

UC San Diego

UC San Diego Electronic Theses and Dissertations

Title

Multiplexed microfluidics utilizing genome-scale dynamics for biosensing and fermentation monitoring

Permalink

<https://escholarship.org/uc/item/4n38g65x>

Author

Stasiowski, Elizabeth

Publication Date

2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Multiplexed microfluidics utilizing genome-scale dynamics for biosensing and fermentation monitoring

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Bioengineering

by

Elizabeth Maria Stasiowski

Committee in charge:

Professor Jeff Hasty, Chair
Professor Pedro Cabrales
Professor Nan Hao
Professor Christian Metallo
Professor Milton Saier

2019

Copyright
Elizabeth Maria Stasiowski, 2019
All rights reserved.

The dissertation of Elizabeth Maria Stasiowski is approved,
and it is acceptable in quality and form for publication on
microfilm and electronically:

Chair

University of California San Diego

2019

DEDICATION

To my loving and supporting family.

EPIGRAPH

Keep it simple, stupid.

—Kelly Johnson

TABLE OF CONTENTS

Signature Page	iii
Dedication	iv
Epigraph	iv
Table of Contents	vi
List of Figures	ix
List of Tables	xi
Acknowledgements	xii
Vita	xiv
Abstract of the Dissertation	xv
Chapter 1	Introduction	1
	1.1 Understanding biology: Top-down vs. bottom-up	1
	1.1.1 Systems biology: The top-down approach	1
	1.1.2 Synthetic biology: The bottom-up approach	2
	1.2 Microfluidics: Technology to advance the understanding of biology	2
Chapter 2	High-throughput microfluidics for continuous culturing and monitoring of <i>E. coli</i> and <i>S. cerevisiae</i>	5
	2.1 Introduction	5
	2.1.1 Current high-throughput microfluidic devices for culturing cells	6
	2.2 Microfluidic device design	8
	2.2.1 Design requirements	8
	2.2.2 Cell trap design	9
	2.2.3 Channel system design	10
	2.3 Device loading	13
	2.4 Cell viability optimization	15
	2.5 Complete loading protocol	16
	2.5.1 Wafer fabrication	16
	2.5.2 PDMS device fabrication	17
	2.5.3 Glass slide preparation	18
	2.5.4 Cell Preparation	19
	2.5.5 6144-density plate and acrylic tool preparation	20
	2.5.6 Aligning the PDMS to the acrylic tool	21
	2.5.7 Oxygen plasma exposure	22

	2.5.8	Cell preparation for <i>S. cerevisiae</i> spotting	22
	2.5.9	Loading and bonding the device	23
	2.5.10	Experimental set-up	23
	2.6	Acknowledgements	23
Chapter 3		Dynamics: A microfluidic and machine learning platform for genome-scale transcriptional dynamics and environmental biosensing	25
	3.1	Introduction	25
	3.1.1	Heavy metal pollution in drinking water	26
	3.1.2	Biosensors for heavy metal detection	27
	3.1.3	Dynamics platform overview	28
	3.1.4	Microfluidic device optimization	29
	3.1.5	Dynamics Optical Enclosure	29
	3.1.6	Screening for responsive promoters to heavy metals	30
	3.1.7	Machine learning	33
	3.1.8	Explainable artificial intelligence	36
	3.1.9	Biosensor validation	41
	3.2	Conclusion	43
	3.3	Materials and methods	45
	3.3.1	Microfluidic device cell trap optimization	45
	3.3.2	Wafer Fabrication	47
	3.3.3	Microfluidic device fabrication	48
	3.3.4	Experimental protocol	48
	3.3.5	Live-cell imaging and data	49
	3.3.6	Microfluidic device validation	52
	3.3.7	Gene ontology results for Figure 3.5	53
	3.3.8	Machine learning	53
	3.3.9	Municipal water experimental set-up	57
	3.3.10	Gold King Mine spill experimental set-up	57
	3.4	Acknowledgements	57
Chapter 4		Real-time monitoring of fermentation processes through genome-scale microfluidics	59
	4.1	Introduction	59
	4.1.1	Process monitoring in the fermentation industry	60
	4.1.2	Fermentation in <i>S. cerevisiae</i>	60
	4.2	Experimental results	61
	4.2.1	Microfluidic device cell trap optimization	61
	4.2.2	Microfluidic device validation	63
	4.2.3	Fermentation monitoring of a diauxic shift	64
	4.3	Conclusion	71
	4.4	Materials and methods	74
	4.4.1	Wafer Fabrication	74

	4.4.2	Microfluidic device fabrication	75
	4.4.3	Batch culture diauxic shift experimental methods	75
	4.4.4	Data processing	76
	4.4.5	Data analysis	77
	4.5	Acknowledgements	78
Chapter 5		Summary	79
Bibliography		81

LIST OF FIGURES

Figure 2.1:	Geometry of the mid-scale 32-strain microfluidic device.	10
Figure 2.2:	Image of the large-scale 2,176-strain device loaded with <i>E. coli</i> . The inset shows a single reservoir before and after cell spotting and revival.	11
Figure 2.3:	Channel systems tested for increasing the throughput microfluidic devices constrained to 1.125 mm spacing between traps.	12
Figure 2.4:	Cell arraying equipment and workflow	14
Figure 2.5:	Cell revival after spotting onto glass or PDMS portion of microfluidic devices.	15
Figure 3.1:	Overview of the Dynamics platform.	28
Figure 3.2:	Overview of the Dynamics microfluidic device.	30
Figure 3.3:	The Dynamics optical enclosure.	31
Figure 3.4:	Responsive strains over the duration of a Dynamics experiment.	32
Figure 3.5:	Dynamics as a screening tool for heavy metal responsive promoters in <i>E. coli</i>	33
Figure 3.6:	Fold change for top responding strains to all metals.	35
Figure 3.7:	Confusion matrices showing the frequency, recall, precision, and F1 score of the LSTM-RNN classifier in predicting six metals across all experimental data (14,332 time points).	37
Figure 3.8:	Confusion matrices showing the frequency, recall, precision, and F1 score of the XGBoost classifier in predicting six metals across all experimental data (14,332 time points).	38
Figure 3.9:	LSTM-RNN classifier applied to time series data for all six detectable metals in two different experiments. Misclassified time points are shown in red.	39
Figure 3.10:	Dynamic SHAP and feature values during metal exposures.	40
Figure 3.11:	Explainable machine learning reveals <i>E. coli</i> transcriptional dynamics contributing to metal classification.	41
Figure 3.12:	SHAP values for the top 10 promoters for Cd(II) and Fe(III) with enriched Gene Ontology (GO) terms.	42
Figure 3.13:	Dynamics and machine learning on environmental samples.	43
Figure 3.14:	Cell trap designs tested to maximize <i>E. coli</i> cell retention, homogeneous growth states, and fluorescence signal.	47
Figure 3.15:	A full Dynamics image taken on custom optics before (a) and after (b) flat-field correction.	51
Figure 3.16:	Data extraction from a Dynamics image.	52
Figure 3.17:	Normalization process for data presented in Figure 3.6. Raw data was median filtered, then divided by the mean traces of promoterless strains U139 and U66. Dark lines show the mean of all time series while shaded regions represent the standard deviation.	53
Figure 3.18:	Device validation with a Dynamics chip with 2,176 replicates of the <i>zntA</i> strain from the <i>E. coli</i> GFP-promoter library exposed to repeated cadmium inductions.	54

Figure 3.19:	Raw and processed fluorescent signals: Our feature engineering eliminates a significant amount of intra-experiment variability by rendering the raw signal into a first derivative-like feature. This variability is due to differing hardware between Dynamics devices, among other sources.	55
Figure 4.1:	Overview of the 2,176-strain microfluidic device for <i>S. cerevisiae</i> a) Design of the device with spotting regions in red, conduits in green, HD biopixels for trapping cells in purple, and media channels in blue and yellow. b) Detailed schematic of four strain banks with arrows showing direction of media flow.	62
Figure 4.2:	Hydrodynamic trapping in HD biopixels.	63
Figure 4.3:	HXT2 expression in response to glucose depletion in the presence of galactose.	64
Figure 4.4:	Batch culture OD and measured glucose and ethanol concentrations over time.	65
Figure 4.5:	Relative intensity of 3,761 strains grown on the supernatant from a batch culture undergoing a diauxic shift.	66
Figure 4.6:	Agglomerative clustering and GO enrichment of the relative intensity of 3,761 strains grown on the supernatant from a batch culture undergoing a diauxic shift.	68
Figure 4.7:	Relative intensity of high-affinity glucose transporters (a) and aldehyde dehydrogenases (b) grown on supernatant from a batch culture undergoing a diauxic shift.	70
Figure 4.8:	Data extraction areas for each strain bank.	76

LIST OF TABLES

Table 2.1:	Singer ROTOR pinning settings for each step of the microfluidic device loading process.	21
Table 3.1:	Gene ontology enrichment analysis of the clusters obtained by agglomerative clustering of Figure 3.5.	34
Table 3.2:	Concentration of metals in HM9 media made with San Juan River samples as measured by ICP-MS at the Environmental Complex Analysis Lab at UC San Diego.	44
Table 3.3:	HM9 minimal media recipe.	49
Table 3.4:	Metal induction count by concentration for fold change analysis and machine learning analysis	50
Table 4.1:	Significant GO terms from up-regulated genes in Clusters 1 and 2 from Figure 4.5 with a p-value less than 0.05.	72
Table 4.2:	Significant GO terms from down-regulated genes in Clusters 3 and 4 from Figure 4.5 with a p-value less than 0.05.	73
Table 4.3:	Significant GO terms from genes in Cluster 5 from Figure 4.5 with a p-value less than 0.05.	74

ACKNOWLEDGEMENTS

Thank you to everyone who has supported me over the past five years on my journey to a Ph.D. I want to first and foremost thank Trevor Johnson who was my lifeline, personal chef, and emotional support system through this journey. I want to thank my family for supporting me all the way from Colorado with their unconditional love. I want to thank friends both near and far for listening to my rants, going on adventures to reset my mind, and for being there for me when I needed them most. Thank you to all of the Hasty Bunch. From helping solve challenges throughout this process, to our weekly Krish and poke lunches, to entertaining all of my crazy lab outings, it has truly been a blast to go through this process with you. We have the most entertaining lab out there and I couldn't have done this without each of you. Finally, thank you to my advisor, Jeff Hasty, who convinced me to get my Ph.D. in the first place and provided me with endless opportunities to grow as a scientist, engineer, event planner, communicator, and person.

-Lizzy Stasiowski, Chancellor of Games, Coordinator of Conferences, Comish of Krish, Mother of Birds

Chapter 2 contains material being prepared for submission as A custom high-throughput microfluidic and optical platform for continuous culturing and measurement of *E. coli* and *S. cerevisiae*. Csicsery, Nicholas*, Stasiowski, Elizabeth*, Thouvenin, Gregoire, Graham, Garrett, O'Laughlin, Richard, Mather, William H., Ferry, Michael, Cookson, Scott, and Hasty, Jeff. (*equal contribution) The dissertation author was one of the primary authors and researchers of this material.

Chapter 3 contains material under review at the *Proceedings of the National Academy of Sciences* as Genome-scale transcriptional dynamics and environmental biosensing. Graham, Garrett*, Csicsery, Nicholas*, Stasiowski, Elizabeth*, Thouvenin, Gregoire*, Mather, William H., Ferry, Michael, Cookson, Scott, and Hasty, Jeff. (*equal contribution) The dissertation author was one of the primary authors and researchers of this material.

Chapter 4 contains material being prepared for submission with the following individuals: Stasiowski, Elizabeth*, O'Laughlin, Richard*, and Hasty, Jeff. (*equal contribution) The dissertation author was one of the primary authors and researchers of this material.

VITA

- 2014 B. S. in Biomedical Engineering *cum laude*, California Polytechnic State University, San Luis Obispo
- 2019 Ph. D. in Bioengineering, University of California, San Diego

PUBLICATIONS

Graham, Garrett*, Csicsery, Nicholas*, Stasiowski, Elizabeth*, Thouvenin, Gregoire*, Mather, William H., Ferry, Michael, Cookson, Scott, and Hasty, Jeff. "Genome-scale transcriptional dynamics and environmental biosensing", *Proceedings of the National Academy of Sciences*. (Under review) (*equal contribution)

ABSTRACT OF THE DISSERTATION

Multiplexed microfluidics utilizing genome-scale dynamics for biosensing and fermentation monitoring

by

Elizabeth Maria Stasiowski

Doctor of Philosophy in Bioengineering

University of California San Diego, 2019

Professor Jeff Hasty, Chair

The rapid growth of synthetic and systems biology has resulted in engineered microbes with impactful applications in industries such as environmental biosensing and fermentation. The coupling of these fields with microfluidic technology has enabled the imaging of microbial colonies under continuous growth conditions with precise environmental control to better understand the dynamics involved in complex molecular networks. The emergence of genome-scale microfluidic devices has bridged the gap between the multiplexing of -omics technology and the dynamics of microfluidics; however, existing high-throughput microfluidic devices are cumbersome and often unable to be applied outside of a laboratory. Towards this end, we have

developed an elegantly simple microfluidic platform capable of monitoring the temporal gene expression of 2,176 unique microbes with both research and industrial application. In Chapter 1, I introduce the role of microfluidics in synthetic and systems biology. In Chapter 2, I describe the high-throughput microfluidic platform we have engineered and the protocol for building these devices. In Chapter 3, I show the platform's application as an environmental biosensor where the dynamics of 1,807 *E. coli* GFP-promoter strains coupled with machine learning algorithms are used to detect the presence of six heavy metals in real-time in both laboratory and real-world settings. Finally in Chapter 4, I show the device's application as a fermentation process monitoring system where the dynamics of 4,156 *S. cerevisiae* strains of a GFP-fusion library show the gene expression profile of a batch culture undergoing a diauxic shift, in real-time, depicting when the batch culture enters different growth states and changes its metabolic profile.

Chapter 1

Introduction

1.1 Understanding biology: Top-down vs. bottom-up

1.1.1 Systems biology: The top-down approach

Over the past two decades, systems biology has emerged as a field to study the complexity of networks in biology and their interactions. This interdisciplinary field has combined cellular biology, computational modeling, proteomics, genomics, and many other fields to begin to understand how many intricate components interact together to form even the simplest biological systems. The emergence of whole-genome sequencing and the decrease in cost of these techniques has led to a large database of DNA with unknown function. High-throughput technologies in systems biology including RNA sequencing [1, 2], ribosome profiling [3], mass spectrometry [4], and microarray-based expression profiling [5] coupled with quantitative modeling, have begun to untangle the web of interactions hidden in the databases of genetic code, leading to discoveries in gene function, cellular behavior, drug response, and more. However, these types of approaches are limiting due to discrete temporal resolution only offering a snapshot of a cell's state. Technologies that acquire single time point data do not capture the information encoded in the dynamics of biomolecular networks, [6] which are complex, time-dependent signals [7–9].

1.1.2 Synthetic biology: The bottom-up approach

Synthetic biology has emerged in parallel to systems biology as an interdisciplinary field with many overlapping disciplines, but with an alternative angle. Instead of striving to understand a biological system as a whole, synthetic biology builds systems from the bottom-up with simpler components, aiming to recreate the complex functions seen at the systems biology level. Incorporating reporters such as fluorescent proteins into simple genetic circuitry allows scientists to study dynamic changes and interactions in gene expression in model organisms such as *E. coli* and *S. cerevisiae* through *in vivo* imaging.

Synthetic biologists and engineers took the field to the next step by building genetic circuits and applying them to real-world problems. This machinery includes, but is not limited to, bacteria engineered to deliver therapeutics to tumor sites [10], bacteria that fluoresce in the presence of heavy metals for water quality monitoring [11, 12], yeast that have been engineered to produce an antimalarial drug [13], greatly improving its commercial availability, and yeast that produce sustainable chemicals used in the fabrication of plastics [14].

1.2 Microfluidics: Technology to advance the understanding of biology

Microfluidics emerged almost forty years ago and has spread across many disciplines as a technology that manipulates and precisely controls fluids at the micro to picoliter scale. Microfluidics integrated with electronics, microscopy, and other systems has allowed many industries to increase multiplexing of processes, decrease the amount of reagents used, have precise control over reactions, and measure small changes. Common microfluidics applications include inkjet printers, genetic sequencers, point-of-care diagnostics, and multiplexed drug-screens.

Coupled with quantitative, time-lapse, fluorescence microscopy, microfluidic platforms have become increasingly common in synthetic biology, providing precise imaging and environmental control to better understand the underlying dynamic biology of microorganisms [15]. Microfluidic platforms have facilitated the evolution of synthetic biology in the Hasty Laboratory, from the development of genetic oscillators [16], to clinically relevant cancer therapies [10], setting the stage for real-world microfluidic applications in synthetic biology. Microfluidic devices have been designed to monitor cells at different scales, ranging from tracking a single yeast cell through its entire lifespan to study their death types [17], to tracking 2.5 million cells and their quorum sensing across a 5 mm device [18].

Conventional microfluidic devices consist of an elastomer polydimethylsiloxane (PDMS) monolith bonded to glass. The PDMS is cast from a photoresist patterned silicon wafer made using standard photolithography. After surface treatment, the PDMS and glass are bonded together and cured at high temperatures (80°C) for several hours. Cells are then loaded into devices. The process of loading cells post-bonding requires each unique strain to have its own loading or waste ports to avoid cross-contamination [15]. Due to space limitations and experimental complications, conventional microfluidic devices are generally low-throughput, allowing users to track the behavior of a few nodes to answer narrow biological questions [19–21].

A handful of studies have demonstrated how microfluidic parallelization and automation allows for the simultaneous tracking of hundreds to thousands of strains from available fluorescent libraries for *S. cerevisiae* [22–24] and *E. coli* [25,26]. Combining microfluidics with large-scale mutant or fusion libraries allows for high-throughput capture and probing of large-scale gene expression dynamics. These high-throughput devices also allow for multiplexed screening of strains or chemicals in a precisely controlled environment.

Current high-throughput microfluidic devices for culturing cells are two-layer devices with pneumatic connections allowing valves to control media flow, cell loading, and chemical delivery [22,27,28]. While these devices are impressively engineered for precise single-cell

tracking, they have complex experimental protocols that are cumbersome to set-up and require hardware that limits the devices' usability both in an industrial setting and in expanding their use to other academic labs.

To this end, I devoted my time in the Biodynamics Lab to developing a simple, one-layer microfluidic device to continuously monitor thousands of strains of *E. coli* or *S. cerevisiae* at a temporal resolution of three minutes. My goal was to create a device with many applications, beneficial to both research and industry, and a device simple enough that others could readily use it and benefit from this research. The following chapters describe the devices created and their initial applications as a heavy metal biosensor and fermentation monitoring system.

