



HAL
open science

Development of a light tunable differentiation system for the creation and control of microbial consortia in *S. cerevisiae*, its single cell characterization for development of predictive models, and application to heterologous expression

Chetan Aditya

► To cite this version:

Chetan Aditya. Development of a light tunable differentiation system for the creation and control of microbial consortia in *S. cerevisiae*, its single cell characterization for development of predictive models, and application to heterologous expression. Quantitative Methods [q-bio.QM]. Université de Paris, 2021. English. NNT: . tel-03612653

HAL Id: tel-03612653

<https://inria.hal.science/tel-03612653>

Submitted on 18 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Université de Paris

Frontières de l'Innovation en Recherche et Education (ED °474)

Laboratoire d'accueil : Méthodes expérimentales et computationnelles pour
la modélisation des processus cellulaires (InBio)
(Institut Pasteur & Inria-Paris, Paris, France)

DEVELOPMENT OF A LIGHT TUNABLE
DIFFERENTIATION SYSTEM FOR THE CREATION AND
CONTROL OF MICROBIAL CONSORTIA IN *S.*
CEREVISIAE, ITS SINGLE CELL CHARACTERIZATION
FOR DEVELOPMENT OF PREDICTIVE MODELS, AND
APPLICATION TO HETEROLOGOUS EXPRESSION

Par **Chetan ADITYA**

Thèse de doctorat de BIOTECHNOLOGIE ET INGENIERIES DES
BIOSYSTEMES MOLECULAIRES (cybergénétique)

Dirigée par Gregory BATT

Présentée et soutenue publiquement à Paris le 1 Octobre 2021

Devant un jury composé de :

Gregory BATT , DR, Inria-Paris & Institut Pasteur	Directeur
Barbara DI VENTURA , Professor, Universität Freiburg	Examinatrice
Sébastien LEON , DR, Institut Jacques Monod	Examinateur
Leonardo RIOS-SOLIS , Assistant Professor, University of Edinburgh	Rapporteur
Delphine ROPERS , DR, Inria-Grenoble	Examinatrice
Jakob RUESS , CR, Inria-Paris & Institut Pasteur	Invité
Gaël YVERT , DR, ENS-Lyon	Rapporteur



Except where otherwise noted, this is work licensed under
<https://creativecommons.org/licenses/by-nc-nd/3.0/fr/>

اپنی لہر ہے اپنا روگ
دریا ہوں اور پیاسا ہوں
ناصر کاظمی

My wave is my disease
I am a river and I am thirsty.
Nasir Kazmi

Développement d'un système de différenciation modulable par la lumière pour la création et le contrôle de consortiums microbiens chez *S. cerevisiae*, sa caractérisation en cellule unique pour le développement de modèles prédictifs, et son utilisation pour l'expression hétérologue

RESUME

Les consortiums microbiens artificiels cherchent à exploiter la division du travail pour optimiser des fonctions et possèdent un immense potentiel pour la bioproduction. Les approches de co-culture, le mode préférentiel pour générer des consortiums, restent limitées dans leur capacité à donner naissance à des consortiums stables ayant des compositions précisément ajustées. J'ai développé ici un système de différenciation artificielle dans la levure boulanger capable de générer à partir d'une seule souche des consortiums microbiens stables avec des fonctionnalités choisies et ayant une composition définie par l'utilisateur dans l'espace et dans le temps, grâce à une modification génétique pilotée par optogénétique. Grâce à une dynamique rapide, reproductible et ajustable par la lumière, mon système permet un contrôle dynamique de la composition des consortiums dans des cultures continues pendant de longues périodes. Je démontre également que notre système peut être étendu de manière simple pour donner naissance à des consortiums avec de multiples sous-populations. Cette stratégie de différenciation artificielle établit un nouveau paradigme pour la création de consortiums microbiens complexes qui sont simples à mettre en œuvre, contrôlables avec précision et polyvalents à utiliser.

En plus de cela, j'ai caractérisé le système au niveau de la cellule unique dans différents contextes en changeant la structure du bruit du facteur de transcription optogénétique qui induit la différenciation. J'ai découvert que le changement de la structure du bruit introduisait un couplage complexe entre les niveaux de la population de cellule et des cellules individuelles, qui ne peut être prédit par un simple modèle d'équations différentielles ordinaires. L'utilisation d'un modèle stochastique bien caractérisé a permis de rétablir la prévisibilité.

Enfin, j'ai couplé le système de différenciation avec un system d'arrêt de croissance et de bioproduction de sorte que les cellules différenciées arrêtent de croître et commencent à produire une protéine d'intérêt. J'ai comparé l'efficacité de l'approche basée sur la différenciation avec des équivalents constitutifs et inductibles. J'ai constaté que la production n'était pas monotone par rapport à la fraction de différenciation mais qu'elle pouvait surpasser l'expression induite par un promoteur constitutif fort.

Mots clefs : Biologie synthétique, biotechnologie, cybergénétique, théorie de contrôle, consortium microbien, modélisation des processus cellulaires, biologie computationnelle, système bimodale, optogénétique, machine d'état basé sur une recombinase contrôlable de manière externe

LONG RESUME

Les consortiums microbiens sont courants dans la nature et coexistent ensemble en exploitant le principe de la division du travail. La construction de consortiums microbiens artificiels est une entreprise séduisante et fréquemment engagée, dont cependant, le contrôle dynamique reste terriblement limité. Les approches de la biologie synthétique utilisées pour générer des consortiums microbiens partagent un thème commun, celui de la communication cellulaire par signalisation avec de petites molécules afin de permettre des interactions spécifiques entre les membres du consortium et ainsi renforcer sa stabilité. Bien que ces approches possèdent certaines qualités comme par exemple le contrôle autonome intégré qui résulte en un comportement programmé qui, du moins en théorie, devrait marcher indépendamment du contexte, elles présentent plusieurs limitations.

- Avant toute chose, la majorité de ces approches ont une dynamique communautaire intrinsèque au système qui nécessite des changements importants de la topologie du circuit à moduler.
- Le peu de solutions disponibles quand cette dynamique peut être modulée par des motifs inductibles et que le mode d'induction est invariablement chimique, empêchent un contrôle dynamique et par conséquent, aucune de ces approches ne démontre un contrôle dynamique de la composition du consortium ou même un contrôle sur le long-terme de la composition à l'équilibre.
- La plupart des solutions proposées restent dépendantes, et ce de façon critique, de la densité cellulaire des cultures et ainsi, se comportent de manière prédictible seulement en conditions calibrées avec soin.
- Toutes les approches présentent des circuits génétiques complexes qui sont un poids pour l'organisme hôte et nécessitent souvent que le programme moléculaire soit distribué parmi deux populations distinctes, si ce n'est plus, ce qui aggrave le risque de défaillance du circuit dû à des mutations dans l'une des espèces constituantes. Cela rend également bancal le passage à plus grande échelle de ces consortiums pour effectuer des tâches complexes.
- Dans le contexte de la bioproduction, des machineries aussi pesantes utilisées seulement pour maintenir un consortium microbien sont sous-optimales puisqu'elles usent des ressources qui auraient pu être employées afin d'augmenter la production.
- Les approches utilisant la mort cellulaire sont particulièrement inappropriées pour la bioproduction car beaucoup de la biomasse est gâchée par la lyse des cellules.

Au cours de ma thèse, je me suis attaqué à ce problème en concevant un système de différenciation inductible par la lumière chez la levure et ai trouvé qu'il était capable de générer un consortium microbien ayant une composition définie par l'utilisateur dans l'espace et le temps. Le système de différenciation repose sur l'expression inductible par optogénétique de la Cre-recombinase par le facteur de transcription EL222. En présence d'une cassette de recombinaison, ce système me permet de passer de l'expression d'une protéine à l'autre dans la population. J'ai caractérisé le système dans des cultures liquides à petite échelle (avec des cellules poussant dans une chambre microfluidique), des cultures liquides à plus grande échelle (en lot et en continu) et des cultures solides à court terme (mono-couche sur lamelle μ lbid) et l'ai trouvé fonctionnel malgré les changements contextuels. Ces expériences de caractérisation m'ont aidé à établir que le système possède plusieurs qualités désirables comme une faible activité basale dans le noir, une efficacité sans-précédent en présence de lumière chez la levure, une réponse graduée à des stimuli variés, et une dynamique rapide, reproductible et ajustable. Ces propriétés m'ont permis de développer un modèle prédictif qui était utilisé dans le cadre d'une commande prédictive pour contrôler la composition du consortium dans le temps. La réponse graduée du système ainsi qu'en particulier, l'absence d'hystérésis, ont été cruciales pour un contrôle dynamique de la composition du consortium sur de longues périodes de temps (3-4 jours). La grande efficacité et la fuite limitée du système m'ont permis de générer un consortium microbien structuré spatialement simplement par illumination d'un motif qui, couplé à une puce DMD, a permis le contrôle de la différenciation en cellule unique. Ceci fut exploité pour démontrer le contrôle des lignages dans des colonies de levures poussant dans une chambre microfluidique. De plus, l'inclusion de plusieurs cassettes de recombinaison dans la même souche m'ont permis de générer des consortiums complexes contenant plusieurs espèces. J'ai démontré que le nombre et la prévalence des espèces dans ces consortiums pouvaient être contrôlés en implémentant des programmes de différenciation distincts. Ceci a pu être réalisé grâce à un ajustement de la longueur relative des cassettes de recombinaison respectives tandis que la longueur absolue a permis de contrôler le taux de différenciation global.

Le système a ensuite été utilisé pour sonder le compromis entre production et croissance pendant l'expression hétérologue en opération continue, en l'employant pour découpler la croissance de la production. Des compromis entre production et croissance apparaissent en conséquence du fardeau métabolique généré par l'expression hétérologue et posent un défi majeur à la bioproduction continue, dû à l'émergence de mutants non-producteurs qui entraînent une diminution du rendement avec le temps. Afin d'éviter ces compromis, j'ai couplé le système de différenciation avec un arrêt de croissance de manière à ce que les cellules arrêtent de croître (en surexprimant FAR1M) après différenciation (GAuDI en anglais) et commencent à produire

une protéine d'intérêt (mScarlet contrôlée par le puissant facteur de transcription orthogonal ATAF1). Le couplage n'était pas simple et a requis que je détecte et résolve des problèmes dans des aspects variés du système. Après avoir fait des ajustements significatifs, j'ai été capable de créer une lignée, GAuDi02, qui arrête de croître après induction par la lumière et commence à produire une protéine d'intérêt. Les cellules se sont arrêtées, formant de larges cellules sans bourgeonnement, et ont continué d'augmenter en taille jusqu'à ce qu'elles meurent. Le circuit, cependant, n'était pas stable et les cellules différenciées ont perdu leur phénotype d'arrêt de croissance après 15-20h d'arrêt complet. J'ai établi par séquençage ciblé du génome et avec des expériences additionnelles que cet échappement était associé à une perte de la cassette transcriptionnelle exprimant FAR1M. Néanmoins, j'ai développé un modèle EDO dédié pour être capable de prédire la dynamique de la population et l'ai ajusté aux données expérimentales dynamiques. Le modèle EDO était suffisant pour prédire la dynamique de la population pour divers apports de lumière. Nous avons déployé le modèle EDO dans un cadre MPC pour contrôler le niveau de différenciation de la population et, par extension, le taux de croissance de la culture. Nous avons été capables de maintenir des niveaux stables de fraction différenciée dans des cultures continues de GAuDi02 et autres lignées associées. En parallèle, j'ai développé des plateformes de référence constitutive et inductible d'expression hétérologue pour comparer les performances du système GAuDi. J'ai trouvé qu'une seule cible contrôlée par optoATAF1 générerait jusqu'à 10 fois plus de fluorescence cellulaire moyenne par rapport à la référence pTDH3. De manière curieuse, l'ajout d'une autre copie du même gène n'a pas résulté en une augmentation détectable des niveaux de fluorescence, cependant, l'ajout d'un autre gène cible a augmenté l'expression plus de 20 fois par rapport à pTDH3. J'ai également trouvé que le taux de croissance diminuait en conséquence de l'expression d'ATAF1 ($32\% \pm 2\%$). Et étrangement, cette diminution n'était pas dépendante des niveaux d'expression hétérologue, ce qui suggère que les cellules pouvaient supporter des niveaux d'expression de pTDH3 20 fois plus forts sans affecter le taux de croissance. Finalement, j'ai trouvé un comportement non-monotone dans deux versions cibles de optoATAF1 par rapport à l'induction. J'ai caractérisé la production de la protéine d'intérêt avec optoATAF1 totalement inductible et les lignées GAuDi maintenues à différents niveaux de différenciation dans des cultures continues. J'ai comparé la performance à pTDH3 en utilisant des indicateurs standards (titre, productivité et rendement). J'ai trouvé que GAuDi02 produisait jusqu'à deux fois plus de protéine que la référence pTDH3 mais n'était pas génétiquement stable et la productivité diminuait avec le temps pour des objectifs élevés de fraction de différenciation. GAuDi04, qui montrait un défaut de croissance moins fort, se prêtait davantage à un contrôle au long terme et s'est montré meilleur que la référence pTDH3 tout du long. Cependant, le système de différenciation s'est montré, au mieux,

aussi bon que le système d'induction. Le système d'induction avec copie unique a maintenu une productivité 4 fois supérieure, était stable génétiquement et, en général, a produit 5 fois plus de protéine par ml de milieu usé, comparé à l'expression constitutive. Il est intéressant de remarquer que, malgré un niveau deux fois plus élevé de fluorescence cellulaire, les deux copies d'optoATAF1 ont réalisé de moins bonnes performances que la copie unique. Finalement, j'ai trouvé que le rendement et la productivité ne pouvaient pas être maximisés en même temps, ce qui en est la première démonstration expérimentale en production continue chez la levure. De manière globale, je suis obligé de conclure que notre approche avec un fort arrêt de croissance n'est pas faisable pour l'expression hétérologue dans des cultures continues mais l'introduction d'un défaut de croissance moins prononcé semble plus prometteur.

Aucune des constructions décrites jusqu'ici n'a montré d'effets de sélection et elles étaient exemptes d'hystérésis, un modèle EDO simple était suffisant pour capturer la dynamique et permettre le contrôle dynamique de la composition du consortium. Néanmoins, je me suis demandé si d'autres instanciations pouvaient nécessiter la caractérisation et le développement de modèles plus nuancés qui seraient capables de capturer les effets de sélection et l'hystérésis en résultant. A cette fin, j'ai créé une lignée où les niveaux d'EL222 et son activité sont directement observables en plus de contenir la cassette de recombinaison à deux couleurs (tétra-rapporteur). Plus spécifiquement, j'ai créé trois versions du tétra-rapporteur (intégré, centromérique et 2-micron) et les ai caractérisées en cultures continues. J'ai trouvé que le comportement du système pour la lignée intégrée était similaire au système originel de différenciation bien que son efficacité soit réduite. Ces résultats impliquaient qu'un simple modèle EDO était suffisant pour capturer la dynamique du système. Cependant, pour les versions plasmides, cela n'était pas le cas et nécessitait le développement d'un modèle stochastique du système de différenciation en plus d'un modèle stochastique des fluctuations du nombre de copie de plasmide. Pour ce dernier, nous avons utilisé les données de caractérisation de la lignée intégrée et ceci nous a permis de modéliser sa dynamique de manière quantitative. Afin d'étendre ce modèle aux versions plasmides, nous avons dû d'abord développer un modèle des fluctuations du nombre de copie de plasmide dans la population. Nous avons utilisé les données de caractérisation des lignées plasmides pour paramétrer le modèle. Nous avons trouvé que ce modèle prédisait de manière satisfaisante les conséquences des fluctuations de plasmide. Par conséquent, nous avons construit un modèle hybride pour prédire la dynamique de la population de la lignée 2-micron. Nous avons trouvé que le modèle hybride rendait des prédictions particulièrement fidèles pour la dynamique de la fraction différenciée en plus d'être quantitativement en accord avec la distribution stationnaire d'EL222 dans la population. De plus, le modèle était capable de capturer la dynamique de la distribution d'EL222 dans des cellules

non-différenciées. Enfin, le modèle a prédit qu'il était possible d'utiliser la lumière de sorte à maintenir la population non-différenciée avec de bas niveaux d'EL222 (expression constitutive), et, par extension, le nombre de copie de plasmide, comparé à la population totale et ce en enlevant de manière spécifique et continue les cellules via différenciation. Nous avons trouvé que les données étaient en accord qualitatif avec les prédictions du modèle, démontrant ainsi que, aussi peu intuitif que cela puisse paraître, il est possible de contrôler l'expression constitutive.

Dans le chapitre 3, je décris la construction, caractérisation, et modélisation d'un système de différenciation optogénétique chez la levure qui permet la création d'un consortium microbien sur-mesure avec fonctions définies par l'utilisateur et une dynamique prédictible, rendant ainsi sa composition contrôlable dans l'espace et le temps, indépendamment de la densité cellulaire de la culture. Ceci constitue une contribution majeure de la thèse.

Dans le chapitre 4, j'ai entrepris une comparaison systématique des différents modes d'expression hétérologue en opération continue. Ainsi, deux nouvelles plateformes d'expression hétérologue furent développées et caractérisées, une plateforme inductible de manière optogénétique (optoATAF1) et une plateforme basée sur la différenciation (GAuDi), qui ont toutes deux surpassé l'expression constitutive basée sur pTDH3.

Le chapitre 5 détaille l'application du système de différenciation afin de comprendre les interactions entre les mécanismes au niveau de la cellule unique et au niveau de la population qui façonnent le comportement de la population. Pour atteindre une compréhension mécanistique, il a été nécessaire de construire et caractériser un tétra-rapporteur dans trois contextes génétiques différents. Les modèles stochastiques décrivant le système de différenciation et la dynamique du plasmide, quand ils sont utilisés ensemble, ont permis la prédiction sans paramètre de la dynamique de la population provenant du tétra-rapporteur et le contrôle de l'expression constitutive du gène (nombre de copie de plasmide) au niveau de la population.

Development of a light tunable differentiation system for the creation and control of microbial consortia in *S. cerevisiae*, its single cell characterization for development of predictive models, and application to heterologous expression

ABSTRACT

Artificial microbial consortia seek to leverage division-of-labour to optimize function and possess immense potential for bioproduction. Co-culturing approaches, the preferred mode of generating a consortium, remain limited in their ability to give rise to stable consortia having finely tuned compositions. Here, I developed an artificial differentiation system in budding yeast capable of generating stable microbial consortia with custom functionalities from a single strain at user-defined composition in space and in time based on optogenetically-driven genetic rewiring. Owing to fast, reproducible, and light-tunable dynamics, my system enables dynamic control of consortia composition in continuous cultures for extended periods independently of the cell density. I further demonstrate that our system can be extended in a straightforward manner to give rise to consortia with multiple subpopulations. This artificial differentiation strategy establishes a novel paradigm for the creation of complex microbial consortia that are simple to implement, precisely controllable, and versatile to use.

In addition to this, I characterized the system at the single cell level in different genetic contexts by changing the noise structure of the optogenetic transcription factor that drives differentiation. I found that changing the noise structure introduced complex coupling between the population and the single cell level, which cannot be predicted by a simple population model. A stochastic model of differentiation composed in a stochastic model of plasmid fluctuations not only restored predictability, but revealed mechanistic insights into the functioning of the system. The latter was exploited to demonstrate control of expression of a constitutively expressed gene (proxy for plasmid copy number).

Lastly, I coupled the differentiation system with a growth arrest and production module such that differentiated cells stop growing and start producing a protein of interest. Growth arrest was effected via hijacking of the mating pheromone pathway and production was carried out by an orthogonal transcription factor. I developed a light inducible reference to assess the increase in production upon growth arrest. Comparing the efficiency of the differentiation-based approach with constitutive and inducible counterparts, I found that production was non-monotonic with respect to differentiation fraction and could outcompete constitutive expression. However, production did not increase upon growth arrest.

Keywords : Synthetic biology, biotechnology, cybergenetics, control theory, microbial consortia, modelling of biological processes, computational biology, bimodal systems, optogenetics, externally controllable recombinase based state machine

TABLE OF CONTENTS

Résumé	5
Abstract	11
Table of Contents	13
Preface	18
Acknowledgements.....	19
Chapter 1: Introduction	25
Historical perspective	25
From microbes to microbial communities to microbiomes.....	27
Social interactions within a community.....	31
Engineering microbial consortia.....	33
Challenges / Limitations	42
Organization of the thesis.....	44
References.....	47
Chapter 2 : State of the art	53
Cybergenetics.....	53
Background.....	53
Population level control.....	57
Single cell control	59
Optogenetic control in multicellular systems	61
Growth production tradeoffs in bioproduction.....	63
Background.....	63
Growth production tradeoffs.....	65
References.....	69
Chapter 3: The differentiation system	75
Introduction.....	75
Problem statement	76
Strategy	77
Results.....	78
Construction of an optogenetic differentiation system and its functional characterization	78
Characterization of differentiation dynamics under the microscope	85
Imprinting patterns for spatial control of population behaviour.....	86
Detailed characterization of the Dual reporter in continuous cultures.....	88

Two reactor MPC control.....	93
Extension to multi species consortia	97
Discussion	99
<i>On efficiency</i>	99
<i>On dynamic control of microbial consortia</i>	100
<i>On opting against existing bimodal systems</i>	101
<i>On differentiation system as a pathway switch for population level metabolic engineering</i>	101
<i>On multi-species consortia</i>	102
<i>On differentiation system as a tool to study complex biological systems</i>	103
References.....	104
Chapter 4: Application to heterologous expression.....	109
Introduction.....	109
Problem statement	110
Strategy	110
Results.....	111
Construction and characterization of constitutive and inducible expression platforms.....	111
optoATAF1 driven FAR1M mediated growth arrest.....	116
Coupling differentiation with growth arrest.....	117
Construction of growth arrest upon Differentiation strains that produce a POI (GAuDi)	121
Development of a predictive model for GAuDi02	131
Single strain single vessel control of population composition.....	135
Assessment of production in continuous mode	137
Discussion	140
<i>On the model</i>	141
<i>On escape</i>	142
<i>On cell death</i>	142
<i>On absence of production increase upon growth arrest</i>	143
<i>On non-monotonic production dynamics of optoATAF1</i>	144
<i>On the apparent optoATAF1 2X transcriptional buffering</i>	144
References.....	146
Chapter 5: Mechanistic interrogations at the single cell level	149
Introduction.....	149
Problem statement	150
Strategy	150
Results.....	151
Integrated tetra reporter	151

Plasmid tetra reporters.....	154
A stochastic model for the differentiation system.....	160
A stochastic model for plasmid copy number dynamics.....	163
Composed model and parameter free prediction of dynamics of the 2-micron tetra reporter.....	169
Control of plasmid copy number.....	173
Discussion	178
<i>On modelling</i>	180
<i>On modelling of the centromeric tetra reporter</i>	180
<i>On plasmid copy number control</i>	181
<i>On plasmid copy numbers</i>	182
References.....	185
Chapter 6 : Materials and methods.....	189
Constructions and cloning	189
Yeast strain manipulation and storage	192
Batch culture.....	194
Experimental setup for continuous cultures.....	194
Pipeline for analysis of flow cytometry data	198
Quantifying differentiation.....	202
Dual reporter	202
GAuDi	203
Multi-species.....	205
Tetra Reporter	206
Model fitting and parameter estimation	210
Dual Reporter.....	210
GAuDi02	210
Tetra reporter and plasmid dynamics.....	210
MPC framework	211
Experimental setup for live cell imaging and pattern formation.....	217
Pipeline for analysis of microscopy data	219
References.....	221
Chapter 7 : Conclusions.....	223
Summary	223
Contributions	226
Shortcomings	227
Perspectives	231
References.....	234

Annex 1 : Single cell control of differentiation	235
Annex 2 : FAR1M mediated growth arrest.....	239
Annex 3 : Curiosities	241
tMFA2 terminator leads to noisy (almost bimodal) expression upon EL222 induction.....	241
A suspicious integrated tetra reporter leads to very fast differentiation	243
Annex 4 & Annex 5 : Submitted manuscripts	246
Annex 6 : Strain genotypes.....	277
Bibliography	281

PREFACE

The work presented here was carried out between February 2018 and June 2021 at the InBio group in Bâtiment Lwoff at Institut Pasteur under the supervision of Gregory Batt. The project was a part of a European consortium COSYBIO and funded by the ERC Horizon 2020 program. It also received funding from Inria under the banner of IPLCOSY and from ANR under the name of CyberCircuits. Gregory approached me with the project when funding for my original project on mistranslation in bacteria with Ivan Matic over at Institut Cochin fell through. At the time, I was working in The Netherlands but I knew Gregory from the days of my Master's and felt that the project sounded cool and challenging. In its original scope, the project was to develop an artificial light inducible differentiation system in yeast that was geared towards optimizing bioproduction by separating growth from production in continuous cultures (what now forms Chapter 4). In addition, the project involved development of a continuous culture platform and deployment of optimal control to maximize bioproduction. For the former, we got lucky because Alan Jacquier and Cosmin Saveanu (Genetics of Macromolecular Interactions, Institut Pasteur) had such a platform lying unutilized which was brought to our attention by a friend from the doctoral school, Lukas Hafner and François Bertaux, a permanent engineer in the group, took the lead on refactoring code and revamping the platform. His efforts culminated in ReacSight. For the optimal control, I was supposed to work together with another PhD student, Elise Weill-Duflos, who unfortunately did not continue her thesis and consequently, due to my elementary mathematical background, we had to drop that aspect. The thesis was written between June 2021 and July 2021. It was a period of intense activity, academic and otherwise. One of the challenges of documenting this work was to decide what to include in the thesis. Primarily because the work is highly interdisciplinary and could have been presented in a number of ways. I settled on the present formulation after discussing the project with a bunch of friends from different walks of life and academia. The intention in writing the thesis as it is written is to provide everyone who reads the thesis with something to think about. This is not to say that I think a lot of people will be reading the thesis.

ACKNOWLEDGEMENTS

My interest in science began at a tender age and was cultivated meticulously by my parents so much so that I never imagined doing anything else. The most outstanding contribution in making me who I am comes from my father who instilled in me a thirst for knowledge and a penchant for questioning. Above all, I cherish the freedom they gave me that transformed into wings for my imagination. Guided by these general principles, my life exemplifies the fluttering of a butterfly in a garden with innumerable flowers. It is a testament to their belief that this butterfly has meandered as far as he has with bearings as eccentric as his. At each juncture, critical or not, I have been buoyed by my parents' unconditional support and sage advice. They have provided me with every single opportunity imaginable to help me realize my esoteric goals. Even if I were to put my gratitude into words, they should fall irremediably short of expressing my feelings but I dedicate my thesis to them, for without their constant motivation and positivity, it would not have been possible to precipitate the PhD. आपका लाख-लाख शुक्रिया कि आपने मुझे ये शब्द लिखने के क़ाबिल बनाया।

I must also confess that another person was crucial for a fruitful outcome of my PhD, especially when things were the hardest, my love and my comrade, Solène (aka sneakySOSSS HERVE). Without her unceasing support and help, I would not be in a position to write these words and possibly been deported back to India. Her selfless love gave me the room to write my thesis full time and she was instrumental in providing critical feedback regarding the thesis manuscript as I was writing it. More generally, when the going got rough, it is through her that I summoned the strength to come to work every single day. And so, I dedicate my thesis to her as well. It will smart like hell, the Solène-shaped hole in my life that you will leave behind.

After arriving in France, my connection with my extended family suffered. I am not as connected to them as I would like to be but I thank them for cheering me up at every possible occasion and to never stop believing in me. In particular, I thank my cousin sisters with whom I grew up, Mangu and Guddu.

I thank my maternal grandfather, who would have been proud of me. I also thank my buddy, Shekhar Rao, who was an influence and a cool uncle, an intergenerational bridge between me and my father. May they rest in peace.

En parlant de familles, je remercie ma famille française, Jacqueline, JF et Loulou (les Faures), qui m'ont adopté et ont pris soin de moi quand j'en avais le plus besoin. Je suis très reconnaissant pour leur soutien (en particulier avec l'administration et la bureaucratie), pour la fabuleuse cuisine faite maison avec l'amour et la tendresse d'une mère et pour m'avoir fourni une myriade de façons de me divertir.

I thank my lab and groupmates from present and past for tolerating me over all these years especially with the red light and aluminum foil everywhere. I thank Gregory for giving me the opportunity to work in his newly founded group and giving me the independence to pursue my own research ideas. Sebas and I have had our fair share of conflicts, but despite all those, I thank him for being my colleague, for all his help in the lab, and I cherish our time taming monkeys together. In the same vein, I thank François for his, infrequent yet valuable, guidance and for directly and indirectly contributing to my python skills and for all his help with the experiments and despite our differences, I hold him in high regard. I thank Virgile for his ever-ready-to-help attitude, for teaching me how to fit models to data, and deep conversations that made staying late in the lab pleasant. Virgile is a stellar colleague. I thank Arthur for sharing his candid thoughts with me and supporting me when things got rough. I thank Steven for being who he is, a chatty bubble of enthusiasm and positivity, and for all his help with the microscope. I thank

Zach for his empathy and his neat microscopy data analysis code. I thank Elise, Achille, Lorenzo, Anđela, Olivier and Davin for engaging conversations about academia and science.

From the group, Jakob Ruess has made the most defining contribution to my PhD. Without Jakob's brutally honest and extremely fast feedback, I would not have been able to write a tight manuscript for my first paper and he basically wrote the second paper all by himself. Above all, I thank him for believing in my abilities even when others questioned them. Working with him during these last 6 months was the most rewarding and enriching time of the PhD, scientifically speaking.

I thank the members of my Thesis Advisory Committee, Gael Yvert and Hyun Youk, for their valuable insights into the work and for providing critical feedback on it. The conversations with them helped me shape the work into a cohesive, consistent and, perhaps most importantly, publishable article.

I thank Didier Mazel and Cosmin Saveanu for scientific advice. I thank Stéphane Descorps Declere for help in professional matters.

I thank Barbara Di Ventura, Sébastien Léon, Leonardo Rios-Solis, Delphine Ropers, and Gaël Yvert for agreeing to be part of the jury on such short notice. I thank Gaël and Leonardo in particular for correcting the thesis and being so accommodating at that.

I also thank Алексáндра Асáновна Элбакя́н for her efforts in making science open and from which this research benefited greatly.

I thank Rogerio Amino for being the most awesome boss I could have hoped for. I arrived in his lab as a clueless chemical engineer during my Bachelors and he taught me most of the things that I know, including microscopy and cytometry (but also mosquito decapitation and opening up mice). He was a constant source of motivation and support. I went to him in times of crisis and his advice always led to results. I thank him again for being so welcoming and cool, for letting me express myself and for moments of great philosophical and scientific exchange. I hope that our interactions will continue over many-many years to come. I thank Vanessa Lagal for her stern tutorship that has served me in good stead and for teaching me mammalian cell culture. I cherish all the evenings we spent trying to troubleshoot the growth of them damn hybridomas.

I thank CRI for giving me a chance, for I was chemical engineer who fell in love with biology and CRI provided the means of turning that into a tangible life path. When I talk of CRI, I think of my Master's batch mates too who have become close friends and who feature profusely in my fondest memories of France.

I also think of Jeanette Nguyen who is singlehandedly the most amazing mentor I have encountered in Europe. Her kindness and generosity are unparalleled and I thank her with all my heart for all the extra work she has had to do on my account. It will not be possible to recount all her help, like waiting in line with me in front of the OFII on a nippy grey February morning or securing funding for me to continue my studies or proofreading my paper or hooking me up with an internship in Amsterdam... there are just too many things to thank her for. I shall try to find ways to return her favours all my life.

I thank my alma mater, BITS Pilani and the Chemical engineering department, for allowing me to pursue my Bachelors thesis in molecular parasitology, which eventually led me to where I am. I thank two professors from BITS Pilani, Paritosh Shukla (organic chemistry) for caressing my interest in chemistry and encouraging me to switch to biology and Hari Nair (modern politics and history) for his thought provoking lectures and for providing new (radical and life-altering) perspectives on the human condition.

I thank Joshi for being a real brother, for becoming a veritable challenger in TT and for making me a part of his musical journey. He has been there for me through thick and thin, from the days of casual insouciance (चाय, सुट्टा और जाँतिई) to periods when I was caught in doldrums. I also thank him for indulging

my obsession with Begum Akhtar and King Crimson. I thank 100m for being an amazing friend. For sharing his house with me when I was homeless and for all the low pH activities that he introduced me to. I remain in his debt. I thank Catherine for being a dear and kind friend, for proofreading my article and for providing general advice regarding diverse matters. It was a pleasure and a privilege to share life experiences over tea and cigarettes with someone of your calibre. I thank Victor for the metal concerts, Oscar for the psy-trance raves, Adrien for delightful picnics, Bilal for his discourses on cultural differences, Yves-Marie Rault Chodankar and Charles (le magnifique) for distracting me with their wry humour and welcoming me with a cold beer after a hard day of work, and Julie Levi-Strauss for the fancy soirées. I thank them all for reminding me of what is good in life. I thank Jennifer Pham and Charlie Kossi for always making me smile and tolerating my oft-abominable French.

I thank my friends from overseas who kept me company from a distance, in particular, I would like to thank M. Lailer, ShittiBoi, Avijit, Cheti, Pappu, Mishra, Khatkar, Moms, Namitha, Bihani, Dariusz, Anastasia and Charles, Giulia, Jacopo, and Amandine for their unconditional support.

Last but not the least, I thank all the smokers in Pasteur who listened to me drone on and on about how difficult life is, over the years. I don't think it is useful to enumerate all (I have been in Pasteur for 6 years) but I point out Jacob Seeler and Patricia Baldacci for counseling me regarding deontological matters, Matthias Wagner for his compassion and for bringing tobacco from Germany, Nir Rozenblum, Pierre Levoe, and Aleksandr Barinov for being part of my amateur chess club, and Yann, Michael, Jota, Regine, and Pierre for all the philosophical conversations about life and politics.

In the rare eventuality that someone reading this thesis should find themselves curious to know more, do not hesitate to write me an email at chetandy@gmail.com

It gives me great pleasure to talk science.

It remains an open global collaborative effort to find the fundamental truths that underpin existence

but needs to become more inclusive and do better in rewarding honesty instead of penalizing it.

We, as scientists, need to foster a culture where self-criticism is not met with rejection and ridicule.

To Gaya, Minakshi, and Solène
and the mutual love that we share.

Chapter 1

INTRODUCTION

“If the first words fail, ten thousand will then not avail.”
Ancient Chinese proverb

Historical perspective ^a

Nowadays, it is common knowledge that microorganisms are omnipresent from most hostile places on earth to inside (and on the surface) of every plant, animal, and human that ever lived (with the possible exception of those with Severe combined immunodeficiency (Sponzilli and Notarangelo 2011)). However, for the major part of recorded history, humans were not privy to this information even though scattered accounts from Roman times (Cato, Varro, and Hooper 1935) and Jain scriptures from ancient India (Dundas 2003) suggest at least some degree of cognizance. The awareness of life beyond what the naked eye could see, did not dawn upon humanity until (Robert Hooke, and, more prolifically,) Antonie van Leeuwenhoek, with microscopes, recorded the presence of tiny “animalcules” in pond water in 1677 (Lane 2015)^b. Owing to the technological improvements in optical microscopy, plethora of animalcules were described in the following century. However, their origins remained mysterious for nearly 200 years. During this period, the Aristotelian paradigm was prevalent that, like all other life, microbes arose spontaneously from abiotic matter and was regarded as the dominant hypothesis to explain the emergence of microbes (Zwier 2018). It was Louis Pasteur who, in 1861, definitively put an end to the spontaneous generation theory through his experiments with the fabled swan-necked flasks (Vallery-Radot 1901). These results allowed Pasteur to prove that microbes were present in the air and needed specific conditions to procreate and, perhaps most memorably, led him to proclaim “*Omne vivum ex vivo*” (“Life only comes from life”) ^c. A contemporary of Louis Pasteur, Robert Koch, established that microbes were the causative agents of a variety of diseases firmly establishing the germ theory of disease. The work done by Louis Pasteur and Robert Koch, widely regarded as the fathers of modern microbiology, culminated in what has come to be known as the golden age of microbiology (Arguelles 2017). However,

^a Being the progeny of two microbiologists and having undertaken the thesis work at Institut Pasteur, it would have been indecent of me to omit this section. However, in including it, I realize that I have simplified the narrative considerably. Those interested in a more nuanced account are referred to Wainwright, 2003.

^b *Ibid.*, 334.

^c *Ibid.*, 340.

the discoveries made during this period focused primarily on microbes that were implicated in human health and disease and did not concern with microbes in natural environments. It was not until the early 20th century, through the works of Martinus Beijerinck and Lourens Baas Beeking, who developed methods for cultivation of natural microbes, that microbiologists confronted microbial diversity present in natural habitats. Sergei Winogradsky, perhaps the first microbiologist to focus solely on microbes outside the context of human health, demonstrated the existence of chemotrophic microbes. His experiments with the eponymous Winogradsky column could be seen as the first study of microbial communities. Through the work of Beijerinck, Beeking, Winogradsky, and others that followed, it came to be known that the disease causing microbes formed but a minuscule minority of the microorganisms that exist in nature (Brock et al. 2003). Dedicated studies describing the diversity of microbes sharing a habitat and their interactions with each other and the environment were first conducted by Robert Hungate and colleagues, who developed a way to culture and observe the microbes that occupy the rumen (primary site of microbial fermentation in ruminants) (reviewed in Hungate, 2013). However, it was suspected that cultivation based techniques revealed a simplified picture as less than 1% of the microbial species found in natural environments could be cultured (Staley and Konopka 1985). Carl Woese and George Fox increased the resolution of such endeavors by developing cultivation independent techniques to measure the phylogenetic differences (classification based on 16S rRNA gene) in microbes inhabiting a shared environment (Woese and Fox 1977) and, in the process, uncovered a new kingdom of life, Archaea. Their efforts culminated in an eventual overhaul of the taxonomic classification of living things (Woese, Kandler, and Wheelis 1990). With the advent of powerful DNA sequencing approaches, which continue to improve in terms of sensitivity and facility (Shendure et al. 2017), the true extent of the diversity in microbial communities that pervades natural environments is ultimately within our grasp. It might be utile, at this point, to provide the reader with an approximate definition of what constitutes a microbial community. A microbial community is loosely defined as a group of phylogenetically distinct microorganisms that inhabit a common environment and possess intra-group interactions. It is useful to note that while, this definition of microbial community has been repudiated or deemed too simplistic (Konopka 2009) because of the problems in clearly delineating a contiguous environment (especially in spatially dynamic configurations), classifying genetic diversity and coming up with a rigorous framework for detecting interactions, for the purposes of this thesis, the simplistic definition would suffice.

From microbes to microbial communities to microbiomes

Microbial communities are omnipresent in nature (Barton and Northup 2011) from air and water to soil to plants and animals. Some have been used by humans for millennia intentionally albeit unknowingly (Dertli and Çon 2017; Melkonian et al. 2019) while those inside our guts have co-evolved with us (Moeller et al. 2016; Ley et al. 2008). More importantly, microbial communities have been found to be of paramount importance for the function of the (micro-) ecosystems they inhabit (Hooper et al. 2005; Strickland and Rousk 2010; Bardgett, Freeman, and Ostle 2008). The importance of microbial communities can be gleaned from the fact that prokaryotes constitute majority of the total nitrogen and phosphorous stored in living organisms, in addition to almost half of the total carbon (Raich and Schlesinger 1992; Galloway et al. 2004). One of the most remarkable features of natural microbial communities is the sheer diversity of constituent species and the wide range of environments they inhabit (Lozupone and Knight 2007; Thompson et al. 2017). Some estimates claim up to a million bacterial species in a pinch of soil (Gans, Wolinsky and Dunbar, 2005^d). Faced with such immense diversity in microbial communities and their crucial role in the functioning of the ecosystem, some have opted for the term microbiome. It must be noted, however, that estimates of diversity in microbiomes vary dramatically due to methodological and technological dissimilarities. Therefore, in 2010, a worldwide consortium on microbiomes by the name of *The Earth Microbiome Project* issued a call for a standardized approach towards assessing microbial diversity (“The Earth Microbiome Project” 2010). A concrete framework for the latter was proposed in a landmark paper that sampled the microbial communities from various parts of the world at an unprecedented scale and catalogued their genetic diversity (Thompson et al., 2017^e). After years of deliberation, yet another consortium of experts (*MicrobiomeSupport*, 2020) agreed upon the following definition for microbiome.

“The microbiome is defined as a characteristic microbial community occupying a reasonably well-defined habitat which has distinct physio-chemical properties. The microbiome not only refers to the microorganisms involved but also encompasses their theatre of activity, which results in the formation of specific ecological niches. The microbiome, which forms a dynamic and interactive micro-ecosystem prone to change in time and scale, is integrated in macro-ecosystems including eukaryotic hosts, and hence crucial for their functioning and health.” (Berg et al. 2020)

^d Other studies estimate the diversity between 400-50000 species in a gram of soil (reviewed in Lozupone and Knight, 2007).

^e Outstanding article that provides a comprehensive database of DNA sequencing data and serves as a reference for future studies. Curious readers are referred to a global initiative with the aim of preserving the microbial diversity (Bello et al. 2018). More information at <https://www.microbiotavault.org/project/>.

Microbiomes represent a trove of information and a holistic understanding of their functioning might hold one of the keys to address the catastrophic environmental changes that are provisioned due to global warming but may also be exploited to usher in a new age of therapeutics.

For the former, soil microbiome is of particular interest because of its role in biogeochemical transformations (Raich and Schlesinger 1992; Galloway et al. 2004), however, a mechanistic understanding of how genetic diversity, especially in the light of functional redundancy (Louca et al. 2018), impacts this role is missing. Furthermore, it is known that the soil microbiome has a major impact on plant physiology and prevalence (Bulgarelli et al. 2013) due to (predominantly) mutualistic associations with the plant in the root microbiome (mycorrhizal (fungi) and rhizobial (bacteria)) (Van Der Heijden, Bardgett and Van Straalen, 2008 ^f). However, an estimate of how anthropic changes (pollution, fertilizer use, depletion of nutrients etc.) affect the nature of these interactions remains elusive. Therefore, a keener understanding of the soil microbiome might lead to deeper insights into the ecological impact of microbial communities. More precisely, it could be helpful to expand the scope of remedial interventions (urgent and long overdue) necessary to counter the global climatic change. For example, such an understanding would enable formulation of tailored microbiomes that increase agricultural productivity (Qiu et al. 2019) and/or encourage forest rejuvenation (Van Der Heijden, Bardgett, and Van Straalen 2008). Custom microbiomes may also be composed for the express purpose of detoxification of contaminated soil (Teng and Chen 2019) especially with respect to persistent organic pollutants. Lastly, a Dutch start-up is developing plant based microbial fuel cells that could provide a source of clean energy (“Plant-E” 2009), and while the intent is noble, the success has been modest (the amount of energy generated at the moment is sufficient to only power LEDs).

There exists a great deal of literature on the potential application of microbial communities for therapeutic purposes (Inda et al. 2019; Bober, Beisel, and Nair 2018; Mimeo, Citorik, and Lu 2016). This is driven by the rapidly unfurling importance of the human microbiome in various aspects human health and disease (Cho and Blaser, 2012; Cryan et al., 2019 ^g). In particular, the gut microbiome has become the epicenter of research due to its role in immunity (Zheng, Liwinski, and Elinav 2020) and metabolism (Tremaroli and Bäckhed 2012) and because there exists ample individual-to-individual variation (Lozupone et al. 2012), which has been associated

^f Fascinating article that provides some semblance of the impact of microbial diversity on the abundance and diversity of plants.

^g This is a mammoth review of all the ways in which the gut microbiome is known to interact with the brain. Curious readers are encouraged to go through the review.

with a host of demographic, environmental and genetic factors (Arumugam et al. 2011; Le Chatelier et al. 2013; Goodrich et al. 2014; Rothschild et al. 2018; Falony et al. 2016). There has been special focus on the connection between the gut microbiome and the brain (Cryan et al. 2019). Correlations have been found between the composition of the gut microbiota and depression (Cruz-Pereira et al. 2020), neurological disorders (Cryan et al. 2020), and social behaviour (Sherwin et al. 2019). Clinical studies assessing microbial community based therapeutics have found moderate success (Bárcena et al. 2019; Tanoue et al. 2019; Davar et al. 2021; Baruch et al. 2021; Tokuhara 2021) but thanks to such studies fecal transplants (Smith, Kelly, and Alm 2014) are now a legitimate form of treatment particularly for chronic *Clostridium difficile* infection (Cammarota, Ianiro, and Gasbarrini 2014) ^h.

Microbial communities have also drawn a lot of interest in recent years in the field of biotechnology. The key realization that has prompted this interest is the potential to exploit division of labour paradigms with microbial consortia, which can be thought of as microbial communities geared towards a useful end. Proof of concept studies have concretely established applications of such consortia in bioproduction (Zhou et al. 2015; Li, Wang, and Zhang 2019) and bioremediation (Zhuang et al. 2014; Zhang et al. 2021).

In the context of bioproduction, microbial consortia possess several advantages over traditional monocultures as functional specialization allows metabolic burden to be shared across different species. Metabolic burden and the ensuing growth production tradeoffs are seen to be one of the major hurdles in expanding the scope of bioproduction (Rugbjerg et al. 2018). Therefore, distributing the burden across different species is expected to optimize the production. Another advantage of using consortia over monocultures is that it gives one the freedom to distribute metabolic flux in a population and this compartmentalization allows yield to be optimized simply through tuning consortia composition, rather than tuning the expression of individual pathways/genes by re-engineering the strain itself (Wu et al. 2015). Moreover, including different species, that are functionally and metabolically distinct, allows the sequestration and/or metabolism of toxic by-products of one species by the other and thereby improving the efficiency of the overall process (LaSarre et al. 2017). It also follows that entirely novel chemical conversions that may not be possible to carry out in a single organism, could be realized with microbial consortia (Zhou et al. 2015; LaSarre et al. 2017; Hom and Murray 2014).

^h An American research consortium, The Human Microbiome Project, was founded in 2007 and carried out research to understand the diversity of the human microbial flora and its impact on health and disease. The consortium published 650 papers between 2007 and 2019. Some of these papers are referred to in the thesis. A complete description of the consortium and their achievements can be found at <http://commonfund.nih.gov/hmp/>.

Zhou et al. (2015) makes for a fascinating case study as they exploit each of the above advantages to optimize the production of oxygenated taxanes by using a consortia of *Escherichia coli* and *Saccharomyces cerevisiae*. Moreover, this study showcases the common hurdles that prevent stable co-cultures. Oxygenated taxanes are a precursor for the anti-cancer drug, Paclitaxel. The motivation for their study came from the fact that the enzyme used for the oxygenation of taxanes inhibits the enzymes involved in the biosynthetic pathway that produces taxadiene (which forms the core of all taxanes). This disfavors the production of oxygenated taxanes from a single organism. The authors first created a strain of *E. coli* (TaxE1) that produced taxadiene and a strain of yeast (TaxS1) that produced oxygenated taxanes if taxadiene was supplied. Monoculture of TaxE1 resulted in moderate levels of taxadiene but no oxygenated taxanes. Monoculture of TaxS1 produced neither taxadiene nor oxygenated taxanes even though the latter was shown to efficiently convert externally supplied taxadiene into oxygenated taxanes (>75%). Upon co-culture of the two in glucose, authors observed that overall taxadiene production decreased by a factor of three, but they were able to obtain some degree of oxygenation (~4%). The authors identified that the decreased production of taxadiene was related to a decrease in the growth rate of *E. coli* which in turn was a consequence of ethanol produced by yeast as waste. In order to address this, they decreased the growth rate of yeast and thereby the ethanol in the culture, by growing the co-culture on xylose. Xylose can be metabolized by *E. coli* but not yeast, thus rendering yeast growth dependent on the acetate byproduct produced by *E. coli* as waste. This turned out to be a mutualistic interaction because acetate inhibits the growth of *E. coli* and thanks to the yeast did not accumulate in the culture. Therefore, the authors were able to stabilize the consortia with mutualism. The stabilized consortia resulted in taxadiene titers that were comparable to ones obtained via monoculture of TaxE1, however, the level of oxygenation remained low. The authors established that this was because of low abundance of yeast (and by extension of the enzyme) and through optimization of yeast inoculum and growth, could obtain oxygenation levels of up to 75%. Somewhat against the grain of this narrative, the same group reported in Biggs et al. (2016) that fine-tuning the expression of the entire pathway in *E. coli* led to 15 fold higher titers of oxygenated taxanes compared to Zhou et al. (2015). However, it needs to be noted that the cell densities achieved in Biggs et al. (2016) were an order of magnitude higher than Zhou et al. (2015). Another study (Walls et al. 2021) that explored the production of oxygenated taxanes in *S. cerevisiae* monocultures also reported two fold higher levels compared to Zhou et al. (2015) but with 5-6 fold higher cell densities.

Notwithstanding these counter examples, microbial consortia hold concrete advantages over monocultures for bioproduction. These advantages are best exemplified by a hypothetical

example. Imagine a consortia of an autotrophic species (that efficiently captures atmospheric carbon dioxide and stores it in small molecules like glucose and converts atmospheric nitrogen into ammonia) with a heterotrophic species (that depends on the autotroph for carbon and nitrogen, and produces a valuable compound). Such a consortia would be extremely desirable, both from an economic and an environmental point of view. This will be possible if we can find the social interactions that would stabilize such a co-culture.

Social interactions within a community

Microbes are rarely found in isolation in nature and therefore several interactions existing between the species of the communities have been described. More often than not, the stability of the community is linked to such interactions. These interactions can be unidirectional or bidirectional and can be grouped into three categories based on the influence on the growth:

- ❖ negative or inhibitory
- ❖ neutral
- ❖ positive or beneficial

Applying combinatorial logic, this leads to a total of 6 distinct types of interactions that can exist between any two members of a microbial community (**Figure 1.1**) (Großkopf and Soyer 2014). Positive interactions are rare and often unidirectional. These could arise when the waste product of one species can be metabolized by another (food chain). These interactions become bidirectional if the waste product inhibits the growth of the species that produced it. However, positive bidirectional interactions are uncommon in natural communities. Neutral interactions are said to occur when neither species has an effect on the other species. It is hard to find truly neutral interactions in nature as the microbes are constantly in competition with each other for space, if not nutrients. However, the competition can be minimized to a degree if species with different metabolic niches are cultured together or if ample nutrients are provided. Negative interactions between species in a community may arise due to a variety of reasons and are perhaps the most common interactions. Such interactions could be bidirectional as is the case for resource driven competition or unidirectional as is the case for predation, parasitism or amensalism.

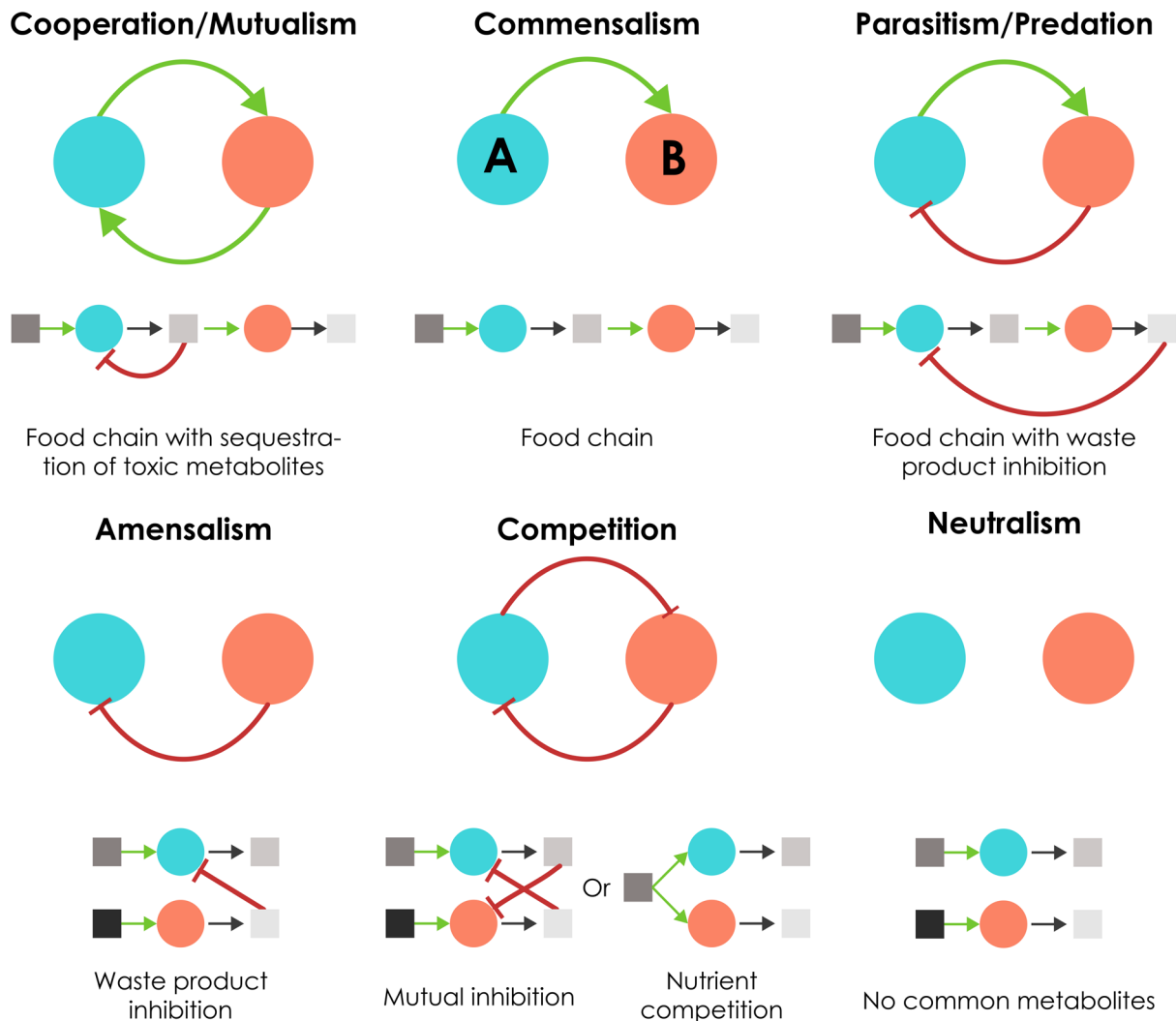


Figure 1.1. Social interactions in microbial communities. Circles represent species and squares represent metabolites. Commensalism entails that Species A promotes the growth of Species B but Species B has no effect on the growth of Species A. This is the case when A produces a metabolite that can be utilized by B. A commensal relationship becomes cooperative (intraspecies) or mutualistic (inter-species) if B also promotes the growth of A. This can happen if the metabolite produced by A inhibits the growth of A. On the other hand, a commensal relationship might become parasitic or predatory if Species B inhibits the growth of Species A. This can be the case if B produces a metabolite that inhibits the growth of A. If Species A did not promote growth of Species B but B still inhibited growth of A, the two would be said to have an amensalistic relationship. Competition is when growth of either species is inhibited by the other. Lastly, a neutral interaction is when neither species influences the growth of the other and could arise due to distinct metabolic needs. Cooperative or mutualistic interactions appear to be rare (Freilich et al. 2011) while competition is the most common one (Foster and Bell 2012). (Figure adapted from Großkopf and Soyer (2014).)

The simplest of natural microbial communities described consist in excess of a hundred species making it difficult to pin down all the complex interactions within the community (Faust and Raes 2012). Therefore, it is desirable to construct simple microbial communities that are directed towards a precise goal or, in other words, engineer microbial consortia.

Engineering microbial consortia

One of the challenges that prevent deployment of consortia on a wide scale is that simply co-culturing different species does not lead to a stable consortium. In the absence of stabilizing interactions, co-culture often lead to extinction of one of the species over time due to competitive exclusion. Co-cultures of different naturally occurring species that possess defined interactions have been shown to lead to stable consortia but curtail our ability to expand the applications of microbial consortia. Stable co-cultures can also be obtained with natural microbes, if their metabolism is well understood, by enforcing cooperation (or mutualism) via metabolic dependencies (Summers et al. 2010) and applied studies have demonstrated that it is possible to engineer mutualistic associations in naturally competitive species. In Summers et al. (2010), the authors found that in certain conditions co-cultures of anaerobic bacteria could develop syntrophic (mutualistic) interactions, highlighting the importance of environment on consortia stability. The same observation was echoed by Hom and Murray (2014), who reported that budding yeast, *S. cerevisiae*, and unicellular green alga, *Chlamydomonas reinhardtii*, could be made to coexist in a culture given appropriate “niche engineering” (fancy way of saying growth conditions). The authors showed that facultative mutualism could be spontaneously induced if nitrite was the only source of nitrogen provided to the co-culture. *C. reinhardtii* is able to reduce nitrite to ammonia, which the *S. cerevisiae* can metabolize (but not the nitrite directly). Moreover, the authors found that the mutualism could be made obligate if access to atmospheric carbon dioxide was limited which was needed by the *C. reinhardtii* and produced by the *S. cerevisiae* from glucose. This means that a co-culture of *S. cerevisiae* and *C. reinhardtii* could grow if nitrite was the sole nitrogen source and atmospheric carbon was made unavailable, but neither could grow on its own, a mutualistic association that has not been found in nature. LaSarre et al. (2017) used a similar metabolic trick to stabilize a microbial consortia between *E. coli* and *Rhodospseudomonas palustris*, a photosynthetic purple non-sulfur bacteria that lives in the soil and can switch between four metabolic modes, fix atmospheric carbon dioxide, fix nitrogen and produce hydrogen (Larimer et al. 2004). *E. coli* grows faster than *R. palustris* and therefore a co-culture of the two would result in competitive exclusion of *R. palustris*. However, since *R. palustris* can fix nitrogen into ammonium ions, the authors conjectured that removing a nitrogen source that the *E. coli* can utilize might result in stable consortia. Conversely, *R. palustris* cannot

metabolise glucose and is dependent on fermentation products of *E. coli* for carbon. Therefore, by supplying nitrogen gas as the only source of nitrogen and glucose the sole source of carbon, the authors demonstrated that a co-culture of *E. coli* and *R. palustris* could grow while both of them in monoculture failed to elicit any growth. It is of interest to note that a consortium of *E. coli* and *R. palustris* could provide a viable platform for bio-hydrogen production. Hammarlund, Chacón and Harcombe (2019) reported a stable microbial consortia when a methionine auxotroph *E. coli* was cultured with *Salmonella enterica* that needed acetate to survive but secreted methionine. *E. coli* grows faster than *S. enterica* but is limited by the amount of methionine present. By regulating the dependency of the faster grower on the slower grower, authors found that it was possible to tune the composition of the consortium. If, however, a shared resource became limiting, stable coexistence could not be maintained. Through mathematical modelling authors argued that changing the concentration of the common resource shifts the relationship between *E. coli* and *S. enterica* from commensal to parasitic and this lies at the heart of the instability. Authors also noted that the composition of the consortium was not stable in time and depended on cell density of the culture (Hammarlund, Chacón, and Harcombe 2019). Interestingly, the same group reported that inclusion of another species to the consortia, another *E. coli* strain that relied on *S. enterica* for arginine instead of methionine, led to a stable three-species consortium independently of the respective growth rates (fitness) of the two *E. coli* strains in monocultures or in co-culture with *S. enterica* (Hammarlund et al. 2021). With the aid of a mathematical model, the authors found that such consortia could be stable if the overall growth rate is limited by *S. enterica* (slowest grower). While these approaches demonstrate our ability to co-culture different species found in nature by imparting mutualistic associations, their composition and function are constrained by the metabolic couplings that stabilise them. Synthetic biology approaches provide more flexibility and control over consortia function and dynamics. Several studies have developed stable microbial consortia by implementing one or more of the defined social interactions listed in the previous section.

Cooperation has been achieved in artificial microbial communities simply by knocking out essential genes to introduce mutually resolvable auxotrophies. For example, Shou, Ram and Villar (2007) generated a synthetic microbial consortia in yeast by engineering two strains that lacked the capability to synthesize adenine and lysine, respectively, while retaining the capacity to produce the other. The authors showed that it was possible to maintain both the strains in a co-culture lacking both adenine and lysine while neither grew on its own. Similarly in *E. coli*, Kerner et al. (2012) used the same approach, but with tryptophan and tyrosine auxotrophic strains, to obtain cooperation driven stabilization of the consortia. In addition, they showed that regulation of the expression of the auxotrophic gene could provide some degree of control of

the consortia composition. However, the proportion did not stabilize in the timeframe of the experiment and appeared to be dependent on the cell density of the culture. The authors attributed this unpredictability to the metabolic burden caused by the circuits used to enable cooperation.

However, the preferred mode for engineering such interactions remains small molecule based signalling, typically mediated by N-Acyl homoserine lactones (AHL) (a class of small molecules that coordinate quorum sensing (QS) in bacteria) primarily because of its key role in establishing and maintaining natural microbial communities (Grandclément et al. 2016; J. Zhou et al. 2016; Montgomery et al. 2013). Admittedly, QS based design is also lucrative because of the autonomous control such circuits afford. The classic study by You et al. (2004) established the thoroughly characterized LuxI/LuxR system (Miller and Bassler 2001) as the system of choice for the studies that followed. The original system from the marine bacterium *Vibrio fischeri*, is composed of two genes LuxI and LuxR and the promoter pLux. LuxI is involved in the synthesis of a freely-diffusible HSL autoinducer (N-(3-oxohexanoyl)-homoserine lactone). LuxR can bind to the autoinducer at a threshold concentration and drive the expression of LuxI constituting a positive feedback loop. At low cell densities, the feedback loop is not activated resulting in basal expression of LuxI and therefore the autoinducer concentration is low. As cell density increases, the autoinducer concentration increases and upon reaching a threshold the feedback loop gets activated leading to full expression of LuxI. Other quorum sensing motifs have been described in bacteria (reviewed in Miller and Bassler, 2001), archaea (reviewed in Montgomery et al., 2013), and fungi (reviewed in Albuquerque and Casadevall, 2012).

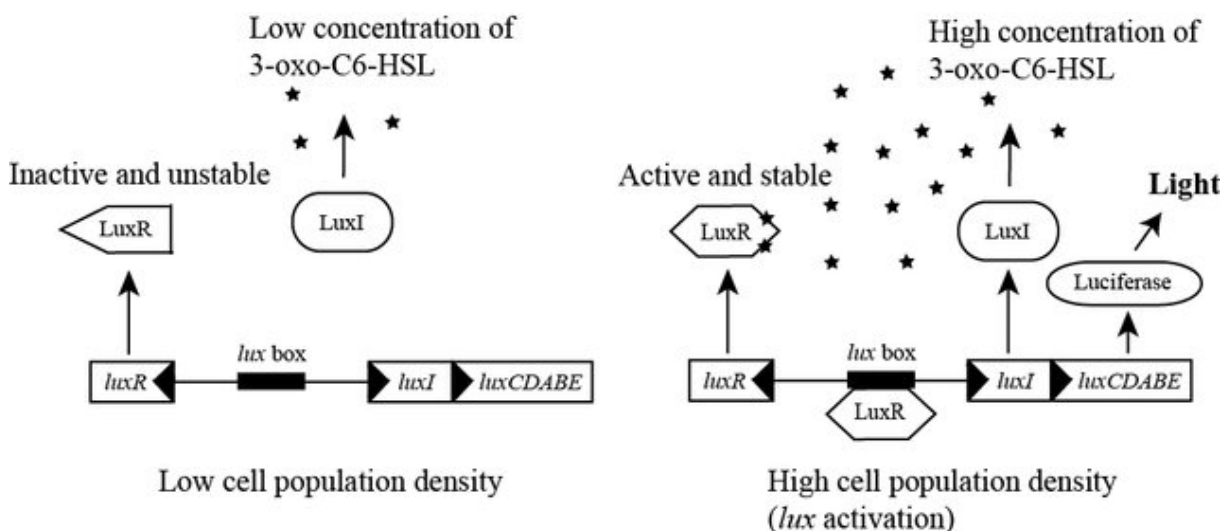


Figure 1.2. The LuxI/LuxR system from *V. fischeri*. (Figure reproduced from Lindell, 2012. All rights belong to Kristoffer Lindell.)

In the natural system, pLux also drives the expression of a gene cluster that codes for luciferase thus making cells generate light upon reaching a critical cell density. You et al. (2004) placed a toxin under the control of pLux and shifted the expression of LuxI and LuxR to an inducible promoter. As the bacteria grow, the autoinducer accumulates in the culture and upon reaching a threshold cell density, the concentration becomes high enough for the LuxR to become activated, thus, triggering the expression of a cytotoxin. However, since the authors shifted the expression of LuxI away from pLux, there is no positive feedback loop meaning once the cell density drops below the threshold, the LuxR becomes inactive thus stopping cell death and resuming growth. After some oscillations, the authors saw cell density stabilizing at the threshold required for activation of LuxR.

Lastly, by altering the pH of the media (and therefore the stability of the AHL molecule), the authors could control the threshold of activation and, in turn, the steady state cell density. This simple and elegant design was used to self-limit the growth of bacteria (You et al. 2004). Quorum sensing based self-killing motif has been exploited in many different ways.

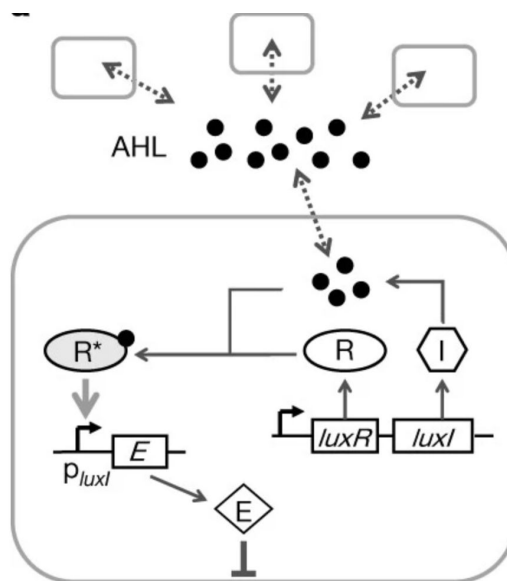


Figure 1.3. Self-limiting growth circuit. (Figure reproduced from You et al. (2004). All rights belong to Nature.)

Balagaddé et al. (2008) for instance, engineered synthetic prey-predator system in *E. coli* by engineering quorum sensing modules to effect differential expression of toxin-antitoxin genes. Briefly, strain A constitutively expressed the toxin gene and required a signaling molecule that strain B produced to express the anti-toxin gene to grow but strain A also produced a signaling molecule that expressed the toxin gene in strain B and therefore growth of strain A resulted in death of strain B. This resulted in oscillatory behaviour where strain B could grow for a certain

amount of time before growth of strain A led to a decrease in growth of strain B which in turn reduced the growth of A making the cycle complete. However, the observed oscillations were slow (period in the order of several days) and noisy. The original cell lysis circuit reported by You et al. (2004) was modified by Din et al. (2016) and Scott et al. (2017) (by shifting back the expression of LuxI to pLux) and used to create populations of bacteria that grew until a threshold cell density and then crashed, leading to oscillatory dynamics. This approach was extended to a three species system by Liao et al. (2019) that cyclically restricted the growth of other strains by toxin-antitoxin-antitoxin systems in addition to a common quorum sensing cell lysis circuit that induced cell lysis above a threshold total cell density (**Figure 1.4, left**). Due to periodic quorum sensing based cell lysis and cyclical repression, the authors were able to achieve fast well-defined sequential oscillations of the three species.

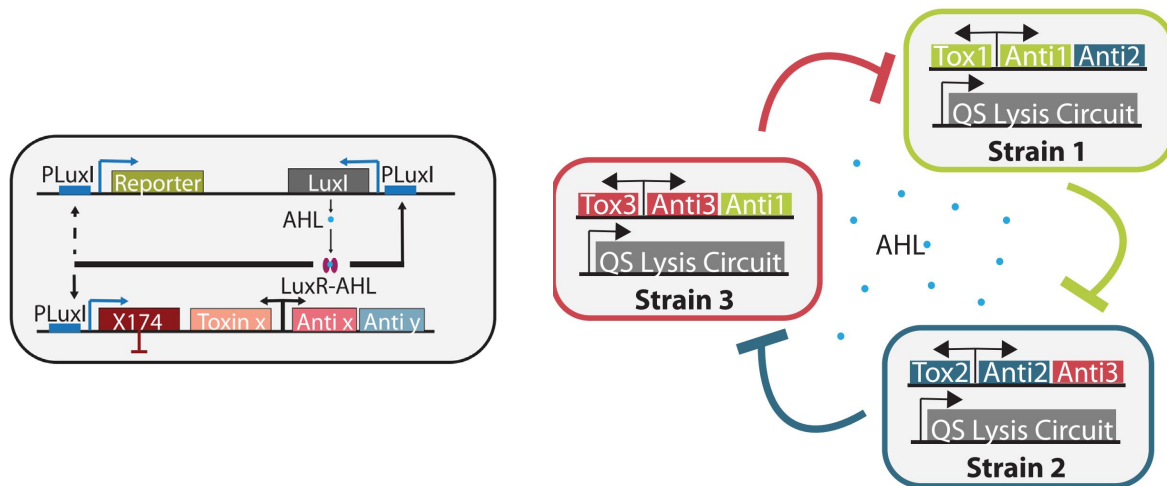


Figure 1.4. Rock, paper, scissors system reported in Liao et al. (**Left**) Coupling of quorum sensing synchronized cell lysis circuit (SLC) and toxin-antitoxin-antitoxin (TA) module. SLC triggers cell density dependent lysis. In the TA module, toxin x and antitoxin x are produced and therefore neutralise each other, however, in addition, antitoxin y is expressed that grants immunity to toxin y (that will be expressed by a second strain) but not against toxin z (that will be expressed by a third strain). (**Right**) Each strain produced its own TA pair while also producing the antitoxin of the following strain. All three strains used the same Lux-AHL quorum-sensing system to drive fluorescent reporter protein expression and self-limiting synchronized lysis. (Figure and text adapted from Liao et al. (2020). All rights belong to Science.)

Along similar lines, an inducible QS system was reported by Miano et al. (2020) that coupled an AHL based system (cell lysis motif) to a precursor of AHL whose concentration can be externally controlled. Using a microfluidic flow cell, the authors demonstrated that their system could be used to maintain stable co-existence of cells that possess the inducible version of the self-killing motif and cells that have the non-inducible version of it. Furthermore, by tuning the concentration of the inducible molecule authors could achieve different dynamics in consortia

composition. However, their results indicated that the composition was highly oscillatory and could not be stably maintained over long time periods (>10h).

These studies noted that the system was not stable and the circuit lost functionality over time suggesting that engineering circuits that result in cell killing/ growth arrest might be constrained by evolutionary timescales (Castle, Grierson, and Gorochofski 2021).

Fedorec et al. (2021) utilized small molecule inducible unidirectional cell killing to stabilize consortia in *E. coli* but, in addition, they introduced a growth defect in the killer strain such that in the absence of cell killing the killer strain would be outcompeted by the WT strain. By co culturing the killer strain and the WT strain at different levels of toxin expression, they were able to maintain stable consortia at distinct compositions. The authors did not demonstrate dynamic control of consortia composition and described the improvements in circuit topology necessary to achieve this goal. Specifically, they constructed an externally (chemically) tunable system coupled to autonomous control via pLux attenuable bacteriocin module described above. The global design principle is that the killer strain expresses the bacteriocin and inhibits the growth of WT strain in the absence of arabinose. Upon addition of arabinose it starts expressing LuxI and therefore it starts producing the AHL autoinducer. As cell density grows, the concentration autoinducer increases and, upon reaching the critical threshold, activates the expression of TetR, which represses the expression of the killer protein, thus, decreasing the inhibition of the WT strain. WT strain, on account of faster growth, outcompetes the killer strain when the bacteriocin is repressed. In this way, the authors hoped to achieve stable maintenance of consortia composition, which could be externally tuned via the addition of arabinose. They used the latter feature to demonstrate that different stable proportions could be maintained by tuning the dilution rate, however, the dynamic range of proportions achievable was limited. Through mathematical modeling, they discovered that the root of this was the leakiness of bacteriocin expression that reduced the region of co-existence of the two strains and they postulated a stronger repressor of the bacteriocin would be crucial for finer control (Fedorec et al. 2021).

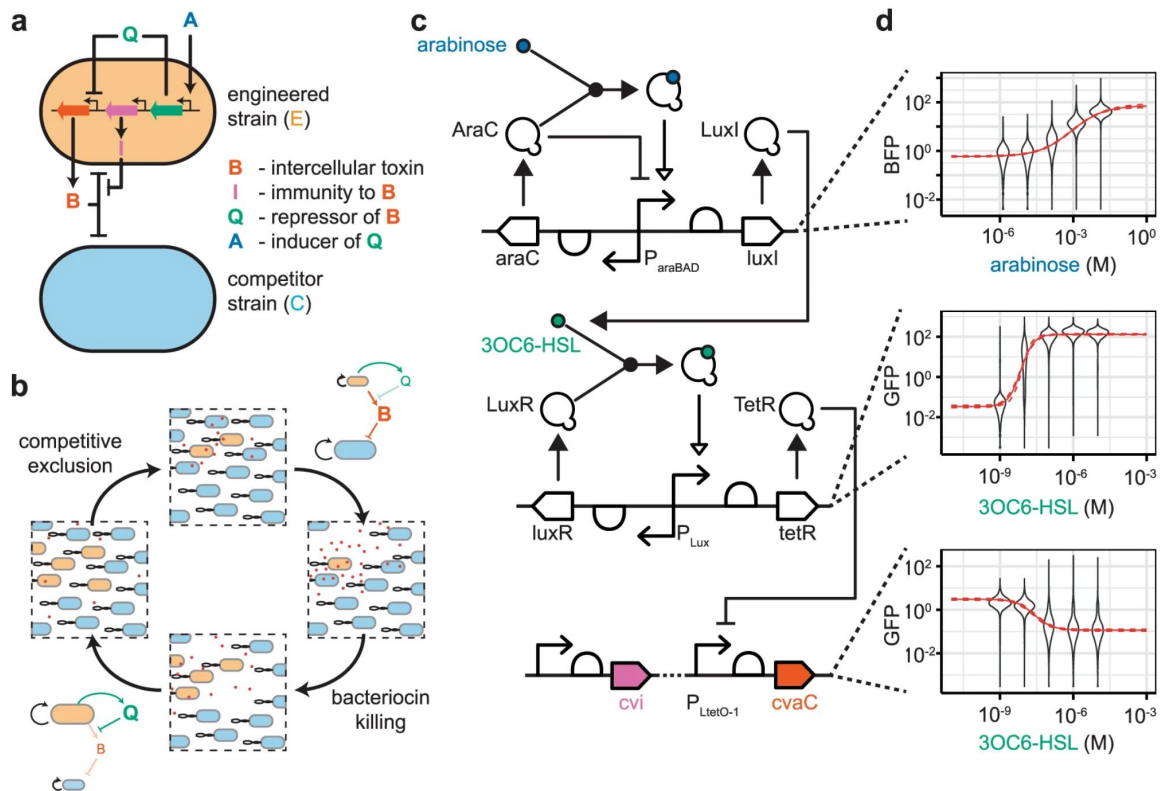


Figure 1.5. Externally tunable quorum sensing attenuable bacteriocin system for single strain control of microbial consortia reported in Fedorec et al.. **a**. Logic of the genetic circuit. **b**. Logic of consortia stabilization. **c**. Genetic circuit. **d**. Readouts of levels of key parts of the circuit with inducer and autoinducer. (Figure and text adapted from Fedorec et al. (2021). All rights belong to Nature.)

The increase in our ability to create stable microbial consortia is perhaps best illustrated by Kong et al. (2018) who developed a modular framework for implementing social interactions in a systematic way by using small signaling molecules to engineer unidirectional or bidirectional relationships between distinct *E. coli* strains. More precisely, they proposed several circuits that reconfigure the pathways of production of nisin and lactococcin A (**Figure 1.6a & b**) and achieved diverse interactions by ascribing differential function to each signaling molecule in different strains (**Figure 1.6c**). Both nisin and lactococcin A are small molecules that are lethal to the cell but can also be used to trigger expression of downstream genes. These toxins can be removed from the cell via the presence of efflux pumps, which grant immunity to the cell expressing them (**Figure 1.6a & b**).

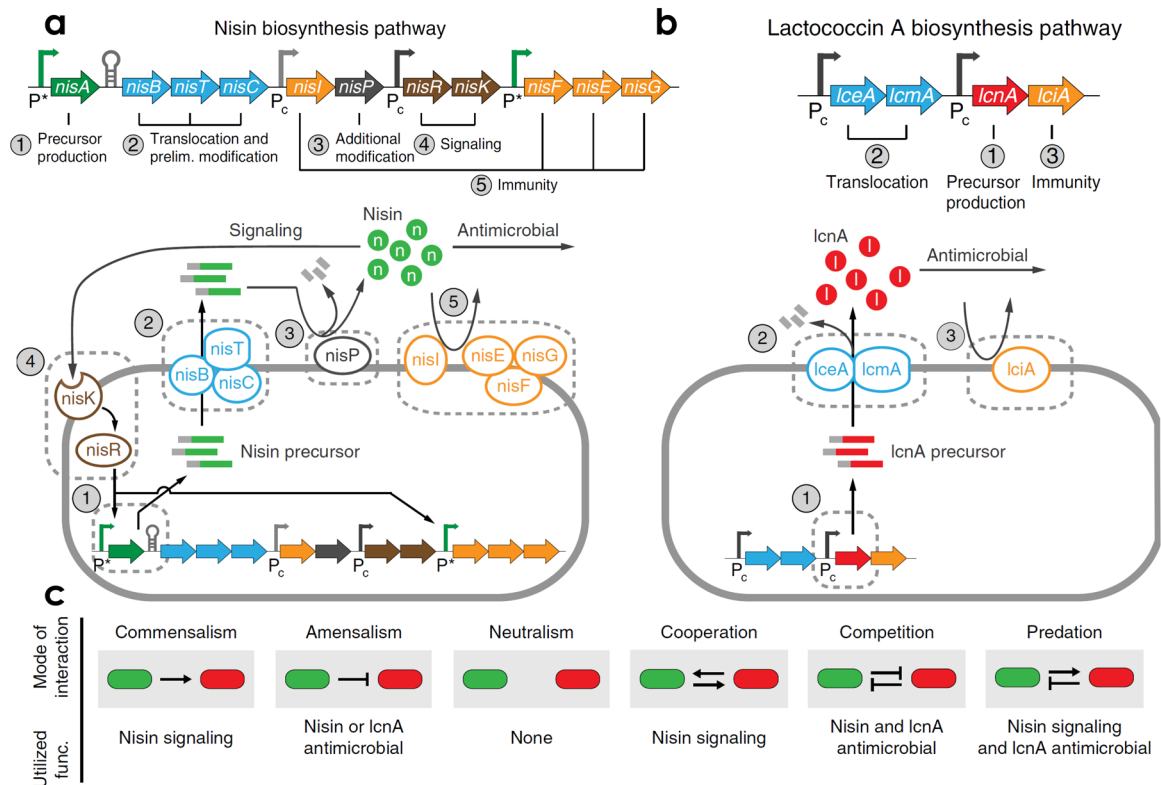


Figure 1.6. Molecular framework for engineering social interaction **a**. Modular organization of the nisin pathway involves five functionally independent modules, including those for precursor production (1), translocation and initial modification (2), additional modification (3), signaling (4) and nisin immunity (5). **b**. Modular organization of the lcnA pathway contains three functionally independent modules, including those for precursor production (1), translocation (2) and lcnA immunity (3). **c**. Design of six two-strain consortia that differentially utilize the signaling and antimicrobial features of the bacteriocins. (Figure and text adapted from Kong et al. (2018). All rights are reserved by Nature.)

For instance, they engineered a cooperative interaction between two strains by distributing the production of nisin across two strains in a context where nisin is a public good (that is used to drive the expression of antibiotic resistance gene in both the strains) and both strains possessed immunity from nisin dependent killing. Strain A produced the precursor of nisin while strain B converted this precursor into nisin. Nisin drove the expression of the TetR gene (**Figure 1.7a**). These two strains could only grow in the presence of each other when the media was complemented with tetracycline because none of them could express the resistance gene on their own. They turned this interaction predatory by shifting the entire biosynthetic pathway for nisin production to Strain A and producing the lcnA from Strain B. Strain B carries immunity against nisin as well as lcnA but depends on Strain A for nisin, which is needed to drive expression of the antibiotic gene. Strain B, in growing, inhibited the growth of Strain A (**Figure 1.7b**). By removing the immunity for nisin from Strain B, they engineered a competitive interaction by driving the

synthesis of bacteriocins and their immunity genes in individual strains, which, in producing their respective bacteriocins, inhibit the growth of the other strain (**Figure 1.7c**).

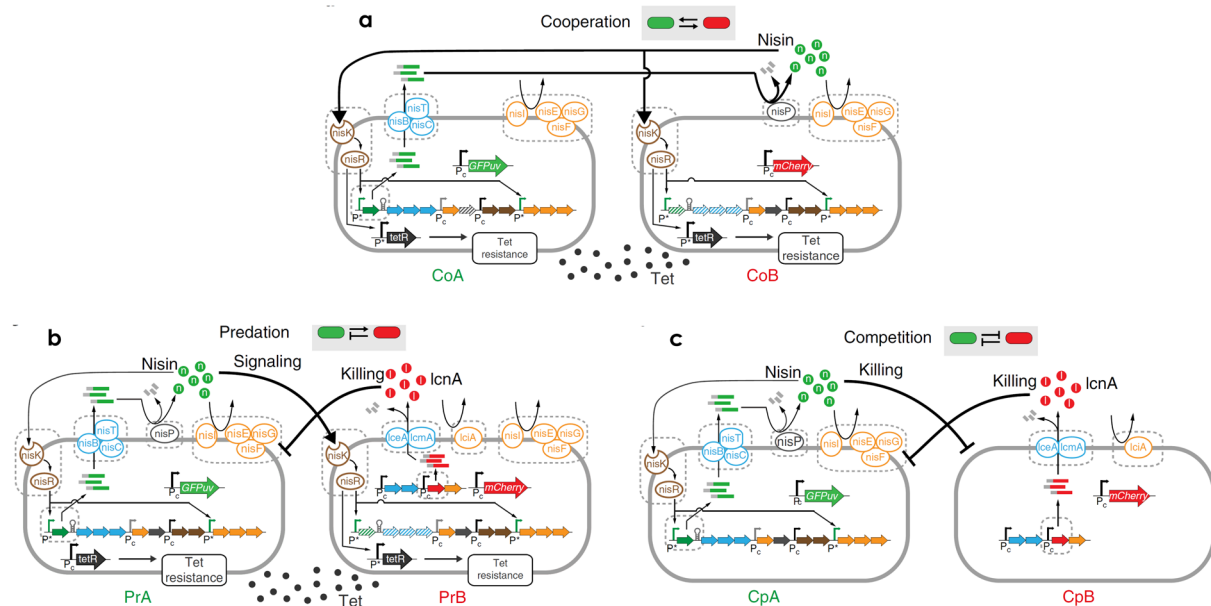


Figure 1.7. Implementations for a. Cooperation b. Predation c. Competition. (Figure and text adapted from Kong et al. (2018). All rights are reserved by Nature.)

Similarly, by reconfiguration of underlying motifs they could scale the number of species in the consortia up to 4. The authors showed some of their circuits allowed them to stably maintain 3 or 4 species. However, similar to previous studies they observed consortia composition changed with cell density and steady state composition (upon reaching stationary phase) relied strongly on initial cell density and composition.

This plug and play approach towards creating diverse microbial consortia using quorum sensing molecules was enriched by the study of Kylilis et al. (2018) who built and characterized a library of acyl-homoserine lactone (AHL)/ transcription factor pairs dubbed AHL receiver devices. The authors developed a software tool to inform the choice of orthogonal AHL receiver devices that could be deployed together without the risk of crosstalk and found 4 orthogonal devices. Three of these were transformed in *E. coli* cells and polyclonal cultures were used to validate that these 3 devices were indeed orthogonal.

These studies demonstrate the capacity to maintain co-cultures of several bacterial subpopulations over extended durations.

Challenges / Limitations

The synthetic biology approaches described in the previous section share a common theme, that is of engineering cell to cell communication via small molecule based signaling to achieve desired interaction between the members of the consortia and thus enforce stability. While such approaches possess certain qualities like inbuilt autonomous control that results in programmed behaviour that, in theory at least, should work independent of the context, they present with several limitations.

- First and foremost, majority of these approaches come with hard wired community dynamics that require extensive alterations to the circuit topology to modulate.
- The small minority of solutions when these dynamics can be modulated via inducible motifs, the mode of induction is invariably chemical, precluding dynamic control in liquid cultures and consequently none of these approaches demonstrate dynamic control of consortia composition or even long-term control of steady state composition.
- Most of the proposed solutions remain critically dependent on the cell density of the cultures and thereby behave predictably only in carefully tuned conditions.
- All approaches feature convoluted genetic circuits that are burdensome for the host organism to carry and often require the molecular program to be distributed across two if not more distinct populations aggravating the risk of circuit failure due to mutations in any one of the constituent species. This also makes scaling up the consortia to perform complex tasks ungainly.
- In the context of bioproduction, such burdensome machinery just to maintain microbial consortia is suboptimal as it takes away resources that could have been employed towards increasing production.
- The approaches featuring cell killing are particularly unsuitable for bioproduction purposes because a lot of biomass is wasted due to cell lysis.

The ubiquitous nature of quorum sensing solutions towards engineering stable consortia make dynamic control a hard problem as the signaling molecules are released by cells, creating de facto a strong dependency of the functioning on growing conditions, and notably on the density of cell cultures, an important aspect for bioproduction applications. For bioproduction and metabolic engineering applications, where precise control of metabolic flux is necessary to optimize yield, dynamic control of consortia composition represents a key challenge to overcome in order to unlock the true potential of artificial microbial consortia. In light of these limitations, an externally controllable differentiation system could be well suited to address this challenge.

Yaoyu Yang, Nemhauser, and Klavins (2019) reported one such system in yeast using a novel bistable switch. They developed an orthogonal transcription factor, ZAVNY, composed of five domains, a zinc finger DNA binding domain (ZDBD), a VP16 activation domain, an auxin degron, a nuclear Localization Signal (NLS), and an enhanced yellow fluorescent protein (EYFP). This transcription factor was introduced in a strain in which ZEV, a β -estradiol inducible transcriptional activator and TIR1^{DM}, an auxin receptor, were constitutively expressed. The authors further introduced a positive feedback loop on ZAVNY such that in the presence of β -estradiol, ZEV localizes to the nucleus and activates the positive feedback loop on ZAVNY triggering its expression. Upon activation, ZAVNY is expressed and can drive its own expression thus eliminating the need of β -estradiol. The system is said to be ON in this state. On the contrary, adding auxin to the system degrades ZAVNY and effectively shuts down the positive feedback loop. In this state, the system is OFF and requires re-addition of β -estradiol to become ON. The authors demonstrated that both the ON and OFF state could be stably maintained and that the system could be flipped by adding the right inducer. As a proof of concept, the authors showed that this bistable switch coupled to antibiotic resistance expression allowed them to control the growth rate of the culture in the presence of antibiotic. I note that the authors did not demonstrate bimodal behaviour but, at least in carefully tuned conditions, such systems can give rise to bimodality (Youk and Lim 2014).

A site-specific recombinase based state machine proposed in Roquet et al. (2016) seems like a more attractive approach towards differentiation as it permits a permanent genetic change. Briefly, site-specific recombinases (SSR) are proteins that recognize specific sequences of DNA (RS) that can, depending on the orientation of these sequences, either excise the DNA between them or invert it (**Figure 1.8**).

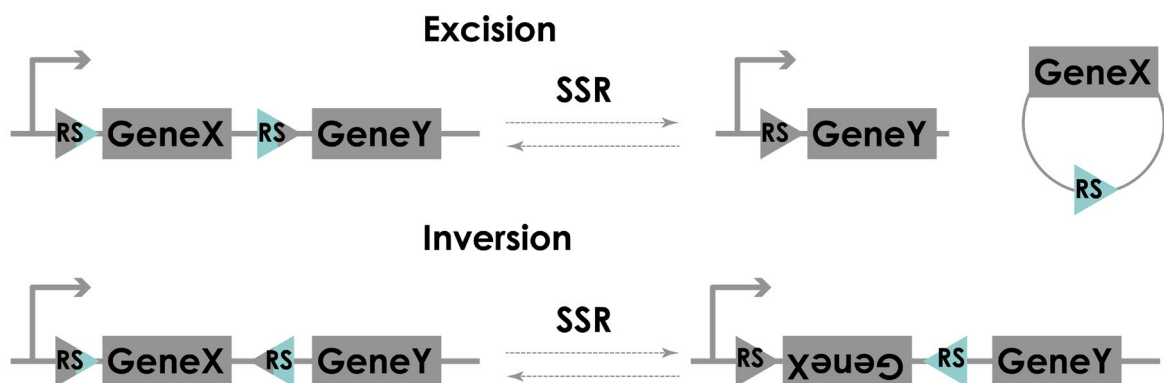


Figure 1.8. SSR based recombination. Presence of SSR results in excision if the RS are oriented in the same direction and inversion if the RS are oriented in the opposite direction to each other. Both processes are

reversible, however, the circular product of excision is lost to dilution if it cannot replicate autonomously (i.e. lacks an origin of replication).

The authors demonstrated, in *E. coli*, that 3 orthogonal chemically inducible recombinases could be used to create a total of 16 distinct states based on the chronological order of the application of inducers. While the authors did not show any functional diversification in this study, they argued that each of these states could potentially be ascribed a distinct function. It is worth remarking that the authors did not report bimodal behaviour meaning system was bistable but incapable of graded response possibly due to inability to remove the inducer fast enough. Guo et al. (2020) reported a scaled down implementation of this approach with two chemically inducible orthogonal recombinases. The authors ascribed distinct metabolic functions to each of these states. More specifically, authors coupled the recombination with genes that regulate the chronological and replicative life span of *E. coli* in addition to key enzymes in butyrate production. The authors demonstrated that changing the state of the culture along production horizon (a pre-established duration during which production takes place) increased the production. Again, the authors did not observe bimodal behaviour with respect to the circuit, however, the authors engineered programmed differences in mother and daughters in order to optimize the production of PLH (poly (lactate-co-3-hydroxybutyrate)) and showed that they could differentiate the population into PLH filled “old” mother cells that died and “rejuvenated” daughter cells for cell growth and production. This was achieved by ensuring asymmetric division of inclusion bodies between mother and daughter cells. Inclusion bodies are known culprits in senescence and aging (Lindner et al. 2008; Nyström and Liu 2014; Steiner 2021). While recombinase based state machines are veritable platforms for programmable cellular differentiation, use of chemical inducers prohibits dynamic control and bimodality (Roquet et al. 2016; Guo et al. 2020).

Organization of the thesis

The introduction extends into Chapter 2 (State of the art), where I briefly introduce the concepts of cybergenetics and growth production tradeoffs that form the background of the work undertaken during the thesis.

The work done during this thesis is organized in three chapters.

In Chapter 3, I address the challenge of dynamic control of consortia composition. We proposed a novel approach of generating a microbial consortia by differentiating an isogenic strain into a distinct cell type by effecting a permanent change. Furthermore, we desired the system to be externally controllable and functionally independent of cell density. Concurrent with these

performance criteria, I designed an artificial differentiation system in yeast that could generate and dynamically control microbial consortia arising out of a single strain. In order to realize this, I used an optogenetic recombination based switch that led to a permanent genetic change upon differentiation. I characterized the system and found it suitable for our task due to the presence of several desirable characteristics like high efficiency, low leakage, graded response to different stimuli, and fast, tunable and predictable response. I demonstrated that the system function is independent of culture density and used it to dynamically control consortia composition.

In Chapter 4, we wondered if the differentiation system could be employed to overcome the traditional limitations of continuous heterologous expression by decoupling growth from production and compartmentalizing the two in different subpopulations. In order to realize this, I designed and characterized a growth arrest module that hijacked the mating factor pathway and, after coupling to the differentiation system, allowed **growth arrest upon differentiation** (GAuDi). To assess the performance of the GAuDi strains, I developed constitutive and inducible reference platforms of heterologous expression. The inducible platform was based on the optogenetic induction of an orthogonal transcription factor (optoATAF1) and led to 10-20 times higher expression when compared to strong constitutive expression. I found that in continuous operation, both GAuDi and optoATAF1 outperformed constitutive expression.

In Chapter 5, we used the differentiation system as a case study to understand the impact of changing the context of the genetic circuit on its functionality and predictability. In order to realize this, I designed a version of the differentiation system that harbored reporters for the levels of optogenetic transcription factor and a readout for its activity. This system was expressed from an integrated copy or from high-copy or low-copy plasmids. The differentiation dynamics of all the strains were characterized at the single cell level for different patterns of light induction. Additionally, for the plasmids strains, the dynamics of plasmid fluctuation were characterized. Differentiation data from the integrated strain and plasmid dynamics from the plasmid carrying strains were then used to parameterize independently developed stochastic models of differentiation and plasmid dynamics, respectively. These models, when composed together, were able to provide parameter-free predictions of population dynamics of the plasmid strain that were in quantitative agreement with the data. Surprisingly, the plasmid version of the differentiation system along with the model allowed us to control constitutive expression and, by extension, plasmid copy number in subpopulations.

Each of these chapters is accompanied by a dedicated discussion pertaining to the specific results.

A summary of the results and contributions achieved during the thesis has been provided at the end. This has been supplemented by enlisting the shortcomings of the system, and the experiments carried out and a general discussion on the potential applications of the system.

In addition to the results presented in Chapters 3-5, I have some scattered results that have been included in the Annexes 1-3.

Annex 1 outlines single cell control of differentiation in a microfluidic chamber using our microscopy platform.

Annex 2 shows preliminary growth rate data with Galactose inducible and light inducible expression of a variant of the downstream effector of the mating pheromone pathway.

Annex 3 is split into two parts, the first part talks about introducing heterogeneity in EL222 driven expression by using tMFA2, a terminator known to increase variability in gene expression, whereas in the second part, I discuss a dubious integrated version of the tetra reporter that has (really) fast recombination dynamics. These dynamics are compared with those obtained with other differentiation constructs.

Finally, in order to facilitate the understanding of the developed systems, I have included the manuscripts of the two manuscripts that have originated from the work done during the thesis.

Annex 4 contains the manuscript of the submitted article based on the work done in Chapter 3 and parts of Chapter 4.

Annex 5 contains the manuscript of the submitted article based on the work done in Chapter 5.

References

- “Advisory Groups - MicrobiomeSupport.” 2020. 2020. <https://www.microbiomesupport.eu/advisory-groups/>.
- Albuquerque, Patricia, and Arturo Casadevall. 2012. “Quorum Sensing in Fungi--a Review.” *Medical Mycology* 50 (4): 337–45.
- Arguelles, Juan-Carlos. 2017. “The Early Days of the Nobel Prize and Golden Age of Microbiology.” *Hektoen International*.
- Arumugam, Manimozhiyan, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel R Mende, Gabriel R Fernandes, et al. 2011. “Enterotypes of the Human Gut Microbiome.” *Nature* 473 (7346): 174–80.
- Bárcena, Clea, Rafael Valdés-Mas, Pablo Mayoral, Cecilia Garabaya, Sylvère Durand, Francisco Rodríguez, María Teresa Fernández-García, et al. 2019. “Healthspan and Lifespan Extension by Fecal Microbiota Transplantation into Progeroid Mice.” *Nature Medicine* 25 (8): 1234–42.
- Bardgett, Richard D, Chris Freeman, and Nicholas J Ostle. 2008. “Microbial Contributions to Climate Change through Carbon Cycle Feedbacks.” *The ISME Journal* 2 (8): 805–14.
- Barton, L L, and D E Northup. 2011. *Microbial Ecology*. Wiley.
- Baruch, Erez N, Ilan Youngster, Guy Ben-Betzalel, Rona Ortenberg, Adi Lahat, Lior Katz, Katerina Adler, et al. 2021. “Fecal Microbiota Transplant Promotes Response in Immunotherapy-Refractory Melanoma Patients.” *Science* 371 (6529): 602–9.
- Bello, Maria G Dominguez, Rob Knight, Jack A Gilbert, and Martin J Blaser. 2018. “Preserving Microbial Diversity.” *Science* 362 (6410): 33–34.
- Berg, Gabriele, Daria Rybakova, Doreen Fischer, Tomislav Cernava, Marie-Christine Champomier Vergès, Trevor Charles, Xiaoyulong Chen, et al. 2020. “Microbiome Definition Re-Visited: Old Concepts and New Challenges.” *Microbiome* 8 (1): 103.
- Bober, Josef R, Chase L Beisel, and Nikhil U Nair. 2018. “Synthetic Biology Approaches to Engineer Probiotics and Members of the Human Microbiota for Biomedical Applications.” *Annual Review of Biomedical Engineering* 20 (1): 277–300.
- Brock, Thomas Dale, Michael T Madigan, John M Martinko, and Jack Parker. 2003. *Brock Biology of Microorganisms*. Upper Saddle River (NJ): Prentice-Hall, 2003.
- Bulgarelli, Davide, Klaus Schlaeppi, Stijn Spaepen, Emiel Ver Loren van Themaat, and Paul Schulze-Lefert. 2013. “Structure and Functions of the Bacterial Microbiota of Plants.” *Annual Review of Plant Biology* 64 (1): 807–38.
- Cammarota, Giovanni, Gianluca Ianiro, and Antonio Gasbarrini. 2014. “Fecal Microbiota Transplantation for the Treatment of Clostridium Difficile Infection: A Systematic Review.” *Journal of Clinical Gastroenterology* 48 (8).
- Castle, Simeon D, Claire S Grierson, and Thomas E Gorochowski. 2021. “Towards an Engineering Theory of Evolution.” *Nature Communications* 12 (1): 1–12.
- Cato, Marcus Porcius, Marcus Terentius Varro, and William D Hooper. 1935. “On Agriculture: With an Engl. Transl. by William Davis Hooper. Rev. by Harrison Boyd Ash.” *Loeb Classical Library*.
- Chatelier, Emmanuelle Le, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, et al. 2013. “Richness of Human Gut Microbiome Correlates with Metabolic Markers.” *Nature* 500 (7464): 541–46.
- Cho, Ilseung, and Martin J Blaser. 2012. “The Human Microbiome: At the Interface of Health and Disease.” *Nature Reviews Genetics* 13 (4): 260–70.
- Cruz-Pereira, Joana S, Kieran Rea, Yvonne M Nolan, Olivia F O’Leary, Timothy G Dinan, and John F Cryan. 2020.

- "Depression's Unholy Trinity: Dysregulated Stress, Immunity, and the Microbiome." *Annual Review of Psychology* 71 (1): 49–78.
- Cryan, John F, Kenneth J O'Riordan, Caitlin S M Cowan, Kiran V Sandhu, Thomaz F S Bastiaanssen, Marcus Boehme, Martin G Codagnone, et al. 2019. "The Microbiota-Gut-Brain Axis." *Physiological Reviews* 99 (4): 1877–2013.
- Cryan, John F, Kenneth J O'Riordan, Kiran Sandhu, Veronica Peterson, and Timothy G Dinan. 2020. "The Gut Microbiome in Neurological Disorders." *The Lancet Neurology* 19 (2): 179–94.
- Davar, Diwakar, Amiran K Dzutsev, John A McCulloch, Richard R Rodrigues, Joe-Marc Chauvin, Robert M Morrison, Richelle N Deblasio, et al. 2021. "Fecal Microbiota Transplant Overcomes Resistance to Anti-PD-1 Therapy in Melanoma Patients." *Science* 371 (6529): 595–602.
- Dertli, Enes, and Ahmet Hilmi Çon. 2017. "Microbial Diversity of Traditional Kefir Grains and Their Role on Kefir Aroma." *LWT-Food Science and Technology* 85: 151–57.
- Dundas, Paul. 2003. *The Jains*. Routledge. Routledge.
- Falony, Gwen, Marie Joossens, Sara Vieira-Silva, Jun Wang, Youssef Darzi, Karoline Faust, Alexander Kurilshikov, et al. 2016. "Population-Level Analysis of Gut Microbiome Variation." *Science* 352 (6285): 560–64.
- Faust, Karoline, and Jeroen Raes. 2012. "Microbial Interactions: From Networks to Models." *Nature Reviews Microbiology* 10 (8): 538–50.
- Fedorec, Alex J H, Behzad D Karkaria, Michael Sulu, and Chris P Barnes. 2021. "Single Strain Control of Microbial Consortia." *Nature Communications* 12 (1): 1–12.
- Foster, Kevin R, and Thomas Bell. 2012. "Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species." *Current Biology* 22 (19): 1845–50.
- Freilich, Shiri, Raphy Zarecki, Omer Eilam, Ella Shtifman Segal, Christopher S Henry, Martin Kupiec, Uri Gophna, Roded Sharan, and Eytan Rupp. 2011. "Competitive and Cooperative Metabolic Interactions in Bacterial Communities." *Nature Communications* 2 (1): 589.
- Galloway, J N, F J Dentener, D G Capone, E W Boyer, R W Howarth, S P Seitzinger, G P Asner, et al. 2004. "Nitrogen Cycles: Past, Present, and Future." *Biogeochemistry* 70 (2): 153–226.
- Gans, Jason, Murray Wolinsky, and John Dunbar. 2005. "Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil." *Science* 309 (5739): 1387–90.
- Goodrich, Julia K, Jillian L Waters, Angela C Poole, Jessica L Sutter, Omry Koren, Ran Blekman, Michelle Beaumont, et al. 2014. "Human Genetics Shape the Gut Microbiome." *Cell* 159 (4): 789–99.
- Grandclément, Catherine, Mélanie Tannières, Solange Moréra, Yves Dessaux, and Denis Faure. 2016. "Quorum Quenching: Role in Nature and Applied Developments." *FEMS Microbiology Reviews* 40 (1): 86–116.
- Großkopf, Tobias, and Orkun S Soyer. 2014. "Synthetic Microbial Communities." *Current Opinion in Microbiology* 18: 72–77.
- Guo, Liang, Wenwen Diao, Cong Gao, Guipeng Hu, Qiang Ding, Chao Ye, Xiulai Chen, Jia Liu, and Liming Liu. 2020. "Engineering Escherichia Coli Lifespan for Enhancing Chemical Production." *Nature Catalysis* 3 (3): 307–18.
- Hammarlund, Sarah P, Jeremy M Chacón, and William R Harcombe. 2019. "A Shared Limiting Resource Leads to Competitive Exclusion in a Cross-Feeding System." *Environmental Microbiology* 21 (2): 759–71.
- Hammarlund, Sarah P, Tomáš Gedeon, Ross P Carlson, and William R Harcombe. 2021. "Limitation by a Shared Mutualist Promotes Coexistence of Multiple Competing Partners." *Nature Communications* 12 (1): 1–8.
- Heijden, Marcel G A Van Der, Richard D Bardgett, and Nico M Van Straalen. 2008. "The Unseen Majority: Soil Microbes as Drivers of Plant Diversity and Productivity in Terrestrial Ecosystems." *Ecology Letters* 11 (3): 296–

- Hom, Erik F Y, and Andrew W Murray. 2014. "Niche Engineering Demonstrates a Latent Capacity for Fungal-Algal Mutualism." *Science* 345 (6192): 94–98.
- Hooper, D U, F S Chapin III, J J Ewel, A Hector, P Inchausti, S Lavorel, J H Lawton, et al. 2005. "Effects Of Biodiversity On Ecosystem Functioning: A Consensus Of Current Knowledge." *Ecological Monographs* 75 (1): 3–35.
- Hungate, Robert E. 2013. *The Rumen and Its Microbes*. Elsevier.
- Inda, Maria Eugenia, Esther Broset, Timothy K Lu, and Cesar de la Fuente-Nunez. 2019. "Emerging Frontiers in Microbiome Engineering." *Trends in Immunology* 40 (10): 952–73.
- Kerner, Alissa, Jihyang Park, Audra Williams, and Xiaoxia Nina Lin. 2012. "A Programmable Escherichia Coli Consortium via Tunable Symbiosis." *PloS One* 7 (3): e34032.
- Konopka, Allan. 2009. "What Is Microbial Community Ecology?" *The ISME Journal* 3 (11): 1223–30.
- Lane, Nick. 2015. "The Unseen World: Reflections on Leeuwenhoek (1677)'Concerning Little Animals.'" *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1666): 20140344.
- Larimer, Frank W, Patrick Chain, Loren Hauser, Jane Lamerdin, Stephanie Malfatti, Long Do, Miriam L Land, et al. 2004. "Complete Genome Sequence of the Metabolically Versatile Photosynthetic Bacterium *Rhodospseudomonas Palustris*." *Nature Biotechnology* 22 (1): 55–61.
- LaSarre, Brea, Alexandra L McCully, Jay T Lennon, and James B McKinlay. 2017. "Microbial Mutualism Dynamics Governed by Dose-Dependent Toxicity of Cross-Fed Nutrients." *The ISME Journal* 11 (2): 337–48.
- Ley, Ruth E, Catherine A Lozupone, Micah Hamady, Rob Knight, and Jeffrey I Gordon. 2008. "Worlds within Worlds: Evolution of the Vertebrate Gut Microbiota." *Nature Reviews Microbiology* 6 (10): 776–88.
- Li, Zhenghong, Xiaonan Wang, and Haoran Zhang. 2019. "Balancing the Non-Linear Rosmarinic Acid Biosynthetic Pathway by Modular Co-Culture Engineering." *Metabolic Engineering* 54: 1–11.
- Lindell, Kristoffer. 2012. "Cell-to-Cell Communication and Virulence in *Vibrio Anguillarum*." Umeåuniversitet.
- Lindner, Ariel B, Richard Madden, Alice Demarez, Eric J Stewart, and François Taddei. 2008. "Asymmetric Segregation of Protein Aggregates Is Associated with Cellular Aging and Rejuvenation." *Proceedings of the National Academy of Sciences* 105 (8): 3076–81.
- Louca, Stilianos, Martin F Polz, Florent Mazel, Michaeline B N Albright, Julie A Huber, Mary I O'Connor, Martin Ackermann, et al. 2018. "Function and Functional Redundancy in Microbial Systems." *Nature Ecology & Evolution* 2 (6): 936–43.
- Lozupone, Catherine A, and Rob Knight. 2007. "Global Patterns in Bacterial Diversity." *Proceedings of the National Academy of Sciences* 104 (27): 11436–40.
- Lozupone, Catherine A, Jesse I Stombaugh, Jeffrey I Gordon, Janet K Jansson, and Rob Knight. 2012. "Diversity, Stability and Resilience of the Human Gut Microbiota." *Nature* 489 (7415): 220–30.
- Melkonian, Chrats, Willi Gottstein, Sonja Blasche, Yongkyu Kim, Martin Abel-Kistrup, Hentie Swiegers, Sofie Saerens, et al. 2019. "Finding Functional Differences between Species in a Microbial Community: Case Studies in Wine Fermentation and Kefir Culture." *Frontiers in Microbiology* 10: 1347.
- Miller, Melissa B, and Bonnie L Bassler. 2001. "Quorum Sensing in Bacteria." *Annual Review of Microbiology* 55 (1): 165–99.
- Mimee, Mark, Robert J Citorik, and Timothy K Lu. 2016. "Microbiome Therapeutics — Advances and Challenges." *Advanced Drug Delivery Reviews* 105: 44–54.
- Moeller, Andrew H, Alejandro Caro-Quintero, Deus Mjungu, Alexander V Georgiev, Elizabeth V Lonsdorf, Martin N Muller, Anne E Pusey, Martine Peeters, Beatrice H Hahn, and Howard Ochman. 2016. "Cospeciation of Gut

- Microbiota with Hominids." *Science* 353 (6297): 380–82.
- Montgomery, Kate, James C Charlesworth, Rebecca LeBard, Pieter T Visscher, and Brendan P Burns. 2013. "Quorum Sensing in Extreme Environments." *Life* 3 (1): 131–48.
- Nyström, Thomas, and Beidong Liu. 2014. "Protein Quality Control in Time and Space--Links to Cellular Aging." *FEMS Yeast Research* 14 (1): 40–48.
- "Plant-E." 2009. 2009. <https://www.plant-e.com/en/hoer-werkt-het/>.
- Qiu, Zhiguang, Eleonora Egidi, Hongwei Liu, Simranjit Kaur, and Brajesh K Singh. 2019. "New Frontiers in Agriculture Productivity: Optimised Microbial Inoculants and in Situ Microbiome Engineering." *Biotechnology Advances* 37 (6): 107371.
- Raich, J W, and W H Schlesinger. 1992. "The Global Carbon Dioxide Flux in Soil Respiration and Its Relationship to Vegetation and Climate." *Tellus B* 44 (2): 81–99.
- Roquet, Nathaniel, Ava P Soleimany, Alyssa C Ferris, Scott Aaronson, and Timothy K Lu. 2016. "Synthetic Recombinase-Based State Machines in Living Cells." *Science* 353 (6297).
- Rothschild, Daphna, Omer Weissbrod, Elad Barkan, Alexander Kurilshikov, Tal Korem, David Zeevi, Paul I Costea, et al. 2018. "Environment Dominates over Host Genetics in Shaping Human Gut Microbiota." *Nature* 555 (7695): 210–15.
- Rugbjerg, Peter, Nils Myling-Petersen, Andreas Porse, Kira Sarup-Lytzen, and Morten O A Sommer. 2018. "Diverse Genetic Error Modes Constrain Large-Scale Bio-Based Production." *Nature Communications* 9 (1): 1–14.
- Shendure, Jay, Shankar Balasubramanian, George M Church, Walter Gilbert, Jane Rogers, Jeffery A Schloss, and Robert H Waterston. 2017. "DNA Sequencing at 40: Past, Present and Future." *Nature* 550 (7676): 345–53.
- Sherwin, Eoin, Seth R Bordenstein, John L Quinn, Timothy G Dinan, and John F Cryan. 2019. "Microbiota and the Social Brain." *Science* 366 (6465).
- Shou, Wenying, Sri Ram, and Jose M G Vilar. 2007. "Synthetic Cooperation in Engineered Yeast Populations." *Proceedings of the National Academy of Sciences* 104 (6): 1877–82.
- Smith, Mark B, Colleen Kelly, and Eric J Alm. 2014. "Policy: How to Regulate Faecal Transplants." *Nature* 506 (7488): 290–91.
- Sponzilli, Ivonne, and Luigi D Notarangelo. 2011. "Severe Combined Immunodeficiency (SCID): From Molecular Basis to Clinical Management." *Acta Bio-Medica: Atenei Parmensis* 82 (1): 5–13.
- Staley, J T, and A Konopka. 1985. "Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats." *Annual Review of Microbiology* 39: 321–46.
- Steiner, Ulrich Karl. 2021. "Senescence in Bacteria and Its Underlying Mechanisms." *Frontiers in Cell and Developmental Biology* 9.
- Strickland, Michael S, and Johannes Rousk. 2010. "Considering Fungal:Bacterial Dominance in Soils – Methods, Controls, and Ecosystem Implications." *Soil Biology and Biochemistry* 42 (9): 1385–95.
- Summers, Zarath M, Heather E Fogarty, Ching Leang, Ashley E Franks, Nikhil S Malvankar, and Derek R Lovley. 2010. "Direct Exchange of Electrons Within Aggregates of an Evolved Syntrophic Coculture of Anaerobic Bacteria." *Science* 330 (6009): 1413–15.
- Tanoue, Takeshi, Satoru Morita, Damian R Plichta, Ashwin N Skelly, Wataru Suda, Yuki Sugiura, Seiko Narushima, et al. 2019. "A Defined Commensal Consortium Elicits CD8 T Cells and Anti-Cancer Immunity." *Nature* 565 (7741): 600–605.
- Teng, Ying, and Wei Chen. 2019. "Soil Microbiomes—a Promising Strategy for Contaminated Soil Remediation: A Review." *Pedosphere* 29 (3): 283–97.

- "The Earth Microbiome Project." 2010. 2010. <https://earthmicrobiome.org/people/working-group-consortium/>.
- Thompson, Luke R, Jon G Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, Kenneth J Locey, Robert J Prill, et al. 2017. "A Communal Catalogue Reveals Earth's Multiscale Microbial Diversity." *Nature* 551 (7681): 457–63.
- Tokuhara, Daisuke. 2021. "Role of the Gut Microbiota in Regulating Non-Alcoholic Fatty Liver Disease in Children and Adolescents." *Frontiers in Nutrition* 8 (June): 700058.
- Tremaroli, Valentina, and Fredrik Bäckhed. 2012. "Functional Interactions between the Gut Microbiota and Host Metabolism." *Nature* 489 (7415): 242–49.
- Vallery-Radot, René. 1901. *La Vie de Pasteur*. Librairie Hachette.
- Wainwright, Milton B T - *Advances in Applied Microbiology*. 2003. "An Alternative View of the Early History of Microbiology." In , 52:333–55. Academic Press.
- Walls, Laura E, Koray Malci, Behnaz Nowrouzi, Rachel A Li, Leo D'Espaux, Jeff Wong, Jonathan A Dennis, et al. 2021. "Optimizing the Biosynthesis of Oxygenated and Acetylated Taxol Precursors in *Saccharomyces Cerevisiae* Using Advanced Bioprocessing Strategies." *Biotechnology and Bioengineering* 118 (1): 279–93.
- Woese, Carl R, and George E Fox. 1977. "Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms." *Proceedings of the National Academy of Sciences* 74 (11): 5088–90.
- Woese, Carl R, Otto Kandler, and Mark L Wheelis. 1990. "Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya." *Proceedings of the National Academy of Sciences* 87 (12): 4576–79.
- Wu, Stephen G, Lian He, Qingzhao Wang, and Yinjie J Tang. 2015. "An Ancient Chinese Wisdom for Metabolic Engineering: Yin-Yang." *Microbial Cell Factories* 14 (1): 1–9.
- Yang, Yaoyu, Jennifer L Nemhauser, and Eric Klavins. 2019. "Synthetic Bistability and Differentiation in Yeast." *ACS Synthetic Biology* 8 (5): 929–36.
- You, Lingchong, Robert Sidney Cox, Ron Weiss, and Frances H Arnold. 2004. "Programmed Population Control by Cell–cell Communication and Regulated Killing." *Nature* 428 (6985): 868–71.
- Youk, Hyun, and Wendell A. Lim. 2014. "Secreting and Sensing the Same Molecule Allows Cells to Achieve Versatile Social Behaviors." *Science* 343 (6171).
- Zhang, Xuwang, Dongli Bao, Maoting Li, Qidong Tang, Minghuo Wu, Hao Zhou, Lifan Liu, and Yuanyuan Qu. 2021. "Bioremediation of Petroleum Hydrocarbons by Alkali-Salt-Tolerant Microbial Consortia and Their Community Profiles." *Journal of Chemical Technology & Biotechnology* 96 (3): 809–17.
- Zheng, Danping, Timur Liwinski, and Eran Elinav. 2020. "Interaction between Microbiota and Immunity in Health and Disease." *Cell Research* 30 (6): 492–506.
- Zhou, Jin, Yihua Lyu, Mindy L Richlen, Donald M Anderson, and Zhonghua Cai. 2016. "Quorum Sensing Is a Language of Chemical Signals and Plays an Ecological Role in Algal-Bacterial Interactions." *Critical Reviews in Plant Sciences* 35 (2): 81–105.
- Zhou, Kang, Kangjian Qiao, Steven Edgar, and Gregory Stephanopoulos. 2015. "Distributing a Metabolic Pathway among a Microbial Consortium Enhances Production of Natural Products." *Nature Biotechnology* 33 (4): 377–83.
- Zhuang, Wei-Qin, Shan Yi, Markus Bill, Vanessa L Brisson, Xueyang Feng, Yujie Men, Mark E Conrad, Yinjie J Tang, and Lisa Alvarez-Cohen. 2014. "Incomplete Wood-Ljungdahl Pathway Facilitates One-Carbon Metabolism in Organohalide-Respiring *Dehalococcoides Mccartyi*." *PNAS* 111 (17): 6419–24.
- Zwier, Karen R. 2018. "Methodology in Aristotle's Theory of Spontaneous Generation." *Journal of the History of Biology* 51 (2): 355–86.