Chapter 2

High-throughput microfluidics for continuous culturing and monitoring of *E. coli* and *S. cerevisiae*

2.1 Introduction

Genome-scale technologies have transformed our understanding of the biomolecular signaling networks that underpin gene function, cellular behavior and drug responses. Omics-level analysis has cemented the view that biological signal processing is not the result of linear pathways, but an emergent property of complex networks whose functions and dynamics we now seek to understand [18, 29–33]. In model organisms, signaling networks are often elucidated by studying the changing patterns of gene expression in reaction to experimentally-induced environmental perturbations [29, 30, 34, 35]. However, the high-throughput, genome-wide techniques (including RNA-seq [1, 2], ribosome profiling [3], mass spectrometry [4], and microarray-based expression profiling [5, 36]) that have successfully been applied to this problem are destructive in nature and only offer snapshots of a cell's state [37]. Technologies that acquire single time point data do

not capture the information encoded in the dynamics of biomolecular networks [6], which are complex, time-dependent signals [7–9].

Over the past two decades, high-throughput multiplexed microfluidic devices coupled to time-lapse fluorescence microscopy have evolved to bridge the gap of Omics-level data and precisely controlled dynamic environments, resulting in high-quality time series data for thousands of signaling network constituents in live cells [38, 39] for a variety of model organisms and cell lines [22–26].

2.1.1 Current high-throughput microfluidic devices for culturing cells

One of the first high-throughput microfluidic devices for cellular analysis was published by Thorsen et. al [40] and consists of 256 chambers controlled by 2,056 microvalves where two separate substrates can be loaded into a compartmentalized chamber, mixed pairwise, and independently retrieved. This device was used as a microfluidic comparator to successfully screen and retrieve *E. coli* cells expressing a cytochrome c peroxidase enzyme from a heterogeneous population. Though this device can only load a maximum of two unique strains, it introduced multiplexed valve systems to microfluidics and paved the way for many valve-based devices.

The Maerkl lab at EPFL further evolved valve-based microfluidic technology to develop a 1,152-chamber microfluidic device for monitoring *S. cerevisiae* at the single cell level [22]. Cells are first arrayed onto an epoxysilane-coated coverslip using a DNA array spotter. The multi-layer PDMS device is then aligned and bonded to the coverslip. The device has four valves that are opened and closed in a specific sequence, controlling the direction of flow to prime the chip with media, prevent cross-contamination, and allow the cells to continuously grow as a monolayer in each chamber. This device is arrayed into 3, 384-chamber sections that are fluidically isolated from each other and allow a medium switch between two sources. The temporal resolution for imaging this device is twenty minutes. The Maerkl group measured gene expression over six hours of growth and seven hours of DNA damage exposure. While this device is the most

high-throughput device currently available for *S. cerevisiae* monitoring, the temporal resolution is not sufficient for measuring dynamics in all signaling networks in *S. cerevisiae*. For example, in the oxidative response network, the time scale of protein expression change is on the order of minutes [41]. Additionally, this device only monitors strains for less than a day, preventing the monitoring of longer dynamics such as fermentation and shifts in metabolism and growth phases. Finally, this platform only allows for one medium shift and does not allow for a dynamic input, despite such control being one of the most salient advantages of microfluidics [15].

The Maerkl lab additionally developed a valve-based microfluidic device with 768 chambers for monitoring *E. coli* [27]. The same spotting technique is used as their *S. cerevisiae* device, where cells are arrayed onto a coverslip using a DNA array spotter and a multi-layer PDMS device is aligned and bonded to the coverslip. While this device can culture and monitor strains for up to one week, it requires a complex sampling process where a series of valves are opened and closed in order control the flow of medium, sample, and lysis buffer every hour throughout the duration of the experiment. During this sampling process, the sample is flowed onto the device for 10 minutes, the cells are then exposed to the sample and media via diffusion for 45 minutes, and then lysis buffer is flown on to clear the channels for ten minutes. This discontinuous exposure to a fresh sample could lead to a change in the microenvironment in each biopixel over time due to nutrient metabolism and waste secretion by the cells. Additionally, this sampling step inhibits a dynamic input of the sample. This device was used to characterize strains and as a biosensor to detect varying levels of arsenic using a custom optical system.

The most complex and multiplexed microfluidic device currently published was developed by the Tay Lab and is a 1,500 chamber microfluidic device built to culture single cells, two-dimensional populations, or 3D neurospheres of stem cells [28]. Each chamber can have independent culture conditions including cell type and density, and can also have a unique set of signaling molecules, growth factors, or drugs delivered to the chamber. The device has an on-chip chemical formulator, peristaltic pump, and valve system to mix various chemicals and deliver

them to individual chambers with sub-minute temporal resolution. Contrary to the previous valve-based devices where cells are spotted prior to bonding the PDMS device to the glass slide, this device is first bonded and then cells are loaded onto the PDMS-glass device through a semi-automatic loading program through the channels and on-chip valve system. A typical week long experiment results in the tracking of 30,000 individual cells with 1,500 dynamic individual conditions. While this device is capable of 1,500 dynamic inputs of chemicals, it is not used in the study to look at many different cell types in one experiment. This is potentially due to the requirement of loading each strain individually using the valve system.

While all three of the devices described above have many strengths with respect to monitoring the gene expression of their organism of interest, all share the disadvantage that these multi-layer, valve-based microfluidic devices have complex experimental protocols that are cumbersome to set-up and require hardware that limits the devices' usability in an industrial setting. These shortcomings and those listed above prevent the democratization of a potentially transformative technology and its application towards signaling pathway dynamics and associated biomedical applications [23]. To this end, we have developed a novel, simpler, high-throughput microfluidic device that can continuously culture and monitor *E. coli* or *S. cerevisiae*. This chapter describes the optimization performed to fabricate, load, and bond these devices and concludes with a full detailed protocol for their manufacturing.

2.2 Microfluidic device design

2.2.1 Design requirements

Contrary to traditional microfluidic loading methods where a single strain is loaded onto a device, it was necessary to develop a method to automatically load strains onto a device when loading more than about 15 strains. To make this device more available to other academic labs and more compatible for industry, it was necessary to keep the device as simple as possible.

Contrary to existing high-throughput devices, the device needed to be a single layer with minimal fluidic connections to the outside and contain no valves or pneumatic connections. To accomplish this, it would be necessary to develop a method to spot or array cells onto either the PDMS half or glass half of the device prior to bonding the two halves together.

There is no off-the-shelf equipment that can load tens to thousands of strains onto a microfluidic device. To meet this need, we adapted the Singer ROTOR spotting robot from its original purpose as a tool to transfer cells between liquid well-plates or agar plates using contact (agar) and surface tension (liquid), to a tool that could transfer cells from agar to a microfluidic device. The microfluidic design requirements were constrained to the capabilities of this robot. The requirements included the following: a cell trap region at least 400 μm in diameter where each initial cell spot would be placed, 1.125 mm spacing between each cell trap in both the x and the y dimension due to SBS-format limitations, and a region for exponential cells to grow and have sufficient diffusion of nutrients where the measurements would be taken.

2.2.2 Cell trap design

The design requirements mentioned above led to the 32-strain device design seen in Figure 2.1. This mid-scale device consists of a 4x8 array of strain banks each containing a cell spotting region connected to an array of 10 cell traps (Figure 2.1c). The strain banks are spaced 1.125 mm apart for compliance with standard SBS-format 6144-spacing. Each strain bank has an equal path length to the main media channel to guarantee even flow to every chamber. A single inlet and single outlet feed media onto the device, resulting in a simple experimental set-up.

A detailed strain bank can be seen in Figure 1a. Each cell spotting reservoir (a1) allows for cell growth down feeder channels (a2) into the cell traps (a3). The cell traps are connected by a linker channel (a5) to allow for the cross-seeding of traps. The media channel (a4) delivers fresh nutrients to cells in the traps and washes away excess cells.

Individual cell traps are seen in Figure 2.1b. Each cell trap is 10 μm wide in order to

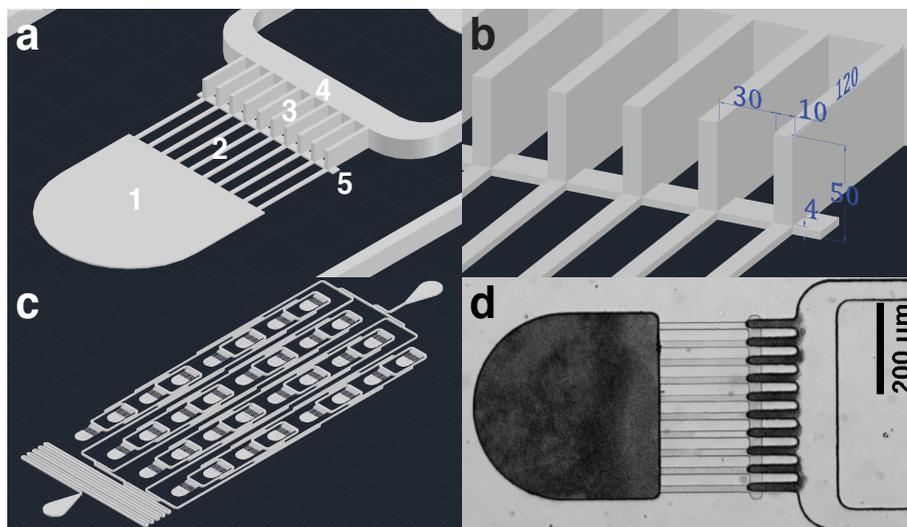


Figure 2.1: Geometry of the mid-scale 32-strain microfluidic device. a) Schematic of single strain bank, with a cell spotting region (1), feeder channels (2), cell traps (3), media channel (4), and linker channel (5). b) Detailed dimensions (μm) of the cell traps, optimized for cell growth, retention, and optical signal. c) Overview of the mid-scale device with 32-strain banks connected in parallel. d) Microscope image of *E. coli* growing exponentially in one strain bank.

decrease convective flow into the trap, improving cell retention. The traps are $30 \mu\text{m}$ s apart for fabrication purposes, and $50 \mu\text{m}$ s tall in order to maximize fluorescence signal. Figure 2.1d shows a strain bank with exponentially growing *E. coli*.

This design was then scaled-up to include 2,176 strains as seen in Figure 2.2. The strain banks are arranged into two halves of 17x64 strain banks. A single inlet splits to feed media to each half of the device, with one outlet combining the flow to remove excess cells and media waste.

2.2.3 Channel system design

Traditionally in microfluidics, a binary splitting channel system is used, where each cell trap has an equal channel path length in order to have identical flow rates and fresh nutrients. However, as the number of desired cell traps increases, this channel system requires large amounts of space (Figure 2.3a). Three additional channel systems were tested in order to decrease

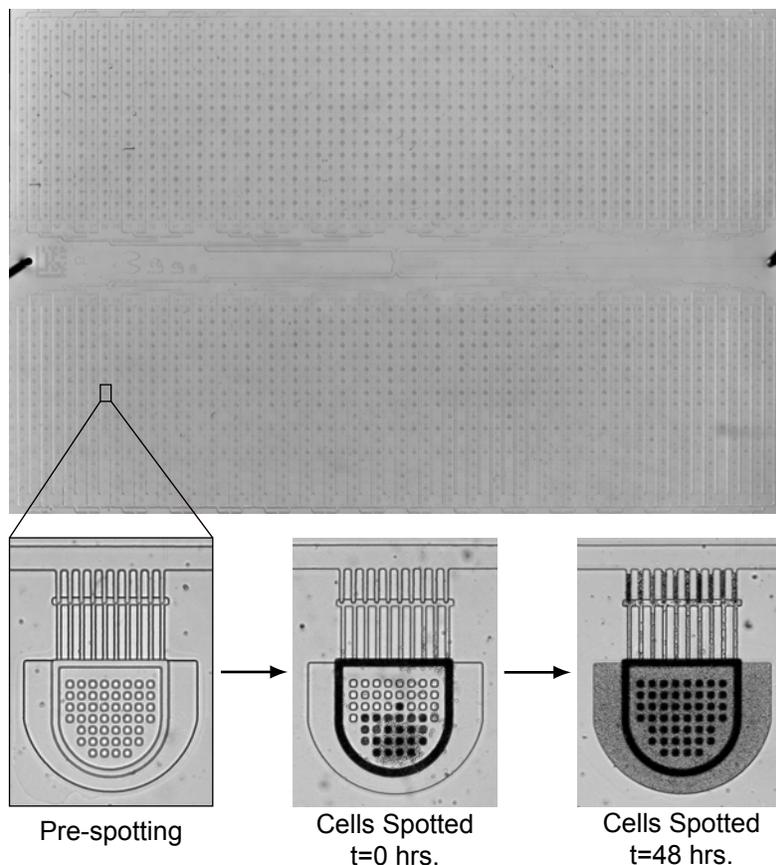


Figure 2.2: Image of the large-scale 2,176-strain device loaded with *E. coli*. The inset shows a single reservoir before and after cell spotting and revival.

the channel footprint between traps, while maintaining sufficient nutrient delivery, preventing excessive build up of cell mass, and preventing cross contamination (Figure 2.3b-d).

The shared channel system (Figure 2.3b) consists of a single inlet channel splitting into multiple channels of equal length. Each channel has many traps in series, with media from upstream traps flowing directly to downstream traps. This design allows for an even flow rate across all traps and with a large enough flow rate, still allows for sufficient nutrient delivery to each cell trap. However, we found that downstream cell traps are cross-contaminated with waste cells from upstream traps. The shared channel system is only successful if replicates of strains are loaded in a single channel, reducing the throughput of the device.

The channel-free system (Figure 2.3c) consists of an inlet channel feeding media into one

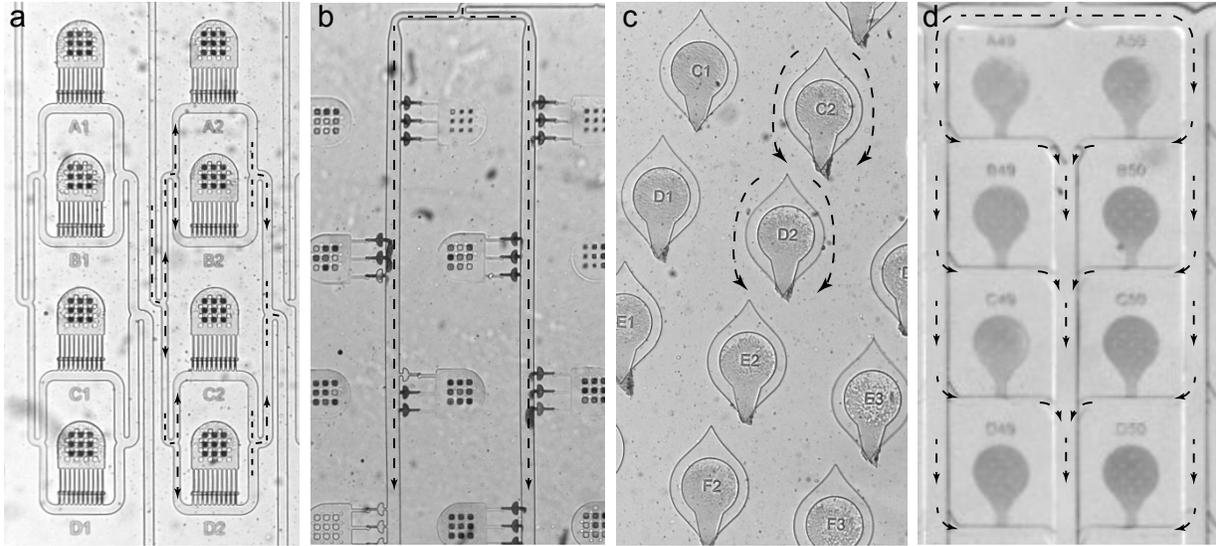


Figure 2.3: Channel systems tested for increasing the throughput microfluidic devices constrained to 1.125 mm spacing between traps. a) Binary splitting channel system traditionally used in microfluidics. b) Shared channel system. c) Channel-free system. d) Manifold channel system which is used in the 2,176-strain device. Black arrows denote flow direction.

large open area with media flowing around the top of cell traps. The traps were rotated 15° so that the laminar flow paths from the mouth of one cell trap would not directly feed another, reducing cross contamination, and maximizing nutrient diffusion. While the channel-free system has the smallest channel footprint, analysis using fluorescent microbeads showed that flow direction unexpectedly changes near some traps, resulting in cross-contamination. The unpredicted changes in direction of flow were likely due to small differences in the height of the open area due to imperfections in the silicon wafer fabrication process.

The manifold channel system (Figure 2.3d) consists of an inlet channel that splits into several vertical major media channels with a height of $230\ \mu\text{ms}$. Major channels split into minor channels in parallel that are $50\ \mu\text{ms}$ tall. Each minor channel feeds one cell trap. The minor channels then combine again into shared outlet media channels. The manifold channel system meets the design requirements best, as cells see fresh media at each trap eliminating cross contamination, and the channel footprint is small enough to fit between the required 1.125 mm spacing between cell traps. The manifold channel system results in about a 20% difference in

flow rate down a column of cell traps; however that did not affect the response of the *E. coli* strains tested during channel optimization. The manifold channel system was implemented in the large-scale device.

2.3 Device loading

To prepare strains of *E. coli* or *S. cerevisiae* for loading onto a microfluidic device, they are rearranged from their original 96-well density format. The strains are handled using the Singer ROTOR (Figure 2.4a). Bulk strain movement uses Singer 96-, 384-, 1536-, or 6144-density RePads (Figure 2.4b) whereas single colonies can be rearranged using the Singer Stinger (Figure 2.4c). The workflow for rearranging cells is shown in Figure 2.4d. First, 96-well liquid plates are spotted to 96-colony-density agar plates using the ROTOR. Next, the Stinger re-arrays strains onto four fresh agar plates at a 1536-density based on the desired arrangement on the microfluidic device. The four 1536-density agar plates are then combined to form a single 6144-density plate. 6144-density RePads are used to transfer cells from the 6144-density agar plate to the microfluidic device(s).

Initial device loading testing involved spotting cells from agar plates onto glass slides using the mid-scale device. PDMS and glass slides were cleaned and then exposed to oxygen plasma. After plasma exposure, glass slides were placed on a 3D printed insert compatible with the Singer ROTOR. Cells were spotted from the 6144-density agar plate onto the glass slides. Finally, PDMS was aligned to the cell array on the glass slides using a photolithography mask aligner, pressed together, and bonded at 37°C for two hours.

Spotting to glass resulted in poor bonding between the glass and PDMS. Regions surrounding spotted cells often unbonded, causing clogging and cross-contamination (Figure 2.5a). We hypothesized that the increased hydrophilicity of the glass slides after oxygen plasma treatment was causing a thin layer of liquid to spread around the cells and prevent bonding. To test this

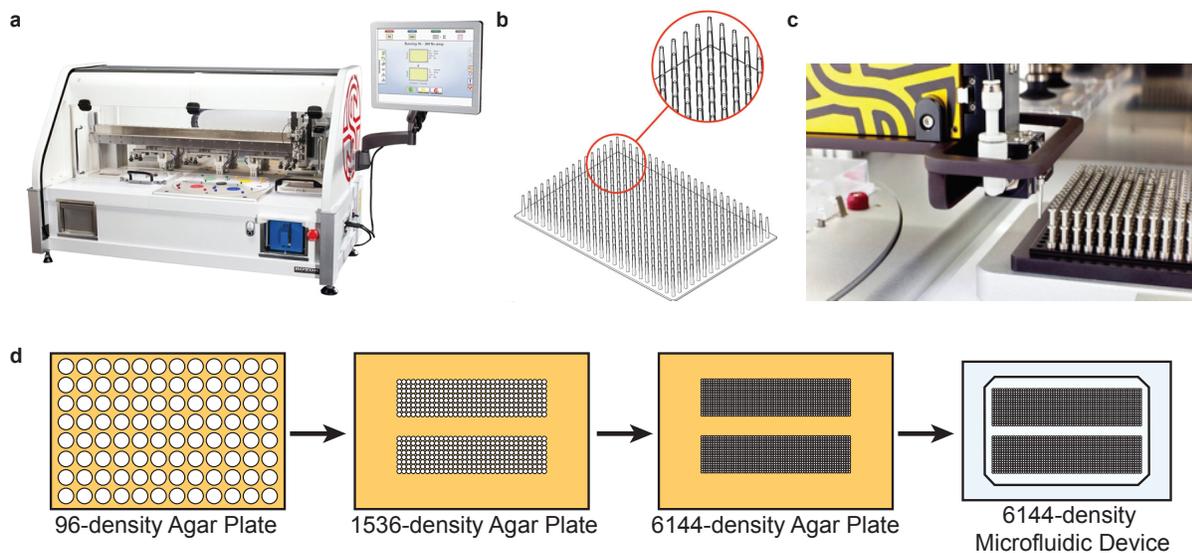


Figure 2.4: Cell arraying equipment and workflow. a) Singer ROTOR HDA. b) A 384-density long-pin RePad, used for bulk movement of strains. c) Singer Stinger, an attachment used for single colony movement. d) The workflow for loading strains onto a large-scale microfluidic device. Cells are re-arrayed from 96-density to 1536-density using the Singer Stinger. 1536-density plates are scaled up to a 6144-density and then spotted to the microfluidic device.

hypothesis, cells were grown on agar plates containing $10 \mu\text{M}$ fluorescein and spotted to both plasma treated and untreated glass slides. Fluorescence imaging of glass slides showed that significantly more liquid spreading occurred on plasma treated glass slides. Adding a five second delay between picking up cells and spotting them resulted in less liquid spreading, but also led to inconsistent spotting due to cells drying on the RePad (Figure 2.5b).

Spotting to PDMS was explored next. Cells were first transferred from a 6144-density plate to an acrylic insert compatible with the Singer ROTOR to establish alignment markers. The PDMS devices were then aligned to the cell spots on the acrylic using a photolithography mask aligner. After alignment, the aligned PDMS and glass slides were exposed to oxygen plasma. Cells were then spotted from a 6144-density agar plate onto the aligned PDMS. PDMS was pressed together with the glass slides and bonded at 37°C for two hours.

Experiments where cells were spotted to PDMS resulted in better bonding than those spotted to glass slides, but the hydrophilicity of the PDMS caused moisture to wick to the edges of

the reservoirs, resulting in some unbonding. This was solved by adding features to the reservoirs to increase surface area (Figure 2.5c). Additionally, spotting parameters on the Singer ROTOR including pin pressure, speed, offset, and overshoot were fine-tuned for spotting the optimal number of cells. Due to decreased process time and higher bonding efficiency, we chose spotting cells to PDMS as our primary loading method.

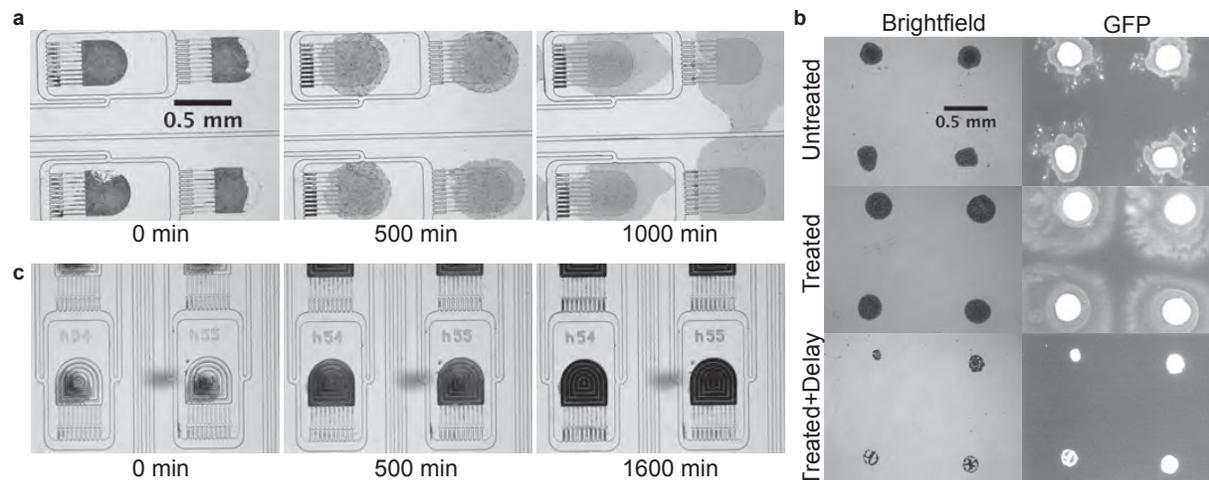


Figure 2.5: Cell revival after loading onto microfluidic devices. a) Experiments with cells spotted to glass slides resulted in significant unbonding. b) Cells grown on 10 μ M fluorescein agar plates that were either treated or untreated with oxygen plasma. Spreading of fluorescent fluid was largest for treated slides. Adding a 5-second delay prevented spreading but resulted in few cells. c) Cells spotted to PDMS patterned reservoirs resulted in well bonded devices.

2.4 Cell viability optimization

Cell health during experiments is essential; therefore, we optimized growth conditions before and after spotting cells to the microfluidic device to ensure that cells survived the spotting process and recovered during the experiment.

During early *S. cerevisiae* experiments, anywhere from 25-97% of spotted reservoirs per chip did not grow up. We hypothesized that this was due to cells drying out between being loaded onto the device and being wetted with media. We verified this using a yeast viability kit [42]. Cells growing on agar plates were spotted onto glass slides and then allowed to dry for

variable periods of time. After drying, cells were stained with green- and red-fluorescent dyes, in which the red dye selectively stained nonviable cells. Using flow cytometry, we found fresh cells growing on agar plates to be 98% viable, whereas cells spotted to glass slides and immediately stained for viability were only 6% viable.

To improve *S. cerevisiae* viability, we increased trehalose production in the cells, which has been shown to increase the tolerance of *S. cerevisiae* to the osmotic stress associated with freezing and drying [43,44]. During the cell preparation pipeline seen in Figure 3d, cells at the 1536-density were grown at 37°C under heat shock conditions. In microfluidic runs loaded from cells grown at these conditions, we observed complete cell growth.

Early *E. coli* experiments had complete revival on microfluidic devices when grown on LB+Kan agar plates. However, both *E. coli* and *S. cerevisiae* viability decreased after replicating the same set of strains for several months. Consequently, we thaw fresh overnight cells stored at -80°C every six months and repeat the preparation pipeline.

2.5 Complete loading protocol

Below is the step by step loading and bonding protocol for mid-scale and large-scale devices, for both *E. coli* and *S. cerevisiae*.

2.5.1 Wafer fabrication

Our group has previously described the steps to design and build a silicon wafer patterned with the features of the microfluidic device [15]. Briefly, the device is designed using AutoCAD or similar software, with each set of features of one desired height designed on an individual mask. Masks are printed by CAD/Art Services, Bandon, OR. A silicon wafer is built using standard soft photolithography techniques with SU-8 negative photoresist, layering each set of features with a unique height. The wafer is then coated with silane to prevent SU-8 features from tearing off the

wafer with successive PDMS pours.

2.5.2 PDMS device fabrication

Timing: 3 hours

1. In a clean weighing dish, mix 70 grams of the Dow Corning Sylgard 184 elastomer base with 7 grams of the Dow Corning Sylgard 184 curing agent. Mix vigorously with a clean stir rod for 5-10 minutes until the PDMS solution is well mixed.
2. Place the PDMS mixture in a vacuum desiccator to remove bubbles. Vent the vacuum desiccator as needed so that the bubbles do not spill over the weighing dish.
3. Place a 5" x 5" x 0.125" glass plate in the center of a two-layer 8" x 8" piece of aluminum foil. Carefully fold up the aluminum to create a dish around the glass plate. Overlap the aluminum foil over the edges of the glass to minimize PDMS leaking underneath the glass dish.
4. Place the patterned silicon wafer on the center of the glass plate and pour the PDMS mixture onto the center of the wafer, using a spatula if necessary to get the viscous mixture onto the wafer.
5. Place the wafer stack into a leveled vacuum desiccator and degas until all bubbles are removed.
6. Once the bubbles are removed, use the blunt ends of two pipette tips to center the wafer if it slid to one side, and gently push down on opposite sides of the wafer to push out PDMS that seeped under the wafer.
7. Place the wafer stack into a level oven and bake at 95°C for 1 hour.

8. Remove the wafer stack from the oven. Use a razor blade to cut off the foil from the wafer stack and cut off the excess PDMS around the wafer.
9. Gently slide a razor blade horizontally between the wafer and glass plate and then remove it. Repeat this around the circumference of the wafer until the wafer separates from the glass plate. *Note:* The razor blade must slide horizontally between the glass and the wafer. Wafers are extremely fragile and if the razor blade is angled then the wafer will break.
10. Using a razor blade, remove any excess PDMS from the bottom of the wafer. Peel the PDMS off of the feature side of the wafer in the direction of the major channels.
11. Place the PDMS on a cutting mat with the feature-side up to keep the PDMS clean. Using a razor blade, cut out each PDMS device. Punch out the inlet and outlet channels using a stainless-steel puncher.
12. Rinse each device with 70% ethanol and blow it dry with compressed air or pressurized nitrogen gas.
13. Remove debris from each device using Scotch tape, cleaning the feature side four times and the non-feature side twice. Use forceps to gently press the tape into the features to remove all debris. Leave tape on each side to keep the devices clean.

2.5.3 Glass slide preparation

Timing: 1 hour

1. Sonicate glass slides in a 2% Helmanex III solution for 30 minutes at 40°C.
2. Rinse glass slides with deionized water, rubbing them with a clean latex glove.
3. Completely dry the glass slides with pressurized nitrogen gas and ensure that no streaks are visible.

4. Store glass slides in a clean, dust-free environment until used.

2.5.4 Cell Preparation

Timing: 48-72 hours

1. Using Singer Plus Plates, pour agar plates with the appropriate cell culture media on a level surface, pouring 42 mL of media + agar into each plate. Allow plates to dry on the benchtop with the lids covered for 48 hours before parafilming and storing at 4°C.
2. Fill 96-well or 384-well liquid plates with overnight cultures of the strains of interest, or thaw glycerol stocks of strains of interest in a 96-well or 384-well density format.
3. Using the Singer ROTOR, spot the liquid plate onto an agar Plus Plate. Use the default pinning settings for both the source and target plates.
 - If spotting *E. coli*, grow the cells at 37°C overnight.
 - If spotting *S. cerevisiae*, grow the cells at 30°C for 2 days.
4. Using the Singer Stinger single colony arrayer, rearray the 96-agar or 384-density agar plate onto a set of 4, 1536-density plates that matches the array of the microfluidic device(s). These plates will later be combined into one, 6144-density plate that will be spotted to the microfluidic devices. *Note:* if spotting devices with fewer strains, multiple devices can be arrayed onto one set of 1536-plates.
5. Grow the cells overnight.
 - If spotting *E. coli*, grow the cells at 30°C overnight.
 - If spotting *S. cerevisiae*, grow the cells on the benchtop overnight.

Note: 1536-density plates can be stored in the fridge and continually used as source plates for up to 6 months.

2.5.5 6144-density plate and acrylic tool preparation

Timing: 1 hour for E. coli, 25 hours for S. cerevisiae

A: If spotting *E. coli*:

1. Using the Singer ROTOR, combine the 4x 1536-density agar source plates onto one 6144-density agar plate using the "1:4 Array" program and the pinning settings listed in Table 2.1. Repeat this for two target plates.
2. Grow one of the target plates for one hour at 37°C if the source plates have been used less than three times, or for 90 minutes at 37°C if the plates have been used more than three times.
3. Using the other 6144-density target plate, the "Replicate" program on the Singer ROTOR, and the pinning settings listed in Table 2.1, spot cells from 6144-density target plate onto the clean acrylic alignment tool. These cells will be used as alignment markers for the PDMS device.

B. If spotting *S. cerevisiae*:

1. Using the Singer ROTOR, replicate the 4x 1536-density agar source plates onto 4x 1536-density agar source plates using the "Replicate" program. Grow the plates at 37°C overnight.
2. Using the Singer ROTOR, combine the 4x 1536-density agar source plates onto one 6144-density agar plate using the "1:4 Array" program and the pinning settings listed in Table 2.1.
3. Using the "Replicate" program on the Singer ROTOR and the pinning setting listed in Table 2.1, spot cells from 6144-density target plate onto the clean acrylic alignment tool. These cells will be used as alignment markers for the PDMS device.

Table 2.1: Singer ROTOR pinning settings for each step of the microfluidic device loading process.

Source to target description	Pinning pressure (%)		Pinning speed (mm/sec)		Pinning overshoot (mm)	
	Source	Target	Source	Target	Source	Target
1536 agar to 6144 agar	58	64	19	10	2	1
6144 agar to acrylic	50	100	10	10	0.6	0.6
6144 agar to microfluidic device for <i>E. coli</i>	55	70	10	10	0.6	0.6
6144 agar to microfluidic device for <i>S. cerevisiae</i>	50	64	10	10	0.6	0.6

2.5.6 Aligning the PDMS to the acrylic tool

Timing: 30 minutes

A. If spotting multiple devices with fewer strains:

1. Using a photolithography mask aligner or similar system, set the acrylic tool on top of the mask holder with the alignment cells facing up. Bring the cells of one device into focus.
2. Remove the Scotch tape from one PDMS device, avoiding touching the feature side of the device.
3. Gently place the PDMS device on top of the alignment cells, feature-side up, such that the center of the spotting regions is centered over the cells.
4. Place tape on top of the PDMS, pressing the PDMS down to ensure adhesion of the PDMS onto the acrylic tool.
5. Repeat this for each device on the acrylic tool.

B. If spotting one large device with many strains:

1. Using a photolithography mask aligner or similar system, place the PDMS device feature-side-down on top of the wafer chuck. Remove the tape from the non-feature side of the PDMS.

2. Place the acrylic tool on top of the mask holder, with the alignment cells facing down, above the PDMS.
3. Using the micrometers on the mask aligner, align the center of the spotting regions with the center of the alignment cells.
4. Bring the PDMS and acrylic tool into contact using the wafer chuck.
5. Remove the tape from the feature-side of the PDMS and check the alignment, ensuring that the PDMS did not shift when it came into contact with the acrylic tool. If necessary, adjust the alignment.
6. Re-tape the feature side of the PDMS until ready to expose the PDMS to oxygen plasma.

2.5.7 Oxygen plasma exposure

Timing: 10 minutes

1. Expose the clean 4x3" glass slide and the PDMS acrylic stack to 30W of oxygen plasma for 30 seconds.
2. Blow any dust off the glass slide and PDMS acrylic stack with compressed nitrogen.

2.5.8 Cell preparation for *S. cerevisiae* spotting

Timing: 10 minutes

1. If spotting *S. cerevisiae*, using the Singer ROTOR, combine the heat shocked 4x 1536-density agar source plates onto one 6144-density agar plate using the "1:4 Array" program and the pinning settings listed in Table 2.1.

Note: This plate should be used immediately to spot cells onto the oxygen plasma exposed PDMS device.

2.5.9 Loading and bonding the device

Timing: 10 minutes

1. Using the Singer ROTOR and the parameters listed in Table 2.1, spot the cells from the 6144-density agar plate to the oxygen plasma exposed PDMS acrylic stack.
2. Peel the spotted PDMS off the acrylic piece and gently place it face down on the center of the oxygen-plasma-exposed glass slide.
3. Gently tap the top of the PDMS, ensuring that the device bonds to the glass.
4. Incubate the device at 37°C for at least two hours.

2.5.10 Experimental set-up

Timing: 40 minutes

1. Place the bonded PMDS device in a vacuum desiccator for at least twenty minutes.
2. Prepare an inlet syringe using methods previously described by our group [15].
3. Prepare an outlet syringe (mid-scale) using methods previously described by our group [15] or prepare an outlet tube and metal connector feeding into a waste bottle.
4. Mount the bonded PDMS device on the desired imaging platform, plugging in the inlet syringe first and then the outlet tube when a bead of liquid has formed on the outlet port.
5. Image the device at the desired temporal resolution.

2.6 Acknowledgements

This chapter is currently being prepared for submission as A custom high-throughput microfluidic and optical platform for continuous culturing and measurement of *E. coli* and *S.*

cerevisiae. Csicsery, Nicholas*, Stasiowski, Elizabeth*, Thouvenin, Gregoire, Graham, Garrett, O’Laughlin, Richard, Mather, William H., Ferry, Michael, Cookson, Scott, and Hasty, Jeff. (*equal contribution) The dissertation author was one of the primary authors and researchers of this material.

Chapter 3

Dynamics: A microfluidic and machine learning platform for genome-scale transcriptional dynamics and environmental biosensing

3.1 Introduction

At the juncture of synthetic biology and the study of genome-scale dynamics has been the development of whole-cell biosensors [11, 45]. Previously, whole-cell sensors have been developed towards applications including the detection heavy metals for environmental safety [11, 18, 46], sensing metabolites relevant to bioproduction [47–50], or operating *in vivo* towards diagnostic applications [51–53]. These sensors often harness transcriptional machinery, using either native or engineered pathways in *E. coli*, to tie specific environmental inputs to the expression of a reporter gene from a transcriptional promoter. Here, we demonstrate the utility of Dynamics, a 2,176-strain microfluidic and imaging platform, in the discovery of these transcriptional elements,

namely, in the case of heavy metals, where *E. coli* have already evolved native sense and response mechanisms [54, 55].

3.1.1 Heavy metal pollution in drinking water

Heavy metal contamination in water sources is a global threat that continues to grow due to industrialization, mining, and aging water supply infrastructures [56]. Heavy metal contamination is detrimental in both drinking and agricultural water supplies.

The contamination of US water supplies has affected hundreds of thousands of Americans in cases such as Flint, Michigan where an estimated 140,000 people were exposed to dangerously high levels of lead in their drinking water for over a year, with thousands of children testing positive for elevated blood levels for years to come [57, 58]. While Flint, MI received much of the media's attention, there are almost 3,000 small counties in the United States with a prevalence of lead poisoning twice that of Flint's [59].