Chapter 2

STATE OF THE ART

“The purpose of science is not to cure us of our sense of mystery and wonder, but to constantly reinvent and reinvigorate it.”
Robert Sapolsky

Cybergenetics

Background

Control of biological systems has been sought for a long time and has been met with incremental degrees of success over the years (Pouzet et al. 2020). A lot of this success is driven by insights afforded by mathematical models that can accurately capture the dynamics of the system, especially at the molecular scale, and can be used to guide the necessary control signal required to drive the biological system in question to the desired target. The development of such a predictive model requires extensive characterization to infer features that dictate dynamics and this in turn informs the choice of model. Once a model has been constructed, it can be fit to the data to estimate the parameters. This parametrized model then can be used to predict the behaviour of the system given different inputs, and after supplementing it with an optimisation algorithm, can output the control signal needed to achieve desired behaviour. Such control is also called open-loop control. One may also choose to close the loop and this is where cybergenetics comes into picture. In recent years, advances in biological control have come from coupling computers with growing cells carrying the engineered system, made possible by special platforms that integrate biological systems with the computer via a feedback loop (Milias-Argeitis et al. 2011; Toettcher et al. 2011; Menolascina et al. 2014; Lugagne et al. 2017; Carrasco-López et al. 2020) (**Figure 2.1**).

This allows the deployment of the model in a model predictive framework. In this framework, data from the ongoing control experiment is analysed in real time and the relevant information is extracted and used to estimate the state of the system vis-à-vis the model and the model is used to query an optimised light signal that would drive the system towards the target for a given time horizon. This light signal can be updated every time a measurement is made (or as frequently as the data analysis and optimisation permits) to constitute what is called receding horizon or model predictive control (MPC).

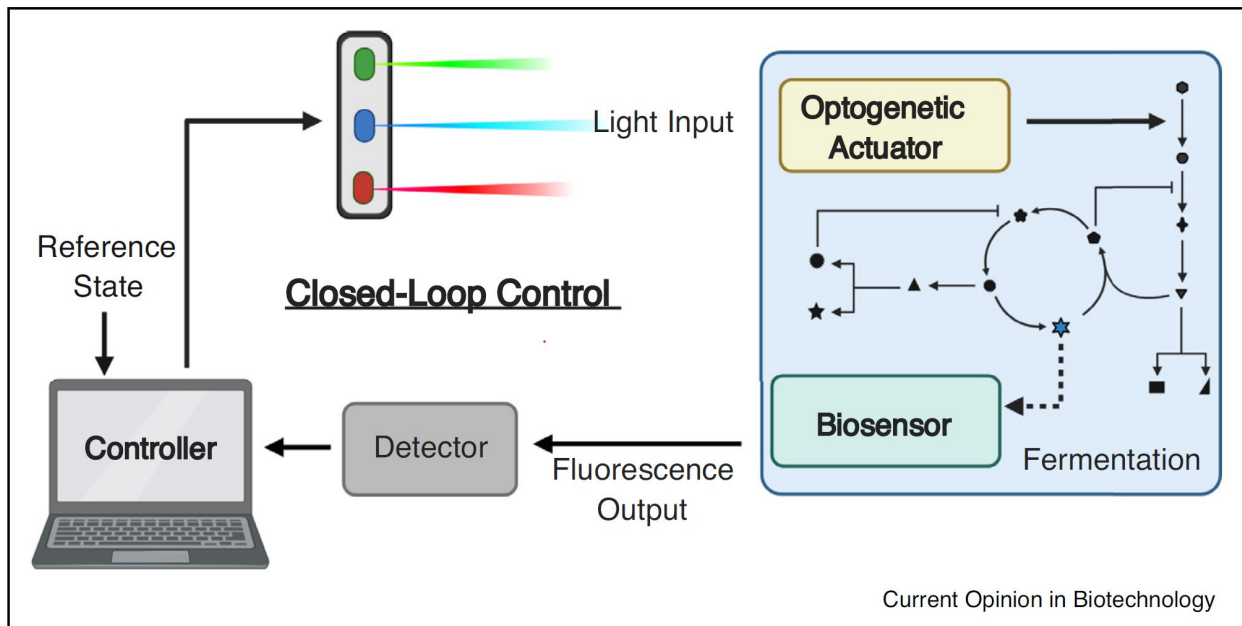


Figure 2.1. Principle of in-silico feedback control. The engineered biological circuit is observed continuously via monitoring the fluorescence output of a biosensor/ activity marker. The data is analysed on-line and the relevant information is sent to the controller. The controller might be model predictive or bang-bang. In the former, the data at each sampling is used to estimate the state of the system, which is then fed into the model to obtain an optimized control signal. Concretely, MPC requires repeatedly solving a constrained optimization problem, using predictions of future errors, over a moving time horizon to decide the control signal. For bang-bang control, the data is used to provide an all-or-none signal to the system depending on whether the observable is above the target or below, respectively. Apart from these two control schemes, the controller response can also be made proportional to the error between the observed and desired levels (P controller). It may further be complemented by taking into account either the derivative of the error (PD controller) or the integral of error over time (PI controller) or both (PID controller). Other control schemes that are not routinely employed for biological control include, fuzzy, neural net, sliding mode etc. (Figure adapted from Carrasco-López et al. (2020))

This process may be done for controlling the population by minimizing the difference between population level mean and the target (called error in control theory parlance) or at the single cell level by minimizing the error of each cell with the target value. The latter necessitates solving the optimisation problem for each cell as well as a means of delivering unique signals to individual cells independently of others. However, the application of MPC is limited to systems where data can be readily measured and analysed in real time. Even then, the resolution of the MPC is constrained by the frequency of measurements, which may in turn be limited by the computational time required for analysis and optimisation. Resolution may further be impacted by the dynamics of the system to be controlled and delays in delivering control signal. While other controllers (PID, bang-bang, etc.) have been deployed to control biological systems with

in-silico feedback, MPC remain the most impressive examples of control of biological systems especially when they are used to control behaviour at the single cell level.

A number of studies have demonstrated robust control of biological processes with MPC using chemical induction. The landmark contribution came from Uhlendorf et al. (2012), in which the authors demonstrated long-term control of gene expression in yeast at the population and the single cell level using the high osmolarity glycerol (HOG) pathway. The gene of interest was placed downstream of the HOG signalling cascade effectively giving authors the ability to change gene expression by changing the glycerol content of the medium. They used a custom microfluidic device to observe growing cells under the microscope and to change the osmolarity of the medium in order to trigger the HOG pathway. Images were analysed online and coupled to a controller that adjusted the osmolarity of the medium. With this experimental setup, the authors could control gene expression for extended durations at both the population level and in single cells. Menolascina et al. (2014) showed that dynamically changing the concentration of galactose and glucose enabled dynamic control of expression from a galactose inducible promoter in yeast cells growing in a microfluidic chamber. Another study in *E. coli* (Lugagne et al. 2017), reported that it was possible to maintain the toggle switch in an unstable configuration by tuning the concentration of two inducers that express the mutually antagonist repressors (basis of toggle switch). The authors used a platform similar to the one reported in Uhlendorf et al. (2012) albeit customized for *E. coli* growth. To rapidly change the concentration of inducers, all three studies relied on microfluidic devices. This limits the control to small-scale cultures and while, it is possible, in theory, to implement the same in liquid cultures, the diffusion of chemical inducers introduces non-trivial delays and their removal remains tedious.

The development of optogenetics, i.e. the use of light to trigger cellular processes, has contributed significantly to control applications by increasing the spatiotemporal resolution of the control signal (Miliadis-Argeitis et al. 2011; Toettcher et al. 2011; Olson et al. 2014; Miliadis-Argeitis et al. 2016; Chait et al. 2017; Rullan et al. 2018; Zhao et al. 2018; Johnson et al. 2020; Perkins et al. 2020; Liu, Chen, and Wang 2020; Bertaux et al. 2020; Lalwani et al. 2021; Fox et al. 2021). Light possesses several advantages over chemical inducers. Light is cheap, can be instantaneously delivered and removed and this process can be easily automated. Last but not the least, light is cleaner i.e. no chemical waste. In addition to this, the spatiotemporal precision with which light can be delivered makes it a more suitable choice for single cell control (Baumschlager and Khammash 2021). A number of different systems have been described in bacteria, yeast and mammalian cells. **Figure 2.2** summarizes these systems.

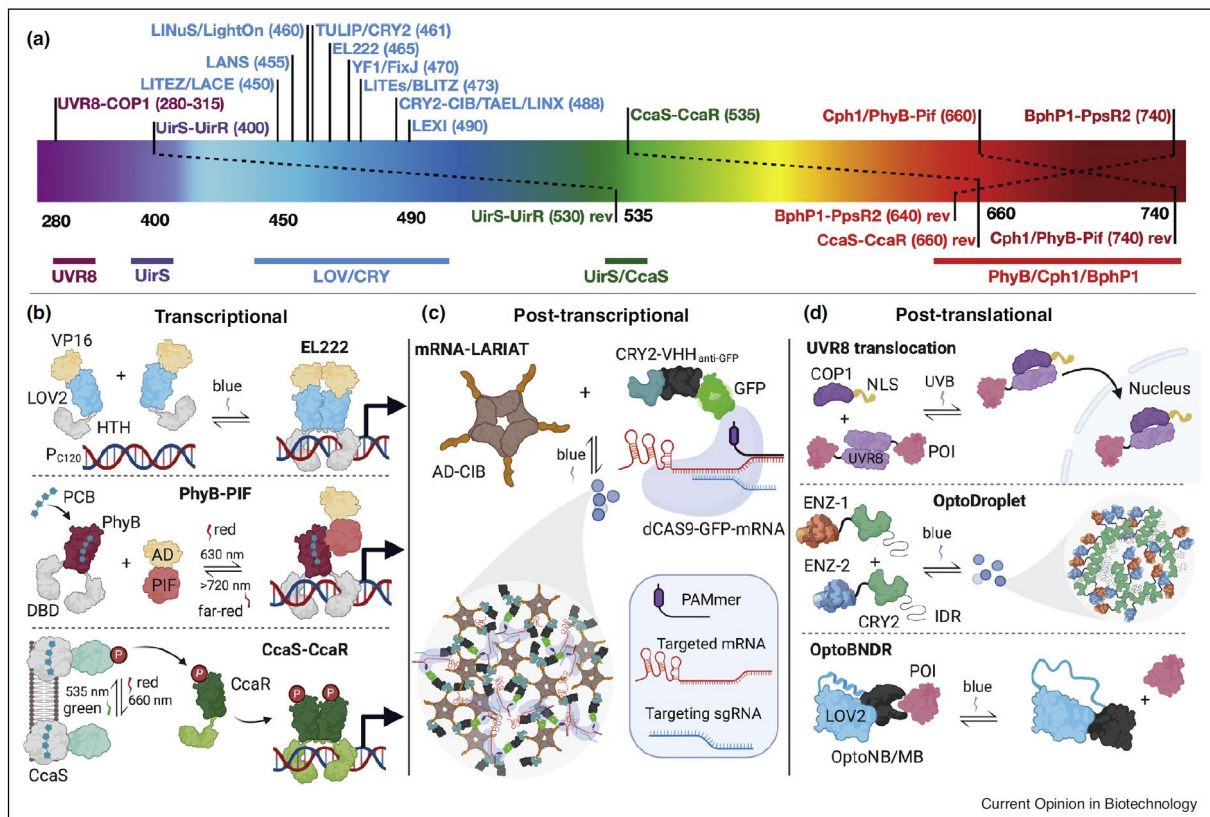


Figure 2.2. Various optogenetic systems characterized to date (not exhaustive). The systems may be classified on the basis of the wavelength of activating light or according to the actuation process (transcriptional, translational, and post translation). The actuation process determines the temporal aspects of the control (transcriptional being the slowest while post translation being the fastest). (Figure adapted from Carrasco-López et al. (2020). All rights belong to Elsevier.)

Several optogenetic systems in addition to those showcased in **Figure 2.2** have been described. It is beyond the scope of this discussion to enumerate all existing systems and the different ways in which they have been utilized to control cellular processes, however, curious readers are encouraged to read a somewhat dated yet pertinent review (Repina et al. 2017). Comparison with a more recent review (Seong and Lin 2021), highlights the rapid pace at which optogenetic tools and technologies are emerging. Notably, a recent study (Romano et al. 2021) in *E. coli* engineered a modified version of AraC system that responds to blue light and the same group reported a dCAS9 based optogenetic repressor in yeast (Geller et al. 2019). Another recent study developed an efficient and fast optogenetic system for protein degradation in human cells (Reynders et al. 2020). Their approach is based on photoactivable proteolysis targeting chimeras that, in addition to carrying an E3 ligase ligand and a ligand for binding to the protein to be degraded, carry a photoswitch.

From here on, I focus on optogenetic control due to superior spatiotemporal resolution of the control signal.

Population level control

Some of the earliest examples of cybergenetics involve control of population behaviour. Milias-Argeitis et al. (2011) demonstrated for the first time, using MPC, that it was possible to control the output of a gene in vivo by observing the system in real time. In order to achieve this, the authors used the Phy/PIF optogenetic system that is activated with red light and can be deactivated with far-red light (**Figure 2.2**). They used it to drive the expression of a fluorescent protein in yeast and developed a model to describe the dynamics of induction and de-induction. Next, they grew cells in a flask equipped with LEDs and took regular bulk measurements of the fluorescence in the culture with a spectrophotometer. Using this setup and integrating it in a real time analysis module, the authors could use an in silico feedback algorithm to compute regulatory inputs that maintained the fluorescence of the culture at the target levels. This proof of concept study while demonstrating the applicability of in silico feedback, showed limited tracking accuracy. The authors also mentioned that the controller was incapable of robust control in the face of external perturbations and day-to-day variability. In Milias-Argeitis et al. (2016), the authors addressed these issues by developing a completely autonomous platform capable of long-term optogenetic feedback control in continuous cultures. Concretely, the upgraded platform permitted continuous measurement of the state of the system, which did not rely on human intervention. Additionally, the measurements were analysed on the fly, followed by state estimation, optimization of control signal and actuation. Thus, having integrated all aspects of MPC in a single framework, the platform became autonomous. The authors used the CcaS/R system in *E. coli* to demonstrate control of gene expression that was robust to external perturbations (which were delivered in the form of change of media or change in temperature). The authors further compared the performance of different controllers (PI, MPC and open-loop) and found that PI and MPC controllers were superior to open-loop control that was noisy and extremely sensitive to perturbations. To further show that control gene expression can be used to dynamically control other properties of the system, the authors controlled the growth rate of the culture by regulating the expression of an essential gene required for growth. Another study in yeast (Harrigan, Madhani, and El-Samad 2018), reported that it was possible to compensate for the absence of biological feedback in an intracellular signalling pathway (mutant lacking a key gene in the cascade) by integrating the biological system under study to computers and using light to control the expression of downstream gene in the cascade. The authors demonstrated that by measuring the levels of the gene that regulates the biological feedback and developing a quantitative model of the system, it was possible to compensate for the lack of biological feedback by using computer-aided light-driven expression. This permitted the

authors to rescue the phenotype of the signalling mutant. Toettcher et al. (2011) used the Phy-PIF system to control intracellular signalling dynamics in mammalian cells. The authors achieved this feat by engineering optogenetically active fusion proteins. PHY was fused to a transmembrane protein involved in intracellular signalling while PIF was fused to a downstream signalling domain. Using red light, the authors showed that the PIF protein fused to the signalling domain could be recruited to the plasma membrane in a light-dependent fashion. Upon recruitment, the signalling domain fused to PIF triggered the recruitment of another signalling protein to the plasma membrane. Next, using an in silico feedback loop similar in principle to the one used in Milias-Argeitis et al. (2016) the authors demonstrated that by tuning the light signal they could maintain the levels of PIF recruited to the membrane and in turn control the levels of the downstream signalling protein recruited. Such control was not possible without the feedback. This, in theory, can be coupled to the expression of a gene, although the authors did not demonstrate this. It is worth noting that the authors used a microscopy platform for measurements and real time image analysis for closing the loop. Olson et al. (2014) demonstrated dynamic control gene expression using the CcaS-CcaR in bacteria. CcaS-CcaR can be activated by green light and deactivated by red light (**Figure 2.2**). The authors coupled this system to the expression of a fluorescent protein. They were able to develop a predictive model by characterizing the induction behaviour in presence of green light and de-induction behaviour in the presence of red light in liquid cultures. The authors then queried the model for regulatory light sequences that would lead to desired dynamics at the population level. The authors validated their control objectives a posteriori by measuring the fluorescence at different timepoints in a flow cytometer after the end of the experiment. Their results relating to control of steady state fluorescence were particularly impressive for open-loop control, however, for dynamic control of population fluorescence, the authors observed a set point error. The authors noted in the discussion that an in-silico feedback should allow for dynamic control in addition to increasing the precision by updating the control signal in real time. Open-loop optogenetic control of protein expression was also exploited by Zhao et al. (2018) and Lalwani et al. (2021) to optimise bio production in yeast and *E. coli*, respectively. In the former, the authors used an EL222 based system, tailored to function in yeast, to regulate metabolic flux in the population and demonstrated higher titers of butanol and isopropanol. In the latter, the authors developed a novel optogenetic system in *E. coli* by engineering the lac repression to become responsive to light. Briefly, they used the pDawn system (Ohlendorf et al. 2012) to repress the expression of LacI that in turn represses the lac operon (that drives the gene of interest). In the absence of light, the expression of LacI cannot take place and, therefore, the gene of interest is expressed from the lac operon. In the presence of light, the pDawn system represses the repressor of LacI leading

to the expression of *Lacl*. *Lacl* expression, then, represses the gene of interest. Using this optoLac system, the authors demonstrated marginal, but significant, increase over the canonical IPTG driven production of isobutanol, mevalonate and a recombinant protein (YFP). Interestingly, the authors observed non-monotonic behaviour in their optoLac constructs with respect to induction timing when producing metabolites (isobutanol and mevalonate) but not with the recombinant protein. The authors discussed the presence of a memory effect in the system that reduces the temporal resolution of their control objective. A recent study from our lab, Bertaux et al (2021), demonstrated robust and dynamic control of gene expression at the population level in yeast by using blue light inducible EL222 transcription factor (**Figure 2.2**). The induction dynamics were characterized by shining repeated pulses of different durations in a custom bioreactor platform capable of automated cytometry measurements and based on this characterization data, a quantitative model was developed. Automated sampling was used to close the loop and the quantitative model, in an MPC framework, allowed for precise dynamic control of gene expression in yeast. Multichromatic control of gene expression has been demonstrated for *E. coli* at the population level (Fernandez-Rodriguez et al. 2017). Using three orthogonal optogenetic systems the authors demonstrated spatial control of gene expression. Furthermore, by coupling biosynthetic pathways to produce red, green, and blue pigments in response to red, green, and blue light, respectively, the authors demonstrated that a colony of *E. coli* may be turned into a RGB colour photograph. A common shortcoming of all the studies described here and for population level control in general is the large variability around the target.

Single cell control

While population level control allows one to tune the mean of the population by providing global control signals, it is limited, however, in controlling the variability in the population around the mean. This is due to the phenotypic heterogeneity that exists in an isogenic population (Raj and Van Oudenaarden 2008). As a consequence of this heterogeneity, individual cells behave slightly differently with respect to the average behaviour of the population. This constitutes a major challenge for control purposes primarily because models typically used for population level control are agnostic to this heterogeneity and consider each cell to be identical. In terms of parametrization of the model, this means that while the behaviour of the population can be described by population level estimates, the bulk parameters cannot predict the behaviour of individual cells accurately. The most straightforward example of how this heterogeneity might affect control, is variability in transcription factor levels. In order to account for phenotypic heterogeneity, either the model needs to be stochastic, such that phenotypic heterogeneity arises mechanistically, or model parameters need to be re-identified for each cell from single cell data.

For the former, computation costs become large as the number of molecules increase and for the latter, the re-identification might be time consuming. Both these aspects contribute to the challenging nature of single cell control by increasing the time required to analyse the data and provide an optimised control signal. Another aspect that makes single cell control challenging, is the need to infer information about the state of single cells and to deliver appropriate control signals individually. Nevertheless, several studies have demonstrated single cell control of cellular processes. A common feature of these studies is that they employ microscope as the platform to measure the output (i.e. fluorescence) and to deliver control signals to single cells (light). The latter has been enabled by the use of digital mirror devices (DMD) that allow one to shine light with a high degree of spatial control.

Chait et al. (2017) reported an experimental setup that allowed the authors to measure and optogenetically control the gene expression at the single cell level in *E. coli*. The authors used CcaS/R system in conjunction with a mother machine (Wang et al. 2010) and a simple stochastic model to achieve single cell control. The authors compared the performance of open-loop, population level and single cell closed-loop feedback controllers and found that in all three cases, the average population fluorescence followed the target levels set by the controller. Even if the performance was comparable, single cell closed-loop displayed the least error from the target followed by population closed-loop and open-loop control. The crucial difference between the three controllers, however, was the tightness of the population distributions around the target level of fluorescence. The authors showed, with the help of long-term experiments that featured a change of controller during the experiment, a marked reduction in the variability with the single cell closed-loop controller. A similar reduction in heterogeneity by using the single cell closed-loop feedback control was reported by our lab in Fox et al. (2021), in which we used the EL222 system to control the expression of a fluorescent protein in individual cells in yeast. In the study, we proposed a new software framework that integrates all aspects of feedback control with the control of the microscope. As a true testament to the ability to control gene expression at the single cell level, the authors maintained cells growing in the same field of view at different levels of fluorescence by delivering individualized light signals (via the DMD). Our results concerning the superior control of the population fluorescence around the target value echoed those reported in Chait et al. (2017). Another study in yeast (Rullan et al. 2018) employing a similar experimental setup, showed single cell control of population fluorescence using the EL222 system and reported a marked decrease in cell-to-cell variability compared to population level control. In addition, the authors demonstrated that the heterogeneity in response to global light signals was a consequence of the bursty nature of transcription factor activity in presence of light. As a follow up to this study, the authors emulated an artificial signalling

pathway in yeast via model guided computer controlled control of single cells (Perkins et al. 2020). More precisely, the authors employed the same experimental pipeline as Rullan et al. (2018) for measuring the levels of fluorescence in single cells (that are driven by light) and to deliver individualised light signals, with one crucial difference. Instead of determining the control signal on the basis of the fluorescence of the cell itself, the control signal was determined by the neighbouring cells. The authors claimed that this could be used to generate patterns in yeast, however, the results they obtained do not provide any concrete evidence for the same. I will allow myself to comment that they used a complicated reporter system to assess the performance of their approach whose rationale is not obvious to me. Briefly, light triggers the expression of a protein consisting of two domains, an NLS domain and a heterodimeric domain SZ1. SZ1 can bind to another domain, SZ2, which is tagged to a constitutively expressed RFP such that upon light induction, RFP signal is enriched in the nucleus. Depending on the ratio of RFP in the cytoplasm to that in the nucleus, a cell might be classified as “on” or “off”. However, the results are not as straightforward to interpret as the fold change in fluorescence is not striking enough to confirm the theoretical results obtained. It would have been more prudent and perhaps more simple to implement both mathematically and experimentally to trigger expression of a fluorescent protein (like mNeonGreen or mCerulean, which have fast maturation dynamics) with EL222. Another recent study (Pedone et al. 2021) developed a computational framework for optogenetic control at the single cell level which, along with similar studies like Fox et al. (2021), is bound to facilitate future research.

Optogenetic control in multicellular systems

Optogenetics has been used to control cellular processes in other contexts, for instance, signalling dynamics (Toettcher, Weiner, and Lim 2013), morphogenesis (Johnson et al. 2020), and neuroscience (Liu, Chen, and Wang 2020). Although not strictly comparable to the *in-silico* feedback applications of optogenetic control described so far, to give a flavour of the possibilities, I talk about the three studies referred to above. These studies were chosen because they accurately capture the broad scope of optogenetic control.

In Toettcher, Weiner and Lim (2013) the authors exploited optogenetic control to understand the functional plasticity of signalling modules in multicellular organisms. In order to achieve this, the authors focussed on elucidating signal transmission in the RAS/ERK MAPK cascade. The choice of pathway was motivated by the fact that it can be activated by a variety of extracellular signals and lead to diverse fates including proliferation, differentiation, and arrest and it not well-understood how the cell distinguishes between different modes of activation. Using an approach similar to the one described for Toettcher et al. (2011) to activate RAS, the authors found that

information needed for the cell to distinguish and decide was encoded in the dynamics of RAS activation. More precisely, the authors observed distinct downstream processes being triggered upon the duration of RAS activation.

Johnson et al. (2020) reported optogenetic control of developmental processes in *Drosophila melanogaster*. Concretely, the authors engineered a patterning mutant of *drosophila*, which possessed a lethal loss of patterning in early embryos that led to incomplete development. Physiologically, the pattern is established due to a terminal-to-interior gradient of a signalling protein and essential for anterior and posterior localized processes. The absence of this gradient results in defective physiology. The authors introduced light controllable expression of this protein and demonstrated that shining light at the poles was sufficient for the rescue of the mutant. Interestingly, the authors found that the an all-or-none signal at the terminal ends of the embryo (as opposed to a gradient in WT *drosophila*) at the right time during embryogenesis could rescue the embryo, suggesting that information encoded in the gradient was not necessary for this process.

Liu, Chen, and Wang (2020) evoked optogenetic control of neuronal activity to decipher the relative contributions of GABAergic motor neurons to locomotor decision making in *Caenorhabditis elegans*. The authors could dissect the role of these neurons due to a blue light sensitive channel rhodopsin (CHR2) expressed specifically in these cells and that could be activated at single cell resolution. Using this approach, the authors found that GABAergic neurons play a major role in determining the behaviour of the worm. *C. elegans* is the model organism for studying the connectedness of the neural system (connectome) because it is composed of only 302 neurons (Yan et al. 2017).

Growth production tradeoffs in bioproduction

Background

Bioproduction presents as a lucrative and powerful approach to produce diverse molecules of interest including (biodegradable) plastic (Franziska et al. 2011), drugs & aromatics (Luo et al. 2019; Liu et al. 2019), vaccines (Smith, Lipsitch, and Almond 2011), and antibiotics (Vary et al. 2007) that are hard or impossible to synthesize chemically and/or difficult to harvest from nature (McElroy and Jennewein 2018) and is the only approach that could potentially fix atmospheric carbon in the process (Gassler et al. 2020; Gleizer et al. 2019). The advantage over chemical synthesis has become more daunting over the past few decades following a tremendous increase in the ease of reading and writing DNA (Church 2020). The former has unlocked a vast natural repertoire of evolutionarily optimized metabolic pathways and enzymes (Omura 2011; Zhang and Elliot 2019; Blakemore et al. 2018) while the latter is culminating in an ability to manipulate genomes on an unprecedented scale with remarkable accuracy (Robertson et al. 2021; Shao et al. 2019). These advances coupled to an accrescent understanding of cellular metabolism (Erickson et al. 2017), in all its biochemical-reaction-network glory (Guimera and Amaral 2005) and nuances in different environmental contexts (Basan et al. 2020), permits one to not only import metabolic pathways to “unnatural” contexts (Paddon et al. 2013) but, together with protein engineering (Cao et al. 2020) or directed evolution (Yang and Arnold 2021), empowers the engineering of altogether novel pathways (Scheffen et al. 2021) and enzymes (Dick et al. 2019) that can spawn “unnatural” products (Kan et al. 2016; Park et al. 2019; Robertson et al. 2021). For all practical purposes, synthetic biologists are modern incarnations of alchemists of yore, the only real difference, perhaps, being that we have found our philosopher’s stone in DNA. This is best evidenced by Eculizumab, a monoclonal antibody sold under the brand name Soliris, that costs a whopping 21,000 \$/g (Kelley 2009) more than 300 times higher than the price of gold (60 \$/g) (London metal exchange, 2021). In theory, knowing the DNA sequence of the gene coding for the antibody should allow an enterprising modern alchemist to transmute cheap media into “liquid gold”. In practice, however, it is easier said than done and this largely contributes to the bloated price. As the exogenous DNA does not bring any functional utility to the host organism and competes with the native program of procreation, it becomes an “extra” metabolic burden that introduces tradeoffs between the host program (growth) and the exogenous program (production). These tradeoffs manifest in a variety of ways but the final consequence is invariably a decrease in the growth rate of cells that produce (producer) and reduced fitness as a “cost” of producing the exogenous protein. In addition to the metabolic cost, the product or an intermediate might be toxic for the host organism and further decrease the

growth. This in and of itself is not the problem. The problem is that the DNA replication machinery is not perfect and there are random mutations. Most of these mutations are neutral or detrimental to the host cell but some, notably those that free the cell from metabolic burden and/or toxicity of exogenous production, are advantageous. This means that any cell that “escapes” from producing the protein (escaper) becomes more fit than the producers and consequently spreads in the population. Over time, these escapers negatively impact the yield and can compromise the quality of the product.

This situation is further exacerbated at the industrial scale where huge volumes of cultures at high cell densities are required to obtain profitable yields (increasing the likelihood of emergence of escapers). Genetic heterogeneity has been found to be ubiquitous in industrial scale fermenters. Rugbjerg et al. (2018) studied the consequences of heterogeneity in an industrial context and found it to be the constraining factor to scale production to industrial fermenters.

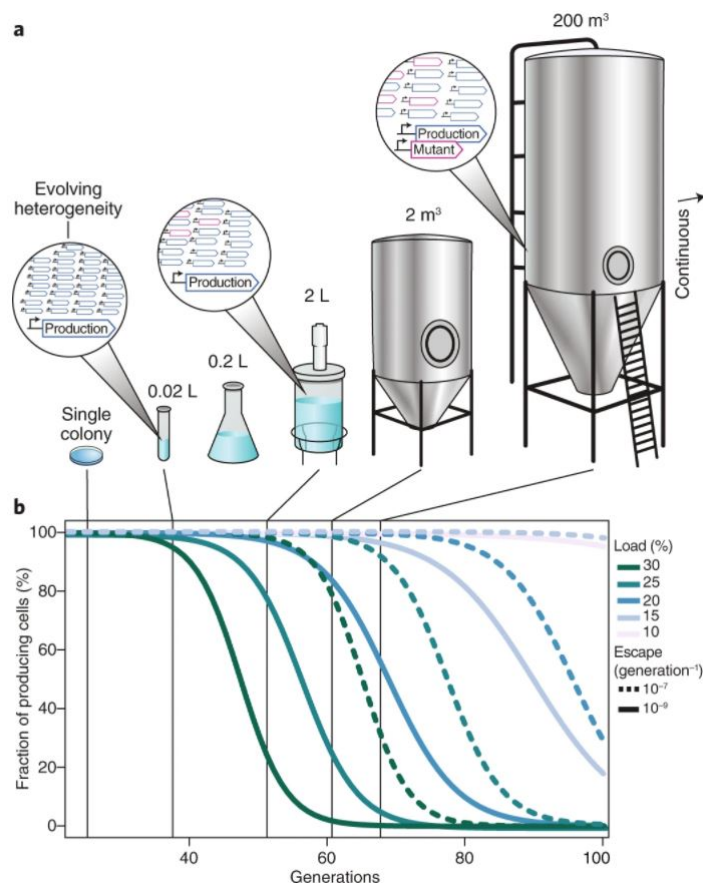


Figure 2.3. Increase in genetic diversity with scale. Solid and dashed lines indicate the fraction of population that produces the POI for different rates of escape. Colours represent the metabolic load imposed by the production. As the metabolic load increases, the production declines earlier. Continuous operation extolls yet greater burden. (Figure adapted from Rugbjerg et al. (2018). All rights belong to Nature.)

In this study, the authors developed models of *E. coli* driven production of mevalonic acid and found that the production was constrained by the emergence of escapers. More specifically, they simulated an industrial bioprocess in the lab that involves at least 60 cell generations to acquire suitable cell density (typically 10^{20} cells in a 200m³ reactor) by making serial dilutions of culture started from a single colony of *E. coli* that produced mevalonic acid. The authors found that from 30 generations onwards there was a steady decrease in the production and a recovery of the growth rate. Through time-lapse deep sequencing, the authors identified that this was due to genetic heterogeneity in the culture (presence of escapers). They used this data to determine the fitness cost (metabolic burden) and escape rate of the producer cells. They then used a mathematical model of population level dynamics vis-a-vis production and escape to predict the limitations of scaling up bioproduction (Rugbjerg et al. 2018). The authors proposed that decreasing the rate of escape, by removing insertion sequence (IS) elements (Griffiths et al. 2000) for example (Csörgő et al. 2012; Choi et al. 2015), or easing the metabolic burden (Ceroni et al. 2018; Gupta et al. 2017) could help scaling up. Notwithstanding, there remain fundamental gaps in our understanding of metabolic burden caused by heterologous expression that limit general solutions (Wu et al. 2016), especially when scaled to production (Rugbjerg et al. 2018), and often necessitate product specific optimization of the strain and/or the expression platform (Van Dien 2013) and tinkering of process variables (for instance, pH, temperature, reactor volume, substrate concentrations, etc.) to obtain relevant titers and profitable yields (Qiao et al. 2017).

However, such metabolic tinkering requires significant capital injection for research and development. To sum up, technological advances in our ability to manipulate the genome of the host, understanding of the host metabolism, and unravelling of the biochemical pathways from a wide spectrum of sources, have empowered us to produce diverse chemicals from living organisms (bioproduction), however, scaling these processes to industrial scale constitutes as the main technological and economic hurdle to overcome for widespread commercialization of bioproduction.

Growth production tradeoffs

Typically, exogenous production has a negative impact on the growth rate of the host organism and consequently on the fitness, primarily because producers direct resources towards the production of useless products that could have been otherwise utilized for cell's growth and proliferation. This constitutes as a metabolic burden meaning that the production comes at a cost. The cost being a combination of all the different resources for which the exogenous program competes with the native program of the cell.

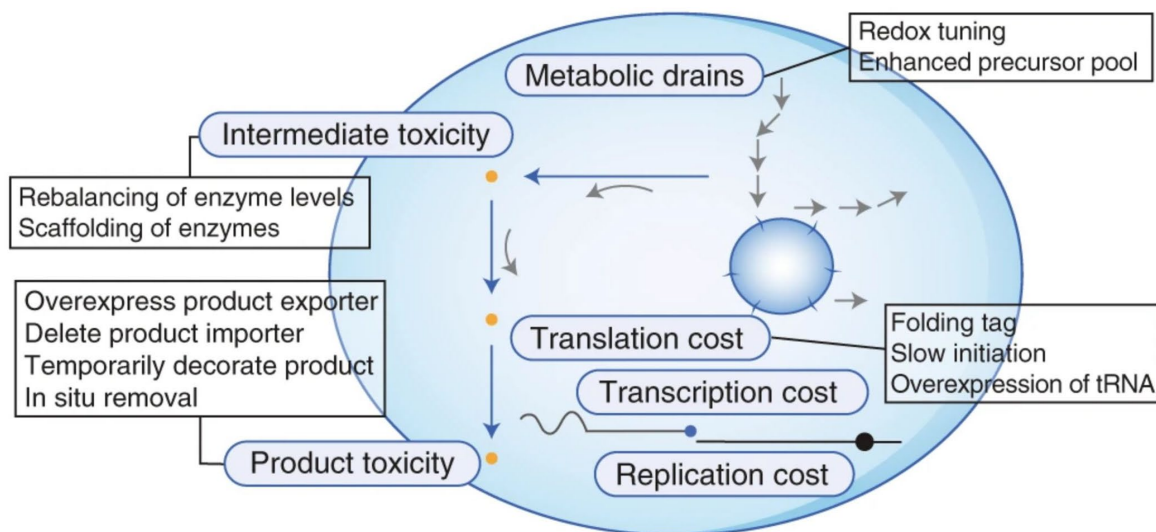


Figure 2.4. Various costs of production. From replicative burden to metabolic toxicity. Ellipsoid bubbles represent the costs while the rectangular boxes indicate proposed solutions to ameliorate the competition/ reduce burden. (Figure adapted from Rugbjerg and Sommer (2019). All rights belong to Nature.)

The most obvious, and perhaps the least problematic of all, is the cost of replicating the exogenous DNA. Metabolically speaking, the cost of replication of DNA is negligible when compared to the total energy consumption of the cell (Bentley et al. 1990, Karim et al. 2012). However, in case of very high copy plasmids in *E. coli*, scarcity of DNA polymerases, depletion of the nucleotide pool and “hoarding” of initiation factors might lead to a significant replication cost (Yano et al. 2016). In this study, the authors found that the retention of plasmids could be improved if either the proteins of the replication machinery were overexpressed or, somewhat counterintuitively, the *ori* sequence of the plasmid was altered to make its interaction with the replication machinery less stable.

The next source of competition comes at the transcriptional level where the exogenous and native programs vie for the transcriptional machinery of the cell. This includes unbound RNA polymerases, pre-initiation complexes, ribonucleotides and other factors. However, it is hard to disentangle the cost of transcription from the cost of translation that breaks down to competition for free ribosomes and charged tRNAs. A dedicated study to address this question in yeast was reported by Kafri et al. (2016). In the study, the authors integrated a fluorescent protein in different copy numbers and correlated the decrease in growth rate with the protein content and the number of copies. The authors further demonstrated by expressing a destabilized mRNA (leading to the same amount of transcription but reduced translation) that translation is the larger contributor to the burden although the transcription costs are not negligible. The authors also

showed that by changing the growth conditions, one could change the rate limiting step from translation (complete media and low nitrogen potentially because of ribosome unavailability and scarcity of charged tRNAs, respectively) to transcription (low phosphorus potentially because of depletion of the ribonucleotide pool). It must be noted, however, that the authors could not observe the consequences of accumulation of mRNA and its role on regulating transcription (Timmers and Tora 2018). Nevertheless, the authors claimed that as the amount of exogenous production increased as a percentage of the total proteome, the growth rate decreased (1% decrease in growth rate for every 2% of the proteome). Similar correlations between protein production and growth rate have been reported for *E. coli* (M. Scott et al. 2010). Both Kafri et al. (2016) and Scott et al. (2010) reported that the intracellular concentration of the exogenous protein became toxic at levels greater than 30% of the total proteome.

Even at lower concentrations, the exogenous protein may become toxic if it requires additional post-translational modifications for proper folding, the availability of chaperones becoming limiting (Geiler-Samerotte et al. 2011) in this case. This problem mostly concerns eukaryotic platforms as bacteria lack the molecular apparatus to make these modifications (Baneyx and Mujacic 2004) although genetically engineered bacteria have been reported that are capable of post-translational modifications (Nguyen et al. 2011; Siller et al. 2010). The situation worsens if the protein is hard to fold or secreted (both requiring the endoplasmic reticulum pathway of translation), often triggering endoplasmic reticulum (ER) stress which in turn results in activation of the unfolded protein response (UPR) that tries to ameliorate the stress by overexpressing chaperones (McMillan, Gething, and Sambrook 1994). If the levels of expression are sufficiently high to not be addressed by the UPR, endoplasmic reticulum associated degradation of unfolded proteins is triggered (ERAD) (Haynes, Titus, and Cooper 2004). Unfolded proteins, in and of themselves, are toxic to the cell because they trigger misfolding of properly folded physiological proteins and therefore their degradation becomes necessary (Haynes, Titus, and Cooper 2004). At still higher levels of expression, analogous to the strength of promoters used to drive industrial production, the proteasome of the cell is saturated leading to accumulation of toxic unfolded proteins. At this point, cells may commit apoptosis (Tabas and Ron 2011).

The fitness is further reduced if the final product or if one of the intermediates is toxic to the cell. This is typically the case for enzymes that have catalytic activity in the host cell or when the exogenous metabolic pathways interfere with the native pathways. The scope of this interference can be general (competing for ATP, NADPH (Yu et al. 2018; Zhao et al. 2018), acetyl-CoA etc.) or specific (competition for intermediate substrates like heme (Michener, Nielsen, and Smolke 2012) or glutamate (Raab et al. 2010)). Specific crosstalk between the exogenous pathway and the native ones, can be predicted and addressed with the help of models of biochemical

networks (Guimera and Amaral 2005). However, if left unaddressed, these may lead to unintentional redirection of the metabolic flux towards undesirable (and often toxic) intermediates (Eiteman and Altman 2006). This, in addition to increasing burden and decreasing yield, might lead to compromised product quality. In conclusion, a number of factors contribute towards growth-production tradeoffs (**Figure 2.4**) in exogenous production in living organisms.

The industry has predominately opted for an umbrella strategy that diminishes the vagaries of product specific metabolic burden and the ensuing growth production trade-offs by temporally separating growth from production in a two-step fermentation process (also called batch or fed-batch operation). While this strategy possesses several advantages that ensure higher productivity, like bulk production, strong existing heuristics, and low susceptibility to mutations (Burg et al. 2016), it still remains suboptimal in time due to reactor downtime and the time spent during growth phase (Rathore et al. 2015) and is plagued with lot-to-lot variability (Langer and Rader 2014). Over recent years, several experts from academia and industry have advocated for the adoption of continuous operation (Croughan, Konstantinov, and Cooney 2015; Warikoo et al. 2012; Walther et al. 2015). Despite, evident benefits (nonstop production, higher productivity, optimality in time, wherewithal to produce labile molecules, and real time quality control (Xu et al. 2017)), adoption at production scale has been sparse but pilot studies indicate improved performance (Cankorur-Cetinkaya et al. 2018; Warikoo et al. 2012; Xu et al. 2017; Godawat et al. 2015; Steinebach et al. 2017). There are several reasons for this hesitancy encompassing technical (development of new infrastructure and standards) and financial (risk and investment in existing infrastructure) aspects as well as perceived biological limitations (severe sensitivity of continuous operation to growth production tradeoffs).

References

- Baneyx, François, and Mirna Mujacic. 2004. "Recombinant Protein Folding and Misfolding in Escherichia Coli." *Nature Biotechnology* 22 (11): 1399–1408.
- Basan, Markus, Tomoya Honda, Dimitris Christodoulou, Manuel Hörl, Yu-Fang Chang, Emanuele Leoncini, Avik Mukherjee, et al. 2020. "A Universal Trade-off between Growth and Lag in Fluctuating Environments." *Nature* 584 (7821): 470–74.
- Baumschlager, Armin, and Mustafa Khammash. 2021. "Synthetic Biological Approaches for Optogenetics and Tools for Transcriptional Light-Control in Bacteria." *Advanced Biology*, 2000256.
- Bertaux, François, Sebastián Sosa Carrillo, Achille Fraisse, Chetan Aditya, Mariela Furstenheim, and Gregory Batt. 2020. "Enhancing Multi-Bioreactor Platforms for Automated Measurements and Reactive Experiment Control." *BioRxiv*.
- Blakemore, David C, Luis Castro, Ian Churcher, David C Rees, Andrew W Thomas, David M Wilson, and Anthony Wood. 2018. "Organic Synthesis Provides Opportunities to Transform Drug Discovery." *Nature Chemistry* 10 (4): 383–94.
- Burg, Jonathan M, Charles B Cooper, Zhixia Ye, Benjamin R Reed, Eirik A Moreb, and Michael D Lynch. 2016. "Large-Scale Bioprocess Competitiveness: The Potential of Dynamic Metabolic Control in Two-Stage Fermentations." *Current Opinion in Chemical Engineering* 14: 121–36.
- Cankorur-Cetinkaya, Ayca, Nathalie Narraïdoo, Ceyda Kasavi, Nigel K H Slater, David B Archer, and Stephen G Oliver. 2018. "Process Development for the Continuous Production of Heterologous Proteins by the Industrial Yeast, *Komagataella Phaffii*." *Biotechnology and Bioengineering* 115 (12): 2962–73.
- Cao, Longxing, Inna Goresnik, Brian Coventry, James Brett Case, Lauren Miller, Lisa Kozodoy, Rita E Chen, et al. 2020. "De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors." *Science* 370 (6515): 426–31.
- Carrasco-López, César, Sergio A García-Echauri, Therese Kichuk, and José L Avalos. 2020. "Optogenetics and Biosensors Set the Stage for Metabolic Cybergenetics." *Current Opinion in Biotechnology* 65: 296–309.
- Ceroni, Francesca, Alice Boo, Simone Furini, Thomas E Gorochowski, Olivier Borkowski, Yaseen N Ladak, Ali R Awan, Charlie Gilbert, Guy-Bart Stan, and Tom Ellis. 2018. "Burden-Driven Feedback Control of Gene Expression." *Nature Methods* 15 (5): 387–93.
- Chait, Remy, Jakob Ruess, Tobias Bergmiller, Gašper Tkačik, and C\ualin C Guet. 2017. "Shaping Bacterial Population Behavior through Computer-Interfaced Control of Individual Cells." *Nature Communications* 8 (1): 1–11.
- Cho, Raymond J., Michael J. Campbell, Elizabeth A. Winzeler, Lars Steinmetz, Andrew Conway, Lisa Wodicka, Tyra G. Wolfsberg, et al. 1998. "A Genome-Wide Transcriptional Analysis of the Mitotic Cell Cycle." *Molecular Cell* 2 (1): 65–73.
- Choi, Jae Woong, Sung Sun Yim, Min Jeong Kim, and Ki Jun Jeong. 2015. "Enhanced Production of Recombinant Proteins with *Corynebacterium Glutamicum* by Deletion of Insertion Sequences (IS Elements)." *Microbial Cell Factories* 14 (1): 207.
- Church, George. 2020. "Technologies for Reading, Writing & Editing Omes." *Journal of Biomolecular Techniques: JBT* 31 (Suppl): S41.
- Croughan, Matthew S., Konstantin B. Konstantinov, and Charles Cooney. 2015. "The Future of Industrial Bioprocessing: Batch or Continuous?" *Biotechnology and Bioengineering* 112 (4):
- Csörgő, Bálint, Tamás Fehér, Edit Tímár, Frederick R Blattner, and György Pósfai. 2012. "Low-Mutation-Rate, Reduced-Genome *Escherichia Coli*: An Improved Host for Faithful Maintenance of Engineered Genetic Constructs." *Microbial Cell Factories* 11 (1): 11.

- Dick, Markus, Nicholas S Sarai, Michael W Martynowycz, Tamir Gonen, and Frances H Arnold. 2019. "Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation." *Journal of the American Chemical Society* 141 (50): 19817–22.
- Dien, Stephen Van. 2013. "From the First Drop to the First Truckload: Commercialization of Microbial Processes for Renewable Chemicals." *Current Opinion in Biotechnology* 24 (6): 1061–68.
- Eiteman, Mark A, and Elliot Altman. 2006. "Overcoming Acetate in Escherichia Coli Recombinant Protein Fermentations." *Trends in Biotechnology* 24 (11): 530–36.
- Erickson, David W, Severin J Schink, Vadim Patsalo, James R Williamson, Ulrich Gerland, and Terence Hwa. 2017. "A Global Resource Allocation Strategy Governs Growth Transition Kinetics of Escherichia Coli." *Nature* 551 (7678): 119–23.
- Fernandez-Rodriguez, Jesus, Felix Moser, Miryoung Song, and Christopher A Voigt. 2017. "Engineering RGB Color Vision into Escherichia Coli." *Nature Chemical Biology* 13 (7): 706–8.
- Fox, Zachary R, Steven Fletcher, Achille Fraisse, Chetan Aditya, Sebastián Sosa-Carrillo, Sébastien Gilles, François Bertaux, Jakob Ruess, and Gregory Batt. 2021. "MicroMator: Open and Flexible Software for Reactive Microscopy." *BioRxiv*.
- Franziska, Hempel, Bozarth Andrew, Lindenkamp Nicole, Klingl Andreas, Zauner Stefan, and others. 2011. "Microalgae as Bioreactors for Bioplastic Production."
- Gassler, Thomas, Michael Sauer, Brigitte Gasser, Michael Egermeier, Christina Troyer, Tim Causon, Stephan Hann, Diethard Mattanovich, and Matthias G Steiger. 2020. "The Industrial Yeast *Pichia Pastoris* Is Converted from a Heterotroph into an Autotroph Capable of Growth on CO₂." *Nature Biotechnology* 38 (2): 210–16.
- Geiler-Samerotte, Kerry A, Michael F Dion, Bogdan A Budnik, Stephanie M Wang, Daniel L Hartl, and D Allan Drummond. 2011. "Misfolded Proteins Impose a Dosage-Dependent Fitness Cost and Trigger a Cytosolic Unfolded Protein Response in Yeast." *Proceedings of the National Academy of Sciences* 108 (2): 680–85.
- Geller, Stephanie H, Enoch B Antwi, Barbara Di Ventura, and Megan N McClean. 2019. "Optogenetic Repressors of Gene Expression in Yeasts Using Light-Controlled Nuclear Localization." *Cellular and Molecular Bioengineering* 12 (5): 511–28.
- Gleizer, Shmuel, Roe Ben-Nissan, Yinon M Bar-On, Niv Antonovsky, Elad Noor, Yehudit Zohar, Ghil Jona, et al. 2019. "Conversion of Escherichia Coli to Generate All Biomass Carbon from CO₂." *Cell* 179 (6): 1255–63.
- Godawat, Rahul, Konstantin Konstantinov, Mahsa Rohani, and Veena Warikoo. 2015. "End-to-End Integrated Fully Continuous Production of Recombinant Monoclonal Antibodies." *Journal of Biotechnology* 213: 13–19.
- Griffiths, A J, J H Miller, D T Suzuki, R C Lewontin, and W M Gelbart. 2000. "Bacterial Insertion Sequences." WH Freeman.
- Guimera, Roger, and Luis A Nunes Amaral. 2005. "Functional Cartography of Complex Metabolic Networks." *Nature* 433 (7028): 895–900.
- Gupta, Apoorv, Irene M Brockman Reizman, Christopher R Reisch, and Kristala L J Prather. 2017. "Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent Quorum-Sensing Circuit." *Nature Biotechnology* 35 (3): 273–79.
- Harrigan, Patrick, Hiten D Madhani, and Hana El-Samad. 2018. "Real-Time Genetic Compensation Defines the Dynamic Demands of Feedback Control." *Cell* 175 (3): 877–86.
- Haynes, Cole M, Eric A Titus, and Antony A Cooper. 2004. "Degradation of Misfolded Proteins Prevents ER-Derived Oxidative Stress and Cell Death." *Molecular Cell* 15 (5): 767–76.
- Johnson, Heath E, Nareg J V Djabrayan, Stanislav Y Shvartsman, and Jared E Toettcher. 2020. "Optogenetic Rescue of a Patterning Mutant." *Current Biology* 30 (17): 3414–24.

- Kafri, Moshe, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. 2016. "The Cost of Protein Production." *Cell Reports* 14 (1): 22–31.
- Kan, S B Jennifer, Russell D Lewis, Kai Chen, and Frances H Arnold. 2016. "Directed Evolution of Cytochrome c for Carbon--Silicon Bond Formation: Bringing Silicon to Life." *Science* 354 (6315): 1048–51.
- Kelley, Brian. 2009. "Industrialization of MAb Production Technology: The Bioprocessing Industry at a Crossroads." In *MAbs*, 1:443–52.
- Lalwani, Makoto A, Samantha S Ip, Cesar Carrasco-Lopez, Catherine Day, Evan M Zhao, Hinako Kawabe, and José L Avalos. 2021. "Optogenetic Control of the Lac Operon for Bacterial Chemical and Protein Production." *Nature Chemical Biology* 17 (1): 71–79.
- Langer, ERIC S, and RONALD A Rader. 2014. "Continuous Bioprocessing and Perfusion: Wider Adoption Coming as Bioprocessing Matures." *Bioprocess J* 13 (1).
- Liu, Ping, Bojun Chen, and Zhao-Wen Wang. 2020. "GABAergic Motor Neurons Bias Locomotor Decision-Making in *C. Elegans*." *Nature Communications* 11 (1): 1–19.
- Liu, Quanli, Tao Yu, Xiaowei Li, Yu Chen, Kate Campbell, Jens Nielsen, and Yun Chen. 2019. "Rewiring Carbon Metabolism in Yeast for High Level Production of Aromatic Chemicals." *Nature Communications* 10 (1): 1–13.
- Lugagne, Jean-Baptiste, Sebastián Sosa Carrillo, Melanie Kirch, Agnes Köhler, Gregory Batt, and Pascal Hersen. 2017. "Balancing a Genetic Toggle Switch by Real-Time Feedback Control and Periodic Forcing." *Nature Communications* 8 (1): 1–8.
- Luo, Xiaozhou, Michael A Reiter, Leo d'Espaux, Jeff Wong, Charles M Denby, Anna Lechner, Yunfeng Zhang, et al. 2019. "Complete Biosynthesis of Cannabinoids and Their Unnatural Analogues in Yeast." *Nature* 567 (7746): 123–26.
- McElroy, Christopher, and Stefan Jennewein. 2018. "Taxol@biosynthesis and Production: From Forests to Fermenters." In *Biotechnology of Natural Products*, 145–85. Springer.
- McMillan, D Randy, Mary-Jane Gething, and Joseph Sambrook. 1994. "The Cellular Response to Unfolded Proteins: Intercompartmental Signaling." *Current Opinion in Biotechnology* 5 (5): 540–45.
- Menolascina, Filippo, Gianfranco Fiore, Emanuele Orabona, Luca De Stefano, Mike Ferry, Jeff Hasty, Mario di Bernardo, and Diego di Bernardo. 2014. "In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks." *PLoS Comput Biol* 10 (5): e1003625.
- Michener, Joshua K, Jens Nielsen, and Christina D Smolke. 2012. "Identification and Treatment of Heme Depletion Attributed to Overexpression of a Lineage of Evolved P450 Monooxygenases." *Proceedings of the National Academy of Sciences* 109 (47): 19504–9.
- Milias-Argeitis, Andreas, Marc Rullan, Stephanie K. Aoki, Peter Buchmann, and Mustafa Khammash. 2016. "Automated Optogenetic Feedback Control for Precise and Robust Regulation of Gene Expression and Cell Growth." *Nature Communications* 7 (1): 1–11.
- Milias-Argeitis, Andreas, Sean Summers, Jacob Stewart-Ornstein, Ignacio Zuleta, David Pincus, Hana El-Samad, Mustafa Khammash, and John Lygeros. 2011. "In Silico Feedback for in Vivo Regulation of a Gene Expression Circuit." *Nature Biotechnology* 29 (12): 1114–16.
- Nguyen, Van Dat, Feras Hatahet, Kirsi E H Salo, Eveliina Enlund, Chi Zhang, and Lloyd W Ruddock. 2011. "Pre-Expression of a Sulfhydryl Oxidase Significantly Increases the Yields of Eukaryotic Disulfide Bond Containing Proteins Expressed in the Cytoplasm of *E. Coli*." *Microbial Cell Factories* 10 (1): 1–13.
- Ohlendorf, Robert, Roe R Vidavski, Avigdor Eldar, Keith Moffat, and Andreas Möglich. 2012. "From Dusk till Dawn: One-Plasmid Systems for Light-Regulated Gene Expression." *Journal of Molecular Biology* 416 (4): 534–42.
- Olson, Evan J, Lucas A Hartsough, Brian P Landry, Raghav Shroff, and Jeffrey J Tabor. 2014. "Characterizing Bacterial Gene Circuit Dynamics with Optically Programmed Gene Expression Signals." *Nature Methods* 11

(4): 449–55.

- Omura, Satoshi. 2011. “Microbial Metabolites: 45 Years of Wandering, Wondering and Discovering.” *Tetrahedron (Oxford. Print)* 67 (35).
- Paddon, Christopher J, P J Westfall, D J Pitera, K Benjamin, K Fisher, D McPhee, M D Leavell, et al. 2013. “High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin.” *Nature* 496 (7446): 528–32.
- Park, Jooyoung, Brinda Selvaraj, Andrew C McShan, Scott E Boyken, Kathy Y Wei, Gustav Oberdorfer, William DeGrado, et al. 2019. “De Novo Design of a Homo-Trimeric Amantadine-Binding Protein.” *Elife* 8: e47839.
- Pedone, Elisa, Irene De Cesare, Criseida G Zamora-Chimal, David Haener, Lorena Postiglione, Antonella La Regina, Barbara Shannon, et al. 2021. “Cheetah: A Computational Toolkit for Cybergenetic Control.” *ACS Synthetic Biology* 10 (5): 979–89.
- Perkins, Melinda Liu, Dirk Benzinger, Murat Arcak, and Mustafa Khammash. 2020. “Cell-in-the-Loop Pattern Formation with Optogenetically Emulated Cell-to-Cell Signaling.” *Nature Communications* 11 (1): 1–10.
- Pouzet, Sylvain, Alvaro Banderas, Matthias Le Bec, Thomas Lautier, Gilles Truan, and Pascal Hersen. 2020. “The Promise of Optogenetics for Bioproduction: Dynamic Control Strategies and Scale-up Instruments.” *Bioengineering* 7 (4): 151.
- Qiao, Kangjian, Thomas M Wasylenko, Kang Zhou, Peng Xu, and Gregory Stephanopoulos. 2017. “Lipid Production in *Yarrowia Lipolytica* Is Maximized by Engineering Cytosolic Redox Metabolism.” *Nature Biotechnology* 35 (2): 173–77.
- Raab, Andreas M, Gabi Gebhardt, Natalia Bolotina, Dirk Weuster-Botz, and Christine Lang. 2010. “Metabolic Engineering of *Saccharomyces Cerevisiae* for the Biotechnological Production of Succinic Acid.” *Metabolic Engineering* 12 (6): 518–25.
- Raj, Arjun, and Alexander Van Oudenaarden. 2008. “Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences.” *Cell* 135 (2): 216–26.
- Rathore, Anurag S, Harshit Agarwal, Abhishek Kumar Sharma, Mili Pathak, and SJPB Muthukumar. 2015. “Continuous Processing for Production of Biopharmaceuticals.” *Preparative Biochemistry and Biotechnology* 45 (8): 836–49.
- Repina, Nicole A., Alyssa Rosenbloom, Abhirup Mukherjee, David V. Schaffer, and Ravi S. Kane. 2017. “At Light Speed: Advances in Optogenetic Systems for Regulating Cell Signaling and Behavior.” *Annual Review of Chemical and Biomolecular Engineering* 8 (1): 13–39.
- Reynders, Martin, Bryan S Matsuura, Marleen Bérouti, Daniele Simoneschi, Antonio Marzio, Michele Pagano, and Dirk Trauner. 2020. “PHOTACs Enable Optical Control of Protein Degradation.” *Science Advances* 6 (8): eaay5064.
- Robertson, Wesley E, Louise F H Funke, Daniel de la Torre, Julius Fredens, Thomas S Elliott, Martin Spinck, Yonka Christova, et al. 2021. “Sense Codon Reassignment Enables Viral Resistance and Encoded Polymer Synthesis.” *Science* 372 (6546): 1057–62.
- Romano, Edoardo, Armin Baumschlager, Emir Bora Akmeriç, Navaneethan Palanisamy, Moustafa Houmani, Gregor Schmidt, Mehmet Ali Öztürk, Leonard Ernst, Mustafa Khammash, and Barbara Di Ventura. 2021. “Engineering AraC to Make It Responsive to Light Instead of Arabinose.” *Nature Chemical Biology*, 1–11.
- Rugbjerg, Peter, Nils Myling-Petersen, Andreas Porse, Kira Sarup-Lytzen, and Morten O A Sommer. 2018. “Diverse Genetic Error Modes Constrain Large-Scale Bio-Based Production.” *Nature Communications* 9 (1): 1–14.
- Rugbjerg, Peter, and Morten O A Sommer. 2019. “Overcoming Genetic Heterogeneity in Industrial Fermentations.” *Nature Biotechnology* 37 (8): 869–76.
- Rullan, Marc, Dirk Benzinger, Gregor W. Schmidt, Andreas Miliás-Argeitis, and Mustafa Khammash. 2018. “An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation.”

Molecular Cell 70 (4): 745–756.e6.

- Scheffen, Marieke, Daniel G Marchal, Thomas Beneyton, Sandra K Schuller, Melanie Klose, Christoph Diehl, Jessica Lehmann, et al. 2021. "A New-to-Nature Carboxylation Module to Improve Natural and Synthetic CO₂ Fixation." *Nature Catalysis* 4 (2): 105–15.
- Scott, Matthew, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. 2010. "Interdependence of Cell Growth and Gene Expression: Origins and Consequences." *Science* 330 (6007): 1099–1102.
- Seong, Jihye, and Michael Z Lin. 2021. "Optobiochemistry: Genetically Encoded Control of Protein Activity by Light." *Annual Review of Biochemistry* 90.
- Shao, Yangyang, Ning Lu, Chen Cai, Fan Zhou, Shanshan Wang, Zhihu Zhao, Guoping Zhao, Jin-Qiu Zhou, Xiaoli Xue, and Zhongjun Qin. 2019. "A Single Circular Chromosome Yeast." *Cell Research* 29 (1): 87–89.
- Siller, Efraín, Diane C DeZwaan, John F Anderson, Brian C Freeman, and José M Barral. 2010. "Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency." *Journal of Molecular Biology* 396 (5): 1310–18.
- Smith, Jon, Marc Lipsitch, and Jeffrey W Almond. 2011. "Vaccine Production, Distribution, Access, and Uptake." *The Lancet* 378 (9789): 428–38.
- Steinebach, Fabian, Nicole Ulmer, Moritz Wolf, Lara Decker, Veronika Schneider, Ruben Wälchli, Daniel Karst, Jonathan Souquet, and Massimo Morbidelli. 2017. "Design and Operation of a Continuous Integrated Monoclonal Antibody Production Process." *Biotechnology Progress* 33 (5): 1303–13.
- Tabas, Ira, and David Ron. 2011. "Integrating the Mechanisms of Apoptosis Induced by Endoplasmic Reticulum Stress." *Nature Cell Biology* 13 (3): 184–90.
- Timmers, H Th Marc, and László Tora. 2018. "Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation." *Molecular Cell* 72 (1): 10–17.
- Toettcher, Jared E, Delquin Gong, Wendell A Lim, and Orion D Weiner. 2011. "Light-Based Feedback for Controlling Intracellular Signaling Dynamics." *Nature Methods* 8 (10): 837–39.
- Toettcher, Jared E, Orion D Weiner, and Wendell A Lim. 2013. "Using Optogenetics to Interrogate the Dynamic Control of Signal Transmission by the Ras/Erk Module." *Cell* 155 (6): 1422–34.
- Uhlendorf, Jannis, Agnès Miermont, Thierry Delaveau, Gilles Charvin, François Fages, Samuel Bottani, Gregory Batt, and Pascal Hersen. 2012. "Long-Term Model Predictive Control of Gene Expression at the Population and Single-Cell Levels." *Proceedings of the National Academy of Sciences* 109 (35): 14271–76.
- Vary, Patricia S, Rebekka Biedendieck, Tobias Fuerch, Friedhelm Meinhardt, Manfred Rohde, Wolf-Dieter Deckwer, and Dieter Jahn. 2007. "Bacillus Megaterium—from Simple Soil Bacterium to Industrial Protein Production Host." *Applied Microbiology and Biotechnology* 76 (5): 957–67.
- Walther, Jason, Rahul Godawat, Chris Hwang, Yuki Abe, Andrew Sinclair, and Konstantin Konstantinov. 2015. "The Business Impact of an Integrated Continuous Biomanufacturing Platform for Recombinant Protein Production." *Journal of Biotechnology* 213: 3–12.
- Wang, Ping, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. 2010. "Robust Growth of Escherichia Coli." *Current Biology* 20 (12): 1099–1103.
- Warikoo, Veena, Rahul Godawat, Kevin Brower, Sujit Jain, Daniel Cummings, Elizabeth Simons, Timothy Johnson, et al. 2012. "Integrated Continuous Production of Recombinant Therapeutic Proteins." *Biotechnology and Bioengineering* 109 (12): 3018–29.
- Wu, Gang, Qiang Yan, J Andrew Jones, Yinjie J Tang, Stephen S Fong, and Mattheos A G Koffas. 2016. "Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications." *Trends in Biotechnology* 34 (8): 652–64.

- Xu, Sen, John Gavin, Rubin Jiang, and Hao Chen. 2017. "Bioreactor Productivity and Media Cost Comparison for Different Intensified Cell Culture Processes." *Biotechnology Progress* 33 (4): 867–78.
- Yan, Gang, Petra E Vértes, Emma K Towilson, Yee Lian Chew, Denise S Walker, William R Schafer, and Albert-László Barabási. 2017. "Network Control Principles Predict Neuron Function in the *Caenorhabditis Elegans* Connectome." *Nature* 550 (7677): 519–23.
- Yang, Yang, and Frances H Arnold. 2021. "Navigating the Unnatural Reaction Space: Directed Evolution of Heme Proteins for Selective Carbene and Nitrene Transfer." *Accounts of Chemical Research* 54 (5): 1209–25.
- Yu, Tao, Yongjin J Zhou, Mingtao Huang, Quanli Liu, Rui Pereira, Florian David, and Jens Nielsen. 2018. "Reprogramming Yeast Metabolism from Alcoholic Fermentation to Lipogenesis." *Cell* 174 (6): 1549–58.
- Zhang, Xiafei, and Marie A Elliot. 2019. "Unlocking the Trove of Metabolic Treasures: Activating Silent Biosynthetic Gene Clusters in Bacteria and Fungi." *Current Opinion in Microbiology* 51: 9–15.
- Zhao, Evan M., Yanfei Zhang, Justin Mehl, Helen Park, Makoto A. Lalwani, Jared E. Toettcher, and José L. Avalos. 2018. "Optogenetic Regulation of Engineered Cellular Metabolism for Microbial Chemical Production." *Nature* 555 (7698): 683–87.

Chapter 3

DIFFERENTIATION SYSTEM FOR MICROBIAL CONSORTIA IN YEAST

“An answer that was right the first time may not be right again the second.”

Italo Calvino

Preface

Portions of this chapter as well as some figures have been adapted from Aditya *et al.* (2021).

Introduction

The evolutionary transition from single cell to multicellular organisms marked a critical turning point in biology (Maynard Smith and Szathmary 1997). Such shift relied on optimising fitness and productivity through division of labour and functional specialization (Ispolatov, Ackermann, and Doebeli 2012; Rueffler, Hermisson, and Wagner 2012). The same principle can be extended to microorganisms living together to form microbial communities or consortia. Engineered microbial consortia hold enormous potential and have been hailed as the next frontier in synthetic biology (Brenner, You, and Arnold 2008; Rapp, Jenkins, and Betenbaugh 2020). Proof of concept studies have concretely established applications in bioproduction (Zhou *et al.* 2015; Li, Wang, and Zhang 2019), bioremediation (Zhuang *et al.* 2014; Zhang *et al.* 2021), and soil microbiome engineering (Panke-Buisse *et al.* 2015), paving the way for therapeutic applications using human microbiome engineering (Inda *et al.* 2019; Chen *et al.* 2020).

Particularly in the context of bioproduction, microbial consortia possess several advantages over traditional monocultures as functional specialization allows metabolic burden to be shared across different species. Diversification thus allows yields to be optimised simply through tuning consortia composition, rather than re-engineering the strain itself (Wu *et al.* 2015). Moreover, by including multiple species, toxic by-products produced by one species can be sequestered and/or metabolised by another, thereby improving the efficiency of the overall process (LaSarre *et al.* 2017). Microbial consortia are typically generated by culturing two or more species together. Such co-culturing approaches rely on various inter-species interactions to ensure the co-existence of different species like mutualism (Shou, Ram, and Vilar 2007), emergent cooperation

(Wintermute and Silver 2010), competitive amensalism (Fedorec et al. 2021), and predation (Balagaddé et al. 2008). Despite considerable advances in our ability to engineer microbial consortia (Zhou et al. 2015; Zhuang et al. 2014; Zhang et al. 2021; Roell et al. 2019; Karkaria, Fedorec, and Barnes 2021; Kong et al. 2018) and in our understanding of community interactions (Wintermute and Silver 2010; Shou, Ram, and Vilar 2007; Kylilis et al. 2018; Kong et al. 2018; Balagaddé et al. 2008), dynamic control of consortium composition remains a key challenge in the field (Roell et al. 2019). Typically, stable consortia are based on syntrophic or quorum sensing interactions that, albeit being autonomous, remain critically dependent on cell density, thus limiting the applicability for dynamic control. Additionally, scaling the consortium to include more than two species requires non-trivial considerations that may not lead to stable co-existence (Karkaria, Fedorec, and Barnes 2021). In light of these limitations, an externally controllable differentiation system could be well suited to address this challenge.

In recent years, advances in biological control have come from coupling computers with growing cells carrying the engineered system, made possible by special platforms that integrate biological systems with the computer via a feedback loop (Miliás-Argeitis et al. 2011; Toettcher et al. 2011; Menolascina et al. 2014; Lugagne et al. 2017; Carrasco-López et al. 2020). The development of optogenetics, i.e. the use of light to trigger cellular processes, has contributed significantly to control applications by increasing the spatiotemporal resolution of the control signal. Control of protein expression using light has been demonstrated both at the population level (Miliás-Argeitis et al. 2011; Olson et al. 2014; Bertaux et al. 2020) and in single cells (Chait et al. 2017; Rullan et al. 2018; Fox et al. 2021; Perkins et al. 2020). Optogenetics has been used to control cellular processes in other contexts, for instance, signalling dynamics (Toettcher et al. 2011), morphogenesis (Johnson et al. 2020), neuroscience (Liu, Chen, and Wang 2020), bioproduction and metabolic engineering (Lalwani et al. 2021; Zhao et al. 2018). However, control of population dynamics using optogenetics in a multispecies environment is yet to be demonstrated.

Problem statement

To construct a differentiation system in yeast capable of giving user defined fractions of distinct subpopulations upon light stimulation. Certain characteristics would be desirable in such a system, for example, the system should be able to leverage the phenotypic heterogeneity in the population to give rise to graded behaviour, the system should be irreversible in the sense that once differentiated, cells cannot turn back, and last, but not the least, the system should show high efficiency upon light exposure and low background activity (leakage) in the dark.

Strategy

Several possible strategies presented themselves to me during my literature review. Notably, cell-to-cell communication (Youk and Lim 2014) was an early inspiration. Yeast, owing to asymmetric division, have well characterized processes that take place differentially in mothers and daughters (Colman-Lerner, Chin, and Brent 2001; Cosma 2004). *ASH1*, for example, codes for an mRNA that is exclusively translated in the daughter cells to repress expression from HO promoter and was used to create a set of inducible mother specific promoters (Pothoulakis and Ellis 2018). Another source of increasing heterogeneity in gene expression could be the use of mRNA destabilizing elements, like the terminator tMFA2 (Decker and Parker 1993), which generate high variability in gene expression at the population level (Annex 3). I opted against these approaches because none of them, at least in theory, would lead to 100% differentiation. I chose to go with optogenetic recombination to effect differentiation because recombination is an irreversible genetic change. For the choice of recombinase, I decided on Cre as it has been heavily characterized and known to be functional in yeast without toxicity. Optogenetic or photoactivable Cre have been reported in the past (Taslimi et al. 2016; Kawano et al. 2016; Hochrein et al. 2018). However, none were deemed appropriate because of their low efficiency. A recent paper from Gael Yvert lab reported a single chain photoactivable Cre that shows high efficiency and low leakage in the dark and is coded by a single CDS (Duplus-Bottin et al. 2021). Unfortunately, this work was not published at the time, so I decided to use EL222, a thoroughly characterized optogenetic transcription factor (Benzinger and Khammash 2018), driving an intact Cre. In particular, the choice was influenced by the finding that EL222 enabled tuning of gene expression by playing with the intensity of light used for induction (Benzinger and Khammash 2018).

To recapitulate, my strategy was to make a promoter trap that leads to a switch in the gene expressed upon recombination that can be controlled with light. In this way, I hoped to create controllable proportions of recombined and non-recombined cells in a population, called species to highlight that this is a microbial consortia composed of functionally (and genetically) distinct constituents.

Results

Construction of an optogenetic differentiation system and its functional characterization

I designed, constructed and cloned an optogenetic differentiation system consisting of a blue light inducible Cre recombinase under the control of a constitutively expressed optogenetic transcription factor, EL222 (Motta-Mena et al. 2014; Benzinger and Khammash 2018) in *S. cerevisiae*. In order to test the functionality of the system, I designed a “dual reporter” recombination cassette composed of a floxed coding sequence (CDS), coding for a fluorescent reporter (mCerulean) that is transcribed constitutively via a pTDH3 promoter upstream of the first LoxP site. Another CDS, coding for a different fluorescent reporter (mNeonGreen), was added downstream of the second LoxP site (**Figure 3.1a**).

Prior to differentiation, mCerulean is constitutively expressed and mNeonGreen is not. After light induction, Cre is expressed causing a recombination event leading to the expression of mNeonGreen and loss of mCerulean expression. I tested three different EL222 promoters published in Benzinger and Khammash (2018) to drive Cre recombinase, namely pEL222 3x binding sites (bs), pEL222 6x bs and pEL222 5x bs Gal1. Differentiation was said to have occurred if the cells expressed mNeonGreen. This was ascertained by simply applying a threshold on mNeonGreen fluorescence (**Figure 3.2**). Of these three, only EL222 5x bs Gal1 led to low leakage with <10% colonies recombined in the dark (data not shown).

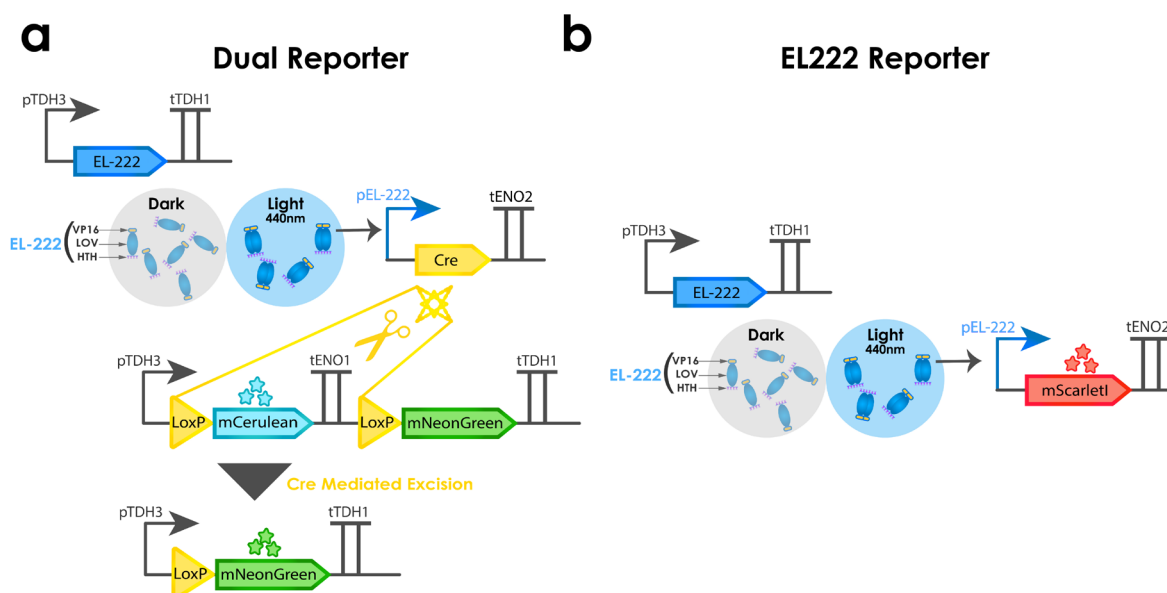


Figure 3.1. **a.** Differentiation system (Dual reporter) and **b.** EL222 activity reporter. EL222 promoter used in both reporters was pEL222 5X bs Gal1.

Cells were cultured to exponential phase and induced with continuous light in the batch reactor. Manual sampling and cytometry measurements revealed that the population fluorescence shifted from cyan to green (**Figure 3.2b**). The same was validated under the microscope by observing cells at different times after induction (data not shown).

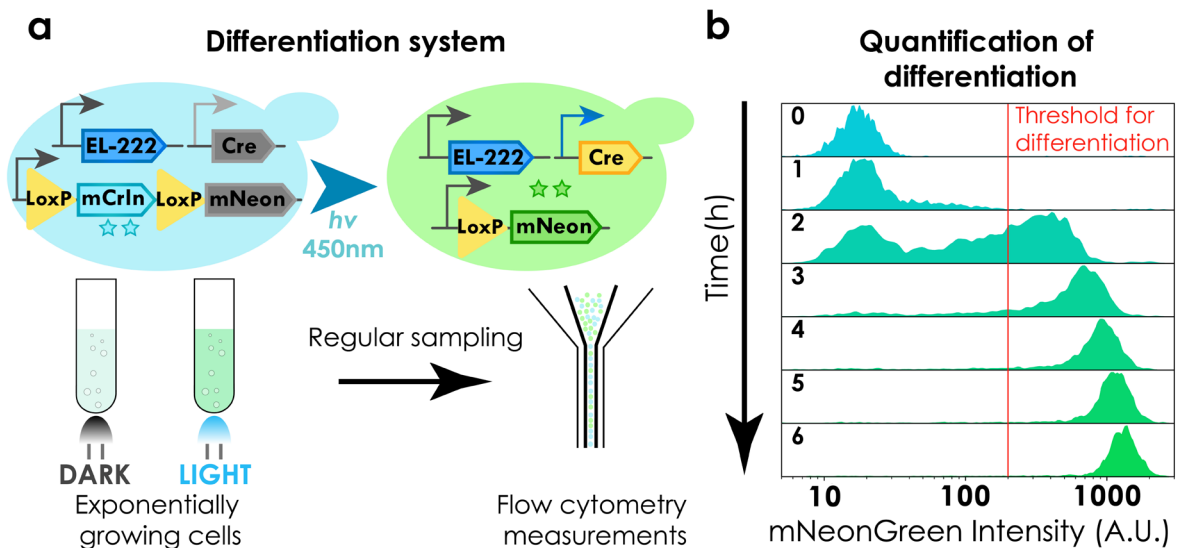


Figure 3.2. **a.** Differentiation system with cellular phenotypes and sampling. **b.** Evolution of population fluorescence under continuous light and quantification of differentiation.

In order to better understand the system, I decided to study the induction behaviour of this promoter. In brief, I put a red fluorescent protein (mScarlet-I) under the control of this promoter (ρ EL222 5X bs Gal1) and characterized response to various light intensities and duty cycles in cells carrying the EL222 reporter in continuous cultures. By duty cycle is meant a quantity between 0 and 1 that reflects the percentage of light shown in a given period. For instance, a duty cycle of 0.5 for a period of 2h signifies 1h of light followed by one hour of darkness. Dynamics of population average fluorescence emerging from various different light intensities and duty cycles are displayed in **Figure 3.3a and 3.3b**, respectively.

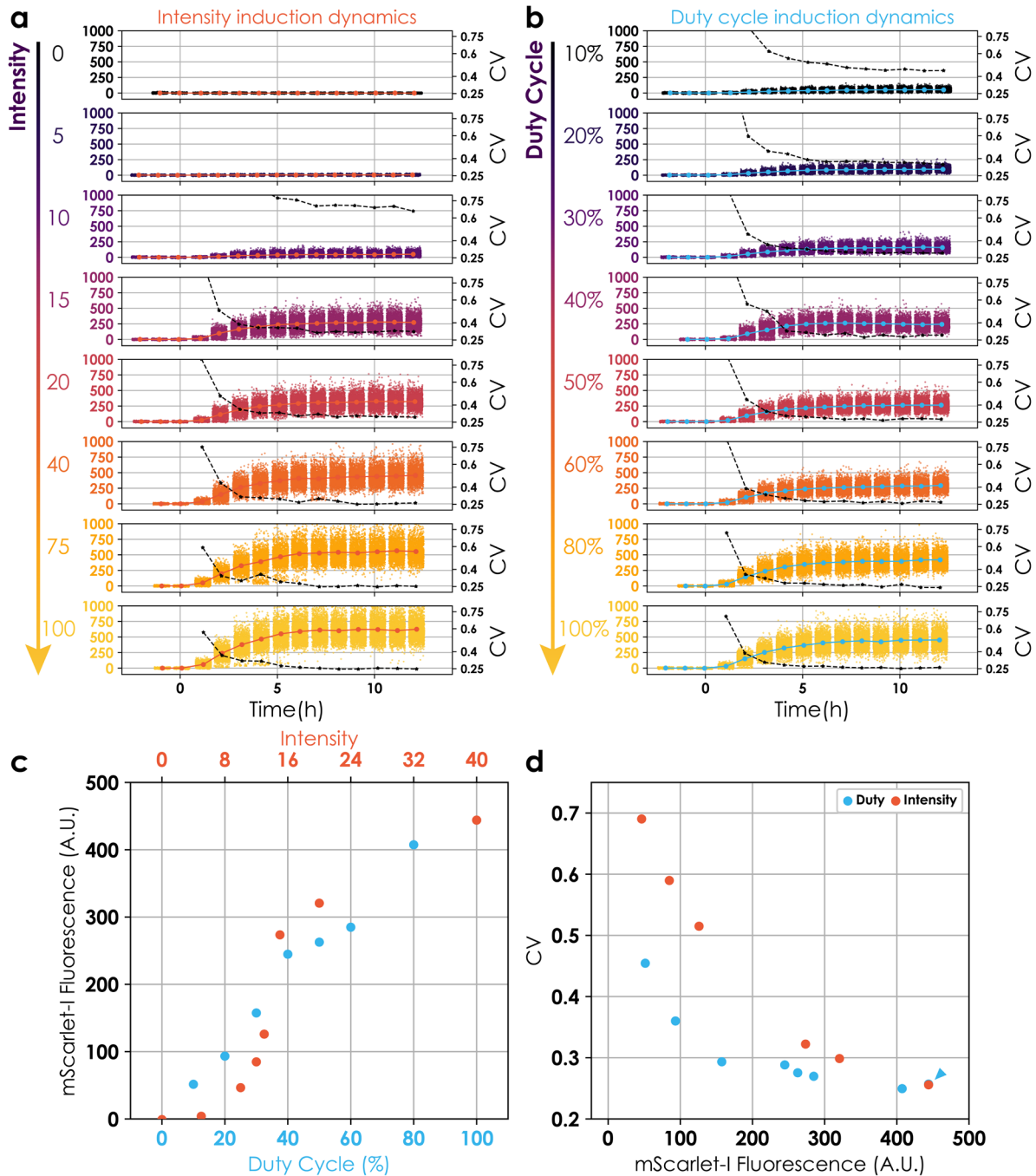


Figure 3.3. Characterizing EL222 induction behaviour. Population level induction dynamics for **a.** intensity modulation and **b.** duty cycle modulation. Small circles represent fluorescence from individual cells. Big circles (red and blue) represent the mean fluorescent values at each time point (connected with red and blue lines, respectively). Black lines marked with stars denote the CV (measured on the right y-axis). CV values prior to induction were not reliable due to low average population fluorescence. **c.** Red circles represent mean population fluorescence for duty cycle modulation and blue circles stand for mean population fluorescence for intensity modulation. Each condition has a single replicate. **d.** CV plotted against mean for intensity modulation (red) and duty cycle modulation (blue).

Overall, we obtained results that are consistent with Benzinger and Khammash (2018):

- ❖ Steady state average fluorescence is linear in the duty cycle and shows a sigmoidal response with increasing intensity of light (**Figure 3.3c**).
- ❖ Cell-to-cell heterogeneity is larger with intensity modulation for the same mean level of fluorescence compared to duty cycle modulation (**Figure 3.3d**).

I note that the maximum levels of expression (~50% of pTDH3 driven constitutive expression) were obtained with an intensity of 100 but at the cost of a slight growth defect. Intensity 40 resulted in slightly lower levels (~40% of pTDH3 driven constitutive expression) but did not exhibit any noticeable growth defect. pTDH3 is used as a reference because it is the strongest native promoter in yeast (Lee et al. 2015).

Having confirmed that the promoter pEL222 5X bs Gal1 behaved in agreement to the findings reported in Benzinger and Khammash (2018) and allowed for modulation of gene expression variability, I was curious to determine how differentiation varied in response to different intensities and duty cycles. For the same, I excited exponentially growing cells in batch in duplicates with single pulses of either fixed duration but at different intensities or fixed intensity but at different durations. These experiments were done in batch because it allowed me to perform multiple induction experiments on the same day. Briefly, overnight cultures could be diluted in a large preculture and partially split to give 16 cultures in falcon tubes that could be induced with different profiles in the batch incubator. Post induction these cultures could be removed from the batch incubator and placed in the dark while the remaining preculture could be used to conduct the next set of experiments.

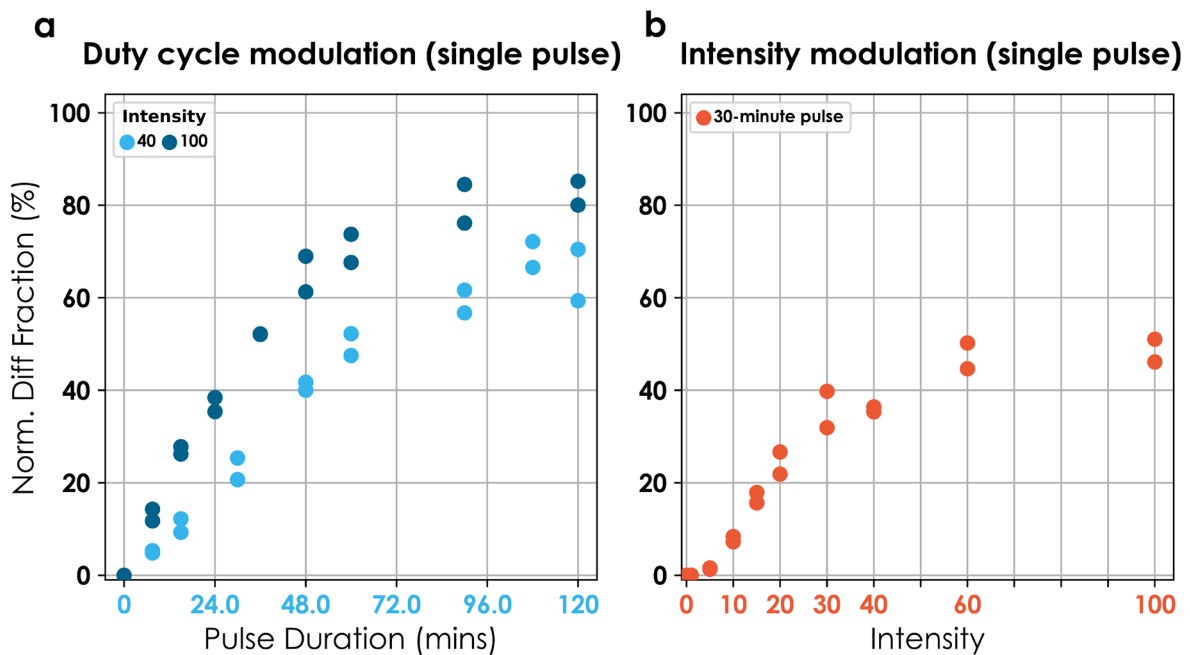


Figure 3.4. Single pulse duration vs intensity modulation of differentiation behaviour. Solid circles are differentiation fractions observed. All conditions have two replicates. Differentiation w.r.t. pulse duration modulation over a period of 2h was assessed for two different intensities, 40 and 100. We noticed that the differentiation fraction increases linearly with pulse duration up to a certain point whence it plateaus. The slope of increase in differentiated fraction with respect to pulse duration was steeper for intensity 100. The linear response range was also smaller for intensity 100. Amplitude modulation led to a more switch like response of the differentiated fraction. Intensity 10 resulted in no significant recombination whereas intensity 30 resulted in 80% differentiation of the maximum intensity (100).

I found that the differentiation dynamics were reminiscent of EL222 induction profiles for different intensities and duty fractions of light (**Figure 3.4**). More precisely, the differentiation fraction increased linearly with the duty cycle up to a point following which it seemed to plateau whereas for the intensity modulation, increase in differentiated fraction was sigmoidal. Interestingly, very low intensities (<5), resulted in no detectable differentiation. For all the results that follow, an intensity of 40 was used for light stimulation because I observed no significant toxicity or growth defect due to expression of EL222, Cre or blue light induction. I also did not notice a growth difference between differentiated and non-differentiated cells (**Figure 3.5**).

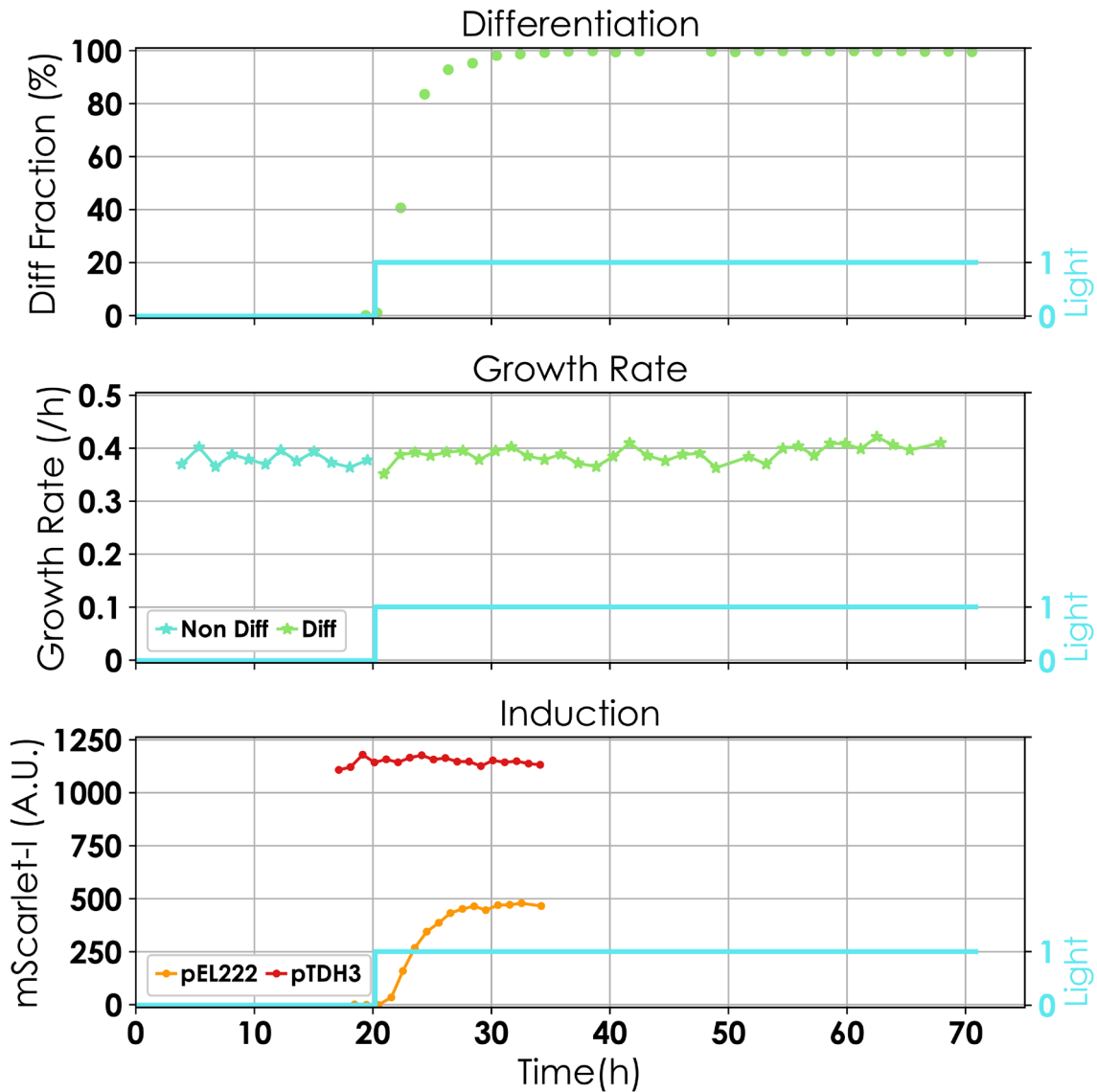


Figure 3.5. No growth defect is found for differentiated cells, growing under blue light and expressing Cre (top and middle panel). The expression strength of the EL22 promoter is approximately 40% of the one of the TDH3 promoter (bottom).

In order to evaluate efficiency and background activity at this intensity, cells harbouring the differentiation system were cultured to exponential phase in batch and then subjected to either blue light or kept in darkness. Samples were taken at regular intervals manually and passed through the cytometer (Methods). I used flow cytometry data to compute the differentiated fraction by applying a threshold on mNeonGreen fluorescence (**Figure 3.2b**). I could observe only a marginal increase in the differentiated fraction after 72h of culture in the dark suggesting low background activity (**Figure 3.6, grey inset**). I note that exposure to light during the sampling could have also led to this increase. Induction with blue light triggered differentiation. Moreover, I found >99 % differentiation after 4h of induction (**Figure 3.6, blue inset**) making the efficiency of the system unprecedented when compared to existing systems in *S. cerevisiae* (Taslimi et al.

2016; Kawano et al. 2016; Hochrein et al. 2018; Duplus-Bottin et al. 2021). High efficiency even at this low intensity permits one to reach high levels of differentiation with minimal light exposure thus eliminating the risk of phototoxicity (**Figure 3.5**).

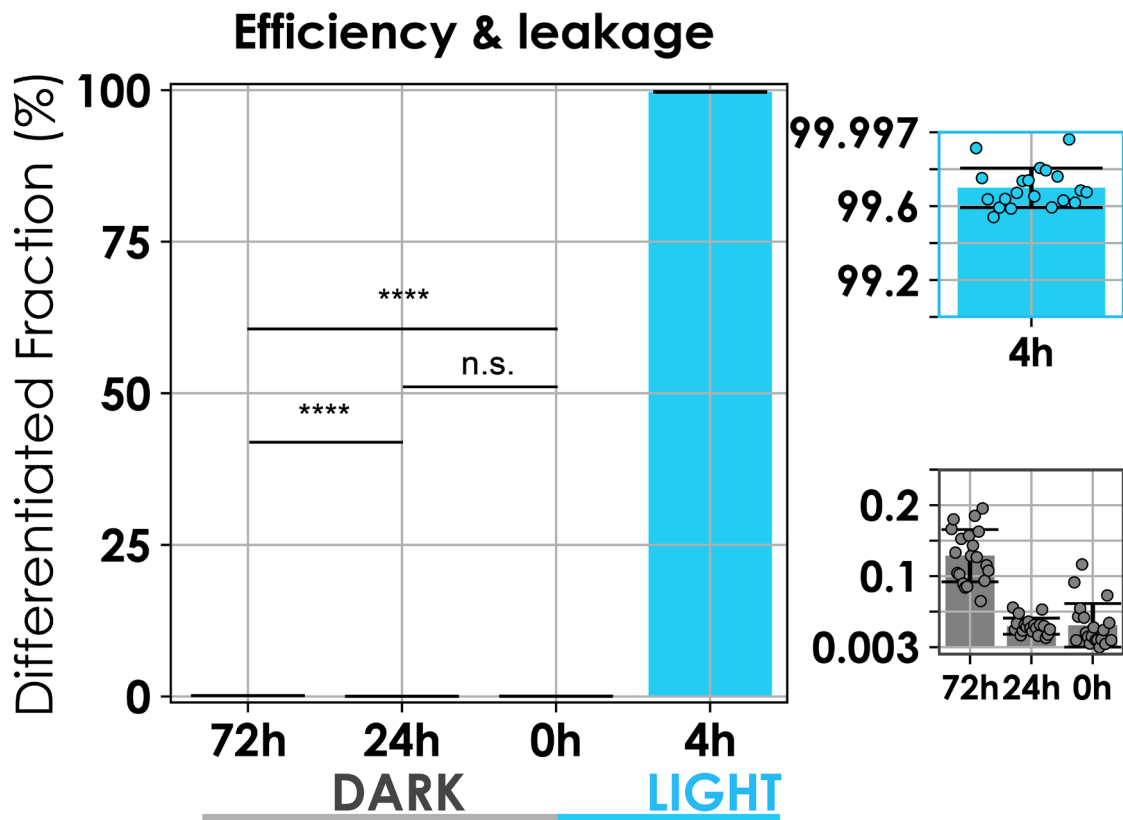


Figure 3.6. Efficiency and background differentiation. 20 colonies were picked and cultured in batch for 72h in dark or induced at $t=0$ for 4h. Measurements were taken at $t=0$, $t=24h$ and $t=72h$. Bars represent means from a single experiment. Errorbars signify standard deviation. Individual data points from colonies are depicted in a scatter plot overlaid on the bar plot (blue & grey inset). Increase in differentiated fraction was not significant (n.s.) after 24h ($p = 0.41$) but became significant at 72h ($p < 10^{-5}$) (black horizontal lines).

I compared the efficiency of my system with existing solutions for optogenetic control of recombination in yeast (**Table 3.1**). Since, the systems reported in (Taslami et al. 2016) and (Kawano et al. 2016) consisted of results for mammalian cells (HEK293 and COS-7, respectively), data for these two is taken from (Duplus-Bottin et al. 2021). Leakage in dark is defined as percentage of cells identified as recombined after culture in the dark.

Table 3.1. Characteristics of optoinducible/photoactivable recombinases in yeast (Kawano et al. 2016; Taslimi et al. 2016; Hochrein et al. 2018; Duplus-Bottin et al. 2021).				
Publication	System	Growth conditions	Efficiency (mean \pm s.d.) - Induction time	Leakage in dark (Mean \pm s.d.)
Duplus-Bottin <i>et al.</i> (Figure 3), adapted from Taslimi <i>et al.</i> [†]	split Cre CIB1_CRY2	Stationary liquid culture	1.6% \pm 0.8 at 90 mins	1.3% \pm 0.5 over 24h of dark culture
Duplus-Bottin <i>et al.</i> (Figure 3), adapted from Kawano <i>et al.</i> ^{††}	split Cre pMag-nMag	Stationary liquid culture	21.2% \pm 5.8 at 90 mins	7.1% \pm 1.1 over 24h of dark culture
Hochrein <i>et al.</i> (Figure 2) original study	split Cre PhyB-PIF3*	Stationary liquid culture	46.7% \pm 5.3 at 24 h**	6.7% \pm 2.6 over 24h of dark culture
Duplus-Bottin <i>et al.</i> (Figure 3 & 4) original study	destabilized Cre fused to asLOV2 (LiCre)	Stationary liquid culture Exponential liquid culture	41.2% \pm 2.8 at 40 mins 66.7% \pm 3.7 at 90 mins 66.8% \pm 3.3 at 180 mins 7.6% \pm 2.1 at 40 mins	0.7% \pm 0.2 over 24h of dark culture
This study	EL222 inducible WT Cre	Exponential liquid culture	43.1% \pm 2.7 at 40 mins 76.8% \pm 1.7 at 90 mins 94.4% \pm 0.9 at 180 mins 99.7% \pm 0.1 at 240 mins	0.06% \pm 0.05 over 72h of dark culture

[†]Original study performed in HEK293 cells.

^{††}Original study performed in COS-7 cells.

*Addition of PCB was necessary to achieve recombination with PhyB-PIF3 system. Results shown here correspond to a concentration of 25 μ M. Efficiency could be improved to 89.3% \pm 4 by increasing the concentration of PCB to 100 μ M. This also resulted in an increased the background activity to 9.7% \pm 2.4.

** Induction was carried out by delivering a single 5-min pulse of red light and followed by 10s pulses every 5 minutes for 24h.

I found that my solution is the most efficient light based recombination system to date and shows the least amount of leakage in the dark. These observations led me to conclude that my system was functional and possessed the desired characteristics of high efficiency, low leakage and irreversibility (by design).

Characterization of differentiation dynamics under the microscope

To further establish that my proposed system remains functional in different experimental contexts, I cultured cells in a microfluidic chamber and stimulated them periodically via short pulses of light on our microscopy platform (Fox et al. 2021). Mean cellular fluorescence (MCF) was calculated by averaging the pixel intensity for all the pixels included within the segmented

area. I, then, used MCF values to assign a differentiation status to cells. Prior to light induction (**Figure 3.7, t=0**) less than 2% of the cells were differentiated (n=817, over 8 fields of from 2 independent experiments) and within 8 hours of induction more than 99% of the cells in the field of view had differentiated (**Figure 3.7 t=6h**).

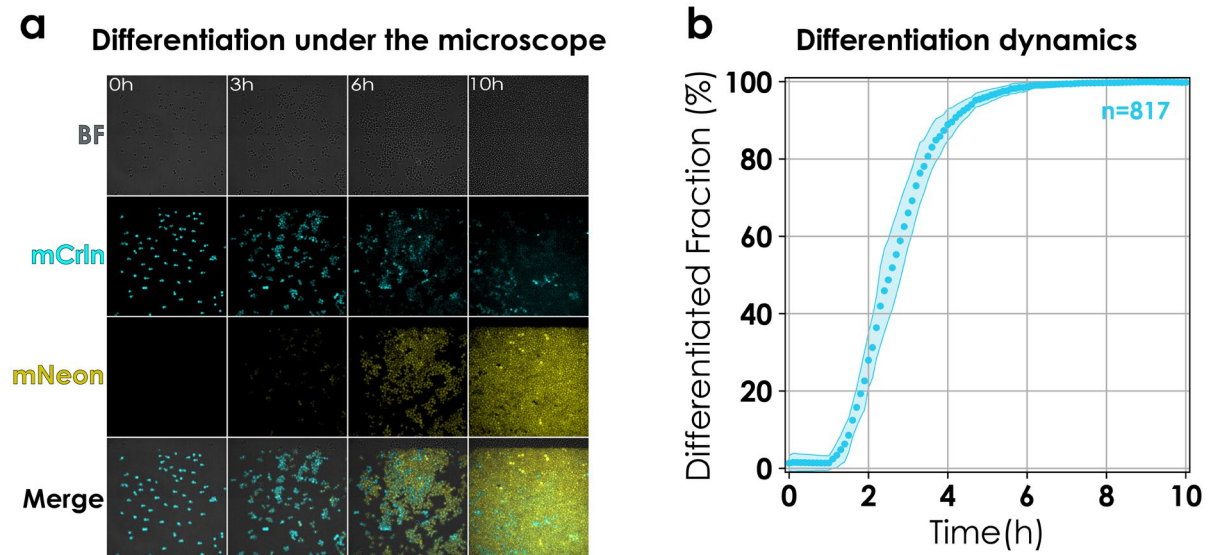


Figure 3.7. a. Snapshots of cell growth and differentiation under the microscope. Images during induction from a representative field of view. Cellular fluorescence changes from mCerulean (cyan) to mNeonGreen (yellow). **b. Differentiation dynamics under the microscope.** Images were segmented and analysed. To be deemed differentiated, median cellular fluorescence had to exceed 300 A.U. mNeonGreen fluorescence. Circles represent mean differentiated fraction over 8 fields of view from two independent experiments. Shaded region shows standard error of mean. The total number of cells at t=0 summed over all fields of view are given by n (n=817).

The differentiation dynamics were reproducible (**Figure 3.7b**). I note the presence of some cells that do not seem to lose their mCerulean fluorescence. A frame-by-frame analysis of the live cell timelapse reveals that these cells are arrested and do not divide. Since mCerulean is not tagged with a degron, dilution is the only way for cells to lose the mCerulean fluorescence. Notwithstanding, I conclude that the differentiation system remains functional under the microscope.

Imprinting patterns for spatial control of population behaviour

Buoyed by this success, I proceeded to ascertain if it were possible to imprint custom patterns in a population such that population behaviour could be controlled spatially. Cells were loaded in a μ lbidi slide and placed under the microscope. I used our microscopy platform that is equipped with a Digital Micromirror Device (DMD) to periodically shine blue light in the shape of different

patterns (**Figure 6.22**). Cells were illuminated with a given pattern for 1s every 3 minutes during 1h. Following this, cells were kept in darkness for an hour prior to imaging. We observed that accurate patterns of differentiated cells emerged (**Figure 3.8**). Some recombination was present outside of the provided pattern but these cells had differentiated prior to the start of experiment as evidenced by their mNeonGreen fluorescence levels and a lack of mCerulean fluorescence. Nevertheless, I remark that brief induction with light pulses was sufficient to create a permanent spatial genetic memory in the population. I also remark that the experiment was performed three independent times and the pattern formation was robust and reproducible.

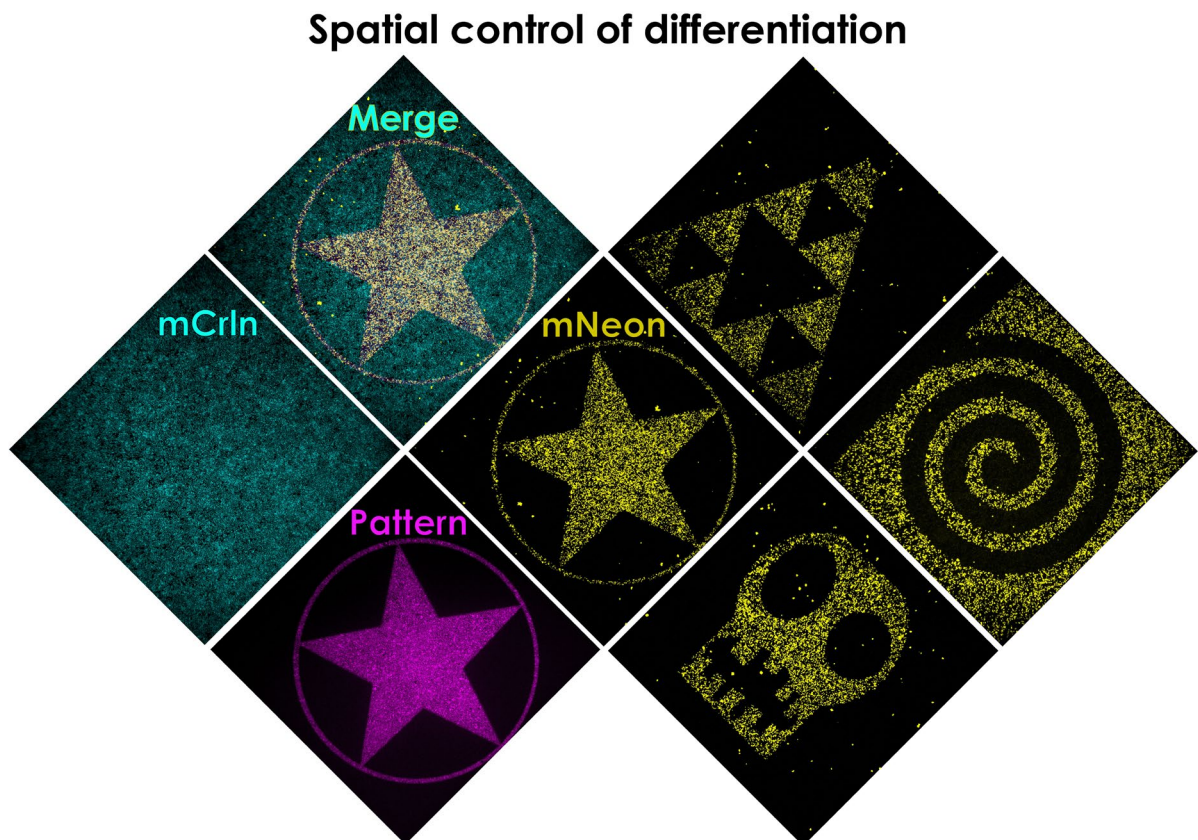


Figure 3.8. Imprinting patterns in the population. Cells were allowed to form a monolayer inside a μ l biddi slide. Light was shone in a user-defined pattern over the monolayer using a digital mirror device (DMD). mCerulean and mNeonGreen fluorescence were ascribed cyan and yellow colour, respectively. The pattern, in magenta, is overlaid on top of the merge. The intensity of the pattern has been decreased to highlight the differentiation.

I conclude that the optogenetic differentiation system enables precise spatial control of differentiation suggesting that it is a practical tool to generate spatially structured heterogeneous microbial communities composed of functionally distinct subpopulations.

Detailed characterization of the Dual reporter in continuous cultures

To characterize the differentiation behaviour in exponentially growing cells, I cultured cells continuously in our LED equipped custom turbidostat platform (Bertaux et al. 2020) (Methods, 6.4) and stimulated them with different light inputs (Figure 3.9). Light inputs were single pulses of different durations ranging from 15 minutes to 180 minutes.

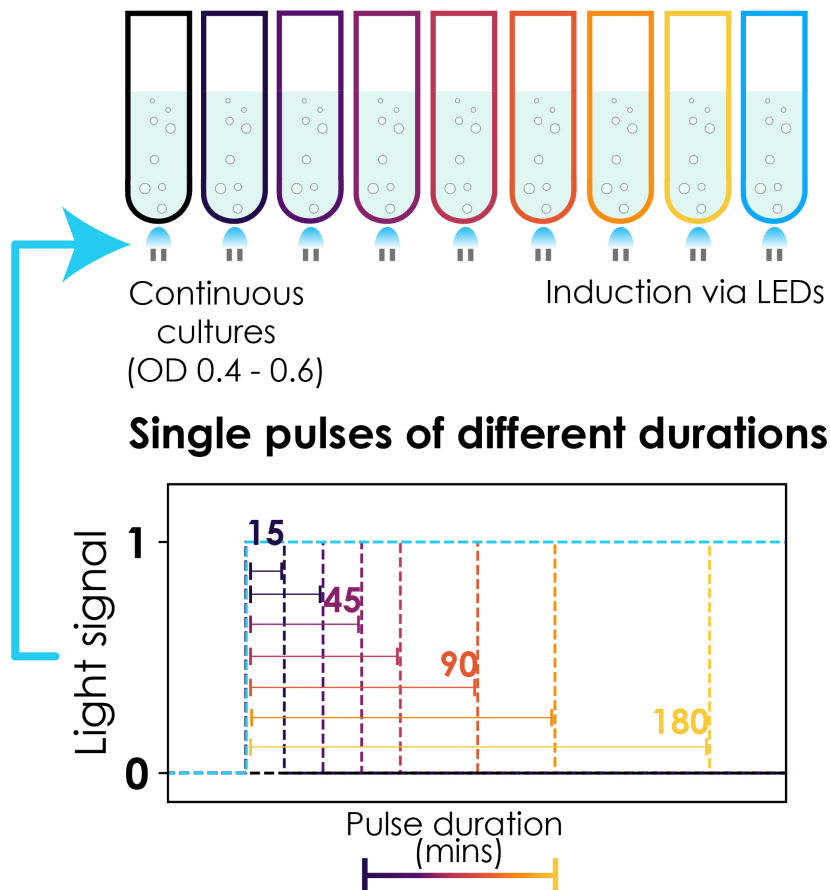


Figure 3.9. Bioreactor platform and induction profiles. Cells were cultured continuously in exponential phase using our bioreactor platform (Bertaux et al. 2020). Cultures were induced via LEDs and flow cytometry measurements were automated. Induction was in the form of single pulses of different durations.

Sampling from the culture and flow cytometry measurements were automated (Methods). I observed that, by modulating duration of light pulses (Figure 3.9), the system could be regulated to reach intermediate levels of differentiation that were stable over time (>48h, Figure 3.10a), highlighting stable maintenance of microbial consortium at different compositions (Figure 3.10b). Moreover, these results reveal an interesting dichotomy of the system: it is capable of eliciting a graded response to different stimuli at the population level as well as a differential response to

the same stimulus at the single cell level (**Figure 3.10 & Figure 3.6**). Both these observations can be explained by heterogeneity in EL222 levels but more on this in Chapter 5.

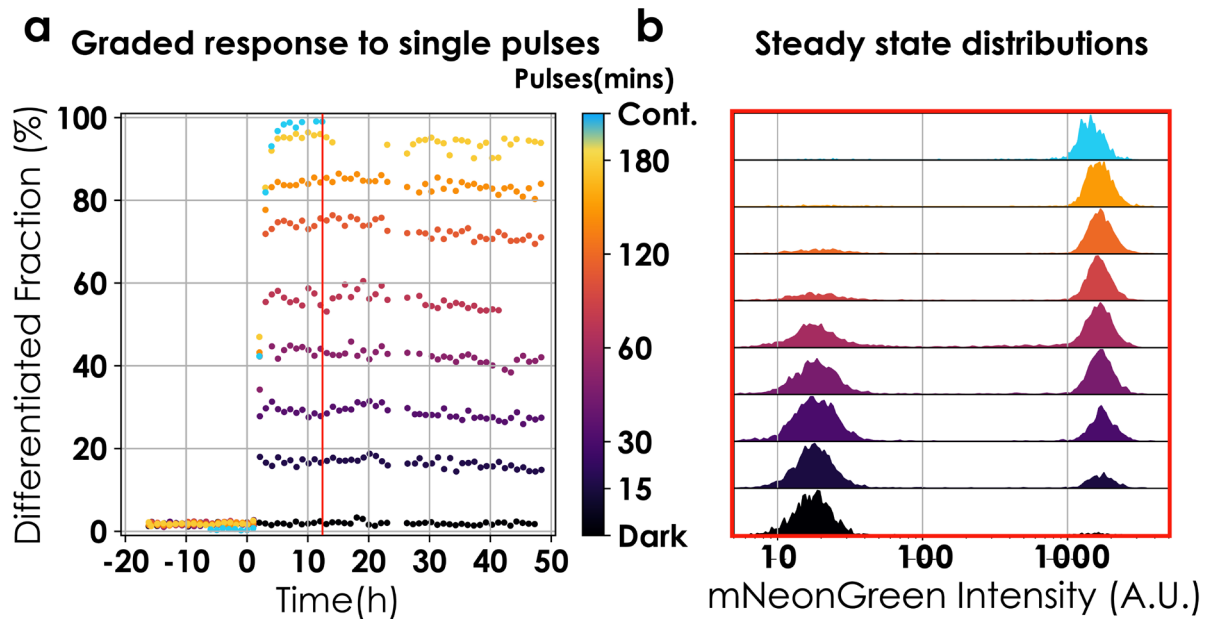


Figure 3.10 a. Differentiation dynamics after single pulse induction. Cultures were induced at $t=0h$ with single pulses of light ranging between 15 to 180 minutes (colourbar). Following induction, cultures were kept in the dark for 48h. Circles represent values from a unique experiment. **b. Tunable population composition.** Snapshots of population mNeonGreen fluorescence after induction with single pulses (**a**, red line indicates the time of each snapshot).