Heavy metal contamination also arises from large-scale spills when dams collapse at mines, dumping toxins into nearby rivers. One recent spill was the Gold King Mine Disaster near Silverton, CO on August 5, 2015 [60]. The Gold King Mine, which was abandoned in 1923, had accumulated a large amount of contaminated groundwater inside of it, a consequence of acid mine drainage [60]. The 2015 spill occurred when the Environmental Protection Agency (EPA) went to clean the site, but accidentally caused a dam plugging the mine's entrance to fail. All of the mine's waste water, including metals such as cadmium, copper, zinc, iron, cadmium, and lead, poured into Cement Creek, and then into the Animas River, followed by the San Juan River. The thousands of members of the Navajo Nation, through which the San Juan flows, had to shut off irrigation from the river to their crops and livestock in the days after the spill, causing widespread damage in the hot summer temperatures and lasting effects on food supply [61].

Heavy metals have long posed a threat to both drinking and agricultural water supplies; however, sufficient equipment for measuring levels of heavy metals is not readily available.

Currently, water analysis requires periodic sampling that is sent to an off-site laboratory for analysis, eliminating the ability to detect early signs of pollution or when pollution has reached safe levels [62]. Not only is such sampling inefficient for large mine spills causing water to be shut off from agriculture for days longer than it needs to be, but such sampling occurs infrequently in municipal areas and rarely in people's homes. With the increasing concern of heavy metal pollution, a real-time, cheap, and portable solution is necessary.

3.1.2 Biosensors for heavy metal detection

With the growth of synthetic biology, biosensors have become a highly researched area for detecting heavy metals. Biosensors can be broadly classified into cell-free systems and whole-cell systems. Protein based biosensors include using aptamers, nucleic acids, peptides, enzymes, antibodies, or proteins as probes for sensing heavy metals. While these systems are able to detect low concentrations of analytes quickly and specifically, they can be expensive and often require immobilization, purifications, and involved analysis [11].

The evolution of genetic engineering has made it possible for creating whole-cell biosensors that can detect samples faster, cheaper, and with less operational expertise required than traditional analytical chemistry techniques. They offer similar detection limits as cell-free biosensors but are cheaper and easier to cultivate. Genetic engineers can also harvest the systems microbes have already evolved for processing heavy metals.

Combining microfluidic technologies with whole-cell biosensors has led to a massive reduction in size and cost for in-line heavy metal detection systems. Many microfluidic biosensors to date have been single-strain, single-toxin devices or devices that require a complex system with multiple external connections, constraining these devices to a lab setting [11, 27]. Instead of single-strain devices, genome-scale microfluidic instruments that can culture, monitor, and measure thousands of strains could leverage biology's natural ability to sense and respond to environmental stimuli, resulting in a non-specific sensor that could sense any substance that

perturbs a microbe. Coupling artificial intelligence (AI) algorithms with the fluorescence outputs of such a device would allow for a robust, simple, field-deployable device to sense toxins of interest in real-time that would output if that toxin is present.

3.1.3 Dynamics platform overview

Using the device developed in Chapter 2, we have developed a high-throughput microfluidic biosensor and associated software platform, Dynamics, that can simultaneously co-culture and monitor 1,807 strains of a promoter-GFP *E. coli* library [24] under time-varying environmental conditions, successfully detecting the presence of Cu(II), Zn(II), Fe(III), Pb(II), Cd(II), and Cr(VI). Initially, time series and fold-change data are used to identify and quantify responsive strains. The data is further leveraged via both deep learning classifiers, for real-time detection of inducers, and XAI algorithms, to quantify each strain’s impact on the classifier’s predictions and understand which strains are responding to each metal (Figure 3.1).

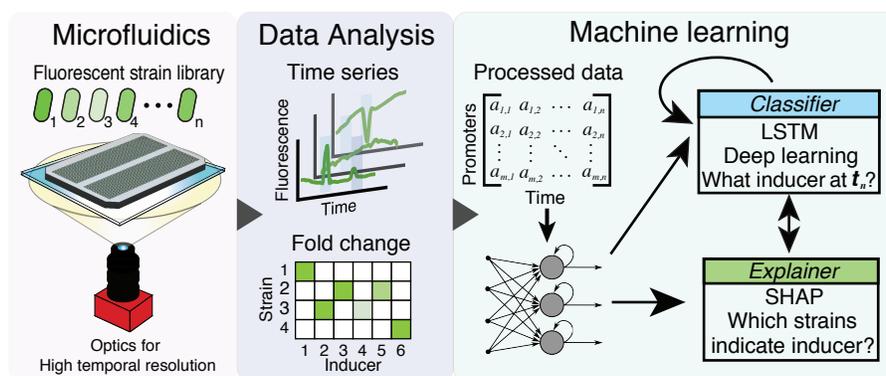


Figure 3.1: The Dynamics platform. Fluorescent strain libraries are loaded onto large-scale microfluidic devices that can be fully captured in a single image using custom optics. Parallel cultures of *E. coli* are subjected to multiple exposures of different stimuli with time series and fold-changes used to quantify responsive strains. Machine learning algorithms are trained on preprocessed data to enable real-time stimulus detection.

3.1.4 Microfluidic device optimization

Based on the initial design in section 2.2.2 the Dynamics microfluidic device was optimized for straightforward experimental setup, reliable trap filling and cell retention, and optimal fluorescence signal from each spotted microcolony (Figure 3.2a-c). The single media inlet-outlet device requires only two fluidic connections after cell spotting and chip bonding. The media inlet channel feeds a total of 2,176 4- μm -tall cell traps. Trap shape and spacing allows a 6,144 SBS-density pin pad to deposit cells into the back of the trap, where they grow towards the tapered opening interfacing with 50 μm -tall minor media-channels. These minor channels branch off of a larger 230 μm -tall major media-channel manifold system, as designed in section 2.2.2. Once spotted cells have reached confluence, inducer compounds can be pulsed in at user-specified frequencies with the dynamic response of each strain measured down to a 3 minute temporal resolution (Figure 3.2d).

3.1.5 Dynamics Optical Enclosure

The Dynamics optical enclosure is a stand-alone microscope for imaging the microfluidic device with a 70 mm x 70 mm field-of-view (Figure 3.3a). The enclosure includes an SBIG STX-16803 CCD camera (a1), a custom lens stack assembly (a2), two blue excitation LEDs (a3) and associated drivers (a7) for GFP imaging, a green LED (a5), associated driver (a7), and diffuser stack assembly (a4) for transmitted light imaging, a temperature-controlled enclosure where the microfluidic device is mounted (a4), all necessary power supplies and wiring (a8), and a Tegra computer with custom software (a7) to control the LEDs, imaging, temperature, and to sync the images onto servers via WiFi. The enclosure costs about \$15,000, compared to an off-the-shelf microscope that can cost an order of magnitude more. The optical enclosure images in both transmitted light and fluorescence channels with a 36 μm optical resolution, comparable to a lab-grade microscope at 4x magnification (Fig. 3.3c-e). With a nominal imaging frequency

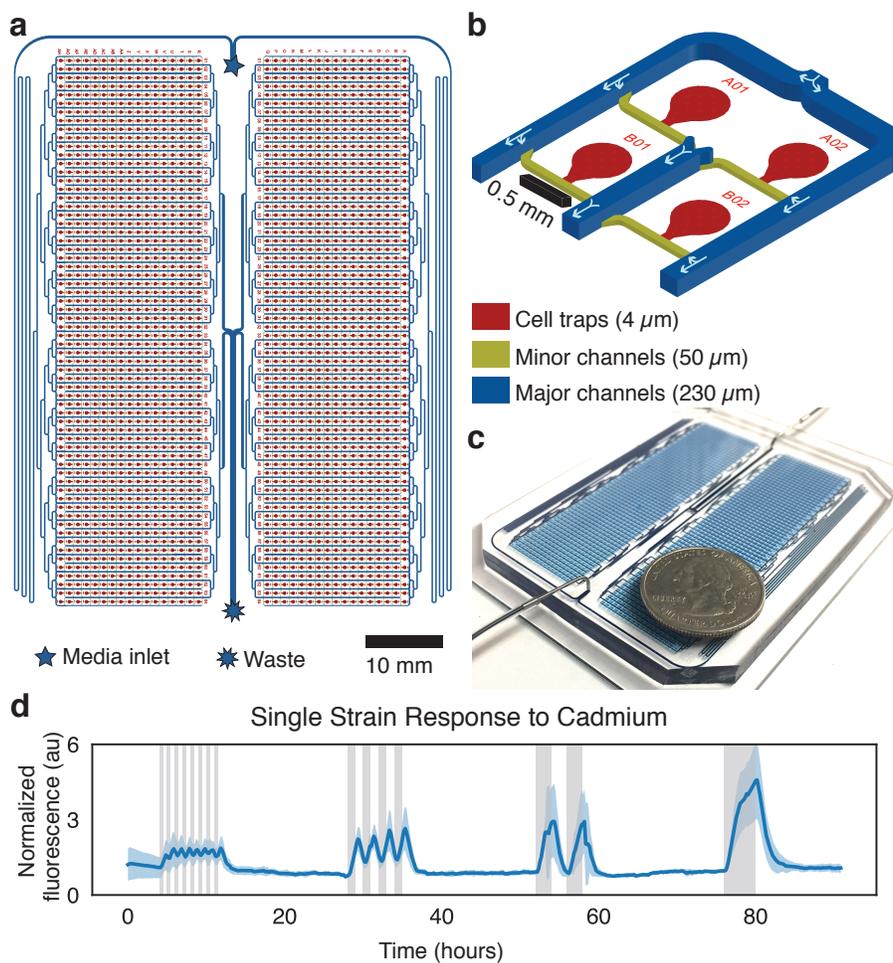


Figure 3.2: The Dynamics microfluidic device. a) Design of the Dynamics 2,176-strain microfluidic device with cell traps in red and media channels in blue and yellow. b) Detailed schematic of four strain banks with arrows showing direction of media flow. c) Mean fluorescence (solid blue) and standard deviation (shaded blue) of the *E. coli zntA* promoter driving GFP to repeated cadmium inductions (gray bars) with periods increasing from left to right (30 min., 2 hrs., 4 hrs., and 8 hrs.)

of 3 minutes, Dynamics allows us to adjust environmental conditions and measure the resulting changes in gene expression at a high temporal resolution.

3.1.6 Screening for responsive promoters to heavy metals

Using the Dynamics platform with a previously developed GFP *E. coli* promoter library [25], 1,807 unique *E. coli* promoters were screened against nine heavy metals (Cu(II), Zn(II),

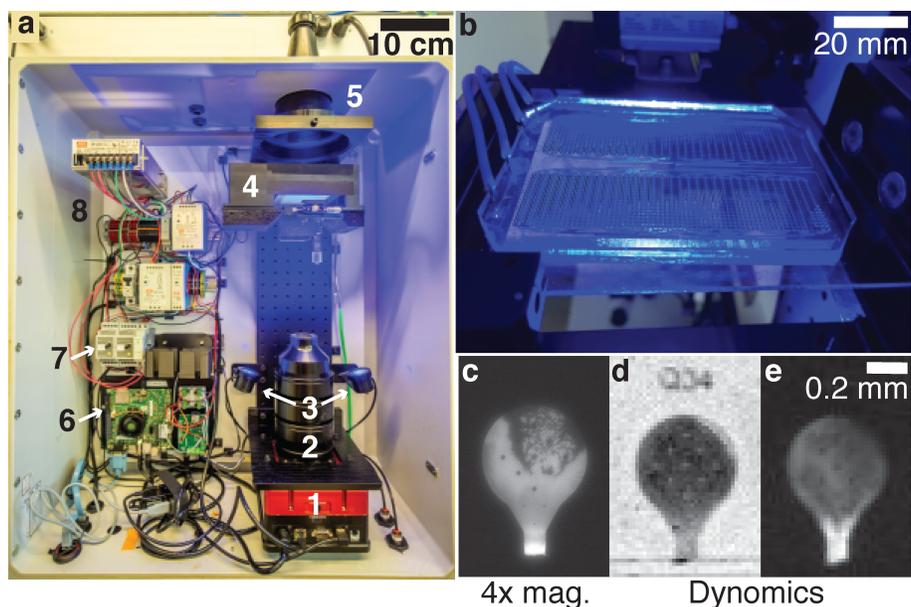


Figure 3.3: The Dynamics custom optical enclosure. a) Components include (1) a SBIG STX-16803 CCD camera, (2) a custom lens stack assembly, (3) blue excitation LEDs for GFP fluorescence imaging, (4) a temperature-controlled enclosure where the microfluidic device is mounted, (5) a green LED and associated diffusers and lenses for transmitted light imaging, (6) a Tegra computer and software for controlling imaging and transferring images to a server, (7) LED drivers, and (8) all associated power supplies and wiring. b) An image of a large-scale microfluidic device mounted in (a4). c-e) The Dynamics microfluidic device imaged on a standard research grade scope with 4x magnification, and on the Dynamics enclosure for both transmitted light (d) and fluorescence (c,e) channels. The Dynamics enclosure has 36 μm resolution.

Fe(III), Pb(II), Cd(II), Cr(VI), Hg(II), As(III), Sb(III)) at environmentally relevant concentrations (Table metal). Screening experiments lasted 7-14 days, with cells exposed to a different heavy metal every 24 hours (Figure 3.4).

Promoters responsive to each metal can be identified through a combination of clustering and fold-change analysis. A high-level view of the 1,807 promoter time traces (Figure 3.5a) and subsequent clustering (Figure 3.5b) reveal distinct classes of transcriptional responses to a single four hour zinc exposure. In (Figure 3.5b), clusters 1 and 2 include promoters that are up- and down-regulated, respectively, in the presence of zinc, but return to baseline expression levels within 15 hours of zinc removal. Clusters 3 and 4 include promoters that are up- and down-regulated, respectively, with slower dynamics. Gene ontology (GO) enrichment analysis

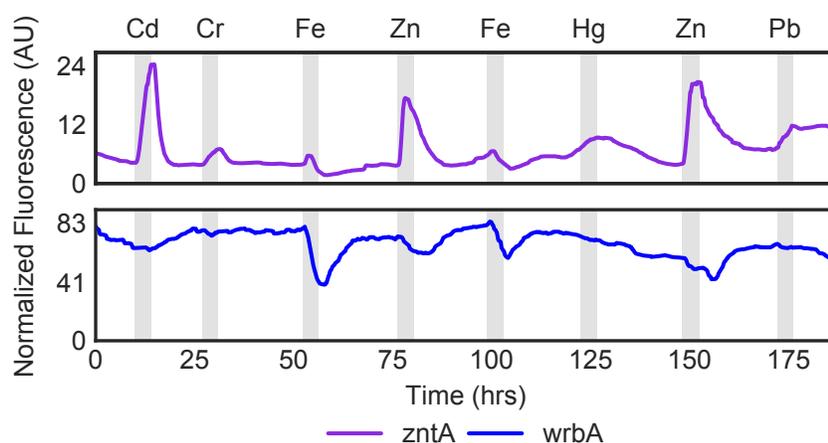


Figure 3.4: Responsive strains over the duration of a Dynamics experiment. Normalized fluorescence for two strains is plotted over the duration of one experiment, with four-hour heavy metal inductions (gray bars) occurring once daily.

suggests that from these four clusters, genes associated with cellular stress are up-regulated (cellular detoxification and antibiotic metabolic process) while genes involved in metabolism and biosynthesis are down-regulated (nitrogen metabolism and glutamine family amino acid biosynthesis). The full list of GO terms and their associated p-values are listed in Table 3.1.

Individual responsive strains for each metal were identified, based on their fold-change response (Figure 3.6a) to four-hour metal exposures, which were repeated in random order once every 24 hours. Fold-change detection highlights the promoters displaying the strongest response to each metal. Subsequent investigation of the most responsive strains (Figure 3.6b) quantitatively elucidates dynamics, such as amplitude, relaxation time, and response rate, all of which are important factors for their use in the study of gene expression regulation and continuous biosensing applications. While many of the Dynamics-identified sensing strains such as *zntA* [63] or *cueO* [64] have well documented metal interactions, others are less studied or poorly annotated, particularly members of *E. coli* ‘y-ome’ [65]. Overall, the methods of data analysis reveal each metal to have responsive promoters with a unique signature of up- and down-regulation. However, in view of the ultimate task of a biosensor, which is to distinguish the presence of a metal based

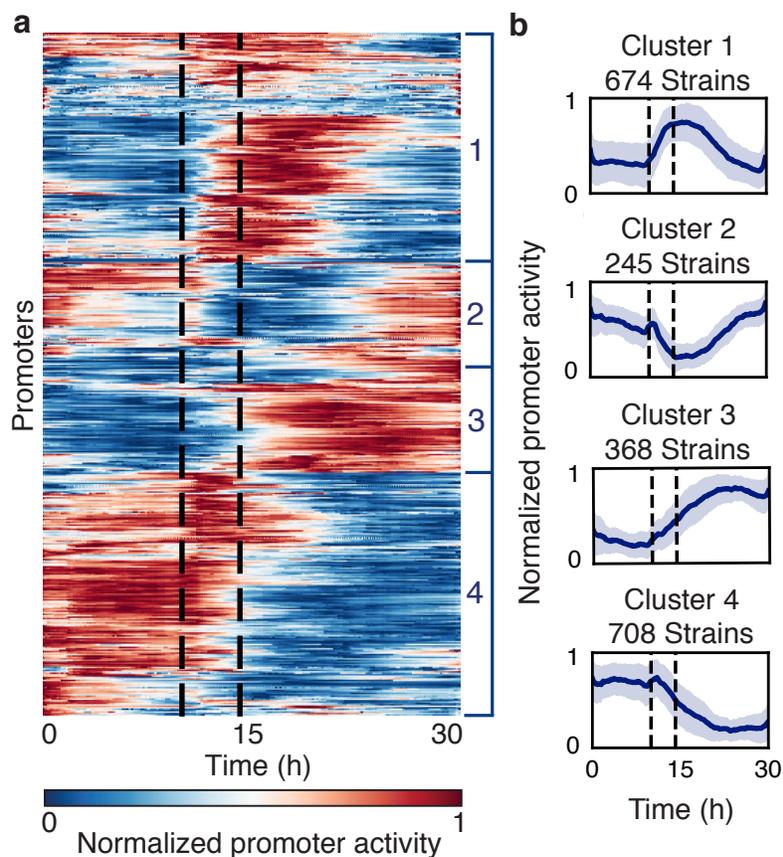


Figure 3.5: Dynamics as a screening tool for heavy metal responsive promoters in *E. coli*. a) Fluorescence response of an *E. coli* promoter library during a four hour 50 ppb Zn induction (dotted window). Each row represents the promoter activity, normalized between 0 and 1, of a single strain, with 1,995 total strains represented. Four clusters from agglomerative clustering are labeled on the right. b) Four clusters of strains calculated from agglomerative clustering from the data in panel a. The mean (dark blue line) and standard deviation (dark blue shading) of all strains in each cluster is plotted. The dotted window denotes when zinc was present.

on a real-time transcriptional data, fold-change alone is difficult due to promoter non-specificity, cross-talk, noise, and low amplitude responses.

3.1.7 Machine learning

To better discriminate between *E. coli*'s response to the heavy metals used in our screening, we trained and tested two types of machine learning models on the Dynamics data. The first model, known as extreme gradient boosted trees (XGBoost), is a popular decision tree ensemble-

Table 3.1: Gene ontology enrichment analysis of the clusters obtained by agglomerative clustering of Figure 3.5.

Cluster	GO biological process complete	Fold Enrichment	P-value
1	localization (GO:0051179)	0.74	4.11E-02
	transport (GO:0006810)	0.71	3.30E-02
	transmembrane transport (GO:0055085)	0.62	4.47E-02
2	protein autoprocessing (GO:0016540)	9.7	4.59E-02
	transcription antitermination (GO:0031564)	9.7	4.59E-02
	carbon utilization (GO:0015976)	9.7	4.59E-02
	cellular response to light stimulus (GO:0071482)	9.7	4.59E-02
	cellular response to radiation (GO:0071478)	9.7	4.59E-02
	regulation of single-species biofilm formation (GO:1900190)	9.7	4.59E-02
	cellular response to UV (GO:0034644)	9.7	4.59E-02
	bacterial transcription (GO:0001121)	9.7	1.29E-02
	one-carbon metabolic process (GO:0006730)	7.28	2.11E-02
	tetrahydrofolate metabolic process (GO:0046653)	5.54	1.43E-02
	folic acid-containing compound metabolic process (GO:0006760)	4.31	2.68E-02
	pteridine-containing compound metabolic process (GO:0042558)	3.88	3.48E-02
	positive regulation of nitrogen compound metabolic process (GO:0051173)	2.11	3.44E-02
	macromolecule metabolic process (GO:0043170)	0.66	4.19E-02
	transmembrane transport (GO:0055085)	0.33	4.02E-02
3	thiamine-containing compound biosynthetic process (GO:0042724)	5.29	2.57E-02
	cellular detoxification (GO:1990748)	5.29	2.57E-02
	thiamine biosynthetic process (GO:0009228)	5.29	2.57E-02
	thiamine-containing compound metabolic process (GO:0042723)	4.23	4.07E-02
	thiamine metabolic process (GO:0006772)	4.23	4.07E-02
	antibiotic catabolic process (GO:0017001)	3.53	2.15E-02
	detoxification (GO:0098754)	3.53	2.15E-02
	cellular response to toxic substance (GO:0097237)	3.53	2.15E-02
	reactive oxygen species metabolic process (GO:0072593)	3.31	4.21E-02
	tetrapyrrole biosynthetic process (GO:0033014)	3.17	2.99E-02
	tetrapyrrole metabolic process (GO:0033013)	3.17	2.99E-02
	response to toxic substance (GO:0009636)	2.91	5.74E-03
	antibiotic metabolic process (GO:0016999)	2.8	1.46E-02
	vitamin biosynthetic process (GO:0009110)	2.65	1.29E-02
	water-soluble vitamin biosynthetic process (GO:0042364)	2.65	1.29E-02
	water-soluble vitamin metabolic process (GO:0006767)	2.43	1.78E-02
	vitamin metabolic process (GO:0006766)	2.43	1.78E-02
	drug catabolic process (GO:0042737)	2.41	2.54E-02
drug metabolic process (GO:0017144)	2.17	1.33E-03	
positive regulation of biological process (GO:0048518)	0.38	4.72E-02	
4	arginine biosynthetic process (GO:0006526)	2.79	4.34E-02
	arginine metabolic process (GO:0006525)	2.54	4.22E-02
	glutamine family amino acid biosynthetic process (GO:0009084)	2.36	3.63E-02
	drug transport (GO:0015893)	1.77	4.98E-02
	transmembrane transport (GO:0055085)	1.48	2.65E-02
	localization (GO:0051179)	1.28	4.48E-02
drug metabolic process (GO:0017144)	0.5	2.90E-02	

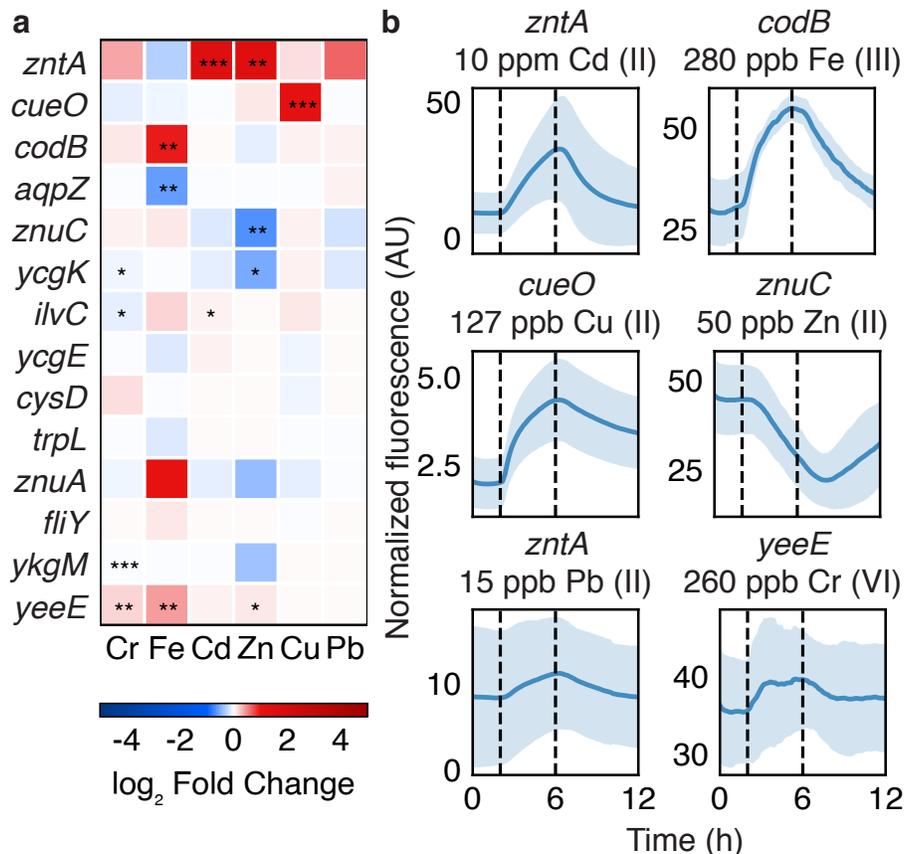


Figure 3.6: a) Fold change for top responding strains to all metals. Log₂ of the average fold change is shown for the top responding strains to each heavy metal. *, **, *** indicate p-values of 0.05, 0.01, 0.001, respectively. b) Significant single strain normalized fluorescence response (blue line) and standard deviation (blue shading) across all inductions for a given metal (dotted window).

based classifier known for its ability to learn nonlinear models [66]. The second, known as a long short-term memory recurrent neural network (LSTM-RNN), is a DNN [67] selected because of its ability to effectively utilize sample sequence history to classify time series data, a property not shared by XGBoost.

Both classification algorithms outperformed random guessing of the majority class (no toxin) on the standardized experiments' feature set, with the LSTM-RNN performing the best overall (Figures 3.7,3.8). The LSTM-RNN was able distinguish both biotic and xenobiotic metal-spiked water from pure water with a high level of reliability. We believe this is the first

instance of a multi-class classifier successfully leveraging genome-wide transcriptional dynamics to predict exposure of a biological organism to an environmental stressor.

The LSTM-RNN found iron and copper to be easily-detectable biotic metals, which is not surprising given their importance to *E. coli* cellular function [64, 68]. Cadmium was the most readily detected xenobiotic metal with the LSTM-RNN classifier, though it was sometimes confused with zinc. *E. coli* are known to use the same sensing and transport systems to capture and export excess amounts of these two metals, which possess the same number of valence electrons [63, 69]. Most classification errors occurred at the beginning and the end of the experiments' induction periods, when the LSTM-RNN occasionally had difficulty determining the exact time that each metal was added or removed from the media (Figure 3.9).

3.1.8 Explainable artificial intelligence

At present, a major obstacle in making scientific conclusions from machine learning results is the "black box problem": as an algorithm's ability to model complex phenomena grows, its decision-making processes become more and more obscured from its operators. Recently, XAI techniques in the life sciences have begun to shed light on the decision making processes of machine learning algorithms [70–72]. Further contributions from coalitional game theory have lead to the development of a mathematically-consistent method for understanding the decision-making process of any AI classifier [73, 74].

Taking advantage of these recent XAI advances, we trained a Shapley additive explanations (SHAP) XAI on both our XGBoost and LSTM learners [73, 75]. Viewing both SHAP values (impact on classifier output) and feature values (data fed to the classifier) with respect to time offers insight into how the classifier operates in real-time (Figure 3.10). The cause of misclassification is made clearer, as SHAP dynamics reveal that the predictive impact of a strain often varies within an induction window, particularly at its start and end. Furthermore, we see how some promoters, such as *zntA*, positively contribute to the detection of multiple metals, which causes

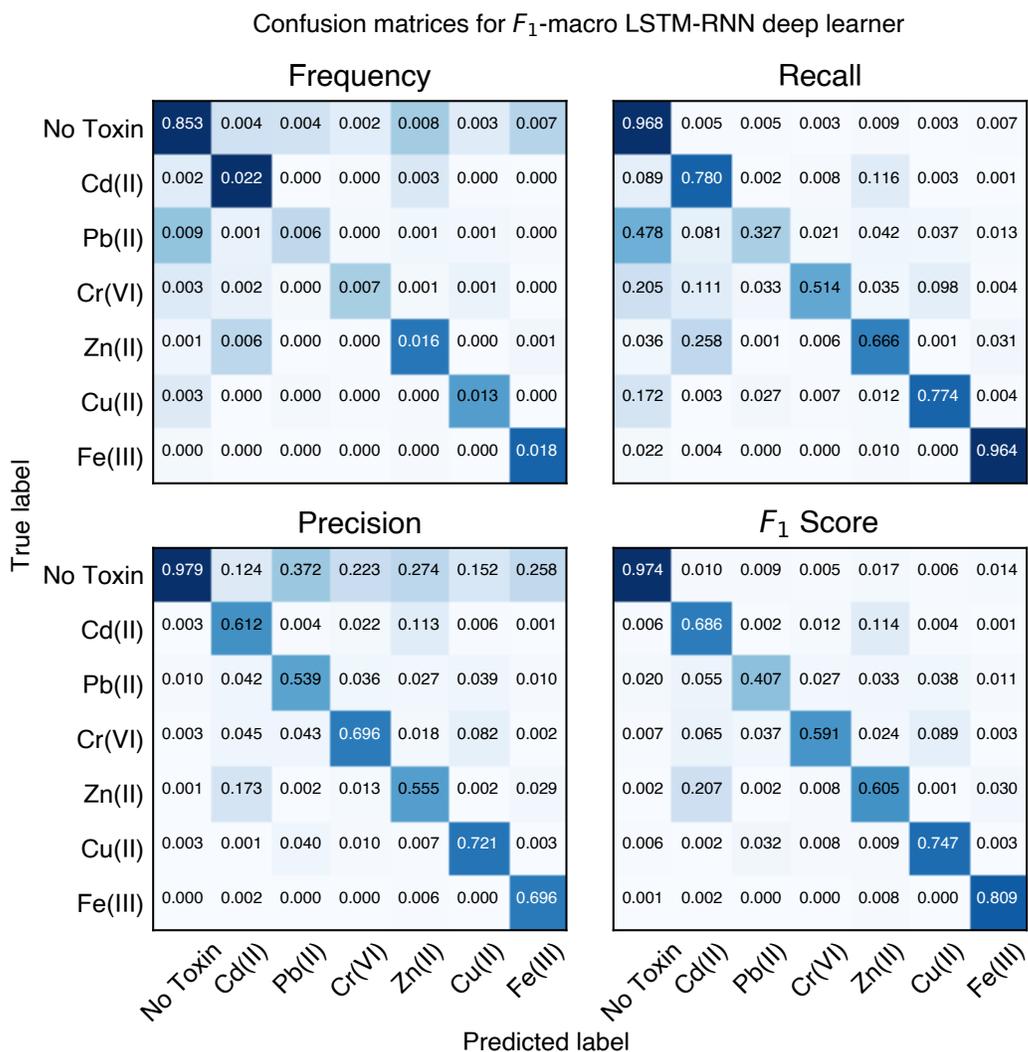


Figure 3.7: Confusion matrices showing the frequency, recall, precision, and F1 score of the LSTM-RNN classifier in predicting six metals across all experimental data (14,332 time points).

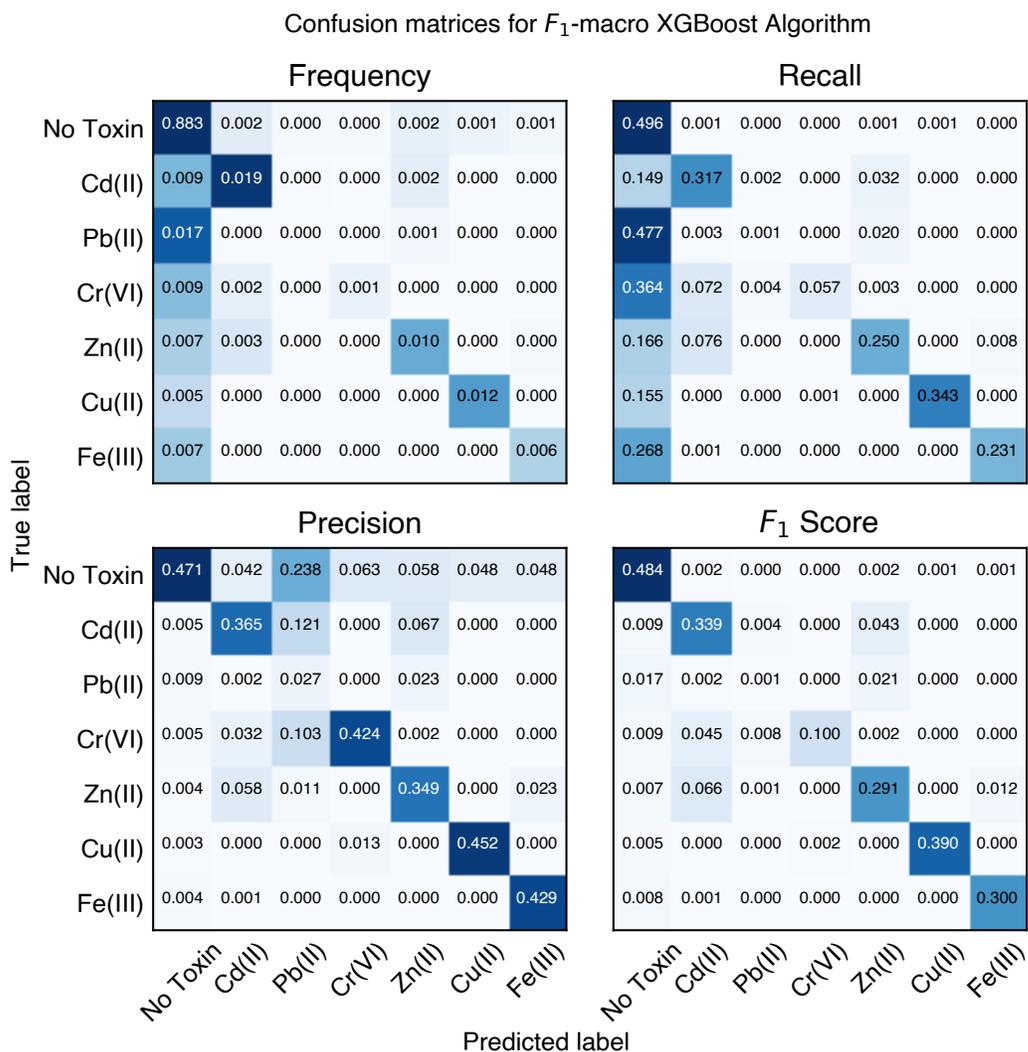


Figure 3.8: Confusion matrices showing the frequency, recall, precision, and F1 score of the XGBoost classifier in predicting six metals across all experimental data (14,332 time points).

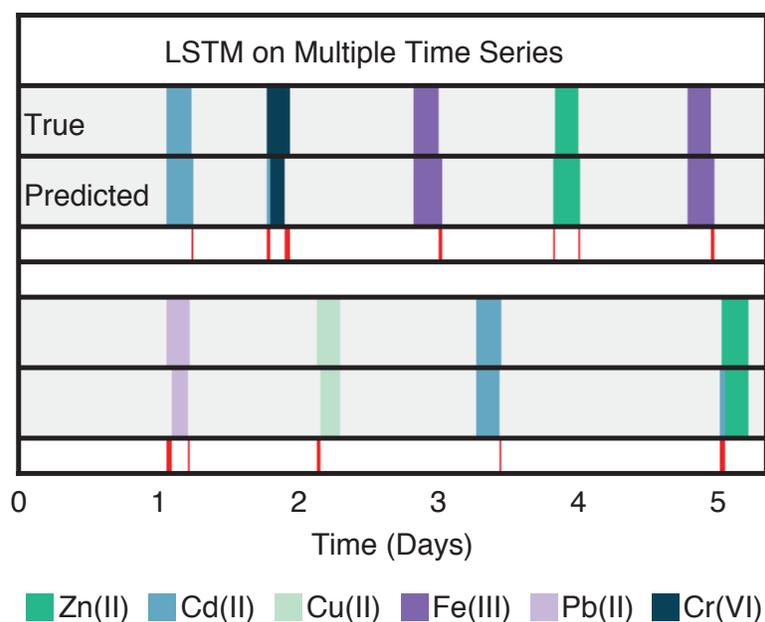


Figure 3.9: LSTM-RNN classifier applied to time series data for all six detectable metals in two different experiments. Misclassified time points are shown in red.

the classifier to rely on promoters with less-pronounced responses to distinguish the exposed metal. As a result of the combination of experimental noise and weaker responses, the SHAP values of these differentiating strains are often smaller and vary more with time, explaining the misclassification of the metals. Finally, promoters that may not have been identified as responsive using fold-change analysis because of subtle, low-amplitude, and noisy responses can be identified via XAI. While these subtle responders may not serve as stand-alone biosensor strains, they provide promising targets for future sensor engineering efforts. These insights highlight the ability of the LSTM-RNN classifier to compile the influence of many strains, prominent and subtle, to make a more often than not accurate prediction of the present metal exposure.