Since the intended use of the system was for the dynamic control of microbial consortia, I wondered if hysteresis could be observed in the system given repeated stimuli. Hysteresis is a phenomenon that occurs in certain dynamical physical systems wherein the system retains a “memory” of previous stimulation. The classic example of such behaviour is a spring. A spring can be stretched by applying force on its ends up to a certain degree such that it reverts back to its original length when the force is removed. However, if the applied force, say F_m , exceeds a threshold (determined by the physical and chemical properties of the spring), the spring does not relax back to its original length upon removal of the force. Furthermore, when one tries to stretch the string again, the behaviour is markedly different from the first time. In this case, the spring could be said to have retained memory of its exposure to the force F_m .

Analogously, hysteresis in my system, would manifest in the form of a reduced capacity of a light stimuli to effect differentiation. If present, hysteresis would strongly inform the modelling choice especially in the context of control applications. Therefore, to gauge the extent to which it might affect differentiation dynamics, I induced cultures with repeated pulses at various interpulse durations. More precisely, the cultures were stimulated with 30-min pulses that were delivered 30, 60, 120, 180, 240, or 300 minutes apart (**Figure 3.11**).

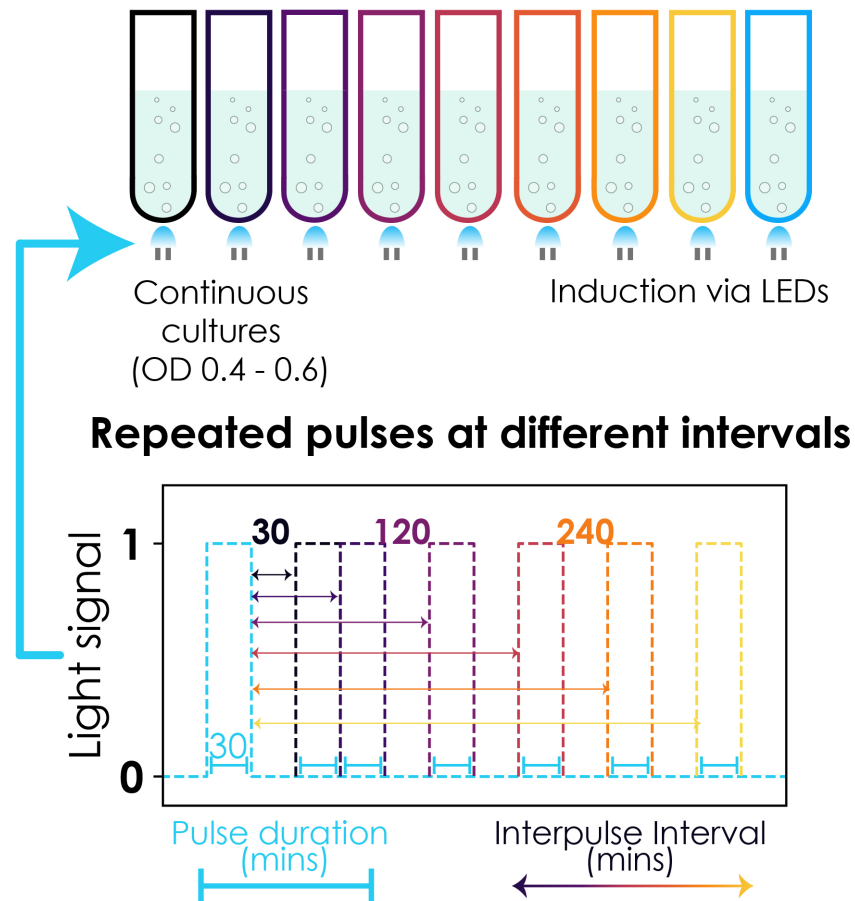


Figure 3.11. Bioreactor platform and induction profiles. Cells were cultured continuously in exponential phase using our bioreactor platform (Bertaux et al. 2020). Cultures were induced via LEDs and flow cytometry measurements were automated. Induction was in the form of repeated pulses of 30 minutes at different interpulse intervals. The figure is representational and only shows two pulses.

If hysteresis were present, we would expect smaller population fractions to differentiate with subsequent pulses compared to the first one. Furthermore, as the interpulse duration is increased, I anticipated the effect to become weaker due to gene expression noise (relaxation). However, I failed to observe either of these two effects. The response was reproducible with each pulse resulting in the same differentiated fraction regardless of prior exposure to light (**Figure 3.12a**) (up to small reactor-to-reactor variability).

Sceptical of these results, I argued that there was another way of detecting the presence of hysteresis in the system. If it were the case, then time-separated, repeated pulses of a given total duration would result in a higher differentiation fraction compared to a single continuous pulse of same duration. Contrary to my intuition, I found that continuous light resulted in similar differentiation fractions as discrete pulses for the same total duration of induction (**Figure 3.12b**). Both results provide strong evidence that hysteresis is absent or masked by reactor-to-reactor variability at the timescales relevant for control purposes. In turn, these results suggest that a simple model could be sufficient to predict differentiation behaviour given a light profile.

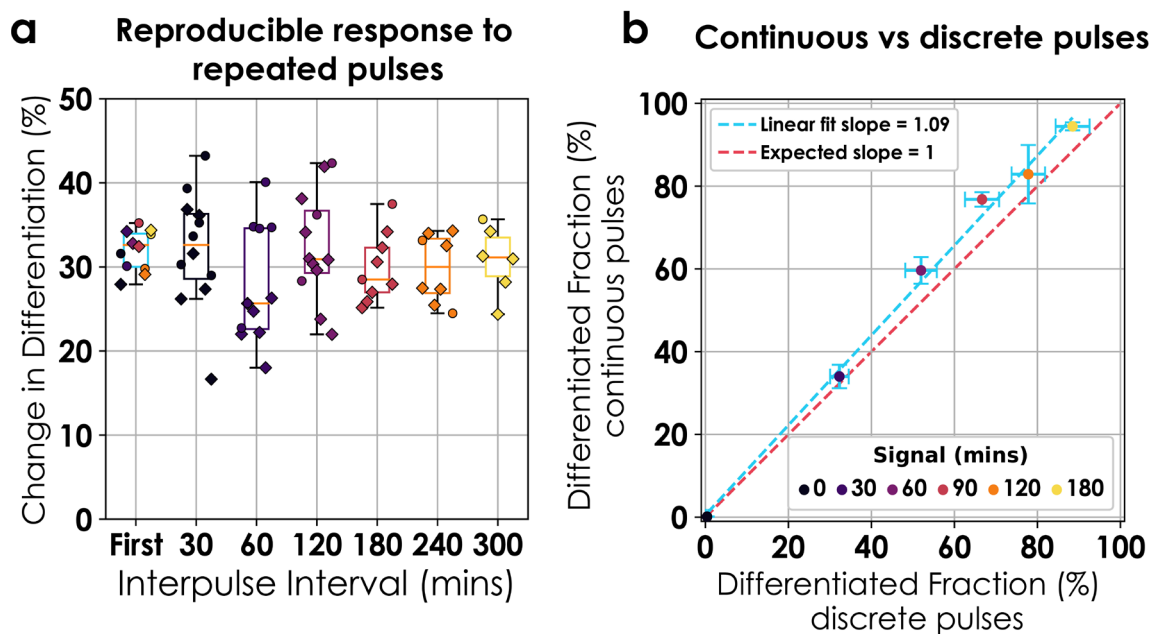


Figure 3.12. a. Reproducible behaviour with repeated pulses. Cultures were stimulated with 30-minute pulses repeated at different interpulse intervals (30 to 300 minutes). Circles and diamonds represent the change in differentiation fraction by individual pulses from two independent experiments. Data was collated over the two experiments for boxplots. The colour of circles and boxplots reflects interpulse interval. The first pulse of each experiment was used for the blue boxplot. **b. No observable memory effect.** Data from **a** & **3.10a** was used to compare the efficiency of continuous light vs discrete pulses. Circles represent differentiation effected by continuous pulses (y-axis) and equivalent duration of induction in form of 30-minute pulses (x-axis). Errorbars signify standard deviation. A linear fit of the data is given by the blue dashed line and compared to expected linear behaviour in the absence of memory (red line).

To be able to predict the differentiated population fraction for a given light input, I developed an ordinary differential equation (ODE) model with a single parameter to calibrate, the differentiation rate. Since the model was primarily developed for the purpose of deploying it in model predictive control of the population composition, I decided to use a simple ODE model that tracks population dynamics.

$$\dot{g}(t) = \mu g(t) - \mathbf{U}(t)r_{diff}g(t) - \lambda g(t)$$

$$\dot{p}(t) = \mu p(t) + \mathbf{U}(t)r_{diff}g(t) - \lambda p(t)$$

$$n = g(t) + p(t)$$

$$\lambda = \mu$$

g and p stand for specific cell density (in O.D. units) of non-differentiated and differentiated cells, respectively. μ is growth rate per hour of the culture. $\mathbf{U}(t)$ is light signal as a function of time and can take values 0 or 1. r_{diff} is the differentiation rate under continuous light. n is the total cell density (in O.D. units) of the culture and is kept constant. λ is the dilution rate. At constant total cell density, dilution rate, λ equals growth rate, μ . No significant difference was observed in the growth rate of differentiated cells and non-differentiated cells, and therefore, these were assumed to be the same (**Figure 3.2**).

This model was fitted (**Methods, 6.7**) to two datasets of steady state differentiation fractions post induction with mutually exclusive single pulses. Both fits captured observed behaviours well and resulted in similar estimates for the differentiation rate (**Figure 3.13a**). I also used the model to predict differentiation dynamics emerging from additional light input sequences and found that model predictions were in good agreement with observed data (**Figure 3.13b**). I note that the predicted dynamics were shifted in time (1 h) to account for observation delays that were not explicitly modelled.

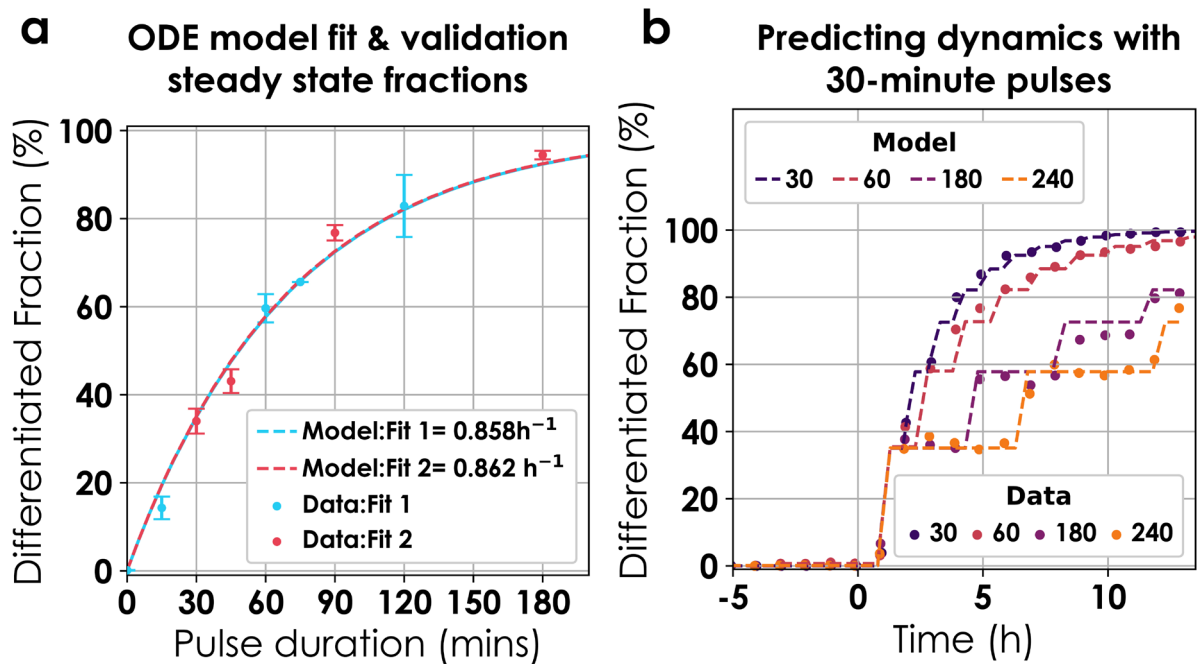


Figure 3.13. a. Model fitting and validation. An ODE model was fitted to single pulse induction data. Circles represent mean steady state differentiation fractions of three independent experiments (except 75-

minute pulse, unique experiment). Error bars signify standard deviation. Blue and red circles were independently used to fit the ODE model (dashed lines) (Methods). **b. Predicting dynamics.** Data (circles) from 3.12a was used to check the predictability of the ODE model (dashed lines). Induction started at $t=0$. Model predictions were shifted in time to account for observation delay.

To ascertain the impact of cell density on functionality and efficiency of the differentiation system, I induced cultures maintained at different cell densities between OD 0.1 and OD 1.5 with repeated pulses. I found that the functioning of our differentiation system is not affected by the increased cell density (variability in the response at different ODs is smaller than reactor-to-reactor variability at the same OD compare Figure 3.12 & 3.14).

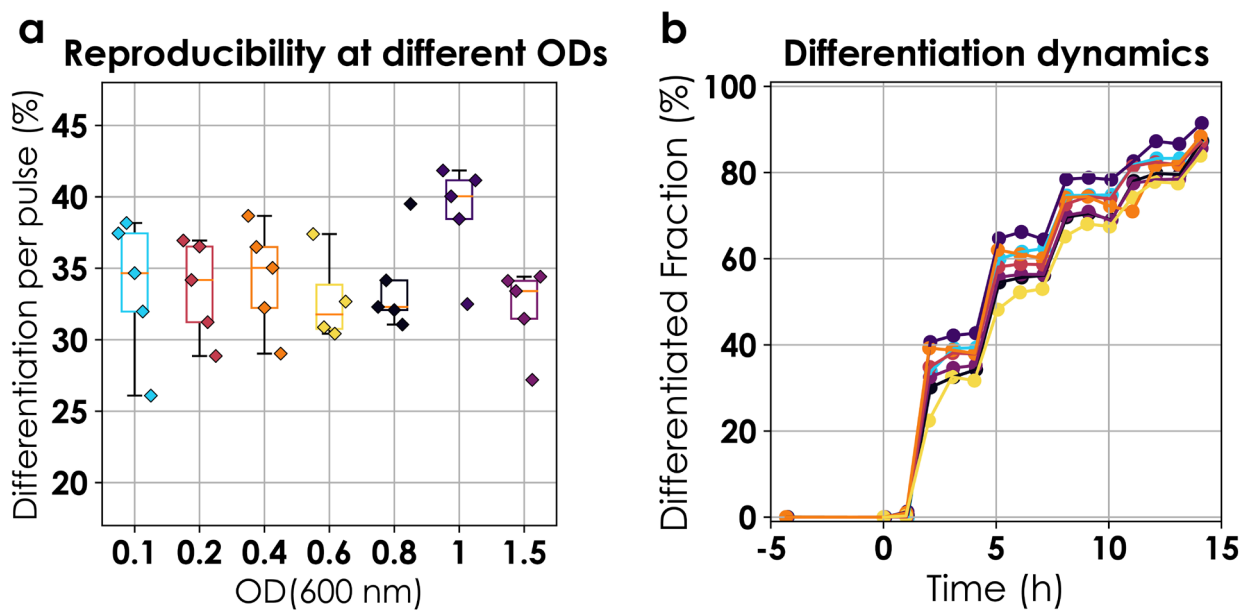


Figure 3.14. a. Comparing the variability in response to the same (30-min) pulse when cells are cultured continuously at **different ODs**. For each OD, a single experiment was performed in which cells have been exposed to five 30min light pulses delivered 3h apart from each other and the differentiated fraction in response to each pulse was quantified (diamonds). **b. Dynamics of differentiation at different ODs.** Colours correspond to the ODs in **a**. Induction started at $t=0$.

I conclude that the differentiation behaviour of my system is predictable and can be captured by a simple ODE model suggesting that open-loop control of consortia composition is trivial with the system and could be achieved any culture density. I clarify that this is only true for when the differentiated cells grow at the same rate as the non-differentiated cells.

Two reactor MPC control

In the previous section, I established that my system rendered the precise control of consortia composition trivial, when both differentiated and non-differentiated cells grow at the same rate. However, this scenario was unlikely to be true in the context of metabolic engineering and

bioproduction applications. Furthermore, this system, on its own, does not permit bidirectional control of consortia composition because, upon differentiation, cells would grow at the same rate as non-differentiated cells. This means if 50% of the population has been differentiated, I cannot bring the level of differentiation back to below 50%.

Therefore, in order to produce a more robust and relevant proof-of-concept demonstration, I introduced an external growth rate difference in the system. Concretely, I increased the apparent growth rate of non-differentiated cells by adding the output of a second culture kept in the dark (a reservoir for non-differentiated cells) to the (control) culture. The control culture was exposed to light signals to maintain the culture at a target set point for the differentiated fraction. Cytometry samples were taken every hour to observe the state of the control reactor. Cultures in both the reservoir and control reactors were maintained at constant cell densities (Figure 3.15).

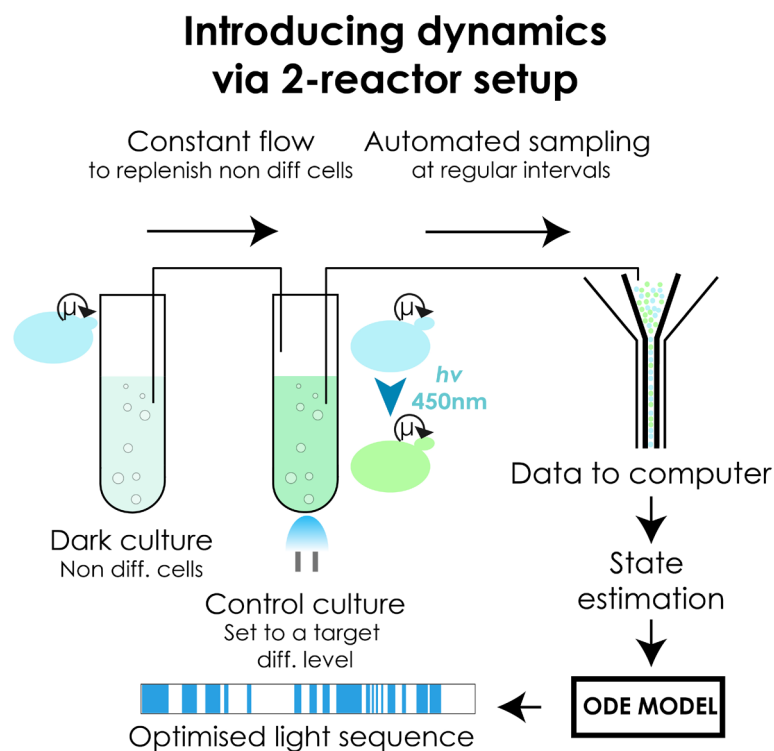


Figure 3.15. Setup for two reactor control. Cells harbouring the original differentiation system were continuously cultured at fixed but different cell densities in two reactors simultaneously. The first reactor was kept in the dark as a reservoir of non-differentiated cells. The output of this vessel was connected to the second 'control' reactor. The control culture was set to a target level of differentiation and continuously monitored via automated flow cytometry measurements that were analysed on-line. The system state was estimated from analysed data and sent to the model predictive control (MPC) module. The MPC module provided an optimized light sequence to maintain the culture at the desired set.

The utility of this two-reactor configuration was twofold. First, it allowed me to test the capacity of the system to operate when growth rate differences are present between the differentiated and non-differentiated cells. Second, it allowed me to verify if bidirectional control of consortia composition was feasible with my differentiation system.

Introducing the second reactor changed the dynamics of the system. Notably, there was a constant flux of non-differentiated cells from the reservoir to the control reactor and the dilution rate of the reactor was no longer equal to the growth rate of culture. In order to account for these changes, I modified the ODE model described in the previous section.

$$\begin{aligned}\frac{dg}{dt} &= \mu_{target}g - \mathbf{U}(t)r_{diff}g + Q - \lambda g \\ \frac{dp}{dt} &= \mu_{target}p + \mathbf{U}(t)r_{diff}g - \lambda p \\ Q &= \mu_{reservoir}n_{reservoir} \\ n_{target} &= g + p\end{aligned}$$

g and p stand for the specific cell density (in O.D. units) of non-differentiated and differentiated cells, respectively. $\mu_{control}$ and $\mu_{reservoir}$ are growth rates in the control and reservoir reactors, respectively. $\mathbf{U}(t)$ is the light signal as a function of time and can take values 0 or 1. n_{target} and $n_{reservoir}$ are O.D. (total cell densities) at which control and reservoir reactors are maintained, respectively. r_{diff} is the differentiation rate under continuous light. Q is the flux of non-differentiated cells from reservoir to control reactor. λ is the dilution rate of the control reactor. I observed that growth rates of cultures at different ODs were not equal. I assumed, however, that reservoir cells instantaneously change their growth rate upon entering the control reactor.

Since this two reactor configuration relied on the reservoir to replenish the non-differentiated population differentially in the control reactor, the OD of the reservoir culture would determine the range of consortia compositions achievable. I wondered how to set the ODs of the two reactors so as to ensure that up to 90% of cells could be differentiated in the control reactor. I calculated that in order to reach a target differentiation level of 90 % in the control reactor, the OD of the reservoir had to be ~ 5 times lower (**Methods, 6.8**). Based on this calculation, the OD of the reservoir reactor was set to 0.1 and that of control reactor to 0.6. Growth rates were determined from the data.

This modified model was then plugged into a model predictive control (MPC) framework described previously in Bertaux et al. (2020) (**Methods, 6.8**) and using this MPC framework, I

targeted population compositions in individual reactors to 10-80% differentiated cells in the population. I was able to control the population compositions in control reactors and maintain them for extended periods (up to 96h) (**Figure 3.16a**) for different target set points. The response was quick and the desired population fractions were reached within approximately 6h of starting the control. Additionally, I targeted two cultures to 40% and 80% differentiated cells in the population, respectively and changed the set points after 66h to 80% and 40%, respectively (**Figure 3.16b**). In both cases, the differentiation fraction followed the target. I note that the control reactor required active control and lost the desired population composition in the absence of an appropriate light signal.

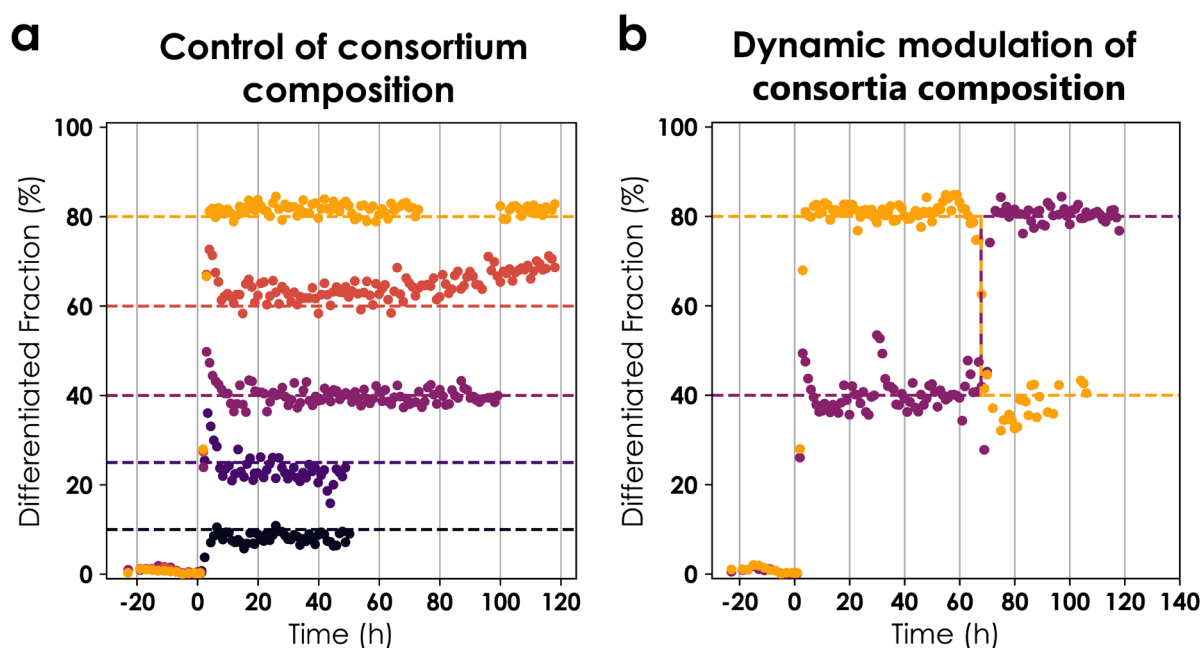


Figure 3.16. a. Control of consortium composition. Cultures were targeted to 10-80% differentiation. Control started at $t=0h$. Circles signify differentiated fractions. Each colour corresponds to a different control experiment and the dashed line reflects the target set point. Note that the figure is composed of independent experiments of different lengths. Data was removed when the OD, either in the reservoir or the control reactor, could not be maintained at the desired target. **b. Bidirectional control of consortium composition.** Cultures were targeted to 40% and 80% differentiation. Data is represented as in **a**. The target was changed at $t=66h$ to 80% and 40%, respectively.

I conclude that, in the presence of growth rate differences between differentiated and non-differentiated cells, my system retains the capability of maintaining microbial consortia at desired compositions and enables dynamic modulation of the population composition in both directions i.e. increase or decrease differentiated cells in the population from a given level.

Extension to multi species consortia

Certain applications, particularly in the context of metabolic engineering, might require microbial consortia composed of more than two species. To probe if my system can be used to engineer differentiation programs that allow one to create complex multi-species consortia, I cloned two recombination cassettes (denoted by C and N) in a single strain along with the differentiation system (**Figure 3.17a**).

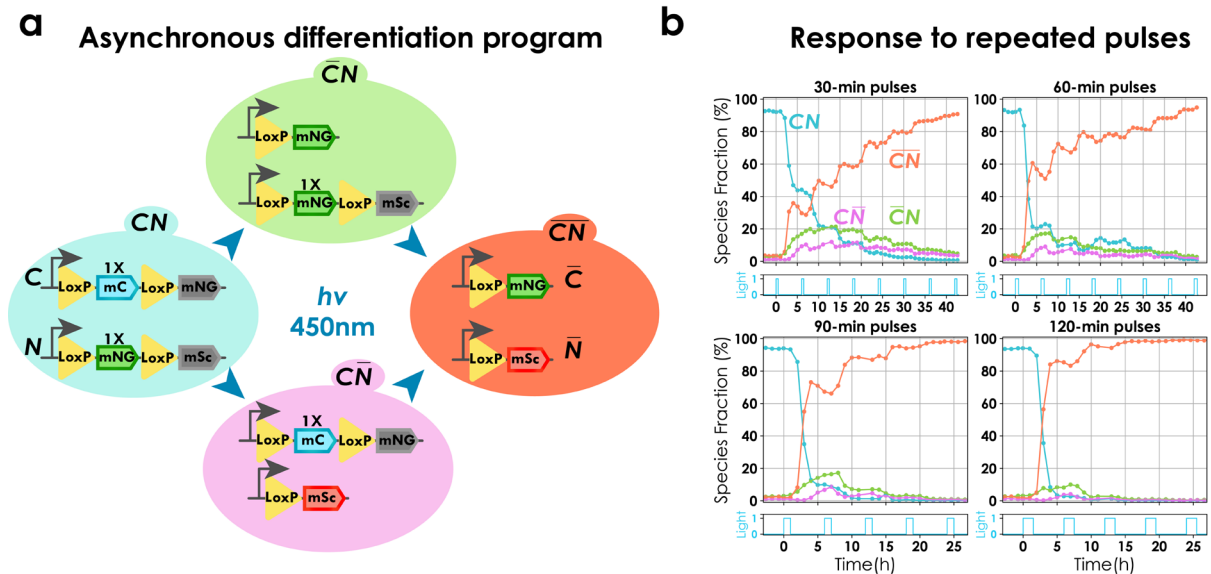


Figure 3.17. a. Asynchronous differentiation program. Two recombination cassettes (C and N) were introduced along with the differentiation system in the same strain. Both cassettes consisted of to-be-excised region of equal lengths ($1X$ of the original system) (Figure 3.1a). All 4 possible species are shown (neither cassettes recombined (CN), single cassette recombined ($\bar{C}N$ & $C\bar{N}$) and both cassettes recombined ($\bar{C}\bar{N}$). Circuits shown in the panel are representational (Methods, Figure 6.6). **b. Response to repeated pulses of different lengths.** Four cultures were induced with repeated light pulses of different durations 6h apart. Each subplot represents data from a single experiment with unique pulse duration given by cyan subplots (clockwise from top left, 30-minute, 60-minute, 120-minute and 90-minute, respectively). Solid lines show dynamics of prevalence of each species (blue, purple, red, and orange for CN , $\bar{C}N$, $C\bar{N}$, and $\bar{C}\bar{N}$, respectively). Circles indicate individual time points. Prevalence of each species was calculated using thresholds on $mNeonGreen$, $mScarlet-I$ and $mCerulean$ fluorescence (**Methods, Figure 6.6**). All four species could be observed.

Cells were cultured and induced in our bioreactor platform with varied pulses repeated every 360 minutes. I was able to detect all 4 species at appreciable prevalence in the culture (microbial consortium), consisting of double recombined cells ($\bar{C}\bar{N}$), single recombined cells ($C\bar{N}$, $\bar{C}N$) and the original non-recombined population recombined (CN). I found that both sites were approximately equally likely to recombine (asynchronous with respect to each other) albeit it is not evident from the curves (**Figure 3.17b**). Seemingly, cassette C recombines at a greater rate

than N, however, this is possibly an artefact of the different maturation times of the reporters for each cassette. Recombination at C is reported by mNeonGreen which has a maturation time of 10 mins compared to 36 minutes for mScarlet-I (Lambert 2019). The composition could be stably maintained at any observed level if subsequent light pulses were not applied (**Figure 3.17b**).

Noting that the length of the to-be-excised region plays a critical role in determining the differentiation rate (**Figure 4.12a**), I set out to test if this could be exploited to modify the differentiation dynamics. We cloned two recombination cassettes of unequal to-be-excised region (C and N') along with the differentiation system (**Figure 3.18a**).

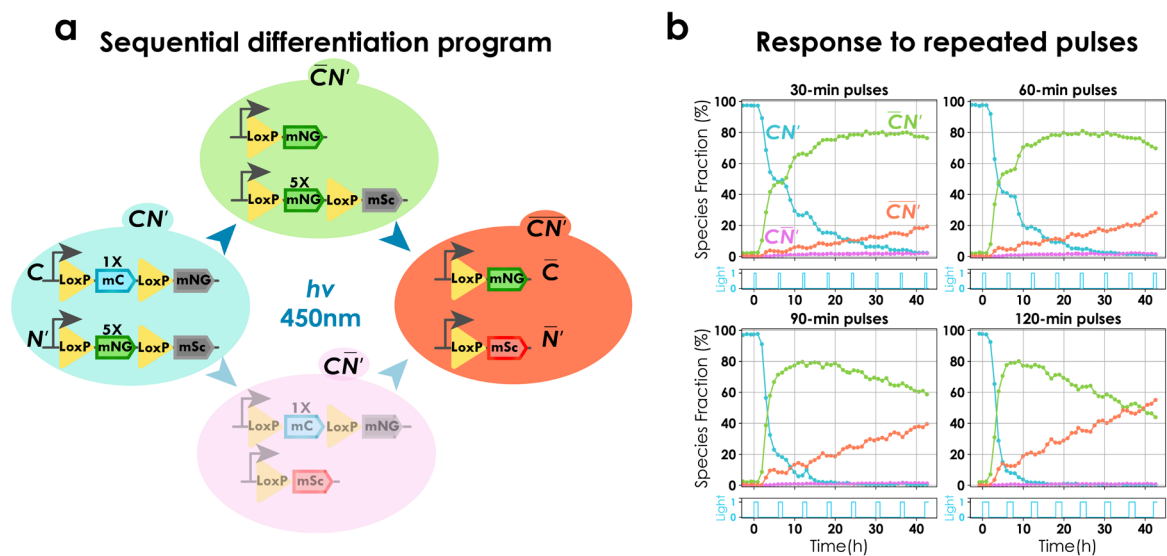


Figure 3.18. a. Sequential differentiation program. Two recombination cassettes were introduced along with the differentiation system in the same strain. Relative length of the to-be-excised region of one cassette (N') was 5X longer than the other (C). All 4 possible species are shown. Notations resemble **a**. Due to increase in the length of to-be-excised-region, the differentiation rate of N' was significantly slower than C leading to inappreciable prevalence of $\bar{C}\bar{N}'$ (faded species). **b. Response to repeated pulses of different lengths.** Four cultures were induced with repeated light pulses of different durations 6h apart. Each subplot represents data from a single experiment with unique pulse duration (clockwise from top right, 30-minute, 60-minute, 120-minute and 90-minute, respectively). Data is represented as in **Figure 3.17b**. Three species could be observed.

I observed that the shorter site recombined first ($\bar{C}N'$) and subsequent pulses led to enrichment of cells with both sites recombined ($\bar{C}\bar{N}'$) (**Figure 3.18b**) constituting a sequential differentiation program as opposed to the asynchronous program obtained with to-be-excised regions of equal lengths. I did not observe appreciable levels of species with only the longer site recombined ($C\bar{N}'$) effectively resulting in a three species microbial consortium.

Both versions of the system with two recombination sites remained capable of eliciting a graded response to light. I conclude that my differentiation system can be extended to implement distinct differentiation programs capable of yielding complex consortia composed of a controllable number of species whose prevalence can be optogenetically tuned.

Discussion

Microbial consortia are expected to be of great utility for biotechnology and hold immense potential for diverse applications (Rapp, Jenkins, and Betenbaugh 2020; Zhou et al. 2015; Li, Wang, and Zhang 2019; Zhuang et al. 2014; Zhang et al. 2021; Panke-Buisse et al. 2015; Inda et al. 2019) and have been hailed as the next frontier in synthetic biology (Brenner, You, and Arnold 2008). However, dynamic control of consortium composition remains relatively unexplored despite being a key challenge in the field (Roell et al. 2019). In this chapter, I addressed this challenge with the help of an artificial differentiation system in *S. cerevisiae* capable of generating microbial consortia with custom composition. The system was based on blue light inducible expression of Cre recombinase driven by EL222 from a non-leaky promoter (Benzinger and Khammash 2018). I characterized the system and established that it possesses several desirable characteristics: fast, reproducible and tunable dynamics, unprecedented efficiency (Taslimi et al. 2016; Kawano et al. 2016; Hochrein et al. 2018; Duplus-Bottin et al. 2021), low leakage, and graded response of the population to light. The efficiency of the system allowed me to achieve high levels of differentiation with short transient pulses that eliminate the risk of phototoxicity. Graded population responses to light were critical for achieving control of consortia composition. Moreover, the high degree of reproducibility in response to light stimuli permitted me to develop a predictive model that could be used as a basis to precisely control microbial community dynamics. Using the developed model in a model predictive control framework allowed us to achieve bidirectional control of consortia composition in a dynamic setup. I showed that the system can be extended to give rise to complex multi-species consortia. Lastly, I demonstrated that it was possible to genetically imprint patterns in a population with single cell precision (Annex 1) using the differentiation system.

On efficiency

The efficiency of the optogenetically inducible recombinase developed in this work exceeds any reported in the literature. Previous optogenetic recombination systems are based on engineering a photoactivable Cre that is typically split into two peptides tagged with the respective photosensitive heterodimers that can be brought together upon light illumination to form a functional Cre (Taslimi et al. 2016; Kawano et al. 2016; Hochrein et al. 2018). However, such

approaches result in massive activity loss as functional Cre is a tetramer and the probability of four dimerized split-Cre molecules to assemble together hinges on the relative concentrations of the two subunits (Duplus-Bottin et al. 2021). A recent study reported a monogenic photoactivable Cre that is based on fusion of a LOV domain with a destabilized Cre variant (Duplus-Bottin et al. 2021). The authors reported higher efficiency and stronger activation when compared to previous systems. Indeed, this monogenic photoactivable Cre matched the efficiency reported for my system up to 40 minutes of induction. After 40 minutes of induction, the activity seemed to plateau for the former. This could perhaps be due to heterogeneity in expression levels of the destabilized Cre and a memory effect might be at play here similar to the one described in Chapter 5. This appears to be absent from my system, at least for the time scales studied here. Nevertheless, I note that applications where genetic burden is important, the monogenic photoactivable Cre might be more suitable.

On dynamic control of microbial consortia

Several solutions for the stable maintenance of microbial consortia have been proposed recently. In particular, Hasty and colleagues (Scott et al. 2017; Miano, Liao, and Hasty 2020), Lu and colleagues (Kong et al. 2018; Liu et al. 2020), and Barnes and colleagues (Fedorec et al. 2021) achieved this by using synthetic biology approaches. These authors demonstrate the capacity to maintain co-cultures of several bacterial subpopulations over extended durations. However, none of these approaches succeeded at precisely controlling consortia composition (State of the art). Moreover, the functioning of these systems relies on the release of signalling molecules in the environment (quorum sensing molecules or bacteriocin) that trigger cell death. The fact that signalling molecules are released by cells creates de facto a strong dependency of the functioning on growing conditions, and notably on the density of cell cultures, an important aspect for bioproduction applications. Lastly, previous designs use elaborate genetic engineering solutions for the molecular implementation of control mechanisms, thus making extension and scaling up of these designs potentially challenging. In summary, in comparison to previously-existing solutions, our system is simple to implement, quantitatively predictable and actionable, and versatile to use. I should like to highlight that this work constitutes as the first report of dynamic control of microbial consortia arising out of a single strain. Previous studies using optogenetic control at the population scale focus on achieving control despite the heterogeneity (Miliadis-Argeitis et al. 2016; Bertaux et al. 2020) or seek to eliminate it altogether (Fox et al. 2021; Rullan et al. 2018), however, in this work I exploited heterogeneity in the population to control population level behaviour. This also happens to be the first report of optogenetic control of microbial consortia.

On opting against existing bimodal systems

Multiple systems have been designed that successfully create bimodality in gene expression at the population level (Gardner, Cantor, and Collins 2000; Becskei, S  raphin, and Serrano 2001; Atkinson et al. 2003; To and Maheshri 2010; Youk and Lim 2014; Yang, Nemhauser, and Klavins 2019; Barbier, Perez-Carrasco, and Sch  erli 2020). In theory (Jia et al. 2017; Del Vecchio et al. 2018; Salzano, Fiore, and di Bernardo 2019; Duddu et al. 2020), such systems could allow one to generate functionally distinct subpopulations from a single population. However, despite these encouraging examples and substantial theoretical advances, bimodal gene expression at the population level, with the exception of (Yang, Nemhauser, and Klavins 2019), has scarcely been exploited to generate and control subpopulations with distinct functionalities. Several obstacles still restrict the applicability of such systems to generate and dynamically control functionally distinct populations emerging from a single strain. First, most engineered systems could only maintain stability under carefully calibrated conditions. These conditions could pertain to the general and relative cell density (Scott et al. 2017; Youk and Lim 2014), inducer concentration (To and Maheshri 2010; Gardner, Cantor, and Collins 2000; Becskei, S  raphin, and Serrano 2001; Barbier, Perez-Carrasco, and Sch  erli 2020; Yang, Nemhauser, and Klavins 2019), and hard to tune genetic circuits (Liao et al. 2019) limiting the feasibility of dynamic control for population level behaviour. Secondly, to change the relative proportion of each phenotype, one must alter circuit topology, a process that is not only laborious but also carries no guarantee of success (Youk and Lim 2014). In essence, limitations of previous studies derive from the difficulty in ensuring appropriate regulation of the decision process. In the present study, I overcame this problem by using external control for differentiation. I note that Klavins and colleagues (Yang, Nemhauser, and Klavins 2019) have developed another differentiation system in yeast, that, in principle at least, could have been used to generate consortia. This system uses a toggle switch to implement memory, and chemical inducers to toggle the switch. I believe that using light as inducer and a DNA implementation of memory allowed me to precisely characterize and select systems with appropriate properties, and drive them with the needed precision to obtain subpopulations in desired proportions. The high degree of reproducibility in response to light stimuli facilitated the development of a predictive model that could be used as a basis to precisely control microbial community dynamics independently of the cell density.

On differentiation system as a pathway switch for population level metabolic engineering

Based on the principle of division of labour, microbial consortia have been employed to increase bioproduction by distributing the metabolic burden (Zhou et al. 2015). Such approaches necessitate functional specialization in the constituent species of consortia. I provide evidence

that our design can be implemented in different physiological contexts by artificially introducing a growth difference between differentiated and non-differentiated populations (a likely scenario for metabolically distinct species) and showed that the system still behaves predictably. In the context of metabolic engineering, our system could serve as a pathway switch, with the potential of compartmentalising metabolic flux in the population. This could be achieved, for instance, by replacing fluorescent proteins by orthogonal TFs that drive entire pathways (Naseri et al. 2019), leading to new division of labour paradigms in microbial consortia and open up possibilities for population level metabolic engineering. One concrete application of the system can be in efficient degradation of polyethylene terephthalate (PET) (Wyeth and Roseveare 1973). PET has been shown to be a major pollutant in both oceans and land (Neufeld et al. 2016). The pathway was first discovered in bacteria and consists of two enzymes that break down PET into its constituent monomers, terephthalic acid and ethylene glycol (Yoshida et al. 2016) and since then the efficiency of the process has been improved by more than 3 orders of magnitude (Tournier et al. 2020). It would be ideal if we could engineer a yeast strain that secretes both these enzymes. However, it is well known that secretion in *S. cerevisiae* is a delicate affair (Idiris et al. 2010). Therefore, it will be preferable to couple the production of these enzymes with the differentiation system such that prior to differentiation, the easy to secrete enzyme is expressed and after differentiation, the hard to secrete enzyme is expressed. It might also be prudent to replace *S. cerevisiae* with *K. phaffii* (erstwhile *P. pastoris*), which is overall more efficient at secreting heterologous proteins (Ahmad et al. 2014). This problem naturally lends itself to optimization and would provide a practical application for modelling biological systems.

On multi-species consortia

More involved metabolic burden distribution strategies might require more than 2 species. To show that complex multi-species consortia can be created using this system, I engineered asynchronous or sequential differentiation programs based on multiple recombination cassettes that extended the core system to generate and stably maintain multi-species consortia from a single strain in continuous liquid cultures. These programs could be scaled exponentially for applications requiring dynamic control of complex multi-species consortia and do not require intricate genetic circuits spread over multiple populations to ensure stability potentially rendering complex division of labour paradigms accessible. Notably, Roquet et al. (2016) demonstrated that as many as 16 distinct cell types could be obtained by using three orthogonal recombinases. However, the extended system does not provide a means of controlling each constituent species independently, making it unfavourable for applications where precise control of each species is

desired. This can, in part, be overcome by using multichromatically controllable orthogonal recombinases. More verbosely, the system described here can be used with another optogenetically driven recombinase system that is triggered with a different coloured light. In this fashion, it is possible to increase the number of species to 8 (theoretically 16), by having three orthogonal recombinases (Lin et al. 2015; Nern et al. 2011; Liu et al. 2018) being driven by three differently coloured light signals, the limiting factor being the number of characterized orthogonal light induction systems (Benzinger and Khammash 2018; Hochrein et al. 2017; Schmidl et al. 2014). Multichromatic control has previously been demonstrated in bacteria (Fernandez-Rodriguez et al. 2017) but is yet to be demonstrated in yeast. A preprint (Jang et al. 2019) reported development of a library of optogenetic tools derived from cyanobacteriochromes (Fushimi and Narikawa 2019) shown to be functional in yeast and operating in green, orange, red, and far-red range.

On differentiation system as a tool to study complex biological systems

I showed pattern generation in 2D cultures in a microfluidic plate. Since we are not restricted to patterns attainable in nature (Basu et al. 2005), my system provides a unique opportunity to study how equilibria are reached in multispecies ecosystems and elucidate microbial interactions in complex spatially structured communities. Alternatively, the capacity to optogenetically control cell fate decisions with spatiotemporal precision has the potential to become a critical tool for dissecting signalling pathways (Toettcher et al. 2011) or understanding developmental programs (Johnson et al. 2020). EL222 has been shown to be functional in mammalian cells and this system, due to high precision and efficiency, could potentially be used to perturb developmental processes in higher mammals in conjunction with ex-utero roller cultures (Aguilera-Castrejon et al. 2021). More concretely, EL222 could be regulated such that it is only active in a subpopulation of cells (Andersson et al. 2014) or during a specific time (Gifford et al. 2013) in development. Light may then be used to effect a genetic change that perturbs the developmental program in cell type specific and spatially or temporally localized fashion.

References

- Aditya, Chetan, François Bertaux, Gregory Batt, and Jakob Ruess. 2021. "A Light Tunable Differentiation System for the Creation and Control of Consortia in Yeast." *BioRxiv*.
- Aguilera-Castrejon, Alejandro, Bernardo Oldak, Tom Shani, Nadir Ghanem, Chen Itzkovich, Sharon Slomovich, Shadi Tarazi, et al. 2021. "Ex Utero Mouse Embryogenesis from Pre-Gastrulation to Late Organogenesis." *Nature*, 1–6.
- Ahmad, Mudassar, Melanie Hirz, Harald Pichler, and Helmut Schwab. 2014. "Protein Expression in *Pichia Pastoris*: Recent Achievements and Perspectives for Heterologous Protein Production." *Applied Microbiology and Biotechnology* 98 (12): 5301–17.
- Andersson, Robin, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, et al. 2014. "An Atlas of Active Enhancers across Human Cell Types and Tissues." *Nature* 507 (7493): 455–61.
- Atkinson, Mariette R, Michael A Savageau, Jesse T Myers, and Alexander J Ninfa. 2003. "Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in *Escherichia Coli*." *Cell* 113 (5): 597–607.
- Balagaddé, Frederick K, Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, Stephen R Quake, and Lingchong You. 2008. "A Synthetic *Escherichia Coli* Predator-Prey Ecosystem." *Molecular Systems Biology* 4 (1): 187.
- Barbier, İçvara, Rubén Perez-Carrasco, and Yolanda Schærli. 2020. "Controlling Spatiotemporal Pattern Formation in a Concentration Gradient with a Synthetic Toggle Switch." *Molecular Systems Biology* 16 (6): e9361.
- Basu, Subhayu, Yoram Gerchman, Cynthia H Collins, Frances H Arnold, and Ron Weiss. 2005. "A Synthetic Multicellular System for Programmed Pattern Formation." *Nature* 434 (7037): 1130–34.
- Becskei, Attila, Bertrand Séraphin, and Luis Serrano. 2001. "Positive Feedback in Eukaryotic Gene Networks: Cell Differentiation by Graded to Binary Response Conversion." *The EMBO Journal* 20 (10): 2528–35.
- Benzinger, Dirk, and Mustafa Khammash. 2018. "Pulsatile Inputs Achieve Tunable Attenuation of Gene Expression Variability and Graded Multi-Gene Regulation." *Nature Communications* 9 (1): 1–10.
- Bertaux, François, Sebastián Sosa Carrillo, Achille Fraisse, Chetan Aditya, Mariela Furstenheim, and Gregory Batt. 2020. "Enhancing Multi-Bioreactor Platforms for Automated Measurements and Reactive Experiment Control." *BioRxiv*.
- Brenner, Katie, Lingchong You, and Frances H Arnold. 2008. "Engineering Microbial Consortia: A New Frontier in Synthetic Biology." *Trends in Biotechnology* 26 (9): 483–89.
- Carrasco-López, César, Sergio A García-Echauri, Therese Kichuk, and José L Avalos. 2020. "Optogenetics and Biosensors Set the Stage for Metabolic Cybergenetics." *Current Opinion in Biotechnology* 65: 296–309.
- Chait, Remy, Jakob Ruess, Tobias Bergmiller, Gašper Tkačik, and C\ualin C Guet. 2017. "Shaping Bacterial Population Behavior through Computer-Interfaced Control of Individual Cells." *Nature Communications* 8 (1): 1–11.
- Chen, Kevin, Yixuan Zhu, Yongrong Zhang, Therwa Hamza, Hua Yu, Ashley Saint Fleur, James Galen, Zhiyong Yang, and Hanping Feng. 2020. "A Probiotic Yeast-Based Immunotherapy against *Clostridioides Difficile* Infection." *Science Translational Medicine* 12 (567).
- Colman-Lerner, Alejandro, Tina E. Chin, and Roger Brent. 2001. "Yeast Cbk1 and Mob2 Activate Daughter-Specific Genetic Programs to Induce Asymmetric Cell Fates." *Cell* 107 (6): 739–50.
- Cosma, Maria Pia. 2004. "Daughter-Specific Repression of *Saccharomyces Cerevisiae* HO: Ash1 Is the Commander." *EMBO Reports* 5 (10): 953–57.

- Decker, Carolyn J, and Roy Parker. 1993. "A Turnover Pathway for Both Stable and Unstable MRNAs in Yeast: Evidence for a Requirement for Deadenylation." *Genes & Development* 7 (8): 1632–43.
- Duddu, Atchuta Srinivas, Sarthak Sahoo, Souvadra Hati, Siddharth Jhunjhunwala, and Mohit Kumar Jolly. 2020. "Multi-Stability in Cellular Differentiation Enabled by a Network of Three Mutually Repressing Master Regulators." *Journal of the Royal Society Interface* 17 (170): 20200631.
- Duplus-Bottin, H el ene, Martin Spichty, G erard Triqueneaux, Christophe Place, Philippe Emmanuel Mangeot, Th eophile Ohlmann, Franck Vittoz, and Ga el Yvert. 2021. "A Single-Chain and Fast-Responding Light-Inducible Cre Recombinase as a Novel Optogenetic Switch." *Elife* 10: e61268.
- Fedorec, Alex J H, Behzad D Karkaria, Michael Sulu, and Chris P Barnes. 2021. "Single Strain Control of Microbial Consortia." *Nature Communications* 12 (1): 1–12.
- Fernandez-Rodr iguez, Jesus, Felix Moser, Miryoung Song, and Christopher A Voigt. 2017. "Engineering RGB Color Vision into Escherichia Coli." *Nature Chemical Biology* 13 (7): 706–8.
- Fox, Zachary R, Steven Fletcher, Achille Fraisse, Chetan Aditya, Sebasti an Sosa-Carrillo, S ebastien Gilles, Fran ois Bertaux, Jakob Ruess, and Gregory Batt. 2021. "MicroMator: Open and Flexible Software for Reactive Microscopy." *BioRxiv*.
- Fushimi, Keiji, and Rei Narikawa. 2019. "Cyanobacteriochromes: Photoreceptors Covering the Entire UV-to-Visible Spectrum." *Current Opinion in Structural Biology* 57: 39–46.
- Gardner, Timothy S, Charles R Cantor, and James J Collins. 2000. "Construction of a Genetic Toggle Switch in Escherichia Coli." *Nature* 403 (6767): 339–42.
- Gifford, Casey A, Michael J Ziller, Hongcang Gu, Cole Trapnell, Julie Donaghey, Alexander Tsankov, Alex K Shalek, et al. 2013. "Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells." *Cell* 153 (5): 1149–63.
- Hochrein, Lena, Fabian Machens, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2017. "PhiReX: A Programmable and Red Light-Regulated Protein Expression Switch for Yeast." *Nucleic Acids Research* 45 (15): 9193–9205.
- Hochrein, Lena, Leslie A. Mitchell, Karina Schulz, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2018. "L-SCRaMbLE as a Tool for Light-Controlled Cre-Mediated Recombination in Yeast." *Nature Communications* 9 (1): 1–10.
- Idiris, Alimjan, Hideki Tohda, Hiromichi Kumagai, and Kaoru Takegawa. 2010. "Engineering of Protein Secretion in Yeast: Strategies and Impact on Protein Production." *Applied Microbiology and Biotechnology* 86 (2): 403–17.
- Inda, Maria Eugenia, Esther Broset, Timothy K Lu, and Cesar de la Fuente-Nunez. 2019. "Emerging Frontiers in Microbiome Engineering." *Trends in Immunology* 40 (10): 952–73.
- Ispolatov, Iaroslav, Martin Ackermann, and Michael Doebeli. 2012. "Division of Labour and the Evolution of Multicellularity." *Proceedings of the Royal Society B: Biological Sciences* 279 (1734): 1768–76.
- Jang, Jaewan, Sherin McDonald, Maruti Uppalapati, and G Andrew Woolley. 2019. "Green, Orange, Red, and Far-Red Optogenetic Tools Derived from Cyanobacteriochromes." *BioRxiv*, 769422.
- Jia, Dongya, Mohit Kumar Jolly, William Harrison, Marcelo Boareto, Eshel Ben-Jacob, and Herbert Levine. 2017. "Operating Principles of Tristable Circuits Regulating Cellular Differentiation." *Physical Biology* 14 (3): 35007.
- Johnson, Heath E, Nareg J V Djabrayan, Stanislav Y Shvartsman, and Jared E Toettcher. 2020. "Optogenetic Rescue of a Patterning Mutant." *Current Biology* 30 (17): 3414–24.
- Karkaria, Behzad D, Alex J H Fedorec, and Chris P Barnes. 2021. "Automated Design of Synthetic Microbial Communities." *Nature Communications* 12 (1): 1–12.

- Kawano, Fuun, Risako Okazaki, Masayuki Yazawa, and Moritoshi Sato. 2016. "A Photoactivatable Cre-LoxP Recombination System for Optogenetic Genome Engineering." *Nature Chemical Biology* 12 (12): 1059–64.
- Kong, Wentao, David R Meldgin, James J Collins, and Ting Lu. 2018. "Designing Microbial Consortia with Defined Social Interactions." *Nature Chemical Biology* 14 (8): 821–29.
- Kylilis, Nicolas, Zoltan A Tuza, Guy-Bart Stan, and Karen M Polizzi. 2018. "Tools for Engineering Coordinated System Behaviour in Synthetic Microbial Consortia." *Nature Communications* 9 (1): 1–9.
- Lalwani, Makoto A, Samantha S Ip, Cesar Carrasco-Lopez, Catherine Day, Evan M Zhao, Hinako Kawabe, and José L Avalos. 2021. "Optogenetic Control of the Lac Operon for Bacterial Chemical and Protein Production." *Nature Chemical Biology* 17 (1): 71–79.
- Lambert, Talley J. 2019. "FPbase: A Community-Editable Fluorescent Protein Database." *Nature Methods* 16 (4): 277–78.
- LaSarre, Breah, Alexandra L McCully, Jay T Lennon, and James B McKinlay. 2017. "Microbial Mutualism Dynamics Governed by Dose-Dependent Toxicity of Cross-Fed Nutrients." *The ISME Journal* 11 (2): 337–48.
- Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. "A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly." *ACS Synthetic Biology* 4 (9): 975–86.
- Li, Zhenghong, Xiaonan Wang, and Haoran Zhang. 2019. "Balancing the Non-Linear Rosmarinic Acid Biosynthetic Pathway by Modular Co-Culture Engineering." *Metabolic Engineering* 54: 1–11.
- Liao, Michael J, M Omar Din, Lev Tsimring, and Jeff Hasty. 2019. "Rock-Paper-Scissors: Engineered Population Dynamics Increase Genetic Stability." *Science* 365 (6457): 1045–49.
- Lin, Qiuhui, Hao Qi, Yi Wu, and Yingjin Yuan. 2015. "Robust Orthogonal Recombination System for Versatile Genomic Elements Rearrangement in Yeast *Saccharomyces Cerevisiae*." *Scientific Reports* 5 (1): 1–8.
- Liu, Feng, Junwen Mao, Wentao Kong, Qiang Hua, Youjun Feng, Rashid Bashir, and Ting Lu. 2020. "Interaction Variability Shapes Succession of Synthetic Microbial Ecosystems." *Nature Communications* 11 (1): 1–13.
- Liu, Ping, Bojun Chen, and Zhao-Wen Wang. 2020. "GABAergic Motor Neurons Bias Locomotor Decision-Making in *C. Elegans*." *Nature Communications* 11 (1): 1–19.
- Liu, Wei, Zhouqing Luo, Yun Wang, Nhan T Pham, Laura Tuck, Irene Pérez-Pi, Longying Liu, et al. 2018. "Rapid Pathway Prototyping and Engineering Using in Vitro and in Vivo Synthetic Genome SCRaMbLE-in Methods." *Nature Communications* 9 (1): 1–12.
- Lugagne, Jean-Baptiste, Sebastián Sosa Carrillo, Melanie Kirch, Agnes Köhler, Gregory Batt, and Pascal Hersen. 2017. "Balancing a Genetic Toggle Switch by Real-Time Feedback Control and Periodic Forcing." *Nature Communications* 8 (1): 1–8.
- Maynard Smith, John, and Eors Szathmary. 1997. *The Major Transitions in Evolution*. Oxford University Press.
- Menolascina, Filippo, Gianfranco Fiore, Emanuele Orabona, Luca De Stefano, Mike Ferry, Jeff Hasty, Mario di Bernardo, and Diego di Bernardo. 2014. "In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks." *PLoS Comput Biol* 10 (5): e1003625.
- Miano, Arianna, Michael J Liao, and Jeff Hasty. 2020. "Inducible Cell-to-Cell Signaling for Tunable Dynamics in Microbial Communities." *Nature Communications* 11 (1): 1–8.
- Milias-Argeitis, Andreas, Marc Rullan, Stephanie K. Aoki, Peter Buchmann, and Mustafa Khammash. 2016. "Automated Optogenetic Feedback Control for Precise and Robust Regulation of Gene Expression and Cell Growth." *Nature Communications* 7 (1): 1–11.
- Milias-Argeitis, Andreas, Sean Summers, Jacob Stewart-Ornstein, Ignacio Zuleta, David Pincus, Hana El-Samad, Mustafa Khammash, and John Lygeros. 2011. "In Silico Feedback for in Vivo Regulation of a Gene Expression Circuit." *Nature Biotechnology* 29 (12): 1114–16.

- Motta-Mena, Laura B., Anna Reade, Michael J. Mallory, Spencer Glantz, Orion D. Weiner, Kristen W. Lynch, and Kevin H. Gardner. 2014. "An Optogenetic Gene Expression System with Rapid Activation and Deactivation Kinetics." *Nature Chemical Biology* 10 (3): 196–202.
- Naseri, Gita, Jessica Behrend, Lisa Rieper, and Bernd Mueller-Roeber. 2019. "COMPASS for Rapid Combinatorial Optimization of Biochemical Pathways Based on Artificial Transcription Factors." *Nature Communications* 10 (1): 1–18.
- Nern, A., B. D. Pfeiffer, K. Svoboda, and G. M. Rubin. 2011. "Multiple New Site-Specific Recombinases for Use in Manipulating Animal Genomes." *Proceedings of the National Academy of Sciences* 108 (34): 14198–203.
- Neufeld, Len, Fabienne Stassen, Ruth Sheppard, and Terry Gilman. 2016. "The New Plastics Economy: Rethinking the Future of Plastics." In *World Economic Forum*. Vol. 7.
- Olson, Evan J, Lucas A Hartsough, Brian P Landry, Raghav Shroff, and Jeffrey J Tabor. 2014. "Characterizing Bacterial Gene Circuit Dynamics with Optically Programmed Gene Expression Signals." *Nature Methods* 11 (4): 449–55.
- Panke-Buisse, Kevin, Angela C Poole, Julia K Goodrich, Ruth E Ley, and Jenny Kao-Kniffin. 2015. "Selection on Soil Microbiomes Reveals Reproducible Impacts on Plant Function." *The ISME Journal* 9 (4): 980–89.
- Perkins, Melinda Liu, Dirk Benzinger, Murat Arcak, and Mustafa Khammash. 2020. "Cell-in-the-Loop Pattern Formation with Optogenetically Emulated Cell-to-Cell Signaling." *Nature Communications* 11 (1): 1–10.
- Pothoulakis, Georgios, and Tom Ellis. 2018. "Construction of Hybrid Regulated Mother-Specific Yeast Promoters for Inducible Differential Gene Expression." *PLoS ONE* 13 (3): 1–14.
- Rapp, Kent M, Jackson P Jenkins, and Michael J Betenbaugh. 2020. "Partners for Life: Building Microbial Consortia for the Future." *Current Opinion in Biotechnology* 66: 292–300.
- Roell, Garrett W, Jian Zha, Rhiannon R Carr, Mattheos A Koffas, Stephen S Fong, and Yinjie J Tang. 2019. "Engineering Microbial Consortia by Division of Labor." *Microbial Cell Factories* 18 (1): 1–11.
- Roquet, Nathaniel, Ava P Soleimany, Alyssa C Ferris, Scott Aaronson, and Timothy K Lu. 2016. "Synthetic Recombinase-Based State Machines in Living Cells." *Science* 353 (6297).
- Rueffler, Claus, Joachim Hermisson, and Günter P Wagner. 2012. "Evolution of Functional Specialization and Division of Labor." *Proceedings of the National Academy of Sciences* 109 (6): E326–35.
- Rullan, Marc, Dirk Benzinger, Gregor W. Schmidt, Andreas Miliadis-Argeitis, and Mustafa Khammash. 2018. "An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation." *Molecular Cell* 70 (4): 745–756.e6.
- Salzano, Davide, Davide Fiore, and Mario di Bernardo. 2019. "Ratiometric Control for Differentiation of Cell Populations Endowed with Synthetic Toggle Switches." In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 927–32.
- Schmidl, Sebastian R, Ravi U Sheth, Andrew Wu, and Jeffrey J Tabor. 2014. "Refactoring and Optimization of Light-Switchable Escherichia Coli Two-Component Systems." *ACS Synthetic Biology* 3 (11): 820–31.
- Scott, Spencer R, M Omar Din, Philip Bittihn, Liyang Xiong, Lev S Tsimring, and Jeff Hasty. 2017. "A Stabilized Microbial Ecosystem of Self-Limiting Bacteria Using Synthetic Quorum-Regulated Lysis." *Nature Microbiology* 2 (8): 1–9.
- Shou, Wenying, Sri Ram, and Jose M G Vilar. 2007. "Synthetic Cooperation in Engineered Yeast Populations." *Proceedings of the National Academy of Sciences* 104 (6): 1877–82.
- Taslimi, Amir, Brian Zoltowski, Jose G. Miranda, Gopal P. Pathak, Robert M. Hughes, and Chandra L. Tucker. 2016. "Optimized Second-Generation CRY2-CIB Dimerizers and Photoactivatable Cre Recombinase." *Nature Chemical Biology* 12 (6): 425–30.

- To, Tsz-Leung, and Narendra Maheshri. 2010. "Noise Can Induce Bimodality in Positive Transcriptional Feedback Loops without Bistability." *Science* 327 (5969): 1142–45.
- Toettcher, Jared E, Delquin Gong, Wendell A Lim, and Orion D Weiner. 2011. "Light-Based Feedback for Controlling Intracellular Signaling Dynamics." *Nature Methods* 8 (10): 837–39.
- Tournier, V, C M Topham, A Gilles, B David, C Folgoas, E Moya-Leclair, E Kamionka, et al. 2020. "An Engineered PET Depolymerase to Break down and Recycle Plastic Bottles." *Nature* 580 (7802): 216–19.
- Vecchio, Domitilla Del, Yili Qian, Richard M Murray, and Eduardo D Sontag. 2018. "Future Systems and Control Research in Synthetic Biology." *Annual Reviews in Control* 45: 5–17.
- Wintermute, Edwin H, and Pamela A Silver. 2010. "Emergent Cooperation in Microbial Metabolism." *Molecular Systems Biology* 6 (1): 407.
- Wu, Stephen G, Lian He, Qingzhao Wang, and Yinjie J Tang. 2015. "An Ancient Chinese Wisdom for Metabolic Engineering: Yin-Yang." *Microbial Cell Factories* 14 (1): 1–9.
- Wyeth, N, and R Roseveare. 1973. "Biaxially Oriented Poly (Ethylene Terephthalate) Bottle." Google Patents.
- Yang, Yaoyu, Jennifer L Nemhauser, and Eric Klavins. 2019. "Synthetic Bistability and Differentiation in Yeast." *ACS Synthetic Biology* 8 (5): 929–36.
- Yoshida, Shosuke, Kazumi Hiraga, Toshihiko Takehana, Ikuo Taniguchi, Hironao Yamaji, Yasuhito Maeda, Kiyotsuna Toyohara, Kenji Miyamoto, Yoshiharu Kimura, and Kohei Oda. 2016. "A Bacterium That Degrades and Assimilates Poly (Ethylene Terephthalate)." *Science* 351 (6278): 1196–99.
- Youk, Hyun, and Wendell A. Lim. 2014. "Secreting and Sensing the Same Molecule Allows Cells to Achieve Versatile Social Behaviors." *Science* 343 (6171).
- Zhang, Xuwang, Dongli Bao, Maoting Li, Qidong Tang, Minghuo Wu, Hao Zhou, Lifen Liu, and Yuanyuan Qu. 2021. "Bioremediation of Petroleum Hydrocarbons by Alkali-Salt-Tolerant Microbial Consortia and Their Community Profiles." *Journal of Chemical Technology & Biotechnology* 96 (3): 809–17.
- Zhao, Evan M., Yanfei Zhang, Justin Mehl, Helen Park, Makoto A. Lalwani, Jared E. Toettcher, and José L. Avalos. 2018. "Optogenetic Regulation of Engineered Cellular Metabolism for Microbial Chemical Production." *Nature* 555 (7698): 683–87.
- Zhou, Kang, Kangjian Qiao, Steven Edgar, and Gregory Stephanopoulos. 2015. "Distributing a Metabolic Pathway among a Microbial Consortium Enhances Production of Natural Products." *Nature Biotechnology* 33 (4): 377–83.
- Zhuang, Wei-Qin, Shan Yi, Markus Bill, Vanessa L Brisson, Xueyang Feng, Yujie Men, Mark E Conrad, Yinjie J Tang, and Lisa Alvarez-Cohen. 2014. "Incomplete Wood-Ljungdahl Pathway Facilitates One-Carbon Metabolism in Organohalide-Respiring *Dehalococcoides Mccartyi*." *Proceedings of the National Academy of Sciences* 111 (17): 6419–24.

Chapter 4

APPLICATION TO CONTINUOUS HETEROLOGOUS EXPRESSION

“Nothing in biology makes sense except in the light of evolution.”
Theodosius Dobzhansky

Preface

Portions of this chapter as well as some figures have been adapted from Aditya *et al.* (2021).

Introduction

Growth production tradeoffs are borne out of an inexorable competition for resources between the cell's native genetic program primed to self-propagate and the exogenously introduced heterologous program designed to produce a valuable molecule that is devoid of any functional utility for the cell. This competition leads to a metabolic burden, which manifests as a decrease in the growth rate of producing cells. Consequently, any mutation that frees the cell from the burden of production (termed reverting mutation) will result in a higher fitness (on account of the faster growth) and confer a selective advantage to the mutated cell (reversion mutant or escaper). Escapers will spread in the population and, over time, supplant the population that still carries the exogenous program (Rugbjerg *et al.* 2018). This problem is exacerbated in continuous operation because constant dilution of the culture exerts a stronger selection pressure, essentially debilitating protein yield over time and rendering this approach ungainful over long periods. Batch operation deemphasizes such considerations by separating growth and production in time. Several strategies have been proposed to increase the resistance to reverting mutations. The most common strategy is to attenuate the effect of mutations at a single loci by using high copy plasmids and employing autoselection markers (Loison *et al.* 1986; Geymonat, Spanos, and Sedgwick 2007) or toxin-antitoxin systems (Kroll *et al.* 2010) or dependencies with the auxotrophic marker (Chen *et al.* 2012) to ensure large copy number of the plasmid. Alternatively, integrative plasmids that integrate in multiple copies, up to 30, have been reported in yeast (Gnügge, Liphardt, and Rudolf 2016).

Conversely, creation of host organisms (also called chassis organisms) with synthetic genomes (J. Liu *et al.* 2020; Richardson *et al.* 2017) tailored to decrease conflicts with the native machinery

or to increase genetic stability (Csörgő et al. 2012; Choi et al. 2015) are active areas of research. Lastly, control strategies to mitigate metabolic burden (Ceroni et al. 2018; Gupta et al. 2017) or induce metabolic addiction (Lv et al. 2020) based on ever increasing array of molecular controllers (Aoki et al. 2019) provide complementary strengths (Carrasco-López et al. 2020) to the genetic approaches described above.

These approaches could prove instrumental in overcoming the biological limitations that preclude heterologous expression in continuous mode; however, despite these developments there has not been a comprehensive characterization of continuous bioproduction in budding yeast.

Problem statement

Our differentiation system offers a novel way of skirting the growth-production tradeoff in continuous cultures since growth and production happen simultaneously but in different cells (growers and producers, respectively). This could provide insularity from reverting mutations by maintaining a population of cells that do not produce and therefore grow at the maximal rate. Consequently, any reverting mutation in the producers would not be able to spread in the population. At the same time, since there is no growth in producers, the competition for resources should be ameliorated. Therefore, we sought to interrogate whether heterologous expression can be increased over constitutive expression and induction based methods in a continuous setting using the differentiation system to compartmentalize growth and production in the population.

Strategy

We wanted to employ the differentiation system to decouple growth from production such that growth and production take place simultaneously but in different cells. Before differentiation, cells grow and do not produce the protein of interest (POI). Light induces differentiation, following which cells arrest and start producing the POI. In order to arrest the growth, I chose the downstream effector of the mating response pathway, FAR1 (Chang and Herskowitz 1990). Several facts informed this choice. First, physiological function of FAR1 is to arrest the growth in nature (Mckinney and Cross 1995) and therefore I expected it to be a more maintainable state than knocking out an essential gene. Second, because the dynamics of the overexpression are more reliable than the degradation and dilution of an essential protein after the corresponding gene has been knocked out. Third, mating factor induced arrest was shown to be a metabolically active state and increased the capacity of yeast cells to produce small metabolites (Williams et al. 2016).

For heterologous expression, I identified a strong orthogonal transcription factor from plants known to be 6 times as strong as pTDH3, one of the strongest promoters known in yeast (Nasari et al. 2017).

I coupled the differentiation system with growth arrest to separate growth from production and probed if this strategy would lead to an increase in heterologous expression over constitutive and inducible counterparts in continuous operation and/or genetic stability.

Results

Construction and characterization of constitutive and inducible expression platforms

To assess the suitability of my system for bioproduction, it was first necessary to establish a reference. For constitutive expression, I cloned a placeholder protein of interest (POI), mScarlet-I, under the control of pTDH3 promoter, the strongest native promoter in *S. cerevisiae*. Although pTDH3 is sensitive to glucose concentration (Jung et al. 1995), for the purpose of this study, ample glucose was readily supplied and therefore posed no threat to the ability of cells to express the POI driven by the pTDH3 promoter. In addition to this strain, other strains were also generated, which expressed different fluorescent proteins constitutively under pTDH3 control (**Figure 4.1**). In addition, I constructed a 4 color strain, which expressed all 4 fluorescent proteins constitutively under the control of pTDH3 promoter, to know how protein production would be affected by the presence of other heterologous proteins.

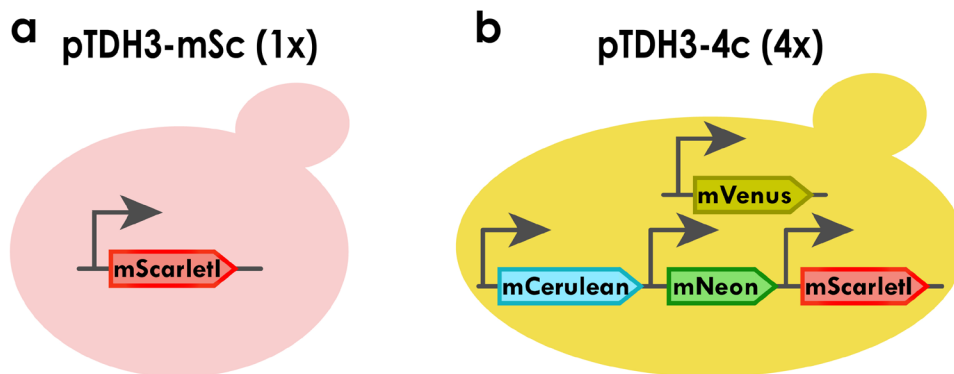


Figure 4.1. Construction of constitutive expression reference. a. pTDH3 driven mScarlet-I (POI). b. 4 color strain, 4 different fluorescent reporters under the control of the pTDH3 promoter. mCerulean, mNeonGreen, and mScarlet-I were integrated in URA locus while mVenus was integrated in the HO locus.

All strains were cultured in the turbidostat and their fluorescence levels were probed using the cytometer. I observed that increase in fluorescence over WT (that carried no fluorescent proteins) varied for each fluorescent protein despite each construction expressing the respective fluorophore from the same promoter and terminator (**Figure 4.2**). This was perhaps because of

each fluorescent protein having different spectral properties (**Methods**) and different sensitivities of the instrument to the various fluorophores. I found that expressing the POI alone or in presence of three other constitutively expressed proteins did not make a significant difference in the final expression levels (**Figure 4.2, bottom left**). I did not observe a significant growth defect in either single color or the 4-color strain.

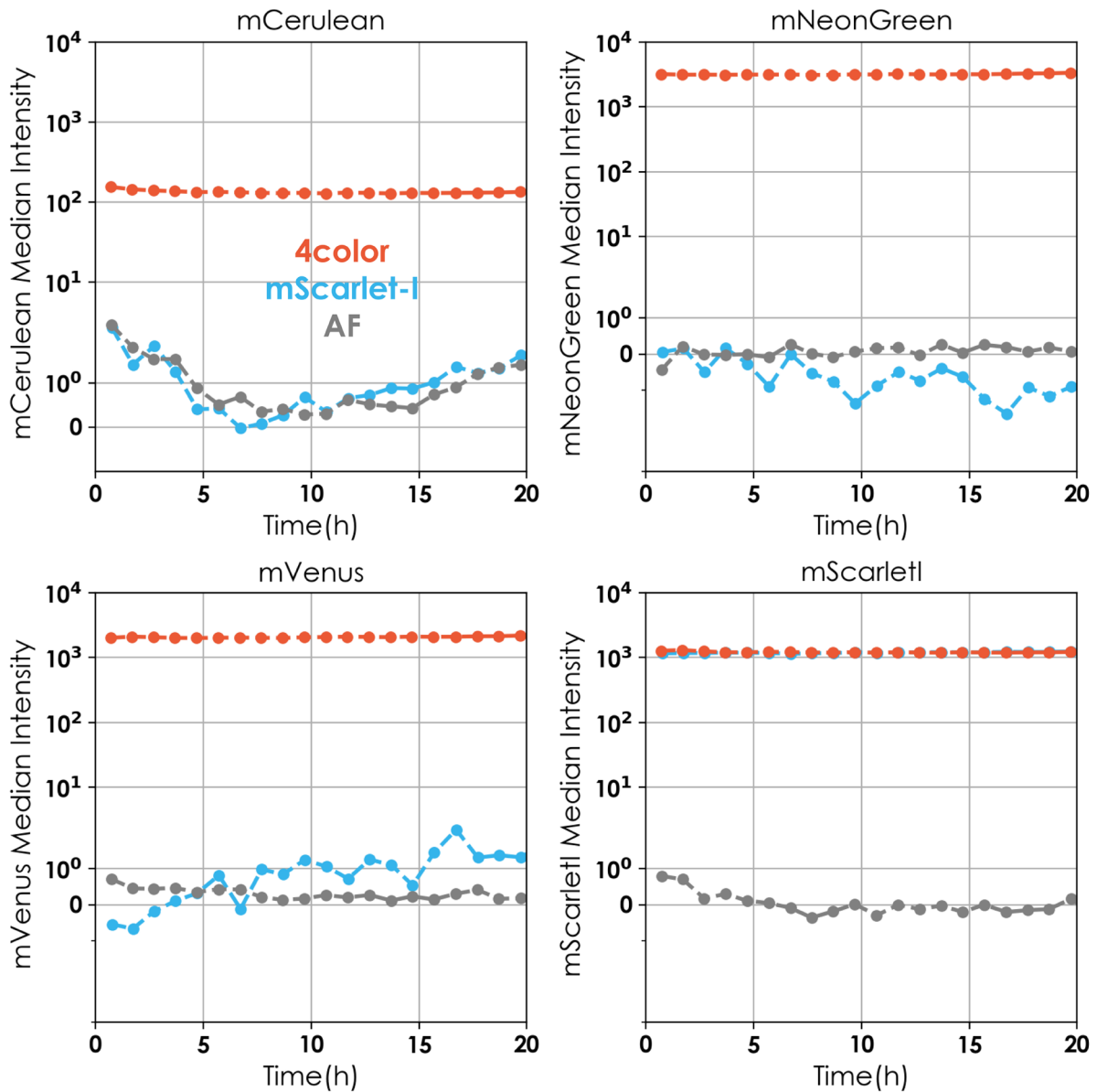


Figure 4.2. Fluorescence levels for **constitutively** driven fluorescent proteins **mScarletI** (blue) and the **4 color strain** (red) for each of the four fluorescent proteins measured in the cytometer over time. WT strain lacking any fluorescent protein is shown in grey. Lines represent median fluorescence values from one experiment and circles signify the time of sampling. For mScarlet-I, mScarlet-I overlaps with the 4-color data. Note that the y-axis is linear between 0 to 1 and logarithmic between 1 to 10,000.

For the induction reference, I identified a strong orthogonal transcription factor, ATAF1, (Naseri et al 2017) that was shown to be 6 times as strong as pTDH3. I first cloned ATAF1 under the control of constitutive promoter, pTDH3, with different fluorescent proteins (mCerulean, mNeonGreen and mScarlet-l) under the control of pATAF1 (ATAF1 specific promoter with 4 binding sites). However, this construction yielded very few transformants and did not lead to any detectable levels of expression in the transformed colonies. Postulating that the cells could not support the burden, I constructed an opto-inducible version of ATAF1 under the control of the strongest EL222 promoter from the Benzinger and Khammash, pEL222 6x bs. The maximum levels of induction with this promoter (at intensity 40) were analogous to pTDH3 driven expression. I introduced the three different fluorescent proteins under the control of 4X bs pATAF1. These strains are referred to as optoATAF1-X strains, where X denotes the protein under the control of ATAF1.

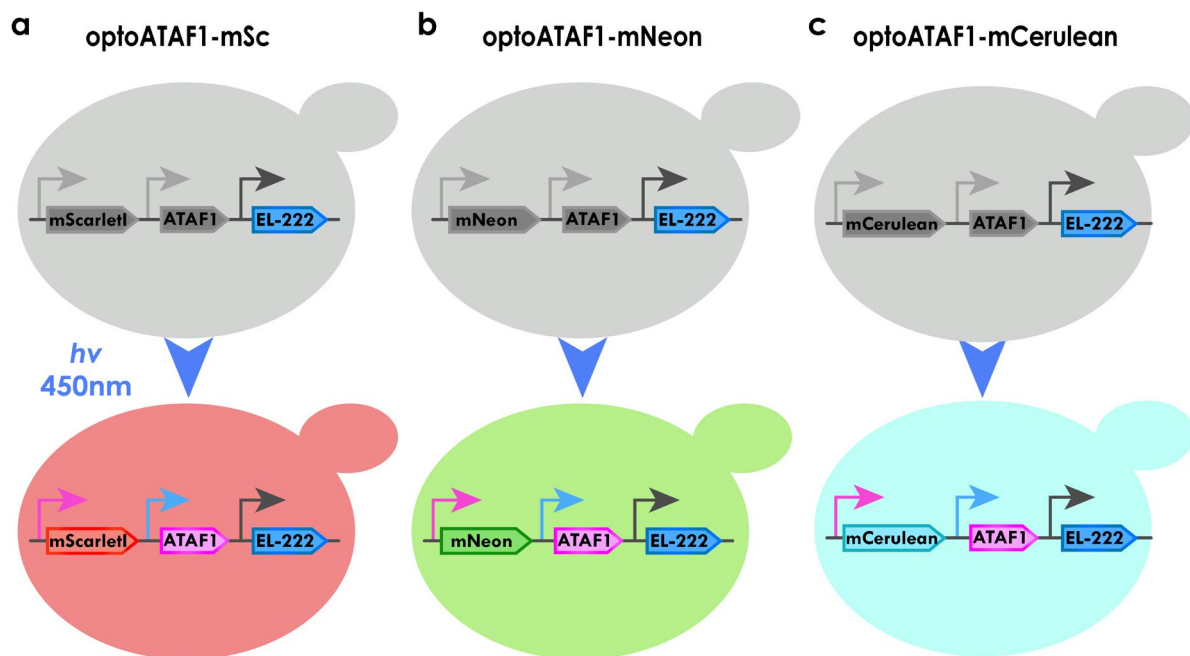


Figure 4.3. Design of optoATAF1-X strains. The optogenetic transcription factor EL-222 is under the control of a constitutive promoter, pTDH3, and expresses ATAF1 transcription factor in presence of blue light. ATAF1 drives the expression of X. Blue, magenta, and dark grey arrows represent pEL222, pATAF1, and pTDH3, respectively. Light grey arrows in the grey cells indicate no active expression.

I cultured cells harbouring the optoATAF1 expression module in the turbidostat under dark conditions overnight before inducing them with continuous light. I noticed that the induction was fast and 3 hours of induction was sufficient to obtain higher levels of expression than pTDH3. The induction dynamics were similar to those of EL222 activation (**Figure 3.3a**) in the first 10 hours of induction. At $t=10h$, OptoATAF1 strains reached much higher levels of expression than

pTDH3 driven constitutive expression (**Figure 4.4c**). Surprisingly, the fold increase over pTDH3 driven expression were different for different fluorescent proteins but, in each case, increased linearly with time until $t=10$ h. The optoATAF1 driven expression of fluorescent proteins at $t = 0$, before any light induction, was marginally higher than the autofluorescence background strain which suggested slight leakage (compare **Figure 4.2**, grey curve and **4.4a, b, c**, timepoint at $t=0$). Fold induction over dark was calculated after deconvolution. Contrary to fold increase over pTDH3 expression, the fold induction over dark showed remarkable similarity. It is evident that optoATAF1 driven expression has a large dynamic range with > 800 fold induction over dark for all three distinct proteins.

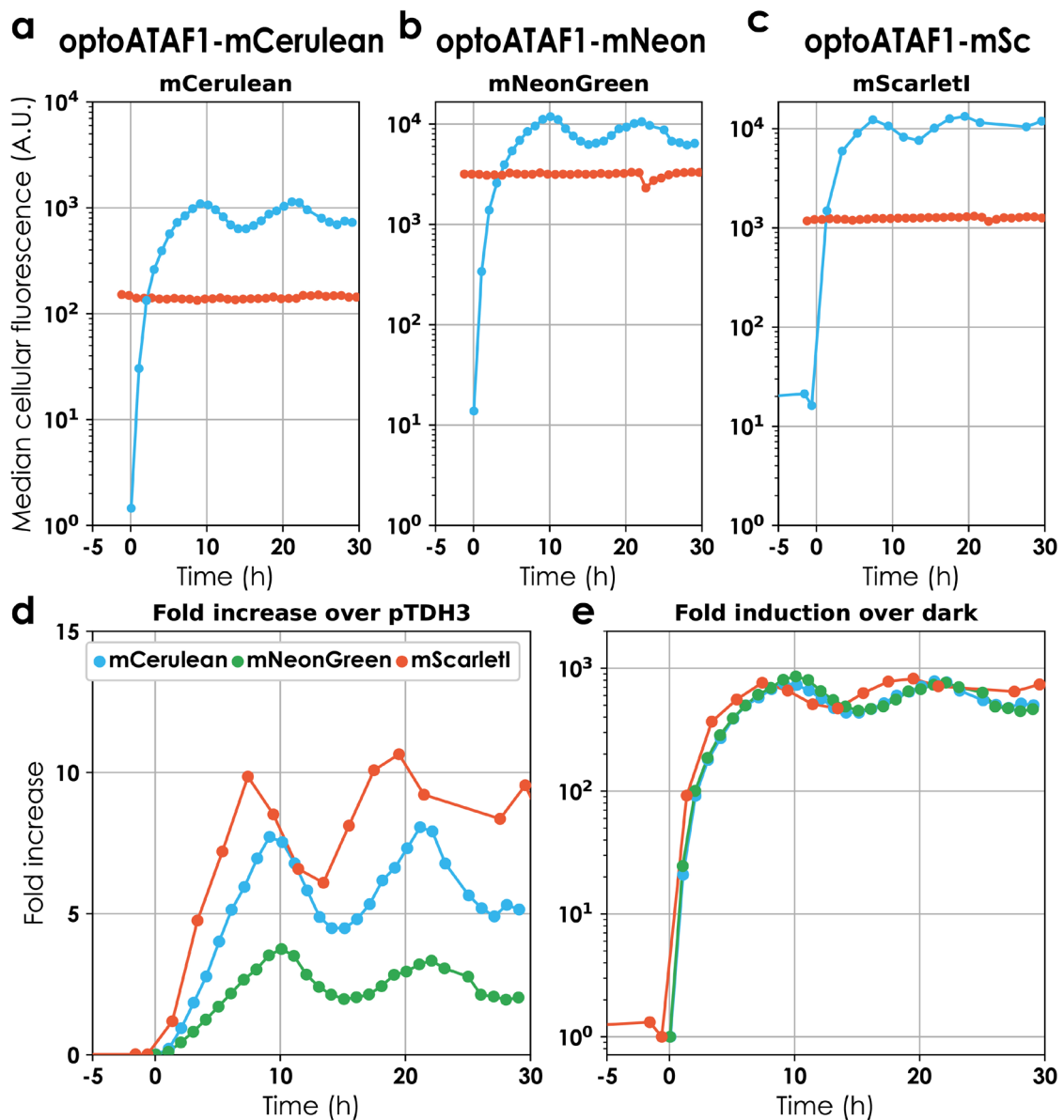


Figure 4.4. optoATAF1 based heterologous expression. Cultures in exponential phase were induced at $t=0$. **a, b, c.** Induction dynamics for optoATAF1-X (blue) compared to pTDH3-X (red). Solid lines represent the

deconvolved median cellular fluorescence from one experiment and circles signify the time of samplings.
d. Fold increase over pTDH3 fluorescence. Line represents values from one experiment and the circles signify times of sampling. Differences might be explained by photobleaching. **e. Fold induction over dark.** Line represents values from one experiment and the circles signify times of sampling.

However, the induction dynamics showed oscillations in median cellular fluorescence post $t=10\text{h}$. This was perhaps because of the strong growth defect that followed induction. Curiously, this slowdown was not associated with heterologous expression as the targetless opto-ATAF1 showed comparable growth defect (**Figure 4.5a**). I note that this growth defect is not due to the presence of blue light because a strain carrying pTDH3-EL222 was used as a control. This made the perceived increase in expression questionable since, even if the production was at the same rate as pTDH3, the decrease in cytoplasmic dilution (due to the slower growth rate) could lead to higher accumulation and therefore, a higher readout of fluorescence. In order to dispel this doubt, I cloned pTDH3-mScarlet in an optoATAF1 strain (without any downstream reporter) and measured the increase in fluorescence in both light and dark conditions in batch after 10h of induction. I observed that the slowdown did contribute to an increase in cellular fluorescence. However, opto-ATAF1 driven expression remained 6 times higher than pTDH3 driven expression (**Figure 4.5b**).

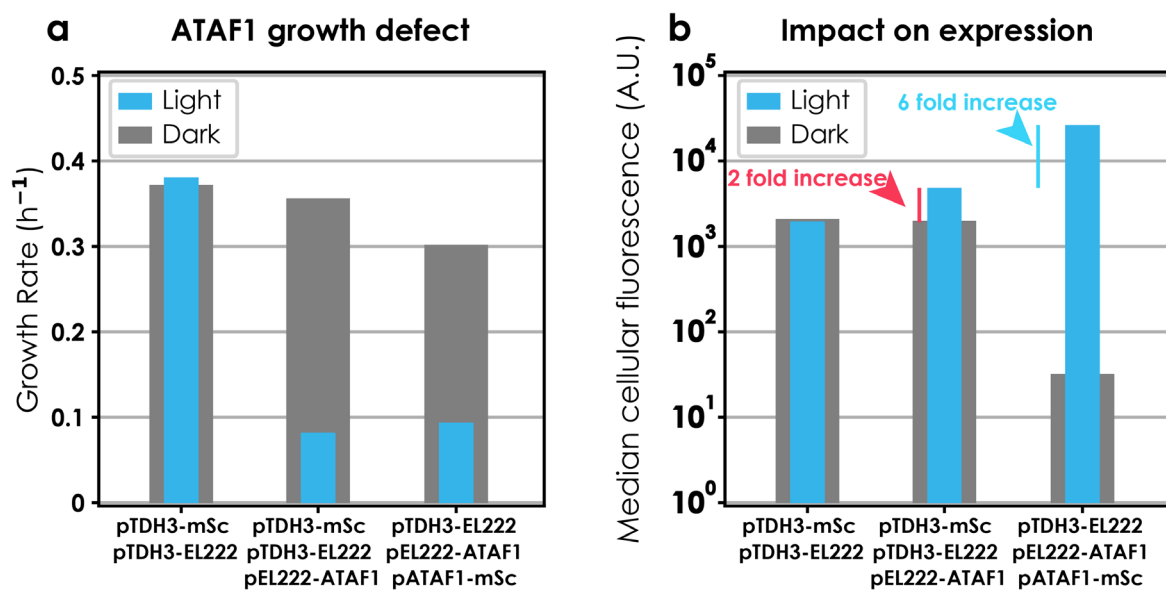


Figure 4.5. optoATAF1 based heterologous expression. **a. Growth defect** due to light induction in optoATAF1 strains compared to strain expressing EL222. Bars represent growth rate (h^{-1}) calculated from a single experiment in batch. **b. Heterologous expression levels** in light and dark conditions. Bars represent median values from a single experiment. Red vertical line denotes the increase in production due to the growth arrest and decrease in cytoplasmic dilution due to cell division. Blue vertical line represent real increase in production after accounting for the growth slowdown.

I conclude that the orthogonal transcription factor, ATAF1, is a potent driver of heterologous expression even if it results in a significant growth defect (~6 fold higher production).

optoATAF1 driven FAR1M mediated growth arrest

In order to arrest the cells, we hijacked the mating factor pathway by overexpressing the downstream effector FAR1 (Chang and Herskowitz 1990). FAR1 is an inhibitor of cyclin dependent kinase activity. Under physiological conditions, FAR1 is expressed in the presence of alpha factor and inhibits the interaction of CDC28 with the G1 cyclins. This is requisite for progression of cell cycle from G1 to S. Consequently, the cells arrest in G1. More specifically, I used a FAR1 variant, FAR1-22 that had been shown to arrest the growth upon overexpression in the absence of alpha-factor (S Henchoz et al. 1997). I performed preliminary experiments with light and galactose inducible FAR1M (FAR1-22 fused at C-terminal with mCerulean) in turbidostat, batch and the microscope. The quality of the generated data was not high enough to merit a place in this chapter, however, curious readers are referred to Annex 2 for more details. Results suggested that the degree of growth defect was dependent on the expression levels of FAR1M. Knowing that ATAF1 driven expression is high, I cloned FAR1M under the control of pATAF1 in an optoATAF1 strain.

optoATAF1-FAR1M

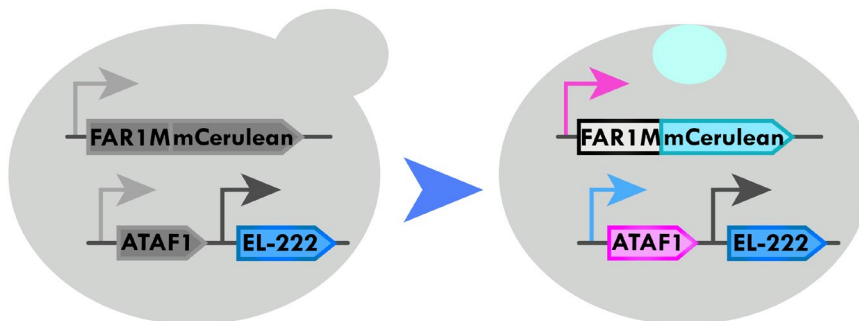


Figure 4.6. Construction of opto-differentiation-ATAF1-FAR1M strain. The optogenetic transcription factor EL-22 is under the control of a constitutive promoter, pTDH3, and expresses ATAF1 transcription factor in presence of blue light. ATAF1 drives the expression of FAR1M. The cyan disk indicates the nuclear localization of FAR1M.

I cultured optoATAF1-FAR1M continuously in the turbidostat in the exponential phase. After overnight culture in dark, cultures were induced with continuous light. I observed that the growth defect for OptoATAF1-FAR1M was more pronounced than optoATAF1-mSc and targetless opto-ATAF1 suggesting that the arrest is effected by FAR1-M overexpression (**Figure 4.7**).

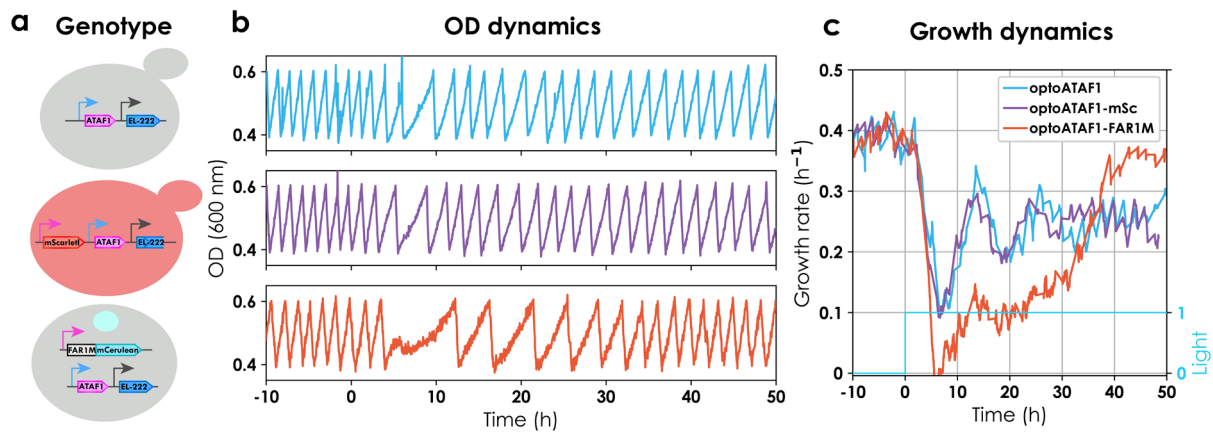


Figure 4.7. Characterization of growth arrest in optoATAF1-FAR1M. **a. Genotype** of the strains used in the experiment. **b. OD dynamics.** Lines represent OD at 600nm. Colors signify strains (see legend of **c**). Cells were induced at $t=0$. All experiments were conducted in the turbidostat grow and dilute mode cycling between OD 0.4-0.6. Data for a single experiment is shown. **c. Growth dynamics** under full induction. Red, blue and violet lines represent instantaneous growth rate (h^{-1}). Cyan line reflects induction.

The growth defect was drastic enough to bring the OD to a complete halt. I note that the optoATAF1-FAR1M cells escaped the growth arrest after 15h. Curiously, the escaped cells were even able to overcome the growth defect due to ATAF1 expression hinting that the mutants were able to shutdown ATAF1 expression.

Coupling differentiation with growth arrest

Seeing that optoATAF1 driven FAR1M was sufficient to completely arrest the growth, I decided to couple it to the differentiation system. In order to achieve that, I designed a construction that differed from the dual reporter (**Figure 3.1**). Notably, EL222 replaced the mCerulean and ATAF1 replaced mNeonGreen such that before recombination, EL222 is expressed constitutively and after recombination, ATAF1 starts being expressed. This construction was integrated in a strain carrying FAR1M under the control of pATAF1. Lastly, a cassette expressing Cre under the control of pEL222 was integrated in a separate locus. The complete design of the circuit is depicted below in (**Figure 4.8**).

L3-17

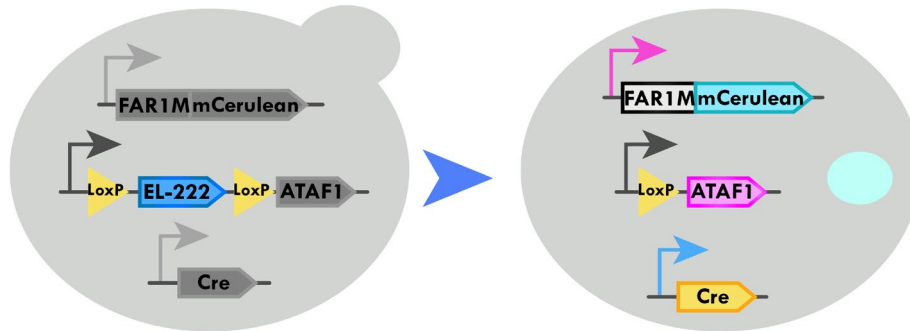


Figure 4.8. Construction of L3-17 strain. EL222 drives Cre and after the recombination ATAF1 is constitutively expressed from a pTDH3 promoter, however, the levels of expression from this pTDH3-LOXP promoter are half of the pEL222 promoter. ATAF1 subsequently drives the expression of FAR1-mCerulean.

In order to estimate the extent to which growth is diminished upon differentiation, I conducted competition experiments in the microscope. The experiment consisted of co-culturing L3-17 with the EL222 reporter strain (**Figure 3.1**) as the reference in a microfluidic chamber starting from an approximately 1:1 ratio. The induction was in the form of 400ms periodic pulses every 10 minutes. This was sufficient to cause 100% recombination in every field of view by 3h (**Figure 4.9a**) as deduced by observing the nuclear localization of mCerulean fluorescence. I estimated the growth rate by calculating the slope of $\log(\# \text{ of cells})$ with respect to time (**Methods, 6.11**). The growth rate was severely impacted (**Figure 4.9b**). I observed that the majority of the differentiated cells become large and stop dividing. A proportion of these cells died over the course of the experiment (not quantified). Some differentiated cells continued to divide, albeit at an increased division time (**Annex 1, Figure A1.2**). Observing the nuclear fluorescence of mCerulean revealed the presence of abnormal nuclear physiology including the presence of micronuclei, anaphase bridges and multinucleated yeast cells.

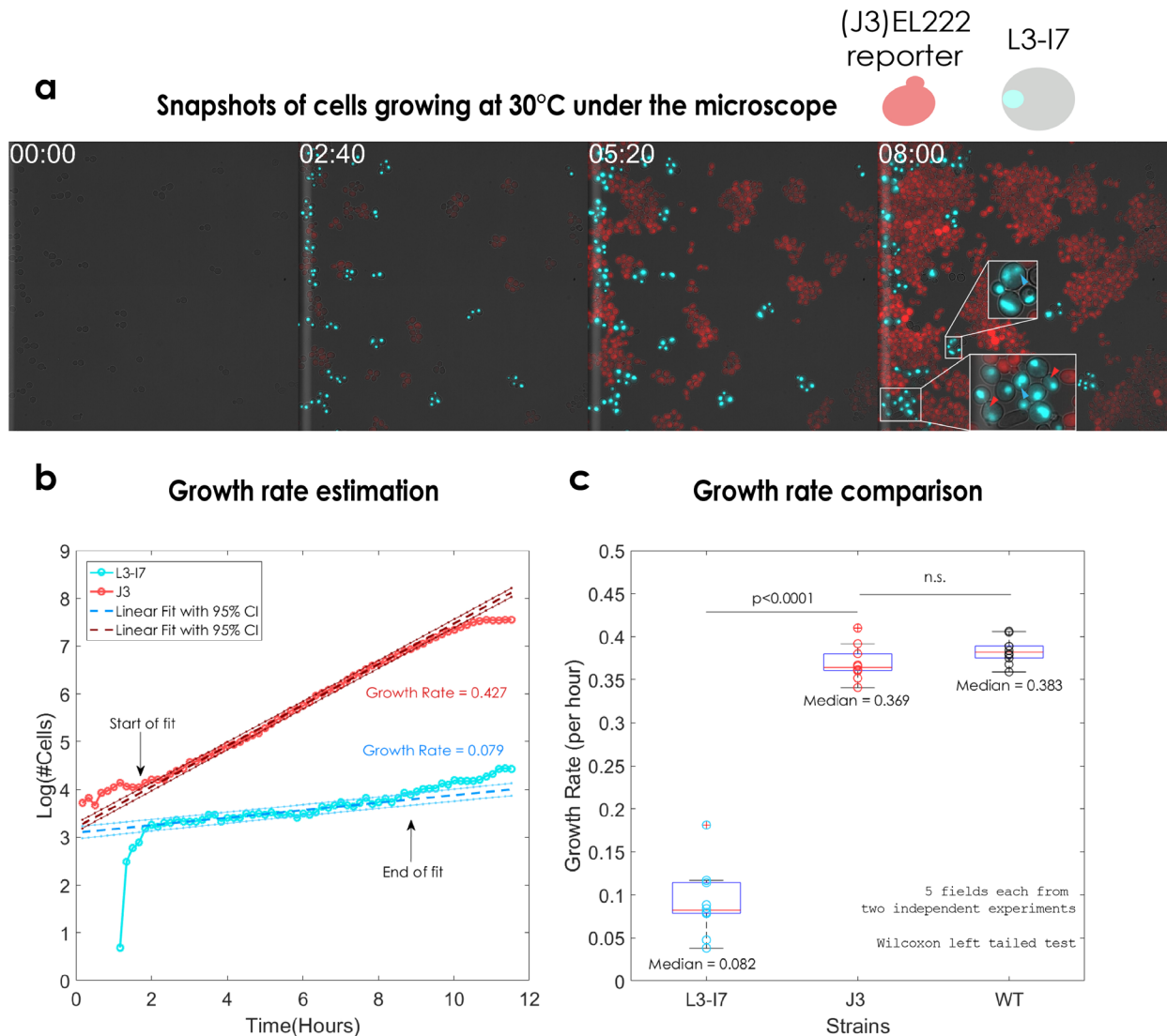


Figure 4.9. Competition experiment in the microfluidic chamber. a. The competition of L3-17 with the EL222 reporter. Cells were co-cultured in roughly 1:1 ratio at $t = 0$ and induction was started in the form of 400ms pulses every 10 minutes. This was sufficient to achieve 100% recombination as evidenced at $t=2:20h$. In the period between 2:20 to 8:00, little to no growth was registered. $t=8h$ is a striking demonstration of the growth arrest. Static images shown here are taken from a representative field. The insets at $t=8h$ indicate abnormal nuclear physiology. These include anaphase bridges (red arrows) and micronuclei (blue arrows). **b. Growth rate estimation from cell counts.** The circles signify log (cell count) trajectories over time for each strain in the representative field of view. Solid lines represent linear fits and dotted lines indicate 95% CI. The fit was started when all the cells became fluorescent and lasted until the field of view remained a monolayer. L3-17 could be discriminated from the EL222 reporter on the basis of cellular fluorescence. **c.** Collated results of fits from two different experiments each with 5 fields of view. The statistical test used is Wilcoxon left tailed test. L3-17 is a moniker for the strain described in Figure 4.8. J3 is a moniker for EL222 reporter described in Figure 3.1.

In order to characterize the differentiation dynamics of this strain and observe the ensuing growth arrest for extended periods, I cultured L3-17 cells in the turbidostat. Exponential cultures

were induced after an overnight of growth in dark with two induction profiles, continuous light and 50% duty cycle and a period of 2h with two replicates each. In both cases, the differentiation rate was high and >90% cells recombined within 8 hours of induction (**Figure 4.10a**). I note that the quantification of differentiation was not perfect because the nuclear localization of mCerulean fluorescence resulted in weaker readout than intercellular expression (of mCerulean). This meant that the fluorescence distribution of the differentiated and non-differentiated cells overlapped leading to the presence of a small population of seemingly differentiated cells in the dark condition. I am certain that these cells are not differentiated based on microscopy data (**Figure 4.9a**). The growth rate also crashed for both conditions however, the defect was stronger and longer for the continuous light condition (**Figure 4.10b**). In both cases, I observed that the differentiated fraction decreased over time along with a recovery in the growth rate despite the presence of light stimulation. In accordance with the differentiated fraction reduction, the amount of fluorescence also decreased for full light condition more than 50% (data not shown) perhaps due to the bleaching effect of blue light on mCerulean fluorescence. The recovery in growth rate was more gradual for the 50% duty cycle as opposed to the “sudden” recovery of growth rate in continuous light conditions. In the case of full light, seemingly a non-fluorescent population overtook the culture around the time of recovery. The dynamics of this population were near identical in the two replicates (**Figure 4.10a**). I suspect that this population was composed of mutants that had lost the ability to express FAR1M. I saved the cultures that escaped for possible analysis of mutations. Oddly, the 50% light seems to be as good as full light in terms of differentiation rate. This suggests that the modified system retains sensitivity to induction.

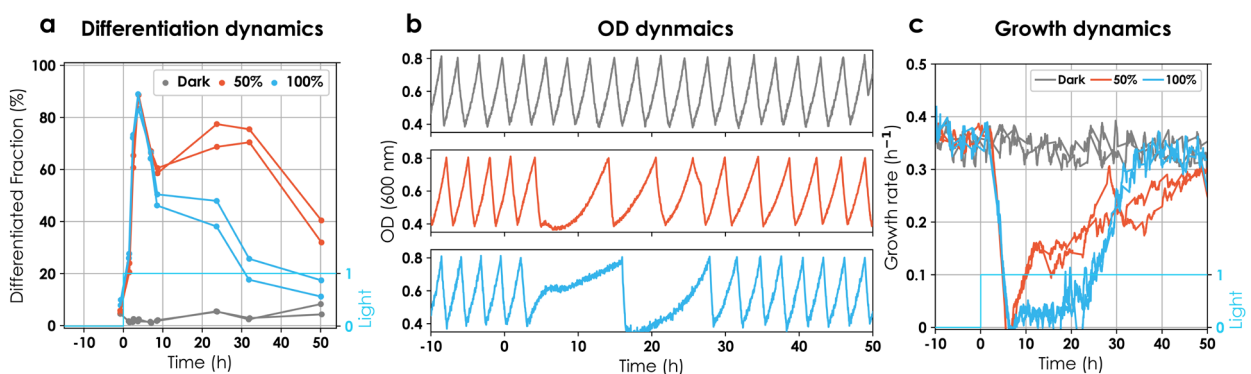


Figure 4.10. Induction in the turbidostat. **a. Differentiation dynamics** of strain L3-17 in duplicates for two different light doses, 50% duty cycle (red) and continuous light (blue) for a period of 2h. Control cultures that were kept in the dark are shown in grey. Lines represent differentiated fraction (calculated by applying a threshold on mCerulean) and circles signify time of sampling. Cyan line indicates the start of induction at $t=0$. **b. OD dynamics.** Lines represent OD at 600nm. Colors signify conditions (see legend of

b). All experiments were conducted in the turbidostat grow and dilute mode cycling between OD 0.4-0.8. Data for a single experiment is shown. **c. Growth rate dynamics** pre and post induction corresponding to two light profiles and dark. In each case, lines represent growth rate (h^{-1}). Cyan line indicates the start of induction at $t=0$.

The growth arrest, though, is markedly different between the two conditions. For both conditions, the growth rate decreases precipitously but the growth arrest is more sustained for full light ($\sim 20h$ doubling time as opposed to $\sim 10h$ (**Figure 4.10 b**)). This is particularly interesting because, post differentiation regardless of what light profile the cells received the observed behaviour should be similar. I also note here that the experiment had multiple dark controls that showed no growth defect or change in differentiated fraction over time. In addition, worthy of note is the fact that the rate of differentiation is faster than the dual reporter. This may simply be explained by a faster maturation time for mCerulean. To finish, I add that there was remarkable reproducibility between replicates.

Given these results, I am led to conclude that the coupling of the differentiation system with the growth arrest was successful and the strain L3-17 elicits a strong growth arrest upon differentiation.

Construction of growth arrest upon Differentiation strains that produce a POI (GAuDi)

In order to couple protein production to growth arrest after differentiation, I adapted the circuit to include a transcriptional unit driving a poi (mScarlet-I) under the control of ATAF1 (**Figure 4.11 a**). Concretely, prior to recombination, cells express mNeonGreen. After recombination, cells start expressing ATAF1 that, in turn, drives expression of FAR1-M (growth arrest) and mScarlet-I (protein production) with pATAF1. The circuit was originally designed such that EL222 and CRE transcriptional units were included within the to-be-excised region. This was done in order to remove the burden of constitutive expression of EL222 transcription factor (as negligible as may be). Since I expected the cell to **Growth Arrest upon Differentiation**, I named it GAuDi00.

I cultured cells continuously in the turbidostat in light and in darkness. Cultures were sampled to the cytometer every hour and the growth rate was monitored via OD measurements. I observed that before induction, the population was predominantly green, suggesting no recombination in the dark (to be precise around 2% of the population was recombined). Induction with blue light resulted in a gradual change in the population from green to red. The entirety of the culture became differentiated after 60h of culture in continuous light (**Figure 4.12a**).

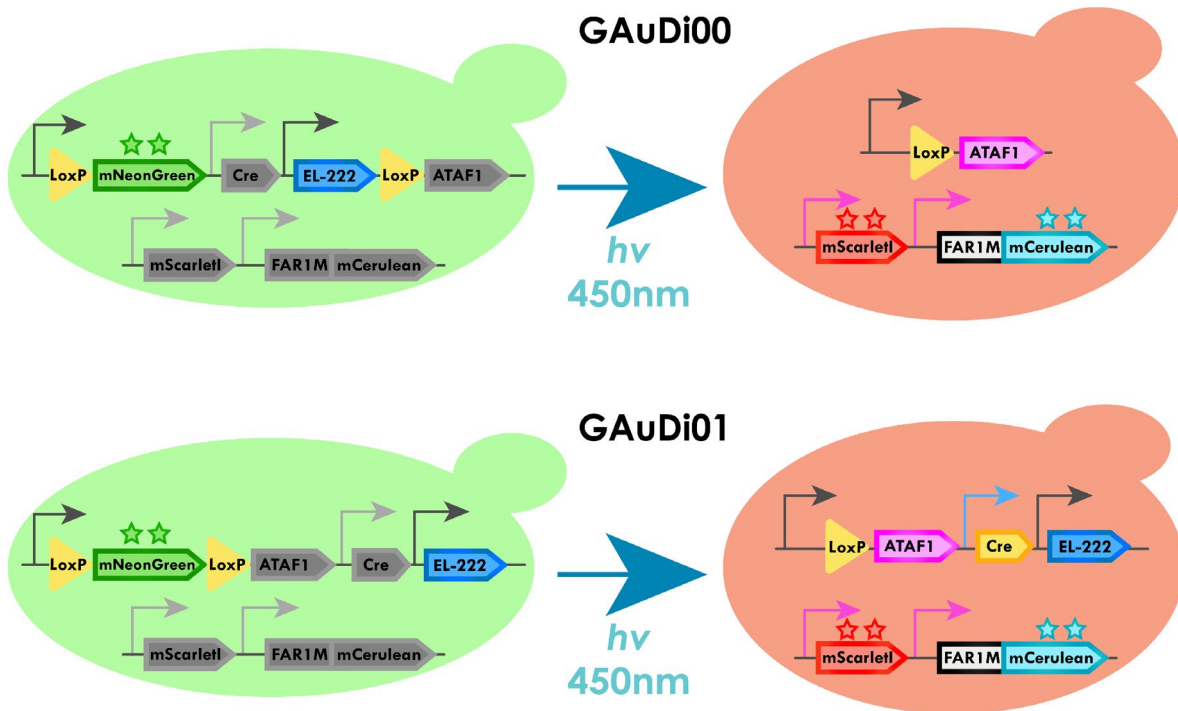


Figure 4.11. Circuits for GAuDi00 and GAuDi01. GAuDi00 contains EL222 and Cre expression cassettes within the region to be excised resulting in slow recombination. To ameliorate this, a shorter version was designed (GAuDi01) which excluded EL222 and Cre transcriptional units. This restored the recombination rate.

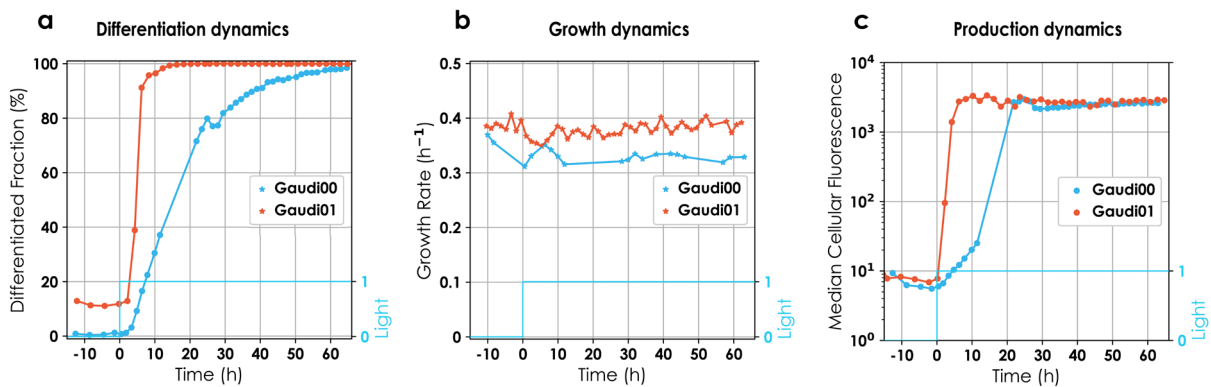


Figure 4.12. a. Differentiation behaviour for GAuDi00 (blue) and GAuDi01 (red) under continuous light. Cells were induced with blue light at $t=0$ as indicated by the cyan line. Blue and red lines represent differentiated fraction over time and circles signify times of sampling for GAuDi00 and GAuDi01, respectively. We observed $>95\%$ recombination for both strains. However, GAuDi00 displayed a much slower response. For the experiment shown here, a significant part of the population for GAuDi01 was recombined prior to light induction. **b. Growth rate under continuous light.** Solid lines represent growth rate. Marginal decrease in the growth rate in response to light was observed in both cases. **c. Production dynamics.** Lines represent median cellular fluorescence from a single experiment. Circles signify times of sampling. GAuDi01 was evidently faster to reach steady state levels of population fluorescence but GAuDi00 eventually settled at the same levels.

This construct, however, had multiple shortcomings:

- 1) The recombination rate was prohibitively slower (8 fold difference) than the previous construct (compare **Figure 4.12a** with **Figure 4.10a** (blue curve) and **Figure 3.10a** (blue curve)).
- 2) The growth arrest was significantly reduced (compare blue line in **Figure 4.12b** with red and blue lines in **Figure 4.8b & 4.10b** respectively).
- 3) The production levels were at least two fold lower than what was previously observed (compare blue lines in **Figure 4.4c and 4.12c**).

I sought advice from the expert on recombination systems in Institut Pasteur, Didier Mazel, who remarked that the length of the to-be-excised fraction plays a critical role in determining the efficiency of recombination (personal conversation). Noting that the length of the to-be-excised fragment was significantly larger than the one used in Dual Reporter and L3-I7 (**Figure 3.1, Figure 4.8**), I constructed another version of the recombination cassette with a smaller to-be-excised fragment (by excluding the EL222 and Cre transcriptional units). This ameliorated the slow recombination rate (**Figure 4.12a**) but growth arrest and production remained limited (**Figure 4.12b & 4.12c**). More concretely, the recombination rate for GAuDi00 is approximately 5 times lower than GAuDi01. It takes GAuDi00 ~40h to reach 90% recombined cells whereas GAuDi01 takes < 8h. The behaviour of the faster circuit was as predicted and analogous to the results obtained with the dual reporter. The final levels of expression are similar as is the observed marginal growth arrest between dark and light in GAuDi00 and GAuDi01.