The SHAP-XAI highlights the similarities and differences between how the LSTM-RNN and XGBoost make decisions. Figure 3.11 shows the 15 promoters with the highest mean impact on the model and the promoterless strain U139. Both methods rely heavily on the metal-sensing promoter *zntA* for the detection and discrimination of multiple metals, especially cadmium and

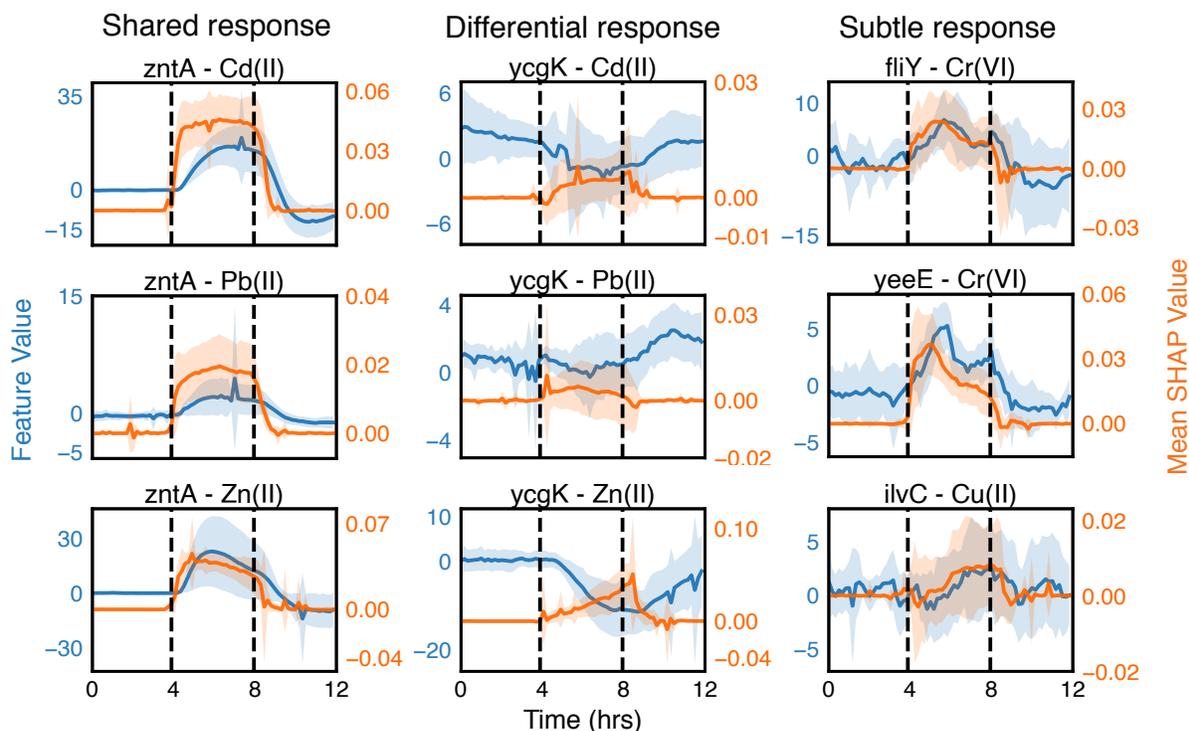


Figure 3.10: Dynamic SHAP and feature values during metal exposures. Feature (blue) and SHAP (orange) time trajectories for individual promoters during metal exposures. Solid lines show the mean value over all inductions for that metal and shaded regions around lines represents standard deviation. Dashed black lines represent the metal exposure window. While some promoters are responsive to many different metals, additional information from other promoters helps the classifier to differentiate each metal. Many promoters with noisy and subtle metal responses also contribute to classifier performance.

zinc. Beyond *zntA*, XGBoost relies heavily on single strains to detect single metals, in a manner comparable to human attention patterns. While XGBoost is not as proficient as the LSTM-RNN classifier at predicting metals, coupled with SHAP-XAI, XGBoost is able to identify top responding strains for each metal. The LSTM-RNN, on the other hand, utilizes many strains of moderate influence in a combinatorial fashion; this tendency to find a different representation than that of the human visual system has been noted in other works [76].

The ability of the XAI-coupled classifiers to identify promoters involved in metal response serves as a valuable scientific tool, suggesting potential pathways for further investigation. This is highlighted by looking at a subset of the ten most-impactful promoters individually for cadmium

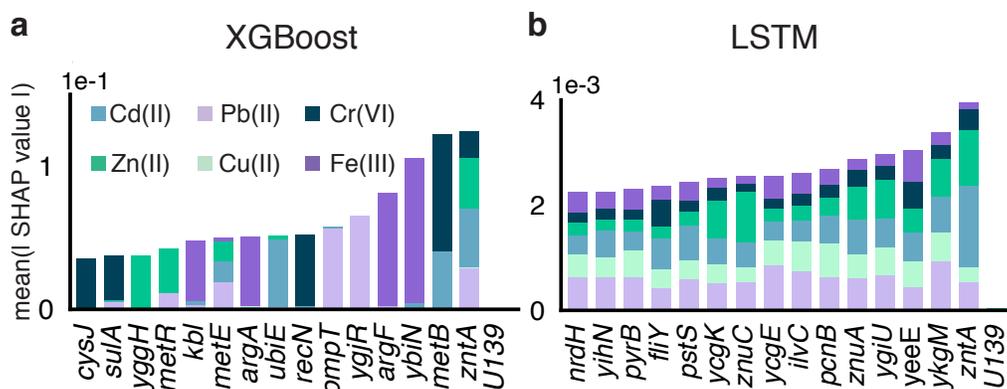


Figure 3.11: Explainable machine learning reveals *E. coli* transcriptional dynamics contributing to metal classification. The bar plots show the cumulative contribution based on the Shapley Additive Explanations (SHAP) values of 15 top promoters and a negative control (promoterless strain U139) to the prediction of each metal for both XGBoost (a) and LSTM-RNN (b) classifiers. Colored bars for each metal represent the mean absolute SHAP value over all experimental time points.

and iron inductions (Figure 3.12). These summary plots illustrate how the two classifiers make similar decisions through different methods. In the case of cadmium, *zntA* plays a significant role for both classifiers, while different sets of genes involved in ion transport or amino acid synthesis are identified for each. Most notably, the *metE* and *metB* promoters which are involved in methionine synthesis, an amino acid known to chelate cadmium, are identified by XGBoost, while the LSTM-RNN only uses the *metE* regulator, *metR*, for detection. Similarly with iron, we see XGBoost rely on members of the arginine synthesis, *argA* and *argC*, while the LSTM-RNN relies on different promoters that are involved in other metabolic or biosynthetic processes.

3.1.9 Biosensor validation

Given the severe impact of heavy metals on human health [77] and the persistence of water quality issues in the US [78] we sought to deploy the Dynamics platform as a real-time water quality biosensor. To validate this device for heavy metal sensing outside of a laboratory setting, we conducted experiments with media made from municipal water samples from San Diego, Seattle, Chicago, Miami-Dade, and New York City with added cadmium. Figure 3.13a

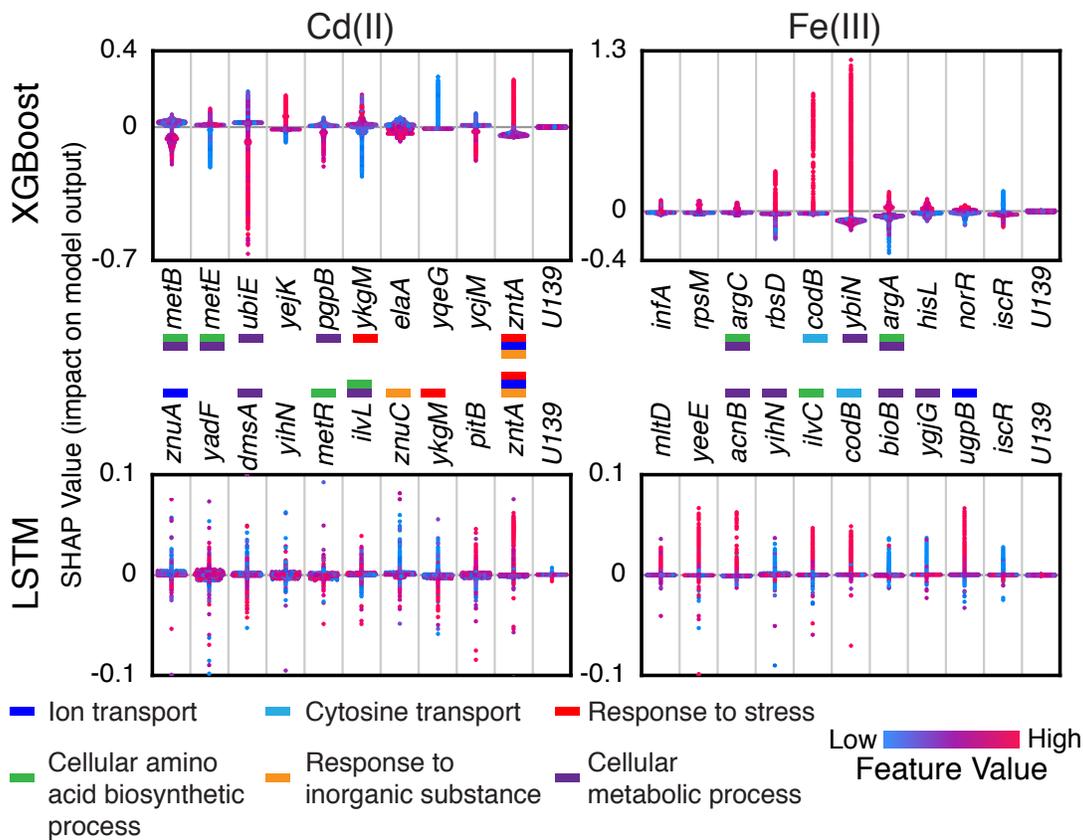


Figure 3.12: SHAP values are shown for 10 top promoters and a negative control (promoterless strain U139) for Cd(II) and Fe(III) for XGBoost and LSTM-RNN. Each point represents the feature value (normalized first derivative) at a time point. Positive SHAP values suggest that a given metal is present while negative values suggest its absence. Upregulated promoters (*zntA*, *codB*) give high SHAP values when feature values are high. Promoters are annotated with prominent gene ontology terms enriched between the two data sets.

shows the LSTM-RNN classifier predictions for cadmium exposures on each city’s water supply. While there is some misclassification of cadmium for zinc, there are few instances of incorrectly predicting the presence of a toxin versus water, even with largely different water compositions between each city.

The Dynamics device was also exposed to samples collected from the Gold King Mine Spill in August 2015. Figure 3.13b shows the predictions of the LSTM-RNN classifier on samples from the spill, collected from the San Juan River. The classifier predictions are output as multi-class, multi-label probability vectors. As the sample was introduced onto the device, the

probability of water decreased significantly while the probabilities of the other metals increased. The metal with the highest probability, iron, was also the most abundant metal in the samples as measured by ICP-MS (Table 3.2). Despite the classifier not being trained on combinations of metals, nor at the concentrations present in these samples, the ability to reliably report the presence of the most prominent metal, and, to a lesser degree, the less abundant metals, suggests the broad applicability of this platform for heavy metal detection.

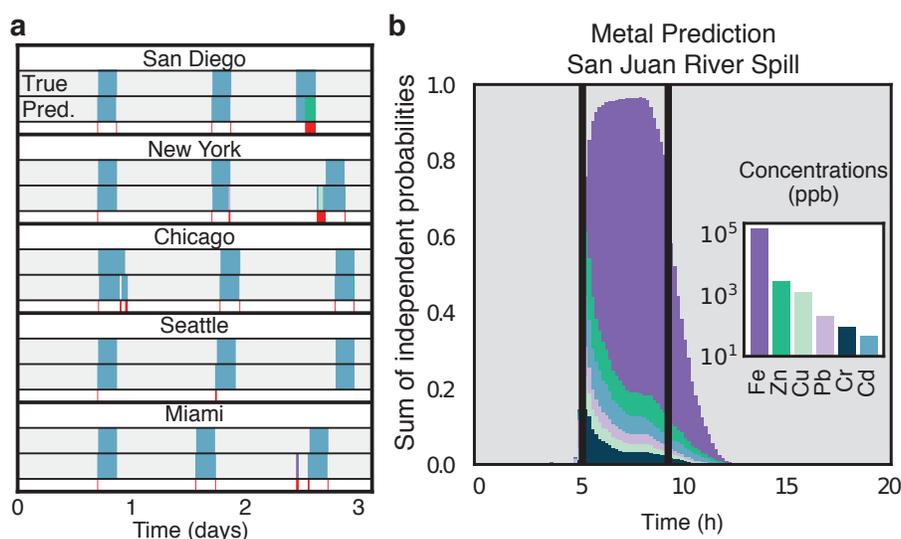


Figure 3.13: Dynamics and machine learning on environmental samples. a) LSTM-RNN classification of cadmium contamination added to five different urban water sources. The colors correspond to the metals in the inset in (b). b) Multi-class, multi-label classification of water samples from the San Juan River during the 2015 Gold King Mine waste water spill. Independent probabilities of each class are determined by the sigmoid activation function. The plot shows the sum of the classifier probabilities, averaged across triplicate sample exposures (addition and removal at vertical black lines). Inset bar chart shows the concentration of detectable metals in San Juan River samples as determined by ICP-MS. The colors of predicted toxins correspond to the metals plotted in the inset.

3.2 Conclusion

Adaptation to rapidly changing environments and external stressors is a hallmark of bacterial life and requires critical dynamic properties, many of which can be traced down to rapid

Table 3.2: Concentration of metals in HM9 media made with San Juan River samples as measured by ICP-MS at the Environmental Complex Analysis Lab at UC San Diego.

Metal	Concentration
Aluminum	100.9 ppm
Vanadium	1.263 ppm
Chromium	88.83 ppb
Manganese	828.3 ppm
Iron	151.8 ppm
Cobalt	1.289 ppm
Nickel	1.634 ppm
Copper	1.215 ppm
Zinc	2.947 ppm
Gallium	63.73 ppb
Arsenic	64.94 ppb
Silver	22.45 ppb
Cadmium	48.16 ppb
Thallium	33.52 ppb
Lead	209.3 ppb

and specific control of gene expression [?]. Transcriptomic technologies [79–81] have identified genes involved in environmental stress response and, more recently, furthered understanding of their mode of regulation. In this work we developed a high-throughput microfluidic platform to track the transcriptional dynamics of thousands of *E. coli* genes in parallel. Our Dynamics platform provides a temporal resolution and degree of experimental control that alternative experimental techniques do not possess [27]. In a high-throughput screen using Dynamics, we simultaneously exposed the 1,807 strains of the promoter-based *E. coli* GFP library to nine different heavy metal stressors. The fine-grained temporal gene-expression data it produced highlighted the unique dynamics of stimuli-specific genes previously identified as heavy-metal responsive [80], and allowed for the identification of gene clusters based on temporal response. Our platform possesses the genome-scale coverage and high sampling frequency needed to probe bacterial gene regulatory networks and screen large strain libraries for motifs difficult to discern using static gene expression data, yet ubiquitous in biology such as nonlinear interaction patterns

and feedback loops [37].

We further illustrate our platform's capacity for exploring the dynamics of transcriptional networks by applying machine learning techniques to detect stimuli-specific features of heavy metal stress response. Here we demonstrate that supervised machine learning can leverage the dynamic change in gene expression to infer the presence of environmental stressors in real-time. Algorithms derived from two formally distinct machine learning frameworks were trained on the time traces of the 1,807 promoter strains and successfully differentiated between metals belonging to biotic (Cu(II), Zn(II), Fe(III)) and xenobiotic classes (Pb(II), Cd(II), Cr(VI)). Deep-learning based methods, known for their ability to leverage time-series data, were more successful. In addition, we exploit recent progress in XAI to gain insight into the biological mechanisms that characterize transcriptomic adaptation to stress.

Finally, we demonstrate the real-world applicability of our platform by testing it on cadmium-spiked water from five municipal areas around the US with varying water composition, as well as water from the Colorado Gold King Mine Spill, possessing multiple metals at concentrations vastly different from those in the training sets. In both cases the machine learning algorithms successfully predicted the presence of the appropriate heavy metals. Approaches such as these involving the combination of high throughput microfluidics and machine learning may outperform single purpose biosensors in accuracy and robustness, and be adaptable to more varied sensing tasks.

3.3 Materials and methods

3.3.1 Microfluidic device cell trap optimization

Several trap designs were tested in order to maximize *E. coli* cell retention, homogeneous growth states, and fluorescence signal while minimizing excess cell mass that would lead to clogging and therefore a shorter device lifetime (Figure 3.14). Figure 3.14a shows the cell traps

initially designed and described in section 2.2.2 known as "gills". Gill traps had a high aspect ratio which resulted in maximized fluorescence signal, however cell retention was poor, due to convective flow into the first gill and flowing out of the tenth gill. The Gill-Shunt design (Figure 3.14b) was designed to reduce convective flow in the gill traps by providing a path of lesser resistance for flow to follow before cells filled. Although this design helped redirect flow, cell retention was still variable. The cell traps in Figure 3.14c were designed to have a dead volume region in order to better retain cells. While cell retention improved, cell growth states were variable, and resulted in variable fluorescence responses. Figure 3.14d shows the simplest trap possible, a single open side trap in which cells are directly spotted. Cells in this 50 μm tall trap accumulated too much biomass and clogged the device quickly. Figure 3.14e shows a design similar to 3.14d with an added spotting post to constrict some biomass while also providing a support pillar for the spotting process. This design also resulted in excessive biomass. Figure 3.14f shows a funneled cone design also 50 μm in height that restricted cell mass. This design was iterated for use in the channel-free design (Figure 3.14g-j). Figure 3.14g has a 4 μm spotting region and 50 μm funneled cone. Figure 3.14h shows a teardrop wall in order to prevent upstream cells from gathering on downstream trap walls. It has a 4 μm spotting region and 50 μm funneled cone similar to Figure 3.14g. The narrow funneled cone prevented clogging and large biomass, but the 50 μm height still resulted in variability in the growth state. Similarly, Figure 3.14i also has a teardrop shape and 4 μm spotting region. Instead of a 50 μm cone, this design has 50 μm gills as the cell trap. The gills reduced cell mass and clogging, but still showed variability in growth state. Figure 3.14j has a 4 μm spotting region and 4 μm cone. This design limited biofilm formation, decreased growth state variability, and had sufficient fluorescence signal; however, this design was not compatible with the manifold channel design. This trap design was adapted to work with the manifold channel design (Figure 3.14k), successfully retaining cells in a homogeneous growth state with detectable fluorescence signal.

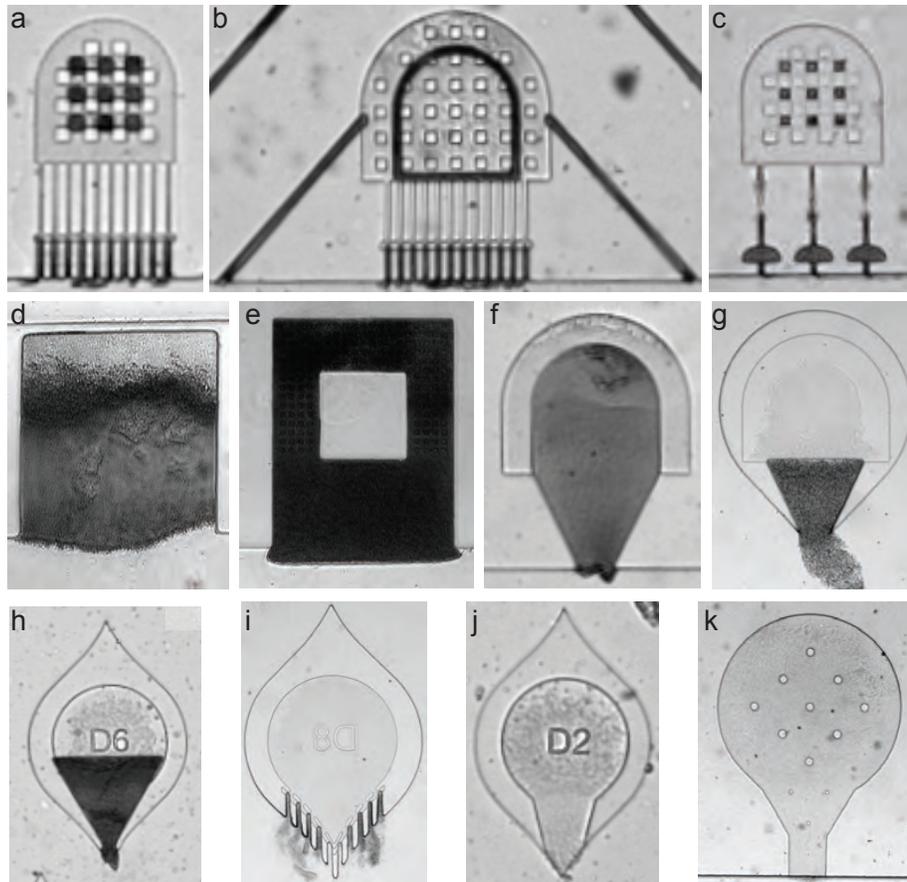


Figure 3.14: Cell trap designs tested to maximize *E. coli* cell retention, homogeneous growth states, and fluorescence signal. Gill traps with (a) and without (b) shunts inconsistently filled traps. Dead end gills (c) and side traps (d,e) lead to heterogeneous growth states. Channel-free traps (f-j) were tested with $4\ \mu\text{m}$ (f,j) and $50\ \mu\text{m}$ (g-i) openings with various opening sizes for cell retention. k) The final trap design used with the manifold channel system for Dynamics data collection.

3.3.2 Wafer Fabrication

The silicon wafer was fabricated using standard photolithography techniques previously described by our group [15]. The cell trap layer was fabricated using 2005 SU-8 photoresist with a spin speed of 5000 rpm and had a resulting height of $3.85\text{-}4.05\ \mu\text{m}$ s, the minor channel layer was fabricated using 2075 SU-8 photoresist with a spin speed of 4250 rpm and had a resulting height of $45\text{-}50\ \mu\text{m}$ s, and the major channel layer was fabricated using 2075 SU-8 photoresist with a spin speed of 1200 on top of an undeveloped major channel layer, resulting in a final height

of 200-260 μms .

3.3.3 Microfluidic device fabrication

The microfluidic devices were fabricated, loaded, and bonded with the *E. coli* promoter library [25] using the protocol in section 2.5.

3.3.4 Experimental protocol

Microfluidic experiments were performed on the Dynamics custom optical enclosure. Continuous imaging occurred every ten minutes, imaging both the transmitted light and GFP fluorescence channels. The inlet port was connected to a 140 mL syringe (Covidien Monoject Syringe) and PTFE tubing (Cole Palmer PTFE24 AWG tubing) with LB media with Kanamycin, 0.075% Tween-20, and 50 mM Methyl α -D-mannopyranoside. The waste port was connected to PTFE tubing and a 1L waste bottle. The height difference between the inlet and outlet was 39" corresponding to a flow rate of 11 mL/hr. Tween-20 and Methyl α -D-mannopyranoside were used in the media to prevent biofilming and therefore increase the longevity of microfluidic experiments. Tween-20 has been used by our group in many experiments without an adverse effect on *E. coli* [15, 18]. Methyl α -D-mannopyranoside inhibits normal surface attachment of type I pili but cannot be metabolized by *E. coli* [82]. After 16-24 hours of growth on LB media for cells to fill the traps, the media was switched to an HM9 minimal media described in Table 3.3, which was based on a previous study [83] and optimized for microfluidic *E. coli* growth with minimal traces of metals. Cells were grown on HM9 for 48 hours before inducing with heavy metals.

Heavy metal inductions occurred once a day for four hours. To induce, the HM9 in the inlet syringe was slowly pipetted out of the syringe and the HM9 + metal media was slowly pipetted in. To remove the heavy metal media after the 4 hour induction, the HM9 + metal media was pipetted

out, and the remaining dead volume was washed with 2, 5 mL HM9 wash steps. Then the HM9 media was pipetted in to fill the syringe. The order of daily metal inductions were randomized for each experiment. Inductions of each metal were performed across multiple experiments, with each experiment lasting 7-14 days. The number of inductions and concentrations for each metal is listed in Table 3.4.

Table 3.3: HM9 minimal media recipe.

Chemical	Concentration (mM)
Potassium Chloride	49.6
MOPS pH 7.2	40
β -Glycerol phosphate disodium salt pentahydrate	4
Dextrose	22.2
Ammonium chloride	18.70
Magnesium Sulfate, 7-Hydrate	0.2
Calcium Chloride Dihydrate	0.01
Iron(III) Chloride hexahydrate	0.001
Kanamycin sulfate from <i>Streptomyces kanamyceticus</i>	0.086
Tween 20	0.611
Methyl α -D-mannopyranoside	50

3.3.5 Live-cell imaging and data

Microfluidic devices were imaged using the Dynamics custom optical enclosure continuously every ten minutes in both the transmitted light and GFP fluorescence channels with a 1 second and 60 second exposure respectively.

Images were synced from the enclosure to a server via WiFi for further data processing. Custom software produced flat-field-corrected images in both channels in real-time to remove

Table 3.4: Metal induction count by concentration for fold change analysis and machine learning analysis

Metal	EPA MCL (ppb)	Concentration Tested (ppb)	Fold Change Induction Count	Machine Learning Induction Count
Arsenic	10	30	7	7
Antimony	6	1000	7	7
Cadmium	5	4.50	12	12
		8.99	0	1
Chromium	100	260	6	6
Copper	1300	127	8	8
Iron	300	279	6	6
		558	0	1
		5580	0	1
Lead	15	14.92	0	3
		29.84	6	6
Mercury	2	298.4	0	1
		2	6	6
Zinc	5000	50	5	5
		500	0	3
		1310	0	1

optical vignetting using the following equation:

$$C = m * \frac{R - D}{F' - D'} \quad (3.1)$$

where R is the raw image to be flat-field corrected (Figure 3.15a), D is the dark-current image for that device, taken at the same exposure settings as R , F' is a raw image taken by the camera with no device present, D' is the dark-current image taken at same exposure as F' , m is the mean value for all values in the array $(F' - D')$, and C is the resulting corrected image (Figure 3.15b).

Flat-field corrected images were then registered to an extraction mask to account for the device drifting from thermal expansion, with mean transmitted light and green fluorescence channels from masked bulb and background regions extracted (Figure 3.16).

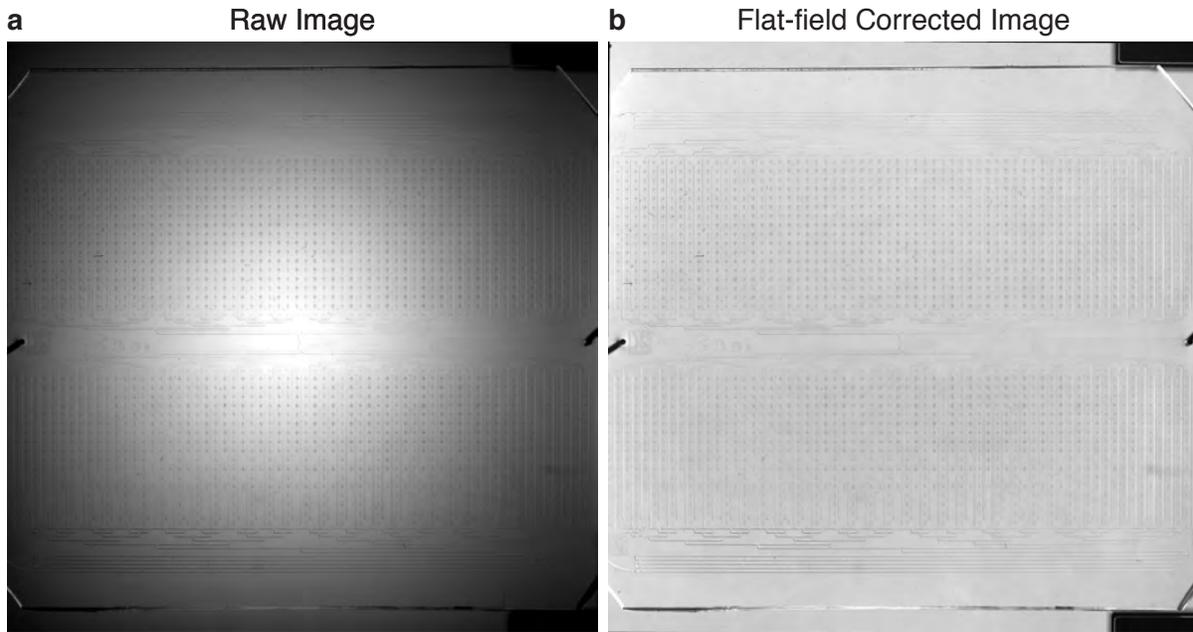


Figure 3.15: A full Dynamics image taken on custom optics before (a) and after (b) flat-field correction.

Raw data was initially processed as:

$$GFP_{Raw} = \frac{GFP_{Bulb} - GFP_{Background}}{GFP_{Background}} \quad (3.2)$$

For fold change analysis, fluorescence values were passed through a median filter (`scipy.signal.medfilt`, `kernel_size=11`) and normalized by promoterless strains (Figure 3.17). Promoterless strains U139 and U66 from the GFP promoter library [25] were spotted at various locations across each device, with 20 device positions of each, for 40 in total. Normalized fluorescence for Figure 3.6 was thus determined as:

$$GFP_{normalized} = \frac{scipy.signal.medfilt(GFP_{Raw}) - mean(scipy.signal.medfilt(GFP_{Promoterless}))}{mean(scipy.signal.medfilt(GFP_{Promoterless}))} \quad (3.3)$$

Fold change in Figure 3.6 was calculated as the quotient of the normalized fluorescence at the first and last time point of each metal exposure. P-values were determined by a dependent

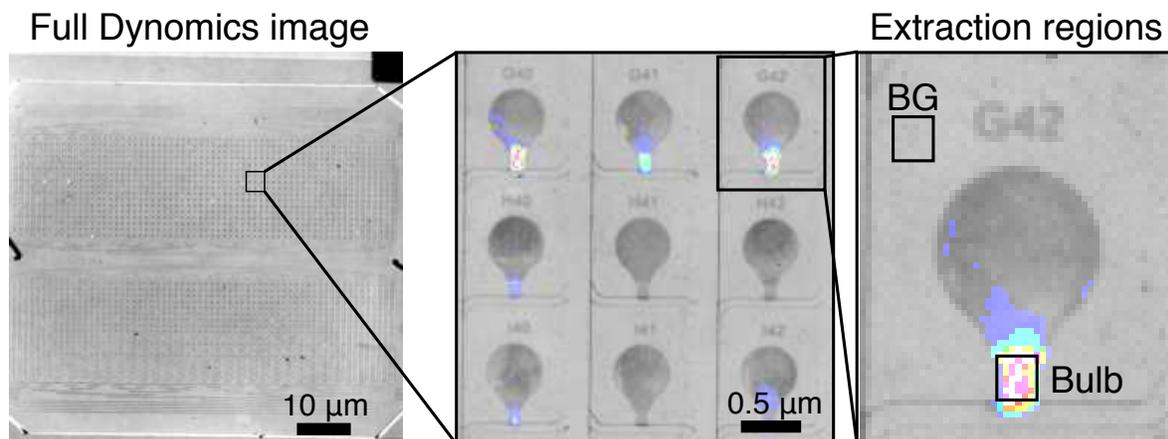


Figure 3.16: Data extraction from a Dynamics image. 2,176 device positions are simultaneously imaged in both transmitted light (gray) and green fluorescence (16 color) channels. At each time point for each device position, the mean fluorescence of the boxed bulb and background (BG) regions are extracted.

two-sided t-test on the the \log_2 fluorescent values at the start and end of each metal exposure (scipy.stats.ttest_rel).

3.3.6 Microfluidic device validation

In order to ensure that measured gene expression was independent of spatial position and to ensure responses weren't affected by previous inductions within a given experiment, a Dynamics chip was spotted with 2,176 replicates of the *zntA* promoter strain from the *E. coli* GFP-promoter library [25] and subjected to multiple replicate cadmium inductions. This strain is a reporter for the zinc and cadmium-responsive gene *zntA* and was chosen for its relevance in *E. coli* heavy metal metabolism and ability to be selectively activated via the heavy metal inducers used in the context of this work. As seen in Figure 3.18, strains across the device maintained statistically similar response parameters (amplitude of response and relaxation time) regardless of day of induction or spatial positioning on the device.

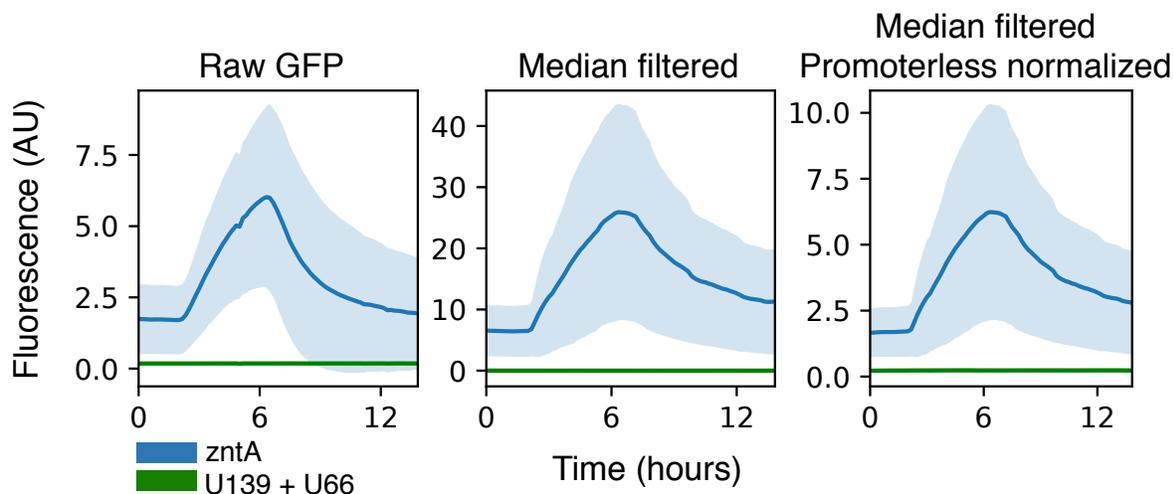


Figure 3.17: Normalization process for data presented in Figure 3.6. Raw data was median filtered, then divided by the mean traces of promoterless strains U139 and U66. Dark lines show the mean of all time series while shaded regions represent the standard deviation.

3.3.7 Gene ontology results for Figure 3.5

We performed enrichment analysis [84] on gene clusters obtained via agglomerative clustering of promoter activity, normalized between 0 and 1, of 1,995 strains responding to a 50 ppb Zn induction (Figure 3.5). Table 3.1 lists the enriched GO terms obtained for each cluster via <http://pantherdb.org/>. The statistical method used was Fisher’s exact test with no correction for multiple testing. No correction was selected as we chose to minimize the number of false negatives over the number of false positives. We note that this increases the likelihood of finding enrichment terms by chance only. The reference list for the GO enrichment analysis was the 1,807 unique genes of the GFP-promoter library [25].

3.3.8 Machine learning

We transformed our eighteen standardized experiments’ time points into a first derivative-based feature for the training and testing feature sets. All data used for machine learning results were pre-processed into first derivative-based features. Intuitively, a first derivative-based feature

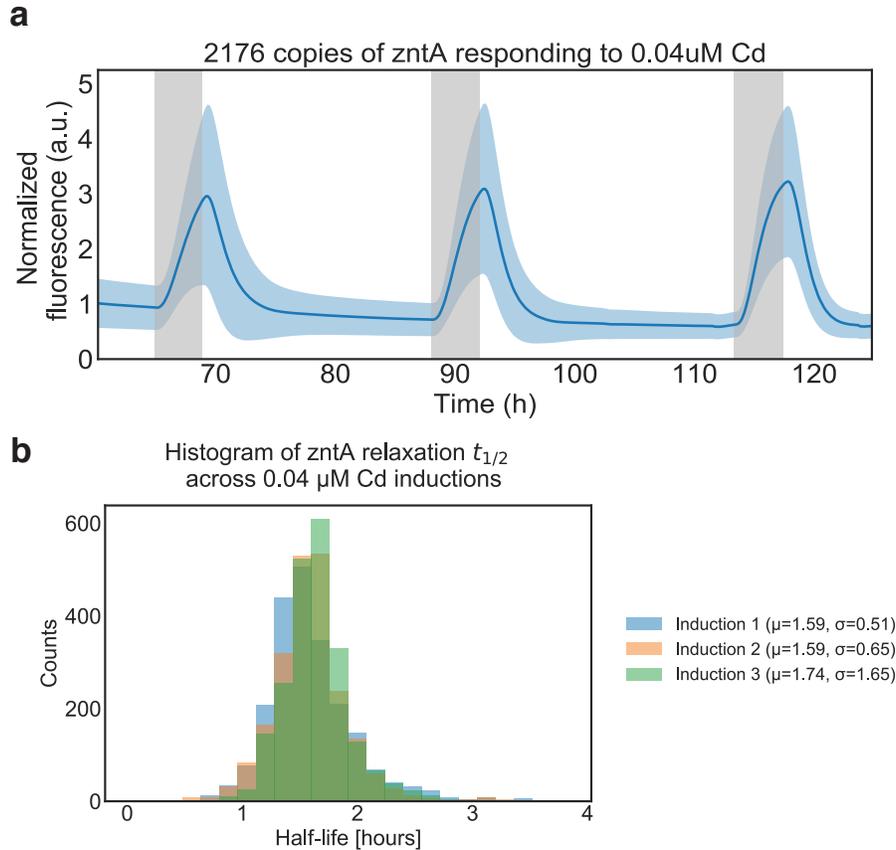


Figure 3.18: Device validation with a Dynamics chip with 2,176 replicates of the *zntA* strain from the *E. coli* GFP-promoter library [25]. a) Mean (dark blue) plus/minus one standard deviation (light blue) for the 2,176 strain responding to three consecutive four-hour 0.04 μM Cadmium inductions. b) Histograms of the decay half-life of post-induction response (i.e the time taken for a strain’s fluorescence levels to decay back down to their half-peak value) are overlaid for the three inductions plotted in a).

is an excellent candidate for any sort of machine learning model because it is what the human brain instinctively monitors when looking for changes in strain-promoter behavior. Any significant modification in the mean or variance of the first derivative of a given promoter while induced or uninduced could signify that the promoter is sensitive to that particular environmental perturbation. The engineering of a first derivative-approximation feature essentially amounts to distilling out the pure changes in the original feature’s behavior, while effectively filtering out any changes that could be due to extraneous local or global environmental influences. Figure 3.19 illustrates the results of the feature engineering process.