To explain the reduced growth arrest and decline in production, I posited that the ATAF1 transcription factor levels post recombination are not high enough to drive the expression of two target genes. This is because expression from pTDH3-LoxP is roughly half of the expression levels observed for fully induced pEL222. To counter this “scarcity” of TF, I added a positive feedback (FB) loop to the circuit in which ATAF1 drives its own expression. In order to assess the efficacy of the feedback loop, I first introduced it in the optoATAF1 strain background. Preliminary experiments in batch suggested that the increase in the production levels was negligible. This led me to believe that the ATAF1 levels in a single target case (optoATAF1-1XS) are enough to saturate the promoter. To assess whether the FB loop is able to maintain the production levels in case of two target genes, I generated a two target optoATAF1 strain driving mScarlet-I with ATAF1 feedback (optoATAF1-2XS FB). optoATAF1-2XS FB resulted in nearly identical median fluorescence levels as optoATAF1-1XS. This was particularly puzzling, as neither the additional copy nor the feedback seemed to increase the levels of expression. To better understand the impact of the feedback loop, I generated two target versions of optoATAF1 driving two different

fluorescent proteins (mScarlet-I & mCerulean (optoATAF1-2XC) or mScarlet-I & mNeonGreen (optoATAF1-2XN)) with or without the ATAF1 feedback loop (FB). These cells were maintained in exponential phase in the turbidostat and induced with continuous light. In order, to compare the production across different fluorescent proteins, all fluorescence were normalized by the pTDH3 reference (**Figure 4.2**) such that the results obtained were in relative promoter units (R.P.U.) and therefore directly comparable with each other.

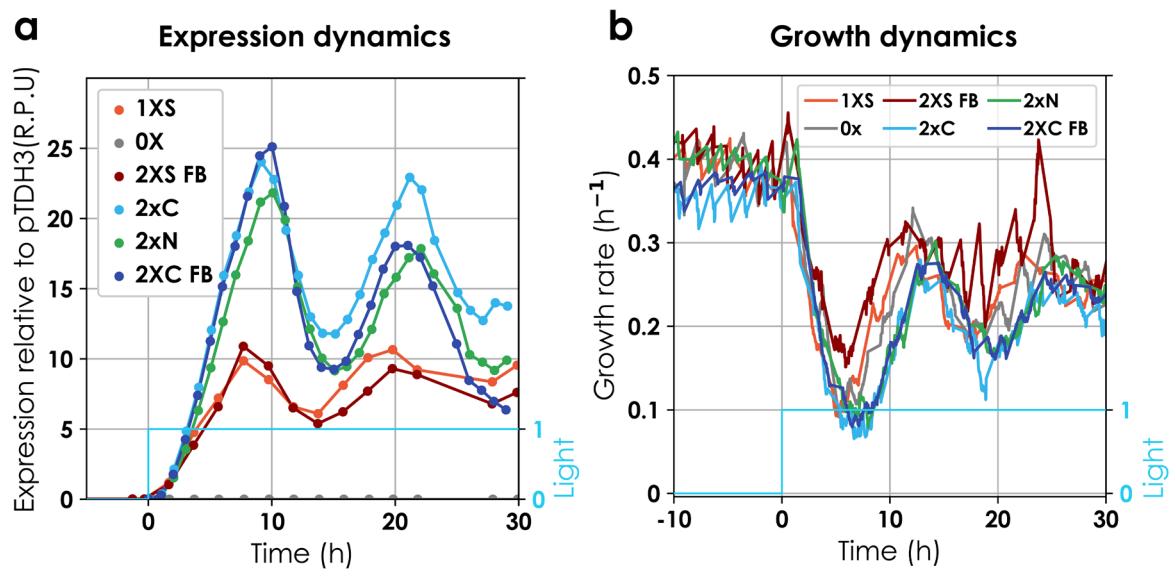


Figure 4.13. a. Dynamics of heterologous expression for different optoATAF1 strains under continuous light. Cells were induced with blue light at $t=0$ as indicated by the cyan line. Lines represent pTDH3 normalized expression over time and circles signify times of sampling. For optoATAF1 2X C&N (FB), expression relative to pTDH3 was calculated by using the respective mean fluorescence from pTDH3 driven controls for each fluorescent protein and summed to plot the circles. **b. Growth rate dynamics of different optoATAF1 strains.** Solid lines represent growth rate. Similar dynamics were observed for all strains except optoATAF1-2XS FB. 0X stands for optoATAF1, 1XS stands for optoATAF1-mSc, 2XS FB stands for optoATAF1-mSc-mSc with a feedback loop, 2XC stands for optoATAF1-mSc-mCerulean, 2XN stands for optoATAF1-mSc-mNeon and lastly, 2XC FB stands for optoATAF1-mSc-mCerulean with a feedback loop.

I found that the optoATAF1 2X (CN) behaviour varied depending on the target protein combination. However, I note that both optoATAF1 2XC (mCerulean & mScarletI) and optoATAF1 2XN (mNeonGreen & mScarletI) resulted in higher relative expression than optoATAF1 2XS FB (mScarlet-I & mScarlet-I + feedback). It is very curious that two copies of mScarlet-I do not lead to double the production (in fact, there is no increase whatsoever) but driving mScarlet-I and a **different protein** leads to double the production. Seemingly, just changing the target protein allows one to overcome whatever mechanism prevents an increase in mScarlet-I production. Another interesting story unfolds if we observe the differences between optoATAF1 2XC and optoATAF1 2XC FB. The two start out being indistinguishable from each other but over time, the

expression levels in optoATAF1 2XC FB drop lower than optoATAF1 2XC and never recover. It would appear that the FB loop seems to decrease production in the long term. A similar inference can be gleaned by comparing optoATAF1 2XS FB and optoATAF1 1XS albeit the effect appears to be less prominent. These results while being hard to interpret indicate two concrete things,

- 1) optoATAF1 without the feedback is capable of driving two distinct targets at full strength when induced with continuous light.
- 2) Expression from two different targets is higher than expression from two copies of the same target.

Growth rate dynamics were also fairly interesting. Remarkably, a two fold increase in expression in optoATAF1 2X (CN) over optoATAF1 1XS (20 fold over pTDH3 vs 10 fold) did not come with a proportional decrease in the growth rate. In fact, the growth rate for both strains dropped almost as much as for optoATAF1 driving no target and the steady state growth rates were indistinguishable. This suggests that for the same level of growth defect it is possible to plateau at different production levels.

Knowing that the EL222 driven expression shows linear behaviour with respect to duty cycle fraction and is nearly as strong as pTDH3 driven expression at full induction, I cultured optoATAF1 2XN in continuous cultures and induced them with 50% duty cycle for a period of 30 minutes to mimic the expression from pTDH3-LOXP (half the levels of pTDH3). I expected the fluorescence levels to drop. Surprisingly, I saw an increase in the relative expression for 50% duty cycle when compared to 100% light (**Figure 4.14a**). Even more surprisingly, this increase did not come at the cost of the growth rate (**Figure 4.14b**). Growth rate was slightly higher for 50% than for 100% light even though at steady state both growth rates were indistinguishable from each other. However, at the steady state there was a twofold difference between the relative expression.

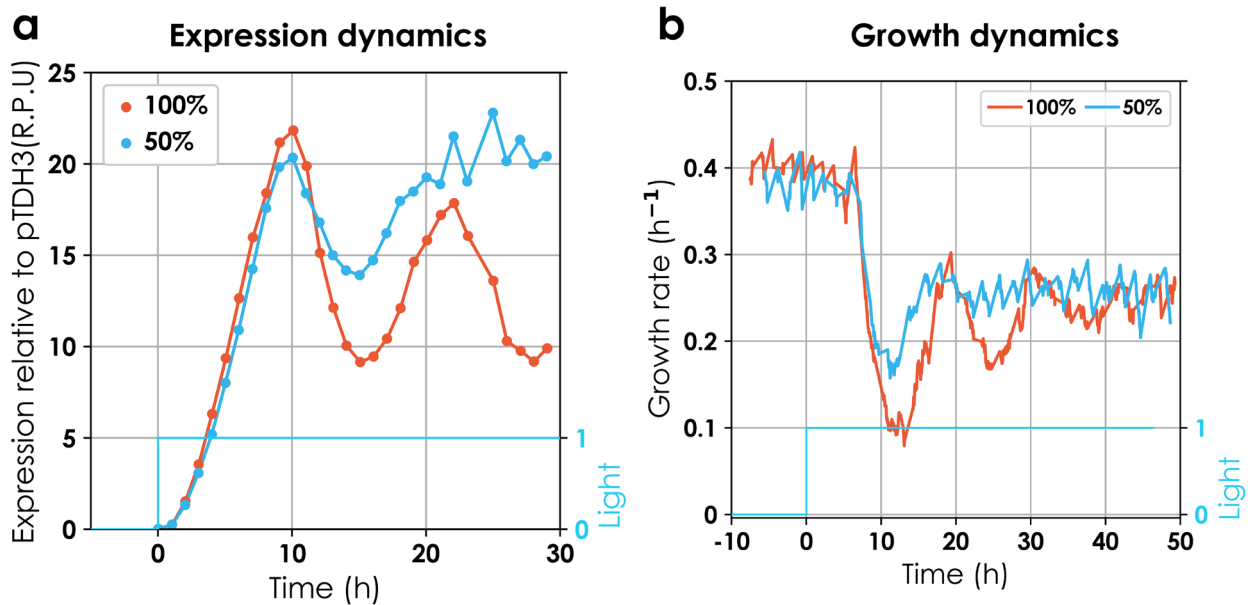


Figure 4.14. a. Dynamics of heterologous expression for optoATAF1 2XN under continuous light and 50% duty cycle. Cells were induced with blue light at $t=0$ as indicated by the cyan line. Data is represented as in Figure 4.13a. **b. Growth rate dynamics of optoATAF1 2XN for different light inputs.** Solid lines represent growth rate.

Notably at steady state, for 50% light induction, on average, cells were producing 20 times the protein obtained by pTDH3 driven expression. I note that at the same time the growth rate was stable around $0.26 h^{-1} \pm 0.03$ (Figure 4.14).

Despite no clear indication as to whether the feedback loop could rescue the growth arrest and production, I proceeded to clone the feedback loop in GAuDi01 in a leap of faith. Concretely, the positive feedback loop on ATAF1 TF was added working under the assumption that it would lead to sufficiently high expression of FAR1M to obtain effective growth arrest upon differentiation. The complete construction of this strain is depicted in Figure 4.15.

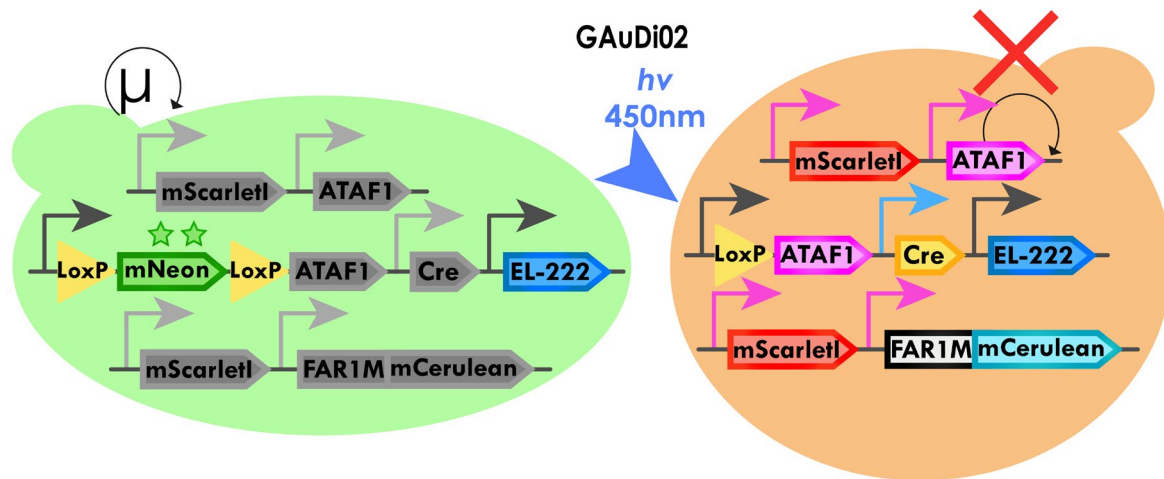


Figure 4.15. Circuit for GAuDi02. I added a positive feedback loop with ATAF1 TF driving its own expression to GAuDi01 to ameliorate the expression levels of FAR1M and the poi.

I characterized the GAuDi02 strain under the microscope and in continuous cultures. Under the microscope, cells were fluorescent in the YFP channel (mNeonGreen) and cellular growth could be observed prior to differentiation, albeit, buds were washed away leading to no change in cell number. After 1h of induction, cells started expressing mScarlet-I, indicating that they had differentiated. Concomitant with the increase in mScarlet-I fluorescence, cellular growth arrest could be observed as cells became large and unbudded. Following the next 14h, no cell divisions were observed (**Figure 4.16**). Sporadic cell death could be observed in some fields of view. I remark that the plate used in this experiment was not quite right. I say this because of its inability to retain more cells during cell loading as well as washout of the new buds.

Differentiation and growth arrest under the microscope

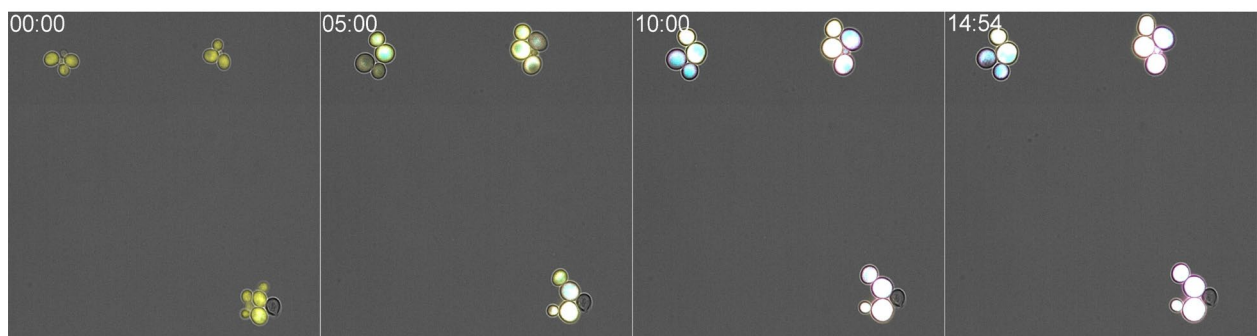


Figure 4.16 Differentiation behaviour and growth arrest of GAuDi02 cells under the microscope in a microfluidic plate. Induction started at $t=0$ with 800ms pulses. Cyan represents mCerulean fluorescence acquired in the CFP channel, yellow represents mNeonGreen fluorescence acquired in the YFP channel, and magenta represents mScarlet-I fluorescence acquired in the Rhodamine channel (**Table 6.3**). The

images shows a merge of brightfield, CFP, YFP, and RHOD channels. Images were cropped to increase resolution while retaining all the cells in the field of view.

To observe the growth arrest for longer durations, I cultured GAuDi02 in the turbidostat. Consistent with the results under the microscope, in presence of continuous light, the growth rate in liquid cultures decreased dramatically ($<0.04 \text{ h}^{-1}$) (**Figure 4.17**). I was not surprised to see that the addition of the feedback loop restored the growth arrest.

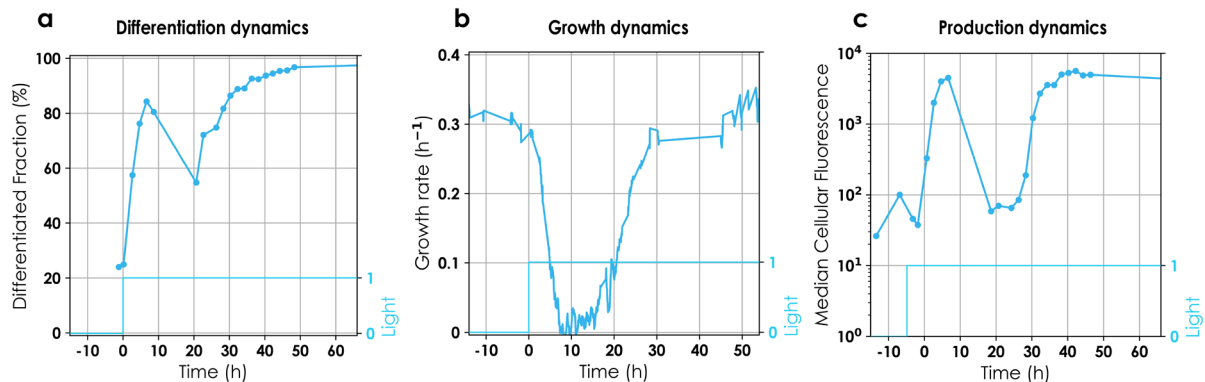


Figure 4.17. a. Differentiation behaviour for GAuDi02 under continuous light. Cells were induced with blue light at $t=0$ as indicated by the cyan line. Blue line represents differentiated fraction over time and circles signify times of sampling. A significant part of the population for GAuDi02 was seemingly differentiated prior to light induction. I note that it was not possible to get samples for cytometry between 9h and 21h due to lack of dilution (arising out of a complete absence of growth). **b. Growth rate under continuous light.** Blue line represents growth rate. **c. Production dynamics.** Blue line represents median cellular fluorescence from a single experiment. Circles signify times of sampling. I note that it was not possible to get samples for cytometry between 9h and 21h due to lack of dilution (arising out of a complete absence of growth).

The differentiation dynamics were not markedly different from Gaudi01. Notably, the differentiation rate seemed comparable with the similar number of cells differentiating after 8h of light induction, however, due to the sustained growth arrest, it was difficult to observe the population composition for Gaudi02. To be more descriptive, the growth arrest caused the OD of the reactor to come to a standstill. In the absence of an increase in the OD, the media could not be replenished (without decreasing the OD) and therefore, after ~ 6 samplings, the level of culture in the reactor dropped below the level required for sampling. I note that in all experiments, a population of cells appeared to be differentiated prior to light exposure. I believe that this population is not actually differentiated but, rather, is plagued by leakiness of the ATAF1 positive feedback loop.

Along with the OD measurements after light induction, I also analysed the population composition and discovered a variable, but consistently present, population of cells that contained neither green nor red. Presumably, these cells are dead cells.

In the turbidostat, however, I observed that the cultures under continuous light escaped the growth arrest 15-20h post differentiation (**Figure 4.17b**). I suspected that the escapers were mutants and extracted the genomes of 5 escapers and sent them for targeted genome sequencing. Results suggested that loss of the integrative cassette was the most common mutation (3/5) that led to escape. I argued that if this were the case, mutants that lose the entire integrative cassette, including the auxotrophic marker, would be selectively disadvantaged. Therefore, I cultured cells under continuous stimulation but this time in selective media. We observed that the cells could be arrested for 50-100% longer (**Figure 4.18a**), suggesting that other mutations could also lead to escape.

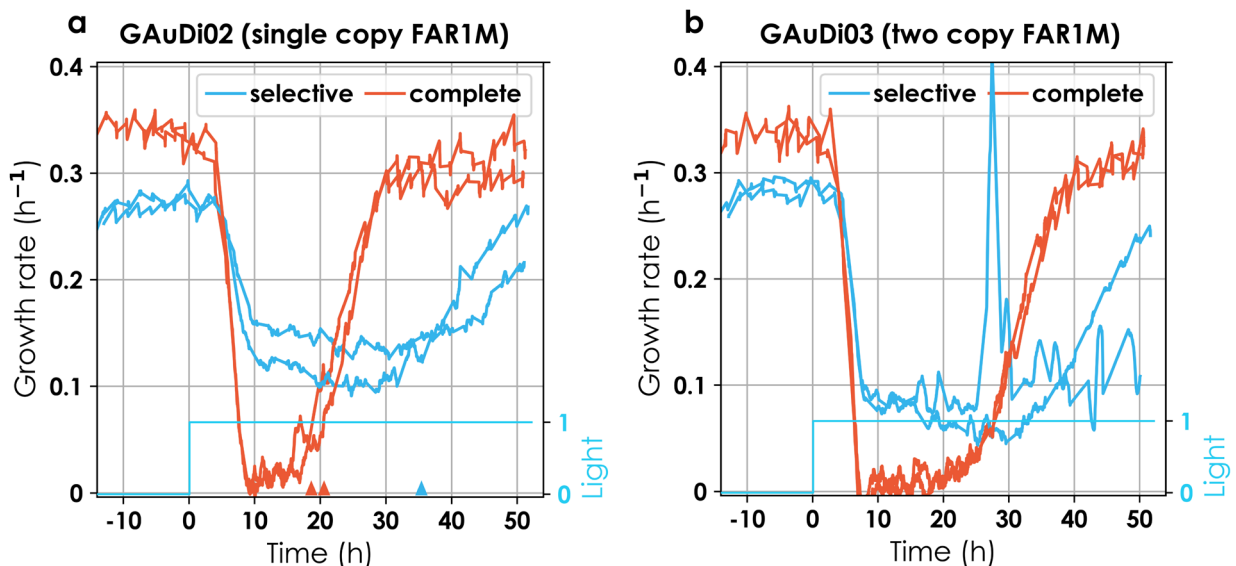


Figure 4.18. a. Growth rate of GAuDi02 under continuous light in non-selective (red) and selective (blue) media in duplicates. The cyan line signifies light induction. Manually placed triangles indicate the time when the culture as a whole escaped the growth arrest. **b. Growth rate of GAuDi03.** Data is represent as in **a**. The spike in one replicate in selective media is probably an artefact of the growth rate estimation due to noisy OD measurements.

Similarly, I reasoned that adding another copy of FAR1M would prolong the arrest because then the mutations needed to happen at two loci. Therefore, I constructed another strain that consisted of two copies of FAR1M being driven by ATAF1 in two different loci as well as the feedback loop. Compatible with the evidence generated so far, I observed that the growth arrest lasted longer in both complete and selective media (**Figure 4.18b**).

Curiously, the growth rate of differentiated cells was higher in selective media (**Figure 4.18**).

Despite, the inability of the system to maintain the growth arrest for >30h, I provide evidence that the escape is linked to genetic instability. There is some evidence in yeast that describes augmentation of DNA damage and chromosomal rearrangements following abnormal nuclear morphologies (anaphase bridges and micronuclei) (Germann et al. 2014). Furthermore, blue light at the given intensity of 40, does not decrease the growth rate but could lead to an augmented rate of DNA damage which coupled to the stress generated due to growth arrest would result in a higher chance of escaping mutations.

Nevertheless, to investigate the suitability of this system for population control purposes, I induced continuously growing cultures to varied pulses of light. I observed that a part of the population differentiated and that the non-differentiated fraction was able to replenish quickly (**Figure 4.19**). I also detected a smaller population of cells that was autofluorescent and rose roughly 6h after induction (**Figure 4.21**). Consistent with the results in the microscope, it is suggestive that only a part of the differentiated population dies.

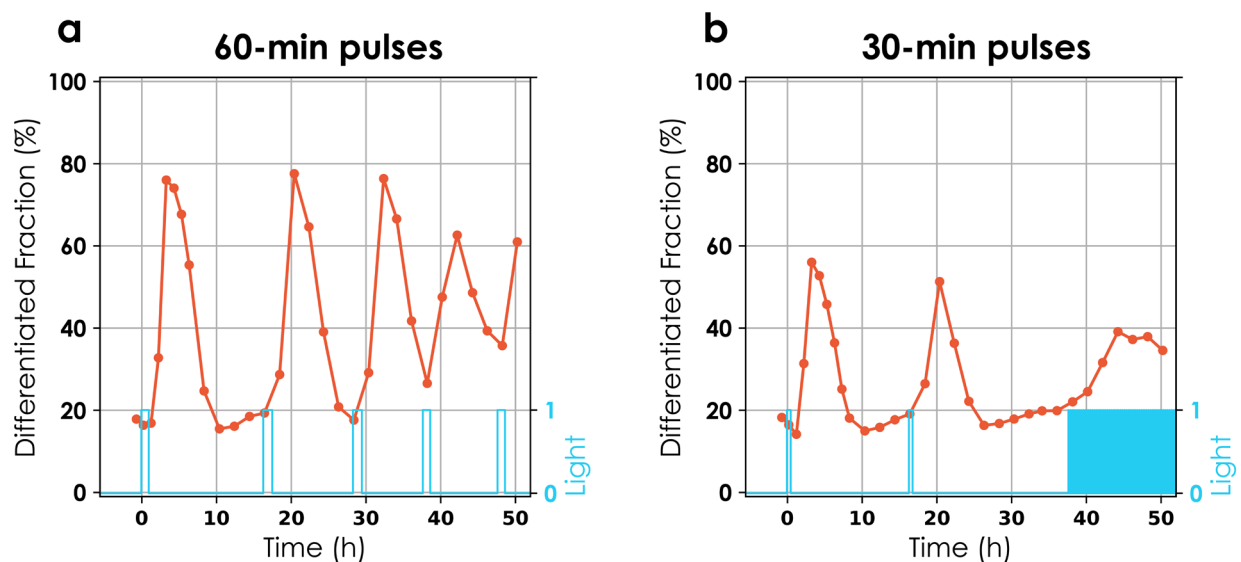


Figure 4.19. Differentiation dynamics of GAudi02 (red dotted line) for periodic light pulses (cyan line). Circles are individual data points. I observed that light pulses led to increases in differentiated fractions that subsided rapidly after removal of light (a, b). When cells were exposed to periodic light pulses, the differentiation fraction seemed to reach a steady state (b last 4 timepoints).

Both observations (quick replenishing of the non-differentiated fraction and cell death in only a subpopulation of growth arrested cells) led me to conclude that the system could be employed for dynamic control of microbial consortium in a self-contained configuration.

Development of a predictive model for GAuDi02

However, before deploying the model predictive framework (MPC) framework, I had to construct an appropriate model and fit the data to it. I adapted the ODE model to account for growth arrest and extended it to include cell death and escape. I assumed that differentiated cells, in addition to growing significantly slower, die and escape from the growth arrest at definite rates. The culture growth rate is then given by a weighted average of the growth rates of individual species and is equal to the dilution rate of the reactor at constant cell density.

The rate of escape and the death rate could not be directly fixed from the data and were treated as unknown model parameters. Furthermore, due to the leakiness of the ATAF1 feedback loop, I observed a persistent fraction of cells that appeared to be initially differentiated given their levels of mScarlet1 fluorescence. Since empirically it was not possible to separate the differentiated cells from those with an overactive feedback loop, we introduced a basal differentiation rate present in the absence of light (and its presence). The growth rates of differentiated, non-differentiated and escapers were fixed from the data.

$$\begin{aligned}\dot{g}(t) &= \mu_g g(t) - (r_{diff}^{\circ} + \mathbf{U}(t)r_{diff})g(t) - \lambda g(t) \\ \dot{p}(t) &= \mu_{diff} p(t) + (r_{diff}^{\circ} + \mathbf{U}(t)r_{diff})g(t) - r_{esc}p(t) - r_{dead}p(t) - \lambda p(t) \\ \dot{e}(t) &= \mu_{esc} e(t) + r_{esc}p(t) - \lambda e(t) \\ \dot{d}(t) &= r_{dead}p(t) - \lambda d(t) \\ n &= g(t) + p(t) + e(t) + d(t) \\ \lambda &= \frac{g(t)}{n} \mu_g + \frac{p(t)}{n} \mu_{diff} + \frac{e(t)}{n} \mu_{esc}\end{aligned}$$

g , p , e , and d are specific cell density (in O.D. units) of non-differentiated, differentiated, escaper, and dead cells, respectively. μ_g , μ_{diff} , and μ_{esc} are growth rates of non-differentiated, differentiated, and escaper cells, respectively. r_{diff}° is basal differentiation rate and r_{diff} is the differentiation rate in presence of light. $\mathbf{U}(t)$ is the light signal as a function of time and can take values 0 or 1. r_{esc} is the rate of escape from growth arrest in differentiated cells. r_{dead} is the rate of death in growth arrested differentiated cells. n is the total cell density (in O.D. units) of the culture. λ is the culture dilution rate and, at constant cell density, equals the weighted sum of growth rates of individual species.

The model was fitted to dynamical data from an experiment with a non-trivial light signal (**Figure 4.20a**). Concretely, the differentiation fraction and the dead fraction were computed from time-

series flow cytometry data to fit the model. However, there were significant delays between differentiation events and the time by which corresponding cells could be classified as differentiated based on their fluorescence. Similarly, the growth arrested cells died after being arrested for some time. To account for these delays, model predictions were shifted in time (2h for differentiated and 6h for dead cells, and 5h for the growth rate). Bounds were imposed on the parameter values (10^{-10} -10) to preserve physiological relevance. The parameter search was conducted locally but starting from different initial parameter values spanning four orders of magnitude using an approach similar to Branch Coleman and Li (1999). In all cases, the search converged to the same parameter values (**Figure 4.20b**).

Parameter	μ_g	μ_{diff}	μ_{esc}	r_{diff}^o	r_{diff}	r_{esc}	r_{dead}
Value	3.9e-01	4.0e-02	3.9e-01	6.98e-02	1.16e+00	9.43e-08	7.12e-02

Parameters in green in **Table 4.1** were fixed directly from the data while those in yellow were estimated.

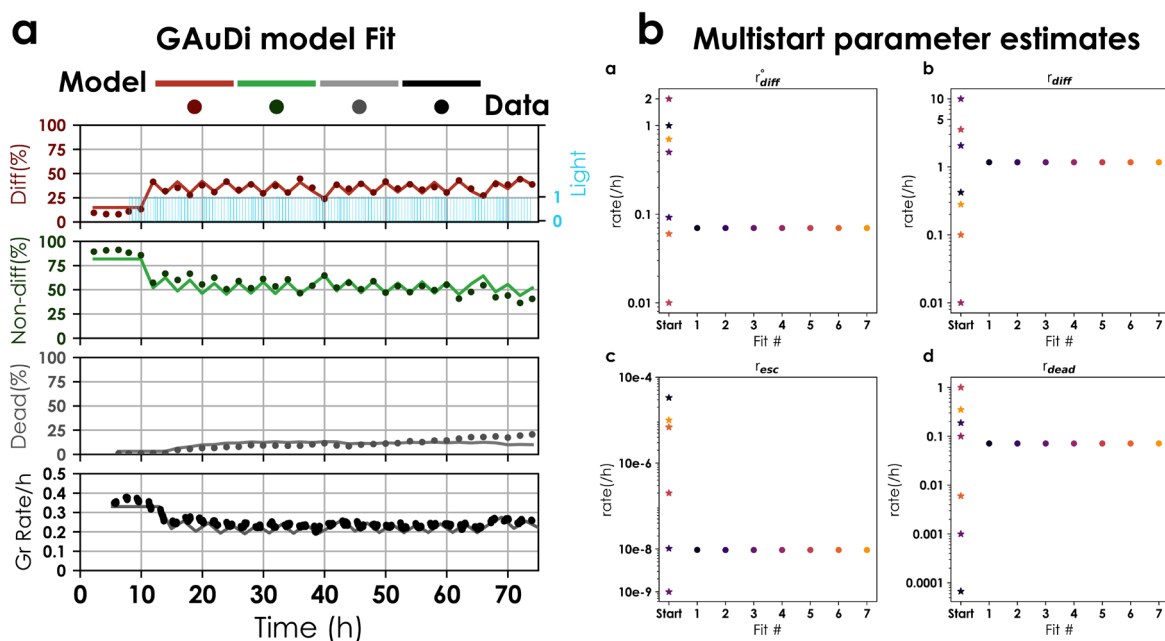


Figure 4.20. a. GAuDi ODE model fit. An ODE model for GAuDi02 population dynamics was developed phenomenologically. The model was fitted to dynamical data with repeated light stimuli of different durations. Solid circles represent data and solid lines signify model predictions. We note that before $t=0$, a persistent non-zero differentiation fraction was present possibly due to leakiness in ATAF1 expression in a subset of the population. **b. Multistart fit for estimating the parameters of the GAuDi02 model** (r_{diff}^o , r_{diff} , r_{esc} , r_{dead}). Regardless of the initial guess estimates (colored stars) of parameters over four orders of

magnitude the search converged to the same parameter value, 0.0698 h^{-1} , 1.16 h^{-1} , $9.43\text{e-}08 \text{ h}^{-1}$, 0.0712 h^{-1} , respectively.

These parameters led to good agreement between model outputs and observed data (**Figure 4.20a**). We further validated the model by predicting the outcome of four other experiments with dynamic non-trivial light inputs. Model predictions were in good agreement with observed data (**Figure 4.21**) validating that the model is predictive.

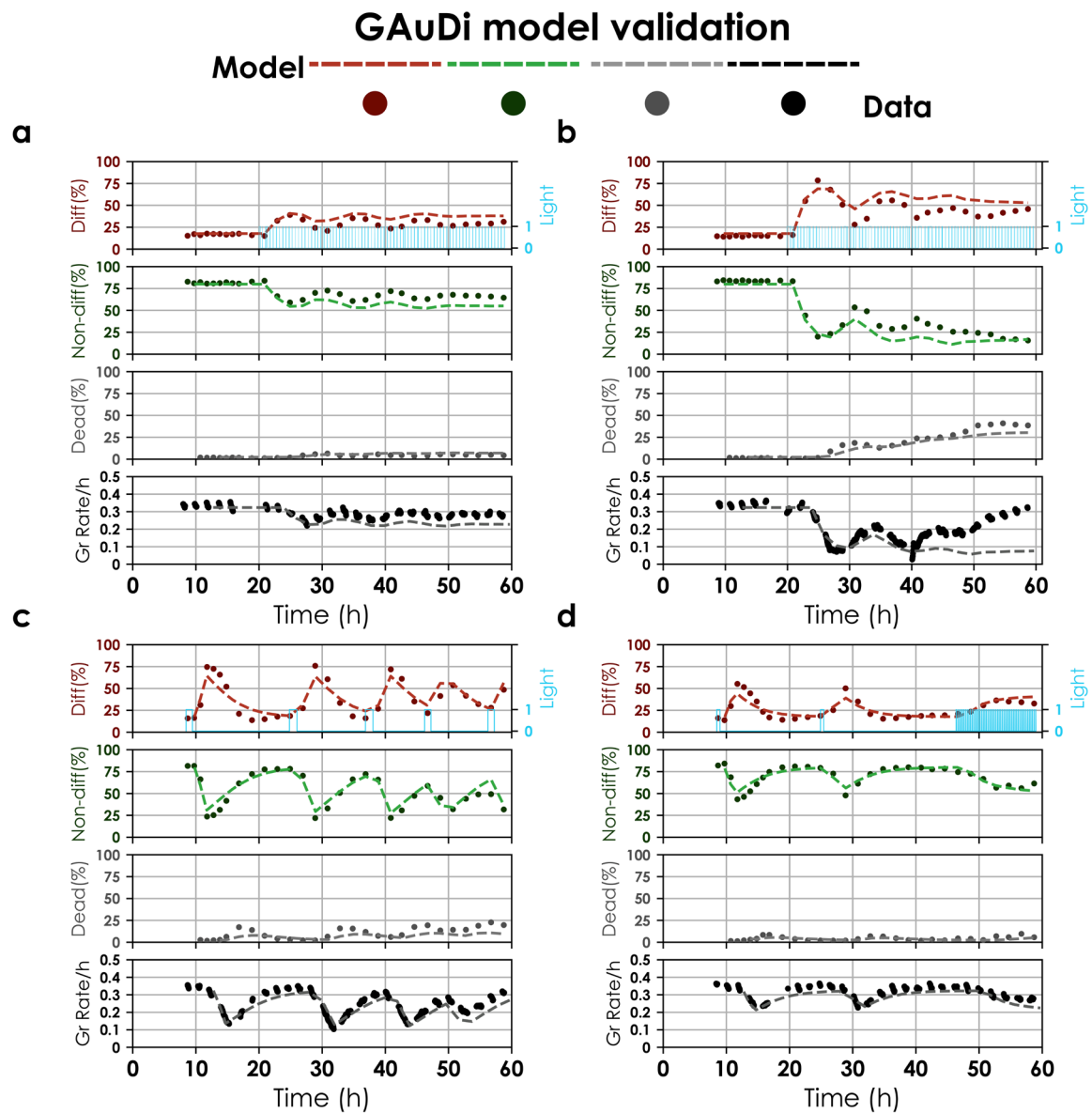


Figure 4.21. GAuDi model predictions and data for 4 light profiles. The fitted model was validated by using it to predict dynamic responses to light signals that were not used for the fit (blue lines). Circles represent differentiation fractions from a single experiment and dotted lines represent model predictions. The red line represents differentiated fraction, green is the non-differentiated fraction, grey is the dead fraction and black is the culture growth rate. Model predictions were shifted in time to account for observation delays.

The model, however, slightly underestimated the differentiated fraction after single light pulses (**Figure 4.21 c & d**). This is perhaps because in the model, growth rate of differentiated cells instantaneously drops to 0.04 h^{-1} whereas in the data we observe that there is a delay between the differentiation event and the growth arrest. Simply put, it means that while in the model cells stop growing immediately after light signal, data indicates that they continue to grow for some time leading to a slightly higher differentiated fraction than the model predictions.

I was curious to know the result of fitting other experiments to data. Therefore, I additionally used the dynamics of these validation experiments to re-identify model parameters to be able to compare the results to the originally identified values (**Table 4.1**). For r_{diff}° , r_{diff} , and r_{dead} , the estimates showed little variation, varying less than 2 fold between the extremes (**Figure 4.22**). Results for r_{esc} showed more variability. Notably, I uncovered that for three of the experiments, the search terminated at the lower bound. Such low estimates suggest that the escape might not play a significant role for the light profiles used for these experiments.

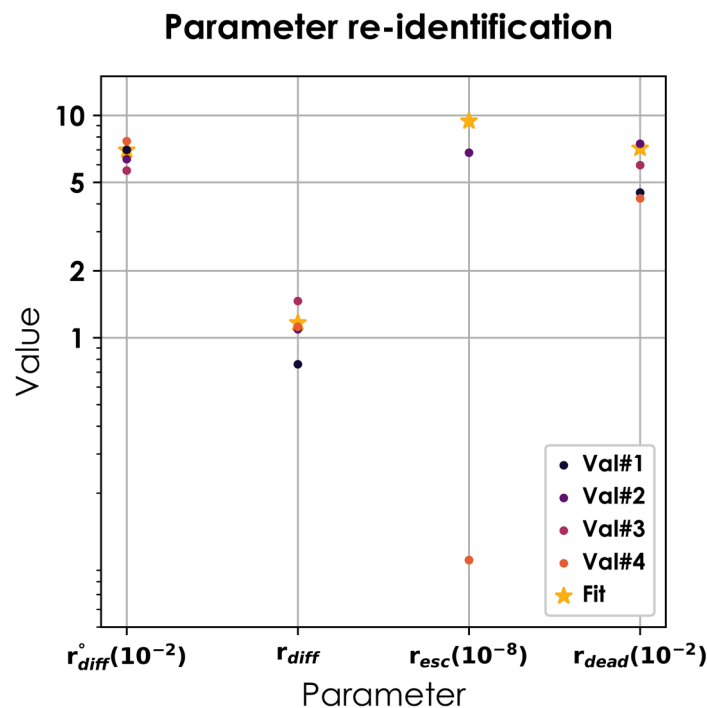


Figure 4.22. Parameter re-identification. Coloured solid circles represent results from different experiments. For all parameters, yellow stars represent the values listed in Table 4.1.

It also suggests that the escape rate might not be constant. Indeed, if escapers are assumed to exclusively arise out of mutations that occur in the differentiated cells, then, the amount of time differentiated cells spend in the reactor and the light signal (proxy for selection pressure imposed) would logically contribute to the escape rate. In line with this hypothesis, continuous light resulted in significant deviations from predicted behaviour (**Figure 4.23a**). Notably, the

escape could not be predicted well, further providing evidence that the escape from growth arrest depends heavily on the selection pressure which is not modeled (continuous light exerts a larger selection pressure on the population than pulses). To test whether fitting the escape rate for the continuous light data would result in a better predictions overall, I fitted the escape rate to match the data and indeed both the differentiated and dead fraction predictions were ameliorated (**Figure 4.23b**).

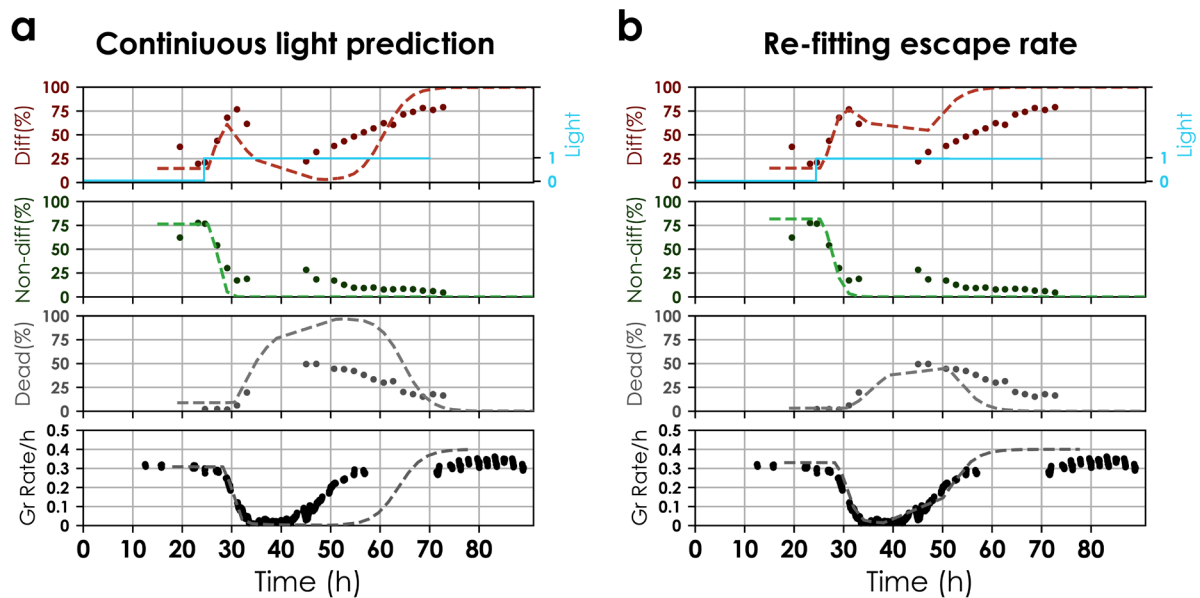


Figure 4.23. a. Deviation from model predictions of the duration of growth arrest of GAuDi02 under continuous light. **b. Escape rate re-fitted to continuous light data.** All other parameters were left identical to Table 4.1 while the escape rate was changed to $8.93e-06$.

Despite the inability to predict the escape rate well, I conclude that the ODE model developed in Chapter 3 could be extended to capture the observed data and that the system remains predictable despite a change in physiological context as long as the selection pressure does not become very high.

Single strain single vessel control of population composition

Next, I used the GAuDi ODE model in conjunction with the MPC framework to demonstrate single reactor control of a microbial consortium originating from a single strain. Concretely, exponentially growing GAuDi cells in the turbidostat were exposed to optimized light signals such that the population composition is maintained at user-defined set points. Flow cytometry measurements were taken at regular intervals to track fraction of differentiated cells (**Figure 4.24a**).

No adaptation to the model was necessary. We could only sample the culture every two hours owing to volume limitations and consequently the observation delay was increased to 2.5 hours (**Methods 6.8**).

I found that the population compositions could be maintained stably for extended periods (up to 50h) (**Figure 4.24b**).

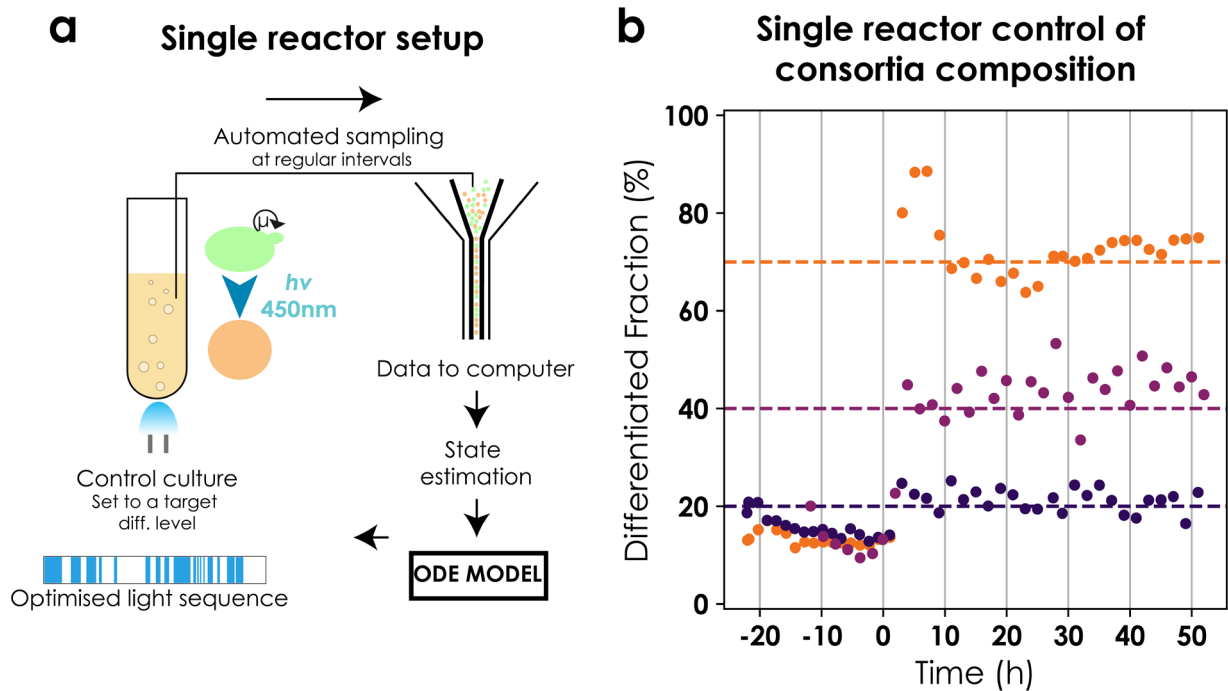


Figure 4.24. a. Setup for single reactor control. *GAuDi02* cells were continuously cultured in exponential phase. The culture was set to a target level of differentiation and continuously monitored via automated flow cytometry measurements that were analysed on-line. The system state was estimated from analysed data and sent to the model predictive control (MPC) module. The MPC module provided an optimized light sequence to maintain the culture at the desired set point. Due to presence of dead cells, data was filtered and only live cells were used for subsequent analysis. **b. Single reactor, single strain control of a microbial consortium.** Cultures were targeted to 25%, 40%, and 70% differentiation. Control started at $t=0$. Circles signify differentiated fractions. Each colour corresponds to a different control experiment and the dashed line reflects the target set point.

I note that due to genetic stability limitations, it was possible to maintain composition control for longer periods only if the target set point was below 50% differentiation. To the best of my knowledge, this is the first report of dynamic control of population composition in a two species artificial microbial consortium arising from a single strain in a single reactor.

Assessment of production in continuous mode

I compared performance of GAuDi strains that express a protein of interest (mScarlet-I) upon differentiation to investigate if our strategy could overcome the shortcomings of heterologous expression in continuous cultures. The strains were cultured in the turbidostat and their differentiated fractions were maintained at fixed set points. Concretely, GAuDi01 was cultured at 100% differentiation, GauDi02 was cultured at 50% and 70% differentiation and lastly, GAuDi04, a strain that displays a moderate growth defect upon differentiation, was maintained at 70%. GauDi02 & GAuDi04 contained two mScarlet-I targets and a positive feedback loop of ATAF1 whereas GAuDi01 only possessed a single target without the feedback loop. In parallel, I cultured strains that express the same protein of interest driven either constitutively by a strong promoter (pTDH3) or by a light inducible ATAF1 transcription factor (optoATAF1) (**Methods, Figure 6.19**). Cells were cultured at a constant Optical Density (OD) and regular cytometry measurements were made to calculate mean cellular fluorescence levels. Even though the cultures were maintained at average constant OD, cell size fluctuations resulted in a change of cell density. Therefore, cell counts from cytometry data were used in addition to OD data to determine the cell density (**Methods, cell density**). Mean cellular fluorescence was calculated from cytometry data. Titers were calculated by multiplying mean cellular fluorescence with cell density. Productivity was defined as protein/ml/h and quantified by multiplying mean cellular fluorescence by the number of cells (cell density from cytometry data times the volume expelled per reactor volume). **Figure 4.25** shows that the differentiation based GAuDi strains outperformed constitutive production in continuous mode but could not beat the induction system (optoATAF1). An extended figure detailing all the observables is included in the Methods section on quantifying protein production.

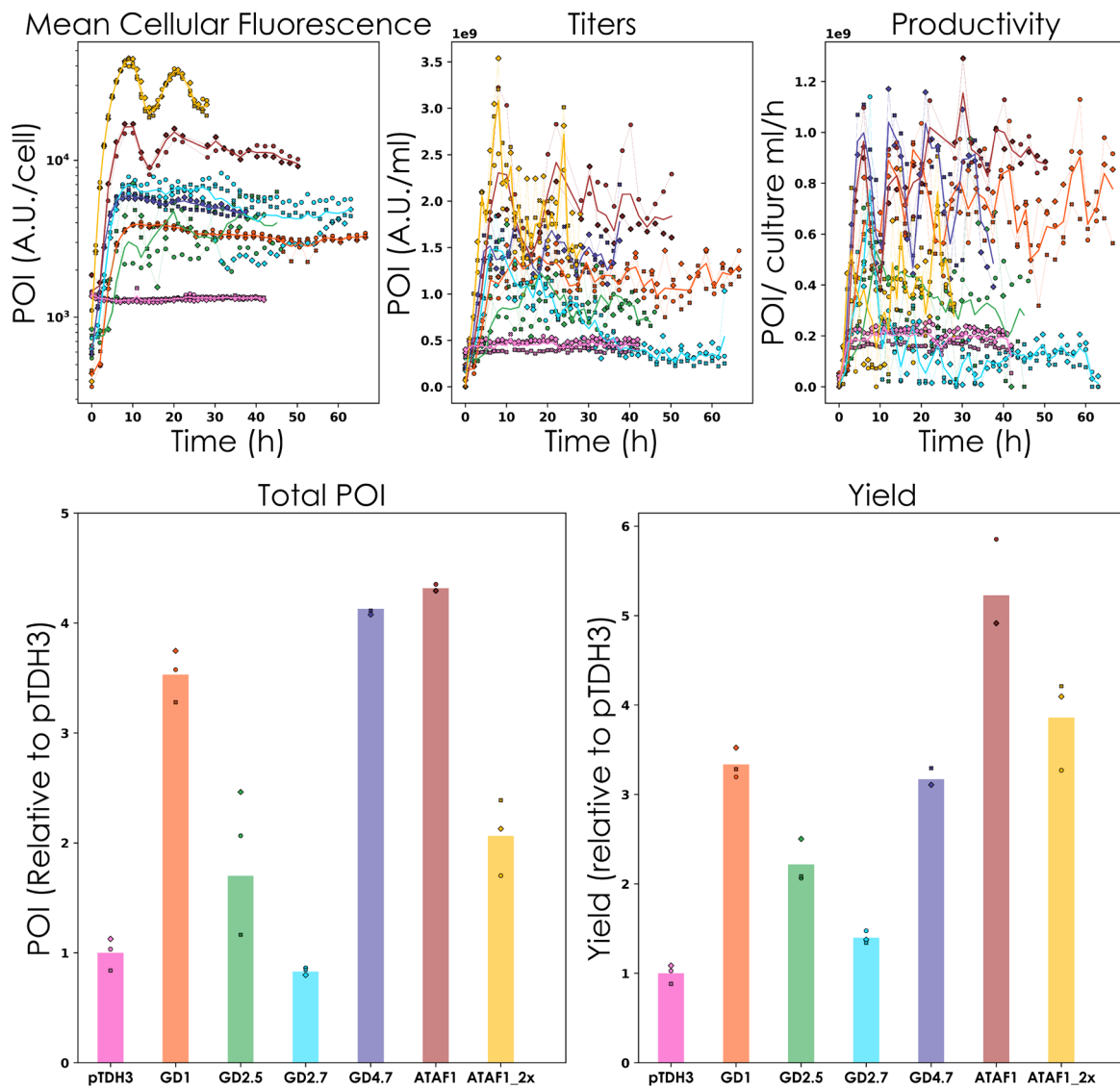


Figure 4.25. Results of heterologous expression in continuous operation. Cells from the reference constitutive (pTDH3-mScarlet-I), reference induction (optoATAF1-mSc & optoATAF1 2X) and GAuDi (GAuDi01, GAuDi02 & GAuDi04) strains were cultured continuously at an average OD of 0.5. GAuDi strains were maintained at different target differentiation levels (Methods). In case of induction and GAuDi strains, light was delivered at $t=0$. Cytometry measurements were taken regularly and used to infer the levels of expression at single cell resolution. **a. Mean cellular fluorescence (POI/cell)** was calculated from cytometry data by taking the mean of viable cells. For optoATAF1 2X, this was done relative to pTDH3 expression and mean cellular fluorescence levels were scaled to be comparable to other single fluorophore strains. The thick lines signify mean values estimated from three independent experiments represented by the thin lines (optoATAF1, GAuDi02@0.7 and GAuDi04@0.7 only had duplicates). The solid circles, diamonds and crosses correspond to individual timepoints from different experiments. Colors map to strains (or conditions) according to Figure 4.25. **b. Protein titers (POI/ml)** were calculated by multiplying cell density with mean cellular fluorescence. Data is represented as in **a**. **c. Productivity (POI/h)** was calculated by multiplying titers with change in volume. This was further normalized by culture volume (see

Methods). Data is represented as in **a**. **d. Overall protein produced (POI)** was estimated by computing the area under curve of productivity vs time. These values were normalized to pTDH3 expression in the time horizon of individual experiments to obtain the plotted data. Bars represent mean values from three (two) different experiments. Circles, diamonds, and crosses signify individual experiments. **e. Yield** was computed by dividing the overall protein produced by total volume consumed. These values were normalized with yield of pTDH3 expression to obtain the plotted data. Data is represented as in **d**.

GAuDi strains had up to four fold higher mean cellular fluorescence than pTDH3. Curiously, the expression levels of GAuDi01 (with one target) were roughly half of GauDi02 and GAuDi04 (two targets). The highest relative mean cellular fluorescence (16-20 fold of pTDH3) was observed in optoATA1 2X (**Figure 4.25a**). Cellular fluorescence was a strong indicator of protein titer. However, due to an increase in cell size, the cell density of optoATAF1 strains dropped significantly (**Methods, 6.9**) resulting in relatively lower titers (**Figure 4.25b**). Productivity trends were quite different from the titers. Notably, productivity also depends on the amount of volume expelled, which, in turn, depends on the growth rate of the culture. Therefore, a growth defect leads to a decrease in productivity. Due to this phenomenon, the productivity of optoATAF1 strains, along with Gaudi02 at 70%, was comparably low, as these cultures showed a marked decrease in growth rate. optoATAF1_2X in particular, despite being 16-20 fold high in mean cellular fluorescence, had marginally higher productivity than pTDH3. At the same time, optoATAF1-mSc with two fold lower mean cellular fluorescence, when compared to optoATAF1_2X had a productivity 2-3 times higher. I could observe a higher total protein output for GAuDi strains, as quantified by area under the productivity vs time curve (**Figure 4.25c**), when compared to pTDH3 (**Figure 4.25d**). Interestingly, I noticed that GAuDi02 maintained at 70% performed worse than pTDH3 whereas GAuDi02 maintained at 50% was twice as good as pTDH3 based expression. This is perhaps due to presence of large number of dead cells that reduce the protein titers (**Methods, 6.9**) and a decrease in growth rate that decreases productivity (Methods, volume change). I also note that GAuDi04 was as good as optoATAF1-mSc (~4 fold higher than pTDH3). Lastly, I computed yield of the process by dividing the total protein produced by the total volume consumed. Yield was positively affected by the growth arrest due to decrease in media consumption. I found that optoATAF1-mSc yielded up to 5.2 fold more protein than pTDH3 based constitutive expression per ml of media used (or per g of glucose) (**Figure 4.25e**). This means that for the same amount of resources optoATAF1-mSc could express 5 times more protein compared to pTDH3. GAuDi strains also yielded more protein per ml of media consumed when compared to pTDH3. GAuDi04 was the best among them (~3 fold). Acknowledging the dense and non-intuitive organization of data in the previous figure, I restate some interesting aspects of the results.

- 1) Both the expression platforms (optoATAF1 & GAuDi) can outperform constitutive expression.
- 2) Protein production appears to be non-monotonic with respect to levels of differentiation for GAuDi02.
- 3) Genetic instability seems to be at the root of this non-monotonic behaviour.
- 4) A strain that exhibits a milder growth defect upon differentiation (GAuDi04) might be better suited for bioproduction purposes.
- 5) Mean cellular fluorescence is not a good indicator of protein production in continuous cultures owing to growth-production tradeoffs (compare optoATAF1 and optoATAF1 2X).
- 6) Due to converse dependence of productivity and yield on the growth rate, there is a tradeoff between the two.

Results from this section when coupled with results reported in **Figure 4.14** suggest that GAuDi and optoATAF1 are able to outcompete constitutive expression and possess non-monotonic tendencies. Dedicated mathematical models of the production process might lead to new insights and could be used to maximize protein output. With this, I conclude that both the platforms developed in this work are worth exploring further in the context of continuous bioproduction.

Discussion

Growth production tradeoffs arise as a consequence of metabolic burden of heterologous expression and pose a major challenge for continuous bioproduction due to emergence of nonproducing mutants that decrease yield over time. In order to ascertain whether distributing growth and production in different subpopulations would lead to increased heterologous expression (due to reduced competition with growth program in producers) or more stable expression (as growers do not suffer from metabolic burden), I coupled the differentiation system to a growth arrest and production module via ATAF1 driven expression of FAR1M and POI. After significant tinkering, I was able to create a strain, GAuDi02, which arrested upon light induction and started producing POI. Cells arrested as large unbudded cells and continued to increase in size until they died. To my dismay, I found that the circuit was not stable and differentiated cells lost the growth arrest phenotype when stimulated with continuous light. I established with targeted genome sequencing and complementary experiments that this escape was associated with the loss of transcriptional cassette expressing FAR1M. Nevertheless, I developed a dedicated ODE model to be able to predict the population dynamics and fitted it to dynamic experimental data. The ODE model was sufficient to predict the population dynamics for a variety of light inputs. We deployed the ODE model in an MPC framework to control the

differentiation level of the population. We were able to maintain stable levels of differentiated fraction in continuous cultures of GAuDi02 and other associated strains.

Concurrently, I developed an optogenetically inducible platform (optoATAF1) for heterologous gene expression as an inducible reference for GAuDi02. I found that a single target driven by optoATAF1 resulted in up to 10 fold higher mean cellular fluorescence over pTDH3 reference. Curiously adding another copy of the same gene did not result in any noticeable increase in the fluorescence levels, however, adding a different target gene increased the expression to >20 times higher than pTDH3 reference. I also found that the growth rate decreased as a consequence of ATAF1 expression ($32\% \pm 2\%$). Oddly, this decrease was not dependent on heterologous expression levels meaning that cells could support 20 fold pTDH3 expression without it affecting the growth rate (any further). Lastly, I found non-monotonic behaviour in two target versions of optoATAF1 with respect to induction.

I characterized the production of the POI for both the fully induced optoATAF1 and GAuDi strains, maintained at different levels of differentiation in continuous cultures. I compared the performance to a pTDH3 reference with respect to standard metrics (titers, productivity, and yield). I found that the GAuDi02 produced up to two fold more protein than the pTDH3 reference but was not genetically stable. GAuDi04, which showed a milder growth defect, was more amenable to long term control and outperformed the pTDH3 reference all around. However, the differentiation system was found to be, at best, as good as the induction system. The induction system with a single copy, maintained 4 fold higher productivity, was genetically stable and, overall, yielded 5 times as much protein per ml of media used. Interestingly, despite having twice the amount of cellular fluorescence, the two copy optoATAF1 performed worse than the single copy. Lastly, I found that yield and productivity cannot be maximized at the same time demonstrating it experimentally for the first time in continuous production in yeast (Zhuang et al. 2013; Machado and Herrgård 2015). Overall, I conclude that our approach with a strong growth arrest is unfeasible for heterologous expression in continuous cultures but introducing a milder growth defect seems to be more promising.

On the model

The simple ODE based model was sufficient to predict population responses to a wide array of light signals and worked well in an MPC framework. However, model predictions had to be shifted in time to account for delays that were not explicitly modeled. In order to eliminate the need of such inelegant adjustments, the population level model can be coupled to a mechanistic model of molecular events taking place inside the cell so as to incorporate the delays. More

concretely, light can be made to produce Cre instead of resulting in direct differentiation. Differentiation can be modeled as a function of Cre concentration. After the recombination event, cells can be placed in a transition state, instead of instantaneously becoming differentiated. In this state, cells start producing ATAF1 that in turn leads to the production of FAR1M and mScarlet-I. mScarlet-I production could be modelled as a two-step process to account for maturation time. Cells would then be considered differentiated if mScarlet-I levels cross a threshold. Growth rate is not explicitly modeled in this work but it is assumed that cells instantaneously change their growth rate upon differentiation. It would be more accurate to map the growth rate as a function of FAR1M expression levels.

The opto-ATAF1 induction platform has the potential to become a strong reference and a similar approach can be used to construct a model and couple it to a model of batch and continuous operation. This model could then be investigated to find the optimal induction pattern to maximize productivity or yield, in a given timeframe.

On escape

The fact that the differentiation system coupled to growth arrest is not stable seems almost inevitable in retrospect. Evolutionary constraints are known to limit the long-term stability of synthetic circuits (Castle, Grierson, and Gorochowski 2021) particularly circuits that are engineered to implement growth arrest / self-killing at the population level (Din et al. 2016; Balagaddé et al. 2008; Scott et al. 2017). I admit that our lack of foresight could be attributed to the flawed assumption that escapers would not have a competitive advantage due to the presence of growers. More precisely, we reasoned that since the growers and escapers would grow at the same rate, escapers would not be able to spread in the population. That would have been the case if there were not regular light signals constantly differentiating the growers into producers. Due to this differentiation, the effective growth rate of growers becomes negative leaving the escapers to romp free. There remains a question as to where do these escapers arise from. Since the evidence from sequencing suggests that the predominant mutations occur away from the recombination cassette and live cell microscopy time lapses indicate presence of aberrant nuclear phenotypes that have been linked to genome instability (Germann et al. 2014), I am inclined to say that the escapers arise from producers but the mutations may have well been present in the grower population. I note that a milder growth defect might be more stable.

On cell death

I noticed small but consistent cell death in GAuDi02 cells after several hours of becoming differentiated. This was enhanced for GAuDi03 that expressed FAR1M from two copies. While at face value this may

seem undesirable but, taking inspiration from the work of Hasty and colleagues (Scott et al. 2017; Din et al. 2016; Miano, Liao, and Hasty 2020), I believe this can be useful. Notably, in the context of bioproduction this could be used to release the protein of interest in the medium and make downstream processing simpler. It would mean avoiding the hassles of cell lysis and protein purification. More concretely, in the spirit of continuous operation, the outflow of the bioreactor could be coupled to a continuous annular chromatography column to yield purified protein (Hilbrig and Freitag 2003). On the other hand, these could serve as probiotic drug delivery agents (Din et al. 2016). The idea would be to engineer a strain that carries the light based differentiation system that makes a cell massively produce a “drug” upon differentiation and explode after a pre-determined amount of time. Then, in theory, someone could add some powder in a glass of water, shine it with some blue light and drink it and cells would start producing the drug and explode after, say 6h, thus delivering the drug. Even though the system has been characterized in budding yeast, *S. boulardii* might be the better option for therapeutic applications (Chen et al. 2020).

On absence of production increase upon growth arrest

It is understandable even apparent why the growth arrest could not be sustained, it is not so clear why a growth arrest did not result in an increase in protein amounts because 90% of the energy is directed towards ribosome biogenesis in exponentially growing yeast (Warner 1999). Mating factor pheromone induced arrest has been shown to be a metabolically active state (Williams et al. 2016). In this study, while the growth arrest was not complete (the cells continued to grow at a reduced rate), the authors reported 3 to 5 times higher amounts of secondary metabolites in addition to a downregulation of genes involved in ribosome biogenesis. They observed that protein production was decreased and ascribed this decrease to mating pheromone phenotype switch and concluded the study by suggesting a growth arrest decoupled from the mating arrest phenotype will be more suitable for bioproduction purposes. Another study from Avalos group using an approach similar to ours had shown that it was possible to obtain higher titers of metabolites by alternating the same population of cells between a growth phase and a production phase (Zhao et al. 2018). Note that, in their case, the arrest was triggered due to depletion of NAD⁺ co-factors instead of genetically contrived arrest but nonetheless, these studies indicate that, at least, metabolite production might be increased during growth arrest.

The answer perhaps lies in what exactly is meant by growth arrest. I make this distinction because cell growth can happen with or without cell division. In case of mating factor arrest, cells stop dividing but the growth continues. Consequently, cells become large and unbudded. A recent article showed how an increase in cell size results in dilution of native machinery that leads to a decrease in the capacity of cell to produce proteins and pushes the cell towards senescence

(Neurohr et al. 2019). The authors used a thermosensitive *cdc28* allele to stop the growth in G1 phase and made quantitative measurements of mRNA and protein content in the cell along with cell volume measurements. The authors found that mRNA and protein content do not scale with cell volume and manifest in lower levels of induction with the same concentration of inducer when compared to cycling cells. The authors also reported that large sized yeast cells displayed repression of ribosome biogenesis and general transcription factors, consistent with the finding of Williams et al. (2016), and established that this was, at least in part, due to the activation of environmental stress response (ESR)(Gasch 2003). Lastly, they showed that cell size increase led to reduction in cellular lifespan and that the lifespan decreased drastically with the duration of growth arrest. Curiously, they start observing cell death after 6h of arrest. This also happens to be the delay in the ODE model of GAuDi02 for the emergence of dead cells. Restoration to permissive temperatures after 2h of arrest, however, allows cells to live longer. So, until we find a way to truly stop cellular growth, perhaps, an approach similar to Zhao et al. (2018) might be more appropriate to decouple growth from production.

On non-monotonic production dynamics of optoATAF1

The finding that protein levels increase non-monotonically for optoATAF1 with light induction is particularly fascinating because this higher protein levels do not occur due to a decrease in the growth rate, on the contrary, the growth rate is higher with less light. These results taken together suggest that there is large scope for optimization and this presents as a non-trivial problem especially in light of the finding that yield and productivity cannot be maximized simultaneously. Similar results have been reported when proteins are secreted (Wittrup et al. 1994) (which may also result in a growth defect).

On the apparent optoATAF1 2X transcriptional buffering

optoATAF1 driven induction by far was the strongest expression realized in this work. It is not obvious to me why the expression with two target copies of the same gene does not result in higher protein levels while two different genes do. Now, I do realize that I only performed these experiments once and before speculating one would be required to repeat the experiments perhaps with more controls, for instance generate other strains that express a different gene from two copies driven by ATAF1. It would be prudent to perform RT-qPCR to measure the transcript levels. However, if these results can be reproduced, they might reveal interesting insights into cell homeostasis during heterologous expression. Notably, the fact that two target mScarlet-I have no observable increase in protein content might first indicate a gene dosage effect, perhaps due to post-transcriptional bottlenecks, for example, scarcity of free ribosomes

or presence of rare codons. The fact that both mNeonGreen and mScarlet-I could be observed at maximal levels precludes ribosome scarcity. Both mScarlet-I and mNeonGreen coding sequences were codon optimized (CAI>0.65) and did not possess any rare codons. A large scale study reported no increase in the abundance of proteins when multiple copies of the gene were present even if the mRNA levels correlated with copy number (Torres et al. 2007). I note that this study only remotely relates to our discussion since they looked at endogenous genes in aneuploid strains, which have distinct functions inside the cell and therefore can be regulated post-translationally. This is not the case here. Here, we have artificial and supposedly orthogonal transcription factors expressing an exogenous gene. These results indicate a gene specific transcript buffering mechanism might be at play. This would explain why I do not see an increase in mScarlet-I fluorescence when I put two targets of mScarlet-I but notice both mNeonGreen and mScarlet-I expressed to full strength (essentially doubling the heterologous expression). A mechanism for transcriptional buffering has been described in *S. cerevisiae* that operates in S-phase to buffer the additional transcripts from the second copy of the genome and is characterized by regulation of global mRNA synthesis and degradation rates (reviewed in Timmers and Tora, 2018) but gene specific buffering has not been reported outside the context of physiologically relevant processes.

Kafri et al. (2016) reported that pTDH3 driven expression accounts for 2.3% of the total proteome and that each copy decreases the growth rate by roughly 1.3-2.5% depending on the environmental context. I am somewhat skeptical of these results given that the authors do not provide any quantification for how they estimate copy number from FACS data nor do they show fluorescence distributions for different copies. Nevertheless, assuming these results to be true, I note that the two target optoATAF1 constitutes 46% of the proteome and at this level of expression we should expect a drop of 24-33% in the growth rate. Interestingly, this is what I observed. However, I observed the same decrease for optoATAF1 1X and for optoATAF1 driving no fluorescent protein. This begs the question, what would happen if I add another target gene to create optoATAF1 3X? I note that a recent study looking at promoter activity at the single cell level in bacteria found that the individual promoter activity decreased with an increase in copy number (B. Shao et al. 2021)(more details in the discussion of Chapter 5, On plasmid copy numbers) but Kafri et al. (2016) did not report any such decrease, again, highlighting the superior capacity of *S. cerevisiae* for heterologous expression.

Of course, a more elegant approach would be to exploit the heterogeneity in copy number (Lillacci, Benenson, and Khammash 2018) afforded by 2 μ plasmid to characterize burden and transcriptional buffering over a wide range of copy numbers. More specifically, the optoATAF1

could be integrated in the genome while the target is expressed from a 2 μ plasmid. This could be coupled with destabilized mRNAs (Muhlrad and Parker 1999; Kafri et al. 2016) and/or protein degradation tags (Lee et al. 2015) to reveal bottlenecks in heterologous protein production.

References

- Aditya, Chetan, François Bertaux, Gregory Batt, and Jakob Ruess. 2021. "A Light Tunable Differentiation System for the Creation and Control of Consortia in Yeast." *BioRxiv*.
- Aoki, Stephanie K, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. 2019. "A Universal Biomolecular Integral Feedback Controller for Robust Perfect Adaptation." *Nature* 570 (7762): 533–37.
- Balagaddé, Frederick K, Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, Stephen R Quake, and Lingchong You. 2008. "A Synthetic Escherichia Coli Predator-Prey Ecosystem." *Molecular Systems Biology* 4 (1): 187.
- Branch, Mary Ann, Thomas F Coleman, and Yuying Li. 1999. "A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems." *SIAM Journal on Scientific Computing* 21 (1): 1–23.
- Carrasco-López, César, Sergio A García-Echauri, Therese Kichuk, and José L Avalos. 2020. "Optogenetics and Biosensors Set the Stage for Metabolic Cybergenetics." *Current Opinion in Biotechnology* 65: 296–309.
- Castle, Simeon D, Claire S Grierson, and Thomas E Gorochofski. 2021. "Towards an Engineering Theory of Evolution." *Nature Communications* 12 (1): 1–12.
- Ceroni, Francesca, Alice Boo, Simone Furini, Thomas E Gorochofski, Olivier Borkowski, Yaseen N Ladak, Ali R Awan, Charlie Gilbert, Guy-Bart Stan, and Tom Ellis. 2018. "Burden-Driven Feedback Control of Gene Expression." *Nature Methods* 15 (5): 387–93.
- Chang, Fred, and Ira Herskowitz. 1990. "Identification of a Gene Necessary for Cell Cycle Arrest by a Negative Growth Factor of Yeast: FAR1 Is an Inhibitor of a G1 Cyclin, CLN2." *Cell* 63 (5): 999–1011.
- Chen, Kevin, Yixuan Zhu, Yongrong Zhang, Therwa Hamza, Hua Yu, Ashley Saint Fleur, James Galen, Zhiyong Yang, and Hanping Feng. 2020. "A Probiotic Yeast-Based Immunotherapy against Clostridioides Difficile Infection." *Science Translational Medicine* 12 (567).
- Chen, Yun, Siavash Partow, Gionata Scalinati, Verena Siewers, and Jens Nielsen. 2012. "Enhancing the Copy Number of Episomal Plasmids in Saccharomyces Cerevisiae for Improved Protein Production." *FEMS Yeast Research* 12 (5): 598–607.
- Choi, Jae Woong, Sung Sun Yim, Min Jeong Kim, and Ki Jun Jeong. 2015. "Enhanced Production of Recombinant Proteins with Corynebacterium Glutamicum by Deletion of Insertion Sequences (IS Elements)." *Microbial Cell Factories* 14 (1): 207.
- Csörgő, Bálint, Tamás Fehér, Edit Tímár, Frederick R Blattner, and György Pósfai. 2012. "Low-Mutation-Rate, Reduced-Genome Escherichia Coli: An Improved Host for Faithful Maintenance of Engineered Genetic Constructs." *Microbial Cell Factories* 11 (1): 11.
- Din, M Omar, Tal Danino, Arthur Prindle, Matt Skalak, Jangir Selimkhanov, Kaitlin Allen, Ellixis Julio, et al. 2016. "Synchronized Cycles of Bacterial Lysis for in Vivo Delivery." *Nature* 536 (7614): 81–85.
- Gasch, Audrey P. 2003. "The Environmental Stress Response: A Common Yeast Response to Diverse Environmental Stresses." In *Yeast Stress Responses*, edited by Stefan Hohmann and Willem H Mager, 11–70. Berlin, Heidelberg: Springer Berlin Heidelberg.

- Germann, Susanne M, Vera Schramke, Rune Troelsgaard Pedersen, Irene Gallina, Nadine Eckert-Boulet, Vibe H Oestergaard, and Michael Lisby. 2014. "TopBP1/Dpb11 Binds DNA Anaphase Bridges to Prevent Genome Instability." *The Journal of Cell Biology* 204 (1): 45–59.
- Geymonat, Marco, Adonis Spanos, and Steven G Sedgwick. 2007. "A *Saccharomyces Cerevisiae* Autoselection System for Optimised Recombinant Protein Expression." *Gene* 399 (2): 120–28.
- Gnügge, Robert, Thomas Liphardt, and Fabian Rudolf. 2016. "A Shuttle Vector Series for Precise Genetic Engineering of *Saccharomyces Cerevisiae*." *Yeast* 33 (3): 83–98.
- Gupta, Apoorv, Irene M Brockman Reizman, Christopher R Reisch, and Kristala L J Prather. 2017. "Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent Quorum-Sensing Circuit." *Nature Biotechnology* 35 (3): 273–79.
- Henchoz, S, Y Chi, B Catarin, I Herskowitz, R J Deshaies, and M Peter. 1997. "Phosphorylation- and Ubiquitin-Dependent Degradation of the Cyclin-Dependent Kinase Inhibitor Far1p in Budding Yeast." *Genes & Dev.*
- Hilbrig, Frank, and Ruth Freitag. 2003. "Continuous Annular Chromatography." *Journal of Chromatography B* 790 (1): 1–15.
- Jung, So Young, Hae Yong Yoo, Young Ho Kim, Jiyoung Kim, and Hyune Mo Rho. 1995. "The Glucose-Dependent Transactivation Activity of ABF1 on the Expression of the TDH3 Gene in Yeast." *Current Genetics* 27 (4): 312–17.
- Kafri, Moshe, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. 2016. "The Cost of Protein Production." *Cell Reports* 14 (1): 22–31.
- Kroll, Jens, Stefan Klintner, Cornelia Schneider, Isabella Voß, and Alexander Steinbüchel. 2010. "Plasmid Addiction Systems: Perspectives and Applications in Biotechnology." *Microbial Biotechnology* 3 (6): 634–57.
- Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. "A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly." *ACS Synthetic Biology* 4 (9): 975–86.
- Lillacci, Gabriele, Yaakov Benenson, and Mustafa Khammash. 2018. "Synthetic Control Systems for High Performance Gene Expression in Mammalian Cells." *Nucleic Acids Research* 46 (18): 9855–63.
- Liu, Jingyi, Xia Wu, Mingdong Yao, Wenhai Xiao, and Jian Zha. 2020. "Chassis Engineering for Microbial Production of Chemicals: From Natural Microbes to Synthetic Organisms." *Current Opinion in Biotechnology* 66: 105–12.
- Loison, Gérard, Martine Nguyen-Juilleret, Sami Alouani, and Magda Marquet. 1986. "Plasmid--Transformed Ura3 Fur1 Double-Mutants of *S. Cerevisiae*: An Autoselection System Applicable to the Production of Foreign Proteins." *BioTechnology* 4 (5): 433–37.
- Lv, Yongkun, Yang Gu, Jingliang Xu, Jingwen Zhou, and Peng Xu. 2020. "Coupling Metabolic Addiction with Negative Autoregulation to Improve Strain Stability and Pathway Yield." *Metabolic Engineering* 61: 79–88.
- Machado, Daniel, and Markus J Herrgård. 2015. "Co-Evolution of Strain Design Methods Based on Flux Balance and Elementary Mode Analysis." *Metabolic Engineering Communications* 2: 85–92.
- Mckinney, J D, and F R Cross. 1995. "Far1 and the g(1) Phase Specificity of Cell-Cycle Arrest by Mating Factor in *Saccharomyces-Cerevisiae* ." *Molecular and Cellular Biology* 15 (N5): 2509–16.
- Miano, Arianna, Michael J Liao, and Jeff Hasty. 2020. "Inducible Cell-to-Cell Signaling for Tunable Dynamics in Microbial Communities." *Nature Communications* 11 (1): 1–8.
- Muhlrad, Denise, and Roy Parker. 1999. "Aberrant MRNAs with Extended 3' UTRs Are Substrates for Rapid Degradation by mRNA Surveillance." *Rna* 5 (10): 1299–1307.
- Naseri, Gita, Salma Balazadeh, Fabian Machens, Iman Kamranfar, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2017. "Plant-Derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces Cerevisiae*." *ACS Synthetic Biology* 6 (9): 1742–56.

- Neurohr, Gabriel E, Rachel L Terry, Jette Lengfeld, Megan Bonney, Gregory P Brittingham, Fabien Moretto, Teemu P Miettinen, et al. 2019. "Excessive Cell Growth Causes Cytoplasm Dilution and Contributes to Senescence." *Cell* 176 (5): 1083–97.
- Richardson, Sarah M, Leslie A Mitchell, Giovanni Stracquadanio, Kun Yang, Jessica S Dymond, James E DiCarlo, Dongwon Lee, et al. 2017. "Design of a Synthetic Yeast Genome." *Science* 355 (6329): 1040–44.
- Rugbjerg, Peter, Nils Myling-Petersen, Andreas Porse, Kira Sarup-Lytzen, and Morten O A Sommer. 2018. "Diverse Genetic Error Modes Constrain Large-Scale Bio-Based Production." *Nature Communications* 9 (1): 1–14.
- Scott, Spencer R, M Omar Din, Philip Bittihn, Liyang Xiong, Lev S Tsimring, and Jeff Hasty. 2017. "A Stabilized Microbial Ecosystem of Self-Limiting Bacteria Using Synthetic Quorum-Regulated Lysis." *Nature Microbiology* 2 (8): 1–9.
- Shao, Bin, Jayan Rammohan, Daniel A Anderson, Nina Alperovich, David Ross, and Christopher A Voigt. 2021. "Single-Cell Measurement of Plasmid Copy Number and Promoter Activity." *Nature Communications* 12 (1): 1–9.
- Timmers, H Th Marc, and László Tora. 2018. "Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation." *Molecular Cell* 72 (1): 10–17.
- Torres, Eduardo M, Tanya Sokolsky, Cheryl M Tucker, Leon Y Chan, Monica Boselli, Maitreya J Dunham, and Angelika Amon. 2007. "Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast." *Science* 317 (5840): 916–24.
- Warner, Jonathan R. 1999. "The Economics of Ribosome Biosynthesis in Yeast." *Trends in Biochemical Sciences* 24 (11): 437–40.
- Williams, Thomas C., Bingyin Peng, Claudia E. Vickers, and Lars K. Nielsen. 2016. "The *Saccharomyces Cerevisiae* Pheromone-Response Is a Metabolically Active Stationary Phase for Bio-Production." *Metabolic Engineering Communications* 3: 142–52.
- Wittrup, K D, A S Robinson, R N Parekh, and K J Forrester. 1994. "Existence of an Optimum Expression Level for Secretion of Foreign Proteins in Yeast A." *Annals of the New York Academy of Sciences* 745 (1): 321–30.
- Zhao, Evan M., Yanfei Zhang, Justin Mehl, Helen Park, Makoto A. Lalwani, Jared E. Toettcher, and José L. Avalos. 2018. "Optogenetic Regulation of Engineered Cellular Metabolism for Microbial Chemical Production." *Nature* 555 (7698): 683–87.
- Zhuang, Kai, Laurence Yang, William R Cluett, and Radhakrishnan Mahadevan. 2013. "Dynamic Strain Scanning Optimization: An Efficient Strain Design Strategy for Balanced Yield, Titer, and Productivity. DySScO Strategy for Strain Design." *BMC Biotechnology* 13 (1): 1–15.

Chapter 5

Mechanistic interrogation of cell-fate decisions

“Disobedience, in the eyes of anyone who has read history, is man's original virtue.”
Oscar Wilde

Preface

I note that this work was performed in collaboration with Jakob Ruess, who developed the theory behind the model and obtained all the modelling results. Jakob's keen eye was also instrumental in data analysis.

Introduction

At the heart of rational circuit design in synthetic biology, lies the assumption that the functionality of complex circuits can be predicted from known properties of their components. Yet in practice, we routinely fail to make predictions of circuit dynamics that would agree with the data at the level expected in physics or engineering. A core reason for this is cell-to-cell variability inside genetically identical cell populations. Such population heterogeneity is often a consequence of the inherent stochasticity of biochemical processes inside cells (Elowitz et al. 2002). Cell-to-cell variability may lead to unexpected and undesirable circuit dynamics and has been identified as one of the major roadblocks for designing synthetic circuits with in-silico predictable functionality (Kwok 2010; Del Vecchio et al. 2018). On the other hand, it has been shown that identifying and carefully characterizing sources of variability at the single-cell scale allows one to design remarkably robust synthetic circuits (Potvin-Trottier et al. 2016) or to exploit stochasticity for creating features of cell populations, such as bimodal phenotype distributions, that would otherwise be difficult to engineer (Lugagne et al. 2017).

While the single-cell perspective has certainly helped to advance our understanding of cellular processes (Raser and O'shea 2005; Zechner et al. 2012; Neuert et al. 2013), what eventually matters for most applications in synthetic biology is how a particular circuit functions inside growing populations. At the population-scale, synthetic circuits may intentionally (e.g. circuits to control growth (Miliadis-Argeitis et al. 2016)) or unintentionally (e.g. burden caused by protein production (Weiße et al. 2015)) affect traits of cells that can be selected during population growth. If this is the case, we need to expect that variability that is generated at the single-cell

scale (e.g. stochastic production of the burdensome protein) will lead to consequences at the population-scale that cannot be predicted solely based on a characterization of the circuit inside single cells. However, despite the apparent prevalence of problems where single-cell and population processes are coupled, multi-scale models that capture both single-cell stochastic chemical kinetics as well as population dynamics have not been used to predict, or even just to explain, how population-scale functionality emerges from single-cell characteristics of synthetic circuits. Presumably, this is because the classically used modeling framework for single-cell processes, that is the chemical master equation (Gillespie 1992), is only directly applicable at the population-scale whenever the population phenotype distribution is equivalent to the distribution of the single cell process and not additionally shaped by population-level processes (Lunz et al. 2020).

Problem statement

The core feature of our system is that varying the duration of applied light pulses allows one to modulate the fraction of cells that differentiates even though the entire population is exposed to the same global light stimulation. The system therefore exploits heterogeneity in the response of cells to light to enable external control of population dynamics but, by doing so, creates a complex coupling between population dynamics and the single-cell process that leads to the production of recombinase. It can therefore serve as an ideal test-bed to study what types of models are needed to explain and predict complex system dynamics. Concretely, in this chapter, we explore the consequences of noise in transcription factor levels on circuit functionality and predictability.

Strategy

In order to understand the role of cell-to-cell heterogeneity in EL222 transcription factor expression on the functioning of the differentiation system, we thought that a direct readout of the EL222 TF levels was crucial. Therefore, we decided to tag the EL222 with a fluorescent protein. Since the N-terminal of the EL222 TF is composed of nuclear localization signal (NLS), C-terminal was decided as the site of fusion. Furthermore, we reasoned it will be helpful if, in addition to the transcription factor levels, the activity of the TF in presence of light could also be monitored. For the same, we introduced the EL222 reporter described in Chapter 3 (**Figure 3.1b**). As the EL222 reporter was mScarlet-I and the dual reporter cassette contained mCerulean and mNeonGreen, mVenus was chosen as the fluorophore to tag the EL222 TF. To change the noise structure, we chose to express the tagged EL222, Cre, and the reporter from different genetic contexts (integrated, centromeric plasmid (low copy), and 2-micron plasmid (high copy)).

Results

Integrated tetra reporter

To be able to study the role of cell-to-cell variability in the design and functionality of the differentiation circuit, we constructed a version of the system that carries reporters for EL222 TF levels and an EL222 activity reporter in addition to the original dual reporter (**Figure 3.1**). The simultaneous presence of four fluorescent reporters led to very good observability of the system dynamics. Concretely, we used the same recombination cassette and circuit design as in the dual reporter (**Figure 3.1a**). Furthermore, mScarlet-I is produced alongside Cre from a second copy of the EL222 promoter and reports on the levels of Cre. The EL222 TF itself is tagged with mVenus, at the C-terminal.

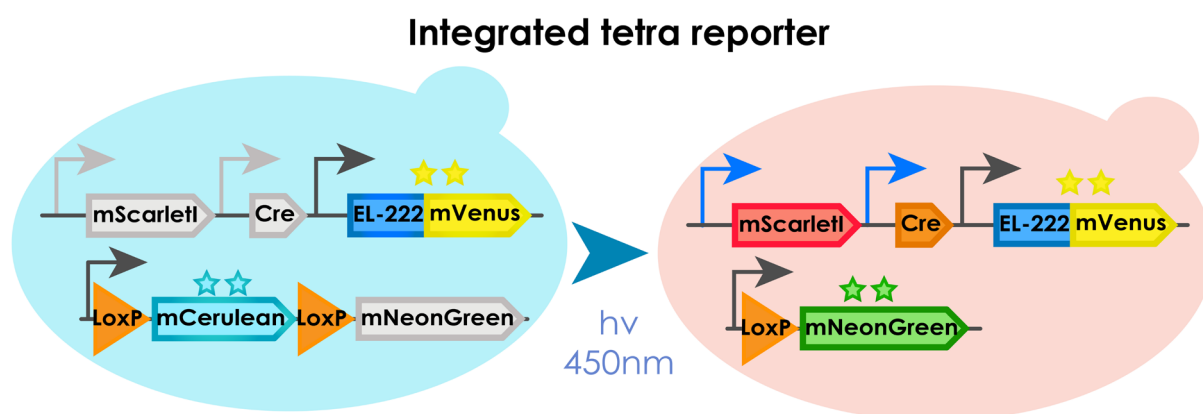


Figure 5.1. Integrated tetra reporter. EL222 tagged with mVenus at the C-terminal is produced in cells from a constitutive promoter (pTDH3, dark grey). Upon exposure of cells to blue light, EL222 switches the EL222-promoter from off (light grey) to on (blue) and cells start to produce mScarlet-I in addition to Cre-recombinase as a measure of EL222 activity. Cre excises a DNA-fragment placed between two target LoxP-sites, which is designed such that cells switch from producing mCerulean to mNeonGreen from a second constitutive promoter (pTDH3, dark grey).

To test system functionality, we performed experiments using the turbidostat platform. Single-cell measurements were automatically taken by flow cytometry with the help of a programmable pipetting robot (**Methods, 6.4**). Using deconvolution to extract amounts of the different reporter proteins in cells from measured spectral signatures (**Methods, 6.5**), we found that cells gradually switched from cyan to green in the presence of blue light (**Figure 5.2**).

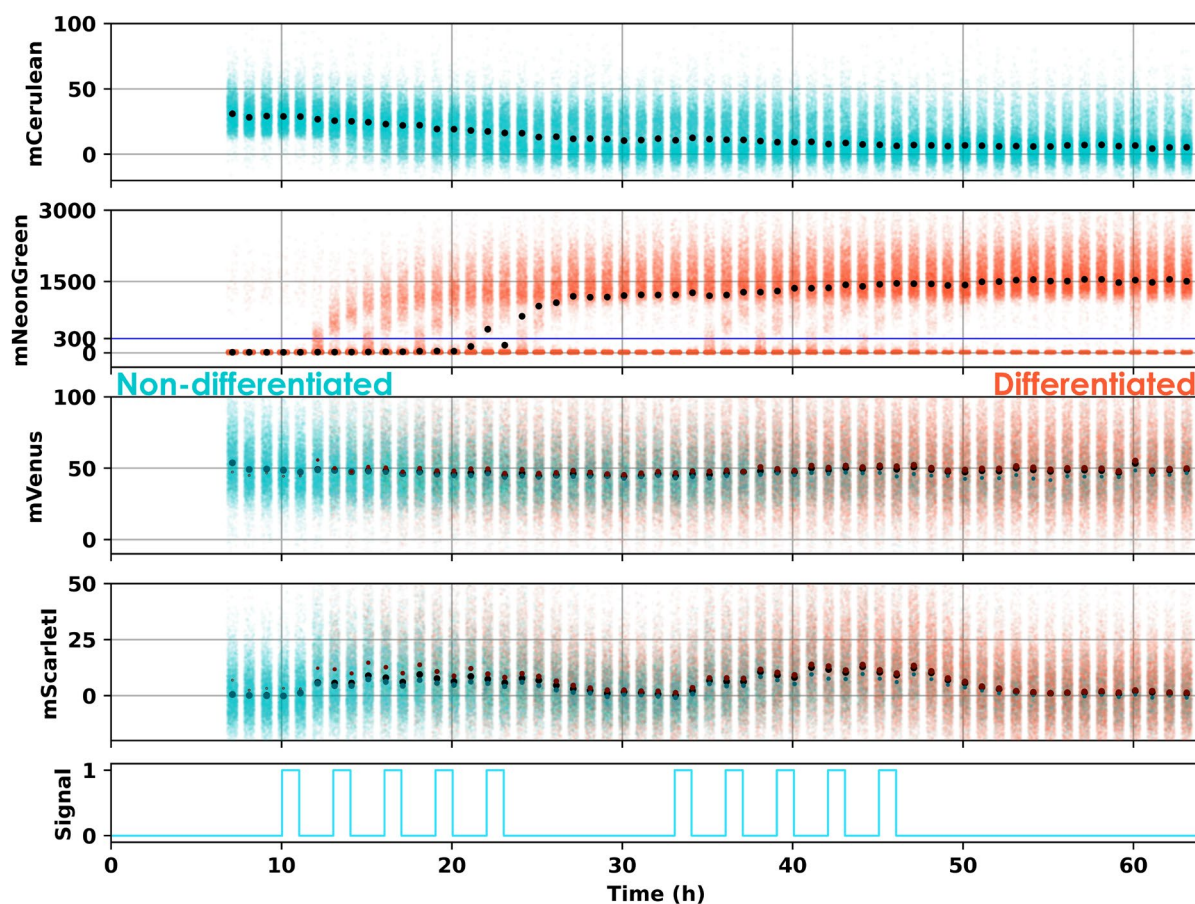


Figure 5.2. Distributions of the different fluorescent proteins in the tetra reporter. The small circles are individual cells. Each cloud of circles corresponds to a single timepoint. The large black circles signify the median of the population. **Cells were classified as differentiated if they exceeded a threshold of 300 A.U.** (blue line) in the **mNeonGreen** channel (see Methods, 6.6). For mVenus and mScarlet-I fluorescence distributions, the differentiated (red) and non-differentiated (blue) cells were plotted in different colors. The variably sized teal and maroon circles indicate the median of non-differentiated and differentiated population, respectively. Size of the circles reflects the subpopulation fraction at that timepoint. The cyan lines at the bottom represent the light signal. 5 1h pulses were delivered every 2h followed by 10h of rest and then another set of 5 1h pulses were delivered every 2h.

Analogous to the original system, exposing cell populations to short pulses of light led to bimodal mCerulean and mNeonGreen distributions (**Figure 5.3, mCerulean & mNeonGreen channels**), which demonstrates that only a fraction of the population recombines in response to light pulses. Varying the duration of light pulses and applying thresholds on mCerulean and mNeonGreen fluorescence to classify cells into recombined and non-recombined, we could quantify the differentiation dynamics of the system (**Methods, 6.6**). To test if the probability for a cell to recombine is correlated with single cell levels of EL222-TF and Cre, we analyzed mVenus and mScarlet-I fluorescence distributions shortly after applied light pulses.

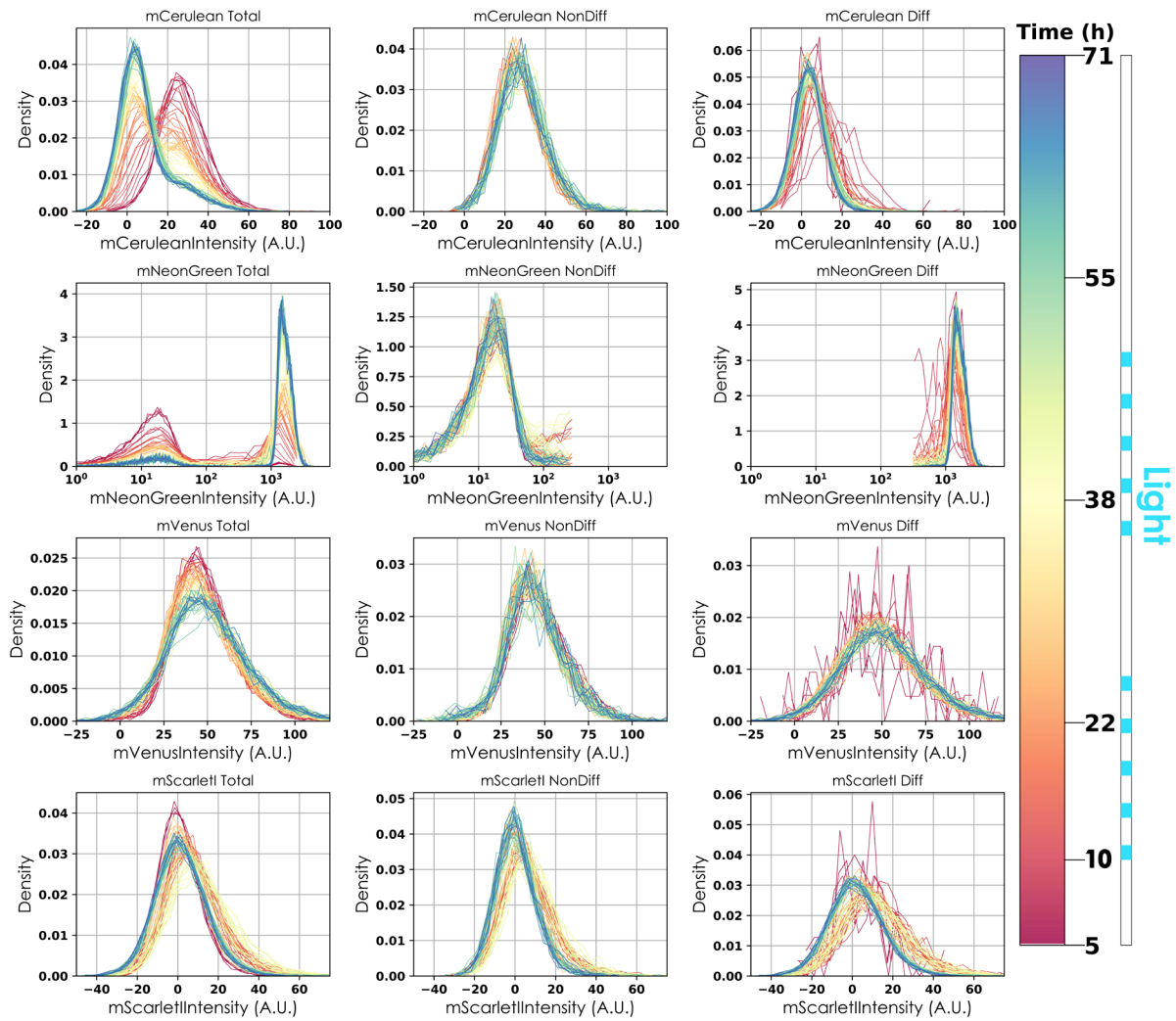


Figure 5.3. Histograms of the different fluorescent reporters. The fluorescence histograms were plotted for each fluorescent protein at each timepoint and further separated by the subpopulations (differentiated and non-differentiated). The color indicate the time of sampling (see colourbar). Light sequence can be found next to the colourbar. Due to deconvolution errors (Methods, 6.6), the final EL222 (*mVenus*) distribution of differentiated cells was slightly different from the initial distribution prior to induction.

However, we found only minor differences in EL222-TF levels of non differentiated cells before and after light induction (**Figure 5.3, *mVenus***) that were difficult to distinguish from small inaccuracies in deconvolution. We observed that population dynamics of differentiated and non-differentiated cells emerge deterministically for given light stimulation patterns and that the system's response to light could be captured fairly well (**Figure 5.4b**) by a simplistic population dynamics model (**Figure 5.4a**) that relates differentiated and non-differentiated cells via a constant differentiation rate in the presence of light similar to the model presented in Chapter 3.

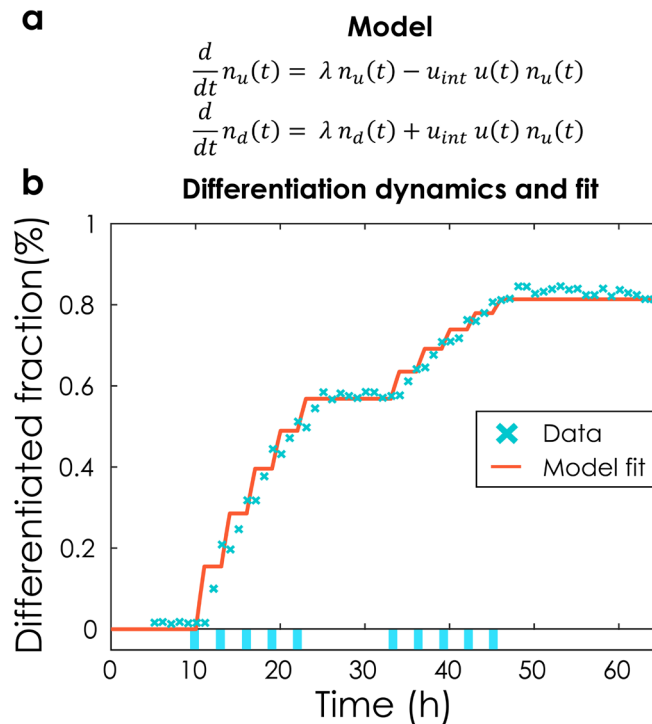


Figure 5.4. a. Simple ODE model. b. Differentiation fraction over time and model fit. Blue crosses represent the differentiation fraction for the experiment shown in Figure 5.2 & 5.3. Solid red line is the model fit. Cyan patches at the bottom indicate light pulses. This simple model is able to capture population dynamics.

Overall, we are led to conclude that cell-to-cell variability in EL222-TF and Cre can be safely ignored and that the functionality of the system can readily be characterized by a simple deterministic model. However, past experience in synthetic biology has shown that most circuits only function reliably in tightly constrained operating conditions and even seemingly good models retain their predictive power only in the precise context that has been used to construct the model.

Plasmid tetra reporters

To test if the differentiation system remains functional when the context of the circuit is modified, we constructed variants of the system in which EL222-TF and Cre genes are placed on 2-micron and centromeric plasmids instead of being integrated into the chromosome (**Figure 5.5a**).

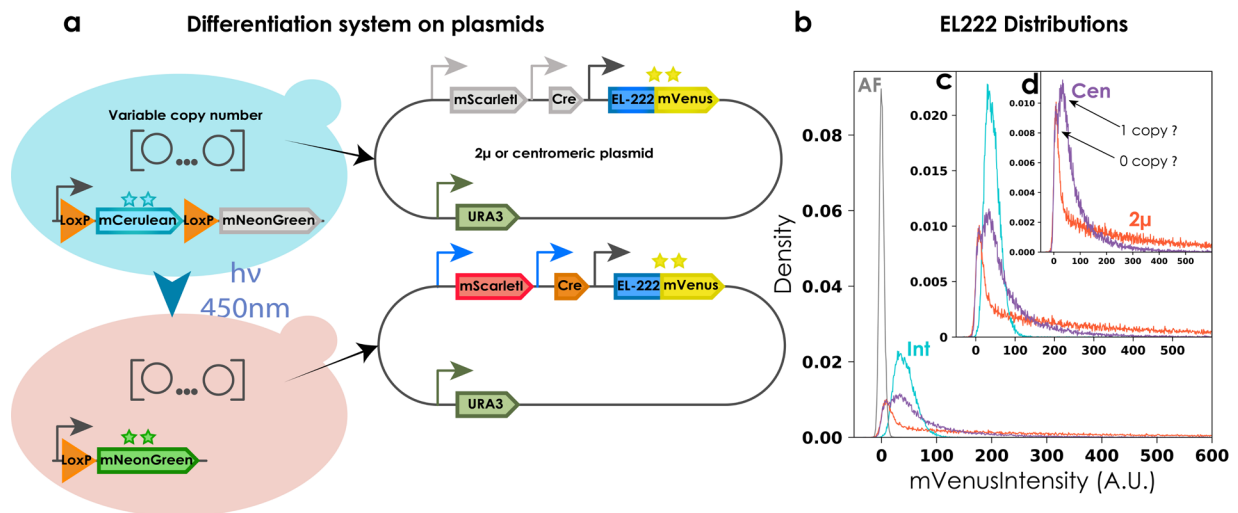


Figure 5.5. Plasmid version of the tetra reporter and variability in EL222 levels. **a.** An integrated recombination cassette was supplemented with the differentiation system on either a 2-micron or a centromeric plasmid. **b. Comparison of mVenus fluorescence in the tetra reporter strains in the dark w.r.t. an autofluorescence strain (AF) lacking fluorescence.** Histograms with different colors represent the strains. **c. Zoomed inset of b.** to highlight the differences between the integrated and plasmid strains. **d. Zoomed inset of c** to highlight the differences between the two plasmid strains. The integrated strain, expected to have a single copy, shows a unimodal distribution. The 2-micron plasmid strain showed a lot of variability in the EL222 levels and showed a peak at zero fluorescence presumably due to absence of plasmids in a subset of the population. Notably, the centromeric version seemed to possess two peaks, one it shared with the integrated strain and the other with the 2-micron strain suggesting that centromeric plasmids are predominantly present in 1 copy and 0 copies.

Since these plasmids are present in variable copy numbers in cells, we expected this change to lead to significant differences in average EL222-TF and Cre levels and variability in cells. Indeed, measuring the fluorescence of exponentially growing cells in the dark via flow cytometry, we found that EL222-TF distributions are markedly different from the integrated system version and characterized by much heavier tails and a mode that is shifted to lower levels (**Figure 5.5b**). Taken together, these two features imply that on average cells in the population contain more EL222-TF but at the same time more cells contain less EL222-TF compared to the integrated system version. We also observed a persistent subpopulation that possessed no mVenus (autofluorescence levels) suggesting an absence of plasmids in these cells. These observations are consistent with the findings reported in Fang *et al.* (2011), Karim, Curran and Alper (2013), and Gnügge, Liphardt and Rudolf (2016).

We may therefore wonder if and how these differences influence the functionality of the circuit and if emerging dynamics of the population composition can still be predicted by the simple model in **Figure 5.4a**. Exposing cells to different light patterns, we found that the same amount of light leads to differentiation of more cells for the plasmid based versions of the system (**Figure**

5.6), which is in line with higher average levels of EL222-TF in the population and the presumable presence of multiple plasmids in cells each carrying a copy of the promoter driving Cre.

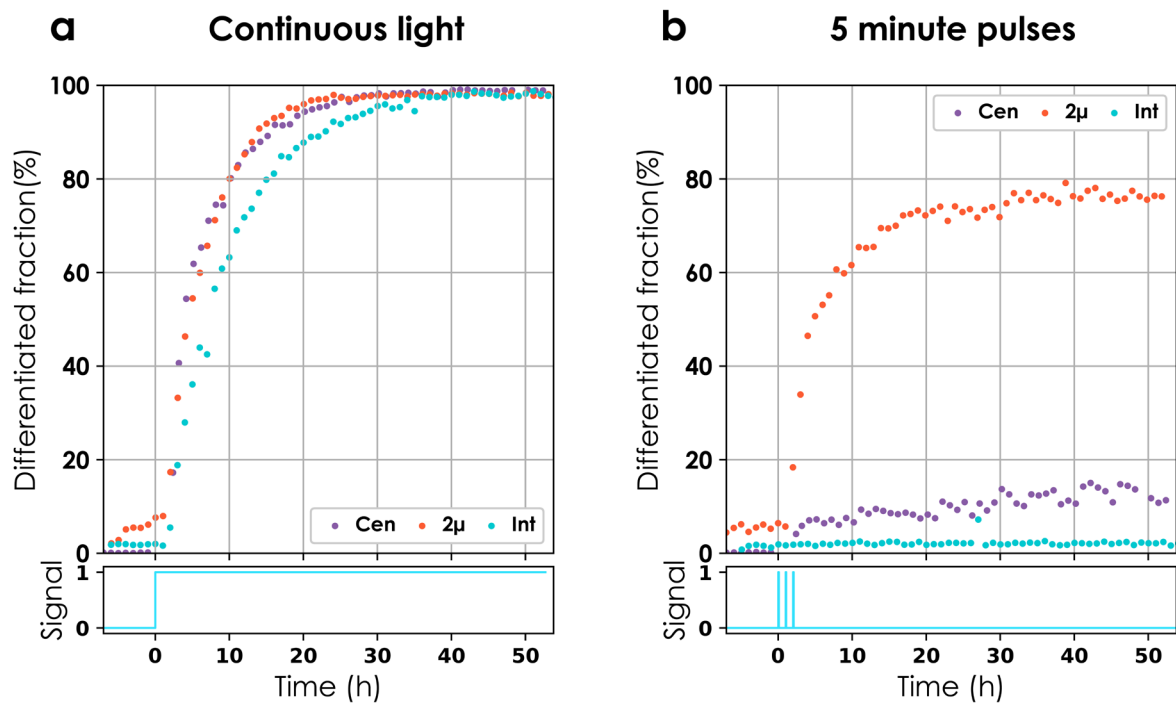


Figure 5.6. Comparison of differentiation behaviour for two light signals between the different tetra reporters. Cells were cultured to exponential phase and exposed to either **a.** continuous light or **b.** 5 minute pulses. Solid circles represent the differentiated fraction and the light input is shown in cyan at the bottom.

Adjusting the differentiation rate parameter in the simple deterministic model to account for the on-average presence of multiple copies of the system, however, does not lead to agreement of model predictions and data. The slow dynamics convinced us that it would not be possible at all to obtain any precise fit of this model to the population dynamics that emerge from the plasmid-based differentiation systems (**Figure 5.7**). Concretely, the increase in differentiation fraction 10h post removal of the light input was impossible to explain using the ODE model.

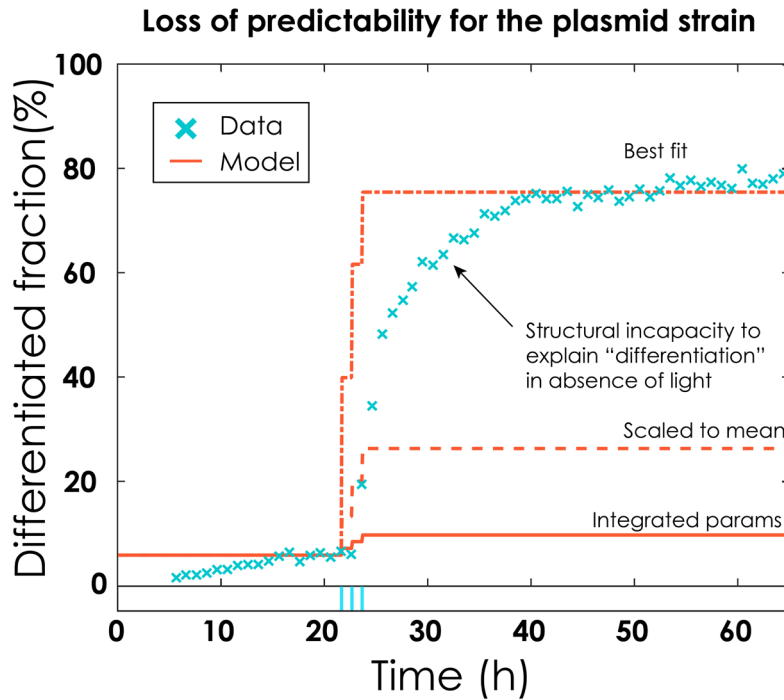


Figure 5.7. Incapacity to accurately predict the differentiation dynamics. The simple model of Figure 5.4a fails to capture population dynamics for the plasmid-based version of the system. Crosses: measured differentiated population fractions. Solid line: model predictions using the same parameter values as in Figure 1d for the integrated system. Dashed line: model predictions after adjusting the differentiation rate parameter to account for the presence of multiple copies of the system. Dash-dotted line: refitting the model to match final stationary differentiated fractions (30-40h after first light induction) leads to significant model mismatch during early transient dynamics. The applied light sequence is shown at the bottom.

Upon analyzing mVenus distributions, we found that differentiated cells shortly after applied light pulses are characterized by high EL222-TF levels whereas the mVenus distribution of the non-differentiated sub-population is shifted to lower levels compared to the total mVenus population distribution (**Figure 5.8 & Figure 5.9 mVenus channel**). Since EL222-TF is constitutively expressed from the same promoter on plasmids in all cells, we conclude that these differences must be caused by selective differentiation of cells with high amounts of EL222-TF. Differences between sub-populations gradually disappear over time but are still clearly noticeable days after the last application of light to the population (**Figure 5.8**). This is quite remarkable as it is difficult to comprehend, at a first glance, how a constitutively expressed gene can display a cellular memory of a stimulus that is retained over several tens of cell generations. In conclusion, cell-to-cell variability in EL222-TF, which previously seemed to be negligible for the characterization of the system, suddenly appears to be of key importance for understanding how population dynamics emerge from the differentiation system.

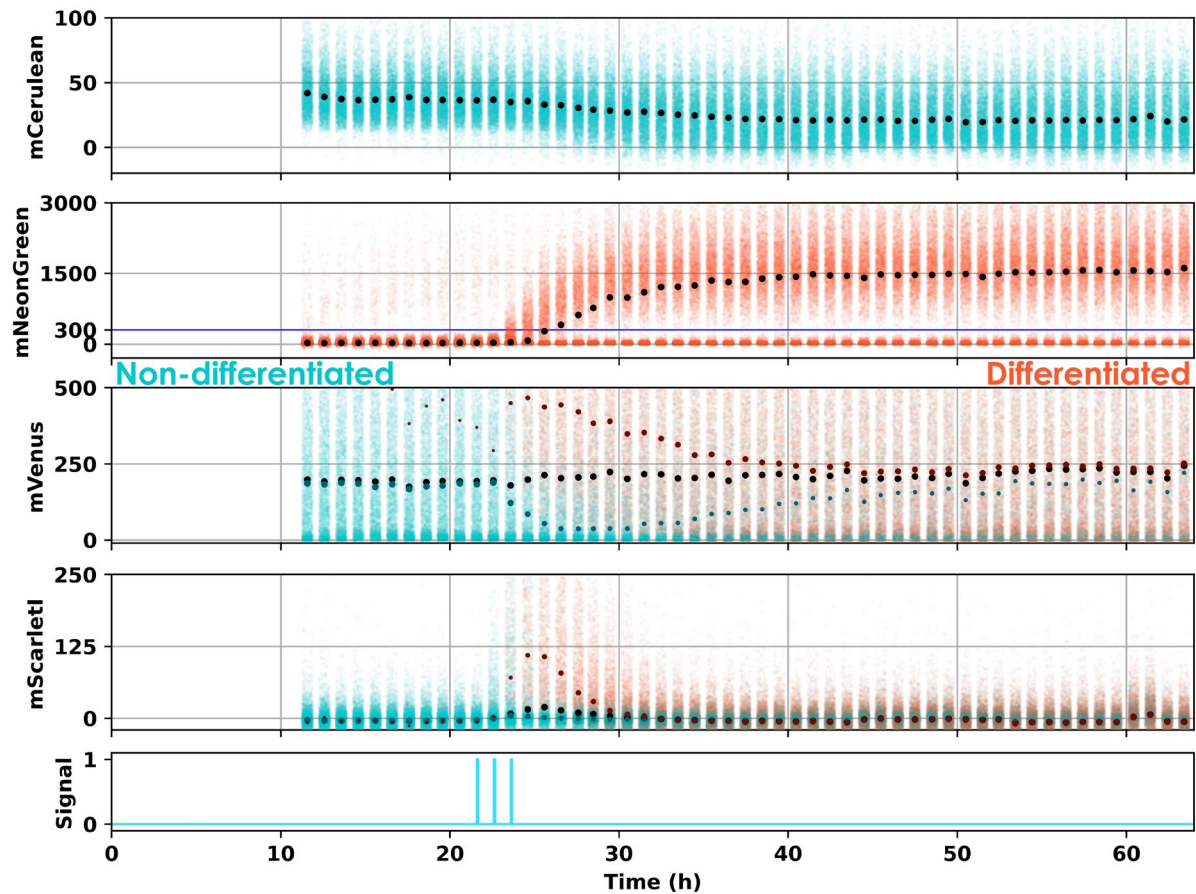


Figure 5.8. Distributions of the different fluorescent reporters for the 2-micron tetra reporter. Data is represented as in Figure 5.2. Large cell-to-cell variability in EL222 propagates to large variability in mScarlet-I (and thus Cre) production with some cells displaying much higher mScarlet-I fluorescence compared to the integrated system version (5-fold increased scale on y-axis for mScarlet-I and mVenus compared to Figure 5.2). Using mNeonGreen fluorescence to classify cells into differentiated and non-differentiated (blue line) shows that light causes a transient split in mVenus distributions of the two sub-populations in response to light. The applied light sequence is shown at the bottom (three 5min pulses with 55min between subsequent pulses).

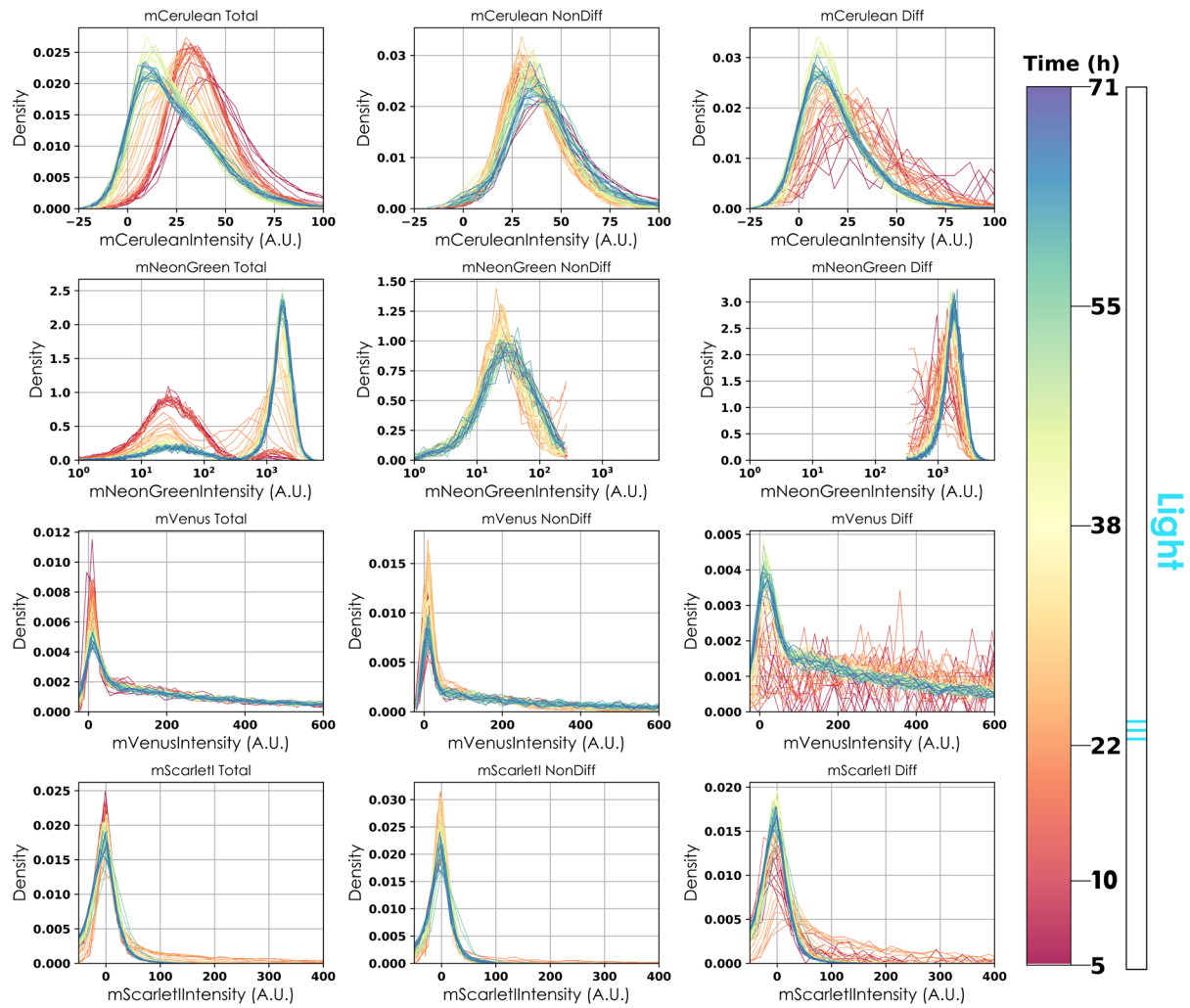
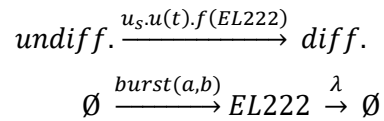


Figure 5.9. Distributions of the different fluorescent reporters. Data is represented as in Figure 5.3. A clear selection effect is present in the population. Upon light induction, cells with high EL222 levels differentiate preferentially leading to a transient shift in the distributions that relax back over time. Due to deconvolution errors (Methods, 6.6), the final EL222 distribution of differentiated cells was slightly different from the initial distribution prior to induction. Note that the scale is different from Figure 5.3.

A stochastic model for the differentiation system

To test if consequences of changes in the context of the system can be understood and predicted, we constructed a stochastic kinetic model of the differentiation system and a model of plasmid copy number fluctuations and asked if the models can be composed to predict emerging single-cell and population dynamics when the differentiation system is expressed from plasmids. Concretely, since variability in EL222-TF appeared to be of key importance, we deployed a model of bursty production of EL222-TF and cell differentiation to represent the differentiation system:



where u_s is the maximal single cell differentiation rate for given fixed light intensity, $u(t)$ is equal to one in the presence of light and zero otherwise, λ is the cells' growth rate, and a is the rate at which protein bursts occur. Protein production bursts are assumed to be geometrically distributed with average burst size b , as dictated by classical results for modeling stochastic gene expression (Friedman, Cai, and Xie 2006; Shahrezaei and Swain 2008). To keep the model as simple as is strictly necessary, we neglected possible delays or noise caused by the production and action of recombinase or the experimental detection of recombined cells. Instead, we assumed that the probability per unit time for a cell to differentiate in the presence of light is directly a function of the amount of EL222-TF, $f(EL222)$, in the cell.

When $u(t) = 0$, the EL222-TF distribution converges to the stationary gamma distribution of the standard bursty protein production model. Growing cells in the darkness and measuring their mVenus fluorescence by flow cytometry, therefore allows one to fix average burst size and frequency in the model by matching the stationary gamma distribution of the model to mean (up to a fluorescence scaling factor) and coefficient of variation of measured fluorescence distributions. When light is applied to the population, $u(t) = 1$, the dynamics of the differentiated population fraction are determined by the specific shape and parameters of the differentiation function $f(EL222)$. Matching emerging population dynamics for the light patterns in **Figure 5.2**, we found that a steep Hill-function with a threshold significantly larger than average amounts of EL222-TF leads to good agreement with the data. Mathematically, the model couples a variability generating process (stochastic production of EL222-TF) to a population process that selectively differentiates cells based on their phenotypes. The coupling

of single cell and population scale tells us that, in response to light, we should expect preferential differentiation of cells with high levels of EL222-TF.

Stochastic model for the differentiation system

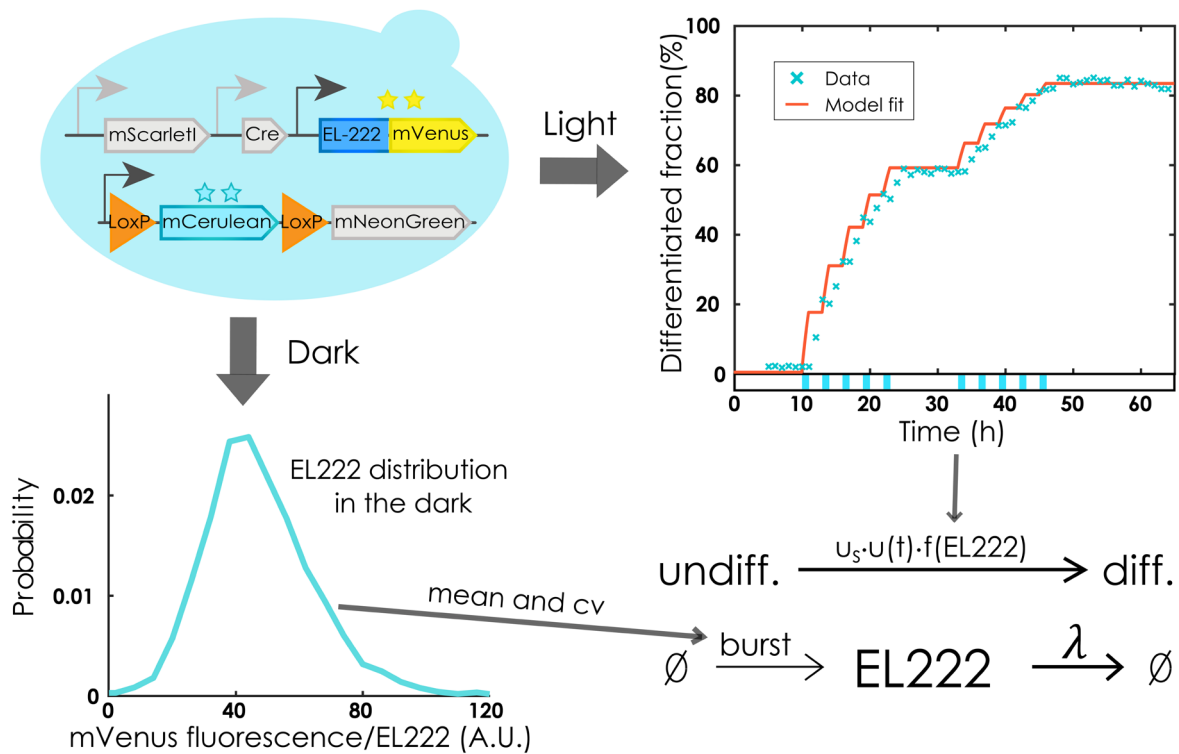


Figure 5.10. Stochastic model for the differentiation system. Bottom: Extracting mean ($= 50.6$ A.U.) and coefficient of variation ($= 0.35$) from measured stationary EL222-TF distributions before light induction allows one to determine burst frequency and average burst size of the EL222 production model up to a fluorescence scaling factor. Right: Exposing cell populations to light pulses and measuring dynamics of the resulting population composition (crosses: data, line: model fit, bottom: applied light sequence) then allows one to determine parameters of the differentiation function $u_s f(EL222)$.

Correspondingly, differentiated cells should display increased amounts of EL222-TF shortly after light while the (sub-) population distribution of non-differentiated cells should shift to lower levels. However, this selective population split is counter-acted by the fact that the same variability generating process is operating in all cells and that this process will always drift to the original gamma distribution of EL222-TF in both sub-populations at a timescale that is determined by the cells' growth rate. In the absence of light, we therefore expect distributions to revert back to the original gamma distribution fairly quickly while the continuous presence of light should lead to a quasi-stationary condition in which single cell and population process are dynamically balanced until eventually all cells will have differentiated. The quasi-stationary EL222-TF distribution for non-differentiated cells in the presence of light can be calculated from the bursty protein production model. These computations are non-trivial but beyond the scope of this chapter and

can be found in Annex 5 (S.2. Stochastic modeling of the integrated differentiation system). Performing this calculation, we found that the maximum possible shift in EL222-TF distributions of non-differentiated cells that can potentially be observed in experiments is fairly small and of similar size as experimental errors due to inaccurate deconvolution or reactor-to-reactor variability.

To test whether such a shift can nevertheless be detected, we exposed cells to continuous light and collected measurements at time points early enough after induction such that sufficiently many cells remain non-differentiated to allow for reliable quantification of EL222-TF distributions (Figure 5.11a). We found that experimental EL222-TF distributions of non-differentiated cells indeed seem to show a small shift towards lower levels in response to continuous. This shift is in good agreement with distribution dynamics predicted from the model (Figure 5.11b). As a side note, the model provides a very good prediction of the increase in the differentiated fraction in response to continuous light (Figure 5.11a).

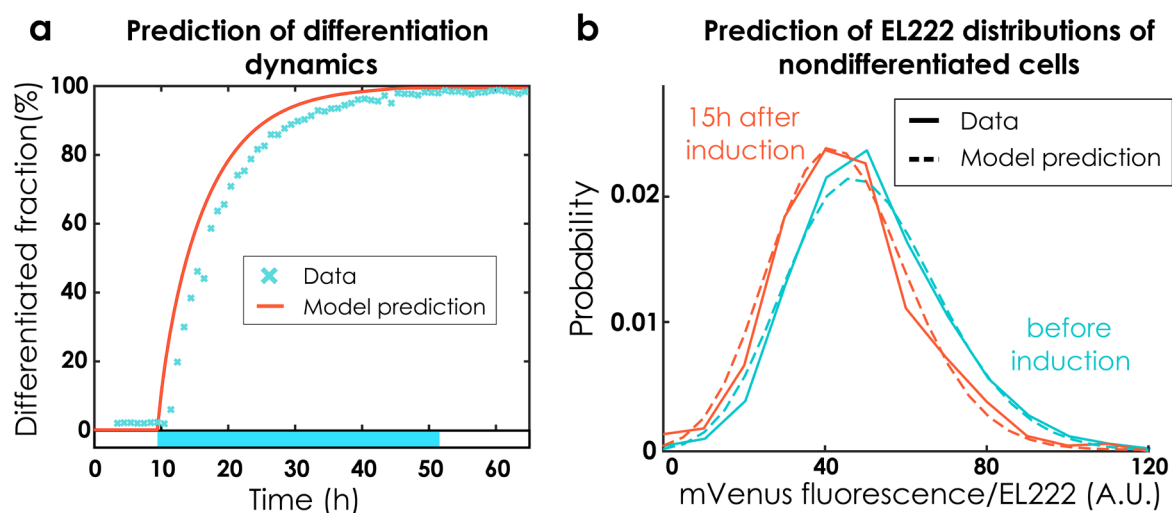


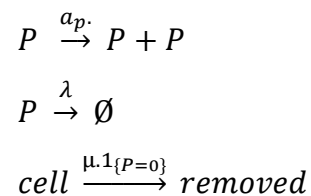
Figure 5.11. Predictions for continuous light experiment. a. Differentiation dynamics and b. Distributions of non-differentiated cells before and after induction. **a.** Exposing cells to continuous light eventually leads to differentiation of the entire population. **b.** Cells that remain non-differentiated after light induction (light blue) seem to be characterized by slightly lower EL222 levels compared to initial levels (dark blue) according to both model predictions (dashed) and data (solid).

Despite the possible presence of small selection effects, we can overall conclude that sufficiently low noise in EL222-TF production coupled to sufficiently fast fluctuations implies low selection. Stated differently, single cell fluctuations in EL222-TF replenish the original gamma distribution faster than the population selection process can shift them. Detecting such small changes reliably is experimentally difficult due to small but persistent errors in deconvolution but we note that the dynamically changing population differentiation rates that emerge from the dynamics of EL222-

TF sub-population distributions in response to light (**Figure 5.11b**) lead to slightly better agreement with observed population dynamics compared to the simple deterministic model (**Figure 5.11a**). Despite the possible presence of small selection effects, we conclude overall that sufficiently low noise in EL222-production coupled to sufficiently fast fluctuations implies that cell-to-cell variability has only small consequences for emerging population dynamics. It is clear, however, that this conclusion will change if either noise levels or time scales of the single cell process are modified.

A stochastic model for plasmid copy number dynamics

To be able to understand and predict emerging population dynamics when the optogenetic differentiation system is expressed from plasmids, we first required a quantitative mathematical description of how plasmid copy number fluctuations affect noise levels of expressed genes. To this end, we started by focusing on 2-micron plasmids. Many, often detailed, models of plasmid copy number fluctuations exist in the literature (Paulsson and Ehrenberg 2001; Gnügge, Liphardt, and Rudolf 2016). In order to keep the system characterization as simple as is strictly necessary, we decided to omit any detailed mechanistic description of processes such as replication failure or unequal division of plasmids between mother and daughter cell (Huh and Paulsson 2011). Instead, we chose to represent plasmid copy number fluctuations by a simple stochastic birth-death process with both birth rate (representing replication) and death rate (representing dilution due to cell growth) being linear in the plasmid copy number:



where P is the number of plasmids in a cell, a_p is the plasmid replication rate, λ is the growth rate of cells, and μ is the rate at which cells that have lost the plasmid are removed from the population when cells are growing in selective media. Replication failure can be implicitly incorporated by choosing the replication rate smaller than the cells' growth rate, which is in any case a necessary feature of a birth-death process model since expected plasmid copy numbers in cells would diverge to infinity if plasmids replicated faster than cells divide. For $a_p < \lambda$ (and $\mu = 0$), on the other hand, the process will eventually reach zero plasmids with probability one. The rate at which cells in the population lose the (last copy of the) plasmid is then determined by the difference between replication and growth rate. When selective media is used for growth,

$\mu > 0$, cells that have lost the plasmid will either die or be outgrown and the plasmid copy number distribution of the population will remain stable. However, selective media implies neither that there exist no cells without plasmids in the population nor do plasmid copy numbers remain constant from the perspective of single cells. Instead, we should expect a dynamic equilibrium and a quasi-stationary plasmid copy number distribution, in which the single cell plasmid loss process is balanced with selective removal of cells without plasmids. We highlight that, from a mathematical perspective, this result is equivalent to what was obtained in the previous section for the EL222-TF distribution in the non-differentiated cell population. In both cases, a variability generating process at the single cell scale (TF fluctuations vs. plasmid copy number fluctuations) is coupled to a state-dependent removal process at the population scale (differentiation vs. removal of cells that have lost the plasmid) and leads to the same type of non-linear master equation for the cells that have not been removed yet. We conclude that static population measurements of plasmid copy numbers or gene expression from plasmids may be misleading, as they do not provide full information on the dynamical aspect of the equilibrium plasmid copy number distribution.

To characterize the full multi-scale dynamics of plasmid copy numbers and cell population fluctuations, we measured the difference in growth rate of our strains between selective and non-selective media. We found that selective media reduces the growth rate of the 2-micron strain by around 15% (**Figure 5.12a, c**). Furthermore, we switched cells from selective to non-selective media and measured how the average abundance of a constitutively expressed fluorescent protein decays over time (**Figure 5.12b**), the fluorescence being an approximate indicator of the plasmid copy number.

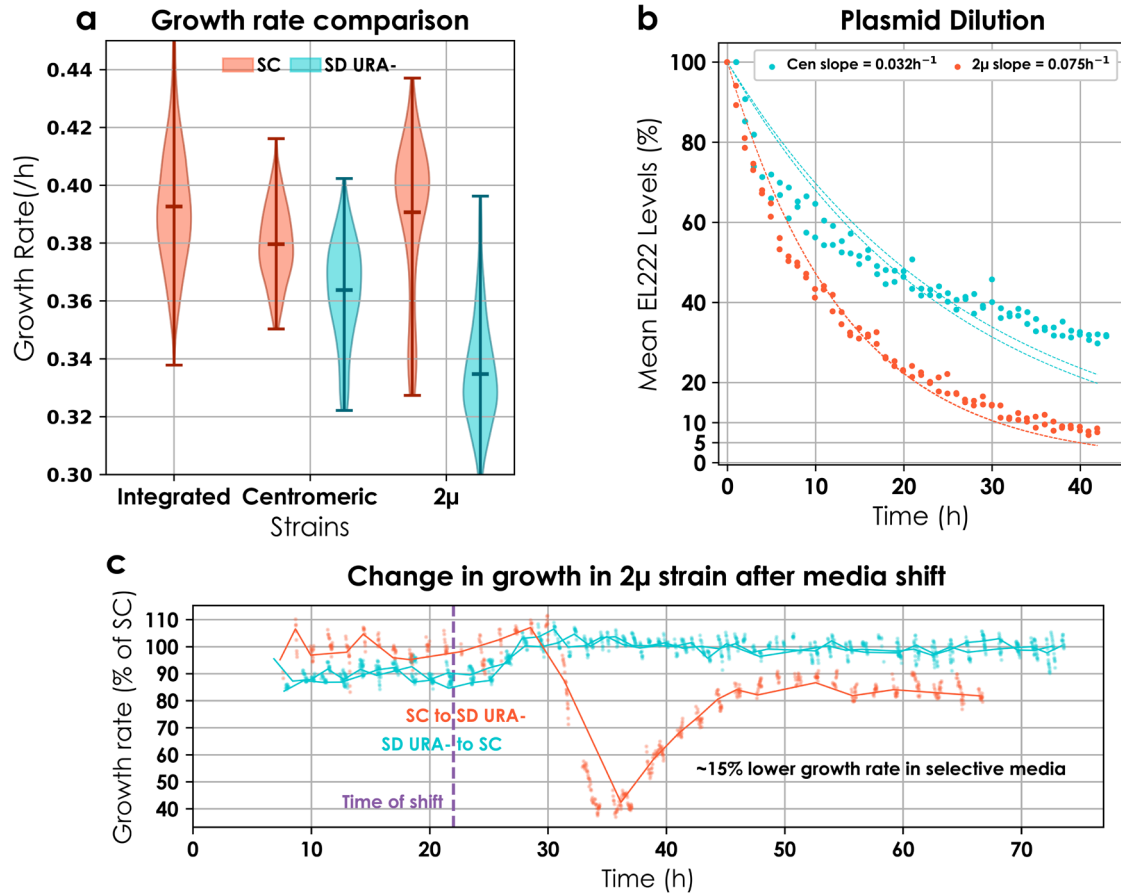


Figure 5.12. Characterization of plasmid strains. a. Growth rate differences in selective and non-selective media. Violin plots are composed of growth rate estimates based on cell density averaged over all reactors of the turbidostat and multiple experiments (in total 19 reactors for the strain with the integrated system, 12 for the strain with the plasmid-based system in selective media and 6 in non-selective media). SC stands for synthetic complete media and SD URA⁻ stands for synthetic defined media lacking Uracil. **b. Plasmid dilution rates.** 2-micron (red) and centromeric (blue) tetra reporter cultures (in duplicates) growing exponentially in selective media were switched to non-selective media at $t=0$. The circles represent the mean population fluorescence normalized to the first timepoint obtained via regular flow cytometry measurements. **c. Growth rate dynamics upon media switch.** 2-micron tetra reporter cultures were cultured to exponential phase in either selective media (blue, duplicates) or nonselective media (red, unique experiment) and at $t=22$ h, the media was switched to non-selective and selective media respectively. Solid lines and faded circles represent the discrete and continuous fit estimates for the growth rate, respectively (Methods). Growth rates are normalized by the growth rate in non-selective media.

Mathematically, the population dynamics of cells that have or have not lost the plasmid can be described by the following ordinary differential equations

$$\frac{dn}{dt} = \lambda n - ln$$

$$\frac{dn_0}{dt} = (\lambda - \mu)n_0 - ln$$

where n is the (expected) number of cells that have not lost the plasmid, n_0 is the number of cells that have lost the plasmid, $\lambda = 0.0067 \text{ min}^{-1}$ is the single cell growth rate (corresponding to a division time of 103min), l is the plasmid loss rate, and μ represents a removal rate of cells that have lost the plasmid. Taking a single cell perspective, plasmid loss is a cellular event and it a priori unclear if it can be appropriately described by single rate parameter since such a description implicitly carries the assumption that the waiting time for the event to occur follows an exponential distribution. However, one can show that the simple population dynamics model can be derived from a mechanistic representation of single cell plasmid copy number fluctuations where the plasmid loss rate emerges as the difference between dilution and replication rate, $l = \lambda - a_p$.

In non-selective media, it holds that $\mu = 0$, the total population grows at a rate λ , and the ratio $\frac{n}{n+n_0}$ will converge exponentially to zero at a rate that is determined by the plasmid loss rate l . Similarly, the average plasmid copy number of cells in the population decreases exponentially at rate l when cells are switched from selective to non-selective media. Assuming that the total rate of protein production of a cell is linear in the number of plasmids implies that expression levels of constitutive proteins approximately decay exponentially at rate l since, from a mechanistic perspective, the plasmid loss rate must be smaller than the growth rate λ . This implies that the plasmid loss rate operates on the slowest time scale and therefore determines changes in protein levels. Since gene expression levels can readily be measured, we can experimentally quantify $l = 10^{-3} \text{ min}^{-1}$ by switching cells from selective to non-selective media and observing the decay rate of the mean of a constitutively expressed protein (Figure 5.12b). From this result, the plasmid replication rate can be derived as $a_p = 0.0057 \text{ min}^{-1} = 0.85\lambda$ i.e. each plasmid is only replicated successfully with a probability of 0.85.

In selective media, μ will take a strictly positive value and must be sufficiently large compared to l to ensure that cells with plasmids can be maintained in the population. If this is the case, the fraction of cells with plasmids converges to $\frac{n}{n+n_0} = \frac{\lambda-l}{\lambda}$ and the total population growth rate, $\lambda_{select} = \lambda - l = a_p$. The seemingly counterintuitive result that the growth rate in selective

media does not depend on μ , and is instead equal to the plasmid replication rate, is because reduced growth rates for cells without plasmids imply that the fraction of these cells will be correspondingly smaller in stationary growth conditions. This implies that growth rate measurements in selective media cannot be used to determine μ but instead provide an independent measurement of a_p . We found that selective media reduces the population growth rate by about 15% (Figure 5.12a), which is in exact agreement with the previous finding that the plasmid replication rate must be 15% smaller than λ .

With the value of l (double-) confirmed we set out to determine the value of $\frac{n}{n+n_0}$ by measuring the fraction of cells without plasmids in selective media. It was necessary to directly measure what fraction of the population has plasmids in stationary growth conditions since it can be shown that this fraction must be equal to $\frac{\mu - \lambda + a_p}{\mu}$ for $a_p < \lambda$ (plasmid numbers do not diverge) and $\gamma > \mu - a_p$ (cells with plasmids can be maintained in the population). To this end, we performed a colony counting experiment. We cultured cells harbouring the three systems to exponential phase in selective media in triplicates. We made serial dilutions to the culture and plated on selective and non-selective media. After 48h of incubation at 30°C in the dark, we counted the number of colonies on each plate manually (Figure 5.13a).

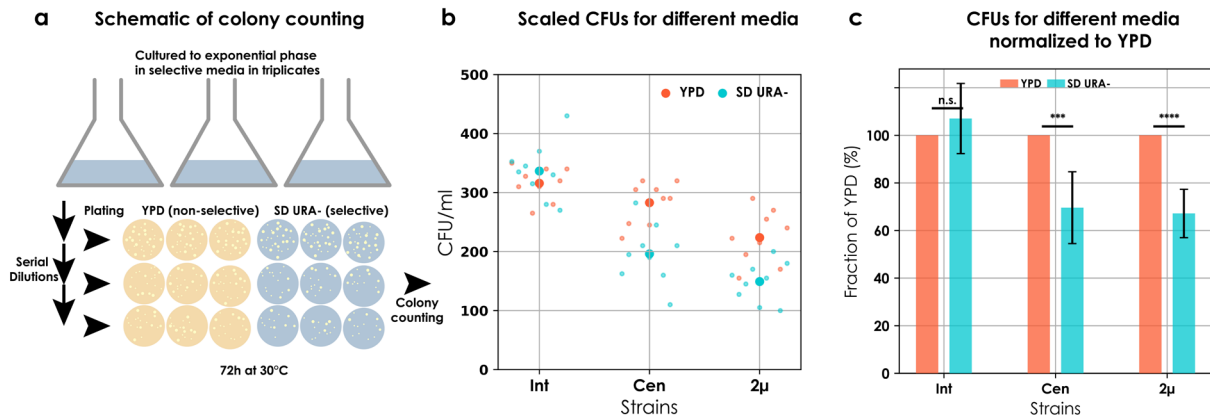


Figure 5.13. a. Schematic of colony counting. b. Cell counts from individual plates. Small circles represent the individual CFUs obtained for each plate scaled by the dilution factor for non-selective (red) and selective (blue). Large circles signify the mean. **c. Normalized CFUs for different strains.** Bars represent mean values and errorbars signify the standard deviation. Significance was tested via applying one-sided *t*-test.

I found that, while there was no significant difference in the number of CFUs at different dilutions between selective and non-selective media for the integrated strain, the number of CFUs for the plasmid strains were always lower than the integrated strain (5.13b). On average, upon normalization with the CFUs observed in the non-selective media at the given dilution, we

observed a significant decrease in the number of CFUs on selective media plates in the plasmid strains (Figure 5.13c). Comparing the values, we estimated that $30.81\% \pm 13.1$ (mean \pm s.d.) of the centromeric tetra reporter and $32.83\% \pm 10.16$ (mean \pm s.d.) of the 2-micron tetra reporter population lacked the plasmid. This allowed us to determine that the removal rate of cells without plasmids in the 2-micron strain, μ must be approximately 45% of the cells growth rate, $\mu = 0.45\lambda$.

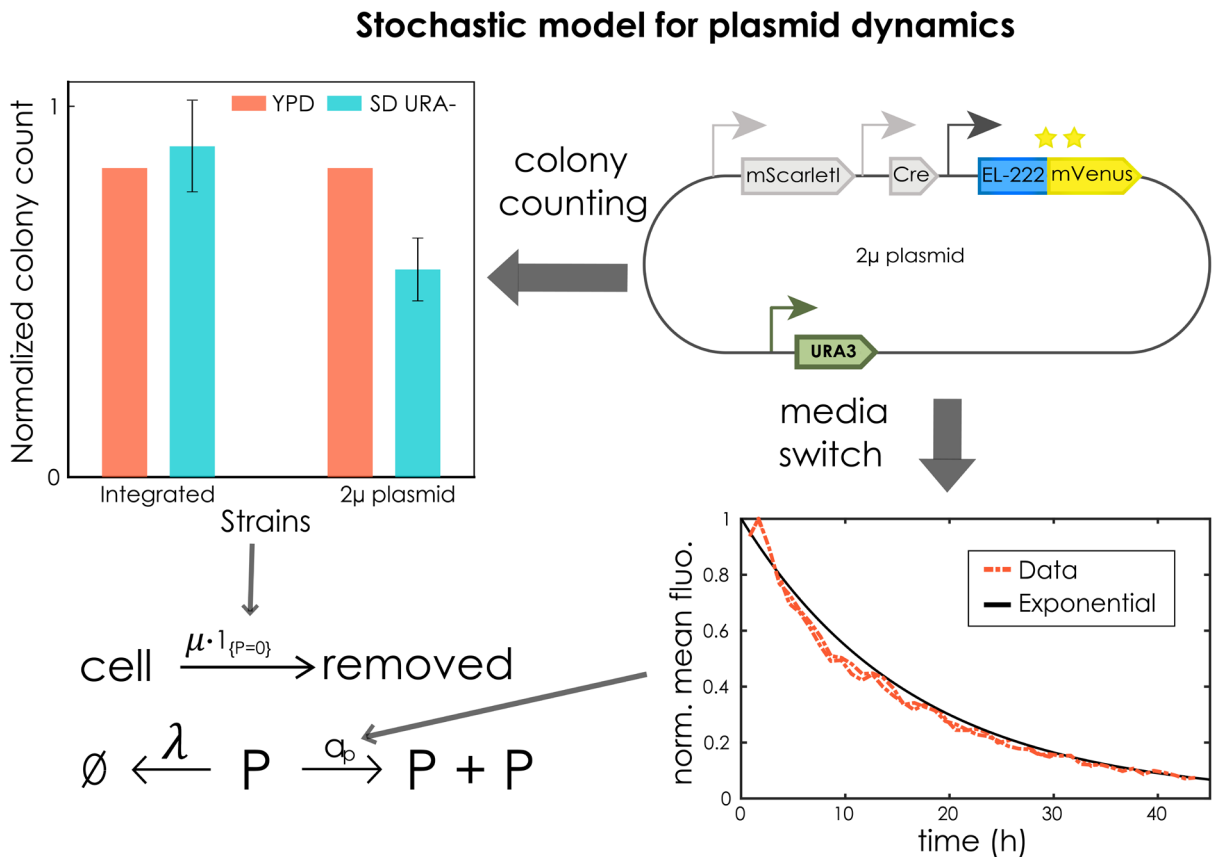


Figure 5.14. Stochastic model of plasmid dynamics (bottom left). Top left: Stationary fractions of cells without plasmids determine the removal rate in selective media of cells that have lost the plasmid. Bottom right: After switching cells from selective to non-selective media, the rate of exponential decay of average levels of a protein that is constitutively expressed from plasmids (here EL222-mVenus) can be measured to determine the population plasmid loss rate and to fix the plasmid replication rate in the model.

Together, the parameters α_p , μ and λ completely characterize the multi-scale dynamics of plasmid copy number distributions and cell populations (Figure 5.14).

Composed model and parameter free prediction of dynamics of the 2-micron tetra reporter

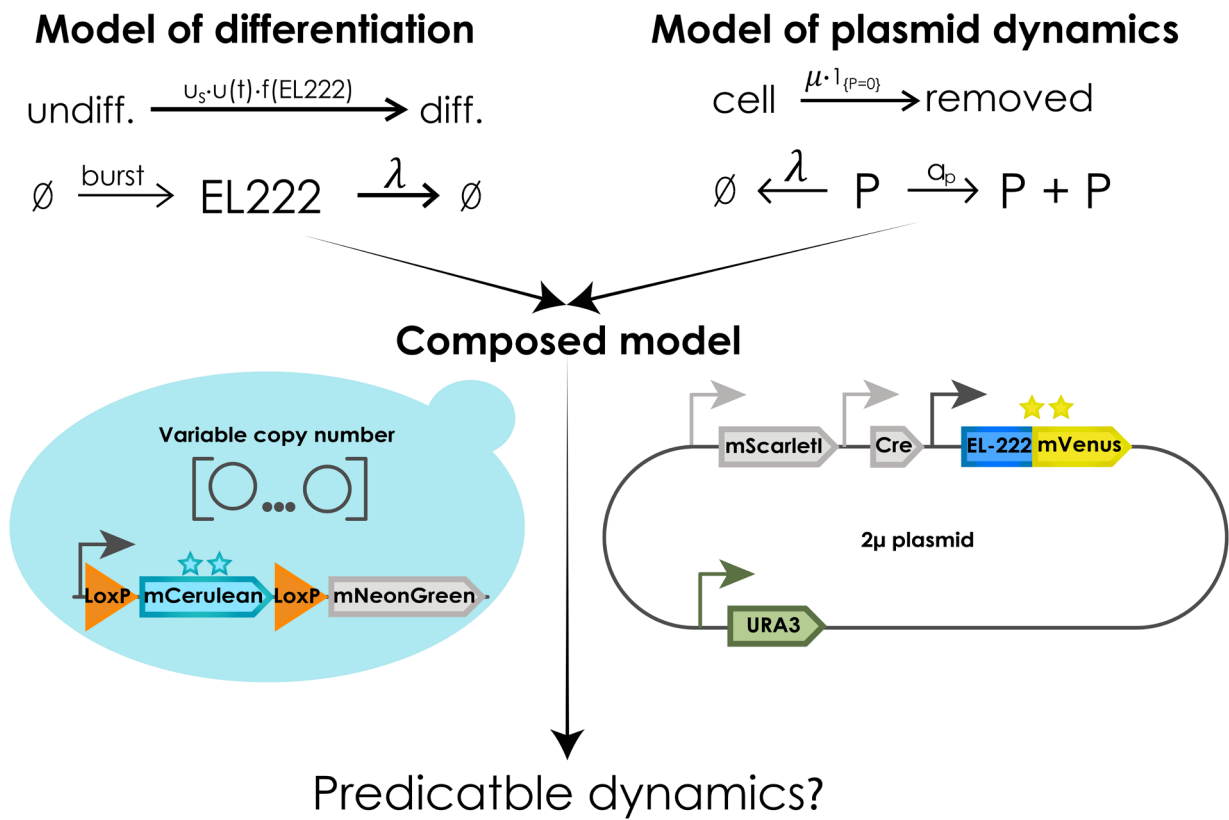


Figure 5.15. Schematic of the composed model. The two models could be composed into a single model but would it be sufficient to predict the population dynamics of the 2-micron tetra reporter?

In order to find out if our rational modelling approach with two stochastic models composed together would restore predictability, we used the composed model to predict the dynamics of the differentiation system in plasmid strains for the light stimulation pattern in **Figure 5.7** and compared the results to data. We found that the population dynamics of differentiated and non-differentiated cells are very well predicted without any adjustment of model parameters (**Figure 5.16a**). Comparing EL222-TF distributions in the two sub-populations between model and data shows that the high quality of population level predictions is a consequence of the fact that the model perfectly predicts how single cell processes will operate in union with population level processes to shape the full dynamics of EL222-TF distributions in sub-populations (**Figure 5.16b**). In particular, the heavy tails of EL222-TF distributions for cells growing in darkness (observed already in **Figure 5.5b**) emerge naturally from the fact that plasmid copy numbers can fluctuate significantly in the model (**Figure 5.16c**). These heavy tails imply that significantly

more cells have EL222-TF levels above the threshold of the differentiation Hill-function $f(TF)$ and will recombine quickly upon light induction. Shortly after light induction, the remaining non-differentiated cell population is therefore shifted to significantly lower EL222-TF levels (the median is reduced by a factor of around 5 to 6, **Figure 5.16b**) while the differentiated cell population inherits the heavy tail of the original distribution (**Figure 5.16d, left**).

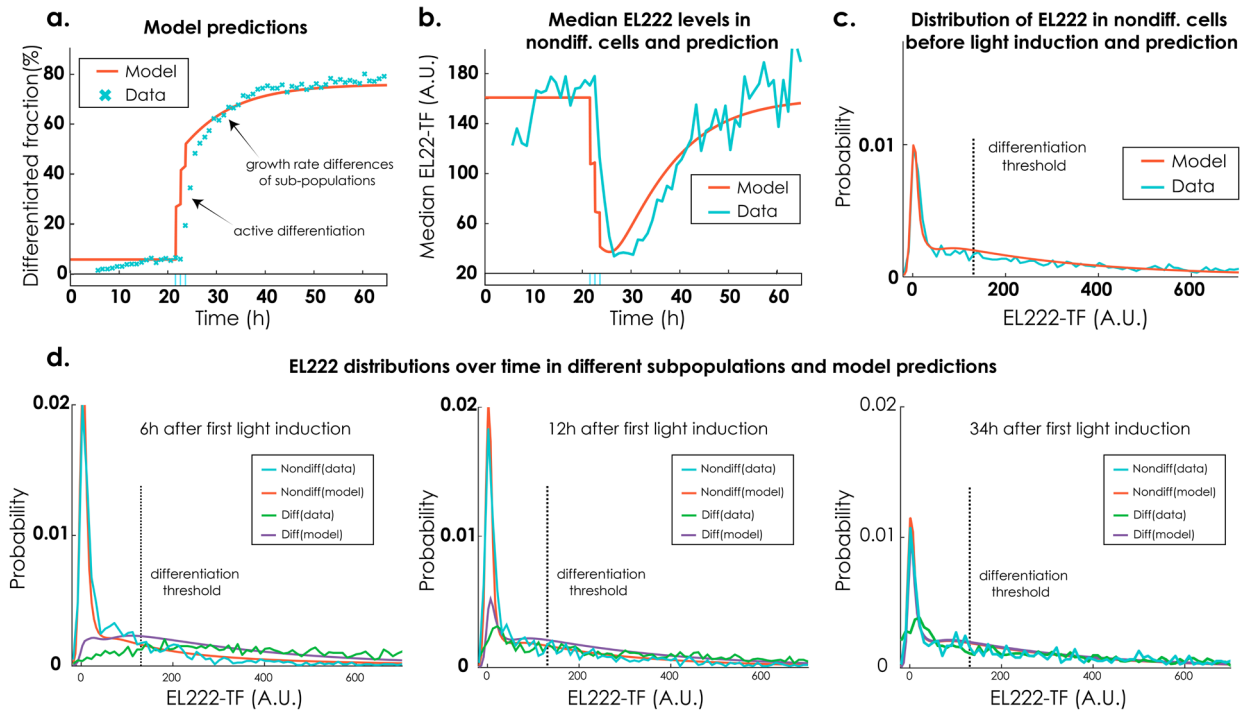


Figure 5.16. Parameter-free prediction of emerging population dynamics for the 2-micron tetra reporter

a. Emerging dynamics of the differentiated fraction according to the composed model (red) for the same light input as in Figure 5.7 are compared to experimental data (blue). According to the model, late increases in the differentiated fraction are due to varying sub-population growth rates and not caused by active differentiation. **b.** Upon light induction, median EL222 levels of non-differentiated cells drop 5-6 fold as predicted by the composed model. **c.** The model-predicted EL222 distribution in the dark agrees very well with experimental data even though this data was not used to parametrize the model. According to the model, the peak of the distribution close to zero corresponds to the ~ 33% of cells without plasmids (see Figure 5.13c). **d.** Shortly after the application of light, cells with EL222 levels above the threshold parameter of the differentiation Hill function (dashed line) are enriched in the differentiated population while the non-differentiated population contains more cells without plasmids (increased peak close to zero). When cells are subsequently kept in the dark (rightmost panel), EL222 distributions converge back to the initial condition. Model predictions show remarkable agreement with measured EL222 distribution dynamics, except for seemingly lower numbers of differentiated cells without plasmids in the data (in particular after 34h, purple distribution is hidden behind blue and red distributions). This small mismatch is presumably caused by inaccuracies in deconvolution (the presence of mNeonGreen in cells makes it difficult to precisely quantify low mVenus levels in cells (Methods, 6.6)).

While not experimentally measurable, it can be deduced from the model that the same holds true for plasmid copy number distributions in sub-populations (**Figure 5.17c & d**). As a consequence, the population differentiation rate spikes very high upon first light induction but is significantly reduced when subsequent light pulses are applied before the EL222-TF distribution of the non-differentiated cell population has converged back to its initial condition (**Figure 5.16b, 5.17b**).

When the light stimulus is maintained for some time, fluctuations in plasmid copy numbers create larger fluctuations in EL222-TF amounts compared to the integrated version of the system, which leads to more frequent threshold crossing events and therefore larger population differentiation rates (**Figure 5.6**). Eventually, the plasmid (2-micron) version of the system reaches differentiated population fractions close to 100% very quickly when light is maintained despite the fact that a large part of the population (around 30 - 40%) displays mVenus levels that are significantly lower than typical values for the integrated strain and indistinguishable from the cells' autofluorescence (**Figure 5.5b, Figure 5.8 mVenus channel**). According to the model, these cells have lost the plasmid (**Figure 5.17d**), cannot differentiate, but are likely to be removed at subsequent time points due to growth in selective media.

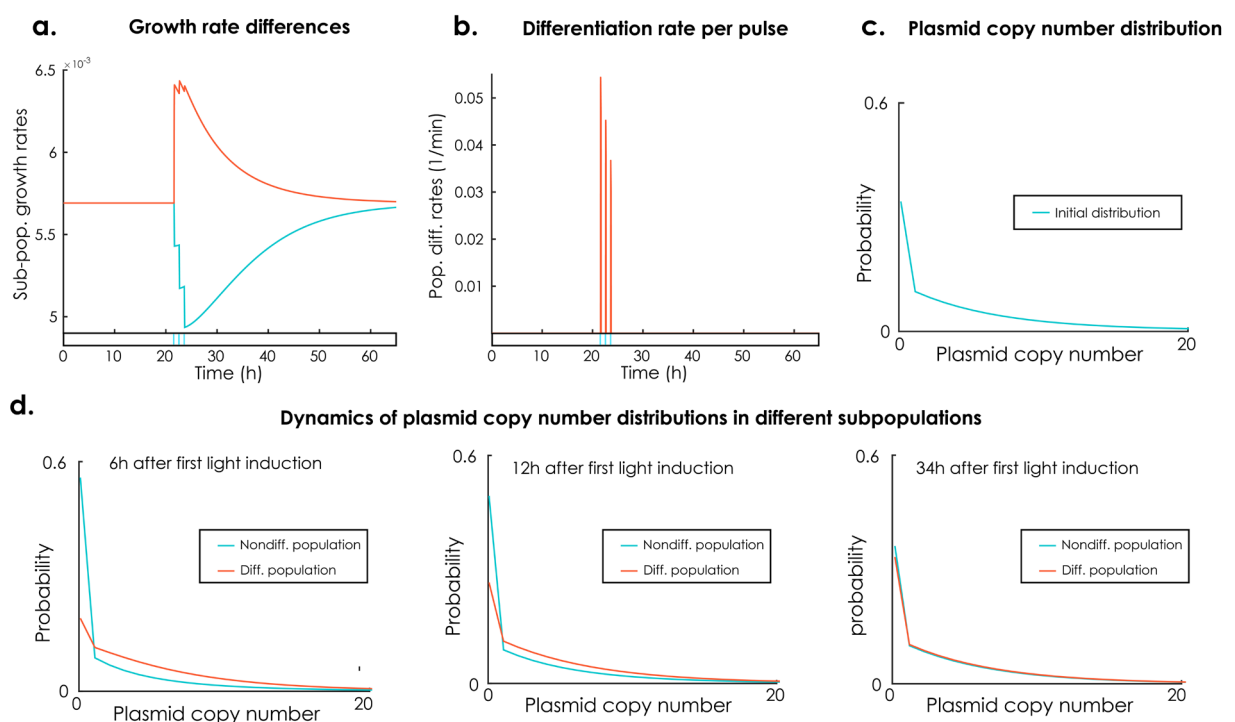


Figure 5.17. Model predictions for the 2-micron tetra reporter a. Subpopulation growth rate differences. **b.** Population differentiation rate. **c.** Initial distribution of plasmids in non-differentiated cells. **d.** Dynamics of plasmid distributions in different subpopulations.

The coupling of plasmid loss dynamics and selective media with the differentiation system therefore leads to unintuitive population dynamics in which seemingly the entire population can recombine despite a continuous presence of 30 - 40% cells that are not carrying any copy of the differentiation system (**Figure 5.5b and Figure 5.9, mVenus channel**). Another complex consequence of the coupling of single cell and population processes is that the split in plasmid copy numbers between sub-populations that follows from selective differentiation of cells with high EL222-TF levels leads to different sub-population growth rates in selective media (**Figure 5.17a**) since cells that have lost the plasmid (or are close to losing it) are enriched in the non-differentiated sub-population. This implies that the differentiated population fraction will continue to increase even if light is removed and no more active differentiation takes place, which explains why the assumption of the simple deterministic model described in **Figure 5.4**, that the differentiated population fraction can only increase due to active differentiation, lead to structural incapacity of the model to explain the slow transient differentiation dynamics and significantly increasing differentiated fractions up to 10h after last light induction (compare **Figure 5.8 & Figure 5.16a**).

If the light stimulus is removed, EL222-TF subpopulation distributions converge back to the original distribution (**Figure 5.16d, right and 5.9, mVenus channel**) but are still noticeably distinct even 30 to 40h after the last light stimulus. This experimental observation is in good agreement with slow convergence of sub-population plasmid copy numbers to their quasi-stationary distribution in selective media, except that the differentiated EL222-TF subpopulation distribution recovers the peak at zero even slightly slower than predicted by the model (**Figure 5.16d, right**).

Before concluding, we note that the system behaviour was reproducible for both the integrated and the 2-micron tetra reporter. This can be appreciated from the results of 3 replicates for each of the light sequences used for characterization (integrated tetra reporter) and validation (2-micron tetra reporter). All three experiments were done on different days. Notably, there was some heterogeneity present in the response, presumably due to reactor-to-reactor variability, (**Figure 5.18**) but despite this variability, two of the three replicates showed remarkable reproducibility. We stress that all three replicates were performed independently of each other and months apart in time.

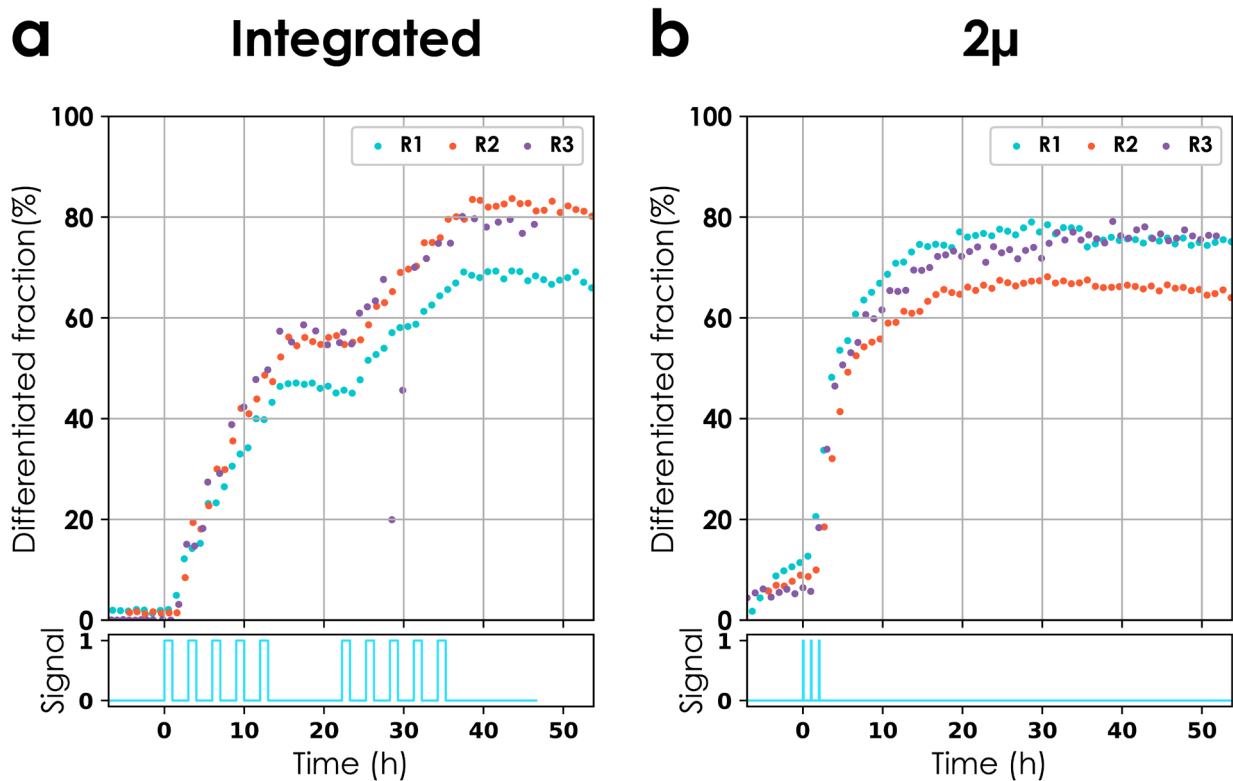


Figure 5.18. Reproducibility in a. Integrated and b. 2-micron tetra reporter. Circles represent the differentiated fraction for different light pulses. Cyan lines at the bottom reflect the light signal applied to the culture. The experiments were done in different reactors on different days. The main source of heterogeneity in response appears to be reactor-to-reactor variability.

We conclude that deploying stochastic chemical kinetics and appropriate mathematical approaches for understanding the interplay of single cell and population processes allows us to understand and predict complex dynamics when the context of the differentiation system is changed. Emerging population dynamics for plasmid strains, despite being shaped by cell-to-cell variability, are deterministically reproducible up to a small reactor-to-reactor variability (Figure 5.18b). Thus, the capacity to predict such dynamics implies that the interplay of single cell and population processes can be exploited for creating features of microbial community dynamics that would otherwise be difficult to engineer.

Control of plasmid copy number

Having demonstrated that we can use a stochastic multi-scale model to reliably characterize the optogenetic differentiation system on 2-micron plasmids, we can now focus on using it to engineer yeast community dynamics. Given the previously demonstrated predictive power of the model, it is clear that it can be used to determine how light stimulation profiles need to be chosen to create desirable population dynamics of differentiated and non-differentiated cells for all strains. A more challenging question is if it is also possible to exploit all the complex consequences

of cell-to-cell variability that we established in the previous sections. Concretely, if light selectively differentiates cells with large amounts of EL222-TF and cellular amounts of EL222-TF correlate with plasmid copy numbers, then it should be possible to maintain EL222 levels in the non-differentiated cell population at reduced constant levels by applying continuous light for sufficiently long. To this end, we queried the model for control of EL222 TF levels in the non-differentiated population and, by extension, the plasmid copy numbers (open loop control). We found that the coupled model allowed us to control the expression levels of EL222 in the non-differentiated population. Mechanistically, light pulses are spaced such that cells with high EL222 levels in the non-differentiated population constantly become differentiated and are “removed” from the population. In other words, light is used to remove cells that have higher EL222 TF levels than a given threshold, thus maintaining the levels of EL222 TF in non-differentiated population. Furthermore, we found that it was possible to maintain the non-differentiated cells at different reduced levels of EL222 levels and by extension suggesting that it was possible to control the plasmid copy number in subpopulations (lower than population level average in non-differentiated fraction and higher than the average in the differentiated fraction). We note that this claim is not entirely backed by the data provided here given that we cannot directly observe the plasmid copy numbers in different subpopulations.

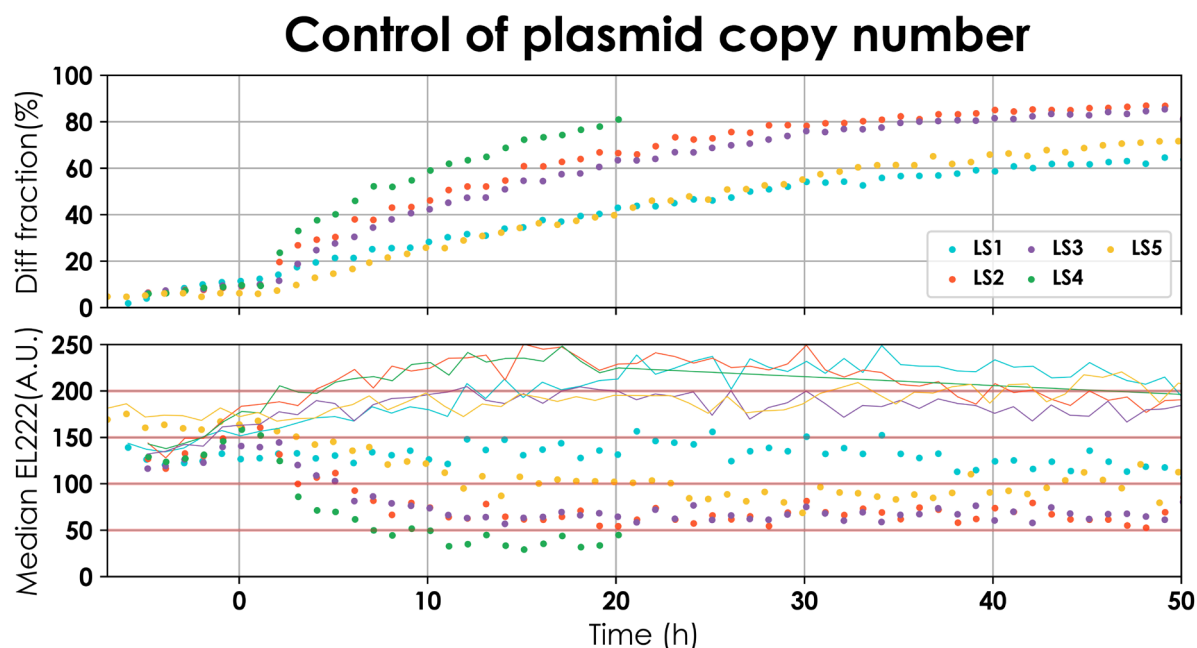


Figure 5.19. Control of constitutive expression. (Top) Differentiation dynamics for different regular light stimulation profiles. Circles represent the differentiated fraction. **(Bottom) median EL222 levels in the non-differentiated** (circles) and total population (lines). Red horizontal lines pertain to expression expected from 1-4 copies of the plasmid (bottom to top). We observe that the median EL222 in the non-

differentiated fraction can be maintained below the total population median by applying different light profiles; 0-1 copy (green, 5 min pulses every 4h), 1-2 copies (red, 1-min pulses every hour & purple 2-min pulses every 4h), 2-3 copies (yellow, 1 min-pulses every hour at intensity 20 (instead of 40)), and ~3 copies (cyan, 30s pulses every four hours). In each case, the total EL222 levels corresponded to 4-5 copies. Induction was started at $t=0$. Please note that the plasmid copy number are rough estimates gathered by comparing the observed fluorescence in these experiments with the fluorescence levels of (single copy) integrated tetra reporter (Figure 5.5b).

I highlight the experiment corresponding to the red circles in **Figure 5.19** (1-min light pulses every hour) by showing the distributions and the medians of all the fluorophores for both the differentiated and non-differentiate populations (**Figure 5.20**).

Observing the levels of EL222 (mVenus fluorescence) in differentiated and non-differentiated populations (**Figure 5.20, mVenus**), we notice a split in the levels consistent with the selection effect observed earlier (**Figure 5.8, mVenus**). Moreover, in accordance with model predictions, non differentiated fraction shows decreased level of mVenus fluorescence that are maintained at lower levels for the duration of the experiment and the differentiated fraction show elevated levels of EL222 immediately following induction that relax back to the population median due to the stochastic process governing plasmid dynamics (**Figure 6.20, mVenus**).

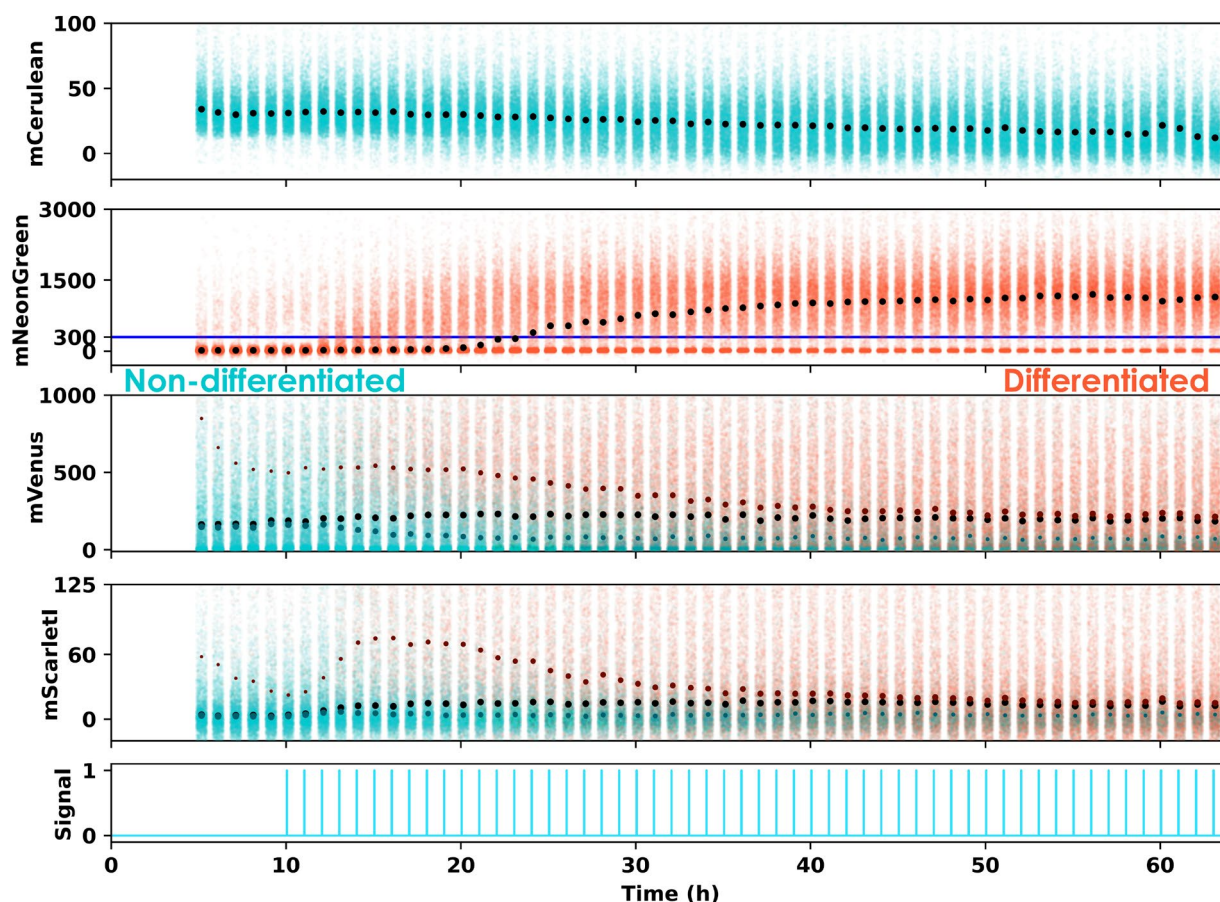


Figure 5.20. Distributions of the different fluorescent reporters for the 2-micron tetra reporter induced with repeated pulses. Data is represented as in Figure 5.2. Median EL222 in the non-differentiated cells could be maintained below the total median of the population for 50h compare small teal circles with large black circles in mVenus channel. The data corresponds to red circles in Figure 5.19 (1-min pulses every hour).

Interestingly, the median levels of EL222 in the differentiated fraction could also be maintained at a higher level compared to the total population (Figure 6.20, mVenus). This was also reflected in the readout of EL222 activity (Figure 6.20, mScarlet-I).

We queried the model with the light sequence employed here and found that model predictions were in qualitative agreement with the data (Figure 5.21).

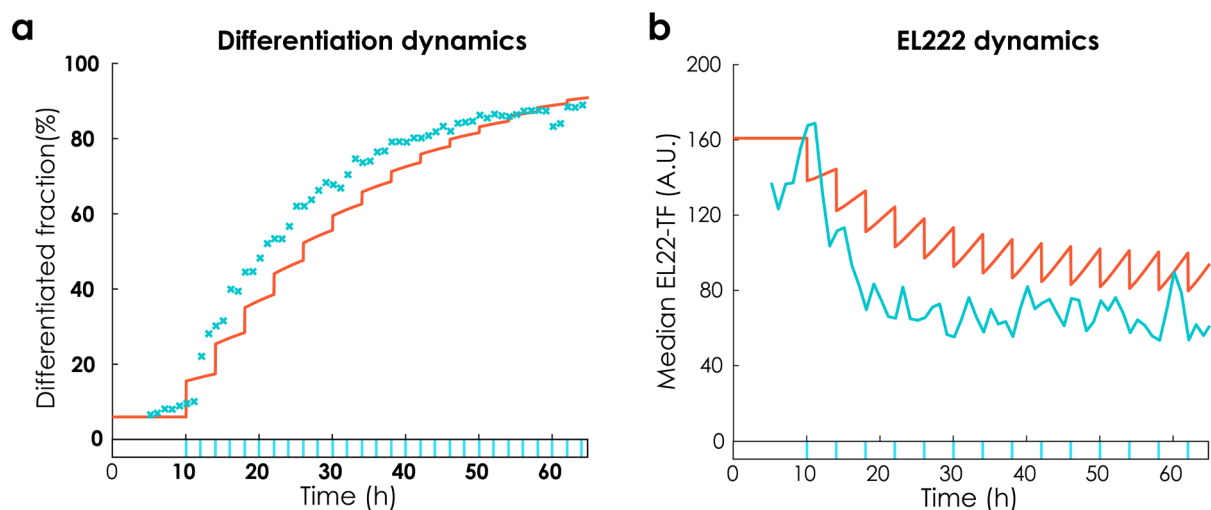


Figure 5.21. Model predictions for optogenetic control of constitutive gene expression a. Regularly repeated short light pulses lead to a slow steady increase in the differentiated population fraction according to both model (red) and data (blue). b. Over longer time periods, repeated selective differentiation and reversion to the mean balance each other, which leads to approximately constant median EL222 levels and quasi-stationary EL222 distributions (Figure 5.22, mVenus).

The results were only in qualitative agreement with the data because we use very short light signals in these experiments and these were not used for constructing the model. Admittedly, there are some non-linearities in the system with respect to the light duration that the model is unable to capture. These could be ameliorated if, mechanistically, EL222 did not directly lead to differentiation but rather produced Cre in the presence of light. The differentiation rate would then be defined as a function of Cre and not of EL222. We did construct one such model in the past and this allowed us to predict the response to different light pulses quantitatively for the integrated tetra reporter, however, this approach became computationally untractable once we

introduced plasmids as an additional species. Therefore, we restricted ourselves to operate in the range of long light pulses for the better part of the chapter.

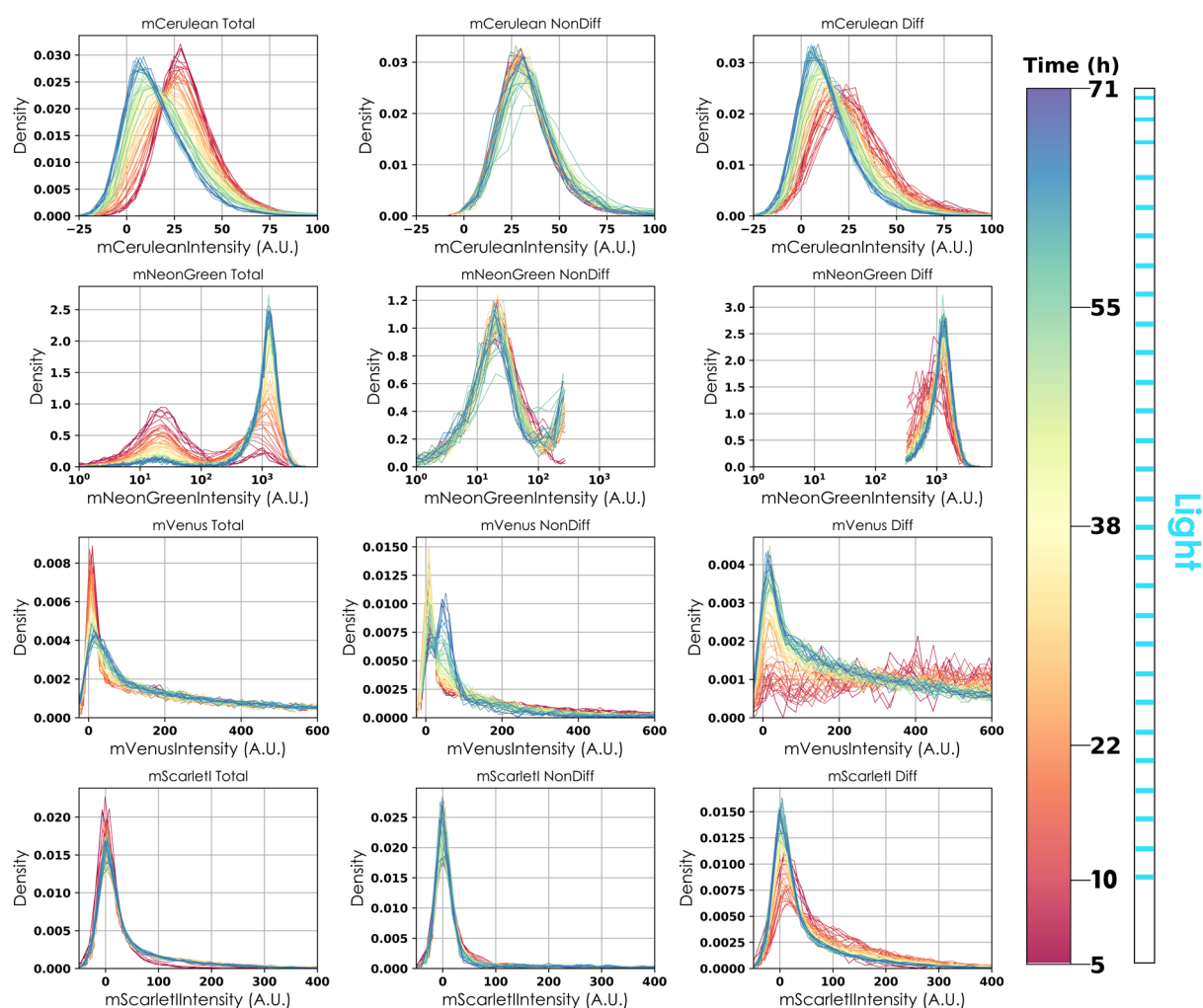


Figure 5.22. Fluorescence distributions. Data is represented as in Figure 5.3. The non-differentiated fraction starts becoming enriched in a population that seemingly possesses 1 copy of the plasmid around 38 h (*mVenus*, *NonDiff*).

Although it is difficult to quantify formally, **Figure 5.22** shows that the distributions of EL222 in the non-differentiated fraction conform to a quasi-steady state distribution for about 30h post induction (*mVenus* *Nondiff* panel, compare yellow with green histograms). Past this point, we seemingly observe an enrichment of the population with one plasmid, in favor of the population with zero copies of the plasmid that is predicted by the model, exclusively in the non-differentiated population (*mVenus* *Diff* panel, the yellow green and blue histograms do not show a marked difference, compare with the corresponding panels in **Figure 5.9**). This enrichment continues to increase as the total number of non-differentiated cells decrease suggesting that this could be an aberrant population. It is difficult to conclude whether the one copy population

is an artefact of the experimental setup or represents a true emergence of a 1 copy population. The case against it being an artefact is as follows: we see this population emerge in every experiment that features repeated pulses and this population is not present in the differentiated population.

The evidence for it being an artefact is exemplified by the fact that this population is also present in some experiments with non-repetitive induction, albeit to a lesser extent. From a theoretical perspective, there is no reason for the selection to change during the experiment and consequently the model does not predict such a population. Accordingly, no mechanism of regulation of 2-micron plasmids has been reported that changes over time and does so specifically in a subpopulation.

However, if true, this result indicates that perhaps there is a long-term non-trivial process at play here, which is not captured by the simple birth and death model of plasmid dynamics. It could well be a regulatory mechanism, like in the case of the centromeric plasmids, acting in a small subset of the population that forbids us from predicting the outcome over long time periods. It would make evolutionary sense to have such a population. Alternatively, the light could somehow prevent plasmid loss in the non-differentiated population because of some effect at the DNA level due to bound transcription factors.

In conclusion, notwithstanding the discrepancies between the data and theory (which may or may not be an experimental artefact), understanding and characterizing the consequences of cell-to-cell variability allowed us to use light to dynamically regulate the (sub-) population copy numbers of the plasmid from which the optogenetic system is expressed, a feat that seems quite counter-intuitive at a first glance and that is not realizable in any obvious way by other means.

Discussion

Quantitatively predicting the dynamics of complex synthetic circuits before the circuit is constructed is possibly the key challenge that needs to be mastered to turn synthetic biology into a true engineering discipline. Yet, while our capacity to construct complex circuits is continuously increasing, our capability to predict circuit functionality from supposedly known and characterized circuit components remains, at best, limited to very tightly constrained operating conditions and qualitative and/or stationary outputs (Nielsen et al. 2016). The concrete reasons for the failure of model predictions may be manifold and caused by unexpected component-to-component interactions (e.g. retroactivity (Del Vecchio et al. 2018) or resource competition (Qian et al. 2017)) or couplings of the circuit to processes of the host (e.g. burden (Miliadis-Argeitis et al. 2011) or saturation of the host's degradation machinery (Potvin-Trottier et al. 2016)).

Eventually, however, a common theme unites these physiologically diverse problems: our incapacity to foresee the consequences of complex inter-dependencies of a circuit in vivo.

We exemplified this problem for the case of our optogenetic differentiation system by showing that predictions obtained from a simple deterministic model break down as soon as the context in which the system is used is modified (Figure 2). More precisely, we created a version of the differentiation system that made it possible to observe levels of the optogenetic transcription factor (EL222) and its activity. We characterized and developed a model of the system that was able to predict the population dynamics. We changed the noise structure of EL222 by expressing it from a centromeric and 2-micron plasmid respectively. Expression from plasmids could not be predicted from the simple model at all, as it led to unforeseen couplings of plasmid copy number fluctuations, growth in selective media, and selective differentiation. Notably, flow cytometry measurements revealed that EL222 distributions were highly variable with around 30% of the population lacking cellular fluorescence seemingly suggesting an absence of plasmids. This was confirmed by a colony-counting assay for both centromeric and 2-micron plasmids. This result seemingly contradicted the finding that under continuous light, the entire population differentiates for both plasmid versions. In the same vein, we observed that the differentiated fraction continued to increase in the culture 10h after removal of light stimulation. Both these results were unintuitive so we analyzed the EL222 levels in the population and found that differentiated cells had more EL222 than non-differentiated cells on average. Understandably, cells with higher amount of EL222 proteins were enriched in the differentiated fraction. Presumably, these cells also had more plasmids. Assuming that cells that have lost the plasmid do not grow (or grow slower) than the cells that have at least one plasmid, we came to an understanding of what was going on. Simply put, the cells that do not differentiate upon induction are composed, in great part, of cells that do not possess any plasmids and therefore are incapable of differentiating. However, since cells are continuously cultured in selective media, the population that does not contain the plasmid will be diluted out of the culture at a faster rate than the differentiated cells, leading to seemingly slow “differentiation dynamics” even in the absence of light (Figure 2d)

We thus developed dedicated single-cell models of both the differentiation system and plasmid dynamics (Figure 3). We found that the composed model of the plasmid-based differentiation system, obtained after merging the two individual models, does not only resolve the previously not understood features but quantitatively predicts the consequences of complex component interactions without any adjustment of model parameters (Figure 4). This is a particularly encouraging finding as it demonstrates that, at least for the differentiation system, our incapacity to foresee the consequences of complex inter-dependencies of synthetic circuits and couplings of

single-cell and population processes can be remedied by appropriate characterization of circuit components. Importantly, with the ability to foresee consequences comes the possibility to exploit such couplings in the design.

On modelling

In the case of our differentiation system on plasmid, complex couplings between single cell processes and population level dynamics necessitated an approach that could capture dynamics of population distributions. The key notion was to derive a non-linear version of the CME that (could be efficiently solved and) kept track of population state along with single-cell dynamics that are typically tracked in stochastic models. We posit that similar approaches would be required to faithfully capture the dynamics of any system where couplings of single-cell and population processes are likely to be at play (Ruess et al. 2019; Shahrezaei and Marguerat 2015; Tan, Marguet, and You 2009). Such couplings notably manifest in the form of growth rate decrease at the population scale due to heterologous expression of either large quantities of (physiologically toxic but commercially valuable) proteins as is the case for bioproduction or convoluted genetic circuits like in synthetic biology. However, similar couplings between scales are likely also present in many natural systems such as selective killing of cancer cells with particular states of internal processes in response to treatments that induce the apoptotic pathway (Bertaux et al. 2014), differential responses of bacteria to antibiotic treatments (Wakamoto et al. 2013) to name a few possible examples. It needs to be noted, however, that for cases where the single-cell model is more high-dimensional than the fairly small models considered here, tracking the entire solution of the corresponding master equation becomes computationally infeasible. Further work is thus necessary to develop and test approaches for approximately calculating with multi-scale stochastic models (Lunz et al. 2020; Duso and Zechner 2020).

On modelling of the centromeric tetra reporter

We found that the differentiation dynamics of the centromeric tetra reporter could not be predicted by any model discussed here. This, primarily, is a consequence of the inability to correctly model plasmid dynamics for the centromeric plasmid with simple birth and death model used for the 2-micron plasmid. Concretely, a significant part of the population appears to contain one plasmid in stationary conditions in addition to a zero plasmid population. The existence of the population with one plasmid indicates some regulatory mechanism, which is not included in our model of plasmid dynamics. Absence of this mechanism would lead to a disagreement of the quasi-stationary TF distributions between data and model predictions and consequently the

model fails to provide accurate predictions for the differentiation dynamics. This one plasmid population was also observed by Gnügge, Liphardt and Rudolf (2015) and presumably suggests a tighter mechanism of plasmid copy number regulation for the centromeric plasmids. The authors provided a predictive model for the dynamics of plasmid copy numbers that showed quantitative agreement with the data, which theoretically could have been employed here. However, their model featured parameters that could not be determined readily from the data at hand.

On plasmid copy number control

As predicted by theory, we found that repeatedly stimulating the culture with short light pulses led to stable maintenance of constitutive median fluorescence (proxy for plasmid copy numbers) in the non-differentiated fractions lower than that of the total population. Furthermore, we found that by tuning the light duration and intensity, it was possible to modulate the level of this decrease. The fact that constitutive expression can be modulated by using the plasmid tetra reporter is as counterintuitive as it is surprising. However, we note that the model cannot accurately predict the extent to which the EL222 levels drop in the non-differentiated population. Looking at the mVenus fluorescence in the cytometry data provided a clue. From the start of the induction to about 25-30h the cells behave predictably, the EL222 levels drop and the non-differentiated cells conform to a stationary distribution. However, around 30-34 hours post induction, we start observing a population with fluorescence levels comparable to those obtained for the integrated strain. It can be hypothesized that this population possesses one plasmid. Surprisingly, this peak is the strongest for the longest light input. We hypothesize that this population emerges due to selection on both fronts, cells that lose the plasmid are removed from the population and cells in which plasmid copy number increases are differentiated, thus leading to an enrichment of plasmids with one copy in the non-differentiated fraction. This hypothesis is further strengthened by the finding that we observe no such enrichment of the one plasmid population in differentiated cells. In light of this finding, we note that further characterization of the dynamics of 2-micron plasmids is required. Theoretician in the study, Jakob Ruess, holds that the deviations from theory arise out of experimental artefacts like media composition differences. However, we cannot discount another hypothesis that explains the emergence of the 1 plasmid population, that is, evolution of the plasmid to persist better in the yeast. Several studies have reported emergence of mutations (either in host chromosome or plasmid sequence or both) that stabilize the presence of plasmids in bacteria under continuous cultivation conditions (Bouma and Lenski, 1988; Modi *et al.*, 1992; Dahlberg and Chao, 2003; De Gelder *et al.*, 2008; Harrison *et al.*, 2015). A rather unintuitive mechanistic explanation was provided by Yano *et al.* (2016) who reported that a part of the burden of carrying the plasmid came from saturating

the replication machinery of the cell. The authors deduced this from the observation that the cultures which showed an increase in plasmid copy numbers carried plasmids with mutations that *decreased* the affinity of plasmid *ori* sequences to bind with the replication machinery (Yano et al. 2016). Such an adaptation over continuous cultivation has not been reported for 2 μ plasmids in yeast to date.

On plasmid copy numbers

Average plasmid copy numbers of cells carrying plasmids similar to the ones used in the study have been reported in the past via bulk measurement methods like DNA extraction of the whole population followed by qPCR (Gnügge, Liphardt, and Rudolf 2016; Karim, Curran, and Alper 2013; Fang et al. 2011). There is a lot of variability in the numbers reported in different studies which is understandable given each study used different plasmid backbones (none of which exactly resemble the plasmid backbone used in the present study) but all studies, when using other approaches to estimate plasmid copy number, remarked the strong dependence of observed copy number on the reporter and the promoter used to drive its expression. Notably in Karim, Curran and Alper (2013), authors reported a 25% decrease in plasmid copy number upon switching from a weak promoter to a medium strength promoter whereas in Fang et al. (2011), authors saw an 8 fold increase in expression for 2-micron plasmids (over low-copy centromeric plasmids (1-1.5 copies on average)) with a weak promoter, which dropped to only a 3 fold increase, when a strong promoter was used to drive the expression. These studies while insinuating that the average plasmid copy numbers are hugely variable, however, do not comment on the distribution of plasmid copy numbers in the population. In Gnügge, Liphardt and Rudolf (2016), an approach with two constitutively expressed fluorescence reporters, one integrated in the genome while the other destabilized and expressed from a plasmid, was employed to indirectly estimate the distribution of 2-micron plasmid copy numbers in the population via flow cytometry. They observed a significant part of the population lacked plasmids (~45%). The fluorescence distribution of the destabilized constitutively expressed fluorescent reporter resembled the mVenus fluorescence distribution of the plasmid version of the tetra reporter granted that the peaks were not as sharply defined in our data. We remark that other studies have reported much larger proportions of cells lacking the 2-micron plasmids. Notably, Wittrup et al. (1994) (overexpression of secreted proteins), and Yoo and O'Malley (2018) (overexpression of membrane proteins) reported loss of plasmid in >90% of the population. Both secretion and membrane proteins are sources of greater burden for the native protein folding machinery in yeast than constitutive expression of fluorescent proteins explored

in Gnügge, Liphardt and Rudolf (2016) and this work. We also note that the plasmid backbones used in Gnügge, Liphardt and Rudolf (2016) were closest to the one used in the present study.

A finer understanding of the plasmid copy number dynamics in yeast could be provided if the approach of Gnügge, Liphardt and Rudolf (2016) is coupled with light inducible expression and regular flow cytometry measurements in selective and non-selective media. Concretely, a fluorescently tagged EL222 transcription factor from an integrated copy in the genome could be used to drive expression from a 2-micron plasmid of a strong fluorescent protein (mScarlet-I) under the control of a tight EL222 promoter tagged with degron tags (Lee et al. 2015) and/or destabilized mRNAs (Kafri et al. 2016) of different strengths. Another strong fluorescent protein (mNeonGreen) with the same degron/destabilized mRNA could be integrated in the genome under the control of the same EL222 promoter. The inducible nature of the strain will provide a handle on the burden imposed by the plasmids while the integrated copy would be used as a reference for the single copy levels as well as a way to gauge the activity of EL222 TF. Moreover, since EL222 itself is tagged, any decrease in the capacity of cells to produce will be reflected in the EL222 levels. At the same time, the destabilized mRNAs and protein degron tags could yield quantitative estimates of the plasmid replication dynamics. Furthermore, such a strain could easily be adapted to study the mechanism of transcript specific buffering observed in Chapter 4. We note that this approach may not yield any direct evidence for plasmid copy number distributions unless cell-sorting techniques are employed to separate the population on the basis of cellular fluorescence followed by sub-population specific qPCR. These measurements could then be compared with those predicted by a model similar to the one reported here.

However, the granularity of such an approach might still be coarse owing to technical and/ or biological limitations. The resolution could be further enhanced by adapting a recently published microscopy based single cell approach (Shao et al. 2021). In the study, the authors directly estimate plasmid copy numbers in bacterial populations by using a tagged DNA binding protein, which interacts specifically with the plasmid origin of replication, and quantifying the fluorescence intensity. The authors also found a persistent population of cells without plasmids (1-3%). Moreover, the authors reported results on promoter activity by observing transcript copy numbers (via quantification of the intensity of a fluorescently tagged PP7 RNA binding protein that specifically binds the transcript in question) and correlating them with the fluorescence intensity of the fluorescent protein coded by the transcript. The authors found that, while plasmid copy number is somewhat correlated with number of transcripts, the protein levels are only poorly correlated with the number of transcripts. Even though the authors did not explore the question of the burden of heterologous expression, they concluded that individual promoter

activity decreases as the number of copies increased, possibly due to resource bottlenecks. This is best exemplified with their finding that a 100-fold higher copy number translates only in a 4-fold increase in the quantity of expressed protein. I note that these results have been obtained in bacteria with plasmids that confer resistance to antibiotics, which are known to be more tightly regulated than 2-micron plasmids in yeast. Their results show that transcription and translation increase nonlinearly with copy number, however, it is well known that individual *E. coli* cells have a much lower capacity for heterologous expression than *S. cerevisiae* and therefore the same results may not stand for yeast (Kafri et al. 2016). Nevertheless, we conclude that such a system (perhaps with a light inducible reporter for gene expression instead of a constitutive one) is worth developing for the 2-micron plasmids in *S. cerevisiae*. Long-term live cell imaging of the hypothetical system in selective and non-selective media would complement the cytometry based approach and might provide definitive answers to the mysterious distribution of 2-micron plasmid copy numbers and their regulation in yeast at different levels of burden.

References

- Aditya, Chetan, François Bertaux, Gregory Batt, and Jakob Ruess. 2021. "A Light Tunable Differentiation System for the Creation and Control of Consortia in Yeast." *BioRxiv*.
- Bertaux, François, Szymon Stoma, Dirk Drasdo, and Gregory Batt. 2014. "Modeling Dynamics of Cell-to-Cell Variability in TRAIL-Induced Apoptosis Explains Fractional Killing and Predicts Reversible Resistance." *PLoS Computational Biology* 10 (10): e1003893.
- Bouma, Judith E, and Richard E Lenski. 1988. "Evolution of a Bacteria/Plasmid Association." *Nature* 335 (6188): 351–52.
- Dahlberg; Chao. 2003. "Amelioration of the Cost of Conjugative Plasmid Carriage in Escherichia Coli K12." *Genetics* 165 (4): 1641–1649.
- Duso, Lorenzo, and Christoph Zechner. 2020. "Stochastic Reaction Networks in Dynamic Compartment Populations." *Proceedings of the National Academy of Sciences* 117 (37): 22674–83.
- Elowitz, Michael B, Arnold J Levine, Eric D Siggia, and Peter S Swain. 2002. "Stochastic Gene Expression in a Single Cell." *Science* 297 (5584): 1183–86.
- Fang, Fang, Kirsty Salmon, Michael W Y Shen, Kimberly A Aeling, Elaine Ito, Becky Irwin, Uyen Phuong C Tran, G Wesley Hatfield, Nancy A Da Silva, and Suzanne Sandmeyer. 2011. "A Vector Set for Systematic Metabolic Engineering in *Saccharomyces Cerevisiae*." *Yeast* 28 (2): 123–36.
- Friedman, Nir, Long Cai, and X Sunney Xie. 2006. "Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression." *Physical Review Letters* 97 (16): 168302.
- Gelder, Leen De, Julia J Williams, José M Ponciano, Masahiro Sota, and Eva M Top. 2008. "Adaptive Plasmid Evolution Results in Host-Range Expansion of a Broad-Host-Range Plasmid." *Genetics* 178 (4): 2179–90.
- Gillespie, Daniel T. 1992. "A Rigorous Derivation of the Chemical Master Equation." *Physica A: Statistical Mechanics and Its Applications* 188 (1–3): 404–25.
- Gnügge, Robert, Thomas Liphardt, and Fabian Rudolf. 2016. "A Shuttle Vector Series for Precise Genetic Engineering of *Saccharomyces Cerevisiae*." *Yeast* 33 (3): 83–98.
- Harrison, E, D Guymier, A J Spiers, S Paterson - Current Biology, and undefined. 2015. "Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum." *Current Biology* 25 (15): 2034–39.
- Huh, Dann, and Johan Paulsson. 2011. "Non-Genetic Heterogeneity from Stochastic Partitioning at Cell Division." *Nature Genetics* 43 (2): 95–100.
- Kafri, Moshe, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. 2016. "The Cost of Protein Production." *Cell Reports* 14 (1): 22–31.
- Karim, Ashty S, Kathleen A Curran, and Hal S Alper. 2013. "Characterization of Plasmid Burden and Copy Number in *Saccharomyces Cerevisiae* for Optimization of Metabolic Engineering Applications." *FEMS Yeast Research* 13 (1): 107–16.
- Kwok, Roberta. 2010. "Five Hard Truths for Synthetic Biology." *Nature News* 463 (7279): 288–90.
- Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. "A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly." *ACS Synthetic Biology* 4 (9): 975–86.
- Lugagne, Jean-Baptiste, Sebastián Sosa Carrillo, Melanie Kirch, Agnes Köhler, Gregory Batt, and Pascal Hersen. 2017. "Balancing a Genetic Toggle Switch by Real-Time Feedback Control and Periodic Forcing." *Nature Communications* 8 (1): 1–8.
- Lunz, Davin, Gregory Batt, Jakob Ruess, and Joseph Frédéric Bonnans. 2020. "Beyond the Chemical Master

Equation: Stochastic Chemical Kinetics Coupled with Auxiliary Processes.”

- Milias-Argeitis, Andreas, Marc Rullan, Stephanie K. Aoki, Peter Buchmann, and Mustafa Khammash. 2016. “Automated Optogenetic Feedback Control for Precise and Robust Regulation of Gene Expression and Cell Growth.” *Nature Communications* 7 (1): 1–11.
- Milias-Argeitis, Andreas, Sean Summers, Jacob Stewart-Ornstein, Ignacio Zuleta, David Pincus, Hana El-Samad, Mustafa Khammash, and John Lygeros. 2011. “In Silico Feedback for in Vivo Regulation of a Gene Expression Circuit.” *Nature Biotechnology* 29 (12): 1114–16.
- Modi, R I, L H Castilla, S Puskas-Rozsa, R B Helling - Genetics, and undefined 1992. 1992. “Genetic Changes Accompanying Increased Fitness in Evolving Populations of Escherichia Coli.” *Genetics* 130 (2): 241–249.
- Neuert, Gregor, Brian Munsky, Rui Zhen Tan, Leonid Teytelman, Mustafa Khammash, and Alexander Van Oudenaarden. 2013. “Systematic Identification of Signal-Activated Stochastic Gene Regulation.” *Science* 339 (6119): 584–87.
- Nielsen, Alec A K, Bryan S Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A Strychalski, David Ross, Douglas Densmore, and Christopher A Voigt. 2016. “Genetic Circuit Design Automation.” *Science* 352 (6281).
- Paulsson, Johan, and Måns Ehrenberg. 2001. “Noise in a Minimal Regulatory Network: Plasmid Copy Number Control.” *Quarterly Reviews of Biophysics* 34 (1): 1–59.
- Potvin-Trottier, Laurent, Nathan D Lord, Glenn Vinnicombe, and Johan Paulsson. 2016. “Synchronous Long-Term Oscillations in a Synthetic Gene Circuit.” *Nature* 538 (7626): 514–17.
- Qian, Yili, Hsin-Ho Huang, José I Jiménez, and Domitilla Del Vecchio. 2017. “Resource Competition Shapes the Response of Genetic Circuits.” *ACS Synthetic Biology* 6 (7): 1263–72.
- Raser, Jonathan M, and Erin K O’shea. 2005. “Noise in Gene Expression: Origins, Consequences, and Control.” *Science* 309 (5743): 2010–13.
- Ruess, Jakob, Maroš Pleška, Čálin C Guet, and Gašper Tkačik. 2019. “Molecular Noise of Innate Immunity Shapes Bacteria-Phage Ecologies.” *PLoS Computational Biology* 15 (7): e1007168.
- Shahrezaei, Vahid, and Samuel Marguerat. 2015. “Connecting Growth with Gene Expression: Of Noise and Numbers.” *Current Opinion in Microbiology* 25: 127–35.
- Shahrezaei, Vahid, and Peter S Swain. 2008. “Analytical Distributions for Stochastic Gene Expression.” *Proceedings of the National Academy of Sciences* 105 (45): 17256–61.
- Shao, Bin, Jayan Rammohan, Daniel A Anderson, Nina Alperovich, David Ross, and Christopher A Voigt. 2021. “Single-Cell Measurement of Plasmid Copy Number and Promoter Activity.” *Nature Communications* 12 (1): 1–9.
- Tan, Cheemeng, Philippe Marguet, and Lingchong You. 2009. “Emergent Bistability by a Growth-Modulating Positive Feedback Circuit.” *Nature Chemical Biology* 5 (11): 842–48.
- Vecchio, Domitilla Del, Yili Qian, Richard M Murray, and Eduardo D Sontag. 2018. “Future Systems and Control Research in Synthetic Biology.” *Annual Reviews in Control* 45: 5–17.
- Wakamoto, Yuichi, Neeraj Dhar, Remy Chait, Katrin Schneider, François Signorino-Gelo, Stanislas Leibler, and John D McKinney. 2013. “Dynamic Persistence of Antibiotic-Stressed Mycobacteria.” *Science* 339 (6115): 91–95.
- Weiß, Andrea Y, Diego A Oyarzún, Vincent Danos, and Peter S Swain. 2015. “Mechanistic Links between Cellular Trade-Offs, Gene Expression, and Growth.” *Proceedings of the National Academy of Sciences* 112 (9): E1038–E1047.
- Wittrup, K D, A S Robinson, R N Parekh, and K J Forrester. 1994. “Existence of an Optimum Expression Level for Secretion of Foreign Proteins in Yeast A.” *Annals of the New York Academy of Sciences* 745 (1): 321–30.

- Yano, Hirokazu, Katarzyna Wegrzyn, Wesley Loftie-Eaton, Jenny Johnson, Gail E Deckert, Linda M Rogers, Igor Konieczny, and Eva M Top. 2016. "Evolved Plasmid-Host Interactions Reduce Plasmid Interference Cost." *Molecular Microbiology* 101 (5): 743–56.
- Yoo, Justin I, and Michelle A O'Malley. 2018. "Tuning Vector Stability and Integration Frequency Elevates Functional GPCR Production and Homogeneity in *Saccharomyces Cerevisiae*." *ACS Synthetic Biology* 7 (7): 1763–72.
- Zechner, Christoph, Jakob Ruess, Peter Krenn, Serge Pelet, Matthias Peter, John Lygeros, and Heinz Koepl. 2012. "Moment-Based Inference Predicts Bimodality in Transient Gene Expression." *Proceedings of the National Academy of Sciences* 109 (21): 8340–45.

Chapter 6

MATERIAL AND METHODS

“Perhaps I am doomed to retrace my steps under the illusion that I am exploring, doomed to try and learn what I simply should recognize..”
André Breton

Preface

Portions of this chapter as well as some figures were previously published as Aditya et al. (2021).

Constructions and cloning

I used Golden Gate cloning (Engler, Kandzia, and Marillonnet 2008) for the constructions used in the experiments described here. Golden Gate is a modular cloning approach that facilitates the generation of new constructs from a library of parts. The underlying principle, is the use of Type IIs restriction enzymes that do not cut the DNA sequence they recognize. This can be used to generate families of DNA sequence with overhangs that, after digestion with a type IIs restriction enzyme, are distinct and complementary to each other. In this regard, the work of (Lee et al. 2015) was instrumental to our research. Briefly, they established a standardized approach to modular cloning in yeast. They designed a library of parts called Yeast Tool Kit (YTK) with specified overhangs depending on the function of the part in the construction (promoters, terminators, coding sequences with the possibility of adding tags on either side). This optimized library was available from AddGene and was used profusely in the current work. Notably, all the backbones used in the study were generated using the YTK. These included three integrative vectors carrying the auxotrophy markers URA3, LEU2, and HIS3. The integrative vectors carried the homology region targeting the endogenous locus of the respective auxotrophy marker except for HIS3 backbone, which targeted the HO locus. In addition, I generated the centromeric and episomal vectors for each of the auxotrophy markers mentioned above. These backbones along with their sequences can be found in the plasmids table. The cloning process to go from DNA to plasmids functional in yeast needed three steps.

- A level 0 assembly to insert the part sequence with overhangs (synthesized or obtained via Phusion PCR) in a standard entry vector. Level 0 assembly required BsmBI as the restriction enzyme (**Figure 6.1**).

- A level 1 assembly to construct an expression cassette composed of Level 0 parts (notably a promoter, a coding sequence (CDS) and a terminator) and constitutes a transcriptional unit in *S. cerevisiae*. However, at this stage, this transcriptional unit is present in a backbone containing only a bacterial origin of replication and hence cannot function in *S. cerevisiae*. Level 1 assembly required BsaI as the restriction enzyme (**Figure 6.1**).
- A level 2 assembly can then be used to introduce the transcriptional unit(s) in yeast vectors, containing either chromosomal homology regions flanking the site of insertion of the transcriptional unit insert (integrative vectors) or containing a yeast origin of replication (CEN6/ARS4 (centromeric) or 2micron(episomal)). In this fashion, up to 3 level 1 transcriptional units were cloned in a single vector backbone. Level 2 assembly required the use of BsmBI restriction enzyme. Level 2 integrative plasmids contained a NotI digestion site that was used to linearize the integrative plasmids prior to transformation in yeast (**Figure 6.1**).

Overview of cloning pipeline

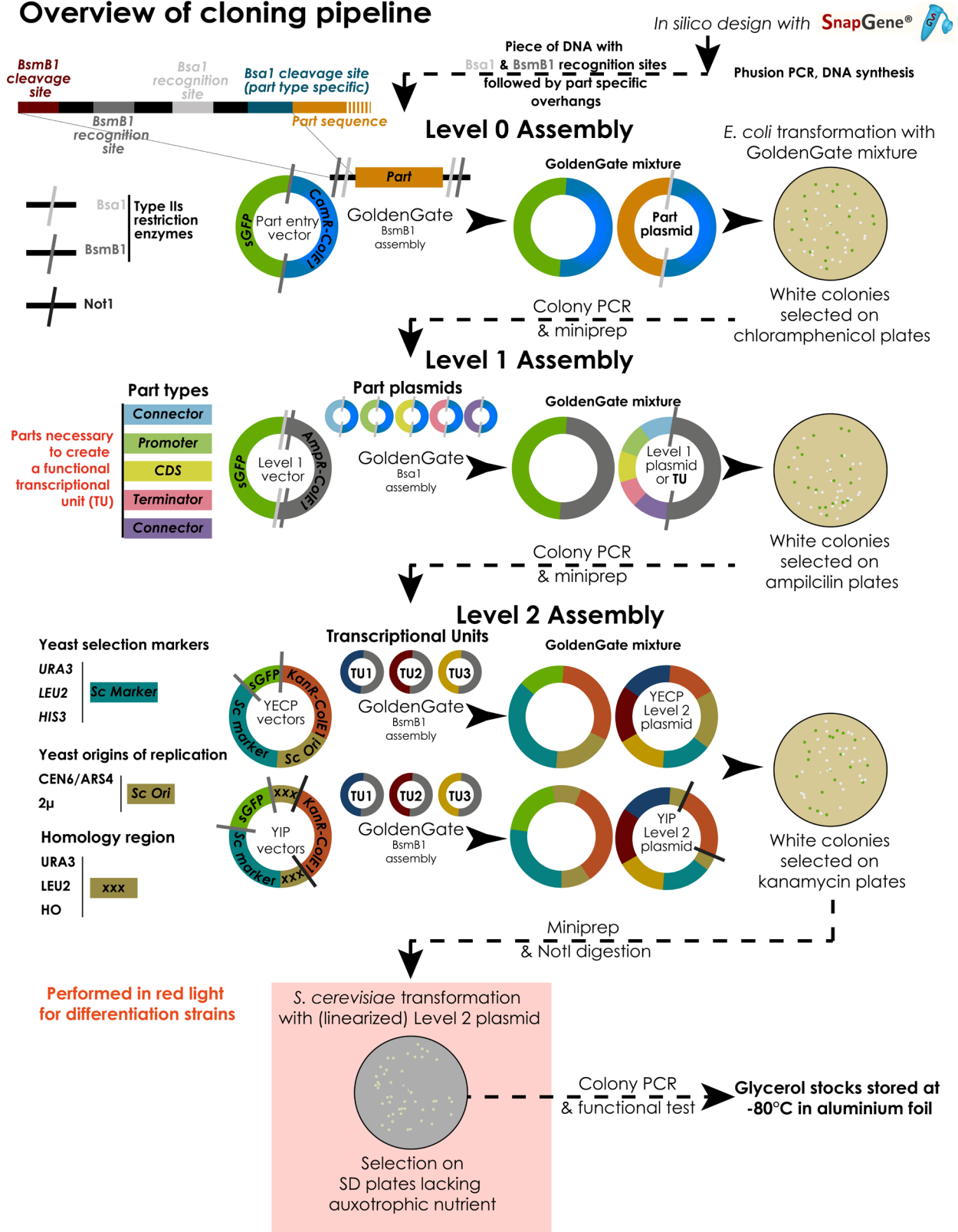


Figure 6.1. Overview of the cloning pipeline. From DNA to strains. If all steps work the first time, it takes 20 days to go from DNA parts to yeast strains. This time is reduced to 12 days if one comes to the lab on the weekends.

New parts, whenever necessary, were generated in the laboratory (DNA synthesis, Phusion PCR). To design new parts and visualize cloning products, I used SnapGene, which allows for simulation of cloning reactions given the sequence of the parts used.

After spending some time on optimizing the protocol for the Golden Gate, I fixed on the following program and concentrations. Bulk of cloning for both level 1 and Level 2 plasmids was done with this optimized protocol.

Table 6.1. Golden Gate Reaction	
Reactant	Volume/sample (μL)
Water	5.6-X
T4 ligase buffer	1
BSA 1X	1
Bsal	1
T4 ligase buffer	0.4
Backbone @ conc. 20fmol	1
Parts @ conc. of (20-40 fmol/ μL)	1 each = X
Total	10

Table 6.2. Golden Gate Program						
37°C	37°C	16°C	37°C	50°C	80°C	4°C
15 mins	2 mins	5 mins	15 mins	5 mins	5 mins	∞
50 cycles						

The Golden Gate mixture was transformed via heat shock transformation in thermocompetent *E. coli* cells and plasmids were isolated using standard miniprep kits (Macherey & Nagel, and Qiagen). Level 1 and Level 2 Plasmids were verified by colony PCR or restriction digestion with NotI enzyme. Only Level 2 plasmids were sequenced.

Yeast strain manipulation and storage

All strains used in this study are derived from BY4741 [MATa his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0]. Cells were transformed with linearized integrative vectors or using standard Lithium Acetate transformation (Gietz and Woods 2002). For selection, common auxotrophic markers, Uracil,

Leucine, and Histidine were used. Integrative plasmids carrying LEU2 and URA3 selection markers were integrated at the endogenous loci while those with HIS3 were integrated at HO locus. Cells were grown in standard defined media (Sigma Aldrich Yeast Nitrogen Base) containing 2% glucose and lacking the respective auxotrophic nutrient during selection in plates (Sigma Aldrich Uracil, Leucine and, Histidine drop-out media supplements). A list of strains used in this study and their genotypes can be found in the strain list (**Annex 6**).

To subdue background activity arising from extraneous exposure to light, I performed all cell manipulations in red light (**Figure 6.2**) including transformation and the creation of stocks for the light sensitive strains. Stocks were covered with aluminium foil during storage. Revivals were performed in red light and yeast plates were encased in aluminium foil during growth in the incubator. Overnight cultures were grown in the dark using a falcon holder covered on all sides.



Figure 6.2. Working in red light. Photo courtesy Virgile Andreani.

Overnight cultures were measured using cytometry prior to the start of preculture to ensure maximum non-differentiated cells. Typically, I observed 0.03%-1% differentiation at this point (for GAuDi I always found 10%-20% background differentiation but as noted in Chapter 4, I believe these cells were present due to leaky expression of ATAF1 TF and not background recombination). Some picked colonies were completely recombined and not used for experiments. I observed that streaked plates more than a week old led to higher levels of

background differentiation in individual colonies. I note that the colonies were extremely sensitive to extraneous light and even opening them in red light led to significant increase in differentiated fraction over the next 96h (5%-10%). Such susceptibility to extraneous light was not observed for liquid cultures. Consequently, only freshly streaked plates were used for the experiments.

Batch culture

From freshly streaked plates, single colonies were picked to start the overnight cultures on the day before experiment. Cells were grown to exponential phase in the dark from ON culture on the day of the experiment in Falcon tubes shaking at 200 rpm at 30°C before starting the experiment. Induction was carried out in the custom Falcon tube holder fitted with LED strips. Sampling for cytometry was performed manually in presence of red light.

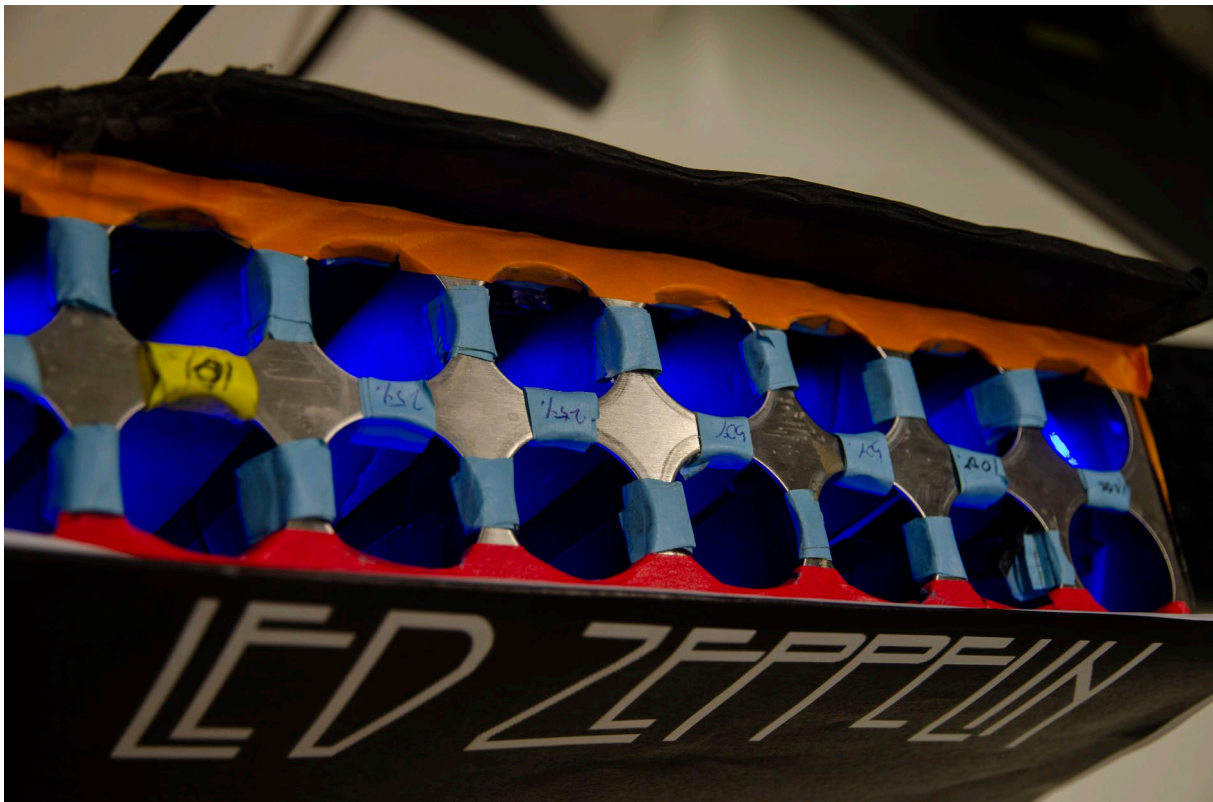


Figure 6.3. Batch incubator dubbed Led Zeppelin. Batch induction was carried out in this contraption as well as cultures in the dark (lid not included in the photograph). Each slot was insulated from other slots. Light signal could be independently controlled for each slot.

Experimental setup for continuous cultures

I used our previously described turbidostat platform (Bertaux et al. 2020) to continuously culture cells and conduct time course experiments. The platform allows the user to monitor 16 cultures in parallel with regular OD measurements and maintain them at a target cell density. The OD

measurements and dilution data were used to estimate the growth rate. Further, culture vessels are equipped with LEDs and can be stimulated independently with light. Samples from the vessels are collected in a 96-well plate and, with the help of a pipetting robot, loaded into the cytometer. The cytometer acquisition is controlled with the help of click and point software. **Figure 6.4** gives a general overview of the experimental platform.

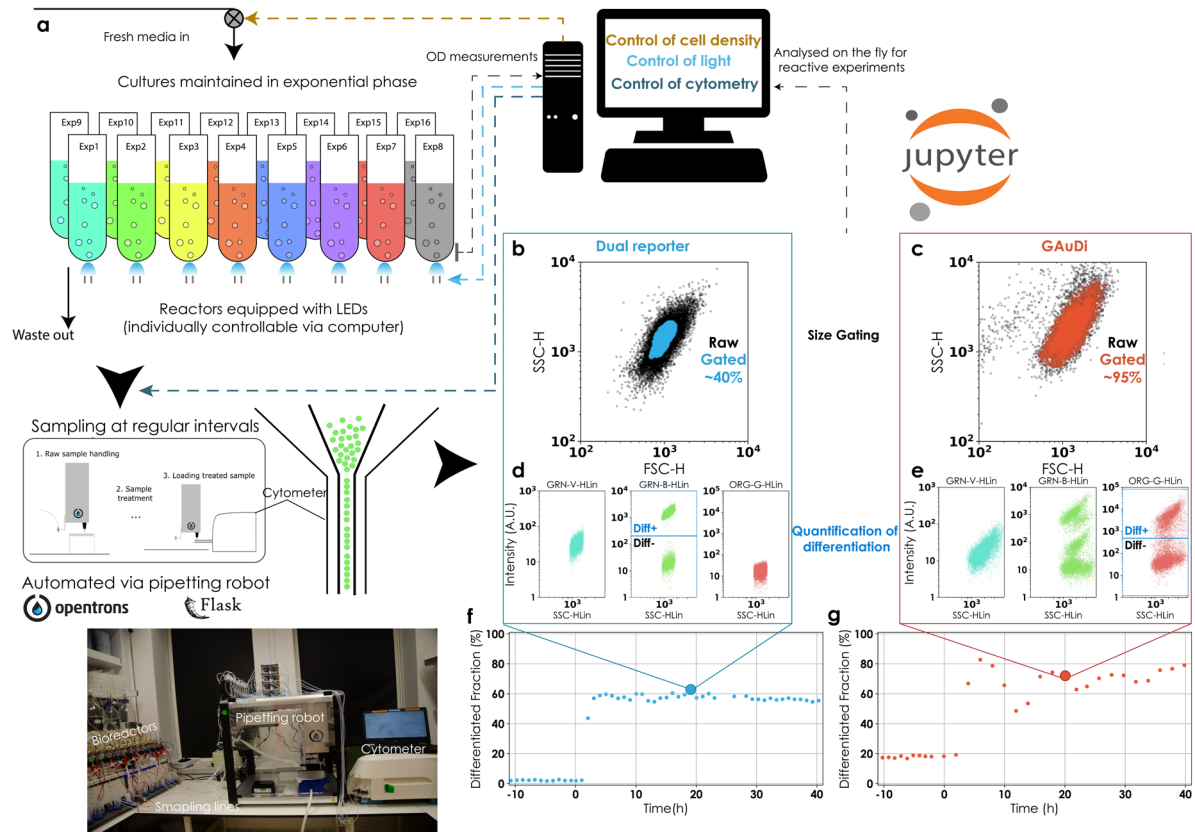


Figure 6.4. a. Experimental platform and b-g. data overview. 16 experiments can be conducted in parallel in reactors with individual OD and LED induction control. The cytometry measurements are automated via a pipetting robot controlled with the Flask app. The raw data from the cytometer is parsed and stored as a csv file or can be analysed online to change experiment conditions such as cell density or induction profiles. All code required for the functioning of the reactors and analysis is developed in Python and implemented using Jupyter notebooks. An image of the physical setup of the experimental platform is shown (a, bottom left). **b-e** Representative data from a single timepoint for one experiment each of the original differentiation system (blue box) and GAuDi strains (red box). The FSC vs SSC scatterplot is shown as an indicator of cell size (**b** and **c**). The GRN-B and ORG-G channels are used to detect differentiated cells for the original differentiation system (mNeonGreen fluorescence) and GAuDi02 (mScarlet-I fluorescence), respectively (blue boxes in **d** and **e**). At this representative timepoint (enlarged circle in **f** and **g**), both exist in two subpopulations.

Even though the bioreactor platform allowed me to grow 16 cultures at the same time, only 8 of these 16 could be used to get reliable cytometry data due to sporadic cross contamination between wells during sampling.

Cells were allowed to grow in the dark until exponential phase following the overnight cultures inside the culture vessels and induction was started only after growth rate stabilised. Samples were taken automatically from the turbidostat at regular intervals, diluted 20 times with a pipetting robot and passed loaded in the cytometer. The entire vessel, including pumps and tubing, was autoclaved before each experiment. Unless stated differently, the experiments used a “grow and dilute” program where cells were allowed to grow until OD 0.6 and then diluted to OD 0.4. The growth rate was computed by calculating the slope of linear curve fit to the log of OD data with time. This could be done for each growth dilution cycle (discrete fit), however, the discrete fit performed poorly when there was a growth arrest and necessitated continuous estimation of the growth rate. This was achieved by fitting a rolling window of 30 measurements with time (continuous fit). Continuous fit could capture other features that could not be seen with the discrete fit like gradual increase in growth rate during the lag phase (**Figure 6.5a**).

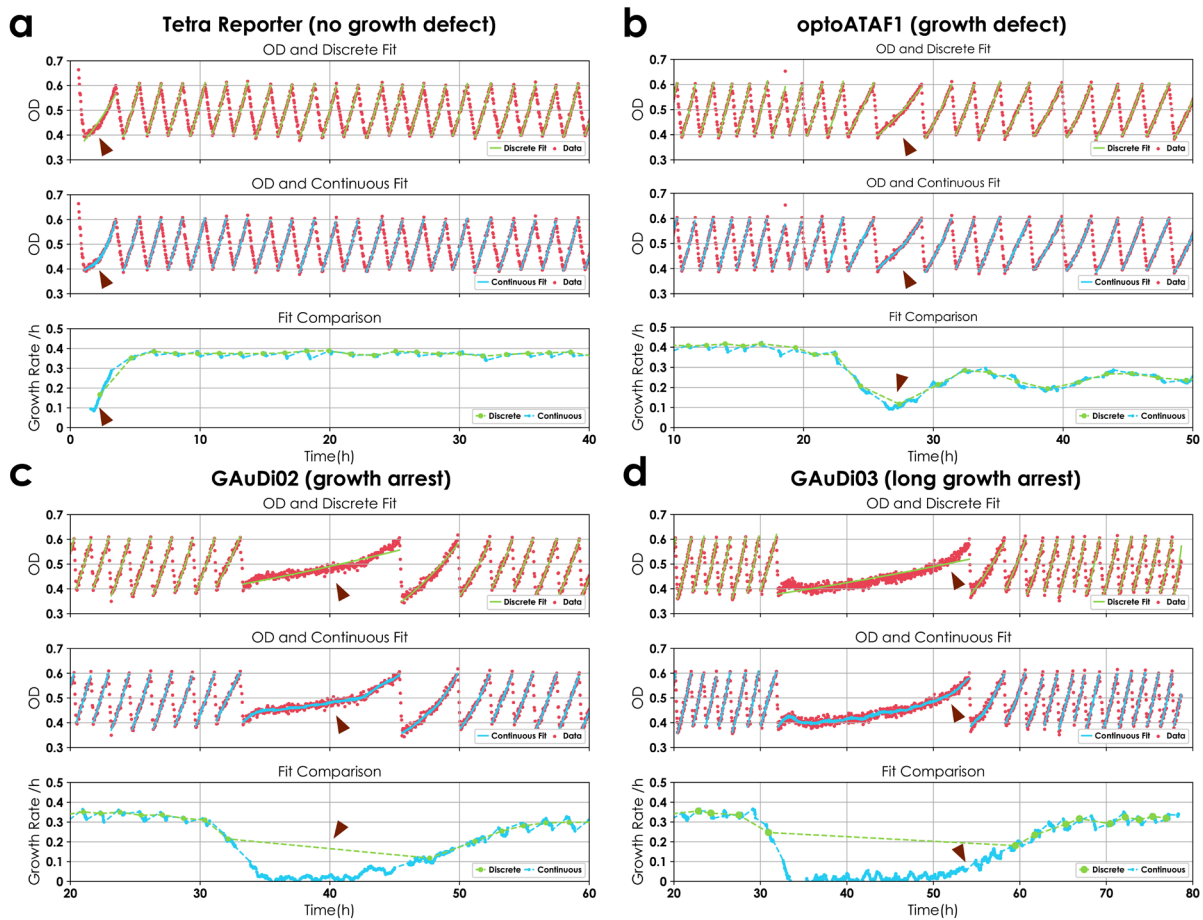


Figure 6.5. OD curves and growth rate computation for 4 different strains. Cultures were grown in a grow and dilute program where they were allowed to grow until OD 0.6 before being diluted back to OD 0.4 (red circles). Growth rate was computed by fitting a linear curve to log of OD with respect to time between dilutions (green dotted line) or by fitting a linear curve to log (OD) with respect to time for a rolling window of 30 measurements (blue dotted line). Green (discrete) and blue (continuous) lines shown on top of the OD data are computed from the fit. **a. Tetra reporter** showed no growth defect and growth was estimated equally well by both the continuous and the discrete fit. The red arrow indicates the lag phase where continuous fit is able to capture the gradual increase in growth rate. **b. optoATAF1** shows a significant growth defect upon light induction. Both fits are able to capture relatively well. Red arrow indicates small differences between the two estimates. **c. GAuDi02** displays a prolonged growth arrest that is not captured by the discrete fit. Red arrow highlights the difference. **d. GAuDi03** shows a massive growth defect. Red arrow notes the escape.

Pipeline for analysis of flow cytometry data

All cytometry measurements were made with a Guava EasyCyte BGV 14HT benchtop flow cytometer. Settings and gains were kept constant for all the experiments.

5000 events were recorded for each sample unless specified differently. No compensation was used during acquisition. Dilutions were made with a pipetting robot so that the cell density was kept between 200 (to have 5000 events in the acquisition window) and 600 cells/ μl (to ensure >90% singlets). Size gating and doublet removal were done using kernel density based methods. Singlets were selected based on deviation from linearity in Forward Scatter Height (FSC-H) vs. Forward Scatter Area (FSC-A). Cells were scored and a threshold was defined above which cells were classified as doublets and removed from analysis (**Figure 6.6b**). For size gating, 2D kernel density estimates were obtained using SciPy Gaussian kde package on Forward Scatter (FSC-H) vs. Side Scatter (SSC-H) and regions of density lower than a threshold were removed (**Figure 6.6a**). The two thresholds were kept constant for all measurements except those made with the GAuDi strain. For the latter, thresholds were increased to include the entire population. This leniency was warranted because of considerable changes in the physiology of growth-arrested cells (compare **Figure 6.4b & 6.4c**).

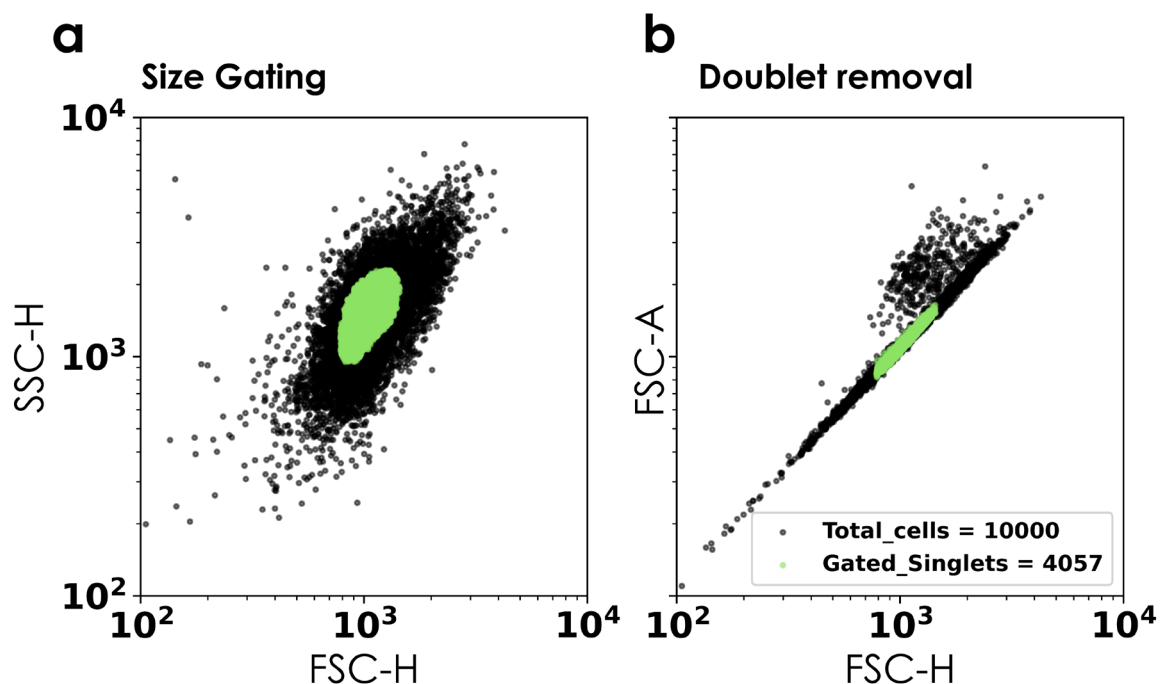


Figure 6.6. Gating. Kernel density based gating of data on **a.** size and **b.** doublet removal. Raw data is shown in black. Gated data is shown in green.

It was difficult to observe all the four proteins together due to significant overlap of fluorescence spectra (**Figure 6.7**). To mitigate this, we implemented a deconvolution approach previously

described in Bertaux et al. (2020). Briefly, 4 single fluorescent protein control strains (mCerulean, mNeonGreen, mVenus, mScarlet-I) with the same promoter and terminator, and integrated in the same locus, were used to determine the spectral signature of each fluorescent protein across the 12 channels of the cytometer (**Figure 6.8**). I also constructed a 4-colour strain in order to validate the results of the deconvolution algorithm.

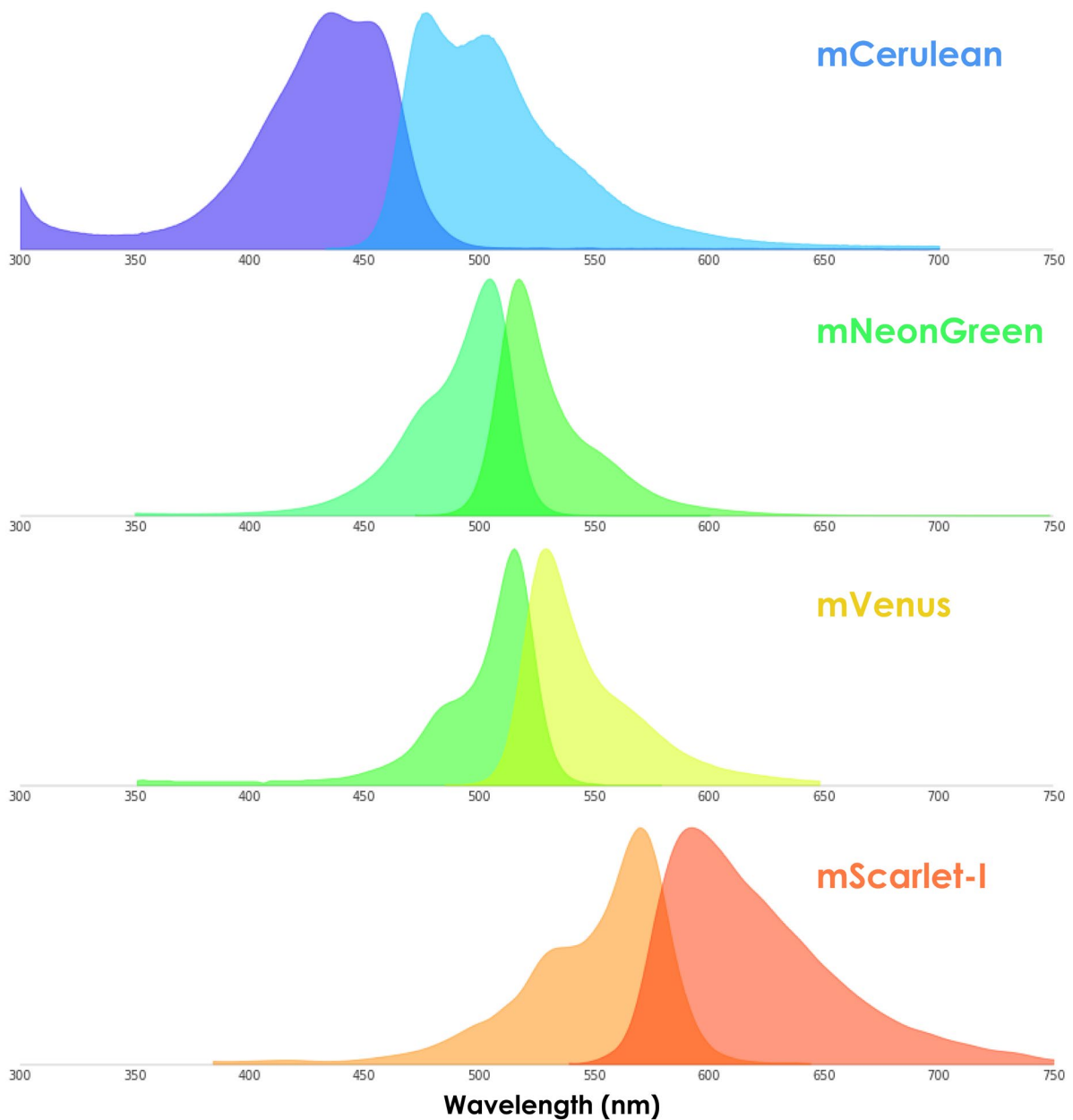


Figure 6.7. *Excitation* (dark) and *emission* (pale) *spectra* of the different fluorophores used in the study (Lambert 2019).

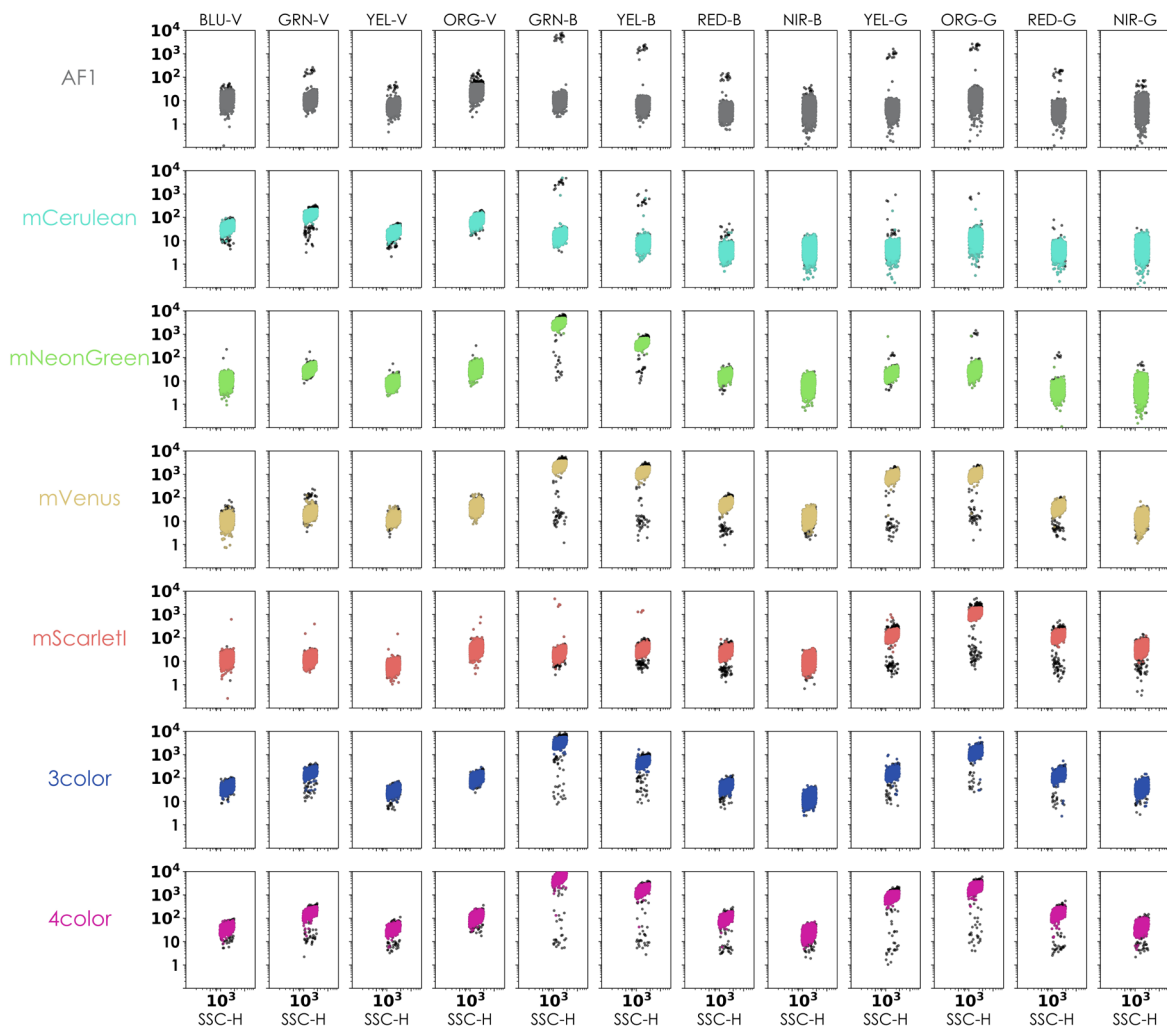


Figure 6.8. Raw data. Population distributions of measured intensity by different channels of the flow cytometer for an auto-fluorescence (grey) strain, four single fluorescent protein strains with identical promoters and terminators, mCerulean, mNeonGreen, mVenus and mScarlet1, and a four colour strain that has all 4 fluorescent proteins listed above. Data was filtered for outliers using mean absolute deviation metric (4.8 medians) to remove sporadic cross-contamination from other cultures during sampling. I note that filtering was performed only for the experiment used to define spectral signatures.

These signatures were then used in a linear algebra framework to calculate the individual fluorescence of each fluorophore in a strain harbouring all 4 of the fluorescent proteins. The values for single colour control and the 4-colour strain were in good agreement after deconvolution (**Figure 6.9 & 6.10b**).

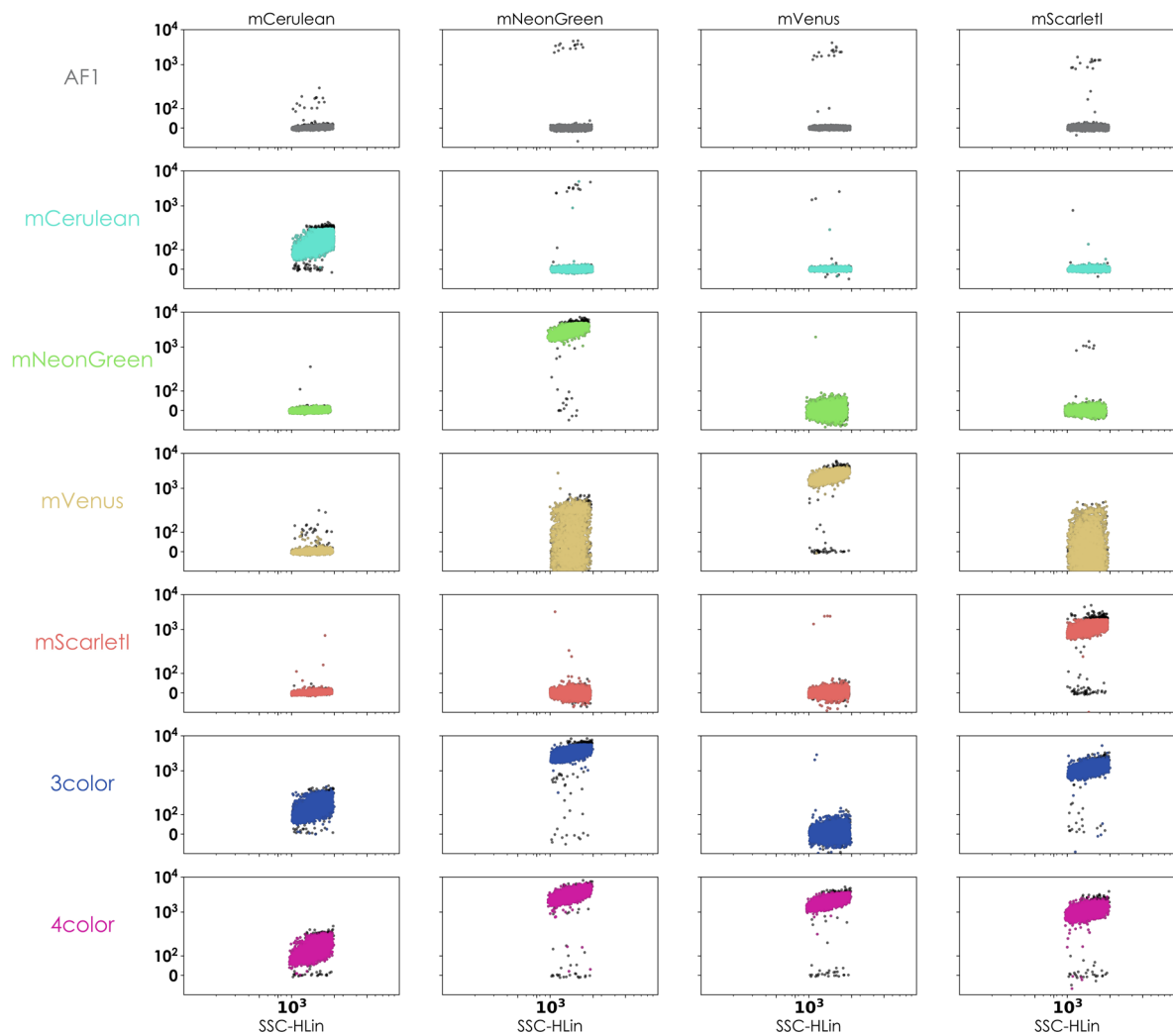


Figure 6.9. Deconvolved data. Filtered channel data was used to compute spectral signatures of each fluorescent protein. These signatures were then used to deconvolve fluorescence in raw data. Deconvolution also reduced size related heterogeneity in fluorescence.

I note that the deconvolved values of mNeonGreen and mScarletI distributions show significant deviations from the AF levels that are expected to be present in the constitutive mVenus strain even though the mean is zero for both distributions. This is perhaps because mVenus excitation spectrum, in addition to the emission spectrum, also overlaps significantly with mScarletI and mNeonGreen. This leads to a strong correlation between the mVenus fluorescence in a cell and the erroneous readout for mNeonGreen and mScarlet-I in the same cell.

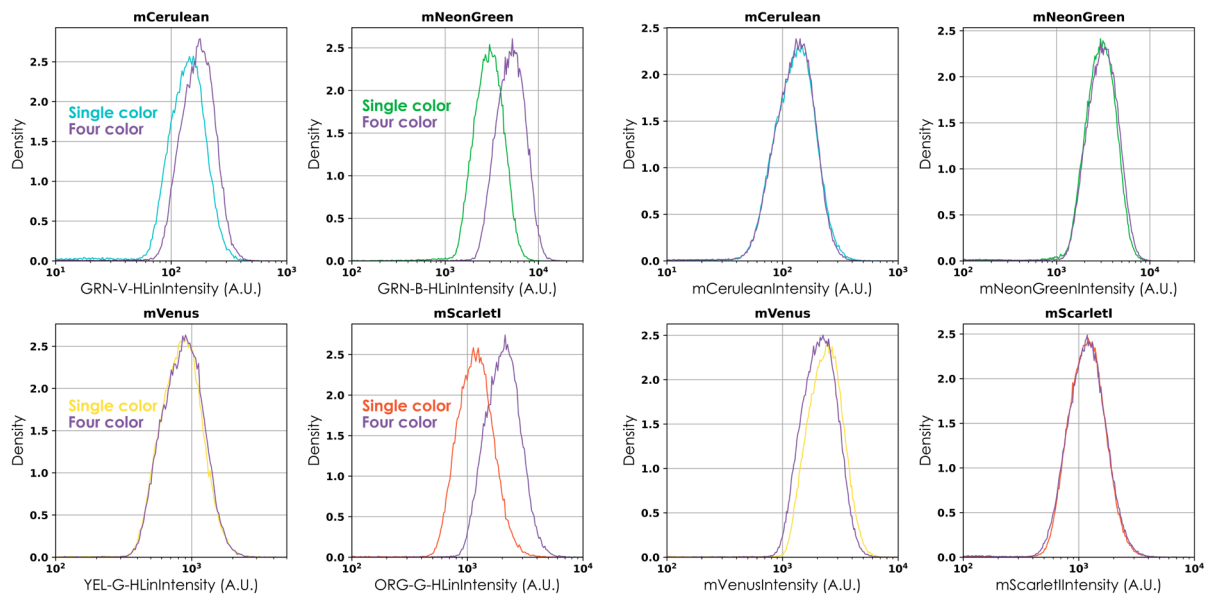


Figure 6.10. Fluorescence histograms for a. raw data and b. deconvolved data for single colour and 4 colour strains. For each fluorescent protein, fluorescence of the most distinguishing channel was plotted in the raw data. These are noted as the x-axis labels in the subplots pertaining to individual fluorophores. Deconvolved values were obtained for each protein. Both **a** and **b** compare the values obtained for single colour strains and the four colour strain. Cells were size gated and doublets were removed. Histograms are composed of >100,000 cells.

There was a slight disagreement between the single colour and 4 colour deconvolved fluorescence for mVenus (5% shift in median value). This is perhaps due to errors in deconvolution but might also stem from epigenetic differences in the gene expression loci for single colour (URA3 locus) and 4 colour (HO locus)(Bai Flagfeldt et al. 2009).

Quantifying differentiation

All differentiation circuits described in this work were designed to show a change of fluorescence upon differentiation. Consequently, the classification of a cell as differentiated or non-differentiated was a rather trivial affair and was achieved by applying a simple threshold on the fluorescence in the relevant channel. This process was slightly more involved for the extended system that could give rise to 3 or 4 species after differentiation. In what follows, I provide a simple description of the various differentiation strains and the quantification of differentiation for each of them.

Dual reporter

The simplest construction of the lot. Raw fluorescence in the GRN-B channel was sufficient to discriminate the two populations. Fluorescence in the GRN-V channel was weak and did not show

two well-separated populations. However, from the microscopy data it is clear that the cells lose mCerulean fluorescence after differentiation.

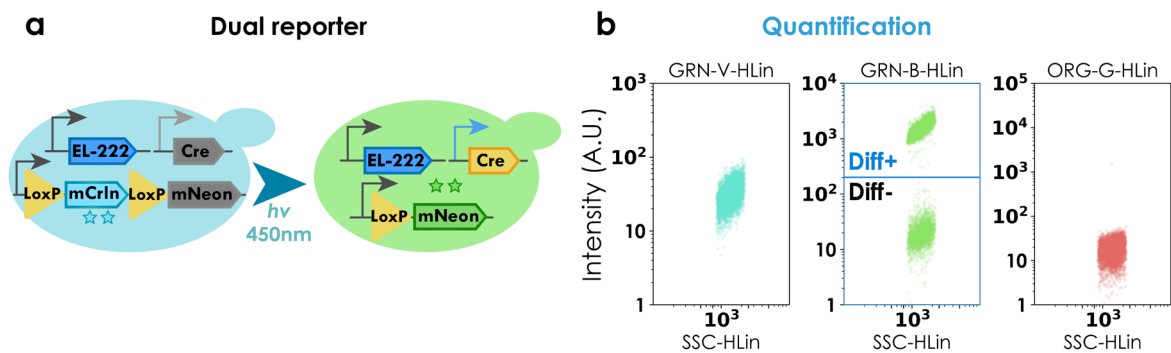


Figure 6.11. a. Dual reporter. b. Fluorescence in relevant channels. Differentiation status was ascertained by determining whether a given cell exceeded the **threshold** intensity of **200 A.U.** in the **GRN-B** channel.

GAuDi

GAuDi strains were equally easy to classify into differentiated and non-differentiated populations by observing raw fluorescence in the ORG-G channel.

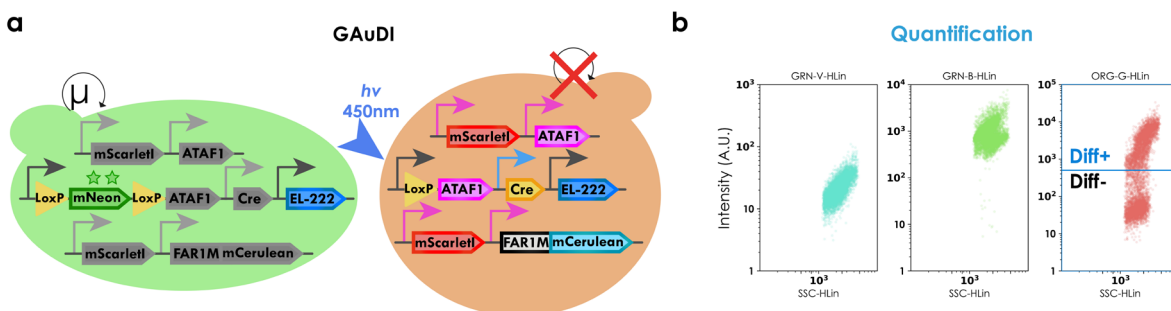


Figure 6.12. a. Construction of GAuDi02. b. Fluorescence in relevant channels. Differentiation status was ascertained by determining whether a given cell exceeded the **threshold** intensity of **500 A.U.** in the **ORG-G** channel.

However, in the case of GAuDi02 & GAuDi03 I could observe a persistent population of seemingly differentiated cells prior to differentiation. To eliminate spurious signals due to crosstalk between the two fluorophores, I deconvolved the data and repeated the analysis. I did not observe a significant change in the quantification however, I found that the population showed a long tail in mScarlet-I distribution before induction (**Figure 6.13a, bottom left**). I also observed a lower and noisier peak in mScarlet fluorescence for the differentiated fraction before induction when compared to differentiated cells after induction. Furthermore, the mNeonGreen fluorescence of the differentiated fraction before induction was higher than the non-differentiated fraction. This would be the case if mNeonGreen expression was still present (i.e. no recombination) but due to the overactive feedback loop the growth rate decreased thus

leading to a higher accumulation of mNeonGreen. These observations have led me to conclude that these cells are not differentiated but are a consequence of leakage from the ATAF1 promoter that triggers the positive feedback loop in a subset of the population. Interestingly, I found that the mNeonGreen fluorescence in the differentiated cells did not decrease to autofluorescence levels suggesting the differentiated cells are arrested. In addition, I could clearly see a peak at roughly half the fluorescence of non-differentiated cells maybe indicating that the growth is arrested after a single division post differentiation.

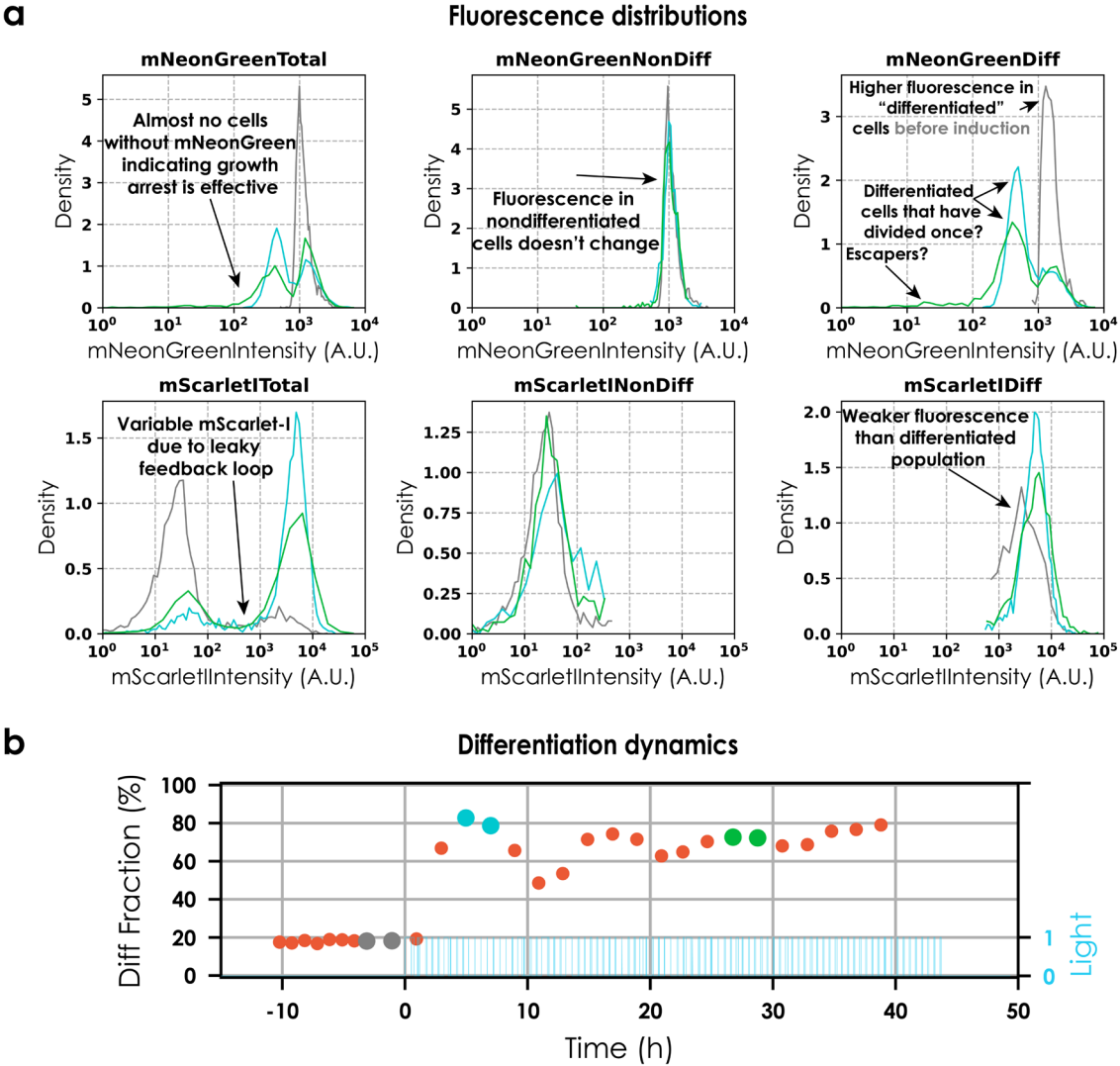


Figure 6.13. Fluorescence histograms of a. deconvolved data for *GauDi02* at three different times after induction for the experiment shown in b. Differentiation dynamics. The colour indicates the time at which histograms are plotted (see enlarged circles in b.). a. compares the fluorescence in mNeonGreen and mScarlet-I channels. Data is represented as in Figure 6.10. Cells were size gated and doublets were removed. 5000 events were acquired for each time point and histograms consist of two timepoints.

Multi-species

It was not possible to use the raw cytometry data to delineate the different species present in the culture. Therefore, the data was deconvolved. This allowed for discrimination between (CN and $\bar{C}\bar{N}$), $C\bar{N}$, and $\bar{C}N$. To separate the populations CN and $\bar{C}N$, it was necessary to look at the ratio of mNeonGreen to mCerulean.

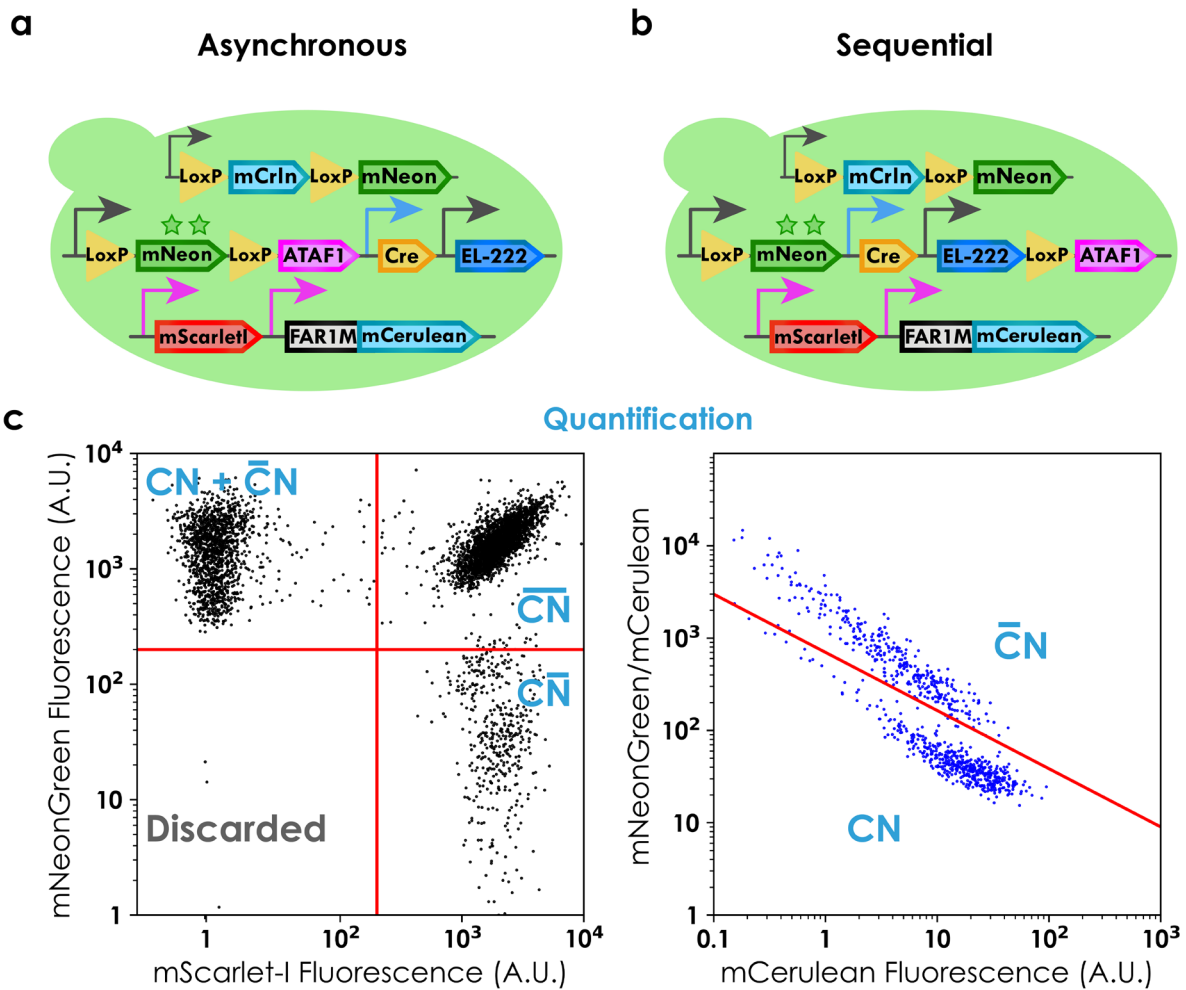


Figure 6.14. a. & b. Constructions of **asynchronous** and **sequential differentiation programs** respectively. **c. Quantification of species prevalence.** Cells possessing **mScarlet-I** fluorescence **over a threshold of 200 A.U.** were either $\bar{C}\bar{N}$ or $C\bar{N}$ and those that did not were CN or $\bar{C}N$. To distinguish between $\bar{C}\bar{N}$ and $C\bar{N}$ **mNeonGreen** fluorescence was enough (**threshold of 200 A.U.**). However, to distinguish between CN and $\bar{C}N$, both mNeonGreen and mCerulean fluorescence had to be analysed to obtain two well-separated populations (high mNeonGreen to mCerulean ratio corresponds to CN and the lower to $\bar{C}N$). Cells not possessing appreciable levels of either mNeonGreen or mScarletI were discarded from analysis (<1%). The figure shows the four subpopulations after stimulation with 3 light pulses of 30 minutes 6h apart at $t=24h$ for the asynchronous differentiation program (Figure 3.16a). Deconvolution was necessary to achieve four well separated populations.

Tetra Reporter

Figure 6.15a and **Figure 6.16a** show raw fluorescence values in the most distinguishing channel for each of the four fluorescent proteins present in the tetra reporter strains in dark (grey), when 99% of the population is undifferentiated, and in continued presence of light (blue) after 99% of the population had differentiated. Prior to deconvolution, the raw data possessed several characteristics suboptimal for analysis. Notably, due to low fluorescence of mCerulean, the undifferentiated population (possessing mCerulean) overlaps significantly with the differentiated population (possessing no mCerulean) in the GRN-V channel. Seemingly, the undifferentiated cells possess 5-6 times higher fluorescence than the WT strain in the GRN-B channel. Furthermore, the fluorescence in the YEL-G channel, misleadingly, suggests significant differences in mVenus fluorescence between differentiated and undifferentiated cells. Lastly, observing the ORG-G channel, we find that cells in the dark possess 3-4 fold higher fluorescence compared to the WT strain prompting the misleading conclusion that EL222 is activated without induction.

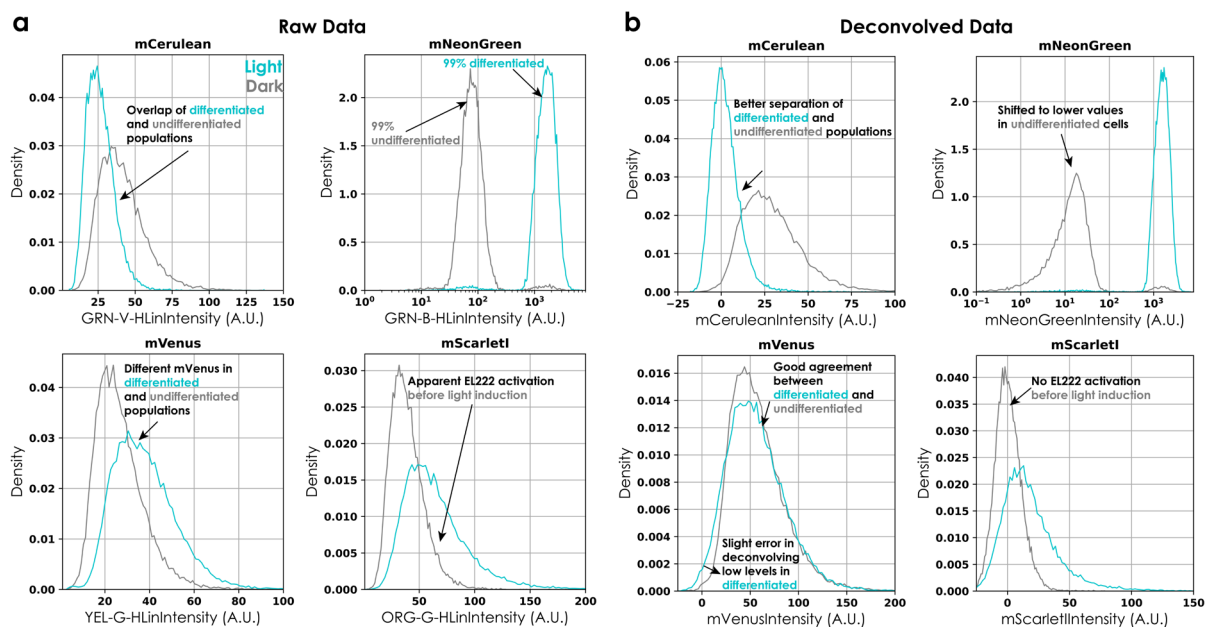


Figure 6.15. Fluorescence histograms of a. raw data and b. deconvolved data for integrated tetra reporter in light (blue) and dark (grey). Data is represented as in Figure 6.10. Both **a** and **b** compare the distributions obtained for tetra reporter in light and dark conditions. Cells were size gated and doublets were removed. Histograms are composed of >10,000 cells.

The situation becomes significantly worse for the plasmid tetra reporter such that we fail to observe two well-separated populations for differentiated and undifferentiated cells (**Figure 6.16a, top right**). Furthermore, it appears from YEL-G fluorescence that cells stop losing the plasmid upon differentiation (**Figure 6.16a, bottom left**).

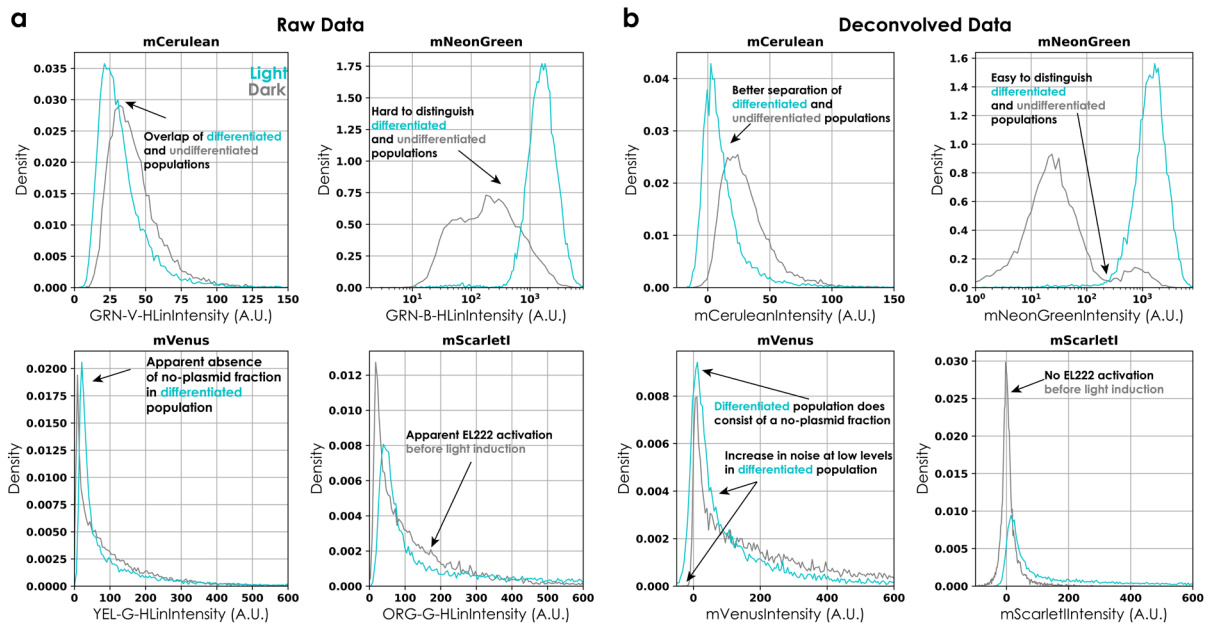


Figure 6.16. Fluorescence histograms of a. raw data and b. deconvolved data for 2-micron tetra reporter in light (blue) and dark (grey). Data is represented as in Figure 6.10. Both **a** and **b** compare the distributions obtained for tetra reporter in light and dark conditions. Cells were size gated and doublets were removed. Histograms are composed of >10,000 cells.

Deploying this deconvolution strategy allows me to deconvolve fluorescence in the tetra reporter. After deconvolution, all the above-mentioned incongruities either disappear completely or are significantly diminished. I note that deconvolution of low levels of mVenus in the presence of mNeonGreen (differentiated cells) yields noisy estimates (**Figure 6.15b, bottom left**) that lead to a broad spread around zero for the plasmid strain (**Figure 6.16b, bottom left**).

For the quantification of differentiation, FSC normalized values were used. These values were scaled by multiplying with the mean FSC of the population to preserve units. I found that normalizing with the FSC led to tighter distributions (**Figure 6.17**), except for mScarlet-I fluorescence, perhaps due to low fluorescence values.

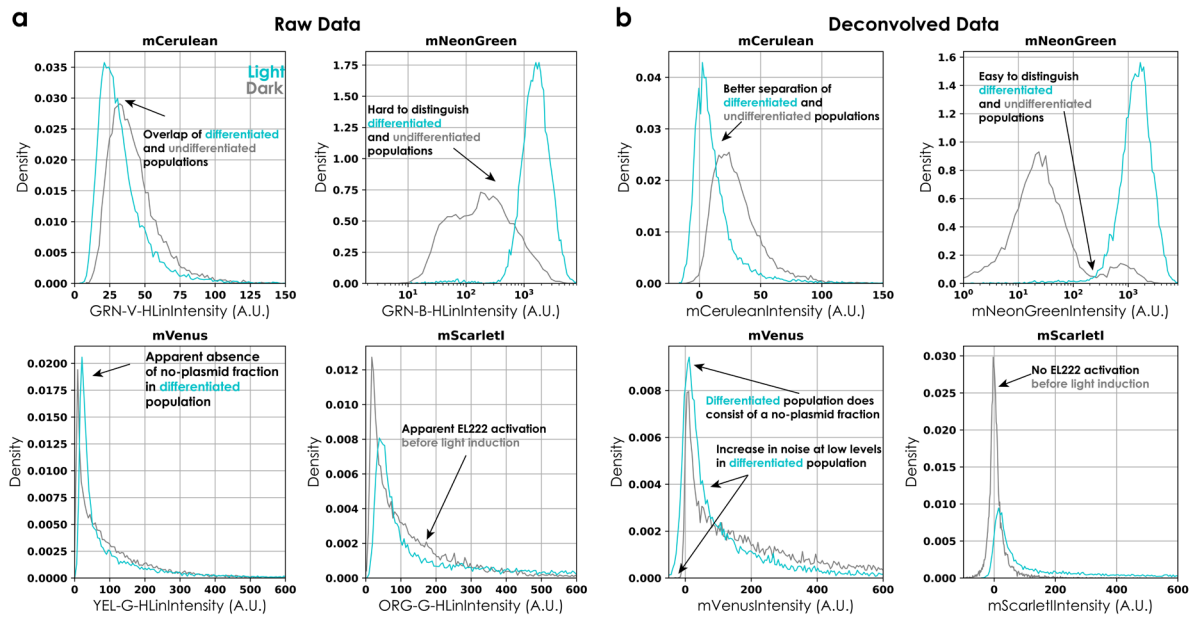


Figure 6.17. Fluorescence histograms of the **four fluorescent proteins** in the **integrated tetra reporter strain** in **a. dark** and **b. light** for deconvolved (grey), gated singlets (green), FSC normalized deconvolved (blue), and FSC normalized gated singlets (purple) data. Histograms are composed of >10,000 cells. Normalization decreases variability.

After normalization, I could observe two sharply defined peaks for both the integrated and the plasmid strain (**Figure 6.18b**).

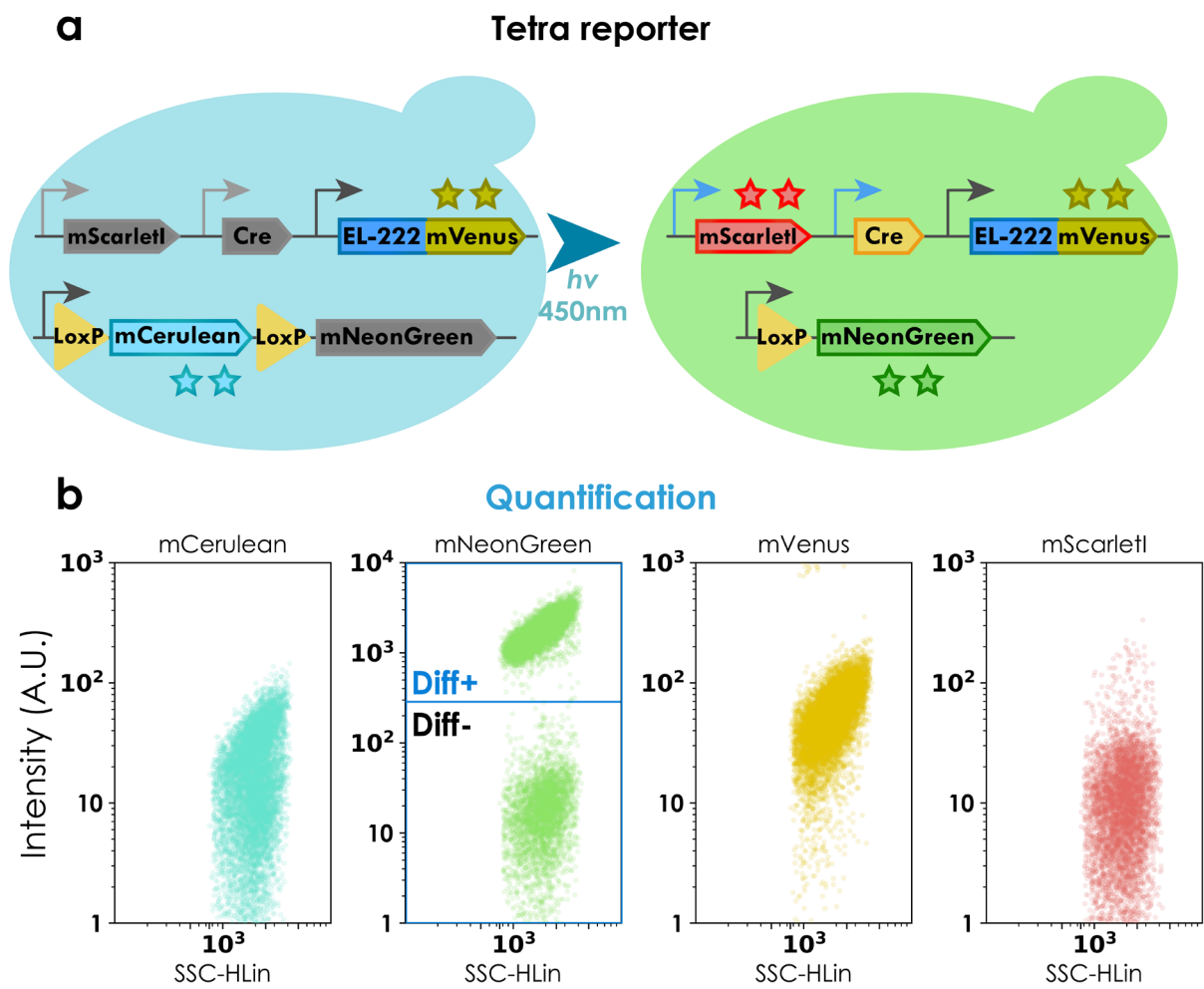


Figure 6.18. a. Construction. b. Deconvolved fluorescence for relevant fluorophores. Differentiation status was ascertained by determining whether a given cell exceeded the **threshold** intensity of **300 A.U.** in the **mNeonGreen** channel.

Model fitting and parameter estimation

ODE models were solved using `solve_ivp` solver from SciPy.integrate library. Models were fitted using the least square method from SciPy.optimize library. Parameter search for each model was done locally (gradient descent) with multiple initial guesses for parameters and bounds on parameter values between 10^{-10} and 10.

Dual Reporter

The ODE model was conceived with only one free parameter, namely the differentiation rate r_{diff} . Initial conditions and the growth rate were fixed from the data (**Figure 3.10a**). Experiments were performed in the bioreactor platform, which showed modest reactor-to-reactor variability.

The model was fit to steady state differentiation fractions after single pulse light inductions of different durations. Data fitting was implemented in Python using SciPy library. Bounds were imposed on the parameter search (10^{-10} -10) to preserve physiological relevance. Parameter searches were conducted locally but starting from different initial parameter values spanning four orders of magnitude using an approach similar to Branch, Coleman and Li (1999). In all cases, parameter searches converged to the same parameter value.

GAuDi02

The model was fitted to dynamical data (differentiated & dead fraction) with non-trivial light signals (**Figure 4.20a**). Concretely, the differentiation fraction and the dead fraction were computed from time-series flow cytometry data to fit the model. Initial conditions and the growth rate were fixed from the data. However, there were significant delays between differentiation events and the time by which corresponding cells could be classified as differentiated based on their fluorescence. Similarly, differentiated cells died only after being arrested for some time and the growth arrest required two cell generations to fully manifest. To account for these delays, model predictions were shifted in time (2h for differentiated and 6h for dead cells, and 5h for the growth rate). Bounds were imposed on the parameter values (10^{-10} -10) to preserve physiological relevance. The parameter search was conducted locally but starting from different initial parameter values spanning four orders of magnitude using an approach similar to Branch Coleman and Li (Branch, Coleman, and Li 1999). In all cases, the search converged to the same parameter values.

Tetra reporter and plasmid dynamics

As these models were not developed by me I have not included this information here, however, for the sake of completeness and for curious readers these can be found in the supplementary information of the submitted manuscript (Annex 5).

MPC framework

ODE models were used in an Model predictive framework described previously (Bertaux et al. 2020). The framework consisted of solving the model given a light sequence. This sequence was then optimised using a least square method from SciPy.optimize library to minimize the error between predictions and the target set point starting from an initial state. The optimization for light sequence was done for a time horizon of 5 hours in the form of 10 duty cycles of 30-minute period. Cultures were sampled every hour for two reactor control experiments (**Figure 3.15b & 3.15c**) and every two hours for single reactor control experiments (**Figure 4.24b**). The light sequence was updated at each timepoint. To estimate the state of the system, cytometry data was analysed online to determine the state at the time of sampling. This estimate of sampling time state was then used as initial conditions for the model and the current state was estimated by solving the model for a time, τ . τ stands for delay, consisting of a sampling delay (time passed between sample acquisition and finalized data analysis) and an observation delay (time required for enough fluorescent protein to accumulate to pass the differentiation threshold). I note that accounting for this delay was critical to avoid oscillatory behaviour. Furthermore, we noticed that the differentiation rate parameter estimated from experiments led to a steady state error. Therefore, the value of this parameter was reduced to 0.4 /h in the MPC model to ensure robust control.

A step-by-step enumeration of the process follows.

1. Flow cytometry data is processed online to yield state estimates.
2. Some variables, like the escaper fraction, are non-measurable and therefore estimated from the model.
3. Conversion of this state estimate into a real-time state estimate.
 - a. Because of delay not captured in the model, state estimates are taken to correspond to;
 $T_0 = \text{sampling time} - \tau$ (*sampling delay + observation delay*)
 - i. Sampling delay was the time between when sample was taken and the data acquisition finished.
 - ii. Observation delay was the delay between change of genetic state of a cell and by the time enough fluorescence was produced to classify it as differentiated cell.
 - b. Forward simulation of the model to go from T_0 to T_{now} (given the actual light sequence applied during that time window)

4. Optimization of the next sequence of 10 duty cycles (horizon = 5h) so as to minimize the predicted error between target and model prediction (gradient descent, scipy.optimize)

Modified ODE model for control of differentiation system

$$\frac{dg}{dt} = \mu_{target}g - \mathbf{U}(t) r_{diff}g + Q - \lambda g$$

$$\frac{dp}{dt} = \mu_{target}p + \mathbf{U}(t) r_{diff}g - \lambda p$$

$$Q = \mu_{reservoir}n_{reservoir}$$

$$n_{target} = g + p$$

g and p stand for the specific cell density (in O.D. units) of non-differentiated and differentiated cells, respectively. $\mu_{control}$ and $\mu_{reservoir}$ are growth rates in the control and reservoir reactors, respectively. $\mathbf{U}(t)$ is the light signal as a function of time and can take values 0 or 1. n_{target} and $n_{reservoir}$ are O.D. (total cell densities) at which control and reservoir reactors are maintained, respectively. r_{diff} is the differentiation rate under continuous light. Q is the flux of non-differentiated cells from reservoir to control reactor. λ is the dilution rate of the control reactor. I observed that growth rates of cultures at different ODs were not equal. I assumed, however, that reservoir cells instantaneously change their growth rate upon entering the control reactor.

Since control culture was maintained at constant OD, the number of cells in the control reactor stays constant at steady state.

$$\frac{dn_{target}}{dt} = 0$$

$$\frac{d(g+p)}{dt} = 0$$

$$\frac{dg}{dt} + \frac{dp}{dt} = \mu_{target}(g+p) + Q - \lambda(g+p)$$

$$0 = \mu_{target}(g+p) + Q - \lambda(g+p)$$

$$\lambda = \mu_{target} + \mu_{reservoir} \frac{n_{reservoir}}{g+p}$$

$$\lambda = \mu_{target} + \mu_{reservoir} \frac{n_{reservoir}}{n_{target}}$$

$$\lambda = \mu_{target} + \alpha \mu_{reservoir}$$

α is the ratio of reservoir OD to control reactor OD.

Let r be equal to the fraction of differentiated cells in the control reactor.

$$r = \frac{p}{g + p}$$

$$\frac{dr}{dt} = \frac{\frac{dp}{dt}}{g + p}$$

$$\frac{dr}{dt} = \frac{\mu_{target}p + \mathbf{U}(t)r_{diff}g - (\mu_{target} + \alpha\mu_{reservoir})p}{g + p}$$

$$\frac{dr}{dt} = \mathbf{U}(t)r_{diff}(1 - r) - \alpha\mu_{reservoir}r$$

$$\frac{dr}{dt} = -(\mathbf{U}(t)r_{diff} + \alpha\mu_{reservoir})r + \mathbf{U}(t)r_{diff}$$

This is a first order non-homogeneous linear differential equation in standard form,

$$\frac{dy}{dt} = -p(t)y + f(t)$$

For maximum differentiation fraction achievable ($\mathbf{U}(t) = 1$), at steady state,

$$\frac{dr}{dt} = 0$$

$$r_{diff} = (r_{diff} + \alpha\mu_{reservoir})r_{max}$$

$$\alpha = \frac{r_{diff}}{\mu_{reservoir}} \left(\frac{1}{r_{max}} - 1 \right)$$

Plugging in values of $r_{diff} = 0.86 \text{ h}^{-1}$, $r_{max} = 0.9$ and $\mu_{reservoir} = 0.41 \text{ h}^{-1}$ we get, $\alpha = 0.23$.

Therefore, to reach a target differentiation level of 90 % in the control reactor, the OD of the reservoir had to be ~5 times lower.

6.9 Estimating protein production in continuous cultures

Cells carrying different heterologous expression constructs were cultured continuously in the turbidostat platform between OD 0.4 and OD 0.6. All constructs expressed the same protein of interest (POI), mScarlet-I. Constitutively expressing strain was cultured at a constant OD of 0.4 in the dark as I did not observe a significant growth defect. ATAF1 induction strains were cultured between OD 0.4 and OD 0.6 in presence of continuous blue light to infer growth rate changes due to induction and subsequent heterologous expression. GAUDI strains were also cultured between OD 0.4 and OD 0.6, however, they were not exposed to continuous light as the idea

was to maintain the culture composition to a target differentiation level (50 or 70%). For all cultures, automated sampling to the cytometer were effected regularly and 5000 events were acquired at each measurement (**Figure 6.19a**).

Gated and deconvolved cytometry data was used to calculate the mean cellular fluorescence (POI/cell) at each timepoint by averaging the fluorescence of all cells. To obtain titers (POI/ml), mean cellular fluorescence was multiplied by the cell density estimated from cell counts in the cytometer. It was further corrected for by the OD of the culture vessel and dead cells were removed from the analysis. This was necessary to obtain true cell densities of viable cells in GauDi02 cultures. Mean protein production (POI/h) was computed by multiplying the volume expelled between two time points with the titers and further normalized by the culture volume to get productivity (POI/ culture ml/h). Total protein produced in the run per ml of culture amounted to the area under the productivity vs time curve necessitated due to missing timepoints in some experiments. Yield was computed dividing the total protein produced in the run by the total media volume consumed (POI/ culture ml/ml media) (**Figure 6.19b**).

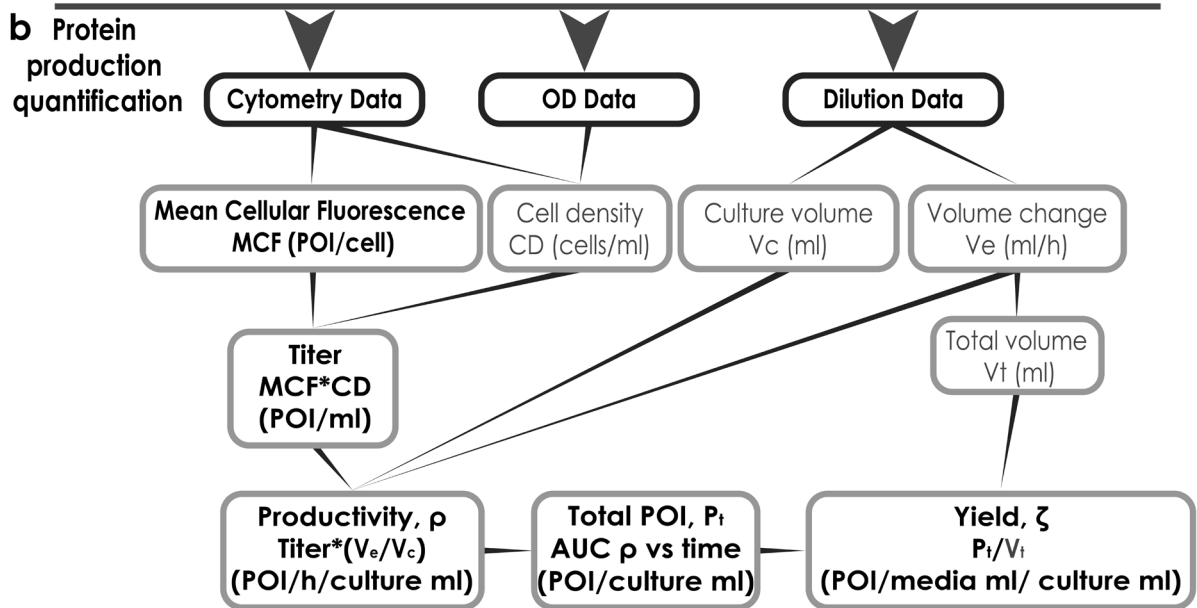
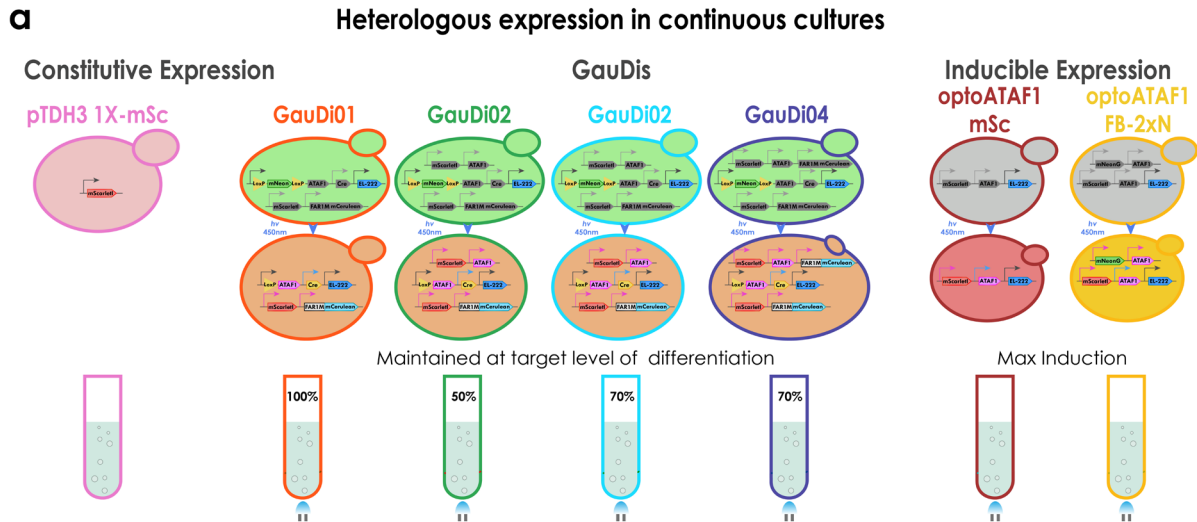


Figure 6.19. a. Description of the experiment. Cells were cultured continuously at fixed ODs. GAuDi cells were maintained at the mentioned composition. optoATAF1 cells were fully induced. pTDH3 cells were cultured in the dark. **b. Pipeline of the protein production quantifications.** Black clouds indicate data that is generated in the bioreactor experiment. Grey clouds signify the data is inferred or obtained from the black data. Bold typefont inside clouds indicate these values were plotted in Figure 4.25.

In order to obtain a value relative to the control pTDH3 driven expression, estimated total protein content was divided by the total protein produced by pTDH3 in the same amount of time. The yield was computed by dividing the total protein content by total media consumed and again normalized by the yield of pTDH3 culture to obtain values relative to pTDH3.

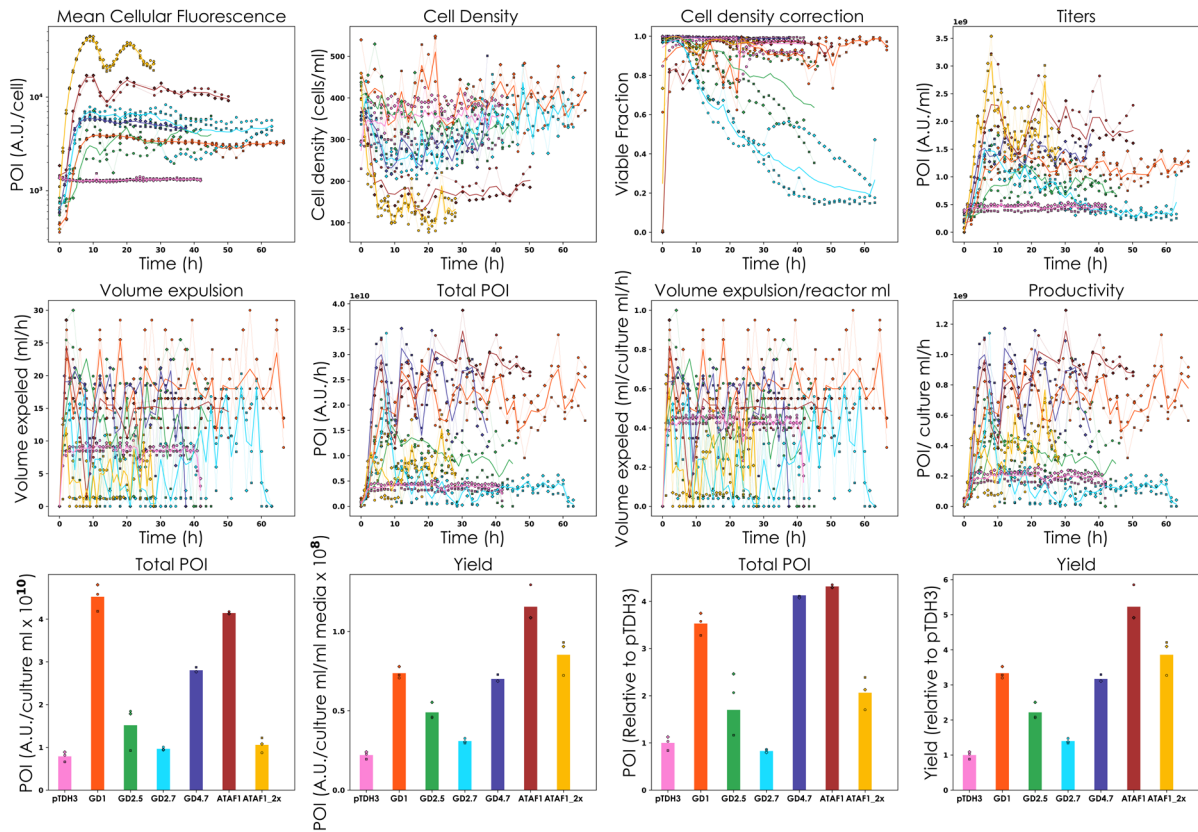


Figure 6.20. Comparison of three platforms of heterologous expression (constitutive, inducible, and differentiation) in continuous operation. Thick lines and bars represent the mean value from 2 or 3 replicates. Individual replicates are represented by thin lines. Data points from the three replicates are represented as circles, crosses, and diamonds. Replicates were conducted independently of each other. Colours indicate the strain (Figure 6.19a). Extended figure for Figure 4.24.

Experimental setup for live cell imaging and pattern formation

Microscopy was done on the inverted microscopy platform Leica DMi8 S. Illumination was effected via CoolLED PE-4000 illumination system (**Figure 6.21, left**).

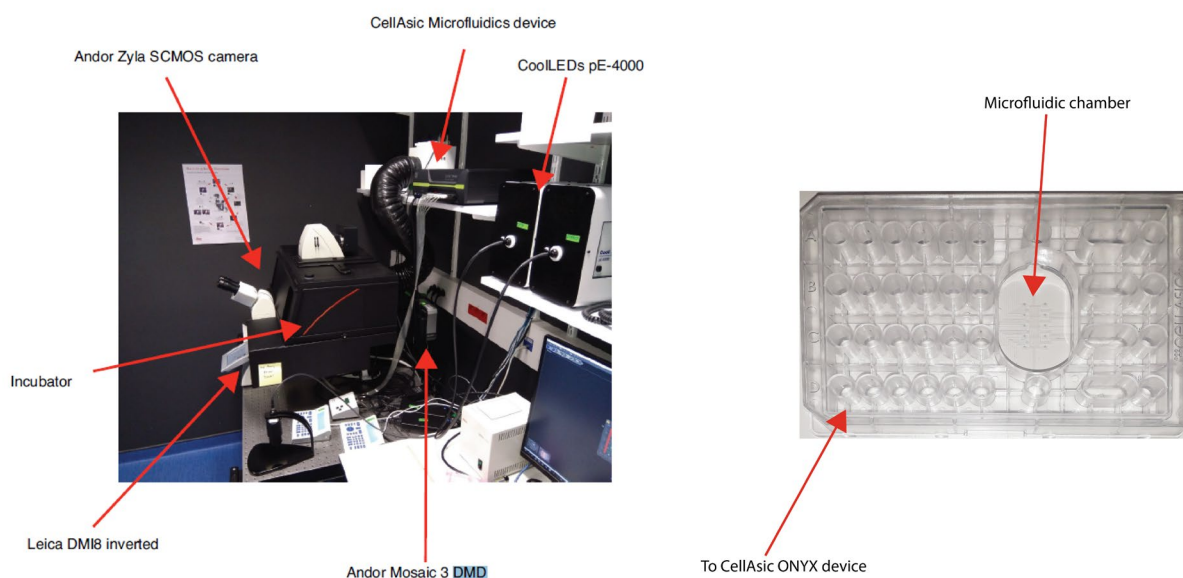


Figure 6.21. Microscopy platform (left) (Fox et al. 2021) and **microfluidic plate** (right).

The spectra of mCerulean, mNeonGreen, and mScarlet-I did not significantly overlap and therefore imaging could be done without the risk of crosstalk or bleed through.

Table 6.3. Illumination and imaging conditions.

Protein	Filter Cube	Excitation (nm)	Emission (nm)	Dichroic (nm)	Light Source (nm) - Intensity	Exposure (ms)	EL222 Induction
mCerulean	CFP	436/20	480/40	455	435-6%	600	Yes
mNeonGreen	YFP	500/20	535/30	515	500-9%	600	Yes
mScarlet-I	Rhodamine	546/10	585/40	560	550-10%	600	No

Prior to time-lapse live cell imaging, the cells were grown as stated in Section 6.3. Cells were allowed to grow to exponential phase on the day of the experiment. During time-lapse live cell imaging the chamber temperature was maintained at 30°C. The flow in the microfluidic chamber was kept constant at 7.5psi for all experiments described in this piece of work. The flow was maintained in the microfluidic chamber using the CellAsic Onyx device.

Briefly, this device, in conjunction with 4 chamber CellASIC plate (**Figure 6.21, right**) for haploid yeast cells (Y04C), allowed me to conduct 4 parallel experiments. Typically, for cells growing at the WT growth rate (0.34-0.44 h⁻¹), the time lapse experiment lasted between 10-16h. The duration was also heavily dependent on the initial cell density. I observed a large variability in the amount of cells that stayed in the microfluidic chamber after loading. I postulated that the

primary reason for this was cell density of the loaded culture (that in turn affects the cell size and physiology); however, after repeated experiments I reached the conclusion that in addition to the cell density, the variability also derived from the differences in the height of the microfluidic chamber. The latter was remarkably inconsistent and depended heavily on the manufacturing lot of plates.

I used an in-house software, called MicroMator, for the automated acquisition and cell tracking (Fox et al. 2021).

Induction was carried out by imaging the cells in the CFP and YFP channels every 6 minutes with wavelengths and intensity noted in the **Table 6.3**.

For pattern formation experiments, the cells were allowed to grow to late exponential phase ($OD \sim 0.8$) before loading 20 μ l of the culture in IBIDI μ -Slide VI 0.4 (Cat.No:80606). This was necessary to ensure a uniform monolayer at the beginning of the experiment. A light of 460 nm was used for induction with the CFP band pass filter (**Table 6.3**). The induction was effected for 1 s at intensity 10% every 3 minutes for 1 h in the shape of custom patterns via the use of a digital micromirror device (**Figure 6.22**). Following the induction, cells were allowed to recover for 1 h prior to imaging. This also gave time for enough mNeonGreen molecules to mature and be detected.

It was difficult to increase the duration of induction period, even though light was not lethal for cells, because of cell growth, and subsequent emergence of bubbles displacing the monolayer. I tried to ameliorate the growing conditions by culturing cells in a microfluidic plate but the media flow disallowed any meaningful quantification of efficiency of pattern formation at low cell density. This was because buds from inside the pattern would be transported with the flow, to regions outside the pattern. When I attempted to do the experiment after cells had formed a monolayer, flow was obstructed leading to growth rate decrease and cell death due to lack of media. We note that the cells could not be imaged during the induction period as both mCerulean and mNeonGreen required illumination that would activate the EL222 and lead to undesirable differentiation.

Experimental setup

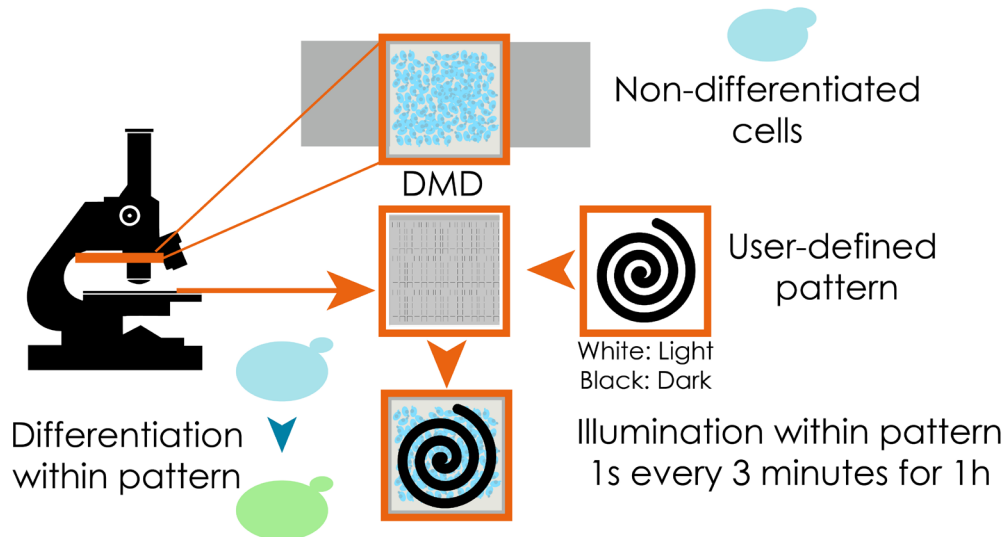


Figure 6.22 *Precise spatial control of heterogeneous microbial consortia* originating from a single strain. Hardware setup for realizing pattern formation under the microscope. Cells carrying the differentiation system were cultured in the dark until late exponential phase ($OD \sim 0.8$) and loaded in a μ bidi slide. Cells were allowed to settle down and form a monolayer. Light was shone in a user-defined pattern over the monolayer using a digital mirror device (DMD). Light was shone as 1s pulses every 3 minutes for 1h (20 pulses). Following light delivery, cells were kept in darkness for 1 hour before imaging them.

Pipeline for analysis of microscopy data

Cell segmentation was achieved via SegMator, an in-house U-net based segmentation algorithm (Fox et al. 2021), on bright field images (**Figure 6.23a**) of cells growing in a microfluidic plate (**Figure 6.23h**). Once cells were segmented (**Figure 6.23b**), median pixel fluorescence was computed for each cell in each channel (**Figure 6.23c & i**). Cellular fluorescence in YFP was used for ascribing differentiation status to cells (**Figure 6.23c**). Cells were considered differentiated once a threshold value for fluorescence was exceeded (**Figure 6.23d**). To estimate the growth rate, I fitted a linear curve to the $\log(\#cells)$ with respect to time (**Figure 6.23f**). I observed that growing in the constrained environment of the microfluidic chamber did not have a significant growth defect (**Figure 6.23g**). All analysis was done in Python.

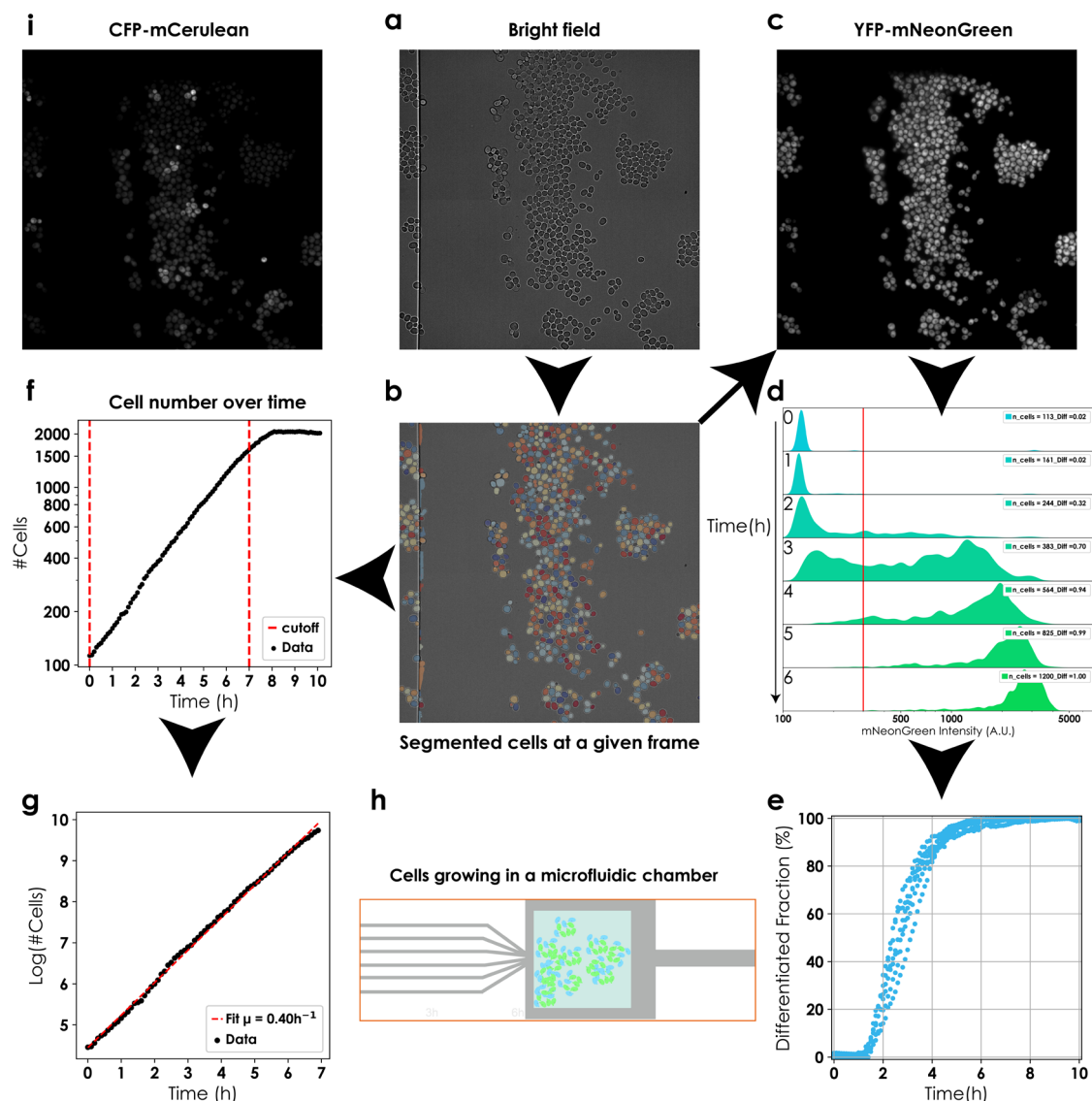


Figure 6.23. Segmentation and growth in the microfluidic plate. SegMator was used for segmentation of brightfield images and subsequent image analysis (Fox et al. 2021). Numbers of cells, over time, in a field of view, were computed by counting the number of segments in each frame. Growth rates were computed by fitting a line to $\log(\# \text{cells})$ as a function of time. Mean growth rate \pm s.d. across 8 fields of view over 2 experiments was found to be $0.394 \text{ h}^{-1} \pm 0.029$. Median cellular fluorescence was calculated from segmented images, and fluorescence in the YFP channel was used to distinguish between differentiated and non-differentiated cells. Cells possessing more than 300 units of median fluorescence (A.U.) were considered differentiated. This differentiation fraction was calculated at each frame to illuminate differentiation dynamics, **e**.

References

- Bai Flagfeldt, Dongmei, Verena Siewers, Le Huang, and Jens Nielsen. 2009. "Characterization of Chromosomal Integration Sites for Heterologous Gene Expression in *Saccharomyces Cerevisiae*." *Yeast* 26 (10): 545–51.
- Bertaux, François, Sebastián Sosa Carrillo, Achille Fraisse, Chetan Aditya, Mariela Furstenheim, and Gregory Batt. 2020. "Enhancing Multi-Bioreactor Platforms for Automated Measurements and Reactive Experiment Control." *BioRxiv*.
- Branch, Mary Ann, Thomas F Coleman, and Yuying Li. 1999. "A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems." *SIAM Journal on Scientific Computing* 21 (1): 1–23.
- Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet. 2008. "A One Pot, One Step, Precision Cloning Method with High Throughput Capability." *PLoS One* 3 (11): e3647.
- Fox, Zachary R, Steven Fletcher, Achille Fraisse, Chetan Aditya, Sebastián Sosa-Carrillo, Sébastien Gilles, François Bertaux, Jakob Ruess, and Gregory Batt. 2021. "MicroMator: Open and Flexible Software for Reactive Microscopy." *BioRxiv*.
- Gietz, R Daniel, and Robin A Woods. 2002. "Transformation of Yeast by Lithium Acetate/Single-Stranded Carrier DNA/Polyethylene Glycol Method." *Methods in Enzymology* 350: 87–96.
- Lambert, Talley J. 2019. "FPbase: A Community-Editable Fluorescent Protein Database." *Nature Methods* 16 (4): 277–78.
- Lee, Michael E., William C. DeLoache, Bernardo Cervantes, and John E. Dueber. 2015. "A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly." *ACS Synthetic Biology* 4 (9): 975–86.

Chapter 7

CONCLUSION AND PERSPECTIVES

"The owl of Minerva spreads its wings only with the falling of the dusk."
Hegel

Summary

Microbial consortia are widespread in nature and co-exist together by exploiting the principle of division of labour. Engineering artificial microbial consortia is an attractive endeavor being undertaken at length, however, the dynamic control of such engineered consortia remain woefully limited.

During my thesis, I tackled this problem by designing a light tunable differentiation system in yeast and found it capable of generating microbial consortia with custom composition in space and in time. The differentiation system relies on an optogenetically inducible expression of Cre-recombinase by the EL222 transcription factor. In presence of a recombination cassette, this enables me to switch expression in the population from one protein to another. I characterized the system in small scale liquid culture (cells growing in a microfluidic chamber), larger scale liquid cultures (batch and continuous), and short term solid cultures (monolayer in μ lbidi slide) and found it to be functional despite changes in the context. These characterization experiments helped me establish that the system possesses several desirable features like low background activity in the dark, unprecedented efficiency in light for yeast, graded response to varying stimuli, and fast, reproducible, and tunable dynamics. These features enabled me to develop a predictive model that was used in a MPC framework to control consortia composition in time. The graded response of the system as well as absence of hysteresis, in particular, were crucial to control consortia composition dynamically over long periods of time (3-4 days). The high efficiency and low leakage permitted me to generate spatially structured microbial consortia simply by illumination of a pattern, which, when coupled to a DMD, afforded single cell control of differentiation. This latter was exploited to demonstrate control of lineage trees in colonies of yeast growing in a microfluidic chamber. Furthermore, including multiple recombination cassettes in the same strain allowed me to generate complex multi-species consortia. I demonstrated that the number and prevalence of species in such consortia could be controlled by implementing distinct differentiation programs. This could be achieved by tuning of the

relative length of the respective recombination cassettes while the absolute length provided a handle on the overall differentiation rate.

The system was then used to probe the growth production tradeoffs during heterologous expression in continuous operation by employing it to decouple growth from production. Growth production tradeoffs arise as a consequence metabolic burden of heterologous expression and pose a major challenge for continuous bioproduction due to emergence of nonproducing mutants resulting in decrease in yield over time. To avoid these tradeoffs, I coupled the differentiation system with growth arrest and production such that the cells growth arrest (via over expression of FAR1M) upon differentiation (GAuDi) and start producing a protein of interest (mScarlet driven by the strong orthogonal transcription factor ATAF1). The coupling was not straightforward and required me to troubleshoot various aspects of the system. After significant tinkering, I was able to create a strain, GAuDi02, which growth arrested upon light induction and started producing POI. Cells arrested as large unbudded cells and continued to increase in size until they died. The circuit, however, was not stable and differentiated cells lost the growth arrest phenotype after 15-20h of complete arrest. I established with targeted genome sequencing and experiments that this escape was associated with the loss of transcriptional cassette expressing FAR1M. Nevertheless, I developed a dedicated ODE model to be able to predict the population dynamics and fitted it to dynamic experimental data. The ODE model was sufficient to predict the population dynamics for diverse light inputs. We deployed the ODE model in an MPC framework to control the differentiation level of the population and by extension the growth rate of the culture. We were able to maintain stable levels of differentiated fraction in continuous cultures of GAuDi02 and other associated strains. In parallel, I developed reference constitutive and inducible platforms of heterologous expression to compare the performance of the GAuDi system. I found that a single target driven by optoATAF1 resulted in up to 10 fold higher mean cellular fluorescence over pTDH3 reference. Curiously adding another copy of the same gene did not result in any noticeable increase in the fluorescence levels, however, adding a different target gene increased the expression to >20 times higher than pTDH3 reference. I also found that the growth rate decreased as a consequence of ATAF1 expression ($32\% \pm 2\%$). Oddly, this decrease was not dependent on heterologous expression levels meaning that cells could support 20 fold pTDH3 expression without it affecting the growth rate (any further). Lastly, I found non-monotonic behaviour in two target versions of optoATAF1 with respect to induction. I characterized the production of the POI for both the fully induced optoATAF1, and GAuDi strains maintained at different levels of differentiation in continuous cultures. I compared the performance to a pTDH3 reference with respect to standard metrics (titers, productivity, and yield). I found that the GAuDi02 produced up to two fold more protein

than the pTDH3 reference but was not genetically stable and the productivity decreased over time for high targets of differentiation fraction. GAuDi04, which showed a milder growth defect, was more amenable to long term control and outperformed the pTDH3 reference all around. However, the differentiation system was found to be, at best, as good as the induction system. The induction system with a single copy, maintained 4 fold higher productivity, was genetically stable and, overall, yielded 5 times as much protein per ml of medium used compared to constitutive expression. Interestingly, despite having twice the amount of cellular fluorescence, the two copy optoATAF1 performed worse than the single copy. Lastly, I found that yield and productivity could not be maximized at the same time demonstrating it experimentally for the first time in continuous production in yeast. Overall, I am compelled to conclude that our approach with a strong growth arrest is unfeasible for heterologous expression in continuous cultures but introducing a milder growth defect seems to be more promising.

None of the constructs described hence far displayed selection effects and were free of hysteresis, a simple ODE model was sufficient to capture the dynamics and permitted dynamic control of consortia composition. However, I wondered if other instantiations might require the characterization and development of more nuanced models that will be able to capture selection effects and the ensuing hysteresis. To this end, I created a strain that possessed observable readout for EL222 levels and its activity in addition to the two color recombination cassette (tetra reporter). Specifically, I created three versions of the tetra reporter (integrated, centromeric and 2-micron) and characterized them in continuous cultures. I found that the system behaviour for the integrated strain was similar to the original differentiation system albeit the efficiency was reduced. This implied that a simple ODE model was sufficient to capture the dynamics. However, for the plasmid versions, this was not the case and necessitated development of a stochastic model of the differentiation system in addition to a stochastic model of plasmid copy number fluctuations. For the former, we used the characterization data from the integrated strain and this allowed us model the dynamics of the integrated strain quantitatively. To extend it to the plasmid versions, we had to first develop a model of plasmid copy number fluctuations in the population. We used the characterization data from the plasmid strains to parameterize the model. We found that this model predicted the consequences of plasmid fluctuations well. Therefore, we built a composed model to predict population dynamics arising from the 2-micron strain. We found that the composed model led to remarkably accurate predictions for the dynamics of the differentiated fraction in addition to quantitative agreement with the stationary distribution of EL222 in the population. Furthermore, the model was able to capture the dynamics of distribution of EL222 in non-differentiated cells. Lastly, the model predicted that it will be possible to apply light in such a fashion so as to maintain the non-differentiated population at

lower levels of EL222 (constitutive expression,) and, by extension, plasmid copy numbers, compared to the whole population by continuously and selectively removing cells via differentiation. We found that data was in qualitative agreement with the model predictions demonstrating that, as unintuitive as it may sound, it is possible to control constitutive expression.

Contributions

In Chapter 3, I describe the construction, characterization and modeling of an optogenetic differentiation system in yeast that allows for creation of tailored microbial consortia with custom functionalities and predictable dynamics rendering its composition controllable in space and in time independent of cell density of the culture. This constitutes as the major contribution of the thesis.

In Chapter 4, I undertook a systematic comparison of different modes of heterologous expression in continuous operation. In doing so, two new platforms for heterologous expression were developed and characterized, an optogenetically inducible platform (optoATAF1) and a differentiation based platform (GAuDi), and both outperformed pTDH3 based constitutive expression.

Chapter 5 details application of the differentiation system to understand the interplay between single cell and population level processes to shape population behaviour. Reaching a mechanistic understanding necessitated construction and characterization of the tetra reporter in three different genetic contexts. Stochastic models describing the differentiation system and the plasmid dynamics, when composed together, enabled parameter free prediction of the population dynamics arising out of the 2-micron tetra reporter and control of constitutive gene expression (plasmid copy number) at the population level.

The following is a list of specific contributions that have emerged out of the work done for the thesis.

The differentiation system

Most efficient optogenetically inducible recombinase characterized in yeast to date

Practical and versatile tool to create and spatiotemporally control multi-species microbial consortia in,

- small (batch) and medium (continuous) scale liquid cultures
- very small scale liquid cultures (μ fluidic devices)
- solid cultures

independently of the cell density

First demonstration of optogenetic control of microbial consortia

First report of dynamic control of microbial consortia in yeast

Continuous heterologous expression

OptoATAF1- light inducible platform

- Results in up to 20 fold higher expression levels at the single cell level when compared to pTDH3 driven expression
- As an optogenetically inducible platform for heterologous expression, outperforms constitutive expression (4 fold higher production and 5 fold higher yield)

GAuDi- light tunable differentiation based platform

- Results in up to 8 fold higher expression levels at the single cell level when compared to pTDH3 driven expression
- As a novel differentiation based platform for heterologous expression, GAuDi outperforms constitutive expression (4 fold higher production and 4 fold higher yield)

First experimental evidence for the yield-productivity pareto frontier in continuous heterologous expression in yeast

Understanding the interplay of single-cell and population processes

Characterization of the consequences of variability on circuit function

Parameter free prediction of circuit function upon change in genetic context

Control of constitutive expression (plasmid copy number)

Shortcomings

The work done here presents with several shortcomings pertaining to various aspects of the work.

In what follows I try to describe these shortcomings categorically. I have included a suggestion (in *Italics*) to improve or solve the shortcoming wherever applicable.

Design of the differentiation system

When the differentiation system is scaled up to give rise to multiple species consortia, light based recombination remains the sole handle to differentiate new species from the non-differentiated cells. While tuning of the relative length of the recombination cassettes allows some degree of control, it is not sufficient to be applicable in some metabolic engineering applications where precise control of metabolic flux is desired.

The system described here can be used with another optogenetically driven recombinase system that is triggered with a different coloured light. In this fashion, it is possible to increase the number of species to 16, by having three orthogonal recombinases (Lin et al. 2015; Nern et al. 2011; W. Liu et al. 2018) being driven by three differently coloured light signals, the limiting factor being the number of characterized orthogonal light induction systems (Benzinger and Khammash 2018; Hochrein et al. 2017; Schmidl et al. 2014). A preprint (Jang et al. 2019) reported development of a library of optogenetic tools derived from cyanobacteriochromes (Fushimi and Narikawa 2019) shown to be functional in yeast and operating in green, orange, red, and far-red range.

Furthermore, the system is perhaps too efficient. This high efficiency results in inadvertent recombination due to ambient light. It is a huge pain to work in red light all the time. It would be preferable to have a system that retains the high efficiency but is less sensitive to ambient light.

This can be addressed, perhaps, by shifting the expression of EL222 to a weak constitutive promoter and including a positive feedback loop on EL222 driving itself at another locus. I expect this to result in less sensitivity to transient ambient light. In terms of efficiency, this would introduce a delay in the time required to reach ~100% differentiation but will eventually get there.

Characterization of the differentiation system

While it is great that the system has been thoroughly characterized in different experimental contexts for a specific length of the to-be-excised fraction, it would have been desirable to characterize the effect of the length of the to-be-excised fraction, at least in liquid cultures, if not in all contexts.

Another thing that remains missing from this work is the characterization of the effects of loci of integration of the recombination cassette on the functioning of the system.

The efficiency of the recombination should have been estimated more quantitatively.

For example, by moving the recombination cassette to a 2 μ plasmid and quantifying the efficiency per copy of cassette.

Control of microbial consortia

The control could have been demonstrated for cultures maintained at different cell densities, as it stands one can only conclude that the cell densities chosen were the only ones that worked (to be completely fair, I did perform experiments with different cell densities and managed to control consortia but these were with a strain with questionable genotype).

Spatial control

Spatial control on a larger scale has not been demonstrated like for example on agar plates.

Design of GAuDi

Design of GAuDi strains could have been more straightforward. Specifically, FAR1-22 should have been expressed without the mCerulean tag. However, I note, experiments with galactose inducible versions showed similar degree of growth arrest with or without the tag and therefore the tagged version was chosen because it could be used at the same time to observe the nuclear physiology.

The positive feedback loop on ATAF1 was leaky and led to the activation of the growth arrest module in a subset of the population. It would have been preferable to avoid such inelegancies.

This could have been achieved by moving ATAF1 downstream of the differentiation event such that another transcription factor (JUB1 (Naseri et al. 2017), for example) is expressed after differentiation, which then drives ATAF1.

The idea of avoiding genetic instability by completely arresting the growth seems ill conceived in retrospect and should have been questioned more systematically at the beginning of the PhD.

Characterization of GAuDi and optoATAF1

This work completely ignores the question of investigating the physiology of the cells that are under the growth arrest. No measurements of DNA content, RNA content, and protein content have been made (except the ones made via cytometry) and is truly a waste of what could have been a rich data set.

No effort has been made to analyze the live cell imaging experiments of cells expressing FAR1 at different strengths in a microfluidic chamber in relation to studying the nuclear architecture in the presence and absence of growth arrest.

Experiments with optoATAF1 driving different copies of fluorescent proteins lacked proper controls and were performed only once. More specifically, all the fluorescent proteins should have been added in two copies to the optoATAF1 strain with and without feedback.

The functioning of the feedback loop should have been independently verified by performing RT-qPCR or western blot.

The characterization of the optoATAF1 strains, in general, has been very poor. It would have been desirable to observe the dynamics of induction and growth rate under different light inputs, preferably in a systematic manner similar to EL222 induction study.

OptoATAF1 targets should have been expressed from plasmids to better understand the buffering mechanism hinted at by the experiments shown in the thesis.

Tetra reporter design and characterization

EL222 tagged with mVenus at the C terminal is less efficient than the untagged EL222. Tagging at the N-terminal (it is unintuitive but that is what Benzinger and Khammash did) should also have been checked for activity.

(I actually did make the constructs and the activity appears to be similar but the data is very low quality because induction was done in batch and the timepoints are not consistent).

Only the plasmids that carry the URA3 auxotrophic marker were explored in the thesis. It would have been worthwhile to verify the results on at least one other auxotrophic marker.

(I actually did make the constructs but could not perform the experiments in the turbidostat).

Lack of replicates

ReacSight is an exceptional platform that is capable of producing very high quality data for long term experiments with minimal human intervention and this is great. ReacSight was in continuous development over the course of the PhD and this necessitated adaptations to experimental protocols. Running automated parallel bioreactor experiments lasting several days with cytometry time points every 1-2 hours is a challenge and not all experiments work flawlessly as one can imagine. A large proportion of my experiments failed and majority of these failed due to errors in software or hardware. These included but were not restricted to, OD control malfunction, cytometry malfunction, pump malfunction, and heating malfunction. Some errors were hard to exactly pinpoint. In short, while the generated data is of very high quality when the platform behaves, there is a significant chance that something would go wrong during the experiment, especially when it is a long term experiment (4-5 days). This, in part, contributed to the lack of replicates for several experiments.

The robustness of the platform might benefit in long term by cataloguing encountered errors especially unresolved mysterious bugs and performing frequent quality checks.

Perspectives

In this section, I talk about some general perspectives that are illuminated by the work. For improvements to the system and additional experiments see Shortcomings and chapter-specific discussions.

Application of the differentiation system as a tool for creating microbial consortia

This work was carried out with the principal application to generate and dynamically control microbial consortia in mind. In this respect, the system presented here possesses several desirable characteristics that make it a lucrative tool.

- **Bioproduction and metabolic engineering**

The most straightforward application of the system is in generating microbial consortia that compartmentalize metabolic flux in the population, which can be achieved by simply replacing the fluorescent proteins used in the original differentiation system by transcription factors that drive entire pathways. Plant based orthogonal transcription factors are good candidates for this application due to the large dynamic range afforded by TF specific promoters of different strengths. In this way, metabolic burden can be distributed and the flux compartmentalized.

It may equally well be employed to decouple growth and production to improve heterologous expression. One realization of the latter was explored in this work and found to be suboptimal. Other realizations, particularly those that arrest cell growth, in addition to cell division, could be worth investigating.

Lastly, coupling the differentiation system to an apoptotic module could serve as an alternative to secretion to obtain proteins in solution. Furthermore, if ATAF1 is used to drive the expression and the apoptosis could be programmed to occur 10h after differentiation, yield may be increased several fold over traditional inducible methods.

- **Bioremediation**

The differentiation system may be used to generate microbial consortia that cooperatively degrade plastic. The approach can be taken one step further by replacing the light induction of recombinase by introducing sensor motifs that in cue to specific environmental cues, for example presence of petroleum products, differentiate cells into “cleaners”. This approach can be multiplexed to prepare strains that are capable of “cleaning” a variety of environmentally undesirable waste products.

- **Microbiome engineering**

The applicability of the system could be translated into therapeutics given the operational yeast is changed to *S. boulardii*. As an example, one could engineer a version of the differentiation system that expresses huge quantities of a bacteria specific toxin (or antitumor drug) upon differentiation and triggers apoptosis when the expression saturates to release the toxin and eliminate drug resistant GI infections. These could be enhanced by replacing light by pH sensitive promoters or engineered motifs that are triggered in the presence of specific bacteria (or (pre-) cancerous cells).

Application of the optoATAF1 platform to determine limits of heterologous expression in yeast

The results described in this work suggest that optoATAF1 driven expression can lead to 20 fold higher protein levels compared to pTDH3 expression, amounting to roughly 46% of the yeast proteome, making optoATAF1 a suitable platform to study the consequences of high heterologous expression that competes with the native machinery for resources. In addition, optoATAF1 strains carrying target genes on 2-micron plasmids can be used to elucidate gene dosage in yeast, characterize the cost of heterologous expression and find the limits to which individual yeast cells can be pushed to produce a heterologous protein.

Application of the optogenetic recombinase as a tool for studying multicellular systems

Owing to high efficiency and low leakage, the differentiation system could be used to study multicellular systems where spatial organization is known to play a key role and perturb them locally.

- **Spatial organization in microbial communities**

Based on the experiments conducted in this thesis, it is clear that the differentiation system can be used to control microbial consortia composition in space. It can, therefore, be used to study how community dynamics are influenced by the spatial organization if the differentiated and non-differentiated species are engineered to interact. Such interactions have previously been engineered in *E. coli* (Kong et al. 2018). Since we are not restricted to patterns attainable in nature (Basu et al. 2005), the system provides a unique opportunity to study how equilibria are reached in multispecies ecosystems and elucidate microbial interactions in complex spatially structured communities.

- **Development of externally controllable organoids**

While the work done in the thesis does not present any evidence to show that the system remains functional in mammalian cells, one can argue that, since both EL222 and Cre are functional, the EL222 inducible Cre should also be functional. If this logic is to be trusted, then, the differentiation system with the capacity to optogenetically control cell fate decisions with spatiotemporal precision has the potential to become a critical tool for dissecting signalling pathways (Toettcher et al. 2011) or understanding developmental programs (Johnson et al. 2020). Taken together, this may allow the development of externally controllable organoids or other multicellular systems.

- **Perturbation of mammalian developmental programs in ex-utero roller cultures**

The results obtained for single cell control of recombination under the microscope are encouraging and indicate that the resolution of the spatial control achievable with this system is sufficiently high to wonder if it can be used to perturb developmental processes in higher mammals in conjunction with ex-utero cultures (Aguilera-Castrejon et al. 2021). More concretely, EL222 could be regulated such that it is only active in a subpopulation of cells (Andersson et al. 2014) or during a specific time (Gifford et al. 2013) in development. Light may then be used to effect a genetic change that perturbs the developmental program in cell type specific and spatially or temporally localized fashion.

References

- Aguilera-Castrejon, Alejandro, Bernardo Oldak, Tom Shani, Nadir Ghanem, Chen Itzkovich, Sharon Slomovich, Shadi Tarazi, et al. 2021. "Ex Utero Mouse Embryogenesis from Pre-Gastrulation to Late Organogenesis." *Nature*, 1–6.
- Andersson, Robin, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, et al. 2014. "An Atlas of Active Enhancers across Human Cell Types and Tissues." *Nature* 507 (7493): 455–61.
- Basu, Subhayu, Yoram Gerchman, Cynthia H Collins, Frances H Arnold, and Ron Weiss. 2005. "A Synthetic Multicellular System for Programmed Pattern Formation." *Nature* 434 (7037): 1130–34.
- Benzinger, Dirk, and Mustafa Khammash. 2018. "Pulsatile Inputs Achieve Tunable Attenuation of Gene Expression Variability and Graded Multi-Gene Regulation." *Nature Communications* 9 (1): 1–10.
- Fushimi, Keiji, and Rei Narikawa. 2019. "Cyanobacteriochromes: Photoreceptors Covering the Entire UV-to-Visible Spectrum." *Current Opinion in Structural Biology* 57: 39–46.
- Gifford, Casey A, Michael J Ziller, Hongcang Gu, Cole Trapnell, Julie Donaghey, Alexander Tsankov, Alex K Shalek, et al. 2013. "Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells." *Cell* 153 (5): 1149–63.
- Hochrein, Lena, Fabian Machens, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2017. "PhiReX: A Programmable and Red Light-Regulated Protein Expression Switch for Yeast." *Nucleic Acids Research* 45 (15): 9193–9205.
- Jang, Jaewan, Sherin McDonald, Maruti Uppalapati, and G Andrew Woolley. 2019. "Green, Orange, Red, and Far-Red Optogenetic Tools Derived from Cyanobacteriochromes." *BioRxiv*, 769422.
- Johnson, Heath E, Nareg J V Djabrayan, Stanislav Y Shvartsman, and Jared E Toettcher. 2020. "Optogenetic Rescue of a Patterning Mutant." *Current Biology* 30 (17): 3414–24.
- Kong, Wentao, David R Meldgin, James J Collins, and Ting Lu. 2018. "Designing Microbial Consortia with Defined Social Interactions." *Nature Chemical Biology* 14 (8): 821–29.
- Lin, Qihui, Hao Qi, Yi Wu, and Yingjin Yuan. 2015. "Robust Orthogonal Recombination System for Versatile Genomic Elements Rearrangement in Yeast *Saccharomyces Cerevisiae*." *Scientific Reports* 5 (1): 1–8.
- Liu, Wei, Zhouqing Luo, Yun Wang, Nhan T Pham, Laura Tuck, Irene Pérez-Pi, Longying Liu, et al. 2018. "Rapid Pathway Prototyping and Engineering Using in Vitro and in Vivo Synthetic Genome SCRaMBLE-in Methods." *Nature Communications* 9 (1): 1–12.
- Naseri, Gita, Salma Balazadeh, Fabian Machens, Iman Kamranfar, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2017. "Plant-Derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces Cerevisiae*." *ACS Synthetic Biology* 6 (9): 1742–56.
- Nern, A., B. D. Pfeiffer, K. Svoboda, and G. M. Rubin. 2011. "Multiple New Site-Specific Recombinases for Use in Manipulating Animal Genomes." *Proceedings of the National Academy of Sciences* 108 (34): 14198–203.
- Schmidl, Sebastian R, Ravi U Sheth, Andrew Wu, and Jeffrey J Tabor. 2014. "Refactoring and Optimization of Light-Switchable *Escherichia Coli* Two-Component Systems." *ACS Synthetic Biology* 3 (11): 820–31.
- Toettcher, Jared E, Delquin Gong, Wendell A Lim, and Orion D Weiner. 2011. "Light-Based Feedback for Controlling Intracellular Signaling Dynamics." *Nature Methods* 8 (10): 837–39.

Annex 1

SINGLE CELL CONTROL OF DIFFERENTIATION

Preface

Portions of this chapter as well as some figures have been adapted from Fox et al. (2021).

Contribution

Strain L3-17 (Figure A1.1a), owing to its fast recombination dynamics, was used to probe control of differentiation at the single cell level. Briefly, prior to recombination the cells express EL222 TF from pTDH3-LoxP. Cre is under the control pEL222 bs 5X Gal1. Blue light triggers the expression of Cre. Cre in turn effects the excision of EL222 TF CDS thus switching expression from EL222 to ATAF1 TF. ATAF1 is an orthogonal transcription factor that drives the expression of FAR1-22-mCerulean (FAR1M). FAR1M arrests the growth.

Cells were grown to exponential phase in the dark and loaded in a microfluidic chamber. A ring-like pattern was shone on the cells in the field of view for 1s every 6 minutes (**Figure A1.1b**) using a digital micromirror device (DMD) (**Figure 6.16**). Cells within the pattern received the light signal. Due to flow and cellular growth, some cells that were initially inside the pattern, moved outside and vice-versa. Despite this low but substantial flux, we observed that recombination predominantly took place within the pattern (**Figure A1.1c**). We note that some cells were erroneously targeted for recombination because of tracking issues, and that only a few cells have not shown the recombined phenotype at the end of the experiment despite having being effectively targeted for recombination (**Figure A1.1d**).

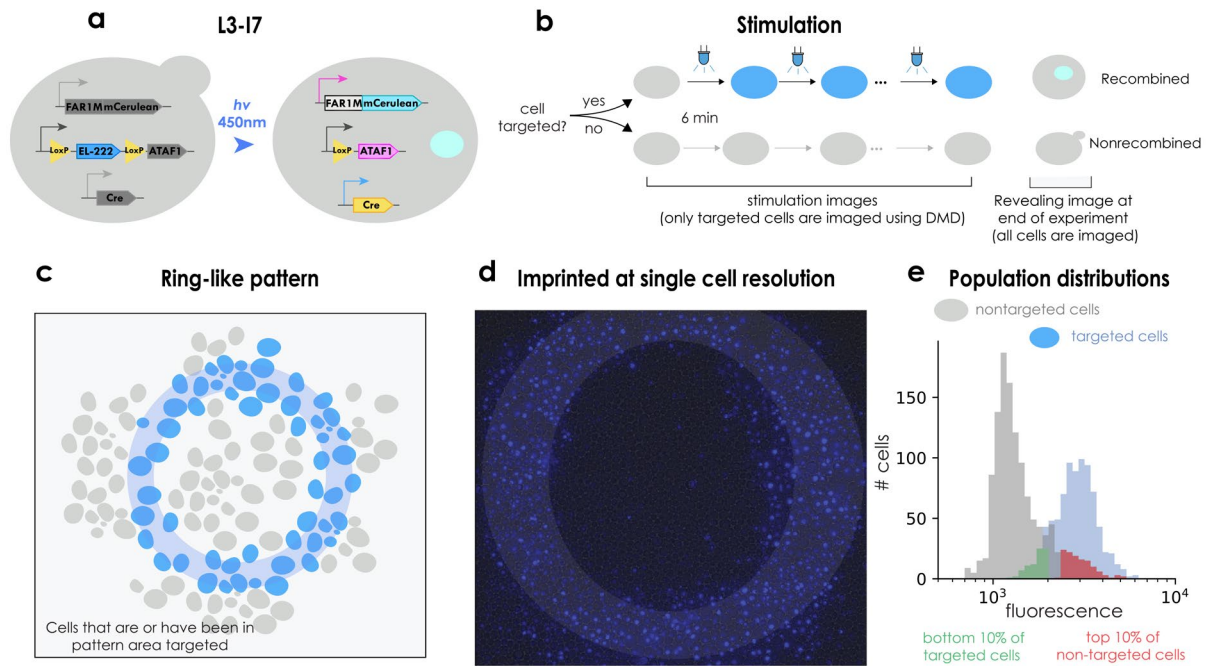


Figure A1.1. Pattern formation in yeast at the single cell level. **a.** Upon light exposure, the Cre recombinase is expressed and triggers recombination, leading to the expression of ATAF1 and then of mCerulean-Far1M. Stars indicate nuclear localization of the protein. **b.** Targeted cells are stimulated for 1 second every 6 minutes until the end of the experiment. Fluorescence levels emitted by targeted cells can be recorded. At the end of the experiment, all cells are imaged and a recombined or non-recombined phenotype is attributed. **c.** A ring-like region in the field of view is selected at the beginning of the experiment and all cells entering the designated region at some time point are targeted for recombination. **e.** The distributions of the fluorescence levels of the targeted and non-targeted cells can be computed at the end of the experiment. The vast majority of cells present the expected phenotype and outliers can be further analyzed.

To exploit our reactive microscopy platform, we tried to create islets of recombined cells. In order to achieve this, we again loaded exponentially growing cells in a microfluidic chamber. Images of cells growing were taken every 6 minutes and segmented in real time using SegMator. We randomly selected half of the cells in the field of view to recombine. To maximize the chances that the chosen cells do recombine, we tracked each chosen cell and targeted it repeatedly with 1s light stimulations (**Figure A1.1b**). However, we observed having two targeted cells close to each other led to undesirable recombination (data not shown). Therefore, we dynamically sought and targeted cells that were far from any previously-targeted cell. (**Figure A1.2a**). We found that this strategy was successful in creating isolated micro colonies of recombined cells without unspecific recombination of non-targeted cells (**Figure A1.2b**). We could reconstruct lineage trees (**Figure A1.2c**) from the time course segmentation data and upon analyzing the lineage trees, we found a clear growth defect in the targeted cells (**Figure A1.2d**) consistent with the phenotype reported for differentiated cells in Chapter 4 (**Figure 4.9**). We note that the previous

works demonstrating optogenetically-driven recombination use static masks for light targeting. Obtaining single-cell resolution as demonstrated in **Figure A1.2b** necessitates real-time image analysis and the use of reactive software.

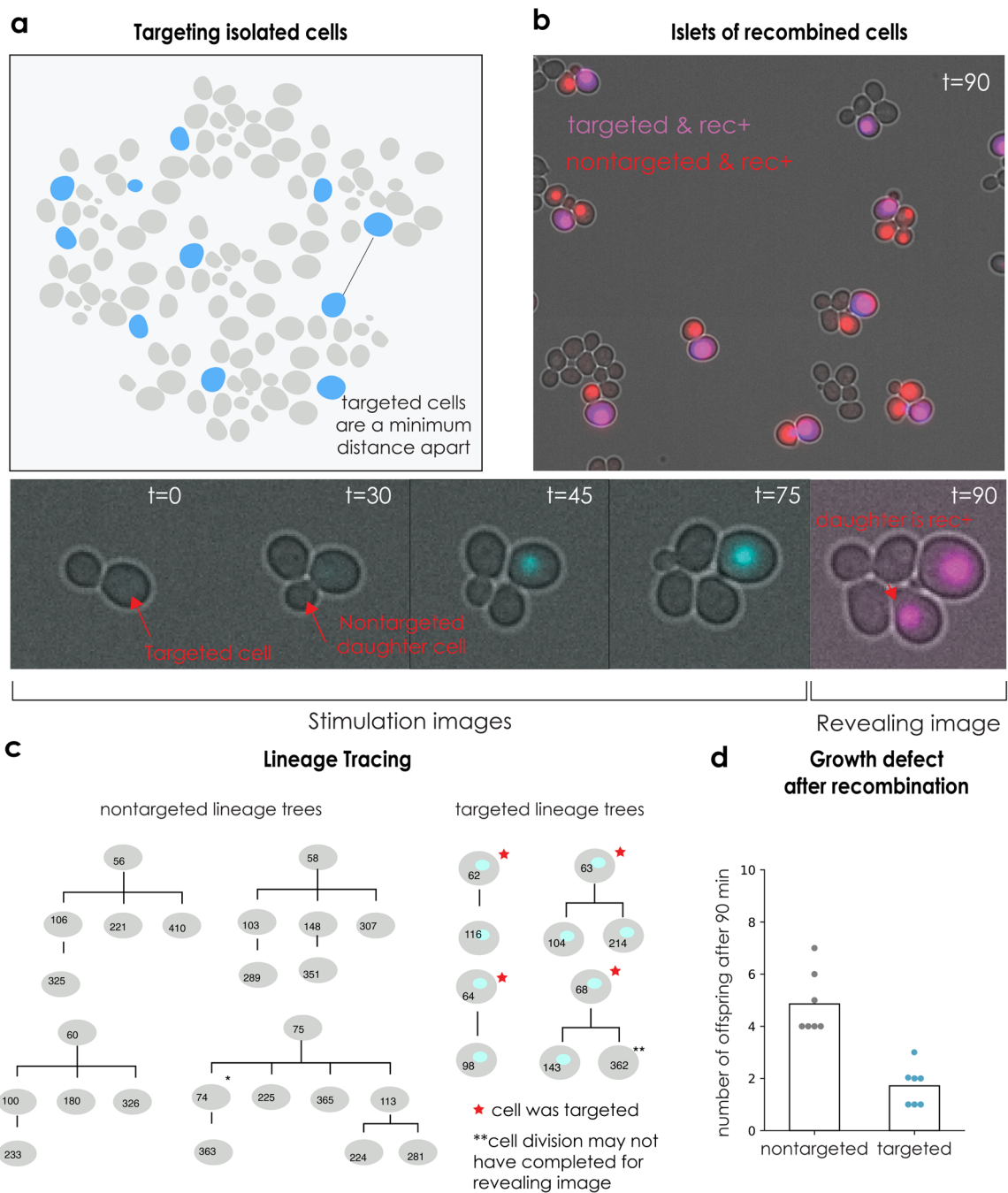


Figure A1.2. Single cell control of recombination in yeast 2D cultures. **a.** Cells are dynamically selected such that no target cells are close to each other. **b.** Image to reveal the targeted recombination **c.** Cell lineages of targeted and non-targeted cells can be manually reconstructed and statistics can be extracted. **d.** Growth rate comparison between targeted and non targeted cells.

Annex 2

FAR1M MEDIATED GROWTH ARREST IS EXPRESSION LEVEL DEPENDENT

I cloned tagged version (mCerulean) of FAR1-22 under the control of a galactose promoter or light inducible (pEL-6x BS) in three different vectors, integrative, centromeric and 2 μ (**Figure A2.1A**) known to be present in 1 copy, 1-2 copies and 5-60 copies respectively. In order to assess the extent of growth arrest in these strains upon induction, they were made to compete with a reference strain in batch, turbidostat and a microfluidic chamber under the microscope (**Figure A2.1B**). The level of growth arrest seemingly depended on the vector of expression of FAR1M (**Figure A2.2**), with the greatest growth defect observed in the high copy vector. Arguably, the cells slowed down considerably, up to 40% of their original growth rate in the case of 2 μ expression vector under pGal (**Figure A2.2**) but the growth arrest was not complete. Under the microscope, I observed that the cells overexpressing FAR1M from an integrated locus under pGal became fluorescent in the nucleus and did not bud. The cell size also increased (**Figure A2.2**). However, the observed growth rate of the reference strain in certain fields of view was up to two-fold less than the WT. This could be due to the observed washout of new buds, which would decrease the apparent growth rate.

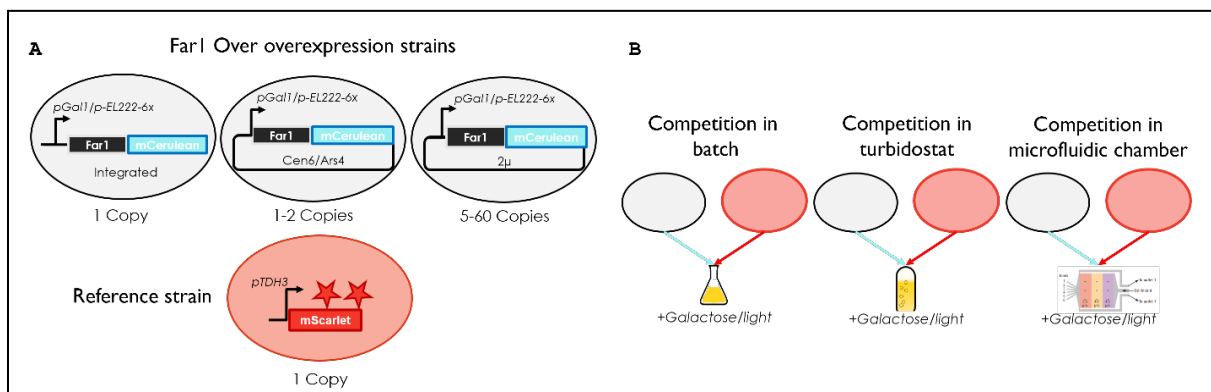


Figure A2.1: Constructions of the optogenetically and galactose inducible FAR1 strains and competition experiment design. A) Strain genotypes. FAR1M was driven either by a galactose inducible promoter or a light inducible (pEL-222 6x BS) promoter. This expression cassette was introduced as an integration (single copy), as centromeric plasmid (1-2 copies) and as a 2 μ plasmid (5-60 copies). **B)** Competition experiments in turbidostat, batch and microfluidic chamber under light were conducted by co-culturing these strains with a reference strain with WT growth rate and constitutive red fluorescence to differentiate them from the FAR1 overexpressing strains. The population was measured at regular intervals via cytometry to get

the fraction of each strain present in the population and the log ratio of these fractions was used to estimate the growth rate of FAR1 overexpressing strains. For galactose induction, pre-culture was done in 2% Raffinose URA- media and 2% Raffinose + 4% galactose media was substituted at the time of induction. For light induction, an intensity of 50 was used. In the microscope, batch LEDs were used instead of the microscope for light induction.

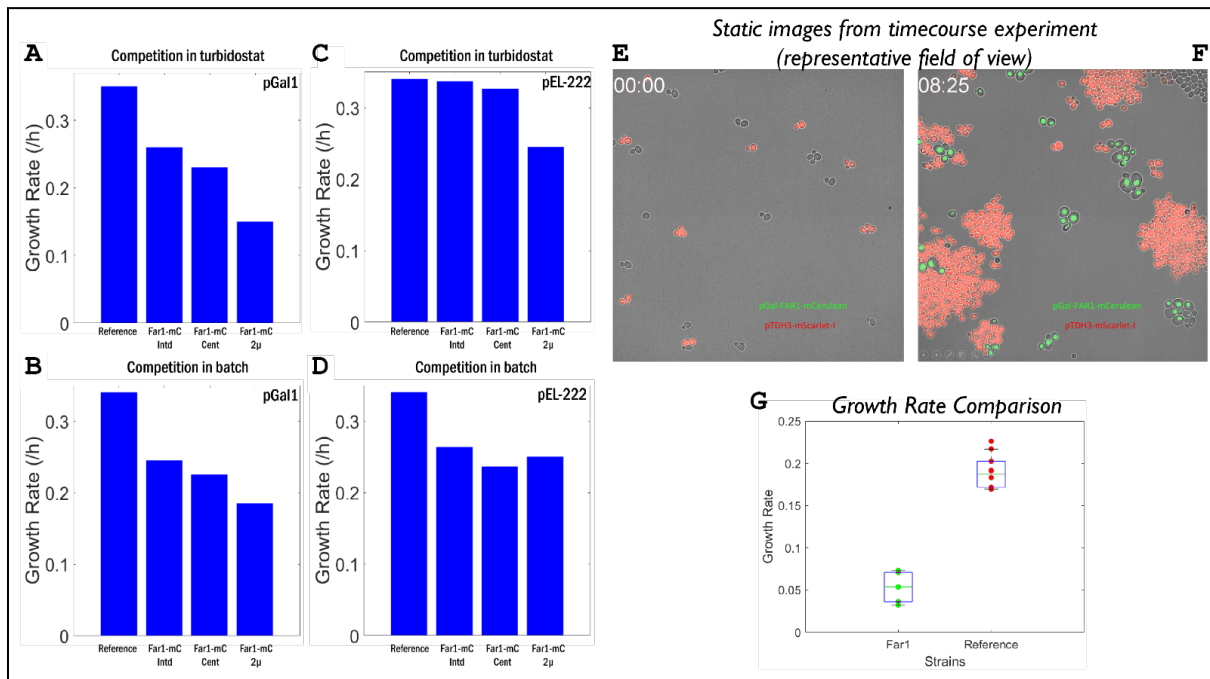


Figure A2.2: Preliminary data on growth arrest via overexpression of FAR1-mCerulean. **A-D)** Cells overexpressing FAR1M via galactose or light were co-cultured with a strain carrying constitutive mScarlet-I in either batch or turbidostat. The culture was monitored by taking cytometry measurements every 2 hours. Growth rate was estimated via the slope of the log ratio of the fractions of two populations over time. In each experiment, a greater degree of growth defect was observed for 2 μ plasmids and pGal promoter. This suggests that FAR1 mediated growth arrest is expression level dependent. **E-G)** Competition experiment in microfluidic chamber with integrated pGal-FAR1M and pTDH3-mScarlet-I. Cells were co-cultured in a microfluidic chamber with 4% galactose induction and the growth rate was calculated by the increase in number of cells over time identified by their fluorescence. Static images at $t = 0$ and $t = 8.25$ from a representative field of view are showed in **E** and **F**. The reference growth rate was 1.5-2 times slower than that observed in turbidostat but this can be explained by the observed washout of new buds. FAR1M expressing cells became fluorescent in the nucleus confirming nuclear localization and arrested as large unbudded or large cells with small buds.

The growth defect was less substantial with strains that expressed FAR1 with optogenetically inducible promoter (EL-222-6x BS-Cyc180) (**Figure A2.2**). This, in addition to vector dependent nature of the growth arrest, was enough to convince me that the growth arrest might increase with FAR1M expression levels. I note that the levels of arrest obtained with gal inducible expression of FAR1M were similar to those obtained by gal inducible expression of FAR1-22.

Annex 3

CURIOSITIES

tMFA2 terminator leads to noisy (almost bimodal) expression upon EL222 induction

tMFA2 is known to destabilize the mRNA (Decker and Parker 1993). To investigate the heterogeneity such a terminator would engender I decided to couple it to EL222 inducible expression.

I cloned a version of the EL222 reporter that contained tMFA2 as a terminator (pEL222-mScarletI-tMFA2) and compared the induction behaviour in the batch with respect to (pEL222-mScarletI-tENO1). In addition to mScarlet-I as the reporter protein, I made constructs that expressed mNeonGreen as the reporter (pEL222-mNeonGreen-X, X stands for tENO1 and tMFA2). All of these constructs were integrated in the Leu2 locus using standard lithium acetate transformation in a strain containing a cassette for the constitutive expression of EL222 integrated in the URA locus. For the first experiment, due to lack of space, I cultured reporters that differed in the fluorescent protein but carried the same terminator together in light (at intensity 40). Induction was started at $t = 0$. Measurements were taken at $t = 0$ and $t = 5h$ (**Figure A3.1**).

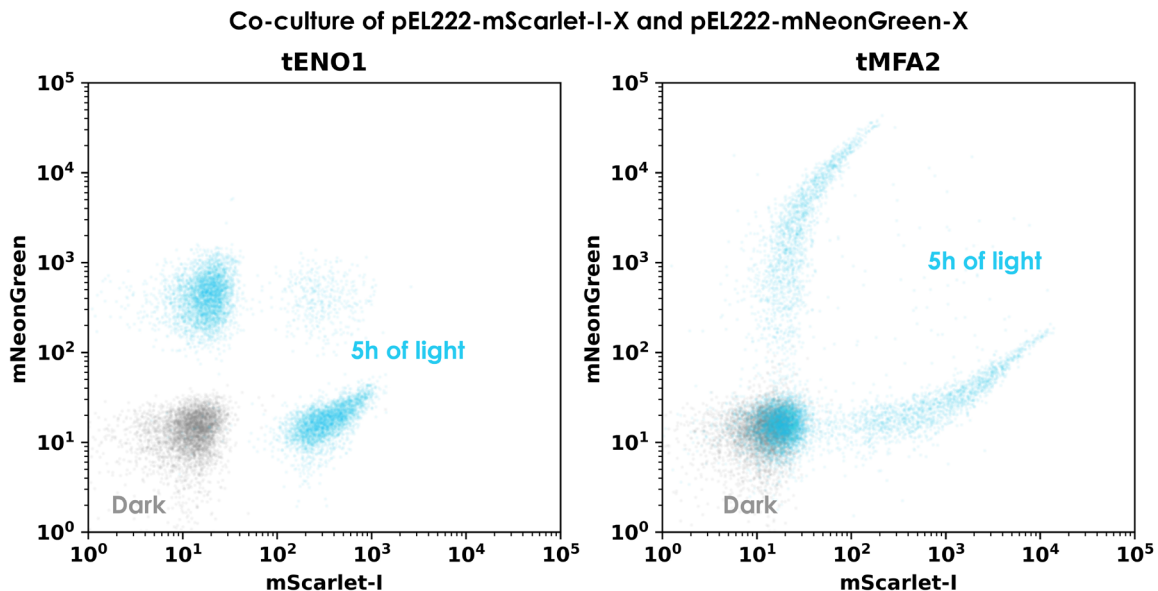


Figure A3.1. Scatter plots of various EL222 reporters in dark (grey) and after 5h of light (blue). Reporters with the same terminator were cultured together. The plotted data is the fluorescence intensity in GRN-B-HLin (mNeonGreen) and ORG-G-HLin (mScarletI). The dark behaviour of the strains is identical, however, tMFA2 leads to much higher noise in expression levels compared to tENO1, upon induction.

I found that expression from tENO1 resembled expression of the EL222 reporter describe in Chapter 3. However, the tMFA2 led to highly variable expression. Notably, some cells reached fluorescence an order of magnitude higher than observed for tENO1 whereas a significant part of the population showed no increase in fluorescence (**Figure A3.1**). Fearing that I had made a mistake in picking the colonies, I repeated the experiment, this time with dark controls for the tMFA2 reporters and replicates. I observed the same behaviour. In order to make sure that the distribution does not change with induction time, I kept the induction going for 22h. I noticed that an increase in the induction period did not result in significant changes in the distribution of the reporter protein (**Figure A3.2**).

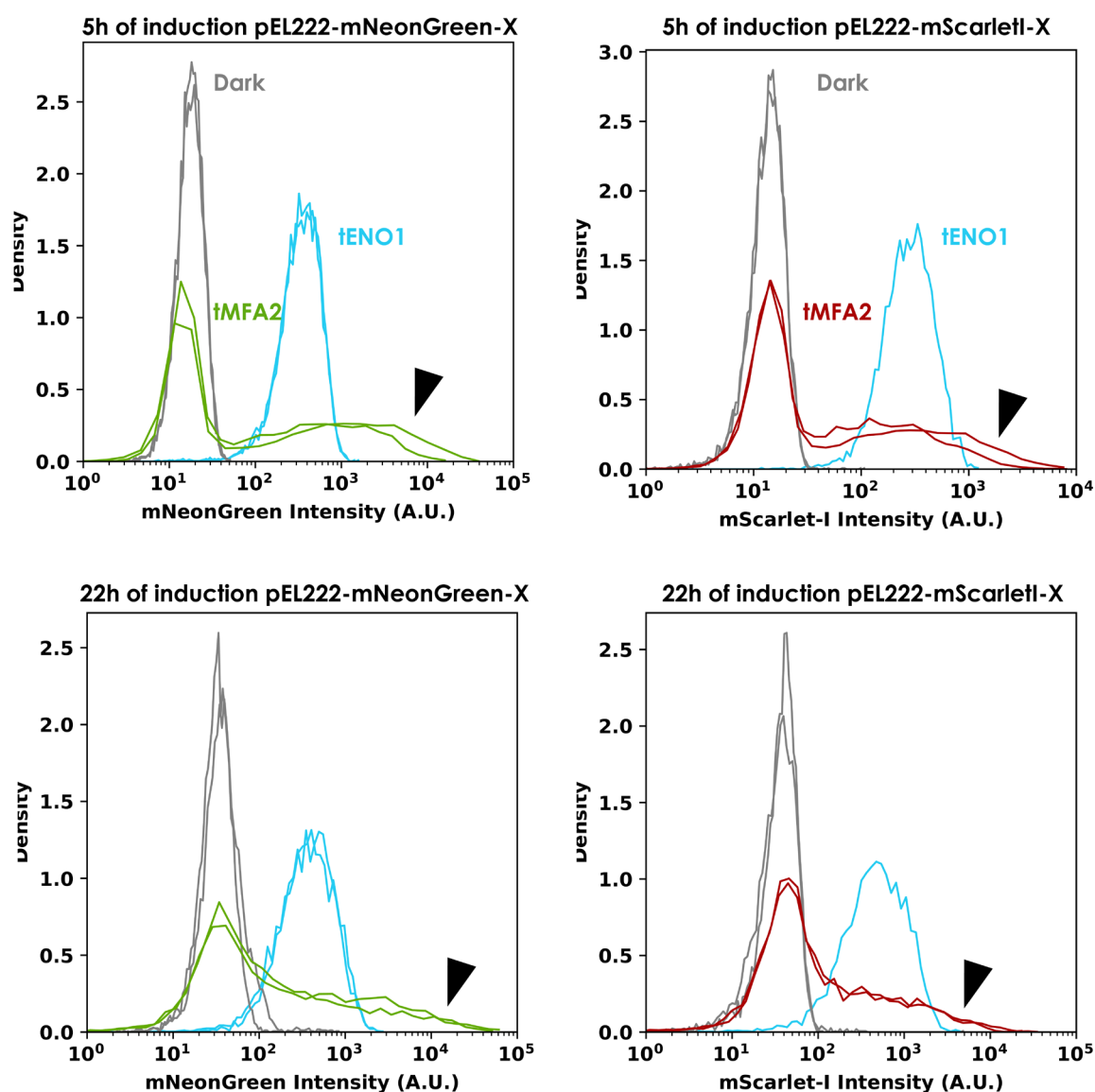


Figure A3.2. Distributions of EL222 reporters. tMFA2 in dark (grey), tMFA2 after 5h (top) & 22h (bottom) of induction (red and green). Corresponding induction levels for tENO1 (blue). All cultures were grown in replicates. Black arrows highlight the long tail of tMFA2. X stands for tENO1 and tMFA2.

A suspicious integrated tetra reporter leads to very fast differentiation

The first tetra reporter construct I made behaved like the dual reporter in full light but I observed that it actually is much more efficient than the dual reporter especially when excited with short pulses of light. I tried to sequence the integrated cassette and found that the part pertaining to the EL222 expression could not be sequenced properly. I postulate that this is due to multiple integrations that took place. I do not know the number of repeats but looking at the fluorescence distribution of mVenus, I am inclined to say that there might be as many as 8 copies of the EL222 expression cassette.

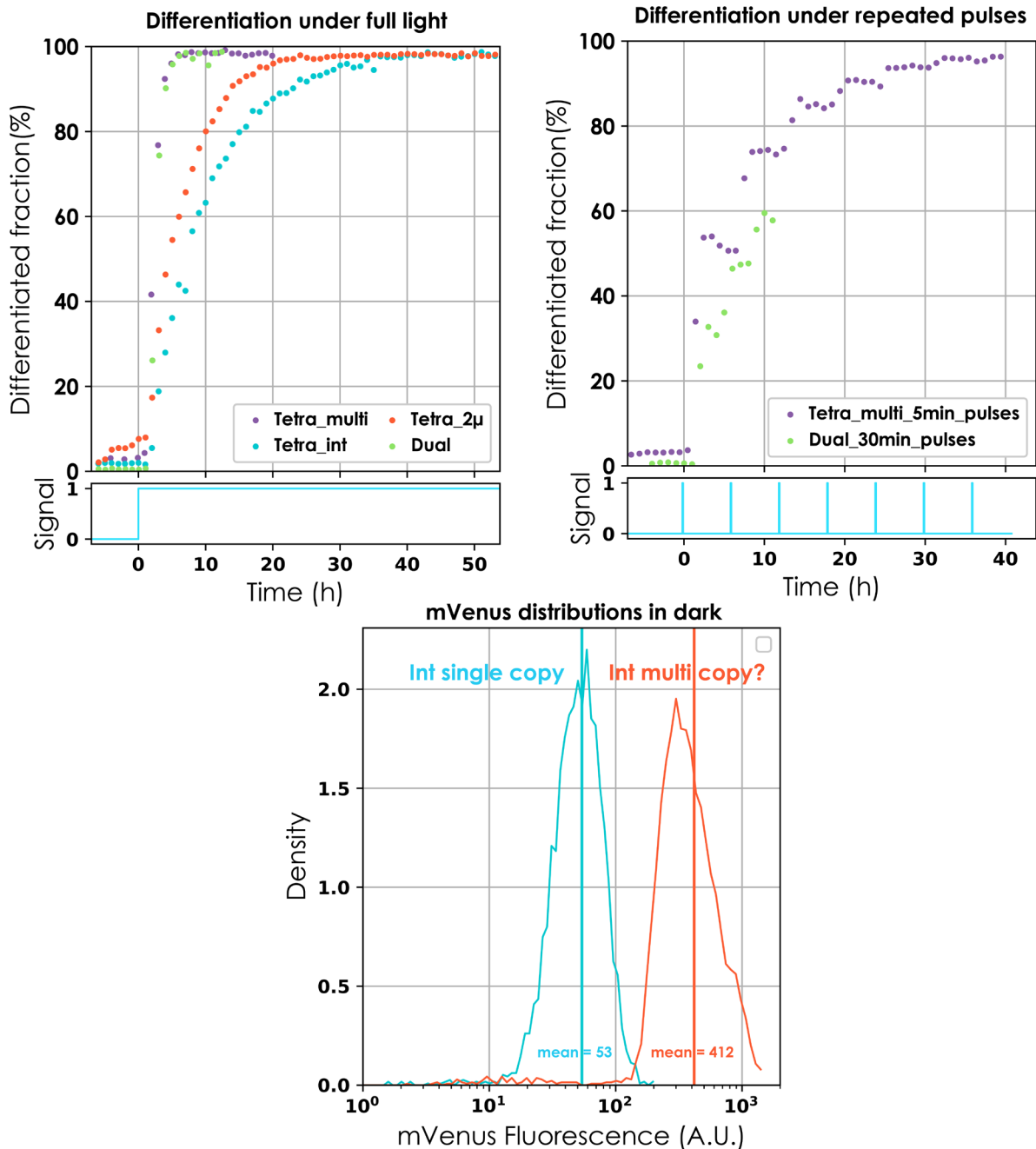


Figure A3.3. Left. Comparison of differentiation behaviour for four strains all with the same recombination cassette but different EL222 expression. Multi_tetra reporter seems as good as the dual reporter (Chapter

3) in full light and much faster than the integrated tetra reporter and the 2μ tetra reporter described in Chapter 5. **Right.** Comparing the behaviour of the Dual reporter and the multi tetra reporter for repeated pulses, I found that 5 minutes pulses (that led to no detectable differentiation for the dual reporter, data not shown) (extremely, Figure A3.4) reproducibly led to ~50% differentiation for the multi tetra reporter. For comparison, I show the behaviour of the dual reporter for 30 minute pulses that still do not lead to 50% differentiation (also see Figure 3.10 & 3.12). **Bottom.** Upon comparing the distributions of mVenus-EL222 in the single copy integrated and the dubious multi-copy strain, I found that the expression was about 8 times higher on average (vertical lines). All cultures were grown in the bioreactor and induced upon reaching exponential phase.

Before I became aware that this strain was aberrant, I thoroughly characterized the differentiation behaviour. I induced the cells growing in the turbidostat with 5-minute pulses at different interpulse intervals and observed remarkable reproducibility despite reactor-to-reactor variability.

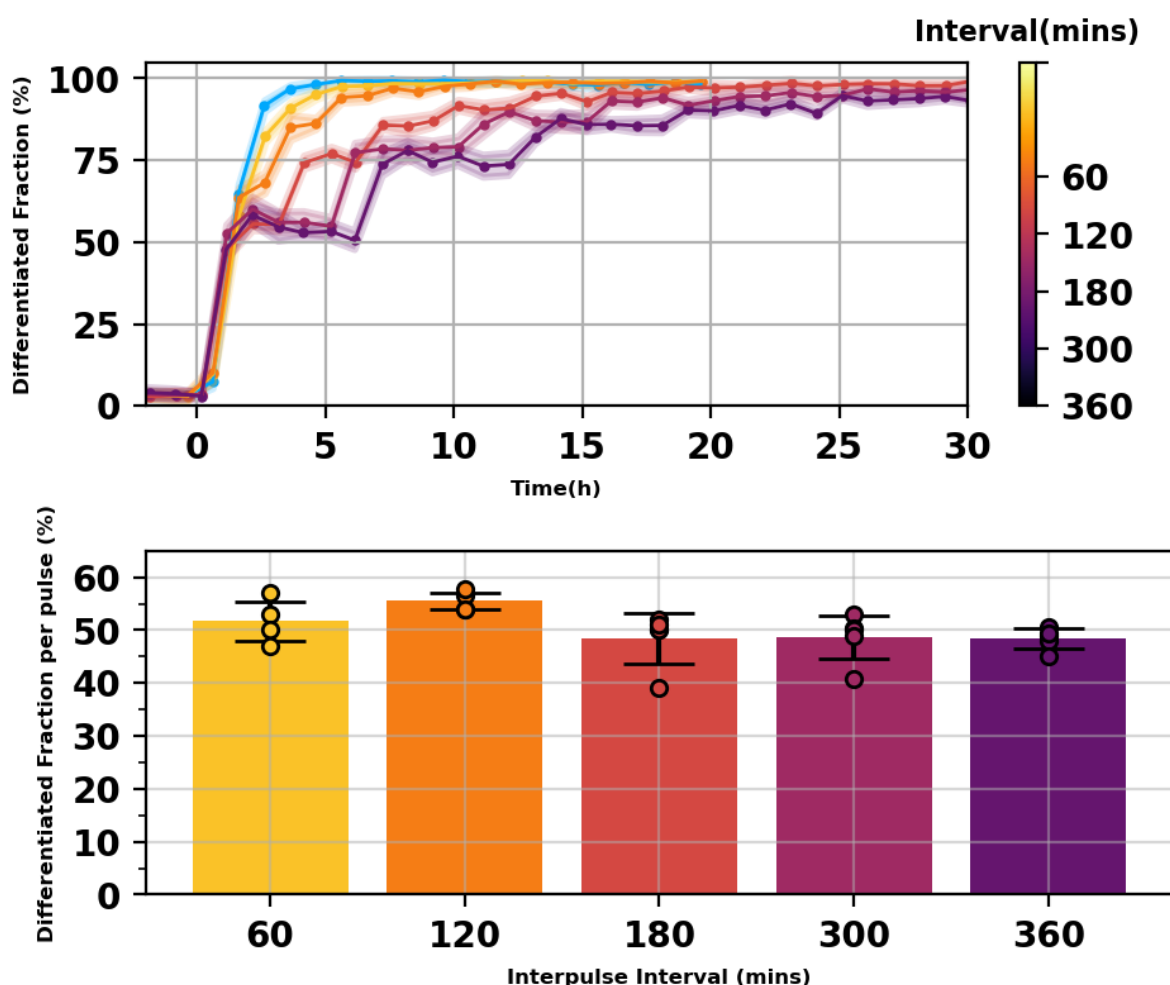


Figure A3.4. Reproducible response to same stimulus (top) and apparent absence of hysteresis (bottom). Cells were cultured continuously in exponential phase with regular cytometry measurements and stimulated with 30-minute light pulse. The pulse was repeated at different interpulse intervals ranging from 60 minutes to 360 minutes. The results show a remarkable tendency of the system to behave predictably with 50% of the cells getting differentiated with each pulse. **(Top)** Dynamics of differentiation for 5-min pulses deliver at different interpulse intervals. The solid lines are means and shaded regions are 95% CI calculated via bootstrapping. **(Bottom)** Bars represent the mean fraction of cells differentiated with each

pulse at a given interpulse interval. Error bars signify standard deviation and individual data points are overlaid on top. There is no statistically significant difference between the different interpulse intervals. Blue represents continuous light.

The absence of a memory effect and high reproducibility between reactors suggests that the high EL222 concentration is enough to counter the effects of variability in induction.

Annex 4: A light tunable differentiation system for the creation and control of consortia in yeast

Chetan Aditya^{1,2,3}, François Bertaux^{2,1}, Gregory Batt^{1,2}, Jakob Ruess^{1,2}

1. Inria Paris, 2 rue Simone Iff, 75012 Paris, France

2. Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France

3. Université de Paris, 85 Boulevard Saint-Germain, 75006 Paris, France

Correspondence: jakob.ruess@inria.fr




ARTICLE



<https://doi.org/10.1038/s41467-021-26129-7>

OPEN

A light tunable differentiation system for the creation and control of consortia in yeast

Chetan Aditya ^{1,2,3}, François Bertaux ^{1,2}, Gregory Batt ^{1,2,4}✉ & Jakob Ruess^{1,2,4}✉

Artificial microbial consortia seek to leverage division-of-labour to optimize function and possess immense potential for bioproduction. Co-culturing approaches, the preferred mode of generating a consortium, remain limited in their ability to give rise to stable consortia having finely tuned compositions. Here, we present an artificial differentiation system in budding yeast capable of generating stable microbial consortia with custom functionalities from a single strain at user-defined composition in space and in time based on optogenetically-driven genetic rewiring. Owing to fast, reproducible, and light-tunable dynamics, our system enables dynamic control of consortia composition in continuous cultures for extended periods. We further demonstrate that our system can be extended in a straightforward manner to give rise to consortia with multiple subpopulations. Our artificial differentiation strategy establishes a novel paradigm for the creation of complex microbial consortia that are simple to implement, precisely controllable, and versatile to use.

¹Inria Paris, 2 rue Simone Iff, 75012 Paris, France. ²Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France. ³Université de Paris, 85 Boulevard Saint-Germain, 75006 Paris, France. ⁴These authors contributed equally: Gregory Batt, Jakob Ruess. ✉email: gregory.batt@inria.fr; jakob.ruess@inria.fr

The evolutionary transition from single cell to multicellular organisms marked a critical turning point in biology¹. Such shift relied on optimizing fitness and productivity through division of labour and specialization^{2,3}. The same principle can be extended to microorganisms living together to form microbial communities or consortia. Engineered microbial consortia hold enormous potential and have been hailed as the next frontier in synthetic biology^{4,5}. Proof of concept studies have concretely established applications in bioproduction^{6,7}, bioremediation^{8,9}, and soil microbiome engineering¹⁰, paving the way for therapeutic applications using human microbiome engineering^{11,12}.

In the context of bioproduction, microbial consortia possess several advantages over traditional monocultures as functional specialization allows metabolic burden to be shared across different species. Diversification thus allows yields to be optimized simply through tuning consortia composition, rather than re-engineering the strain itself¹³. Moreover, by including multiple species, toxic by-products produced by one species can be sequestered and/or metabolized by another, thereby improving the efficiency of the overall process¹⁴. Microbial consortia are typically generated by culturing two or more species together. Such co-culturing approaches rely on various inter-species interactions to ensure the co-existence of different species like mutualism¹⁵, emergent cooperation¹⁶, competitive amensalism¹⁷, and predation¹⁸. Despite considerable advances in our ability to engineer microbial consortia^{6,8,9,19–21} and in our understanding of community interactions^{15,16,18,21,22}, dynamic control of consortium composition remains a key challenge in the field¹⁹. Typically, stable consortia are based on syntrophic or quorum sensing interactions that, albeit being autonomous, remain critically dependent on cell density, thus limiting the applicability for dynamic control. Additionally, scaling the consortium to include more than two species requires non-trivial considerations that may not lead to stable co-existence²⁰. In light of these limitations, an externally controllable differentiation system could be well suited to address this challenge.

In recent years, advances in biological control have come from coupling computers with growing cells carrying the engineered system, made possible by special platforms that integrate biological systems with the computer via a feedback loop^{23–27}. The development of optogenetics, i.e. the use of light to trigger cellular processes, has contributed significantly to control applications by increasing the spatiotemporal resolution of the control signal^{23,24,28–38}. Control of protein expression using light has been demonstrated both at the population level^{23,28,31} and in single cells^{30,33,34,37}. Optogenetics has been used to control cellular processes in other contexts, for instance, signalling dynamics²⁴, morphogenesis³⁶, neuroscience³⁸, bioproduction, and metabolic engineering^{29,35}. However, control of population dynamics using optogenetics in a multispecies environment has not been demonstrated yet.

Here, we present an artificial differentiation system in *S. cerevisiae* capable of generating a microbial consortium composed of functionally different subpopulations emerging from a single population akin to differentiation in multicellular organisms. Concretely, we achieve differentiation into genetically distinct subpopulations—henceforth referred to as species to highlight the analogy to natural microbial consortia—via recombination-based genetic rewiring that can be externally controlled via light. We demonstrate that our system shows desirable features including low background activity, high efficiency for optogenetic recombinases in budding yeast, graded response to varying light signals, absence of hysteresis, and dynamics that are fast, predictable, and tunable. The system reaches >99% differentiation after 4 h of light stimulation and can be stably maintained at any given intermediate level of differentiation for long periods of time (>48 h).

Owing to its fast and predictable dynamics, our differentiation system enables rapid and robust bidirectional control of a microbial consortium arising from a single strain at user-defined compositions in continuous cultures for extended periods in dynamic setups. Coupling the system to a growth arrest module allows us to control population growth rates in continuous culture in different physiological contexts. We show that our system can be extended to give rise to complex multispecies microbial consortia. We engineer two differentiation programmes that can be used to control the total number of species. Finally, we show that our system allows for spatial structuring of microbial consortia by imprinting patterns in 2D cultures with high resolution. To the best of our knowledge, this is the first report of light-driven system for control of a microbial consortium.

Results

An optogenetic synthetic differentiation system in *S. cerevisiae*.

We constructed an optogenetic differentiation system consisting of a blue light-inducible Cre recombinase under the control of a constitutively expressed optogenetic transcription factor, EL222 (refs. 39,40). In order to test the functionality of the system, we designed a recombination cassette composed of a floxed coding sequence (CDS), coding for a fluorescent reporter (mCerulean) that is transcribed constitutively via a pTDH3 promoter upstream of the first LoxP site. Another CDS, coding for a different fluorescent reporter (mNeonGreen), was added downstream of the second LoxP site (Fig. 1a, top and Supplementary Fig. 7).

Prior to differentiation, mCerulean is constitutively expressed and mNeonGreen is not. After light induction, Cre is expressed causing a recombination event leading to the expression of mNeonGreen and loss of mCerulean expression.

Fast and efficient differentiation in light, low background activity in the dark.

In order to evaluate efficiency and background activity, cells harbouring the differentiation system were cultured to the exponential phase in batch and then subjected to either blue light or kept in darkness. Samples were taken at regular intervals and passed through the cytometer (Fig. 1a, bottom). We used flow cytometry data to compute the differentiated fraction by applying a threshold on mNeonGreen fluorescence (Fig. 1b). We observed only a marginal increase in the differentiated fraction after 72 h of culture in the dark suggesting low background activity (Fig. 1c, grey inset). Induction with blue light triggered differentiation. Moreover, the efficiency of the system in budding yeast was superior when compared to existing light inducible systems in *S. cerevisiae*, leading to >99% differentiation after 4 h of induction (Fig. 1c, blue inset) (Supplementary Table 2)^{41–44}. The efficiency of our system allows us to achieve high levels of differentiation with minimal light exposure thus eliminating the risk of phototoxicity (Supplementary Note 3 and Supplementary Fig. 8).

Dynamics of the recombined fraction could be modulated by varying either the light intensity or the duration of applied light pulses. We note that the differentiated fractions are reminiscent of EL222 inducible fluorescent protein levels for varying light intensities or the frequency or duration of applied light pulses (Supplementary Note 3 and Supplementary Figs. 9 and 10)⁴⁰.

To further establish that our system remains functional in different experimental contexts, we cultured cells in a microfluidic chamber and stimulated them periodically via short pulses of light on our microscopy platform³⁰. Through regular imaging, cellular fluorescence was used to classify cells as differentiated (Fig. 1d). Prior to light induction (Fig. 1d, $t = 0$) less than 2% of the cells were differentiated ($n = 817$, over eight fields from two independent experiments) and within 8 h of induction more than

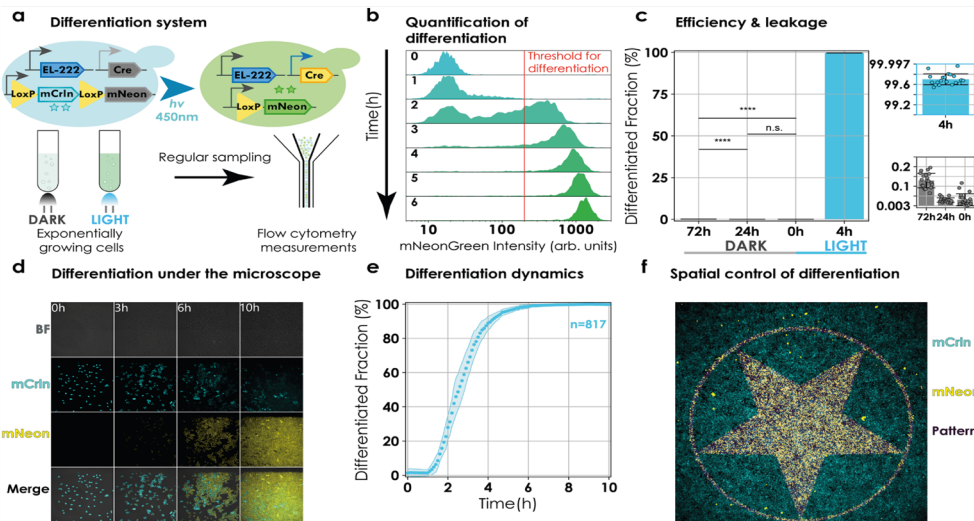


Fig. 1 Design and functional characterization of the differentiation system in liquid and 2D cultures. **a** Design and experimental setup. Cells constitutively express EL222 optogenetic transcription factor. Blue light triggers expression of Cre and recombination, and a change in fluorescence from mCerulean to mNeonGreen. Cells carrying the differentiation system were cultured to exponential phase in batch and induced via LEDs. Cytometry measurements were made by sampling at regular intervals. **b** Quantification of differentiation. Cells are classified as differentiated if cellular mNeonGreen fluorescence exceeds 200 arb. units (red line). The plot shows the evolution of population mNeonGreen fluorescence as a function of time. The threshold was set such that only cells that have expressed detectable amounts of mNeonGreen are classified as differentiated. **c** Efficiency and background differentiation. Twenty colonies were picked and cultured in batch for 72 h in dark or induced at $t = 0$ for 4 h. Measurements were taken at $t = 0$, $t = 24$ h and $t = 72$ h. Bars represent means from a single experiment. Error bars signify standard deviation. Individual data points from colonies are depicted in a scatter plot overlaid on the bar plot (blue and grey insets). Increase in differentiated fraction was not significant (n.s.) after 24 h ($p = 0.41$) but became significant at 72 h ($p = 5.3e-08$; two-sided paired t -test) (black horizontal lines). **d** Snapshots of cell growth and differentiation under the microscope. Images during induction from a representative field of view. Cellular fluorescence changes from mCerulean (cyan) to mNeonGreen (yellow) (Supplementary Movie 1). **e** Differentiation dynamics under the microscope. Images were segmented and analysed. To be deemed differentiated, median cellular fluorescence had to exceed 300 arb. units mNeonGreen fluorescence. Circles represent mean differentiated fraction over eight fields of view from two independent experiments. Shaded region shows standard error of mean. The total number of cells at $t = 0$ summed over all fields of view are given by n ($n = 817$). **f** Imprinting patterns in the population. Cells were allowed to form a monolayer inside a μ bidi slide. A user-defined pattern was illuminated over the monolayer using a digital mirror device (DMD). Merge consists of mCerulean fluorescence (cyan), mNeonGreen fluorescence (yellow), and the pattern (magenta).

99% of the cells in the field of view had differentiated (Fig. 1d, $t = 6$ h and Supplementary Movie 1). The differentiation dynamics were reproducible (Fig. 1e and Supplementary Fig. 6).

practical tool to generate spatially structured heterogeneous microbial communities composed of functionally distinct subpopulations.

Spatial control of differentiation and pattern formation in 2D cultures. To control population composition in space, we grew cells harbouring the differentiation system to the late exponential phase. Cells were then loaded in a μ bidi slide and placed under the microscope. We used our microscopy platform equipped with a digital micromirror device to periodically shine blue light in the shape of different patterns (Supplementary Fig. 20). Cells were illuminated with a given pattern for 1 s every 3 min during 1 h (Supplementary Note 6). Following this, cells were kept in darkness for an hour prior to imaging to ensure a good assessment of the differentiation state of cells (time for the mNeonGreen protein to be produced and mature). We observed that accurate patterns of differentiated cells emerged (Fig. 1f). Some recombination was present outside of the provided pattern, but it is very likely that these cells had differentiated long before the start of the experiment since they lacked mCerulean fluorescence. We conclude that our optogenetic differentiation system enables precise spatial control of differentiation, suggesting that it is a

Characterization of the system and development of a predictive model. To characterize the differentiation behaviour, cells were cultured continuously in the exponential phase in our LED-equipped custom turbidostat platform²⁸ and induced with different light inputs (Fig. 2a). Sampling from the culture and flow cytometry measurements were automated (“Methods”). We observed that, by modulating duration of light pulses (Fig. 2a, top), the system could be regulated to reach intermediate levels of differentiation that were stable over time (>48 h, Fig. 2b), showing that we can stably maintain microbial consortium at different compositions (Fig. 2c). Moreover, these results reveal an interesting dichotomy of our system: it is capable of eliciting a graded response to different stimuli at the population level as well as a differential response to the same stimulus at the single-cell level (Figs. 1d and 2c).

Next, we wondered if hysteresis is present in our system given repeated stimuli. To gauge the extent to which hysteresis affects differentiation dynamics, we induced cultures with repeated

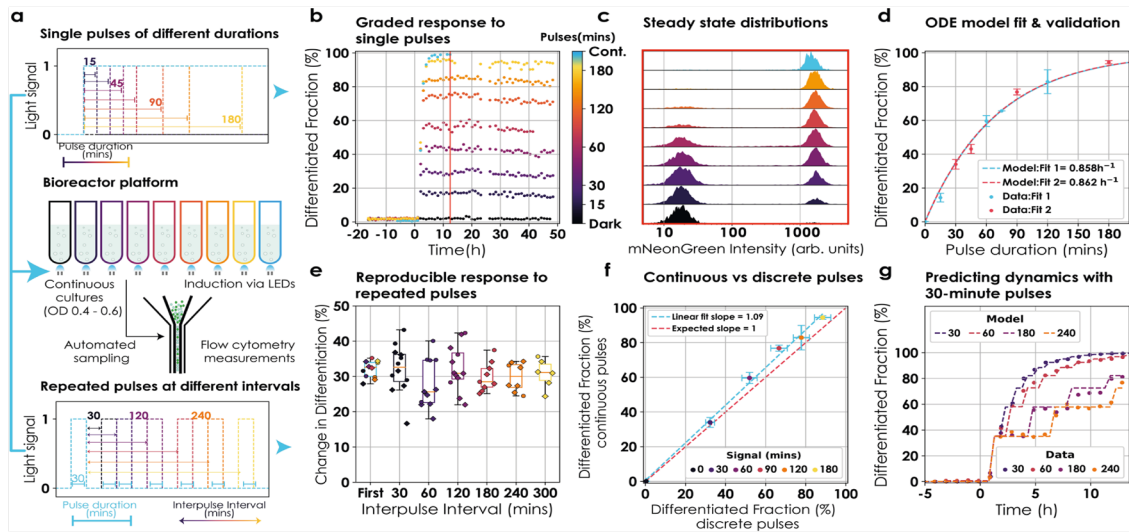


Fig. 2 Characterization and modelling in continuous cultures. **a** Bioreactor platform and induction profiles. Cells were cultured continuously in the exponential phase using our bioreactor platform²⁸. Cultures were induced via LEDs and flow cytometry measurements were automated. Induction was in the form of either single pulses of different durations (top) or repeated pulses of 30 min at different interpulse intervals (bottom; only the first two pulses are represented). **b** Differentiation dynamics after single pulse induction. Cultures were induced at $t = 0$ with single pulses of light ranging between 15 and 180 min (colour bar). Following induction, cultures were kept in the dark for 48 h. Circles represent values from a unique experiment. **c** Tunable population composition. Snapshots of population mNeonGreen fluorescence after induction with single pulses (**b** red line indicates the time of each snapshot). **d** Model fitting and validation. An ODE model was fitted to single pulse induction data. Circles represent mean steady-state differentiation fractions of three independent experiments (except 75-min pulse, unique experiment). Error bars signify s.d. Blue and red circles were independently used to fit the ODE model (dashed lines). **e** Reproducible behaviour with repeated pulses. Cultures were stimulated with 30-min pulses repeated at different interpulse intervals (30–300 min). Circles and diamonds represent the change in differentiation fraction by individual pulses from two independent experiments. Data were collated over the two experiments for boxplots. The colour of circles and boxplots reflects interpulse interval. The first pulse of each experiment was used for the blue boxplot. Lines, boxes, and whiskers denote median, quartiles, and extreme values, respectively. **f** No observable memory effect. Data from (**d** and **e**) were used to compare the efficiency of continuous light vs discrete pulses. Circles represent differentiation effected by continuous pulses (y axis) and equivalent duration of induction in form of 30-min pulses (x-axis). Data are presented as mean values \pm s.d.; $n = 12$ for discrete pulses and $n = 3$ for continuous pulses, where n is the number of pulses delivered. A linear fit of the data is given by the blue dashed line and compared to expected linear behaviour in the absence of memory (red line). **g** Predicting dynamics. Data (Circles) from (**e**) was used to check the predictability of the ODE model (dashed lines). Induction started at $t = 0$. Model predictions were shifted in time to account for observation delay.

pulses at various interpulse durations (Fig. 2a, bottom). If hysteresis were present, we would expect smaller population fractions to recombine with subsequent pulses compared to the first one. The response was reproducible with each pulse resulting in the same differentiated fraction regardless of prior exposure to light (Fig. 2e) (up to small reactor-to-reactor variability). Furthermore, we investigated if time-separated, repeated pulses of a given total duration would result in a higher differentiation fraction compared to a single continuous pulse of same duration. We found that continuous light resulted in similar differentiation fractions as discrete pulses for the same total duration of induction (Fig. 2f).

To be able to predict the differentiated population fraction for a given light input, we developed an ordinary differential equations (ODE) model with a single parameter, the differentiation rate (Supplementary Note 3). This model was fitted to two datasets of steady-state differentiation fractions post induction with mutually exclusive single pulses. Both fits captured observed behaviours well and resulted in similar estimates for the differentiation rate (Fig. 2d). A full description of the model as well as our fitting strategy are presented in the Supplementary Note 3. We also used the model to predict differentiation

dynamics emerging from additional light input sequences and found that model predictions were in good agreement with observed data (Fig. 2g). We conclude that the differentiation dynamics of our system is predictable and can be captured by a simple ODE model.

Robust dynamic control of heterogeneous synthetic communities emerging from a single strain. Having established that our differentiation system exhibits fast and predictable dynamics, we investigated if we could employ it to control consortia composition at user-defined levels in a dynamic setting. We contrived an experimental setup in which two reactors were coupled such that the output of one reactor was connected to the input of the other to be able to dynamically control the population composition despite the absence of growth rate differences between differentiated and non-differentiated cells. The first reactor was kept in the dark as a ‘reservoir’ for non-differentiated cells and the second ‘control’ reactor was exposed to light signals to maintain the culture at a target set point for the differentiated fraction. Cytometry samples were taken every hour to observe the state of the control reactor and adjust the control signal (Fig. 3a). Cultures in

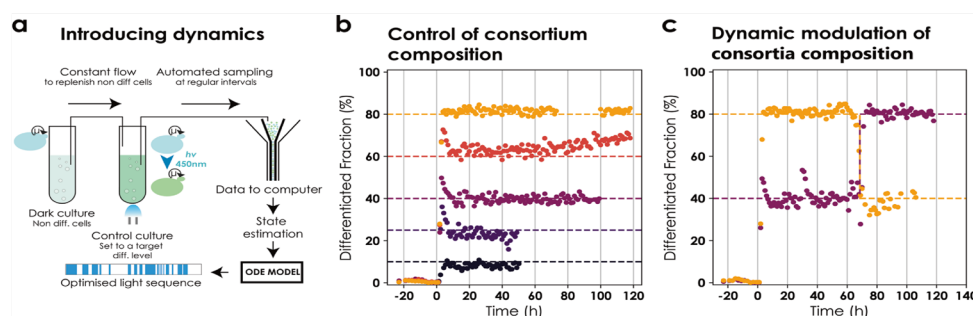


Fig. 3 Dynamic control of population composition in a microbial consortium. **a** Setup for two reactor control. Cells harbouring the original differentiation system were continuously cultured at fixed but different cell densities in two reactors simultaneously. The first reactor was kept in the dark as a reservoir of non-differentiated cells. The output of this vessel was connected to the second ‘control’ reactor. The control culture was set to a target level of differentiation and continuously monitored via automated flow cytometry measurements that were analysed online. The system state was estimated from analysed data and sent to the model predictive control (MPC) module. The MPC module provided an optimized light sequence to maintain the culture at the desired set point (Supplementary Note 5). **b** Control of consortium composition. Cultures were targeted to 10–80% differentiation. Control started at $t = 0$. Circles signify differentiated fractions. Each colour corresponds to a different control experiment and the dashed line reflects the target set point. Note that the figure is composed of independent experiments of different lengths. Data were removed when the OD, either in the reservoir or the control reactor, could not be maintained at the desired target. Light signals are provided in Supplementary Fig. 17. **c** Bidirectional control of consortium composition. Cultures were targeted to 40 and 80% differentiation. Data are represented as in (b). The target was changed at $t = 60$ h to 80% and 40%, respectively.

both the reservoir and control reactors were maintained at constant cell densities (Supplementary Note 5). In addition to the feed from the reservoir reactor, the control reactor was fed fresh media to maintain the culture at the target cell density. Since there was a constant flux of non-differentiated cells from the reservoir to the control reactor, we adjusted the ODE model accordingly.

Using a model predictive control (MPC) framework²⁸ along with the modified ODE model (Supplementary Note 5 and Supplementary Fig. 16), we attempted to maintain user-defined consortia composition (differentiation fractions) in the control reactor. The framework consisted of sampling cells from the control reactor at regular intervals, performing online data analysis to estimate the state of the system (fraction of differentiated cells), and updating the light signal to maintain the fraction of differentiated cells at the desired set point (Supplementary Note 5). We were able to control the population compositions in the control reactor and maintain them for extended periods (up to 96 h) (Fig. 3b) for different target set points ranging from 10 to 80% of differentiated cells. The response was quick and the desired population fractions were reached within approximately 6 h of starting the control. Additionally, our setup enabled us to dynamically modulate the population composition in both directions, i.e., increase or decrease differentiated cells in the population from a given level. Concretely, we targeted two cultures to 40% and 80% differentiated cells in the population, respectively and changed the set points after 60 h to 80% and 40%, respectively (Fig. 3c). The control reactor required active control and lost the desired population composition in the absence of an appropriate light signal.

To eliminate the need for a reservoir of non-differentiated cells, we sought a strategy that would allow the non-differentiated fraction to be replenished over time (non-differentiated cells enrich in absence of light signal). To this end, we coupled our system to a growth arrest module such that cells growth arrest upon differentiation (GAuDi), forcing them to be outcompeted by the non-differentiated cohort after a single exposure to light.

To achieve the coupling, we hijacked the mating factor pathway in *S. cerevisiae* by overexpressing a FAR1 variant after

differentiation⁴⁵. FAR1, a CDK inhibitor, is the downstream effector of the mating factor growth arrest and arrests the cells in G1–S transition by blocking the interaction between CDC28 and G1 cyclins^{46,47}. Briefly, differentiated cells express ATAF1, an orthogonal transcription factor⁴⁸, that, in turn, drives the overexpression of the FAR1 variant. A positive feedback loop on ATAF1 TF leading to higher expression of the FAR1 variant was necessary to obtain effective growth arrest upon differentiation. This came at the cost of higher leakage in the dark. The complete construction of this strain is depicted in Supplementary Fig. 12.

We characterized the GAuDi strain under the microscope and in continuous cultures (Supplementary Note 4, Supplementary Fig. 13 and Supplementary Movie 2) and concluded that our approach allowed us to optogenetically induce growth arrest (Fig. 4b, c) and replenish the non-differentiated fraction in self-contained configuration after transient light induction (Fig. 4b and Supplementary Fig. 14). We observed that under the microscope, growth arrest was complete, and no cell divisions were detected upon differentiation for 15 h (Supplementary Movie 2). Similarly, in the presence of continuous light, the growth rate in liquid cultures decreased dramatically ($<0.04 \text{ h}^{-1}$) (Supplementary Fig. 13). In the turbidostat, we observed that the cultures under continuous light escaped the growth arrest 15–20 h post differentiation. We establish by targeted genome sequencing and culture in selective media that the escape is linked to loss of the integrative cassette (Supplementary Fig. 13). We also observed a subpopulation that possessed neither green nor red fluorescence and were deemed dead. We adapted the ODE model to account for growth arrest and extended it to include cell death and escape. We assumed that differentiated cells, in addition to growing significantly slower, die and escape from the growth arrest at definite rates. The culture growth rate is then given by a weighted average of the growth rates of individual species and is equal to the dilution rate of the reactor at constant cell density. The model was fitted to dynamical data from an experiment with a non-trivial light signal comprised of pulses of varying durations

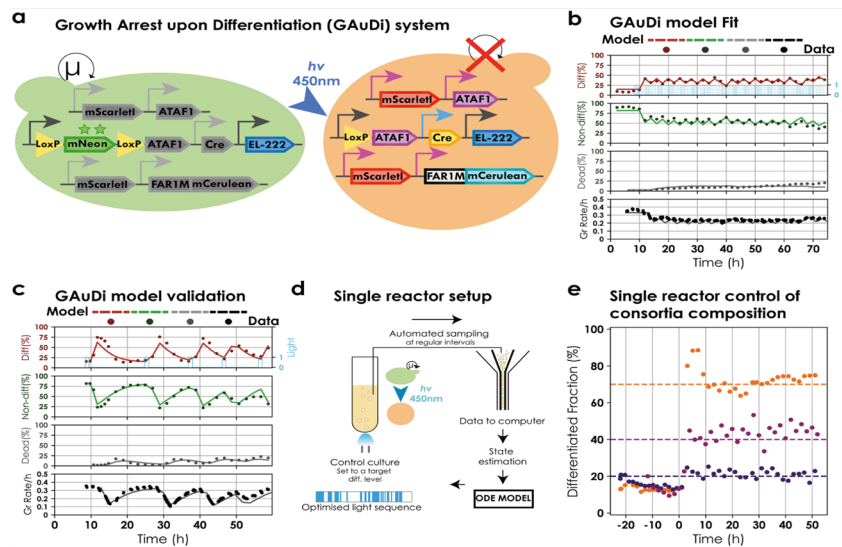


Fig. 4 Characterization and modelling of GAuDi. **a** Design. GAuDi cells constitutively express EL222. Blue light triggers expression of Cre and recombination, and expression of ATAF1 transcription factor. ATAF1 drives the overexpression of a FAR1 mutant (that arrests cells in G1) and mScarlet-I, a “marker” for differentiation. An additional copy of ATAF1 driving itself was necessary for effective arrest. **b** GAuDi ODE model fit. An ODE model for GAuDi was developed phenomenologically. The model was fitted to dynamical data with repeated light stimuli of different durations. Solid circles represent data and solid lines signify model predictions. We note that before $t = 0$, a persistent non-zero differentiation fraction was present possibly due to leakiness in ATAF1 expression in a subset of the population (Supplementary Note 4). **c** GAuDi model validation. The fitted model was validated by predicting dynamic responses to light signals that were not used for the fit. Data are represented as in **(b)**. Globally, model predictions were in good agreement with the data (Supplementary Fig. 15). **d** Setup for single reactor control. Cells harbouring the GAuDi system were continuously cultured in the exponential phase. The culture was set to a target level of differentiation and continuously monitored via automated flow cytometry measurements that were analysed online. The system state was estimated from analysed data and sent to the model predictive control (MPC) module. The MPC module provided an optimized light sequence to maintain the culture at the desired set point (Supplementary Note 5). Due to presence of dead cells, data were filtered and only live cells were used for subsequent analysis. **e** Single reactor, single strain control of a microbial consortium. Cultures were targeted to 25, 40, and 70% differentiation. Control started at $t = 0$ h. Circles signify differentiated fractions. Each colour corresponds to a different control experiment and the dashed line reflects the target set point.

at different intervals (Fig. 4b and Supplementary Note 4). The fitted model was validated by comparing model predictions to data collected in four additional experiments in which cultures were exposed to different light signals (Fig. 4c and Supplementary Fig. 15b). We note that the predictive power of the model is limited when the system operates under strong selection pressure, that is, when light is applied over extended durations.

We used this modified ODE model in conjunction with the MPC framework to demonstrate single reactor control of a microbial consortium originating from a single strain. Concretely, exponentially growing GAuDi cells in the turbidostat were exposed to optimized light signals such that the population composition is maintained at user-defined set points. Flow cytometry measurements were taken at regular intervals to track the fraction of differentiated cells (Fig. 4d). We found that population compositions could be maintained stably for extended periods (up to 48 h) (Fig. 4e). We note that due to genetic stability limitations, it was possible to maintain composition control for longer periods only if the target set point was below 50% differentiation. Evolutionary constraints are known to limit the long-term stability of synthetic circuits⁴⁹ particularly circuits that are engineered to implement growth arrest/self-killing at the population level^{18,50}. To the best of our knowledge, this is the first report of dynamic control of population composition in a

two species artificial microbial consortium arising from a single strain.

Generation of complex microbial consortia with multiple subpopulations. Certain applications, particularly in metabolic engineering, might require microbial consortia composed of more than two species. To probe if our system can be used to engineer differentiation programmes that allow one to create complex multispecies consortia, we cloned two recombination cassettes (denoted by C and N) in a single strain along with the differentiation system (Fig. 5a). Cells were cultured and induced in our bioreactor platform with varied pulses repeated every 360 min. We found that both sites were approximately equally likely to recombine (asynchronous) (Fig. 5b) resulting in a four species microbial consortium, consisting of double recombined cells (CN), single recombined cells (CN, CN), and the original non-recombined population (CN).

Noting that the length of the to-be-excised region plays a critical role in determining the differentiation rate (Supplementary Fig. 21), we set out to test if this could be exploited to modify the differentiation dynamics. We cloned two recombination cassettes of unequal to-be-excised region (C and N') along with the differentiation system (Fig. 5c). We observed that the shorter site recombined first (CN') and subsequent pulses led to

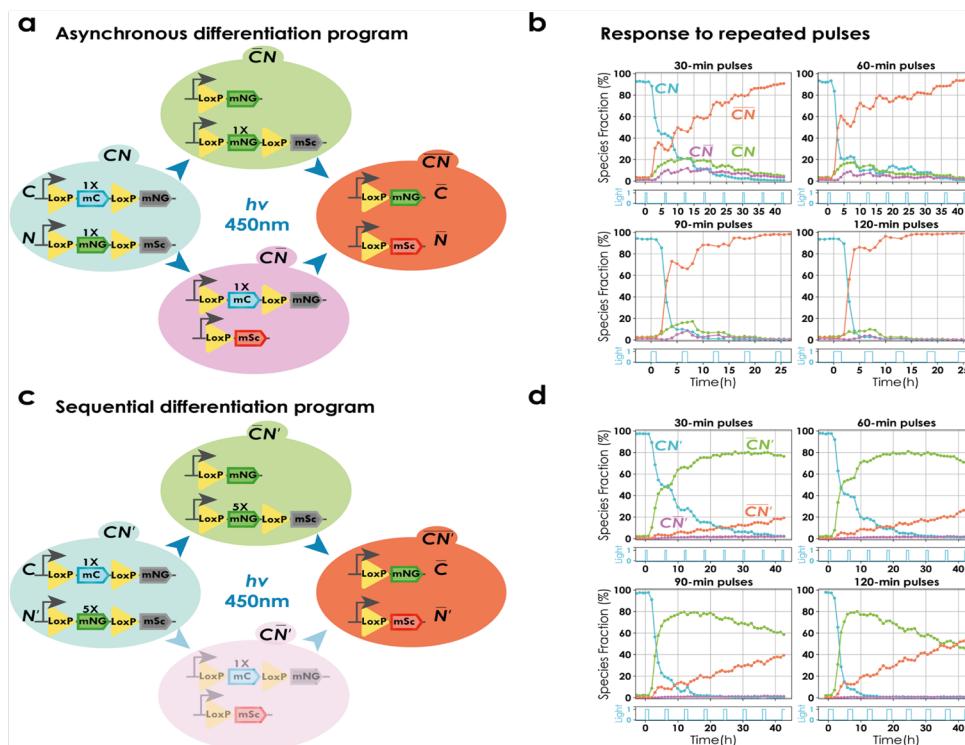


Fig. 5 Engineering differentiation programmes for complex consortia with multiple subpopulations. **a** Asynchronous differentiation programme. Two recombination cassettes (C and N) were introduced along with the differentiation system in the same strain. Both cassettes consisted of to-be-excised region of equal lengths (1× of the original system (Fig. 1a)). All four possible cell types are shown (neither cassettes recombined (CN), single cassette recombined (\overline{CN} & \overline{CN}), and both cassettes recombined (\overline{CN})). Circuits shown in the panel are representational (Supplementary Fig. 22). **b** Response to repeated pulses of different durations. Four cultures were induced with repeated light pulses of different durations, 6 h apart. Each subplot represents data from a single experiment with unique pulse duration given by cyan subplots (clockwise from top right, 30, 60, 90, and 120 min, respectively). Solid lines show dynamics of prevalence of each subpopulation (blue, purple, red, and orange for CN, \overline{CN} , \overline{CN} , and \overline{CN} , respectively). Circles indicate individual time points. Prevalence of each population was calculated using thresholds on mNeonGreen, mScarlet-I, and mCerulean fluorescence (Supplementary Note 7 and Supplementary Fig. 23). All four subpopulations could be observed. We note that the small decrease in the fraction of \overline{CN} cells just after differentiation and a corresponding increase in CN is an artefact of the threshold-based quantification that is limited by the dilution of mNeonGreen. **c** Sequential differentiation programme. Two recombination cassettes were introduced along with the differentiation system in the same strain. Relative length of the to-be-excised region of one cassette (N') was 5× longer than the other (C). All four possible cell types are shown. Notations resemble **a**. Due to increase in the length of to-be-excised region, the differentiation rate of N' was significantly slower than C leading to inappreciable prevalence of $\overline{CN'}$ (faded cell type) (Supplementary Fig. 22). **d** Response to repeated pulses of different lengths. Four cultures were induced with repeated light pulses of different durations, 6 h apart. Each subplot represents data from a single experiment with unique pulse duration (clockwise from top right, 30, 60, 90, and 120 min, respectively). Data are represented as in **(b)**. Three subpopulations could be observed.

enrichment of cells with both sites recombined ($\overline{CN'}$) (Fig. 5d) constituting a sequential differentiation programme as opposed to the asynchronous programme obtained with to-be-excised regions of equal lengths. We did not observe appreciable levels of species with only the longer site recombined ($\overline{CN'}$) effectively resulting in a three species microbial consortium.

Both versions of the system with two recombination sites remained capable of eliciting a graded response to light. These results demonstrate that our differentiation system can be extended to implement distinct differentiation programmes capable of yielding complex consortia composed of a controllable number of species whose prevalence can be optogenetically tuned.

Discussion

Microbial consortia are expected to be of great utility for biotechnology and hold immense potential for diverse applications^{4–11}. However, dynamic control of consortium composition remains relatively unexplored despite being a key challenge in the field¹⁹. In the present study, we address this challenge with the help of an artificial differentiation system in *S. cerevisiae* capable of generating microbial consortia with custom composition. The system is based on blue light inducible expression of Cre recombinase driven by EL222 from a non-leaky promoter⁴⁰. We characterized the system in small-scale liquid culture (cells growing in a microfluidic chamber), larger-scale liquid cultures (batch and continuous), and short-term solid cultures (monolayer

in μ lbi slide) and found it to be functional despite changes in the context. Moreover, we established that it possesses several desirable characteristics: fast, reproducible and tunable dynamics, high efficiency for light inducible recombinases in budding yeast (Supplementary Table 2)^{41–44}, low leakage, and graded response of the population to light (Figs. 1 and 2). The efficiency of our system allowed us to achieve high levels of differentiation with short transient pulses that eliminate the risk of phototoxicity. Graded population responses to light were critical for achieving control of consortia composition (Fig. 3). Moreover, the high degree of reproducibility in response to light stimuli allowed us to develop a predictive model that could be used as a basis to precisely control microbial community dynamics. Using the developed model in an MPC framework allowed us to achieve bidirectional control of consortia composition in a dynamic setup (Fig. 3). We note that Klavins and colleagues⁵¹ have developed another differentiation system in yeast, that, in principle at least, could have been used to generate consortia. This system uses a toggle switch to implement memory, and chemical inducers to toggle the switch. We believe that using light as inducer and a DNA implementation of memory allowed us to precisely characterize and select systems with appropriate properties, and drive them with the needed precision to obtain subpopulations in desired organization and proportions in space and in time, respectively.

Several solutions for the stable maintenance of microbial consortia have been proposed recently. In particular, Hasty and colleagues^{52,53}, Lu and colleagues^{21,54}, and Barnes and colleagues¹⁷ achieved this by using synthetic biology approaches. These authors demonstrate the capacity to maintain co-cultures of several bacterial subpopulations over extended durations. However, none of these approaches succeeded at precisely controlling consortia composition. Moreover, the functioning of these systems relies on the release of signalling molecules in the environment (quorum-sensing molecules or bacteriocin) that trigger cell death. The fact that signalling molecules are released by cells creates de facto a strong dependency of the functioning on growing conditions, and notably on the density of cell cultures, an important aspect for bioproduction applications. Lastly, previous designs use elaborate genetic engineering solutions for the molecular implementation of control mechanisms, thus making extension and scaling up of these designs potentially challenging. Moreover, external control by light is inexpensive and compatible with most media composition. In summary, in comparison to previously existing solutions, our system is simple to implement, quantitatively predictable and actionable, and versatile to use.

The efficiency of the optogenetically inducible recombinase developed in this work exceeds any reported in the literature for optogenetic recombinases in yeast (Supplementary Table 2). Previous optogenetic recombination systems are based on engineering a photoactivable Cre that is typically split into two subdomains tagged with the respective photosensitive heterodimers that can be brought together upon light illumination to form a functional Cre^{41–43}. However, such approaches result in activity loss as functional Cre is a tetramer and the probability of four dimerized split-Cre molecules to assemble together hinges on the relative concentrations of the two subunits⁴⁴. A recent study reported a monogenic photoactivable Cre that is based on fusion of a LOV domain with a destabilized Cre variant⁴⁴. The authors reported higher efficiency and stronger activation when compared to previous systems. This monogenic photoactivable Cre matched the efficiency reported for our system for up to 40 min of induction. After 40 min of induction, the activity seemed to plateau. An optogenetically inducible recombinase has been recently reported in bacteria which uses split-Cre tagged to vivid homodimers⁵⁵. The authors demonstrate the high efficiency and

low leakage of their system at the population level. However, these properties are not quantified at the single-cell level, so a precise comparison of efficiency and leakage is not possible. Chemically inducible recombinase based systems have been employed more prominently in bacteria^{56–58} and show high efficiency (>90%) but graded response or bimodal behaviour has not been reported.

Based on the principle of division of labour, microbial consortia have been employed to increase bioproduction by distributing the metabolic burden. Such approaches necessitate functional specialization in the constituent species of consortia. We provided evidence that our design can be implemented in different physiological contexts by coupling it to a growth arrest module (GAuDi system) to allow optogenetic control of growth rate and consortium composition in self-contained continuous cultures (Fig. 4). A GAuDi-like system has the potential to facilitate the switch to continuous bioproduction, touted to be the future of bioproduction⁵⁹, by separating growth and production across different subpopulations. In the context of metabolic engineering, our system could serve as a pathway switch, with the potential of compartmentalizing metabolic flux in the population. This could be achieved, for instance, by replacing fluorescent proteins by orthogonal TFs that drive entire pathways⁶⁰, leading to division of labour paradigms in consortia engineering and opening up possibilities for population level metabolic engineering.

To show that complex multispecies consortia can be created using our system, we engineered asynchronous or sequential differentiation programmes based on multiple recombination cassettes that extended the core system to generate and stably maintain multispecies consortia from a single strain in continuous liquid cultures (Fig. 5). These programmes could be scaled exponentially for applications requiring dynamic control of complex multispecies consortia and do not require intricate genetic circuits spread over multiple populations to ensure stability.

Finally, the capacity to optogenetically control cell fate decisions with spatiotemporal precision has the potential to become a critical tool for dissecting signalling pathways²⁴ or understanding developmental programmes³⁶. Here, we showed pattern generation in 2D cultures in a microfluidic plate (Fig. 1 and Supplementary Fig. 20). Since we are not restricted to patterns attainable in nature⁶¹, our system can provide a unique tool to study how equilibria are reached in multispecies ecosystems and elucidate microbial interactions in complex spatially structured communities.

In conclusion, we show that the system has highly desirable characteristics making it a practical tool for robustly generating and maintaining functionally distinct subpopulations both in space and in time.

Methods

Cloning: All plasmids were cloned using the Golden Gate method. The majority of the used parts came from the Yeast Tool Kit (YTK)⁶². New parts, whenever necessary, were generated in the laboratory (DNA synthesis, Phusion PCR). The Golden Gate mixture was transformed via heat-shock transformation in thermo-competent *E. coli* cells and plasmids were isolated using standard miniprep kits (Machery & Nagel, and Qiagen). Sequences of integrative plasmids and backbones can be found in Supplementary Data 1. Primers used in the study can be found in Supplementary Table 4.

Yeast strains: All strains used in this study are derived from BY4741 [MATA his3 Δ 1 leu2 Δ 0 met15 Δ 0 ura3 Δ 0]. Cells were transformed with linearized integrative vectors using standard lithium acetate transformation. For selection, common auxotrophic markers uracil, leucine, and histidine were used. Integrative plasmids carrying *LEU2* and *URA3* were integrated at the endogenous loci while those with *HIS3* were integrated at *HO* locus. Cells were grown in standard defined media (Sigma Aldrich Yeast Nitrogen Base) containing 2% glucose and lacking the respective auxotrophic nutrient during selection in plates (Sigma Aldrich, uracil,

leucine and, histidine drop-out media supplements). A list of strains used in this study and their genotypes can be found in Supplementary Table 5.

Cell handling and induction: Cells were grown overnight (ON) before the day of the experiment by picking a single colony in 50 ml Falcon tubes shaking at 200 r.p.m. at 30 °C. Following the ON, a preculture was started by diluting the ON culture 1:50. Depending on the volume of culture required, this was done in either 50 ml Falcon tubes (<15 ml) or 250 ml flat bottom Erlenmeyer flasks (<150 ml). The preculture was allowed to grow for at least 3 h (~2 cell generations). Care was taken to do all the experiments in the exponential phase. Induction was done by using RGB led strips (Adafruit NeoPixel Digital RGB LED Strip). Max LED intensity was set at 40 for all experiments unless specified (max led intensity possible is 255). The control was performed via Arduino microcontrollers (Genuno Uno and Mega). Cells were grown in synthetic complete low fluorescence media with 2% glucose for all experiments. Light-sensitive strains were grown in the dark. All manipulations were performed in the presence of red light (Supplementary Note 8).

Batch culture: Cells were grown to the exponential phase in the dark from ON culture in Falcon tubes shaking at 200 r.p.m. at 30 °C in a custom Falcon tube before starting the experiment. Induction was carried out in a custom Falcon tube holder fitted with LED strips.

Turbidostat: Turbidostat refers to our custom LED-equipped continuous culture platform and control software, ReaSight (Supplementary Note 1 and Supplementary Fig. 1)²⁸. Experiments done in the turbidostat followed a similar protocol as described above (see section on cell handling). Cells were allowed to grow in the dark until exponential phase and induction was started only after the growth rate stabilized. Samples were taken automatically from the turbidostat at regular intervals, diluted 20 times with a pipetting robot and passed through the cytometer. The entire vessel, including pumps and tubing, was autoclaved before each experiment. Unless stated differently, the experiments used a “grow and dilute” programme where cells were allowed to grow until OD 0.6 and then diluted to OD 0.4. The growth rate was computed by calculating the slope of linear curve fit to the log of OD data with time. Information regarding the times when the dilutions took place was stored, in addition to the OD and LED status as csv files. Data were acquired and preprocessed using the ReaSight software framework. Subsequent treatment of the data was done in Python.

Cytometry and data analysis: All cytometry measurements were acquired with a Guava EasyCyte BGV 14HT benchtop cytometer using the InCyte software (version 3.3). The settings were kept constant for all experiments. For turbidostat experiments 5000 events were recorded for each sample, except for leakage and efficiency experiments (Fig. 1c), for which, 50,000 events were recorded. No compensation was used during acquisition. Data were deconvolved after acquisition (Supplementary Figs. 4 and 5). Data were gated using kernel density-based methods (Supplementary Fig. 3). Python was used for data analysis and visualization (Supplementary Note 1).

Microscopy: Microscopy was performed on the inverted microscopy platform Leica DMI8 S. Live cell imaging was either performed in Ibidi μ -slide VI 0.4 (80606) or CellASIC ONIX platform with the microfluidic plates (Y04C) provided by the vendors. The details for the exact excitation and emission spectra of the fluorophores and filters used can be found in Supplementary Table 1. Unless specified differently, the interframe interval was 6 min. During time-lapse live cell imaging, the chamber temperature was maintained at 30 °C. We used an in-house software, called MicroMator, for the automated acquisition and cell tracking. Cell segmentation was achieved via SegMator, an in-house neural net-based segmentation algorithm³⁰. Python was used for data analysis and visualization (Supplementary Note 2). Fiji (ImageJ 1.52i) was used for the processing of images shown in the manuscript.

ODE model fitting and parameter estimation: ODE models were solved using solve_ivp solver from SciPy.integrate library. Models were fitted using the least-square method from SciPy.optimize library. Parameter search for each model was done locally (gradient descent) with multiple initial guesses for parameters and bounds on parameter values between 10^{-10} and 10. To account for the delays, mean squared deviations were calculated after shifting model predictions by 60 min for the differentiation system ODE model and 120 min (differentiated & non-differentiated), 360 min (dead), and 300 min (growth rate) for the GAuDi ODE model.

MPC experiments: ODE models were used in an MPC framework²⁸. The framework consisted of solving the model given a light sequence. This sequence was then optimized using a least-square method from SciPy.optimize library to minimize the error between predictions and the target set point starting from an initial state. The optimization for light sequence was done for a time horizon of 5 h in the form of 10 duty cycles of 30-min period. Cultures were sampled every hour for two reactor control experiments (Fig. 3b, c) and every 2 h for single reactor control experiments (Fig. 3e). The light sequence was updated at each timepoint. To estimate the state of the system, cytometry data were analysed online to determine the state at the time of sampling. This estimate of sampling time state was then used as initial conditions for the model and the current state was estimated by solving the model for a time, τ . τ stands for a delay, consisting of a sampling delay (time passed between sample acquisition and finalized data analysis) and an observation delay (time required for enough fluorescent protein to accumulate to pass the differentiation threshold) (Supplementary Note 5).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Raw cytometry and OD data generated in this study and processed microscopy data have been deposited on Zenodo under the accession code 4923833. Plasmids and strains are available from the corresponding authors upon request.

Code availability

Jupyter notebooks used for data analysis are available online in the GitLab repository (https://gitlab.inria.fr/InBio/Public/YeastOptogeneticDifferentiation_YODA). The notebooks also include code for model simulation and fitting.

Received: 10 June 2021; Accepted: 7 September 2021;

Published online: 05 October 2021

References

1. Maynard Smith, J. & Szathmari, E. *The Major Transitions in Evolution* (Oxford University Press, 1997).
2. Ispolatov, I., Ackermann, M. & Doebeli, M. Division of labour and the evolution of multicellularity. *Proc. R. Soc. B Biol. Sci.* **279**, 1768–1776 (2012).
3. Rueffler, C., Hermisson, J. & Wagner, G. P. Evolution of functional specialization and division of labor. *Proc. Natl Acad. Sci. USA* **109**, E326–E335 (2012).
4. Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26**, 483–489 (2008).
5. Rapp, K. M., Jenkins, J. P. & Betenbaugh, M. J. Partners for life: building microbial consortia for the future. *Curr. Opin. Biotechnol.* **66**, 292–300 (2020).
6. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* **33**, 377–383 (2015).
7. Li, Z., Wang, X. & Zhang, H. Balancing the non-linear rosmarinic acid biosynthetic pathway by modular co-culture engineering. *Metab. Eng.* **54**, 1–11 (2019).
8. Zhuang, W.-Q. et al. Incomplete Wood-Ljungdahl pathway facilitates one-carbon metabolism in organohalide-respiring *Dehalococcoides mccartyi*. *Proc. Natl Acad. Sci. USA* **111**, 6419–6424 (2014).
9. Zhang, X. et al. Bioremediation of petroleum hydrocarbons by alkali-salt-tolerant microbial consortia and their community profiles. *J. Chem. Technol. Biotechnol.* **96**, 809–817 (2021).
10. Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E. & Kao-Kniffin, J. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME J.* **9**, 980–989 (2015).
11. Inda, M. E., Broset, E., Lu, T. K. & de la Fuente-Nunez, C. Emerging frontiers in microbiome engineering. *Trends Immunol.* **40**, 952–973 (2019).
12. Chen, K. et al. A probiotic yeast-based immunotherapy against *Clostridioides difficile* infection. *Sci. Transl. Med.* **12**, eaax4905 (2020).
13. Wu, S. G., He, L., Wang, Q. & Tang, Y. J. An ancient Chinese wisdom for metabolic engineering: Yin-Yang. *Microb. Cell Fact.* **14**, 39 (2015).
14. LaSarre, B., McCully, A. L., Lennon, J. T. & McKinlay, J. B. Microbial mutualism dynamics governed by dose-dependent toxicity of cross-fed nutrients. *ISME J.* **11**, 337–348 (2017).
15. Shou, W., Ram, S. & Vilar, J. M. G. Synthetic cooperation in engineered yeast populations. *Proc. Natl Acad. Sci. USA* **104**, 1877–1882 (2007).
16. Wintermute, E. H. & Silver, P. A. Emergent cooperation in microbial metabolism. *Mol. Syst. Biol.* **6**, 407 (2010).
17. Fedorec, A. J. H., Karkaria, B. D., Sulu, M. & Barnes, C. P. Single strain control of microbial consortia. *Nat. Commun.* **12**, 1977 (2021).
18. Balagaddé, F. K. et al. A synthetic *Escherichia coli* predator-prey ecosystem. *Mol. Syst. Biol.* **4**, 187 (2008).
19. Roell, G. W. et al. Engineering microbial consortia by division of labor. *Microb. Cell Fact.* **18**, 1–11 (2019).
20. Karkaria, B. D., Fedorec, A. J. H. & Barnes, C. P. Automated design of synthetic microbial communities. *Nat. Commun.* **12**, 672 (2021).
21. Kong, W., Meldgin, D. R., Collins, J. J. & Lu, T. Designing microbial consortia with defined social interactions. *Nat. Chem. Biol.* **14**, 821–829 (2018).
22. Kyllis, N., Tuza, Z. A., Stan, G.-B. & Polizzi, K. M. Tools for engineering coordinated system behaviour in synthetic microbial consortia. *Nat. Commun.* **9**, 2677 (2018).
23. Miliadis-Argeitis, A. et al. In silico feedback for in vivo regulation of a gene expression circuit. *Nat. Biotechnol.* **29**, 1114–1116 (2011).

24. Toettcher, J. E., Gong, D., Lim, W. A. & Weiner, O. D. Light-based feedback for controlling intracellular signaling dynamics. *Nat. Methods* **8**, 837–839 (2011).
25. Menolascina, F. et al. In-vivo real-time control of protein expression from endogenous and synthetic gene networks. *PLoS Comput. Biol.* **10**, e1003625 (2014).
26. Lugagne, J.-B. et al. Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nat. Commun.* **8**, 1671 (2017).
27. Carrasco-López, C., García-Echauri, S. A., Kichuk, T. & Avalos, J. L. Optogenetics and biosensors set the stage for metabolic cybergenetics. *Curr. Opin. Biotechnol.* **65**, 296–309 (2020).
28. Bertaux, F. et al. Enhancing multi-bioreactor platforms for automated measurements and reactive experiment control. Preprint at *bioRxiv* <https://doi.org/10.1101/2020.12.27.424467> (2020).
29. Lalwani, M. A. et al. Optogenetic control of the lac operon for bacterial chemical and protein production. *Nat. Chem. Biol.* **17**, 71–79 (2021).
30. Fox, Z. R. et al. MicroMator: Open and flexible software for reactive microscopy. Preprint at *bioRxiv* <https://doi.org/10.1101/2021.03.12.435206> (2021).
31. Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R. & Tabor, J. J. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods* **11**, 449–455 (2014).
32. Milias-Argeitis, A., Rullan, M., Aoki, S. K., Buchmann, P. & Khammash, M. Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nat. Commun.* **7**, 12546 (2016).
33. Chait, R., Ruess, J., Bergmiller, T., Tkačik, G. & Guet, C. C. Shaping bacterial population behavior through computer-interfaced control of individual cells. *Nat. Commun.* **8**, 1535 (2017).
34. Rullan, M., Benzinger, D., Schmidt, G. W., Milias-Argeitis, A. & Khammash, M. An optogenetic platform for real-time, single-cell interrogation of stochastic transcriptional regulation. *Mol. Cell* **70**, 745–756. e6 (2018).
35. Zhao, E. M. et al. Optogenetic regulation of engineered cellular metabolism for microbial chemical production. *Nature* **555**, 683–687 (2018).
36. Johnson, H. E., Djabrayan, N. J. V., Shvartsman, S. Y. & Toettcher, J. E. Optogenetic rescue of a patterning mutant. *Curr. Biol.* **30**, 3414–3424 (2020).
37. Perkins, M. L., Benzinger, D., Arcaik, M. & Khammash, M. Cell-in-the-loop pattern formation with optogenetically emulated cell-to-cell signaling. *Nat. Commun.* **11**, 1355 (2020).
38. Liu, P., Chen, B. & Wang, Z.-W. GABAergic motor neurons bias locomotor decision-making in *C. elegans*. *Nat. Commun.* **11**, 5076 (2020).
39. Motta-Mena, L. B. et al. An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat. Chem. Biol.* **10**, 196–202 (2014).
40. Benzinger, D. & Khammash, M. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat. Commun.* **9**, 3521 (2018).
41. Taslimi, A. et al. Optimized second-generation CRY2-CIB dimerizers and photoactivatable Cre recombinase. *Nat. Chem. Biol.* **12**, 425–430 (2016).
42. Kawano, F., Okazaki, R., Yazawa, M. & Sato, M. A photoactivatable Cre-loxP recombination system for optogenetic genome engineering. *Nat. Chem. Biol.* **12**, 1059–1064 (2016).
43. Hochrein, L., Mitchell, L. A., Schulz, K., Messerschmidt, K. & Mueller-Roeber, B. L-SCRaMBLE as a tool for light-controlled Cre-mediated recombination in yeast. *Nat. Commun.* **9**, 1931 (2018).
44. Duplus-Bottin, H. et al. A single-chain and fast-responding light-inducible Cre recombinase as a novel optogenetic switch. *Elife* **10**, e61268 (2021).
45. Henchoz, S. et al. Phosphorylation-and ubiquitin-dependent degradation of the cyclin-dependent kinase inhibitor Far1p in budding yeast. *Genes Dev.* **11**, 3046–3060 (1997).
46. Chang, F. & Herskowitz, I. Identification of a gene necessary for cell cycle arrest by a negative growth factor of yeast: FAR1 is an inhibitor of a G1 cyclin, CLN2. *Cell* **63**, 999–1011 (1990).
47. McKinney, J. D. & Cross, F. R. FAR1 and the G1 phase specificity of cell cycle arrest by mating factor in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 2509–2516 (1995).
48. Naseri, G. et al. Plant-derived transcription factors for orthologous regulation of gene expression in the yeast *Saccharomyces cerevisiae*. *ACS Synth. Biol.* **6**, 1742–1756 (2017).
49. Castle, S. D., Grierson, C. S. & Gorochofski, T. E. Towards an engineering theory of evolution. *Nat. Commun.* **12**, 3326 (2021).
50. Din, M. O. et al. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature* **536**, 81–85 (2016).
51. Yang, Y., Nemhauser, J. L. & Klavins, E. Synthetic bistability and differentiation in yeast. *ACS Synth. Biol.* **8**, 929–936 (2019).
52. Scott, S. R. et al. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nat. Microbiol.* **2**, 17083 (2017).
53. Miano, A., Liao, M. J. & Hasty, J. Inducible cell-to-cell signaling for tunable dynamics in microbial communities. *Nat. Commun.* **11**, 1–8 (2020).
54. Liu, F. et al. Interaction variability shapes succession of synthetic microbial ecosystems. *Nat. Commun.* **11**, 1193 (2020).
55. Sheets, M. B., Wong, W. W. & Dunlop, M. J. Light-inducible recombinases for bacterial optogenetics. *ACS Synth. Biol.* **9**, 227–235 (2020).
56. Roquet, N., Soleimany, A. P., Ferris, A. C., Aaronson, S. & Lu, T. K. Synthetic recombinase-based state machines in living cells. *Science* **353**, aad8559 (2016).
57. Guo, L. et al. Engineering *Escherichia coli* lifespan for enhancing chemical production. *Nat. Catal.* **3**, 307–318 (2020).
58. Sheth, R. U. & Wang, H. H. DNA-based memory devices for recording cellular events. *Nat. Rev. Genet.* **19**, 718–732 (2018).
59. Langer, E. S. & Rader, R. A. Continuous bioprocessing and perfusion: wider adoption coming as bioprocessing matures. *Bioprocess J.* **13**, 43–49 (2014).
60. Naseri, G., Behrend, J., Rieper, L. & Mueller-Roeber, B. COMPASS for rapid combinatorial optimization of biochemical pathways based on artificial transcription factors. *Nat. Commun.* **10**, 2615 (2019).
61. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
62. Lee, M. E., DeLoache, W. C., Cervantes, B. & Dueber, J. E. A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).

Acknowledgements

We thank Sebastian Sosa Carrillo for providing plasmid backbones, Steven Fletcher and Achille Fraise for assistance with the microscope acquisition software, Virgile Andreani for help with model fitting, and Zachary Fox for providing software for image analysis. We thank the members of the Batt group, Gael Yvert, and Hyun Youk for conducive discussions. We are grateful to Mustafa Khammash (ETH Zurich), Bernd Mueller-Roeber (University of Potsdam), and Matthias Peter (ETH Zurich) for providing plasmids. C.A. is enrolled in the Frontières de l'Innovation en Recherche et Education (FIRE) doctoral school hosted by Université de Paris. This work was supported by the H2020 Fet-Open COSY-BIO grant (grant agreement no. 766840), by the Inria grant IPL COSY and by ANR grants CyberCircuits (ANR-18-CE91-0002), MEMIP (ANR-16-CE33-0018), and Cogex (ANR-16-CE12-0025).

Author contributions

C.A. conceived and cloned the biological circuits and performed experiments. C.A. and F.B. did mathematical modelling and analysed data. C.A., J.R. and G.B. wrote the paper with inputs from F.B. J.R. and G.B. conceived and supervised the study.

Competing interests

The authors declare no competing interests.

Additional information


Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-26129-7>.

Correspondence and requests for materials should be addressed to Gregory Batt or Jakob Ruess.

Peer review information *Nature Communications* thanks Alexander Fedorec and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Annex 5: Using single-cell models to predict the functionality of synthetic circuits at the population scale

Chetan Aditya^{1,2,3}, François Bertaux^{1,2}, Gregory Batt^{1,2}, and Jakob Ruess*^{1,2}

¹ Inria Paris, 2 rue Simone Iff, 75012 Paris, France

² Institut Pasteur, 28 rue du Docteur Roux, 75015 Paris, France

³ Université de Paris, 85 Boulevard Saint-Germain, 75006 Paris, France

Keywords: Optogenetics, synthetic differentiation circuits, composability, chemical master equation, population dynamics

Abstract

Mathematical modeling has become a major tool to guide the characterization and synthetic construction of cellular processes. However, models typically lose their capacity to explain or predict experimental outcomes as soon as any, even minor, modification of the studied system or its operating conditions is implemented. This limits our capacity to fully comprehend the functioning of natural biological processes and is a major roadblock for the de novo design of complex synthetic circuits. Here, using a specifically constructed yeast optogenetic differentiation system as an example, we show that a simple deterministic model can explain system dynamics in given conditions but loses validity when modifications to the system are made. On the other hand, deploying theory from stochastic chemical kinetics and developing models of the system's components that simultaneously track single-cell and population processes allows us to quantitatively predict emerging dynamics of the system without any adjustment of model parameters. We conclude that carefully characterizing the dynamics of cell-to-cell variability using appropriate modeling theory may allow one to unravel the complex interplay of stochastic single-cell and population processes and to predict the functionality of composed synthetic circuits in growing populations before the circuit is constructed.

Introduction

At the heart of rational circuit design in synthetic biology lies the assumption that the functionality of complex circuits can be predicted from known properties of their components. Yet in practice, we routinely fail to make predictions of circuit dynamics that would agree with the data at the level expected in physics or engineering. A core reason for this is cell-to-cell variability inside genetically identical cell populations. Such population heterogeneity is often a consequence of the inherent stochasticity of biochemical processes inside cells [1]. Cell-to-cell variability may lead to unexpected and undesirable circuit dynamics and has been identified as one of the major roadblocks for designing synthetic circuits with in-silico predictable functionality [2, 3]. On the other hand, it has been shown that identifying and carefully characterizing sources of variability at the single-cell scale allows one to design remarkably robust synthetic circuits [4] or to exploit

stochasticity for creating features of cell populations, such as bimodal phenotype distributions, that would otherwise be difficult to engineer [5].

While the single-cell perspective has certainly helped to advance our understanding of cellular processes [6, 7, 8, 9], what eventually matters for most applications in synthetic biology is how a particular circuit functions inside growing populations. At the population-scale, synthetic circuits may intentionally (e.g. circuits to control growth [10]) or unintentionally (e.g. burden caused by protein production [11]) affect traits of cells that can be selected during population growth. If this is the case, we need to expect that variability that is generated at the single-cell scale (e.g. stochastic production of the burdensome protein) will lead to consequences at the population-scale that cannot be predicted solely based on a characterization of the circuit inside single cells. However, despite the apparent prevalence of problems where single-cell and population processes are coupled, multi-scale models that capture both single-cell stochastic chemical kinetics as well as population dynamics have not been used to predict, or even just to explain, how population-scale functionality emerges from single-cell characteristics of synthetic circuits. Presumably, this is because the classically used modeling framework for single-cell processes, that is the chemical master equation [12, 13], is only directly applicable at the population-scale whenever the population phenotype distribution is equivalent to the distribution of the single cell process and not additionally shaped by population level processes [14].

Here, we construct an artificial yeast differentiation system in which a Cre-recombinase is expressed from a light-inducible promoter and used to create dynamically controllable yeast communities of differentiated and undifferentiated cell types [15]. The system is equipped with four fluorescent reporter proteins to allow for simultaneous measurements of cellular processes and the differentiation state of cells. The core feature of our system is that varying the duration of applied light pulses allows one to modulate the fraction of cells that differentiates even though the entire population is exposed to the same global light stimulation. The system therefore exploits heterogeneity in the response of cells to light to enable external control of population dynamics but, by doing so, creates a complex coupling between population dynamics and the single-cell process that leads to the production of recombinase. It can therefore serve as an ideal test-bed to study what types of models are needed to explain and predict complex dynamics of synthetic circuits at the population scale. We find that a simple deterministic model that ignores all sources of variability can explain dynamics of a chromosomally integrated system version fairly well. However, modifying sources of variability and growth conditions, by expressing the system from plasmids and growing cells in selective media, leads to structural incapacity of the deterministic model to match measured population dynamics. To remedy this shortcoming, we derive a non-linear version of the chemical master equation (CME) that tracks conditional probabilities and that can be used to simultaneously model single-cell and population processes. Subsequently, we develop one such non-linear CME model to represent the chromosomally integrated differentiation system and another to track plasmid copy number fluctuations, plasmid loss, and growth in selective media. We find that composing these two models allows one to not only match, but to in-silico predict, complex emerging dynamics of the plasmid-based differentiation system without any modification of model parameters. These results suggest that the frequently encountered failure of model predictions for composed synthetic circuits may be

resolvable with appropriate component models that are carefully constructed to track the coupled dynamics of single-cell and population processes via multi-scale stochastic kinetic models.

A yeast optogenetic differentiation system

To be able to study the role of cell-to-cell variability in the design and functionality of synthetic circuits, we constructed a yeast optogenetic differentiation system whose dynamics can be well-observed at the single-cell level thanks to the simultaneous presence of four different fluorescent reporters. Concretely, we started from a circuit design that we recently published [15] and used a Cre-recombinase (Cre) under control of the EL222 promoter [16] to trigger differentiation in controllable population fractions using global light stimulation patterns (Figure 1). When expressed, Cre excises a DNA fragment that is designed such that, upon recombination, cells switch from constitutively producing blue fluorescent protein (mCerulean) to producing green fluorescent protein (mNeonGreen). Furthermore, red fluorescent protein (mScarlet-I) is produced alongside Cre from a second copy of the EL222 promoter and provides an observable readout that reports on Cre expression levels in response to light. Finally, EL222 transcription factor (EL222) is constitutively expressed from a pTDH3 promoter and fused to a yellow fluorescent protein (mVenus). In summary, the system allows one to trigger and monitor recombination in cells and to observe correlations of the probability for cells to recombine with cellular amounts of EL222. To test system functionality, we performed experiments using a platform of LED-equipped and fully computer-controlled parallel bioreactors [17]. Single-cell measurements are automatically taken by flow cytometry with the help of a programmable pipetting robot (Figure 1b). Using deconvolution to extract amounts of the different reporter proteins in cells from measured spectral signatures (see Supporting Information, Section S.7), we find that all cells gradually switch from blue to green when sufficient light is applied. On the other hand, exposing cell populations to pulses of light leads to bimodal mNeonGreen distributions (Figure 1c and Supporting Information, Figure S.1), which shows that only a fraction of the population recombines in response to light pulses. Applying a threshold in mNeonGreen fluorescence to classify cells into recombined and non-recombined, we can quantify the differentiation dynamics of the system (Figure 1c,d). We find that the system's response to light can be captured fairly well by a simplistic population dynamics model that relates differentiated and undifferentiated cells via a constant differentiation rate in the presence of light (Figure 1d and Supporting Information, Section S.1). To test if the probability for a cell to recombine is correlated with single-cell levels of EL222 and Cre, we analyzed mVenus and mScarlet-I fluorescence distributions shortly after applied light pulses. We find only minor differences in EL222 levels of undifferentiated cells before and after light induction that are difficult to distinguish from small inaccuracies in deconvolution or reactor-to-reactor variability of the experimental platform (Figure 1c and Figure S.1). Overall, we are led to conclude that cell-to-cell variability in EL222 and Cre can be safely ignored and that the functionality of the system can readily be characterized by a simple deterministic model. However, past experience in synthetic biology has shown that most circuits only function reliably in tightly constrained operating conditions and even seemingly good models retain their predictive power only in the precise context that has been used to construct the model.

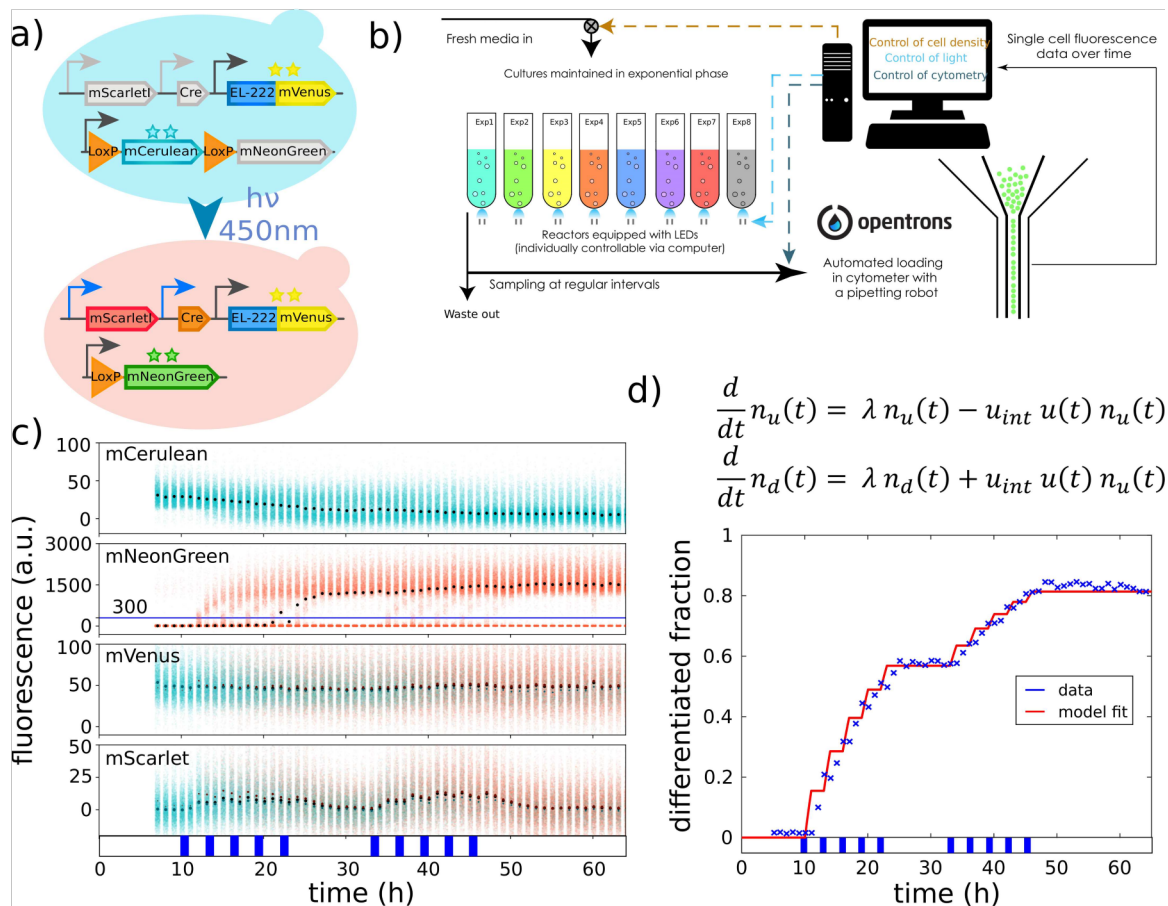


Figure 1: A yeast optogenetic differentiation system. (a) EL222 is produced in cells from a constitutive promoter (pTDH3, dark grey). Upon exposure of cells to blue light, EL222 switches the EL222-promoter from off (light grey) to on (blue) and cells start to produce mScarlet-I and Cre-recombinase. Cre excises a DNA-fragment placed between two target LoxP-sites which is designed such that cells switch from producing mCerulean to mNeonGreen from a second constitutive promoter (pTDH3, dark grey). (b) Yeast cells harboring the differentiation system are grown in parallel, fully automated bioreactors that are equipped with controllable blue LEDs. Optical density of cultures is maintained constant and flow cytometry measurements are taken at regular intervals via a custom-programmed pipetting robot. (c) Upon exposure to light, mScarlet-I fluorescence of some cells increases quickly, which indicates that also Cre is produced. Accordingly, mCerulean fluorescence decreases and mNeonGreen fluorescence increases. mNeonGreen fluorescence is then used to classify cells into differentiated and un-differentiated (blue line). EL222, fused to mVenus, is constitutively expressed in all cells and not affected by recombination. Small changes in mVenus levels in response to light that are noticeable in the data (smeared out, wider distributions towards the end) are due to the presence of mNeonGreen in cells, which leads to small errors in the deconvolution of mVenus fluorescence (Supporting Information, Section S.7). Dots show median fluorescence in all panels (black: all cells, cyan: undifferentiated cells, orange: differentiated cells). The applied light sequence is shown at the bottom (two times five 1h pulses with 2h between subsequent pulses). The data is also visualized in the form of distributions in the Supporting Information, Figure S.1. (d) A simple model (top) is capable of capturing the population dynamics of differentiated and undifferentiated cells very well. The applied light sequence is shown at the bottom.

Modifying circuits leads to unpredictable functionality

To test if the functionality of our differentiation system remains predictable when the circuit is modified, we constructed a variant of the system in which EL222 and Cre genes are placed on 2-micron plasmids instead of being chromosomally integrated (Figure 2a).

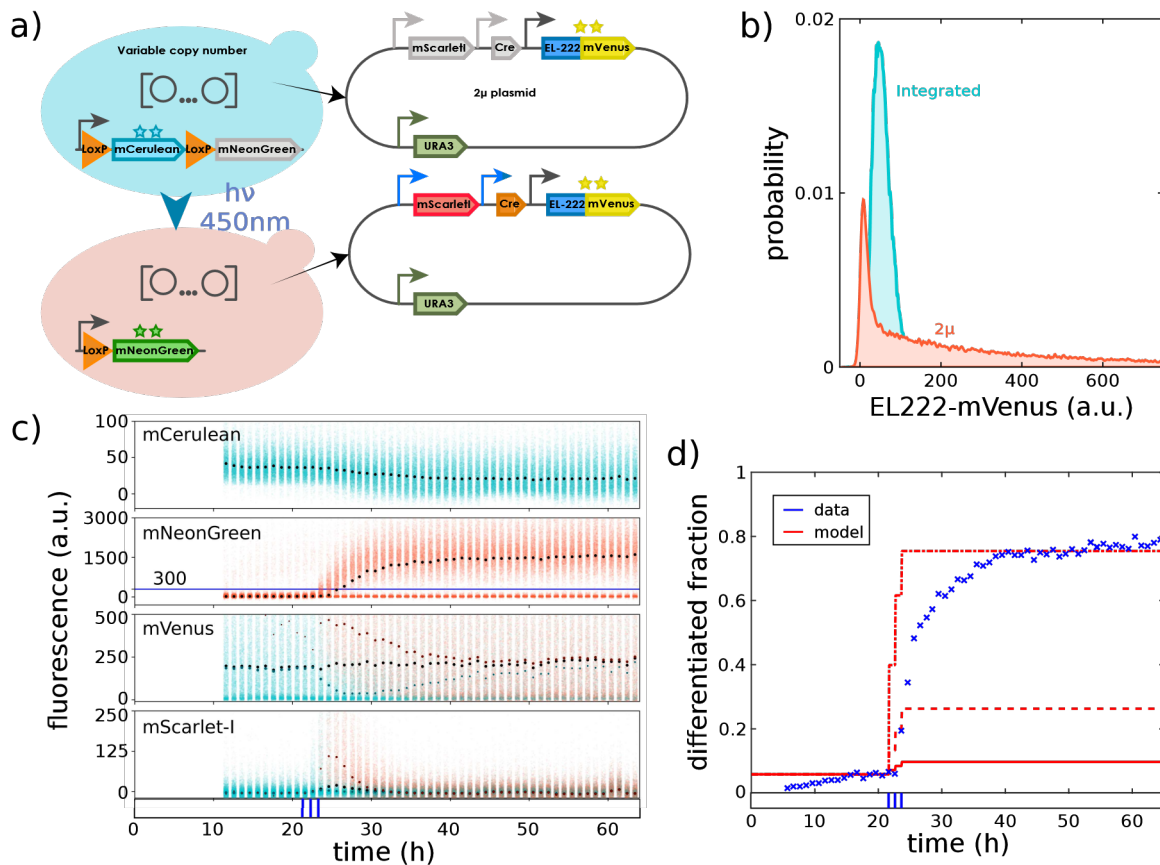


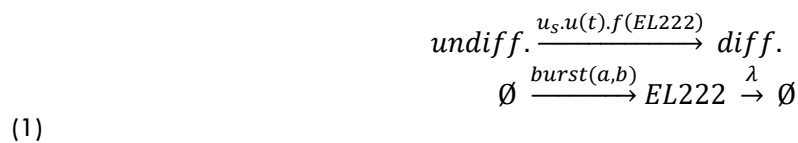
Figure 2: Expressing the differentiation system from plasmids breaks straightforward predictability. (a) To analyze the consequences of changes in cell-to-cell variability, a new strain was constructed in which Cre and EL222 genes are placed on 2-micron plasmids instead of being chromosomally integrated. (b) Measuring stationary EL222 distributions in the dark for the plasmid-based version of the system (red) shows that cell-to-cell variability in the amount of EL222 is strongly increased compared to the integrated version of the system (blue). (c) Large cell-to-cell variability in EL222 propagates to large variability in mScarlet-I (and thus Cre) production with some cells displaying much higher mScarlet-I fluorescence compared to the integrated system version (5-fold increased scale on y-axis for mScarlet-I and mVenus compared to Figure 1c). Using mNeonGreen fluorescence to classify cells into differentiated and undifferentiated (blue line) shows that light causes a transient split in mVenus distributions of the two sub-populations in response to light. The applied light sequence is shown at the bottom (three 5min pulses with 55min between subsequent pulses). (d) The simple model of Figure 1d fails to capture population dynamics for the plasmid-based version of the system. Crosses: measured differentiated population fractions. Solid line: model predictions using the same parameter values as in Figure 1d for the integrated system. Dashed line: model predictions after adjusting the differentiation rate parameter to account for the presence of multiple copies of the system. Dashdotted line: re-fitting the model to match final stationary differentiated fractions (30-40h after first light induction) leads to significant model mismatch during early transient dynamics. The applied light sequence is shown at the bottom.

Since plasmid copy numbers vary between cells, we expect this change to lead to significant differences in EL222 and Cre average levels and additional cell-to-cell variability. Indeed, growing cells in the dark, we find that EL222 distributions are very different from the integrated system version and characterized by much heavier tails and a mode that is shifted to lower levels (Figure 2b). Taken together, these two features imply that on average cells in the population contain more EL222 (almost 6-fold) but at the same time more cells contain less EL222 compared to the integrated system version (which notably falsifies the a priori expectation that all cells

would have higher EL222 levels due to the presence of multiple plasmids in cells). We may therefore wonder if and how these differences impact the functionality of the circuit and if emerging dynamics of the population composition can still be predicted by the simple model in Figure 1d. Exposing cells to different light patterns, we find that the same amount of light leads to differentiation of more cells for the plasmid based version of the system (Figure S.3), which is in line with higher average levels of EL222 in the population and the presumable presence of multiple plasmids in cells, each carrying a copy of the promoter driving Cre. Adjusting the differentiation rate parameter in the simple deterministic model to account for the on-average presence of multiple copies of the system, however, does not lead to agreement of model predictions and data, nor is it at all possible to obtain any precise fit of this model to the population dynamics that emerge from the plasmid-based differentiation system (Figure 2d). Analysing mVenus distributions, we find that differentiated cells shortly after applied light pulses are characterized by high EL222 levels whereas the mVenus distribution of the undifferentiated subpopulation is shifted to lower levels compared to the total mVenus population distribution (Figure 2c). Since EL222 is constitutively expressed from the same promoter and plasmids in all cells, we conclude that these differences must be caused by selective differentiation of cells with high amounts of EL222. Differences between sub-populations gradually disappear over time but are still noticeable up to days after the last application of light to the population (Figure 2c). This is quite remarkable as it is difficult to comprehend, at a first glance, how a constitutively expressed gene can display a cellular memory of a stimulus that is retained over several tens of cell generations. In conclusion, cell-to-cell variability in EL222, which previously seemed to be negligible for the characterization of the system, suddenly appears to be of key importance for understanding how population dynamics emerge from the differentiation system. We may thus ask ourselves if a dedicated characterization of the system with a multi-scale stochastic kinetic model that takes into account both single-cell and population processes would have allowed us to retain predictable functionality.

Single cell modeling of the integrated differentiation system

To test if consequences of changes in the system can be understood and predicted, we constructed a multi-scale stochastic kinetic model of the integrated differentiation system and a model of plasmid copy number fluctuations and asked if the models can be composed to predict emerging single-cell and population dynamics when the differentiation system is expressed from plasmids. Concretely, since variability in EL222 appeared to be of key importance, we deployed a model of bursty production of EL222 and cell differentiation to represent the differentiation system:



where u_s is the maximal single cell differentiation rate for given fixed light intensity, $u(t)$ is equal to one in the presence of light and equal to zero otherwise, λ is the cells' growth rate, and a is the rate at which protein bursts occur. Protein production bursts are of size Z and assumed to be geometrically distributed with average burst size b , $Z \sim \text{Geo}(\frac{1}{b})$, as dictated by classical results for modeling stochastic gene expression [18, 19]. To keep the model as simple as possible, we neglected possible delays or noise caused by the production and action of recombinase or the experimental detection of recombined cells. Instead, we assumed that the probability per unit time for a cell to differentiate in the presence of light is directly a function of the amount of EL222, $f(\text{EL222})$, in the cell.

When $u(t) = 0$, the EL222 distribution converges approximately to a stationary Gamma distribution that is determined by average burst size and frequency [18, 19]. Growing cells in the dark and measuring their mVenus fluorescence by flow cytometry then allows one to determine burst size and frequency (up to a fluorescence scaling factor) from mean and coefficient of variation of measured fluorescence distributions (Figure 3a, left). When light is applied to the population, $u(t) = 1$, the dynamics of the differentiated population fraction are determined by the specific shape and parameters of the differentiation function $f(\text{EL222})$. Matching emerging population dynamics of the integrated system for the light pattern in Figure 1d, we find that a steep Hill-function with a threshold significantly larger than average amounts of EL222 leads to good agreement with the data (Figure 3a, right and Supporting Information, Section S.2 A).

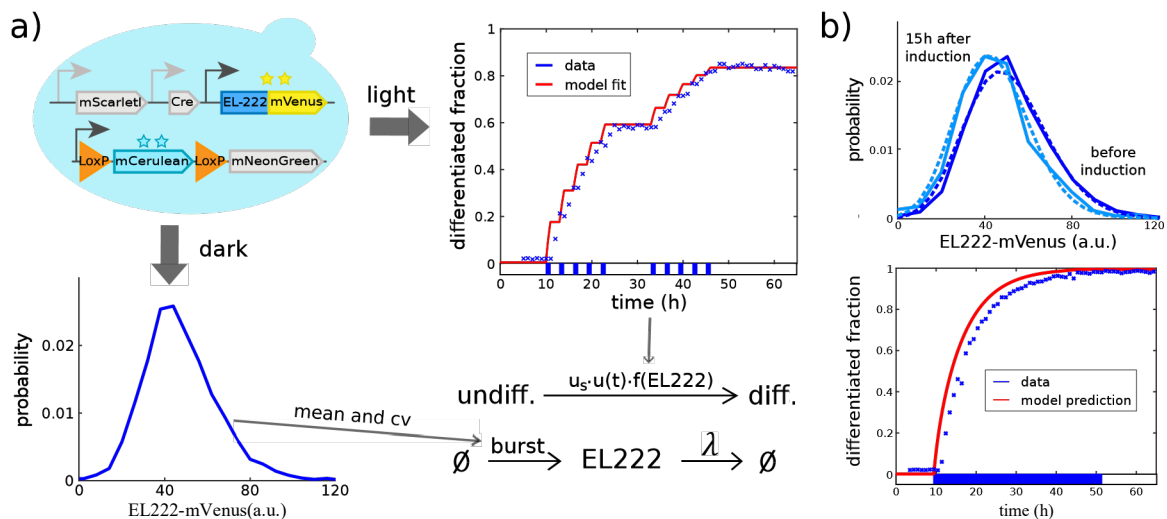


Figure 3: **(a) Characterization of the differentiation system.** *Bottom:* Extracting mean ($= 50.6\text{a.u.}$) and coefficient of variation ($= 0.35$) from measured stationary EL222 distributions before light induction allows one to determine burst frequency and average burst size of the EL222 production model up to a fluorescence scaling factor (Supporting Information, Section S.2). *Right:* Exposing cell populations to light pulses and measuring dynamics of the resulting population composition (crosses: data, line: model fit, bottom: applied light sequence) then allows one to determine parameters of the differentiation function $u_s \cdot f(\text{EL222})$ (Supporting Information, Section S.2). **(b) Dynamics in continuous light.** *Bottom:* Exposing cells to continuous light eventually leads to differentiation of the entire population. *Top:* Cells that remain undifferentiated after light induction (light blue) seem to be characterized by slightly lower EL222 levels

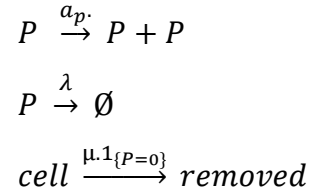
compared to initial levels (dark blue) according to both model predictions (dashed) and data (solid). Data from further time points is provided in the Supporting Information, Figure S.2.

Mathematically, the model couples a variability generating process (stochastic production of EL222) to a population process that selectively differentiates cells based on their phenotypes. The coupling of single-cell and population scale tells us that, in response to light, we should expect preferential differentiation of cells with high levels of EL222. Correspondingly, differentiated cells should display increased amounts of EL222 shortly after light while the (sub-)population distribution of undifferentiated cells should shift to lower levels. However, this selective population split is counter-acted by the fact that the same variability generating process is operating in all cells and that this process will always drift to the original EL222 distribution in both sub-populations at a time scale that is determined by the cells' growth rate (see Supporting Information, Section S.2 B for details). In the absence of light, we therefore expect distributions to revert back to the original EL222 distribution while the continuous presence of light should lead to a maximal shift of the distribution to a quasi-stationary condition in which single cell and population process are dynamically balanced until eventually all cells will have differentiated. The quasi-stationary EL222 distribution for undifferentiated cells in the presence of light can be obtained via an eigenvector calculation from a non-linear master equation that can be derived from the original master equation of the bursty protein production model (see Supporting Information, Section S.2 B). Performing this calculation, we find that the maximum possible shift in EL222 distributions of undifferentiated cells that can potentially be observed in experiments is fairly small and of similar size as experimental errors due to inaccurate deconvolution or reactor-to-reactor variability. To test whether such a shift can nevertheless be detected, we exposed cells to continuous light and collected measurements at time points early enough after induction such that sufficiently many cells remain undifferentiated to allow for reliable quantification of EL222 distributions. We find that experimental EL222 distributions of undifferentiated cells indeed seem to show a small shift towards lower levels in response to continuous light (Supporting Information, Figure S.2). This shift is in good agreement with distribution dynamics predicted from the model (Figure 3b, top and Supporting Information, Figure S.4). As a side note, the model provides a very good prediction of the increase in the differentiated fraction in response to continuous light (Figure 3b, bottom). Despite the possible presence of small selection effects, we can overall conclude that sufficiently low noise in EL222 production coupled to sufficiently fast fluctuations implies that cell-to-cell variability has only small consequences for emerging population dynamics. It is now clear, however, that this conclusion will change if either noise levels or time scales of the single cell process are modified.

Consequences of plasmid copy number fluctuations

In addition to a single-cell model of the differentiation system, we require a model that captures cell-to-cell variability in plasmid copy numbers. Many, often detailed, models of plasmid copy number fluctuations exist in the literature [20, 21]. In order to keep the system characterization as simple as possible, we decided to omit any detailed mechanistic description of processes such as replication failure or unequal division of plasmids between mother and daughter cell [22].

Instead, we chose to represent plasmid copy number fluctuations by a simple stochastic birth-death process with both birth rate (representing replication) and death rate (representing dilution due to cell growth) being linear in the plasmid copy number:



(2)

where α_p is the plasmid replication rate, λ is the growth rate of cells, and μ is the rate at which cells that have lost the plasmid ($1_{\{P=0\}}$) are removed from the population when cells are growing in selective media. Replication failure can be implicitly incorporated by choosing the replication rate smaller than the cells' growth rate, which is in any case a necessary feature of a birth-death process model since expected plasmid copy numbers in cells would diverge to infinity if plasmids are replicated faster than cells divide. For $\alpha_p < \lambda$ (and $\mu = 0$), on the other hand, the process will eventually reach zero plasmids with probability one. The rate at which cells in the population lose the (last copy of the) plasmid is then determined by the difference between replication and growth rate. When selective media is used for growth, $\mu > 0$, cells that have lost the plasmid will either die or be outgrown and the plasmid copy number distribution of the population will remain stable. However, this neither implies that there exist no cells without plasmids in the population nor do plasmid copy numbers remain constant from the perspective of single cells. Instead, we should expect a dynamic equilibrium and a quasi stationary plasmid copy number distribution in which the single cell plasmid loss process is balanced with selective removal of cells without plasmids (Supporting Information, Section S.3 B). From a mathematical perspective, this result is equivalent to what was obtained previously for EL222 distributions in the undifferentiated cell population. In both cases, a variability generating process at the single-cell scale (EL222 fluctuations vs. plasmid copy number fluctuations) is coupled to a state-dependent removal process at the population scale (differentiation vs. removal of cells that have lost the plasmid) and leads to the same type of non-linear master equation for the cells that have not been removed yet (compare Supporting Information, Sections S.2 B and S.3 B).

To characterize the full multi-scale dynamics of plasmid copy numbers and cell populations, we switched cells from selective to non-selective media and measured how the average abundance of a constitutively expressed protein decays over time. We found that average fluorescence decays approximately exponentially with a rate that is 15% of the cells' growth rate λ (Figure 4a, bottom). Mathematical analysis of the single-cell model (Supporting Information, Section S.3 B) shows that this is a direct consequence of the plasmid replication rate, α_p , being 15% smaller than the cells' growth rate $\alpha_p = 0.85\lambda$. Somewhat counter-intuitively, the net population growth rate in selective media is independent of μ . Faster removal of cells that have lost the plasmid means faster reduction (or slower increase) in the total number of cells but this is compensated by the fact that, in stationary growth conditions, the population fraction of

cells without plasmids becomes smaller when their removal rate is increased. Therefore, to determine μ , stationary growth rate measurements are insufficient. Instead, it is necessary to directly measure what fraction of the population has plasmids in stationary growth conditions since it can be shown (Supporting Information, Section S.3 A) that this fraction must be equal to $\frac{\mu - \lambda + \alpha_p}{\mu}$ for $\alpha_p < \lambda$ (plasmid numbers do not diverge) and $\mu > \lambda - \alpha_p$ (cells with plasmids can be maintained in the population). Correspondingly, we performed a colony counting experiment (see Supporting Information, Section S.3 A) and observed that approximately one third of all cells do not contain plasmids in stationary growth conditions (Figure 4a, top), which allowed us to determine that the removal rate of cells without plasmids, μ , must be approximately 45% of the cells growth rate, $\mu = 0.45\lambda$. Together, the parameters α_p , μ and λ completely characterize the multi-scale dynamics of plasmid copy number fluctuations and growth in selective media.

To test the model and to better understand possible consequence of plasmid copy number fluctuations, we experimentally determined effective population growth rates of our cells in selective and non-selective media. We found that, after pre-culture in selective media, populations of cells carrying the plasmid-based differentiation system quickly adopt the same growth rate as populations of cells carrying the integrated system version when transferred to non-selective media. Subsequently, the population growth rate remained constant and no dependence on the fraction of cells that still carry plasmids was detected. When grown in selective media, however, populations of cells carrying the plasmid-based differentiation system display an effective growth rate that is reduced by approximately 15% (Figure 4b, right). Analyzing the model, we find that the reduction in growth is a consequence of removal of cells that have lost the (last copy of) the plasmid and that the observed reduction by 15% is in quantitative agreement with model predictions and another consequence of

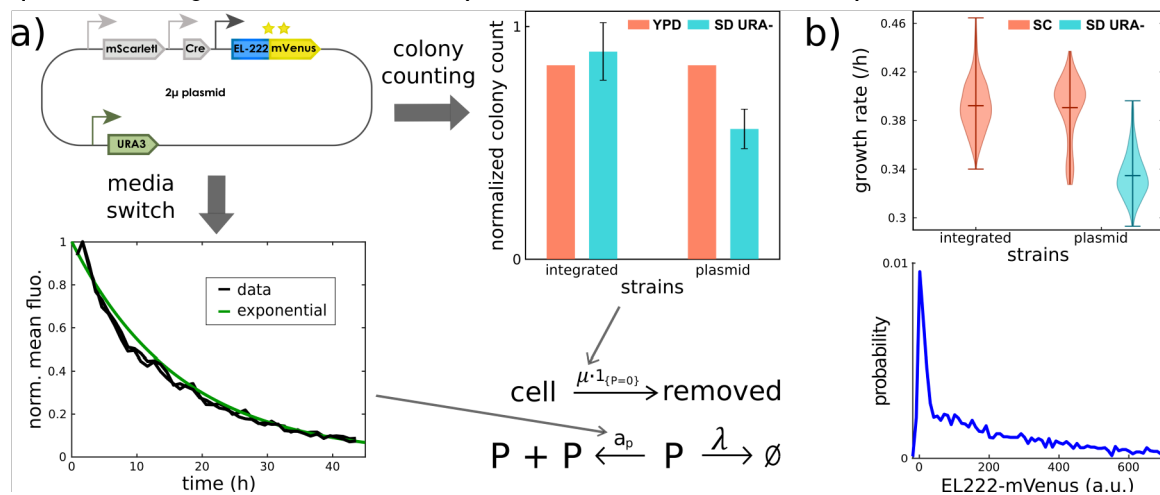


Figure 4: (a) Characterization of plasmid copy number dynamics. *Bottom:* After switching cells from selective to non-selective media, the rate of exponential decay of average levels of a protein that is constitutively expressed from plasmids (here EL222-mVenus) can be measured to determine the population plasmid loss rate and to fix the plasmid replication rate in the model. Mean fluorescence (normalized by its maximum) of two replicates is shown. *Right:* Stationary fractions of cells without plasmids determine the removal rate in selective media of cells that have lost the plasmid. These fractions can be measured by a colony counting experiment as described in the Supporting Information, Section S.3 A. **(b) Consequences of plasmid copy number fluctuations.** Expressing the differentiation system

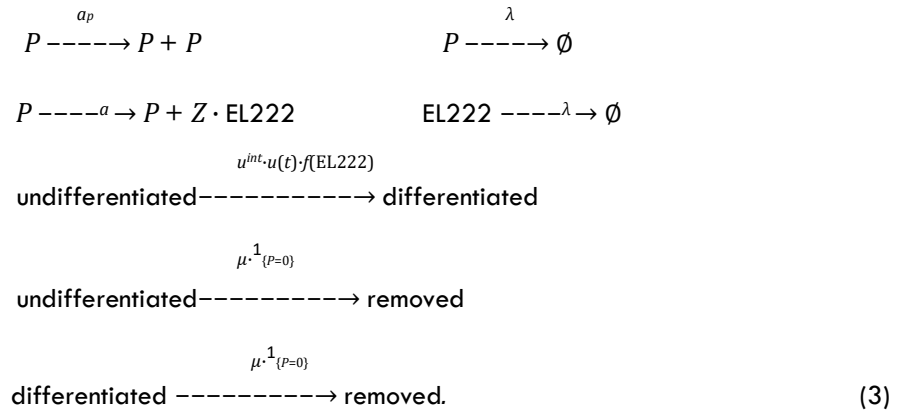
from plasmids leads to modified population distributions of EL222 (bottom) as well as changes in the effective population growth rate (top). Violin plots are composed of growth rate estimates based on cell density averaged over all reactors of the turbidostat and multiple experiments (in total 19 reactors for the strain with the integrated system, 12 for the strain with the plasmid-based system in selective media and 6 in non-selective media).

the plasmid replication rate, α_p , being 15% smaller than the single-cell growth rate λ . Overall, we should expect that placing the differentiation system on plasmids will significantly alter variability in EL222 levels (Figure 4b, bottom) and also modify effective population growth.

Having constructed single cell models of the differentiation system and plasmid copy number fluctuations, we are now in a position to ask the central question of this manuscript: is it possible to quantitatively predict functionality of the plasmid-based differentiation system without having to change the model or even just to re-identify model parameters for the new conditions?

Parameter-free prediction of circuit functionality

To test if the dynamics of the plasmid-based differentiation system can be predicted by combining models calibrated on parts of the system, we coupled the single-cell model of the integrated differentiation system, Eq.(1), to the model of plasmid copy number fluctuations, Eq.(2), to obtain a composed model that can be stated in reaction network form as



We highlight that the coupling of the two models is very natural and does not introduce any new parameters: the protein burst rate is now scaled by the plasmid copy number since production can occur from any copy of the plasmid while all other model parts are exactly the same as previously characterized. To test if complex functionalities of our circuit such as population dynamics emerging from single-cell stochastic biochemical processes can be predicted without any adjustment of model parameters, we used the composed model to predict the dynamics of the plasmid-based version of the differentiation system for the light stimulation pattern in Figure 2d and compared the results to data. We find that population dynamics of differentiated and undifferentiated cells are very well predicted without any adjustment of model parameters (Figure 5b).

Comparing EL222 distributions in the two sub-populations between model and data shows that the high quality of population-level predictions is a consequence of the fact that the model

predicts correctly how single-cell processes will operate in union with population-level processes to shape the full dynamics of EL222 distributions in undifferentiated cells (Figure 5c,d and Supporting Information, Figure S.5). In particular, the heavy tails of EL222 distributions for cells growing in darkness (observed already in Figure 2b) emerge naturally from the fact that plasmid copy numbers can fluctuate significantly in the model. These heavy tails imply that significantly more cells have EL222 levels above the threshold of the differentiation Hill function $f(\text{EL222})$ and will recombine quickly upon light induction. Shortly after light induction, the remaining undifferentiated cell population is therefore shifted to significantly lower EL222 levels (5-6 fold reduction of the median, Figure 5c) while the differentiated cell population inherits the heavy tail of the original distribution (Figure 5d). While not experimentally measurable, it can be deduced that according to the model the same holds true for plasmid copy number distributions in sub-populations (Supporting Information, Figure S.6). As a consequence, the population differentiation rate spikes very high upon first light induction but is significantly reduced when subsequent light pulses are applied before the EL222 distribution of the undifferentiated cell population has converged back to its initial condition (Supporting Information, Figure S.7). When the light stimulus is maintained for some time, fluctuations in plasmid copy numbers create larger fluctuations in EL222 amounts compared to the integrated version of the system, which leads to more frequent threshold crossing events and therefore larger population differentiation rates.

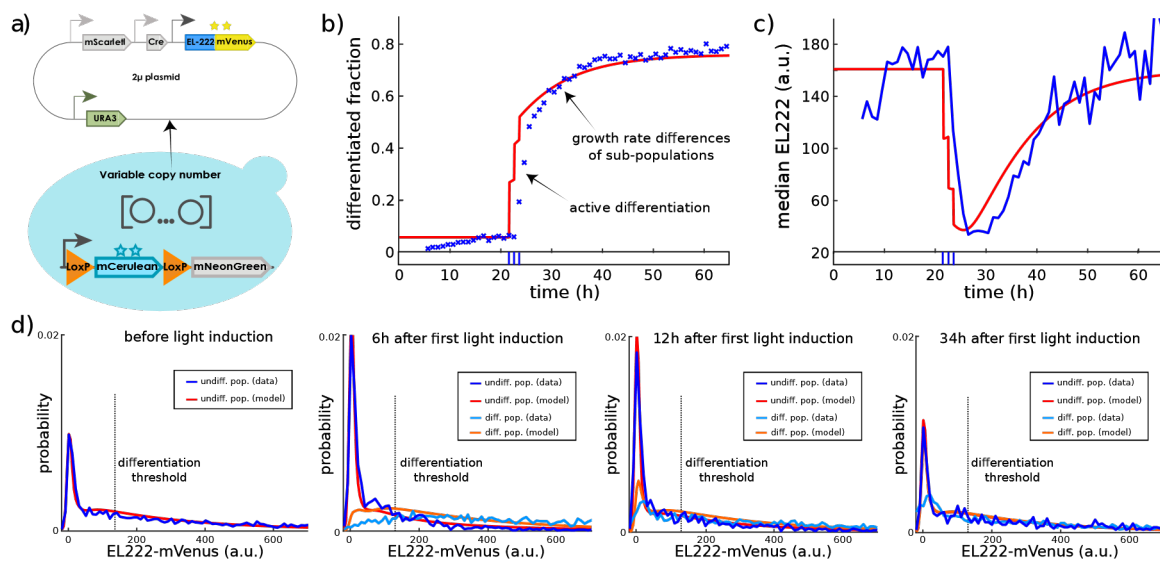


Figure 5: Parameter-free prediction of emerging population dynamics for the plasmid-based differentiation system. (a) Sketch of the plasmid-based differentiation system. (b) Emerging dynamics of the differentiated fraction according to the composed model (red) for the same light input as in Figure 2d are compared to experimental data (blue). According to the model, late increases in the differentiated fraction are due to varying sub-population growth rates and not caused by active differentiation. (c) Upon light induction, median EL222 levels of undifferentiated cells drop 5-6 fold as predicted by the composed model. (d) *Left panel:* The model-predicted EL222 distribution in the dark agrees very well with experimental data even though this data was not used to parametrize the model. According to the model, the peak of the distribution close to zero corresponds to the $\sim 33\%$ of cells without plasmids (see Figure 4a). *Middle panels:* Shortly after the application of light, cells with EL222 levels above the threshold parameter of the differentiation Hill function (dashed line) are enriched in the differentiated population while the undifferentiated population contains more cells without plasmids (increased peak close to zero). *Right panel:* When cells are subsequently kept in the dark, EL222 distributions converge back to the initial

condition. Model predictions show remarkable agreement with measured EL222 distribution dynamics, except for seemingly lower numbers of differentiated cells without plasmids in the data (in particular after 34h). This small mismatch is presumably caused by inaccuracies in deconvolution (the presence of mNeonGreen in cells makes it difficult to precisely quantify low mVenus levels in cells, Supporting Information, Section S.7). Full time-varying distribution data of all fluorescent reporters for this experiment is displayed in the Supporting Information, Figure S.8.

Eventually, the plasmid-based version of the system reaches differentiated population fractions close to 100% very quickly when light is maintained (Supporting Information, Figure S.10) despite the fact that a large part of the population (around a third) displays mVenus levels that are close to zero at all time points. According to the model, these are cells that have lost the plasmid (Supporting Information, Figure S.6), cannot differentiate, but are likely to be removed at subsequent time points due to growth in selective media. The coupling of plasmid loss dynamics and selective media with the differentiation system therefore leads to unintuitive population dynamics in which seemingly the entire population can recombine despite a continuous presence of many cells that are not carrying any copy of the differentiation system. Another complex consequence of the coupling of single-cell and population processes is that the split in plasmid copy numbers between sub-populations that follows from selective differentiation of cells with high EL222 levels leads to different sub-population growth rates in selective media (Supporting Information, Figure S.7) since cells that have lost the plasmid (or are close to losing it) are enriched in the undifferentiated subpopulation. This implies that the differentiated population fraction will continue to increase even if light is removed and no more active differentiation takes place, which explains why the assumption of the simple deterministic model in Figure 1d, that the differentiated population fraction can only increase due to active differentiation, led to structural incapacity of the model to explain the slow transient differentiation dynamics and significantly increasing differentiated fractions up to 10h after last light induction (Figure 5b). If the light stimulus is removed, EL222 sub-population distributions converge back to the original distribution (Figure 5d) but remain noticeably different for at least a day after last light induction. This experimental observation is in good agreement with slow convergence of sub-population plasmid copy numbers to their quasi-stationary distribution in selective media, which creates a surprisingly long sub-population memory of applied light stimuli (Supporting Information, Figure S.6). We conclude that deploying multi-scale stochastic chemical kinetics models for understanding the interplay of single-cell and population processes allows us to understand and predict complex emerging dynamics when the differentiation system is placed on plasmids and cells are grown in selective media. Population dynamics for plasmid strains, despite being shaped by cell-to-cell variability, are deterministically reproducible (Supporting Information, Figure S.12). Thus, the capacity to predict such dynamics implies that the interplay of single-cell and population processes can be exploited for creating features of microbial community dynamics that would otherwise be difficult to engineer.

Optogenetic control of constitutive gene expression

In the previous section, we demonstrated that coupling of single-cell and population processes may lead to outcomes that are fairly unintuitive such as dynamically changing population

distributions of a constitutively expressed gene or increasing fractions of differentiated cells in the absence of active differentiation. In practice, such couplings will often be seen as a nuisance but we may also ask if the interplay of stochastic single-cell processes and population dynamics can be exploited to create and control features of cell populations that would otherwise be impossible to engineer. For instance, in light of the results of this paper, we may ask if it is possible to use light to regulate constitutive gene expression, plasmid copy numbers, and growth rates of sub-populations via targeted differentiation of cells with high EL222 levels. Since neither plasmid copy numbers nor sub-population growth rates are directly measurable on our experimental platform, we set ourselves the goal to regulate the constitutively expressed EL222. Concretely, according to our theoretical results (convergence to a quasi-stationary distribution, Supporting Information, Sections S.2 B and S.3 B), it should be possible to maintain EL222 levels in the undifferentiated cell population at reduced constant levels by applying continuous light for sufficiently long. Indeed, we find that EL222 distributions in undifferentiated cells quickly reach a distribution that remains invariant when the population is exposed to continuous light (Figure 6c, left). However, this distribution is characterized by a median of almost zero (Figure 6b), which indicates that only cells that have no EL222 (and presumably no plasmids) remain undifferentiated. Furthermore, 15 – 20h after the start of light application, almost the entire population is differentiated. To test if also low but non-zero EL222 levels in undifferentiated cells can be stably maintained, we exposed cells to a number of different light sequences that mimic continuous light with short pulses that are regularly repeated with an interpulse duration that is short in comparison to the duration of the experiment. We find that applying light for 2min every 4h or for 1 min every 1h leads to approximately halved median EL222 levels in undifferentiated cells (Figure 6b). In both cases, the full EL222 distribution of undifferentiated cells remains approximately invariant after around 12h have passed from first light application (Figure 6c, middle, see also Supporting Information, Figure S.13). Finally, we applied a light sequence with only half-minute light pulses applied every 4h and found that, in this case, the EL222 distribution of undifferentiated cells differs only very moderately from the initial population distribution at all

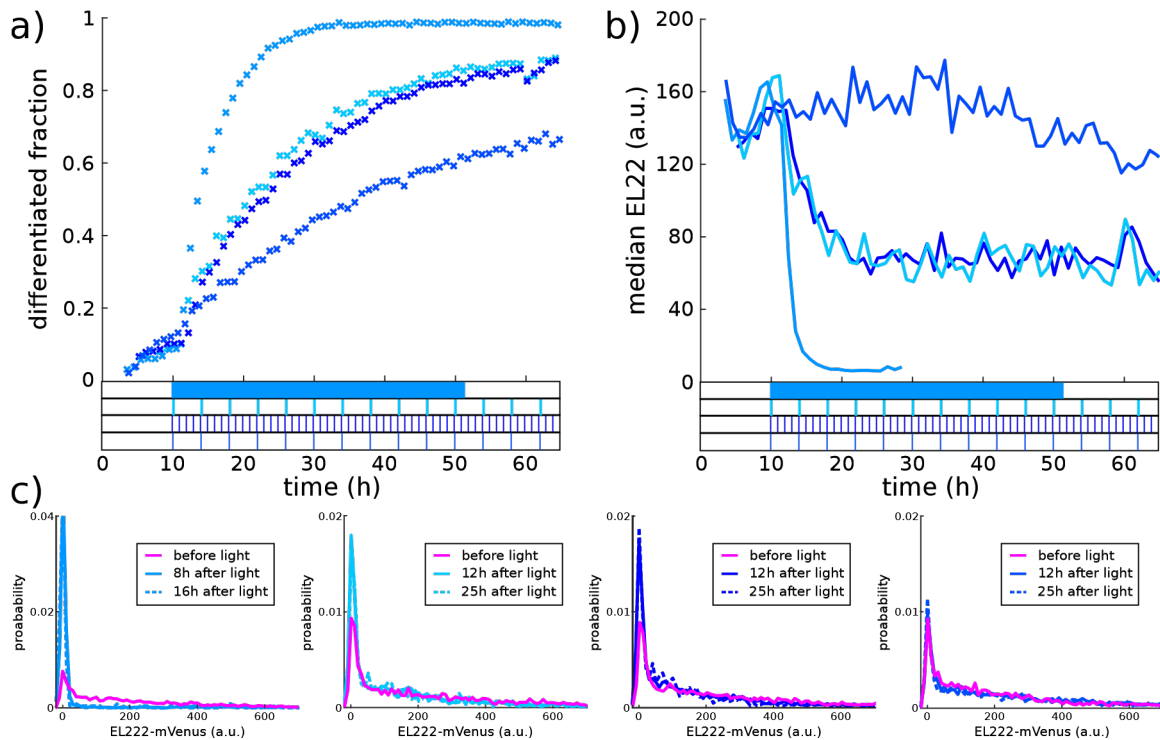


Figure 6: **Optogenetic control of constitutive gene expression** (a) Regularly repeated short light pulses (shown at the bottom) lead to a slow steady increase in the differentiated population fraction while the differentiated fraction increases fairly quickly in response to continuous light. (b) Over intermediate time periods, repeated selective differentiation and reversion to the mean balance each other, which leads to approximately constant EL222 medians in all cases. (c) After transient dynamics in response to the start of the light sequence, EL222 distributions in undifferentiated cells reach a quasi-stationary form that remains invariant for extended time periods (see also Supporting Information, Figure S.13). Color coding of distributions labels correspondence to the light sequences and data in panels a and b.

time points (Figure 6c, right) despite the fact that the differentiated population fraction increases (Figure 6a).

To conclude, understanding and characterizing the consequences of cell-to-cell variability allowed us to use light to optogenetically regulate the (sub-population) level of a constitutively expressed gene, a feat that seems quite counter-intuitive at a first glance and that is not realizable in any obvious way by other means.

Discussion

Quantitatively predicting the dynamics of complex synthetic circuits before the circuit is constructed is the key challenge that needs to be mastered to turn synthetic biology into a true engineering discipline. Yet, while our capacity to construct complex circuits is continuously increasing, our capability to predict circuit functionality from supposedly known and characterized circuit components remains, at best, limited to very tightly constrained operating conditions and qualitative and/or stationary outputs [23]. We exemplified this problem for the case of our optogenetic differentiation system by showing that predictions obtained from a simple deterministic model break down as soon as the context in which the system is used

(plasmid-based expression of proteins and growth in selective media) is modified (Figure 2). The concrete reasons for the failure of model predictions may be manifold and caused by unexpected component-to-component interactions (e.g. retroactivity [24] or resource competition [25]) or couplings of the circuit to processes of the host (e.g. burden [11] or saturation of the host's degradation machinery [4]). Eventually, however, all these reasons are but different facets of a common problem: our incapacity to foresee the consequences of complex inter-dependencies of a circuit in vivo.

For our optogenetic differentiation system, expressing proteins from plasmids led to unforeseen couplings of plasmid copy number fluctuations, growth in selective media, and selective differentiation, which, for instance, created seemingly slow “differentiation dynamics” that remained mysterious (even to us) in the absence of an explanatory model (Figure 2d). We thus set out to construct dedicated multi-scale stochastic kinetic models of the circuit's components in their population context (Figures 3 and 4). Since the chemical master equation (CME) in its standard form does not capture population processes, it was necessary to augment the CME and to derive a non-linear version for conditional probabilities to correctly capture dynamics of population distributions (Supporting Information, Sections S.2 B and S.3 B). We then used our theoretical results to determine experiments that can be used to readily extract parameters of the two component models from data (Figures 3a and 4a). Merging the so obtained models to construct a model of the plasmid-based differentiation system (Eq.[3]), we found that the composed model does not only resolve the previously not understood features but quantitatively predicts the consequences of complex component interactions without any adjustment of model parameters (Figure 5). This is a particularly encouraging finding as it demonstrates that, at least for the system studied in this paper, our incapacity to foresee the consequences of complex inter-dependencies of circuit components and couplings of single-cell and population processes can be remedied by appropriate characterization of circuit components. Importantly, with the ability to foresee consequences comes the possibility to exploit such couplings in the design [26]. This is evidenced by our result that the circuit allows one to modulate constitutive gene expression (and presumably plasmid copy numbers and growth rates) in sub-populations via the application of light (Figure 6) - a feat that seems impossible upon a first glance at the circuit's wiring diagram (Figure 1a).

To conclude, we highlight that for the application considered in this paper, the crucial ingredient for obtaining a faithful model was to augment the CME to incorporate population processes in addition to single-cell dynamics that are classically tracked in stochastic models. It is to be expected that the same will hold true for many other applications where couplings of single-cell and population processes are likely to be at play [27, 28, 29, 30, 31]. This is notably the case for synthetic circuits that are constructed to produce proteins in large quantities since this creates a burden for cells that may lead to growth rate variability between cells and a dynamical enrichment of low-producing cells upon induction of protein production. Furthermore, similar couplings between scales are likely also present in many natural systems such as selective killing of cancer cells with particular states of internal processes in response to treatments that induce the apoptotic pathway [32] or differential responses of bacteria to antibiotic treatments [33] to name but two possible examples. We therefore expect that our approach for calculating

with multi-scale stochastic kinetic models will be of use far beyond the particular case study considered here. It needs to be noted, however, that for cases where the single-cell model is more high-dimensional than the fairly small models considered here, tracking the entire solution of the corresponding master equation will be computationally infeasible. Further work is thus necessary to develop and test approaches for approximately calculating with multi-scale stochastic kinetic models [34, 14].

Author contributions

C.A. performed experimental work. J.R. performed modeling and mathematical work. C.A. and J.R. analyzed data. F.B. helped with experimental platforms and data analysis. G.B. and J.R. conceived and supervised the project. All authors wrote the paper.

Acknowledgements

We thank Sebastian Sosa Carrillo for providing plasmid backbones and Zachary Fox for providing software for efficient master equation solving. C.A. is enrolled in the Frontières de l'Innovation en Recherche et Education (FIRE) doctoral school hosted by Université de Paris. This work was supported by the H2020 Fet-Open COSYBIO grant (grant agreement no. 766840) and by the ANR grant CyberCircuits (ANR-18-CE91-0002).

Data, material and code availability

Data can be found on Zenodo (DOI:10.5281/zenodo.5155290), and code used for data analysis in the GitLab repository (<https://gitlab.inria.fr/InBio/Public/predicting-selection-effects>). Plasmids and strains are available from the authors upon request.

References

- [1] Elowitz M, Levine A, Siggia E, Swain P (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.
- [2] Kwok R (2010) Five hard truths for synthetic biology. *Nature* 463:288–290.
- [3] Del Vecchio D, Qian Y, Murray R, Sontag E (2018) Future systems and control research in synthetic biology. *Annual Reviews in Control* 45:5–17.
- [4] Potvin-Trottier L, Lord N, Vinnicombe G, Paulsson J (2016) Synchronous long-term oscillations in a synthetic gene circuit. *Nature* 538:514.
- [5] Lugagne JB, et al. (2017) Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nature Communications* 8:1671.
- [6] Raser J, O'Shea E (2005) Noise in gene expression: Origins, consequences, and control. *Science* 309:2010–2013.
- [7] Zechner C, et al. (2012) Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences of the USA* 109:8340–8345.
- [8] Minsky B, Neuert G, van Oudenaarden A (2012) Using gene expression noise to understand gene regulation. *Science* 336:183–187.

- [9] Neuert G, et al. (2013) Systematic identification of signal-activated stochastic gene regulation. *Science* 339:584–587.
- [10] Miliadis-Argeitis A, Rullan M, Aoki S, Buchmann P, Khammash M (2016) Automated optogenetic feedback control for precise and robust regulation of gene expression and cell growth. *Nature Communications* 7:12546.
- [11] Weiße A, Oyarzún D, Danos V, Swain P (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences of the USA* 112:E1038–E1047.
- [12] Gillespie D (1992) A rigorous derivation of the chemical master equation. *Physica A* 188:404–425.
- [13] Schnoerr D, Sanguinetti G, Grima R (2017) Approximation and inference methods for stochastic biochemical kinetics - a tutorial review. *Journal of Physics A: Mathematical and Theoretical* 50:093001.
- [14] Lunz D, Batt G, Ruess J, Bonnans J (2021) Beyond the chemical master equation: stochastic chemical kinetics coupled with auxiliary processes. *PLoS Computational Biology* 17:e1009214.
- [15] Aditya C, Bertaux F, Batt G, Ruess J (2021) A light tunable differentiation system for the creation and control of consortia in yeast. *bioRxiv*.
- [16] Baumschlager A, Aoki S, Khammash M (2017) Dynamic blue light-inducible $\tau 7$ rna polymerases (opto- $\tau 7$ rnaps) for precise spatiotemporal gene expression control. *ACS synthetic biology* 6:2157–2167.
- [17] Bertaux F, et al. (2020) Enhancing bioreactor arrays for automated measurements and reactive control with reacSight. *bioRxiv*.
- [18] Friedman N, Cai L, Xie S (2006) Linking stochastic dynamics to population distribution: An analytical framework of gene expression. *Physical Review Letters* 97:168302.
- [19] Shahrezaei V, Swain P (2008) Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences of the USA* 105:17256–17261.
- [20] Paulsson J, Ehrenberg M (2001) Noise in a minimal regulatory network: plasmid copy number control. *Quarterly Reviews of Biophysics* 34:1–59.
- [21] Gnu'gge R, Liphardt T, Rudolf F (2016) A shuttle vector series for precise genetic engineering of *Saccharomyces cerevisiae*. *Yeast* 33:83–98.
- [22] Huh D, Paulsson J (2011) Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics* 43:95–100.
- [23] Nielsen A, et al. (2016) Genetic circuit design automation. *Science* 352:aac7341.
- [24] Del Vecchio D, Ninfa A, Sontag E (2008) Modular cell biology: retroactivity and insulation. *Molecular systems biology* 4:161.
- [25] Qian Y, Huang H, Jimnez J, Del Vecchio D (2017) Resource competition shapes the response of genetic circuits. *ACS Synthetic Biology* 6:12631272.
- [26] Hasty J, Pradines J, Dolnik M, Collins J (2000) Noise-based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences of the USA* 97:2075–2080.
- [27] Tan C, Marguet P, Lingchong Y (2009) Emergent bistability by a growth-modulating positive feedback circuit. *Nature Chemical Biology* 5:842–848.
- [28] Shahrezaei V, Marguerat V (2015) Connecting growth with gene expression: of noise and numbers. *Current Opinion in Microbiology* 25:127–135.

- [29] Ghusinga K, Dennehy J, Singh A (2017) First-passage time approach to controlling noise in the timing of intracellular events. *Proceedings of the National Academy of Sciences of the USA* 114:693–698.
- [30] Ruess J, Plešska M, Guet C, Tkačik G (2019) Molecular noise of innate immunity shapes bacteria-phage ecologies. *PLoS Computational Biology* 15:e1007168.
- [31] Miano A, Liao M, Hasty J (2020) Inducible cell-to-cell signaling for tunable dynamics in microbial communities. *Nature Communications* 11:1193.
- [32] Bertaux F, Stoma S, Drasdo D, Batt G (2014) Modeling dynamics of cell-to-cell variability in trail-induced apoptosis explains fractional killing and predicts reversible resistance. *PLoS Computational Biology* 10:e1003893.
- [33] Wakamoto Y, et al. (2013) Dynamic persistence of antibiotic-stressed mycobacteria. *Science* 339:91–95.
- [34] Duso L, Zechner C (2020) Stochastic reaction networks in dynamic compartment populations. *Proceedings of the National Academy of Sciences of the USA* 117:22674–22683.

Annex 6

STRAIN GENOTYPES

Strain Name (in the order of appearance in the text)	Genotype
Wild type	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0
Dual Reporter	MATa his3Δ:: pEL222_5x_Gal-Cre-tENO2-HIS3 leu2Δ::pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS - VP16AD - EL222-tSSA1-URA3
EL222 Reporter	MATa his3Δ1 leu2Δ::pEL222_5x_Gal-mscarlet1-tENO2-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS - VP16AD - EL222-tSSA1-URA3
Asynchronous program	MATa his3Δ:: pTDH3_LoxP-mNeonGreen-tENO1_LoxP-ATAF1-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tSSA1-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-URA3
Sequential program	MATa his3Δ:: pTDH3_LoxP-mNeonGreen-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tENO1_LoxP-ATAF1-tTDH1-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-URA3
pTDH3-mScarlet-I (1x)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ:: pTDH3-mScarlet_I-tTDH1-URA3
pTDH3-mCerulean (1x)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ:: pTDH3-mCerulean-tTDH1-URA3
pTDH3-mNeonGreen (1x)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ:: pTDH3-mNeonGreen-tTDH1-URA3
pTDH3-mVenus (1x)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ:: pTDH3-mVenus-tTDH1-URA3
pTDH3-3c (3x)	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ:: pTDH3-mCerulean-tTDH1-pTDH3-mNeonGreen-tTDH1- pTDH3- mScarlet_I -tTDH1-URA3
pTDH3-4c (4x)	MATa his3Δ:: pTDH3-mVenus-tTDH1-HIS3 leu2Δ0 met15Δ0 ura3Δ:: pTDH3-mCerulean-tTDH1-pTDH3-mNeonGreen-tTDH1- pTDH3-mScarlet_I -tTDH1-URA3
optoATAF1-mCerulean	MATa his3Δ:: pATAF1_4x-mCerulean-tDIT1 leu2Δ::pEL222_6x_ATAF1-tTDH1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS-VP16AD-EL222-tSSA1-URA3
optoATAF1-mNeonGreen	MATa his3Δ:: pATAF1_4x-mNeonGreen-tDIT1 leu2Δ::pEL222_6x_ATAF1-tTDH1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS-VP16AD-EL222-tSSA1-URA3
optoATAF1-mScarlet-I	MATa his3Δ:: pATAF1_4x-mScarlet_I-tDIT1 leu2Δ:: pEL222_6x_ATAF1-tTDH1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS-VP16AD-EL222-tSSA1-URA3
pTDH3-mScarlet-I pTDH3-EL222	MATa his3Δ:: pTDH3-mScarlet_I-tTDH1-HIS3 leu2Δ0 met15Δ0 ura3Δ:: pTDH3-SV40 NLS-VP16AD-EL222-tSSA1-URA3
optoATAF1 pTDH3-mScarlet-I	MATa his3Δ:: pTDH3-mScarlet_I-tTDH1-HIS3 leu2Δ:: pEL222_6x_ATAF1-tTDH1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS-VP16AD-EL222-tSSA1-URA3

optoATAF1-FAR1M	MATa his3Δ:: pATAF1_4x-FAR1M_mCerulean -tDIT1 leu2Δ:: pEL222_6x_ATAF1-tTDH1-LEU2 met15Δ0 ura3Δ:: pTDH3-SV40 NLS-VP16AD-EL222-tSSA1-URA3
L3-I7	MATa his3Δ::pATAF1_4x-FAR1M_mCerulean-tDIT1-HIS3 leu2Δ:: EL222_5X_Gal-CRE-tENO2-LEU2 met15Δ0 ura3Δ::pTDH3-LoxP-EL222-tENO1-LoxP-ATAF1-tTDH1-URA3
GAuDi00	MATa his3Δ:: pTDH3_LoxP-mNeonGreen-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tENO1_LoxP-ATAF1-tTDH1-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ0
GAuDi01	MATa his3Δ1 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3_LoxP-mNeonGreen-tENO1_LoxP-ATAF1-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tSSA1-URA3
optoATAF1-mSc-1X	MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ:: pATAF1_4x-mScarlet-tDIT1-EL222_6x-ATAF1-tTDH1-pTDH3-EL222-tSSA1-URA3
optoATAF1-mSc-2X FB	MATa his3Δ:: pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-ATAF1-tRPL41b-HIS3 leu2Δ0 met15Δ0 ura3Δ:: pATAF1_4x-mScarlet-tDIT1-EL222_6x-ATAF1-tTDH1-pTDH3-EL222-tSSA1-URA3
optoATAF1 2XC	MATa his3Δ:: pATAF1_4x-mCerulean-tDIT1 -HIS3 leu2Δ0 met15Δ0 ura3Δ:: pATAF1_4x-mScarlet-tDIT1-EL222_6x-ATAF1-tTDH1-pTDH3-EL222-tSSA1-URA3
optoATAF1 2XC FB	MATa his3Δ:: pATAF1_4x-mCerulean-tDIT1-pATAF1_4x-ATAF1-tRPL41b-HIS3 leu2Δ0 met15Δ0 ura3Δ:: pATAF1_4x-mScarlet-tDIT1-EL222_6x-ATAF1-tTDH1-pTDH3-EL222-tSSA1-URA3
optoATAF1 2XN	MATa his3Δ:: pATAF1_4x-mNeonGreen-tDIT1 -HIS3 leu2Δ0 met15Δ0 ura3Δ:: pATAF1_4x-mScarlet-tDIT1-EL222_6x-ATAF1-tTDH1-pTDH3-EL222-tSSA1-URA3
GAuDi02	MATa his3Δ:: pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-ATAF1-tRPL41b-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3_LoxP-mNeonGreen-tENO1_LoxP-ATAF1-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tSSA1-URA3
GAuDi03	MATa his3Δ:: pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-ATAF1-tRPL41b-pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M-tDIT1-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3_LoxP-mNeonGreen-tENO1_LoxP-ATAF1-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tSSA1-URA3
GAuDi04	MATa his3Δ:: pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-ATAF1-tRPL41b-pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean -tDIT1-HIS3 leu2Δ::pATAF1_4x-mScarlet-tDIT1-pATAF1_4x-Far1M_mCerulean-tDIT1-LEU2 met15Δ0 ura3Δ:: pTDH3_LoxP-mNeonGreen-tENO1_LoxP-ATAF1-tTDH1-pEL222_5x_Gal-Cre-tENO2-pTDH3-EL222-tSSA1-URA3
Tetra reporter integrated	MATa his3Δ1 leu2Δ::pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-LEU2 met15Δ0 ura3Δ::pEL222_5x_Gal-mScarlet-I-tENO2-pEL222_5xG-Cre-tENO2-pTDH3-EL222-mVenus-tTDH1-URA3
Tetra reporter centromeric	MATa his3Δ1 leu2Δ::pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-LEU2 met15Δ0 ura3Δ0 centromeric plasmid (pEL222_5x_Gal-mScarlet-I-tENO2-pEL222_5xG-Cre-tENO2-pTDH3-EL222-mVenus-tTDH1-URA3-ARS4/CEN6)

Tetra reporter 2micron	MATa his3 Δ 1 leu2 Δ ::pTDH3-LoxP-mCerulean-tENO1-LoxP-mNeonGreen-tTDH1-LEU2 met15 Δ 0 ura3 Δ 0 2micron plasmid (pEL222_5x_Gal-mScarlet-I-tENO2-pEL222_5xG-Cre-tENO2-pTDH3-EL222-mVenus-tTDH1-URA3-2 μ ori)
------------------------	--

BIBLIOGRAPHY

- Aditya, Chetan, François Bertaux, Gregory Batt, and Jakob Ruess. 2021. "A Light Tunable Differentiation System for the Creation and Control of Consortia in Yeast." *BioRxiv*.
- "Advisory Groups - MicrobiomeSupport." 2020. 2020. <https://www.microbiomesupport.eu/advisory-groups/>.
- Aguilera-Castrejon, Alejandro, Bernardo Oldak, Tom Shani, Nadir Ghanem, Chen Itzkovich, Sharon Slomovich, Shadi Tarazi, et al. 2021. "Ex Utero Mouse Embryogenesis from Pre-Gastrulation to Late Organogenesis." *Nature*, 1–6.
- Ahmad, Mudassar, Melanie Hirz, Harald Pichler, and Helmut Schwab. 2014. "Protein Expression in *Pichia Pastoris*: Recent Achievements and Perspectives for Heterologous Protein Production." *Applied Microbiology and Biotechnology* 98 (12): 5301–17.
- Albuquerque, Patricia, and Arturo Casadevall. 2012. "Quorum Sensing in Fungi--a Review." *Medical Mycology* 50 (4): 337–45.
- Andersson, Robin, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, et al. 2014. "An Atlas of Active Enhancers across Human Cell Types and Tissues." *Nature* 507 (7493): 455–61.
- Aoki, Stephanie K, Gabriele Lillacci, Ankit Gupta, Armin Baumschlager, David Schweingruber, and Mustafa Khammash. 2019. "A Universal Biomolecular Integral Feedback Controller for Robust Perfect Adaptation." *Nature* 570 (7762): 533–37.
- Arguelles, Juan-Carlos. 2017. "The Early Days of the Nobel Prize and Golden Age of Microbiology." *Hektoen International*.
- Arumugam, Manimozhiyan, Jeroen Raes, Eric Pelletier, Denis Le Paslier, Takuji Yamada, Daniel R Mende, Gabriel R Fernandes, et al. 2011. "Enterotypes of the Human Gut Microbiome." *Nature* 473 (7346): 174–80.
- Atkinson, Mariette R, Michael A Savageau, Jesse T Myers, and Alexander J Ninfa. 2003. "Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in *Escherichia Coli*." *Cell* 113 (5): 597–607.
- Bai Flagfeldt, Dongmei, Verena Siewers, Le Huang, and Jens Nielsen. 2009. "Characterization of Chromosomal Integration Sites for Heterologous Gene Expression in *Saccharomyces Cerevisiae*." *Yeast* 26 (10): 545–51.
- Balagaddé, Frederick K, Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, Stephen R Quake, and Lingchong You. 2008. "A Synthetic *Escherichia Coli* Predator-Prey Ecosystem." *Molecular Systems Biology* 4 (1): 187.
- Baneyx, François, and Mirna Mujacic. 2004. "Recombinant Protein Folding and Misfolding in *Escherichia Coli*." *Nature Biotechnology* 22 (11): 1399–1408.
- Barbier, Içvara, Rubén Perez-Carrasco, and Yolanda Schaerli. 2020. "Controlling Spatiotemporal Pattern Formation in a Concentration Gradient with a Synthetic Toggle Switch." *Molecular Systems Biology* 16 (6): e9361.
- Bárcena, Clea, Rafael Valdés-Mas, Pablo Mayoral, Cecilia Garabaya, Sylvère Durand, Francisco Rodríguez, María Teresa Fernández-García, et al. 2019. "Healthspan and Lifespan Extension by Fecal Microbiota Transplantation into Progeroid Mice." *Nature Medicine* 25 (8): 1234–42.
- Bardgett, Richard D, Chris Freeman, and Nicholas J Ostle. 2008. "Microbial Contributions to Climate Change through Carbon Cycle Feedbacks." *The ISME Journal* 2 (8): 805–14.
- Barton, L L, and D E Northup. 2011. *Microbial Ecology*. Wiley.
- Baruch, Erez N, Ilan Youngster, Guy Ben-Betzalel, Rona Ortenberg, Adi Lahat, Lior Katz, Katerina Adler, et al. 2021. "Fecal Microbiota Transplant Promotes Response in Immunotherapy-Refractory Melanoma Patients." *Science* 371 (6529): 602–9.

- Basan, Markus, Tomoya Honda, Dimitris Christodoulou, Manuel Hörl, Yu-Fang Chang, Emanuele Leoncini, Avik Mukherjee, et al. 2020. "A Universal Trade-off between Growth and Lag in Fluctuating Environments." *Nature* 584 (7821): 470–74.
- Basu, Subhayu, Yoram Gerchman, Cynthia H Collins, Frances H Arnold, and Ron Weiss. 2005. "A Synthetic Multicellular System for Programmed Pattern Formation." *Nature* 434 (7037): 1130–34.
- Baumschlager, Armin, and Mustafa Khammash. 2021. "Synthetic Biological Approaches for Optogenetics and Tools for Transcriptional Light-Control in Bacteria." *Advanced Biology*, 2000256.
- Becskei, Attila, Bertrand Séraphin, and Luis Serrano. 2001. "Positive Feedback in Eukaryotic Gene Networks: Cell Differentiation by Graded to Binary Response Conversion." *The EMBO Journal* 20 (10): 2528–35.
- Bello, Maria G Dominguez, Rob Knight, Jack A Gilbert, and Martin J Blaser. 2018. "Preserving Microbial Diversity." *Science* 362 (6410): 33–34.
- Benzinger, Dirk, and Mustafa Khammash. 2018. "Pulsatile Inputs Achieve Tunable Attenuation of Gene Expression Variability and Graded Multi-Gene Regulation." *Nature Communications* 9 (1): 1–10.
- Berg, Gabriele, Daria Rybakova, Doreen Fischer, Tomislav Cernava, Marie-Christine Champomier Vergès, Trevor Charles, Xiaoyulong Chen, et al. 2020. "Microbiome Definition Re-Visited: Old Concepts and New Challenges." *Microbiome* 8 (1): 103.
- Bertaux, François, Sebastián Sosa Carrillo, Achille Fraisse, Chetan Aditya, Mariela Furstenheim, and Gregory Batt. 2020. "Enhancing Multi-Bioreactor Platforms for Automated Measurements and Reactive Experiment Control." *BioRxiv*.
- Bertaux, François, Szymon Stoma, Dirk Drasdo, and Gregory Batt. 2014. "Modeling Dynamics of Cell-to-Cell Variability in TRAIL-Induced Apoptosis Explains Fractional Killing and Predicts Reversible Resistance." *PLoS Computational Biology* 10 (10): e1003893.
- Blakemore, David C, Luis Castro, Ian Churcher, David C Rees, Andrew W Thomas, David M Wilson, and Anthony Wood. 2018. "Organic Synthesis Provides Opportunities to Transform Drug Discovery." *Nature Chemistry* 10 (4): 383–94.
- Bober, Josef R, Chase L Beisel, and Nikhil U Nair. 2018. "Synthetic Biology Approaches to Engineer Probiotics and Members of the Human Microbiota for Biomedical Applications." *Annual Review of Biomedical Engineering* 20 (1): 277–300.
- Bouma, Judith E, and Richard E Lenski. 1988. "Evolution of a Bacteria/Plasmid Association." *Nature* 335 (6188): 351–52.
- Branch, Mary Ann, Thomas F Coleman, and Yuying Li. 1999. "A Subspace, Interior, and Conjugate Gradient Method for Large-Scale Bound-Constrained Minimization Problems." *SIAM Journal on Scientific Computing* 21 (1): 1–23.
- Brenner, Katie, Lingchong You, and Frances H Arnold. 2008. "Engineering Microbial Consortia: A New Frontier in Synthetic Biology." *Trends in Biotechnology* 26 (9): 483–89.
- Brock, Thomas Dale, Michael T Madigan, John M Martinko, and Jack Parker. 2003. *Brock Biology of Microorganisms*. Upper Saddle River (NJ): Prentice-Hall, 2003.
- Bulgarelli, Davide, Klaus Schlaeppi, Stijn Spaepen, Emiel Ver Loren van Themaat, and Paul Schulze-Lefert. 2013. "Structure and Functions of the Bacterial Microbiota of Plants." *Annual Review of Plant Biology* 64 (1): 807–38.
- Burg, Jonathan M, Charles B Cooper, Zhixia Ye, Benjamin R Reed, Eirik A Moreb, and Michael D Lynch. 2016. "Large-Scale Bioprocess Competitiveness: The Potential of Dynamic Metabolic Control in Two-Stage Fermentations." *Current Opinion in Chemical Engineering* 14: 121–36.
- Cammarota, Giovanni, Gianluca Ianiro, and Antonio Gasbarrini. 2014. "Fecal Microbiota Transplantation for the Treatment of Clostridium Difficile Infection: A Systematic Review." *Journal of Clinical Gastroenterology* 48 (8).
- Cankorur-Cetinkaya, Ayca, Nathalie Narraidoo, Ceyda Kasavi, Nigel K H Slater, David B Archer, and Stephen G Oliver. 2018. "Process Development for the Continuous Production of Heterologous Proteins by the Industrial Yeast, *Komagataella Phaffii*." *Biotechnology and Bioengineering* 115 (12): 2962–73.

- Cao, Longxing, Inna Goreshnik, Brian Coventry, James Brett Case, Lauren Miller, Lisa Kozodoy, Rita E Chen, et al. 2020. "De Novo Design of Picomolar SARS-CoV-2 Miniprotein Inhibitors." *Science* 370 (6515): 426–31.
- Carrasco-López, César, Sergio A García-Echauri, Therese Kichuk, and José L Avalos. 2020. "Optogenetics and Biosensors Set the Stage for Metabolic Cybergenetics." *Current Opinion in Biotechnology* 65: 296–309.
- Castle, Simeon D, Claire S Grierson, and Thomas E Gorochowski. 2021. "Towards an Engineering Theory of Evolution." *Nature Communications* 12 (1): 1–12.
- Cato, Marcus Porcius, Marcus Terentius Varro, and William D Hooper. 1935. "On Agriculture: With an Engl. Transl. by William Davis Hooper. Rev. by Harrison Boyd Ash." *Loeb Classical Library*.
- Ceroni, Francesca, Alice Boo, Simone Furini, Thomas E Gorochowski, Olivier Borkowski, Yaseen N Ladak, Ali R Awan, Charlie Gilbert, Guy-Bart Stan, and Tom Ellis. 2018. "Burden-Driven Feedback Control of Gene Expression." *Nature Methods* 15 (5): 387–93.
- Chait, Remy, Jakob Ruess, Tobias Bergmiller, Gašper Tkačik, and Valin C Guet. 2017. "Shaping Bacterial Population Behavior through Computer-Interfaced Control of Individual Cells." *Nature Communications* 8 (1): 1–11.
- Chang, Fred, and Ira Herskowitz. 1990. "Identification of a Gene Necessary for Cell Cycle Arrest by a Negative Growth Factor of Yeast: FAR1 Is an Inhibitor of a G1 Cyclin, CLN2." *Cell* 63 (5): 999–1011. [https://doi.org/10.1016/0092-8674\(90\)90503-7](https://doi.org/10.1016/0092-8674(90)90503-7).
- Chatelier, Emmanuelle Le, Trine Nielsen, Junjie Qin, Edi Prifti, Falk Hildebrand, Gwen Falony, Mathieu Almeida, et al. 2013. "Richness of Human Gut Microbiome Correlates with Metabolic Markers." *Nature* 500 (7464): 541–46.
- Chen, Kevin, Yixuan Zhu, Yongrong Zhang, Therwa Hamza, Hua Yu, Ashley Saint Fleur, James Galen, Zhiyong Yang, and Hanping Feng. 2020. "A Probiotic Yeast-Based Immunotherapy against *Clostridioides Difficile* Infection." *Science Translational Medicine* 12 (567).
- Chen, Yun, Siavash Partow, Gionata Scalcinati, Verena Siewers, and Jens Nielsen. 2012. "Enhancing the Copy Number of Episomal Plasmids in *Saccharomyces Cerevisiae* for Improved Protein Production." *FEMS Yeast Research* 12 (5): 598–607.
- Cho, Ilseung, and Martin J Blaser. 2012. "The Human Microbiome: At the Interface of Health and Disease." *Nature Reviews Genetics* 13 (4): 260–70.
- Choi, Jae Woong, Sung Sun Yim, Min Jeong Kim, and Ki Jun Jeong. 2015. "Enhanced Production of Recombinant Proteins with *Corynebacterium Glutamicum* by Deletion of Insertion Sequences (IS Elements)." *Microbial Cell Factories* 14 (1): 207.
- Church, George. 2020. "Technologies for Reading, Writing & Editing Omes." *Journal of Biomolecular Techniques: JBT* 31 (Suppl): S41.
- Colman-Lerner, Alejandro, Tina E. Chin, and Roger Brent. 2001. "Yeast Cbk1 and Mob2 Activate Daughter-Specific Genetic Programs to Induce Asymmetric Cell Fates." *Cell* 107 (6): 739–50. [https://doi.org/10.1016/S0092-8674\(01\)00596-7](https://doi.org/10.1016/S0092-8674(01)00596-7).
- Cosma, Maria Pia. 2004. "Daughter-Specific Repression of *Saccharomyces Cerevisiae* HO: Ash1 Is the Commander." *EMBO Reports* 5 (10): 953–57. <https://doi.org/10.1038/sj.embor.7400251>.
- Croughan, Matthew S., Konstantin B. Konstantinov, and Charles Cooney. 2015. "The Future of Industrial Bioprocessing: Batch or Continuous?" *Biotechnology and Bioengineering* 112 (4): 648–51. <https://doi.org/10.1002/bit.25529>.
- Cruz-Pereira, Joana S, Kieran Rea, Yvonne M Nolan, Olivia F O'Leary, Timothy G Dinan, and John F Cryan. 2020. "Depression's Unholy Trinity: Dysregulated Stress, Immunity, and the Microbiome." *Annual Review of Psychology* 71 (1): 49–78.
- Cryan, John F, Kenneth J O'Riordan, Caitlin S M Cowan, Kiran V Sandhu, Thomaz F S Bastiaanssen, Marcus Boehme, Martin G Codagnone, et al. 2019. "The Microbiota-Gut-Brain Axis." *Physiological Reviews* 99 (4): 1877–2013.
- Cryan, John F, Kenneth J O'Riordan, Kiran Sandhu, Veronica Peterson, and Timothy G Dinan. 2020. "The Gut Microbiome in Neurological Disorders." *The Lancet Neurology* 19 (2): 179–94.

- Csörgő, Bálint, Tamás Fehér, Edit Tímár, Frederick R Blattner, and György Pósfai. 2012. "Low-Mutation-Rate, Reduced-Genome Escherichia Coli: An Improved Host for Faithful Maintenance of Engineered Genetic Constructs." *Microbial Cell Factories* 11 (1): 11.
- Dahlberg; Chao. 2003. "Amelioration of the Cost of Conjugative Plasmid Carriage in Escherichia Coli K12." *Genetics* 165 (4): 1641–1649.
- Davar, Diwakar, Amiran K Dzutsev, John A McCulloch, Richard R Rodrigues, Joe-Marc Chauvin, Robert M Morrison, Richelle N Deblasio, et al. 2021. "Fecal Microbiota Transplant Overcomes Resistance to Anti-PD-1 Therapy in Melanoma Patients." *Science* 371 (6529): 595–602.
- Decker, Carolyn J, and Roy Parker. 1993. "A Turnover Pathway for Both Stable and Unstable MRNAs in Yeast: Evidence for a Requirement for Deadenylation." *Genes & Development* 7 (8): 1632–43.
- Dertli, Enes, and Ahmet Hilmi Çon. 2017. "Microbial Diversity of Traditional Kefir Grains and Their Role on Kefir Aroma." *LWT-Food Science and Technology* 85: 151–57.
- Dick, Markus, Nicholas S Sarai, Michael W Martynowycz, Tamir Gonen, and Frances H Arnold. 2019. "Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation." *Journal of the American Chemical Society* 141 (50): 19817–22.
- Dien, Stephen Van. 2013. "From the First Drop to the First Truckload: Commercialization of Microbial Processes for Renewable Chemicals." *Current Opinion in Biotechnology* 24 (6): 1061–68.
- Din, M Omar, Tal Danino, Arthur Prindle, Matt Skalak, Jangir Selimkhanov, Kaitlin Allen, Ellixis Julio, et al. 2016. "Synchronized Cycles of Bacterial Lysis for in Vivo Delivery." *Nature* 536 (7614): 81–85.
- Duddu, Atchuta Srinivas, Sarthak Sahoo, Souvadra Hati, Siddharth Jhunjhunwala, and Mohit Kumar Jolly. 2020. "Multi-Stability in Cellular Differentiation Enabled by a Network of Three Mutually Repressing Master Regulators." *Journal of the Royal Society Interface* 17 (170): 20200631.
- Dundas, Paul. 2003. *The Jains*. Routledge. Routledge.
- Duplus-Bottin, H el ene, Martin Spichty, G erard Triqueneaux, Christophe Place, Philippe Emmanuel Mangeot, Th eophile Ohlmann, Franck Vittoz, and Ga el Yvert. 2021. "A Single-Chain and Fast-Responding Light-Inducible Cre Recombinase as a Novel Optogenetic Switch." *Elife* 10: e61268.
- Duso, Lorenzo, and Christoph Zechner. 2020. "Stochastic Reaction Networks in Dynamic Compartment Populations." *Proceedings of the National Academy of Sciences* 117 (37): 22674–83.
- Eiteman, Mark A, and Elliot Altman. 2006. "Overcoming Acetate in Escherichia Coli Recombinant Protein Fermentations." *Trends in Biotechnology* 24 (11): 530–36.
- Elowitz, Michael B, Arnold J Levine, Eric D Siggia, and Peter S Swain. 2002. "Stochastic Gene Expression in a Single Cell." *Science* 297 (5584): 1183–86.
- Engler, Carola, Romy Kandzia, and Sylvestre Marillonnet. 2008. "A One Pot, One Step, Precision Cloning Method with High Throughput Capability." *PloS One* 3 (11): e3647.
<https://doi.org/10.1371/journal.pone.0003647>.
- Erickson, David W, Severin J Schink, Vadim Patsalo, James R Williamson, Ulrich Gerland, and Terence Hwa. 2017. "A Global Resource Allocation Strategy Governs Growth Transition Kinetics of Escherichia Coli." *Nature* 551 (7678): 119–23.
- Falony, Gwen, Marie Joossens, Sara Vieira-Silva, Jun Wang, Youssef Darzi, Karoline Faust, Alexander Kurilshikov, et al. 2016. "Population-Level Analysis of Gut Microbiome Variation." *Science* 352 (6285): 560–64.
- Fang, Fang, Kirsty Salmon, Michael W Y Shen, Kimberly A Aeling, Elaine Ito, Becky Irwin, Uyen Phuong C Tran, G Wesley Hatfield, Nancy A Da Silva, and Suzanne Sandmeyer. 2011. "A Vector Set for Systematic Metabolic Engineering in Saccharomyces Cerevisiae." *Yeast* 28 (2): 123–36.
- Faust, Karoline, and Jeroen Raes. 2012. "Microbial Interactions: From Networks to Models." *Nature Reviews Microbiology* 10 (8): 538–50.
- Fedorec, Alex J H, Behzad D Karkaria, Michael Sulu, and Chris P Barnes. 2021. "Single Strain Control of Microbial Consortia." *Nature Communications* 12 (1): 1–12.
- Fernandez-Rodriguez, Jesus, Felix Moser, Miryoung Song, and Christopher A Voigt. 2017. "Engineering RGB

- Color Vision into Escherichia Coli." *Nature Chemical Biology* 13 (7): 706–8.
- Foster, Kevin R, and Thomas Bell. 2012. "Competition, Not Cooperation, Dominates Interactions among Culturable Microbial Species." *Current Biology* 22 (19): 1845–50.
- Fox, Zachary R, Steven Fletcher, Achille Fraisse, Chetan Aditya, Sebastián Sosa-Carrillo, Sébastien Gilles, François Bertaux, Jakob Ruess, and Gregory Batt. 2021. "MicroMator: Open and Flexible Software for Reactive Microscopy." *BioRxiv*.
- Franziska, Hempel, Bozarth Andrew, Lindenkamp Nicole, Klingl Andreas, Zauner Stefan, and others. 2011. "Microalgae as Bioreactors for Bioplastic Production."
- Freilich, Shiri, Raphy Zarecki, Omer Eilam, Ella Shtifman Segal, Christopher S Henry, Martin Kupiec, Uri Gophna, Roded Sharan, and Eytan Ruppin. 2011. "Competitive and Cooperative Metabolic Interactions in Bacterial Communities." *Nature Communications* 2 (1): 589.
- Friedman, Nir, Long Cai, and X Sunney Xie. 2006. "Linking Stochastic Dynamics to Population Distribution: An Analytical Framework of Gene Expression." *Physical Review Letters* 97 (16): 168302.
- Fushimi, Keiji, and Rei Narikawa. 2019. "Cyanobacteriochromes: Photoreceptors Covering the Entire UV-to-Visible Spectrum." *Current Opinion in Structural Biology* 57: 39–46.
- Galloway, J N, F J Dentener, D G Capone, E W Boyer, R W Howarth, S P Seitzinger, G P Asner, et al. 2004. "Nitrogen Cycles: Past, Present, and Future." *Biogeochemistry* 70 (2): 153–226.
- Gans, Jason, Murray Wolinsky, and John Dunbar. 2005. "Computational Improvements Reveal Great Bacterial Diversity and High Metal Toxicity in Soil." *Science* 309 (5739): 1387–90.
- Gardner, Timothy S, Charles R Cantor, and James J Collins. 2000. "Construction of a Genetic Toggle Switch in Escherichia Coli." *Nature* 403 (6767): 339–42.
- Gasch, Audrey P. 2003. "The Environmental Stress Response: A Common Yeast Response to Diverse Environmental Stresses." In *Yeast Stress Responses*, edited by Stefan Hohmann and Willem H Mager, 11–70. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Gassler, Thomas, Michael Sauer, Brigitte Gasser, Michael Egermeier, Christina Troyer, Tim Causon, Stephan Hann, Diethard Mattanovich, and Matthias G Steiger. 2020. "The Industrial Yeast *Pichia Pastoris* Is Converted from a Heterotroph into an Autotroph Capable of Growth on CO₂." *Nature Biotechnology* 38 (2): 210–16.
- Geiler-Samerotte, Kerry A, Michael F Dion, Bogdan A Budnik, Stephanie M Wang, Daniel L Hartl, and D Allan Drummond. 2011. "Misfolded Proteins Impose a Dosage-Dependent Fitness Cost and Trigger a Cytosolic Unfolded Protein Response in Yeast." *Proceedings of the National Academy of Sciences* 108 (2): 680–85.
- Gelder, Leen De, Julia J Williams, Jose M Ponciano, Masahiro Sota, and Eva M Top. 2008. "Adaptive Plasmid Evolution Results in Host-Range Expansion of a Broad-Host-Range Plasmid." *Genetics* 178 (4): 2179–90.
- Geller, Stephanie H, Enoch B Antwi, Barbara Di Ventura, and Megan N McClean. 2019. "Optogenetic Repressors of Gene Expression in Yeasts Using Light-Controlled Nuclear Localization." *Cellular and Molecular Bioengineering* 12 (5): 511–28.
- Germann, Susanne M, Vera Schramke, Rune Troelsgaard Pedersen, Irene Gallina, Nadine Eckert-Boulet, Vibe H Oestergaard, and Michael Lisby. 2014. "TopBP1/Dpb11 Binds DNA Anaphase Bridges to Prevent Genome Instability." *The Journal of Cell Biology* 204 (1): 45–59.
- Geymonat, Marco, Adonis Spanos, and Steven G Sedgwick. 2007. "A *Saccharomyces Cerevisiae* Autoselection System for Optimised Recombinant Protein Expression." *Gene* 399 (2): 120–28.
- Gietz, R Daniel, and Robin A Woods. 2002. "Transformation of Yeast by Lithium Acetate/Single-Stranded Carrier DNA/Polyethylene Glycol Method." *Methods in Enzymology* 350: 87–96.
- Gifford, Casey A, Michael J Ziller, Hongcang Gu, Cole Trapnell, Julie Donaghey, Alexander Tsankov, Alex K Shalek, et al. 2013. "Transcriptional and Epigenetic Dynamics during Specification of Human Embryonic Stem Cells." *Cell* 153 (5): 1149–63.
- Gillespie, Daniel T. 1992. "A Rigorous Derivation of the Chemical Master Equation." *Physica A: Statistical Mechanics and Its Applications* 188 (1–3): 404–25.
- Gleizer, Shmuel, Roe Ben-Nissan, Yinon M Bar-On, Niv Antonovsky, Elad Noor, Yehudit Zohar, Ghil Jona, et al.

2019. "Conversion of Escherichia Coli to Generate All Biomass Carbon from CO₂." *Cell* 179 (6): 1255–63.
- Gnügge, Robert, Thomas Liphardt, and Fabian Rudolf. 2016. "A Shuttle Vector Series for Precise Genetic Engineering of *Saccharomyces Cerevisiae*." *Yeast* 33 (3): 83–98.
- Godawat, Rahul, Konstantin Konstantinov, Mahsa Rohani, and Veena Warikoo. 2015. "End-to-End Integrated Fully Continuous Production of Recombinant Monoclonal Antibodies." *Journal of Biotechnology* 213: 13–19.
- Goodrich, Julia K, Jillian L Waters, Angela C Poole, Jessica L Sutter, Omry Koren, Ran Blekhman, Michelle Beaumont, et al. 2014. "Human Genetics Shape the Gut Microbiome." *Cell* 159 (4): 789–99.
- Grandclément, Catherine, Mélanie Tannières, Solange Moréra, Yves Dessaux, and Denis Faure. 2016. "Quorum Quenching: Role in Nature and Applied Developments." *FEMS Microbiology Reviews* 40 (1): 86–116.
- Griffiths, A J, J H Miller, D T Suzuki, R C Lewontin, and W M Gelbart. 2000. "Bacterial Insertion Sequences." WH Freeman.
- Großkopf, Tobias, and Orkun S Soyer. 2014. "Synthetic Microbial Communities." *Current Opinion in Microbiology* 18: 72–77.
- Guimera, Roger, and Luis A Nunes Amaral. 2005. "Functional Cartography of Complex Metabolic Networks." *Nature* 433 (7028): 895–900.
- Guo, Liang, Wenwen Diao, Cong Gao, Guipeng Hu, Qiang Ding, Chao Ye, Xiulai Chen, Jia Liu, and Liming Liu. 2020. "Engineering Escherichia Coli Lifespan for Enhancing Chemical Production." *Nature Catalysis* 3 (3): 307–18.
- Gupta, Apoorv, Irene M Brockman Reizman, Christopher R Reisch, and Kristala L J Prather. 2017. "Dynamic Regulation of Metabolic Flux in Engineered Bacteria Using a Pathway-Independent Quorum-Sensing Circuit." *Nature Biotechnology* 35 (3): 273–79.
- Hammarlund, Sarah P, Jeremy M Chacón, and William R Harcombe. 2019. "A Shared Limiting Resource Leads to Competitive Exclusion in a Cross-Feeding System." *Environmental Microbiology* 21 (2): 759–71.
- Hammarlund, Sarah P, Tomáš Gedeon, Ross P Carlson, and William R Harcombe. 2021. "Limitation by a Shared Mutualist Promotes Coexistence of Multiple Competing Partners." *Nature Communications* 12 (1): 1–8.
- Harrigan, Patrick, Hiten D Madhani, and Hana El-Samad. 2018. "Real-Time Genetic Compensation Defines the Dynamic Demands of Feedback Control." *Cell* 175 (3): 877–86.
- Harrison, E, D Guymer, A J Spiers, S Paterson - Current Biology, and undefined 2015. 2015. "Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum." *Current Biology* 25 (15): 2034–39.
- Haynes, Cole M, Eric A Titus, and Antony A Cooper. 2004. "Degradation of Misfolded Proteins Prevents ER-Derived Oxidative Stress and Cell Death." *Molecular Cell* 15 (5): 767–76.
- Heijden, Marcel G A Van Der, Richard D Bardgett, and Nico M Van Straalen. 2008. "The Unseen Majority: Soil Microbes as Drivers of Plant Diversity and Productivity in Terrestrial Ecosystems." *Ecology Letters* 11 (3): 296–310.
- Henchoz, S, Y Chi, B Catarin, I Herskowitz, R J Deshaies, and M Peter. 1997. "Phosphorylation- and Ubiquitin-Dependent Degradation of the Cyclin-Dependent Kinase Inhibitor Far1 p in Budding Yeast." *Genes & Dev.* <https://doi.org/10.1101/gad.11.22.3046>.
- Henchoz, Sandra, Yong Chi, Barbara Catarin, Ira Herskowitz, Raymond J Deshaies, and Matthias Peter. 1997. "Phosphorylation-and Ubiquitin-Dependent Degradation of the Cyclin-Dependent Kinase Inhibitor Far1 p in Budding Yeast." *Genes & Development* 11 (22): 3046–60.
- Hilbrig, Frank, and Ruth Freitag. 2003. "Continuous Annular Chromatography." *Journal of Chromatography B* 790 (1): 1–15.
- Hochrein, Lena, Fabian Machens, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2017. "PhiReX: A Programmable and Red Light-Regulated Protein Expression Switch for Yeast." *Nucleic Acids Research* 45 (15): 9193–9205. <https://doi.org/10.1093/nar/gkx610>.
- Hochrein, Lena, Leslie A. Mitchell, Karina Schulz, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2018. "L-SCRaMble as a Tool for Light-Controlled Cre-Mediated Recombination in Yeast." *Nature Communications* 9

(1): 1–10. <https://doi.org/10.1038/s41467-017-02208-6>.

- Hom, Erik F Y, and Andrew W Murray. 2014. "Niche Engineering Demonstrates a Latent Capacity for Fungal-Algal Mutualism." *Science* 345 (6192): 94–98.
- Hooper, D U, F S Chapin III, J J Ewel, A Hector, P Inchausti, S Lavorel, J H Lawton, et al. 2005. "Effects Of Biodiversity On Ecosystem Functioning: A Consensus Of Current Knowledge." *Ecological Monographs* 75 (1): 3–35.
- Huh, Dann, and Johan Paulsson. 2011. "Non-Genetic Heterogeneity from Stochastic Partitioning at Cell Division." *Nature Genetics* 43 (2): 95–100.
- Hungate, Robert E. 2013. *The Rumen and Its Microbes*. Elsevier.
- Idiris, Alimjan, Hideki Tohda, Hiromichi Kumagai, and Kaoru Takegawa. 2010. "Engineering of Protein Secretion in Yeast: Strategies and Impact on Protein Production." *Applied Microbiology and Biotechnology* 86 (2): 403–17.
- Inda, María Eugenia, Esther Broset, Timothy K Lu, and Cesar de la Fuente-Nunez. 2019. "Emerging Frontiers in Microbiome Engineering." *Trends in Immunology* 40 (10): 952–73.
- Ispolatov, Iaroslav, Martin Ackermann, and Michael Doebeli. 2012. "Division of Labour and the Evolution of Multicellularity." *Proceedings of the Royal Society B: Biological Sciences* 279 (1734): 1768–76.
- Jang, Jaewan, Sherin McDonald, Maruti Uppalapati, and G Andrew Woolley. 2019. "Green, Orange, Red, and Far-Red Optogenetic Tools Derived from Cyanobacteriochromes." *BioRxiv*, 769422.
- Jia, Dongya, Mohit Kumar Jolly, William Harrison, Marcelo Boareto, Eshel Ben-Jacob, and Herbert Levine. 2017. "Operating Principles of Tristable Circuits Regulating Cellular Differentiation." *Physical Biology* 14 (3): 35007.
- Johnson, Heath E, Nareg J V Djabrayan, Stanislav Y Shvartsman, and Jared E Toettcher. 2020. "Optogenetic Rescue of a Patterning Mutant." *Current Biology* 30 (17): 3414–24.
- Jung, So Young, Hae Yong Yoo, Young Ho Kim, Jiyoung Kim, and Hyune Mo Rho. 1995. "The Glucose-Dependent Transactivation Activity of ABF1 on the Expression of the TDH3 Gene in Yeast." *Current Genetics* 27 (4): 312–17.
- Kafri, Moshe, Eyal Metzli-Raz, Ghil Jona, and Naama Barkai. 2016. "The Cost of Protein Production." *Cell Reports* 14 (1): 22–31.
- Kan, S B Jennifer, Russell D Lewis, Kai Chen, and Frances H Arnold. 2016. "Directed Evolution of Cytochrome c for Carbon-Silicon Bond Formation: Bringing Silicon to Life." *Science* 354 (6315): 1048–51.
- Karim, Ashty S, Kathleen A Curran, and Hal S Alper. 2013. "Characterization of Plasmid Burden and Copy Number in *Saccharomyces Cerevisiae* for Optimization of Metabolic Engineering Applications." *FEMS Yeast Research* 13 (1): 107–16.
- Karkaria, Behzad D, Alex J H Fedorec, and Chris P Barnes. 2021. "Automated Design of Synthetic Microbial Communities." *Nature Communications* 12 (1): 1–12.
- Kawano, Fuun, Risako Okazaki, Masayuki Yazawa, and Moritoshi Sato. 2016. "A Photoactivatable Cre-LoxP Recombination System for Optogenetic Genome Engineering." *Nature Chemical Biology* 12 (12): 1059–64. <https://doi.org/10.1038/nchembio.2205>.
- Kelley, Brian. 2009. "Industrialization of MAb Production Technology: The Bioprocessing Industry at a Crossroads." In *MAbs*, 1:443–52.
- Kong, Wentao, David R Meldgin, James J Collins, and Ting Lu. 2018. "Designing Microbial Consortia with Defined Social Interactions." *Nature Chemical Biology* 14 (8): 821–29.
- Konopka, Allan. 2009. "What Is Microbial Community Ecology?" *The ISME Journal* 3 (11): 1223–30.
- Kroll, Jens, Stefan Klinter, Cornelia Schneider, Isabella Voß, and Alexander Steinbüchel. 2010. "Plasmid Addiction Systems: Perspectives and Applications in Biotechnology." *Microbial Biotechnology* 3 (6): 634–57.
- Kwok, Roberta. 2010. "Five Hard Truths for Synthetic Biology." *Nature News* 463 (7279): 288–90.
- Kylilis, Nicolas, Zoltan A Tuza, Guy-Bart Stan, and Karen M Polizzi. 2018. "Tools for Engineering Coordinated

- System Behaviour in Synthetic Microbial Consortia.” *Nature Communications* 9 (1): 1–9.
- Lalwani, Makoto A, Samantha S Ip, Cesar Carrasco-Lopez, Catherine Day, Evan M Zhao, Hinako Kawabe, and José L Avalos. 2021. “Optogenetic Control of the Lac Operon for Bacterial Chemical and Protein Production.” *Nature Chemical Biology* 17 (1): 71–79.
- Lambert, Talley J. 2019. “FPbase: A Community-Editable Fluorescent Protein Database.” *Nature Methods* 16 (4): 277–78.
- Lane, Nick. 2015. “The Unseen World: Reflections on Leeuwenhoek (1677) ‘Concerning Little Animals.’” *Philosophical Transactions of the Royal Society B: Biological Sciences* 370 (1666): 20140344.
- Langer, ERIC S, and RONALD A Rader. 2014. “Continuous Bioprocessing and Perfusion: Wider Adoption Coming as Bioprocessing Matures.” *Bioprocess J* 13 (1).
- Larimer, Frank W, Patrick Chain, Loren Hauser, Jane Lamerdin, Stephanie Malfatti, Long Do, Miriam L Land, et al. 2004. “Complete Genome Sequence of the Metabolically Versatile Photosynthetic Bacterium *Rhodospseudomonas Palustris*.” *Nature Biotechnology* 22 (1): 55–61.
- LaSarre, Breah, Alexandra L McCully, Jay T Lennon, and James B McKinlay. 2017. “Microbial Mutualism Dynamics Governed by Dose-Dependent Toxicity of Cross-Fed Nutrients.” *The ISME Journal* 11 (2): 337–48.
- Lee, Michael E., William C. Deloache, Bernardo Cervantes, and John E. Dueber. 2015. “A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly.” *ACS Synthetic Biology* 4 (9): 975–86. <https://doi.org/10.1021/sb500366v>.
- Ley, Ruth E, Catherine A Lozupone, Micah Hamady, Rob Knight, and Jeffrey I Gordon. 2008. “Worlds within Worlds: Evolution of the Vertebrate Gut Microbiota.” *Nature Reviews Microbiology* 6 (10): 776–88.
- Li, Zhenghong, Xiaonan Wang, and Haoran Zhang. 2019. “Balancing the Non-Linear Rosmarinic Acid Biosynthetic Pathway by Modular Co-Culture Engineering.” *Metabolic Engineering* 54: 1–11.
- Liao, Michael J, M Omar Din, Lev Tsimring, and Jeff Hasty. 2019. “Rock-Paper-Scissors: Engineered Population Dynamics Increase Genetic Stability.” *Science* 365 (6457): 1045–49.
- Lillacci, Gabriele, Yaakov Benenson, and Mustafa Khammash. 2018. “Synthetic Control Systems for High Performance Gene Expression in Mammalian Cells.” *Nucleic Acids Research* 46 (18): 9855–63.
- Lin, Qihui, Hao Qi, Yi Wu, and Yingjin Yuan. 2015. “Robust Orthogonal Recombination System for Versatile Genomic Elements Rearrangement in Yeast *Saccharomyces Cerevisiae*.” *Scientific Reports* 5 (1): 1–8. <https://doi.org/10.1038/srep15249>.
- Lindell, Kristoffer. 2012. “Cell-to-Cell Communication and Virulence in *Vibrio Anguillarum*.” Umeåuniversitet.
- Lindner, Ariel B, Richard Madden, Alice Demarez, Eric J Stewart, and François Taddei. 2008. “Asymmetric Segregation of Protein Aggregates Is Associated with Cellular Aging and Rejuvenation.” *Proceedings of the National Academy of Sciences* 105 (8): 3076–81.
- Liu, Feng, Junwen Mao, Wentao Kong, Qiang Hua, Youjun Feng, Rashid Bashir, and Ting Lu. 2020. “Interaction Variability Shapes Succession of Synthetic Microbial Ecosystems.” *Nature Communications* 11 (1): 1–13.
- Liu, Jingyi, Xia Wu, Mingdong Yao, Wenhai Xiao, and Jian Zha. 2020. “Chassis Engineering for Microbial Production of Chemicals: From Natural Microbes to Synthetic Organisms.” *Current Opinion in Biotechnology* 66: 105–12.
- Liu, Ping, Bojun Chen, and Zhao-Wen Wang. 2020. “GABAergic Motor Neurons Bias Locomotor Decision-Making in *C. Elegans*.” *Nature Communications* 11 (1): 1–19.
- Liu, Quanli, Tao Yu, Xiaowei Li, Yu Chen, Kate Campbell, Jens Nielsen, and Yun Chen. 2019. “Rewiring Carbon Metabolism in Yeast for High Level Production of Aromatic Chemicals.” *Nature Communications* 10 (1): 1–13.
- Liu, Wei, Zhouqing Luo, Yun Wang, Nhan T Pham, Laura Tuck, Irene Pérez-Pi, Longying Liu, et al. 2018. “Rapid Pathway Prototyping and Engineering Using in Vitro and in Vivo Synthetic Genome SCRaMbLE-in Methods.” *Nature Communications* 9 (1): 1–12.
- Loison, Gérard, Martine Nguyen-Juilleret, Sami Alouani, and Magda Marquet. 1986. “Plasmid--Transformed Ura3 Fur1 Double-Mutants of *S. Cerevisiae*: An Autoselection System Applicable to the Production of

- Foreign Proteins." *Bio/Technology* 4 (5): 433–37.
- Louca, Stilianos, Martin F Polz, Florent Mazel, Michaeline B N Albright, Julie A Huber, Mary I O'Connor, Martin Ackermann, et al. 2018. "Function and Functional Redundancy in Microbial Systems." *Nature Ecology & Evolution* 2 (6): 936–43.
- Lozupone, Catherine A, and Rob Knight. 2007. "Global Patterns in Bacterial Diversity." *Proceedings of the National Academy of Sciences* 104 (27): 11436–40.
- Lozupone, Catherine A, Jesse I Stombaugh, Jeffrey I Gordon, Janet K Jansson, and Rob Knight. 2012. "Diversity, Stability and Resilience of the Human Gut Microbiota." *Nature* 489 (7415): 220–30.
- Lugagne, Jean-Baptiste, Sebastián Sosa Carrillo, Melanie Kirch, Agnes Köhler, Gregory Batt, and Pascal Hersen. 2017. "Balancing a Genetic Toggle Switch by Real-Time Feedback Control and Periodic Forcing." *Nature Communications* 8 (1): 1–8.
- Lunz, Davin, Gregory Batt, Jakob Ruess, and Joseph Frédéric Bonnans. 2020. "Beyond the Chemical Master Equation: Stochastic Chemical Kinetics Coupled with Auxiliary Processes."
- Luo, Xiaozhou, Michael A Reiter, Leo d'Espaux, Jeff Wong, Charles M Denby, Anna Lechner, Yunfeng Zhang, et al. 2019. "Complete Biosynthesis of Cannabinoids and Their Unnatural Analogues in Yeast." *Nature* 567 (7746): 123–26.
- Lv, Yongkun, Yang Gu, Jingliang Xu, Jingwen Zhou, and Peng Xu. 2020. "Coupling Metabolic Addiction with Negative Autoregulation to Improve Strain Stability and Pathway Yield." *Metabolic Engineering* 61: 79–88.
- Machado, Daniel, and Markus J Herrgård. 2015. "Co-Evolution of Strain Design Methods Based on Flux Balance and Elementary Mode Analysis." *Metabolic Engineering Communications* 2: 85–92.
- Maynard Smith, John, and Eors Szathmary. 1997. *The Major Transitions in Evolution*. Oxford University Press.
- McElroy, Christopher, and Stefan Jennewein. 2018. "Taxol® biosynthesis and Production: From Forests to Fermenters." In *Biotechnology of Natural Products*, 145–85. Springer.
- Mckinney, J D, and F R Cross. 1995. "Far1 and the g(1) Phase Specificity of Cell-Cycle Arrest by Mating Factor in *Saccharomyces-Cerevisiae* ." *Molecular and Cellular Biology* 15 (N5): 2509–16.
- McKinney, John D, and Frederick R Cross. 1995. "FAR1 and the G1 Phase Specificity of Cell Cycle Arrest by Mating Factor in *Saccharomyces Cerevisiae*." *Molecular and Cellular Biology* 15 (5): 2509–16.
- McMillan, D Randy, Mary-Jane Gething, and Joseph Sambrook. 1994. "The Cellular Response to Unfolded Proteins: Intercompartmental Signaling." *Current Opinion in Biotechnology* 5 (5): 540–45.
- Melkonian, Chrats, Willi Gottstein, Sonja Blasche, Yongkyu Kim, Martin Abel-Kistrup, Hentie Swiegers, Sofie Saerens, et al. 2019. "Finding Functional Differences between Species in a Microbial Community: Case Studies in Wine Fermentation and Kefir Culture." *Frontiers in Microbiology* 10: 1347.
- Menolascina, Filippo, Gianfranco Fiore, Emanuele Orabona, Luca De Stefano, Mike Ferry, Jeff Hasty, Mario di Bernardo, and Diego di Bernardo. 2014. "In-Vivo Real-Time Control of Protein Expression from Endogenous and Synthetic Gene Networks." *PLoS Comput Biol* 10 (5): e1003625.
- Miano, Arianna, Michael J Liao, and Jeff Hasty. 2020. "Inducible Cell-to-Cell Signaling for Tunable Dynamics in Microbial Communities." *Nature Communications* 11 (1): 1–8.
- Michener, Joshua K, Jens Nielsen, and Christina D Smolke. 2012. "Identification and Treatment of Heme Depletion Attributed to Overexpression of a Lineage of Evolved P450 Monooxygenases." *Proceedings of the National Academy of Sciences* 109 (47): 19504–9.
- Milias-Argeitis, Andreas, Marc Rullan, Stephanie K. Aoki, Peter Buchmann, and Mustafa Khammash. 2016. "Automated Optogenetic Feedback Control for Precise and Robust Regulation of Gene Expression and Cell Growth." *Nature Communications* 7 (1): 1–11. <https://doi.org/10.1038/ncomms12546>.
- Milias-Argeitis, Andreas, Sean Summers, Jacob Stewart-Ornstein, Ignacio Zuleta, David Pincus, Hana El-Samad, Mustafa Khammash, and John Lygeros. 2011. "In Silico Feedback for in Vivo Regulation of a Gene Expression Circuit." *Nature Biotechnology* 29 (12): 1114–16. <https://doi.org/10.1038/nbt.2018>.
- Miller, Melissa B, and Bonnie L Bassler. 2001. "Quorum Sensing in Bacteria." *Annual Review of Microbiology* 55 (1): 165–99.

- Mimee, Mark, Robert J Citorik, and Timothy K Lu. 2016. "Microbiome Therapeutics — Advances and Challenges." *Advanced Drug Delivery Reviews* 105: 44–54.
- Modi, R I, L H Castilla, S Puskas-Rozsa, R B Helling - Genetics, and undefined 1992. 1992. "Genetic Changes Accompanying Increased Fitness in Evolving Populations of Escherichia Coli." *Genetics* 130 (2): 241–249.
- Moeller, Andrew H, Alejandro Caro-Quintero, Deus Mjungu, Alexander V Georgiev, Elizabeth V Lonsdorf, Martin N Muller, Anne E Pusey, Martine Peeters, Beatrice H Hahn, and Howard Ochman. 2016. "Cospeciation of Gut Microbiota with Hominids." *Science* 353 (6297): 380–82.
- Montgomery, Kate, James C Charlesworth, Rebecca LeBard, Pieter T Visscher, and Brendan P Burns. 2013. "Quorum Sensing in Extreme Environments." *Life* 3 (1): 131–48.
- Motta-Mena, Laura B., Anna Reade, Michael J. Mallory, Spencer Glantz, Orion D. Weiner, Kristen W. Lynch, and Kevin H. Gardner. 2014. "An Optogenetic Gene Expression System with Rapid Activation and Deactivation Kinetics." *Nature Chemical Biology* 10 (3): 196–202. <https://doi.org/10.1038/nchembio.1430>.
- Muhlrad, Denise, and Roy Parker. 1999. "Aberrant MRNAs with Extended 3' UTRs Are Substrates for Rapid Degradation by mRNA Surveillance." *Rna* 5 (10): 1299–1307.
- Naseri, Gita, Salma Balazadeh, Fabian Machens, Iman Kamranfar, Katrin Messerschmidt, and Bernd Mueller-Roeber. 2017. "Plant-Derived Transcription Factors for Orthologous Regulation of Gene Expression in the Yeast *Saccharomyces Cerevisiae*." *ACS Synthetic Biology* 6 (9): 1742–56.
- Naseri, Gita, Jessica Behrend, Lisa Rieper, and Bernd Mueller-Roeber. 2019. "COMPASS for Rapid Combinatorial Optimization of Biochemical Pathways Based on Artificial Transcription Factors." *Nature Communications* 10 (1): 1–18.
- Nern, A., B. D. Pfeiffer, K. Svoboda, and G. M. Rubin. 2011. "Multiple New Site-Specific Recombinases for Use in Manipulating Animal Genomes." *Proceedings of the National Academy of Sciences* 108 (34): 14198–203. <https://doi.org/10.1073/pnas.1111704108>.
- Neuert, Gregor, Brian Munsy, Rui Zhen Tan, Leonid Teytelman, Mustafa Khammash, and Alexander Van Oudenaarden. 2013. "Systematic Identification of Signal-Activated Stochastic Gene Regulation." *Science* 339 (6119): 584–87.
- Neufeld, Len, Fabienne Stassen, Ruth Sheppard, and Terry Gilman. 2016. "The New Plastics Economy: Rethinking the Future of Plastics." In *World Economic Forum*. Vol. 7.
- Neurohr, Gabriel E, Rachel L Terry, Jette Lengefeld, Megan Bonney, Gregory P Brittingham, Fabien Moretto, Teemu P Miettinen, et al. 2019. "Excessive Cell Growth Causes Cytoplasm Dilution and Contributes to Senescence." *Cell* 176 (5): 1083–97.
- Nguyen, Van Dat, Feras Hatahet, Kirsi E H Salo, Eveliina Enlund, Chi Zhang, and Lloyd W Ruddock. 2011. "Pre-Expression of a Sulfhydryl Oxidase Significantly Increases the Yields of Eukaryotic Disulfide Bond Containing Proteins Expressed in the Cytoplasm of *E. Coli*." *Microbial Cell Factories* 10 (1): 1–13.
- Nielsen, Alec A K, Bryan S Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A Strychalski, David Ross, Douglas Densmore, and Christopher A Voigt. 2016. "Genetic Circuit Design Automation." *Science* 352 (6281).
- Nyström, Thomas, and Beidong Liu. 2014. "Protein Quality Control in Time and Space--Links to Cellular Aging." *FEMS Yeast Research* 14 (1): 40–48.
- Ohlendorf, Robert, Roe R Vidavski, Avigdor Eldar, Keith Moffat, and Andreas Möglich. 2012. "From Dusk till Dawn: One-Plasmid Systems for Light-Regulated Gene Expression." *Journal of Molecular Biology* 416 (4): 534–42.
- Olson, Evan J, Lucas A Hartsough, Brian P Landry, Raghav Shroff, and Jeffrey J Tabor. 2014. "Characterizing Bacterial Gene Circuit Dynamics with Optically Programmed Gene Expression Signals." *Nature Methods* 11 (4): 449–55.
- Omura, Satoshi. 2011. "Microbial Metabolites: 45 Years of Wandering, Wondering and Discovering." *Tetrahedron (Oxford. Print)* 67 (35).
- Paddon, Christopher J, P J Westfall, D J Pitera, K Benjamin, K Fisher, D McPhee, M D Leavell, et al. 2013. "High-Level Semi-Synthetic Production of the Potent Antimalarial Artemisinin." *Nature* 496 (7446): 528–32.

- Panke-Buisse, Kevin, Angela C Poole, Julia K Goodrich, Ruth E Ley, and Jenny Kao-Kniffin. 2015. "Selection on Soil Microbiomes Reveals Reproducible Impacts on Plant Function." *The ISME Journal* 9 (4): 980–89.
- Park, Jooyoung, Brinda Selvaraj, Andrew C McShan, Scott E Boyken, Kathy Y Wei, Gustav Oberdorfer, William DeGrado, et al. 2019. "De Novo Design of a Homo-Trimeric Amantadine-Binding Protein." *Elife* 8: e47839.
- Paulsson, Johan, and Måns Ehrenberg. 2001. "Noise in a Minimal Regulatory Network: Plasmid Copy Number Control." *Quarterly Reviews of Biophysics* 34 (1): 1–59.
- Pedone, Elisa, Irene De Cesare, Criseida G Zamora-Chimal, David Haener, Lorena Postiglione, Antonella La Regina, Barbara Shannon, et al. 2021. "Cheetah: A Computational Toolkit for Cybergenetic Control." *ACS Synthetic Biology* 10 (5): 979–89.
- Perkins, Melinda Liu, Dirk Benzinger, Murat Arcaç, and Mustafa Khammash. 2020. "Cell-in-the-Loop Pattern Formation with Optogenetically Emulated Cell-to-Cell Signaling." *Nature Communications* 11 (1): 1–10.
- "Plant-E." 2009. 2009. <https://www.plant-e.com/en/hoewerker-het/>.
- Pothoulakis, Georgios, and Tom Ellis. 2018. "Construction of Hybrid Regulated Mother-Specific Yeast Promoters for Inducible Differential Gene Expression." *PLoS ONE* 13 (3): 1–14. <https://doi.org/10.1371/journal.pone.0194588>.
- Potvin-Trottier, Laurent, Nathan D Lord, Glenn Vinnicombe, and Johan Paulsson. 2016. "Synchronous Long-Term Oscillations in a Synthetic Gene Circuit." *Nature* 538 (7626): 514–17.
- Pouzet, Sylvain, Alvaro Banderas, Matthias Le Bec, Thomas Lautier, Gilles Truan, and Pascal Hersen. 2020. "The Promise of Optogenetics for Bioproduction: Dynamic Control Strategies and Scale-up Instruments." *Bioengineering* 7 (4): 151.
- Qian, Yili, Hsin-Ho Huang, José I Jiménez, and Domitilla Del Vecchio. 2017. "Resource Competition Shapes the Response of Genetic Circuits." *ACS Synthetic Biology* 6 (7): 1263–72.
- Qiao, Kangjian, Thomas M Wasylenko, Kang Zhou, Peng Xu, and Gregory Stephanopoulos. 2017. "Lipid Production in *Yarrowia lipolytica* Is Maximized by Engineering Cytosolic Redox Metabolism." *Nature Biotechnology* 35 (2): 173–77.
- Qiu, Zhiguang, Eleonora Egidi, Hongwei Liu, Simranjit Kaur, and Brajesh K Singh. 2019. "New Frontiers in Agriculture Productivity: Optimised Microbial Inoculants and in Situ Microbiome Engineering." *Biotechnology Advances* 37 (6): 107371.
- Raab, Andreas M, Gabi Gebhardt, Natalia Bolotina, Dirk Weuster-Botz, and Christine Lang. 2010. "Metabolic Engineering of *Saccharomyces cerevisiae* for the Biotechnological Production of Succinic Acid." *Metabolic Engineering* 12 (6): 518–25.
- Raich, J W, and W H Schlesinger. 1992. "The Global Carbon Dioxide Flux in Soil Respiration and Its Relationship to Vegetation and Climate." *Tellus B* 44 (2): 81–99.
- Raj, Arjun, and Alexander Van Oudenaarden. 2008. "Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences." *Cell* 135 (2): 216–26.
- Rapp, Kent M, Jackson P Jenkins, and Michael J Betenbaugh. 2020. "Partners for Life: Building Microbial Consortia for the Future." *Current Opinion in Biotechnology* 66: 292–300.
- Raser, Jonathan M, and Erin K O'shea. 2005. "Noise in Gene Expression: Origins, Consequences, and Control." *Science* 309 (5743): 2010–13.
- Rathore, Anurag S, Harshit Agarwal, Abhishek Kumar Sharma, Mili Pathak, and SJPB Muthukumar. 2015. "Continuous Processing for Production of Biopharmaceuticals." *Preparative Biochemistry and Biotechnology* 45 (8): 836–49.
- Repina, Nicole A., Alyssa Rosenbloom, Abhirup Mukherjee, David V. Schaffer, and Ravi S. Kane. 2017. "At Light Speed: Advances in Optogenetic Systems for Regulating Cell Signaling and Behavior." *Annual Review of Chemical and Biomolecular Engineering* 8 (1): 13–39. <https://doi.org/10.1146/annurev-chembioeng-060816-101254>.
- Reynders, Martin, Bryan S Matsuura, Marleen Bérouti, Daniele Simoneschi, Antonio Marzio, Michele Pagano, and Dirk Trauner. 2020. "PHOTACs Enable Optical Control of Protein Degradation." *Science Advances* 6 (8): eaay5064.

- Richardson, Sarah M, Leslie A Mitchell, Giovanni Stracquadanio, Kun Yang, Jessica S Dymond, James E DiCarlo, Dongwon Lee, et al. 2017. "Design of a Synthetic Yeast Genome." *Science* 355 (6329): 1040–44.
- Robertson, Wesley E, Louise F H Funke, Daniel de la Torre, Julius Fredens, Thomas S Elliott, Martin Spinck, Yonka Christova, et al. 2021. "Sense Codon Reassignment Enables Viral Resistance and Encoded Polymer Synthesis." *Science* 372 (6546): 1057–62.
- Roell, Garrett W, Jian Zha, Rhiannon R Carr, Mattheos A Koffas, Stephen S Fong, and Yinjie J Tang. 2019. "Engineering Microbial Consortia by Division of Labor." *Microbial Cell Factories* 18 (1): 1–11.
- Romano, Edoardo, Armin Baumschlager, Emir Bora Akmeriç, Navaneethan Palanisamy, Moustafa Houmani, Gregor Schmidt, Mehmet Ali Öztürk, Leonard Ernst, Mustafa Khammash, and Barbara Di Ventura. 2021. "Engineering AraC to Make It Responsive to Light Instead of Arabinose." *Nature Chemical Biology*, 1–11.
- Roquet, Nathaniel, Ava P Soleimany, Alyssa C Ferris, Scott Aaronson, and Timothy K Lu. 2016. "Synthetic Recombinase-Based State Machines in Living Cells." *Science* 353 (6297).
- Rothschild, Daphna, Omer Weissbrod, Elad Barkan, Alexander Kurilshikov, Tal Korem, David Zeevi, Paul I Costea, et al. 2018. "Environment Dominates over Host Genetics in Shaping Human Gut Microbiota." *Nature* 555 (7695): 210–15.
- Rueffler, Claus, Joachim Hermisson, and Günter P Wagner. 2012. "Evolution of Functional Specialization and Division of Labor." *Proceedings of the National Academy of Sciences* 109 (6): E326–35.
- Ruess, Jakob, Maroš Pleška, Čálin C Guet, and Gašper Tkačik. 2019. "Molecular Noise of Innate Immunity Shapes Bacteria-Phage Ecologies." *PLoS Computational Biology* 15 (7): e1007168.
- Rugbjerg, Peter, Nils Myling-Petersen, Andreas Porse, Kira Sarup-Lytzen, and Morten O A Sommer. 2018. "Diverse Genetic Error Modes Constrain Large-Scale Bio-Based Production." *Nature Communications* 9 (1): 1–14.
- Rugbjerg, Peter, and Morten O A Sommer. 2019. "Overcoming Genetic Heterogeneity in Industrial Fermentations." *Nature Biotechnology* 37 (8): 869–76.
- Rullan, Marc, Dirk Benzinger, Gregor W. Schmidt, Andreas Miliadis-Argeitis, and Mustafa Khammash. 2018. "An Optogenetic Platform for Real-Time, Single-Cell Interrogation of Stochastic Transcriptional Regulation." *Molecular Cell* 70 (4): 745–756.e6. <https://doi.org/10.1016/j.molcel.2018.04.012>.
- Salzano, Davide, Davide Fiore, and Mario di Bernardo. 2019. "Ratiometric Control for Differentiation of Cell Populations Endowed with Synthetic Toggle Switches." In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 927–32.
- Scheffen, Marieke, Daniel G Marchal, Thomas Beneyton, Sandra K Schuller, Melanie Klose, Christoph Diehl, Jessica Lehmann, et al. 2021. "A New-to-Nature Carboxylation Module to Improve Natural and Synthetic CO₂ Fixation." *Nature Catalysis* 4 (2): 105–15.
- Schmidl, Sebastian R, Ravi U Sheth, Andrew Wu, and Jeffrey J Tabor. 2014. "Refactoring and Optimization of Light-Switchable Escherichia Coli Two-Component Systems." *ACS Synthetic Biology* 3 (11): 820–31.
- Scott, Matthew, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. 2010. "Interdependence of Cell Growth and Gene Expression: Origins and Consequences." *Science* 330 (6007): 1099–1102.
- Scott, Spencer R, M Omar Din, Philip Bittihn, Liyang Xiong, Lev S Tsimring, and Jeff Hasty. 2017. "A Stabilized Microbial Ecosystem of Self-Limiting Bacteria Using Synthetic Quorum-Regulated Lysis." *Nature Microbiology* 2 (8): 1–9.
- Seong, Jihye, and Michael Z Lin. 2021. "Optobiochemistry: Genetically Encoded Control of Protein Activity by Light." *Annual Review of Biochemistry* 90.
- Shahrezaei, Vahid, and Samuel Marguerat. 2015. "Connecting Growth with Gene Expression: Of Noise and Numbers." *Current Opinion in Microbiology* 25: 127–35.
- Shahrezaei, Vahid, and Peter S Swain. 2008. "Analytical Distributions for Stochastic Gene Expression." *Proceedings of the National Academy of Sciences* 105 (45): 17256–61.
- Shao, Bin, Jayan Rammohan, Daniel A Anderson, Nina Alperovich, David Ross, and Christopher A Voigt. 2021. "Single-Cell Measurement of Plasmid Copy Number and Promoter Activity." *Nature Communications* 12 (1):

- Shao, Yangyang, Ning Lu, Chen Cai, Fan Zhou, Shanshan Wang, Zhihu Zhao, Guoping Zhao, Jin-Qiu Zhou, Xiaoli Xue, and Zhongjun Qin. 2019. "A Single Circular Chromosome Yeast." *Cell Research* 29 (1): 87–89.
- Sheets, Michael B, Wilson W Wong, and Mary J Dunlop. 2020. "Light-Inducible Recombinases for Bacterial Optogenetics." *ACS Synthetic Biology* 9 (2): 227–35.
- Shendure, Jay, Shankar Balasubramanian, George M Church, Walter Gilbert, Jane Rogers, Jeffery A Schloss, and Robert H Waterston. 2017. "DNA Sequencing at 40: Past, Present and Future." *Nature* 550 (7676): 345–53.
- Sherwin, Eoin, Seth R Bordenstein, John L Quinn, Timothy G Dinan, and John F Cryan. 2019. "Microbiota and the Social Brain." *Science* 366 (6465).
- Sheth, Ravi U, and Harris H Wang. 2018. "DNA-Based Memory Devices for Recording Cellular Events." *Nature Reviews Genetics* 19 (11): 718–32.
- Shou, Wenying, Sri Ram, and Jose M G Vilar. 2007. "Synthetic Cooperation in Engineered Yeast Populations." *Proceedings of the National Academy of Sciences* 104 (6): 1877–82.
- Siller, Efraim, Diane C DeZwaan, John F Anderson, Brian C Freeman, and José M Barral. 2010. "Slowing Bacterial Translation Speed Enhances Eukaryotic Protein Folding Efficiency." *Journal of Molecular Biology* 396 (5): 1310–18.
- Smith, Jon, Marc Lipsitch, and Jeffrey W Almond. 2011. "Vaccine Production, Distribution, Access, and Uptake." *The Lancet* 378 (9789): 428–38.
- Smith, Mark B, Colleen Kelly, and Eric J Alm. 2014. "Policy: How to Regulate Faecal Transplants." *Nature* 506 (7488): 290–91.
- Sponzilli, Ivonne, and Luigi D Notarangelo. 2011. "Severe Combined Immunodeficiency (SCID): From Molecular Basis to Clinical Management." *Acta Bio-Medica: Atenei Parmensis* 82 (1): 5–13.
- Staley, J T, and A Konopka. 1985. "Measurement of in Situ Activities of Nonphotosynthetic Microorganisms in Aquatic and Terrestrial Habitats." *Annual Review of Microbiology* 39: 321–46.
- Steinebach, Fabian, Nicole Ulmer, Moritz Wolf, Lara Decker, Veronika Schneider, Ruben Wälchli, Daniel Karst, Jonathan Souquet, and Massimo Morbidelli. 2017. "Design and Operation of a Continuous Integrated Monoclonal Antibody Production Process." *Biotechnology Progress* 33 (5): 1303–13.
- Steiner, Ulrich Karl. 2021. "Senescence in Bacteria and Its Underlying Mechanisms." *Frontiers in Cell and Developmental Biology* 9.
- Strickland, Michael S, and Johannes Rousk. 2010. "Considering Fungal:Bacterial Dominance in Soils – Methods, Controls, and Ecosystem Implications." *Soil Biology and Biochemistry* 42 (9): 1385–95.
- Summers, Zarath M, Heather E Fogarty, Ching Leang, Ashley E Franks, Nikhil S Malvankar, and Derek R Lovley. 2010. "Direct Exchange of Electrons Within Aggregates of an Evolved Syntrophic Coculture of Anaerobic Bacteria." *Science* 330 (6009): 1413–15.
- Tabas, Ira, and David Ron. 2011. "Integrating the Mechanisms of Apoptosis Induced by Endoplasmic Reticulum Stress." *Nature Cell Biology* 13 (3): 184–90.
- Tan, Cheemeng, Philippe Marguet, and Lingchong You. 2009. "Emergent Bistability by a Growth-Modulating Positive Feedback Circuit." *Nature Chemical Biology* 5 (11): 842–48.
- Tanoue, Takeshi, Satoru Morita, Damian R Plichta, Ashwin N Skelly, Wataru Suda, Yuki Sugiura, Seiko Narushima, et al. 2019. "A Defined Commensal Consortium Elicits CD8 T Cells and Anti-Cancer Immunity." *Nature* 565 (7741): 600–605.
- Taslimi, Amir, Brian Zoltowski, Jose G. Miranda, Gopal P. Pathak, Robert M. Hughes, and Chandra L. Tucker. 2016. "Optimized Second-Generation CRY2-CIB Dimerizers and Photoactivatable Cre Recombinase." *Nature Chemical Biology* 12 (6): 425–30. <https://doi.org/10.1038/nchembio.2063>.
- Teng, Ying, and Wei Chen. 2019. "Soil Microbiomes—a Promising Strategy for Contaminated Soil Remediation: A Review." *Pedosphere* 29 (3): 283–97.

- "The Earth Microbiome Project." 2010. 2010. <https://earthmicrobiome.org/people/working-group-consortium/>.
- Thompson, Luke R, Jon G Sanders, Daniel McDonald, Amnon Amir, Joshua Ladau, Kenneth J Locey, Robert J Prill, et al. 2017. "A Communal Catalogue Reveals Earth's Multiscale Microbial Diversity." *Nature* 551 (7681): 457–63.
- Timmers, H Th Marc, and László Tora. 2018. "Transcript Buffering: A Balancing Act between mRNA Synthesis and mRNA Degradation." *Molecular Cell* 72 (1): 10–17.
- To, Tsz-Leung, and Narendra Maheshri. 2010. "Noise Can Induce Bimodality in Positive Transcriptional Feedback Loops without Bistability." *Science* 327 (5969): 1142–45.
- Toettcher, Jared E, Delquin Gong, Wendell A Lim, and Orion D Weiner. 2011. "Light-Based Feedback for Controlling Intracellular Signaling Dynamics." *Nature Methods* 8 (10): 837–39.
- Toettcher, Jared E, Orion D Weiner, and Wendell A Lim. 2013. "Using Optogenetics to Interrogate the Dynamic Control of Signal Transmission by the Ras/Erk Module." *Cell* 155 (6): 1422–34.
- Tokuhara, Daisuke. 2021. "Role of the Gut Microbiota in Regulating Non-Alcoholic Fatty Liver Disease in Children and Adolescents." *Frontiers in Nutrition* 8 (June): 700058.
- Torres, Eduardo M, Tanya Sokolsky, Cheryl M Tucker, Leon Y Chan, Monica Boselli, Maitreya J Dunham, and Angelika Amon. 2007. "Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast." *Science* 317 (5840): 916–24.
- Tournier, V, C M Topham, A Gilles, B David, C Folgoas, E Moya-Leclair, E Kamionka, et al. 2020. "An Engineered PET Depolymerase to Break down and Recycle Plastic Bottles." *Nature* 580 (7802): 216–19.
- Tremaroli, Valentina, and Fredrik Bäckhed. 2012. "Functional Interactions between the Gut Microbiota and Host Metabolism." *Nature* 489 (7415): 242–49.
- Uhlendorf, Jannis, Agnès Miermont, Thierry Delaveau, Gilles Charvin, François Fages, Samuel Bottani, Gregory Batt, and Pascal Hersen. 2012. "Long-Term Model Predictive Control of Gene Expression at the Population and Single-Cell Levels." *Proceedings of the National Academy of Sciences* 109 (35): 14271–76.
- Vallery-Radot, René. 1901. *La Vie de Pasteur*. Librairie Hachette.
- Vary, Patricia S, Rebekka Biedendieck, Tobias Fuerch, Friedhelm Meinhardt, Manfred Rohde, Wolf-Dieter Deckwer, and Dieter Jahn. 2007. "Bacillus Megaterium—from Simple Soil Bacterium to Industrial Protein Production Host." *Applied Microbiology and Biotechnology* 76 (5): 957–67.
- Vecchio, Domitilla Del, Yili Qian, Richard M Murray, and Eduardo D Sontag. 2018. "Future Systems and Control Research in Synthetic Biology." *Annual Reviews in Control* 45: 5–17.
- Wainwright, Milton B T - *Advances in Applied Microbiology*. 2003. "An Alternative View of the Early History of Microbiology." In , 52:333–55. Academic Press.
- Wakamoto, Yuichi, Neeraj Dhar, Remy Chait, Katrin Schneider, François Signorino-Gelo, Stanislas Leibler, and John D McKinney. 2013. "Dynamic Persistence of Antibiotic-Stressed Mycobacteria." *Science* 339 (6115): 91–95.
- Walls, Laura E, Koray Malca, Behnaz Nowrouzi, Rachel A Li, Leo D'Espaux, Jeff Wong, Jonathan A Dennis, et al. 2021. "Optimizing the Biosynthesis of Oxygenated and Acetylated Taxol Precursors in *Saccharomyces Cerevisiae* Using Advanced Bioprocessing Strategies." *Biotechnology and Bioengineering* 118 (1): 279–93.
- Walther, Jason, Rahul Godawat, Chris Hwang, Yuki Abe, Andrew Sinclair, and Konstantin Konstantinov. 2015. "The Business Impact of an Integrated Continuous Biomanufacturing Platform for Recombinant Protein Production." *Journal of Biotechnology* 213: 3–12.
- Wang, Ping, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. 2010. "Robust Growth of *Escherichia Coli*." *Current Biology* 20 (12): 1099–1103.
- Warikoo, Veena, Rahul Godawat, Kevin Brower, Sujit Jain, Daniel Cummings, Elizabeth Simons, Timothy Johnson, et al. 2012. "Integrated Continuous Production of Recombinant Therapeutic Proteins." *Biotechnology and Bioengineering* 109 (12): 3018–29.
- Warner, Jonathan R. 1999. "The Economics of Ribosome Biosynthesis in Yeast." *Trends in Biochemical Sciences* 24 (11): 437–40.

- Weißbe, Andrea Y, Diego A Oyarzún, Vincent Danos, and Peter S Swain. 2015. "Mechanistic Links between Cellular Trade-Offs, Gene Expression, and Growth." *Proceedings of the National Academy of Sciences* 112 (9): E1038--E1047.
- Williams, Thomas C., Bingyin Peng, Claudia E. Vickers, and Lars K. Nielsen. 2016. "The *Saccharomyces Cerevisiae* Pheromone-Response Is a Metabolically Active Stationary Phase for Bio-Production." *Metabolic Engineering Communications* 3: 142–52. <https://doi.org/10.1016/j.meteno.2016.05.001>.
- Wintermute, Edwin H, and Pamela A Silver. 2010. "Emergent Cooperation in Microbial Metabolism." *Molecular Systems Biology* 6 (1): 407.
- Wittrup, K D, A S Robinson, R N Parekh, and K J Forrester. 1994. "Existence of an Optimum Expression Level for Secretion of Foreign Proteins in Yeast A." *Annals of the New York Academy of Sciences* 745 (1): 321–30.
- Woese, Carl R, and George E Fox. 1977. "Phylogenetic Structure of the Prokaryotic Domain: The Primary Kingdoms." *Proceedings of the National Academy of Sciences* 74 (11): 5088–90.
- Woese, Carl R, Otto Kandler, and Mark L Wheelis. 1990. "Towards a Natural System of Organisms: Proposal for the Domains Archaea, Bacteria, and Eucarya." *Proceedings of the National Academy of Sciences* 87 (12): 4576–79.
- Wu, Gang, Qiang Yan, J Andrew Jones, Yinjie J Tang, Stephen S Fong, and Mattheos A G Koffas. 2016. "Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications." *Trends in Biotechnology* 34 (8): 652–64.
- Wu, Stephen G, Lian He, Qingzhao Wang, and Yinjie J Tang. 2015. "An Ancient Chinese Wisdom for Metabolic Engineering: Yin-Yang." *Microbial Cell Factories* 14 (1): 1–9.
- Wyeth, N, and R Roseveare. 1973. "Biaxially Oriented Poly (Ethylene Terephthalate) Bottle." Google Patents.
- Xu, Sen, John Gavin, Rubin Jiang, and Hao Chen. 2017. "Bioreactor Productivity and Media Cost Comparison for Different Intensified Cell Culture Processes." *Biotechnology Progress* 33 (4): 867–78.
- Yan, Gang, Petra E Vértés, Emma K Towson, Yee Lian Chew, Denise S Walker, William R Schafer, and Albert-László Barabási. 2017. "Network Control Principles Predict Neuron Function in the *Caenorhabditis Elegans* Connectome." *Nature* 550 (7677): 519–23.
- Yang, Yang, and Frances H Arnold. 2021. "Navigating the Unnatural Reaction Space: Directed Evolution of Heme Proteins for Selective Carbene and Nitrene Transfer." *Accounts of Chemical Research* 54 (5): 1209–25.
- Yang, Yaoyu, Jennifer L Nemhauser, and Eric Klavins. 2019. "Synthetic Bistability and Differentiation in Yeast." *ACS Synthetic Biology* 8 (5): 929–36.
- Yano, Hirokazu, Katarzyna Wegrzyn, Wesley Loftie-Eaton, Jenny Johnson, Gail E Deckert, Linda M Rogers, Igor Konieczny, and Eva M Top. 2016. "Evolved Plasmid-Host Interactions Reduce Plasmid Interference Cost." *Molecular Microbiology* 101 (5): 743–56.
- Yoo, Justin I, and Michelle A O'Malley. 2018. "Tuning Vector Stability and Integration Frequency Elevates Functional GPCR Production and Homogeneity in *Saccharomyces Cerevisiae*." *ACS Synthetic Biology* 7 (7): 1763–72.
- Yoshida, Shosuke, Kazumi Hiraga, Toshihiko Takehana, Ikuo Taniguchi, Hironao Yamaji, Yasuhito Maeda, Kiyotsuna Toyohara, Kenji Miyamoto, Yoshiharu Kimura, and Kohei Oda. 2016. "A Bacterium That Degrades and Assimilates Poly (Ethylene Terephthalate)." *Science* 351 (6278): 1196–99.
- You, Lingchong, Robert Sidney Cox, Ron Weiss, and Frances H Arnold. 2004. "Programmed Population Control by Cell-cell Communication and Regulated Killing." *Nature* 428 (6985): 868–71.
- Youk, Hyun, and Wendell A. Lim. 2014. "Secreting and Sensing the Same Molecule Allows Cells to Achieve Versatile Social Behaviors." *Science* 343 (6171). <https://doi.org/10.1126/science.1242782>.
- Yu, Tao, Yongjin J Zhou, Mingtao Huang, Quanli Liu, Rui Pereira, Florian David, and Jens Nielsen. 2018. "Reprogramming Yeast Metabolism from Alcoholic Fermentation to Lipogenesis." *Cell* 174 (6): 1549–58.
- Zechner, Christoph, Jakob Ruess, Peter Krenn, Serge Pelet, Matthias Peter, John Lygeros, and Heinz Koepl. 2012. "Moment-Based Inference Predicts Bimodality in Transient Gene Expression." *Proceedings of the National Academy of Sciences* 109 (21): 8340–45.

- Zhang, Xiafei, and Marie A Elliot. 2019. "Unlocking the Trove of Metabolic Treasures: Activating Silent Biosynthetic Gene Clusters in Bacteria and Fungi." *Current Opinion in Microbiology* 51: 9–15.
- Zhang, Xuwang, Dongli Bao, Maoting Li, Qidong Tang, Minghuo Wu, Hao Zhou, Lifen Liu, and Yuanyuan Qu. 2021. "Bioremediation of Petroleum Hydrocarbons by Alkali-Salt-Tolerant Microbial Consortia and Their Community Profiles." *Journal of Chemical Technology & Biotechnology* 96 (3): 809–17.
- Zhao, Evan M., Yanfei Zhang, Justin Mehl, Helen Park, Makoto A. Lalwani, Jared E. Toettcher, and José L. Avalos. 2018. "Optogenetic Regulation of Engineered Cellular Metabolism for Microbial Chemical Production." *Nature* 555 (7698): 683–87. <https://doi.org/10.1038/nature26141>.
- Zheng, Danping, Timur Liwinski, and Eran Elinav. 2020. "Interaction between Microbiota and Immunity in Health and Disease." *Cell Research* 30 (6): 492–506.
- Zhou, Jin, Yihua Lyu, Mindy L Richlen, Donald M Anderson, and Zhonghua Cai. 2016. "Quorum Sensing Is a Language of Chemical Signals and Plays an Ecological Role in Algal-Bacterial Interactions." *Critical Reviews in Plant Sciences* 35 (2): 81–105.
- Zhou, Kang, Kangjian Qiao, Steven Edgar, and Gregory Stephanopoulos. 2015. "Distributing a Metabolic Pathway among a Microbial Consortium Enhances Production of Natural Products." *Nature Biotechnology* 33 (4): 377–83.
- Zhuang, Kai, Laurence Yang, William R Cluett, and Radhakrishnan Mahadevan. 2013. "Dynamic Strain Scanning Optimization: An Efficient Strain Design Strategy for Balanced Yield, Titer, and Productivity. DySScO Strategy for Strain Design." *BMC Biotechnology* 13 (1): 1–15.
- Zhuang, Wei-Qin, Shan Yi, Markus Bill, Vanessa L Brisson, Xueyang Feng, Yujie Men, Mark E Conrad, Yinjie J Tang, and Lisa Alvarez-Cohen. 2014. "Incomplete Wood-Ljungdahl Pathway Facilitates One-Carbon Metabolism in Organohalide-Respiring *Dehalococcoides Mccartyi*." *Proceedings of the National Academy of Sciences* 111 (17): 6419–24.
- Zwier, Karen R. 2018. "Methodology in Aristotle's Theory of Spontaneous Generation." *Journal of the History of Biology* 51 (2): 355–86.

The butterfly, a cabbage-white,
(His honest idiocy of flight)
Will never now, it is too late,
Master the art of flying straight,
Yet has- who knows so well as I ?-
A just sense of how not to fly:
He lurches here and here by guess
And God and hope and hopelessness.
Even the acrobatic swift
Has not his flying-crooked gift.

Robert Graves

(Introduced to me by Giridhar Rao, an influence.)