Raw FL signal and processed FL feature

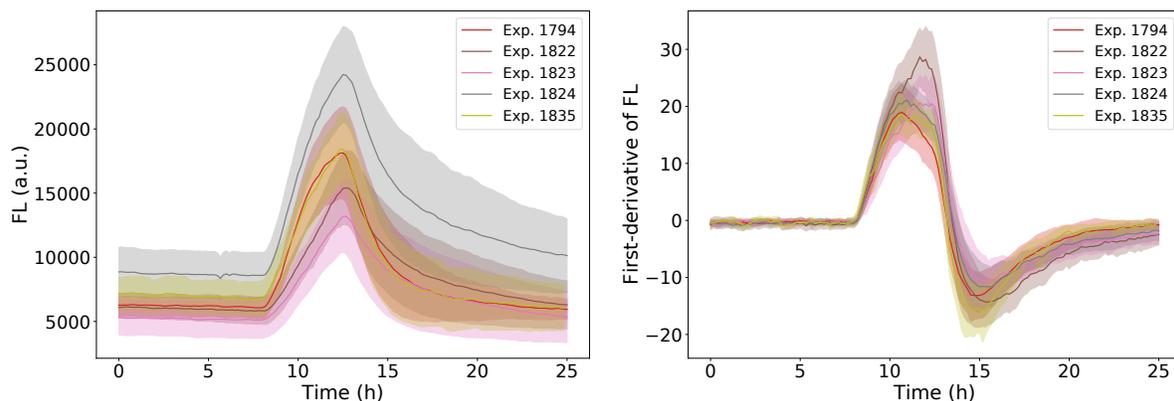


Figure 3.19: Raw and processed fluorescent signals: Our feature engineering eliminates a significant amount of intra-experiment variability by rendering the raw signal into a first derivative-like feature. This variability is due to differing hardware between Dynamics devices, among other sources.

In order to optimize the classifiers, extensive Bayesian Optimization searches were used to find optimal hyperparameter combinations [85]. Throughout our hyperparameter searches, we used leave-one-out cross-validation on a per-experiment basis and appropriate overfitting-prevention strategies to ensure that any resultant classifier would generalize to future data sets. All classifiers were evaluated using the F_1 -macro scoring metric. The F_1 -macro score, which is the per-class average of the harmonic mean of precision and recall, was especially well-suited because of our data set’s large multi-class imbalances, with water making up approximately 86% of the final feature set [86]. Finally, all generalization evaluations were performed by recording the results of using leave-one-out cross-validation with early stopping and then taking the mean prediction across the cross-validation’s output.

Additional feature preprocessing

In addition to the data preprocessing and feature engineering that were explicitly enumerate in the previous subsection, other steps were often taken when training and testing any machine learning algorithms. These additional steps each dropped some portion of the final feature set, but

only after the features had been calculated using the entire original experiments. Dropping these time points before calculating the features would have introduced potential discontinuities to any features approximating the first-derivative. All features were processed and cached in permanent memory. Only the cached features were used for any further analysis.

All experiments included transient periods over the first several days of the experiment. These transients were caused by the colonies' recoveries post-spotting shock, their growth-to-effluence within their individual traps, and their second recovering following the switch to minimal HM9 growth media. In addition, since most experiments were run until the microfluidic chips were deemed unusable due to clogging by biofilms, the final hours of most experiments did not yield high-quality water data. Since the beginnings and the ends of the experiments represented non-steady state local and global growth conditions, the features from these periods were dropped prior to analysis and machine learning. This trimming was done by removing features more than eighteen hours before the first induction and more than four hours after the final metal induction.

In addition, inductions using undetectable metals, such as arsenic, mercury, and antimony, and special inductions that introduced any non-standard inducers were dropped from the feature set post-processing.

Mercury, arsenic, and antimony exposures were also fed into the classifiers, but despite extensive feature-engineering efforts, no classifier was able to successfully detect these metals. This negative result was most likely a consequence of the absence of the *E. coli ars* and *mer* operons from the Alon promoter library [87]. These operons are known to be highly efficient at sensing and exporting arsenic and mercury, respectively, from their cells [88, 89]. In addition, upon further review, we found that the tested concentrations of both metals were over an order of magnitude lower than the known in-batch minimum inhibitory concentrations for *E. coli* [90]. To focus classification on detectable metals, features during these metal inductions were relabeled as No Toxin.

3.3.9 Municipal water experimental set-up

Water samples were obtained from the Department of Water Management at the City of Chicago, in Chicago, IL, the Alex Orr Water Treatment Plant in Miami, FL, the New York City Department of Environmental Protection and Bureau of Water Supply in Corona, NY, the Seattle Public Utilities Water Quality Lab in Seattle WA, and the Alvarado Water Treatment Plant in San Diego, CA. HM9 media for each city water experiment was prepared by diluting 5x HM9 concentrate made from milliQ water with the water obtained from each city. The microfluidic device was initially grown on LB media with Kanamycin, 0.075% Tween-20, and 50 mM Methyl α -D-mannopyranoside until traps were filled to confluence and then switched to HM9 made with city water for the remainder of the experiment. Cadmium diluted in the HM9 city water media was used to perform inductions as described in Section 3.3.4.

3.3.10 Gold King Mine spill experimental set-up

Water was collected from Mexican Hat, Utah in August, 2015 when the Gold King Mine Spill plume reached the collection point in the San Animas River. Samples were stored in 0.5% HCl Acid until tested. HM9 media was prepared by diluting 5x HM9 concentrate made from milliQ water with filtered San Animas samples. The pH was adjusted to 7.05. The metal concentrations of the HM9 San Animas samples were tested by ICP-MS at the Environmental and Complex Analysis Laboratory (ECAL) at UC San Diego. Four hour inductions were performed as described in Section 3.3.4.

3.4 Acknowledgements

This chapter contains material under review at the *Proceedings of the National Academy of Sciences* as Genome-scale transcriptional dynamics and environmental biosensing. Graham, Garrett*, Csicsery, Nicholas*, Stasiowski, Elizabeth*, Thouvenin, Gregoire*, Mather, William

H., Ferry, Michael, Cookson, Scott, and Hasty, Jeff. (*equal contribution) The dissertation author was one of the primary authors and researchers of this material.

Chapter 4

Real-time monitoring of fermentation processes through genome-scale microfluidics

4.1 Introduction

Synthetic and systems biology have been improving the global fermentation market for decades by engineering microbes to produce compounds in energy, plastics, pharmaceuticals, and food and beverage. The global fermentation chemicals market was valued at \$58.68 billion in 2018 and is expected to grow to \$85.77 billion by 2025 [91]. The ability to engineer organisms has led to unique products not natively produced in fermentation such as the antimalarial drug artemisinin [13]. One organism, *S. cerevisiae*, has been prominent in the industry due to its ability to rapidly convert sugars into ethanol.

4.1.1 Process monitoring in the fermentation industry

Although fermentation has been utilized by humans for centuries, and even with the industry growing rapidly with the emergence of synthetic and systems biology, there still remains a lack of continuous process monitoring for many substrates. Although reactors can measure oxygen levels, carbon dioxide levels, overall sugar content, and many other parameters, there is limited technology that can determine the concentration and rate of consumption of different metabolites such as glucose, sucrose, diacetyl, and acetaldehyde in real time. The gold standard of the industry is to use highly-specialized and expensive pieces of equipment such as a YSI Biochemistry Analyzer. While this equipment is accurate, it requires an operator to sample the batch culture and manually insert the sample into the equipment. Because of this grab-sample stationary measurement, parameters such as the rate of feed can only be adjusted as often as metabolite concentrations are measured. Additionally, many smaller facilities do not have the funds to purchase this equipment. The industry would greatly benefit from a technology that allows for continuous measurement of a variety of metabolites in order to more accurately adjust input parameters for the fermentation process, to understand why a fermentation fails, and to be able to react to unpredicted changes in fermentation to save time, money, and product.

4.1.2 Fermentation in *S. cerevisiae*

In *S. cerevisiae*, glucose is the preferred carbon source and the metabolism of other sugars is suppressed until glucose is depleted, at which point other carbon sources are metabolized in a sequential fashion. Typically, under aerobic conditions organisms metabolize glucose through respiration as this pathway results in a higher energy output. In *S. cerevisiae*, however, this is only true in the presence of oxygen under glucose-limited conditions. When glucose levels are normal or rich, *S. cerevisiae* undergo the Crabtree effect. In this phenomenon, regardless of oxygen content, glucose goes through glycolysis and is converted into pyruvate. However

instead of continuing to the tricarboxylic acid (TCA) cycle, the pyruvate is converted into ethanol. This continues until all of the glucose has been depleted. At this point, the cells undergo a diauxic shift, adjusting their metabolism to be able to metabolize the ethanol back into pyruvate and then through the TCA cycle and oxidative phosphorylation. While *S. cerevisiae* has been widely studied to understand the pathways involved in glucose repression and the Crabtree effect, other yeast strains are much less understood, though still used in the fermentation industry [92]. Because of this, it is ideal to measure the glucose and ethanol production and consumption rates in real time during a fermentation process.

To this end, we have developed a high-throughput 2,176-strain microfluidic device spotted with a 4,156-strain GFP-fusion library of *S. cerevisiae* [24], that coupled with fluorescence imaging can monitor the metabolism of a batch culture in real time. By growing the library strains on the microfluidic device on the same supernatant from the batch culture in real-time we are able to observe the dynamic changes in gene expression from the exponential growth, through the diauxic shift, and all the way into stationary phase. Not only can this device be used to monitor fermentation processes, but it simultaneously provides a deeper insight into the genetic pathways involved.

4.2 Experimental results

4.2.1 Microfluidic device cell trap optimization

Based on the microfluidic design described in section 2.2.2, a 2,176-strain microfluidic device was optimized for healthy growth, retention, and fluorescence signal from each spotted microcolony of *S. cerevisiae* (Figure 4.1). The device has the same cell trap spotting region and manifold channel layout as the device designed for *E. coli* in section 3.1.4; however, fluorescence is measured in downstream traps instead of at the mouth of the spotting region. These downstream traps, known as hydrodynamic biopixels (HD biopixels) consist of four 70 μm by 70 μm biopixels

that are $13\ \mu\text{m}$ tall. A $2\ \mu\text{m}$ tall conduit region at the back of the trap fluidically connects the biopixels with the minor media channel, resulting in convective flow through empty biopixels. The convective flow of media also pushes *S. cerevisiae* cells into the trap. Because of the $2\ \mu\text{m}$ tall conduits, the $4\ \mu\text{m}$ tall cells get trapped against the back wall and are able to divide and fill the traps to confluence (Figure 4.2). This concept is known as hydrodynamic trapping [93]. The HD biopixels are angled to increase cell retention in the traps.

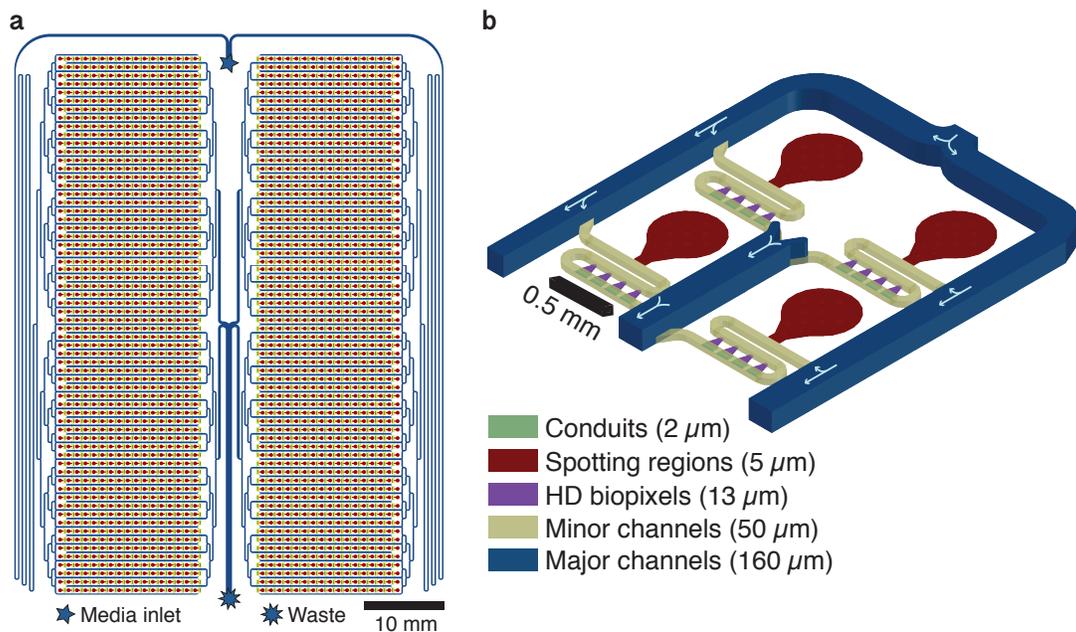


Figure 4.1: Overview of the 2,176-strain microfluidic device for *S. cerevisiae* a) Design of the device with spotting regions in red, conduits in green, HD biopixels for trapping cells in purple, and media channels in blue and yellow. b) Detailed schematic of four strain banks with arrows showing direction of media flow.

The biopixels foster healthier cells than those in the mouth of the spotting region, which are affected by nutrient-limited cells in the back of the spotting region pushing the healthy cells out of the mouth of the trap. The height of the biopixels also results in a larger fluorescence signal, imaged by the Dynamics optical enclosure described in section 3.1.5.

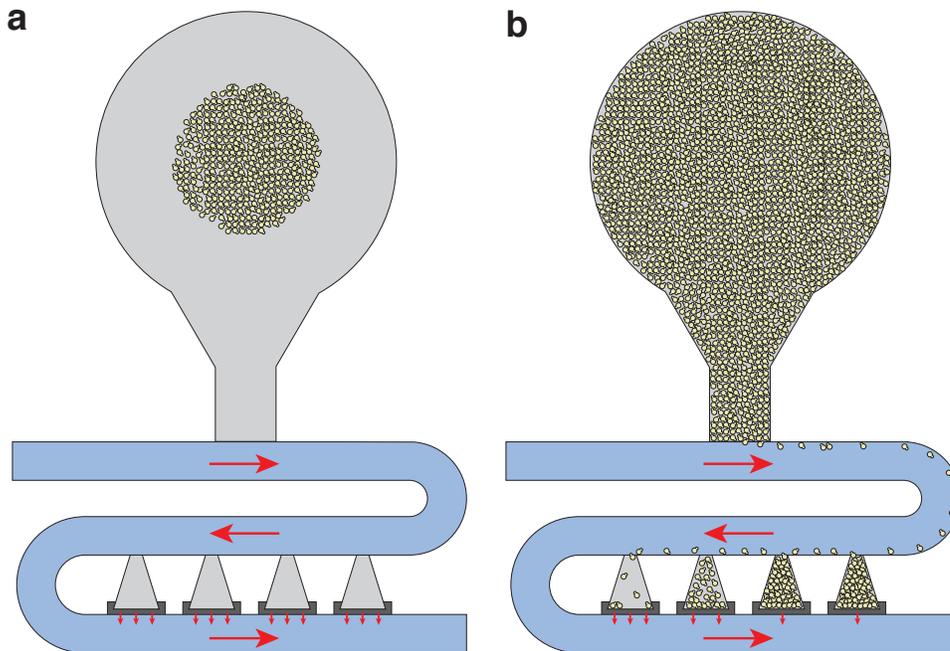


Figure 4.2: Hydrodynamic trapping in HD biopixels. a) A strain bank immediately after media is introduced onto the chip before the initial spotted cells revive. b) A strain bank with cells growing out of the cell spotting region and filling the downstream HD biopixels. Red arrows denote direction of media flow. HD biopixels filled with more cells have an increased resistance and therefore less media flow through them, causing subsequent traps to fill.

4.2.2 Microfluidic device validation

In order to ensure that the measured gene expression is independent of each device position, a single strain, a high-affinity glucose transporter, HXT2, was spotted in every position. Once traps filled to confluence on CSM-His+2% glucose+ 0.04% tween media, the cells were switched to CSM-His + 2% glucose + 2% galactose + 0.04% tween. After 12 hours, the glucose level was dropped serially every two hours. Once steady state was reached on 2% galactose, the media was switched back to 2% glucose. Gene expression levels were extracted for each biopixel, background normalized, and averaged across all 8,704 HD biopixels on the device. As seen in Figure 4.3, replicates across the device maintained statistically similar response patterns through the duration of the experiment. HXT2 expression increased between 0.2-0.02% glucose, but decreased when glucose levels dropped below 0.02%, consistent with the literature [94].

Different baselines of expression are also apparent when cells were grown on 2% glucose versus 2% galactose.

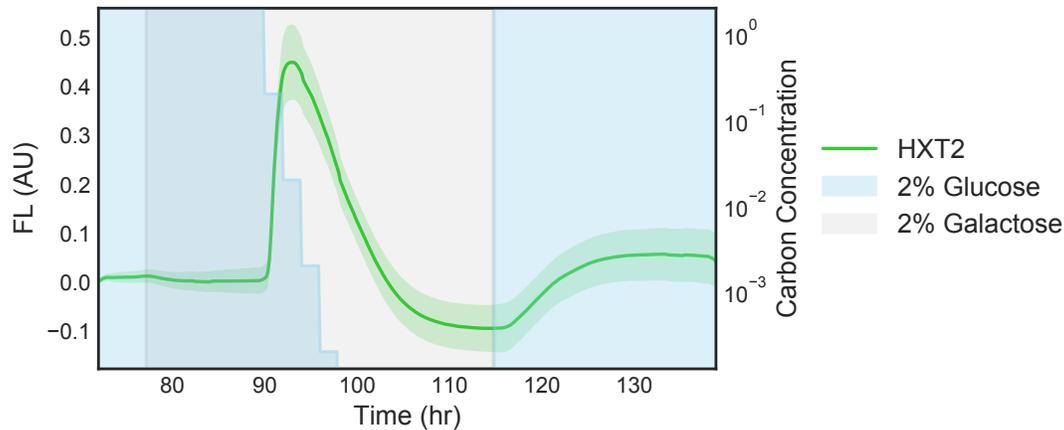


Figure 4.3: HXT2 expression in response to glucose depletion in the presence of galactose. 2,176 strain banks were spotted with HXT2 for a total of 8,704 replicate biopixels across one microfluidic device. The concentrations of glucose (blue shading) and galactose (gray shading) varied throughout the experiment. The green line denotes the average expression of all replicates while the shaded region denotes the standard deviation.

4.2.3 Fermentation monitoring of a diauxic shift

The diauxic shift is one of the most well studied pathways in *S. cerevisiae* and encompasses glucose metabolism, ethanol production, and a shift in machinery and metabolism for ethanol consumption. As a proof-of-concept, two 2,176-strain HD biopixel devices loaded with a 4,156-strain GFP-fusion library [24] were used to monitor the diauxic shift of a batch culture in real time.

Briefly, the two devices with 4,156 strains were grown to confluence on YPD media and imaged on the Dynamics optical enclosure described in Section 3.1.5. Separately, a batch culture of an S288C derived strain, DBY12020, was grown on YPD media. Every hour, supernatant from the batch culture was pipetted onto the microfluidic devices, and the gene expression of the GFP-fusion library was measured. As the supernatant composition changed due to the metabolism of

the batch culture, the library gene expression also changed and reflected that media composition.

The OD of the batch culture and measured ethanol and glucose concentrations are shown in Figure 4.4. As expected, the glucose concentration remains around 2% while cells are in lag before exponential growth. As exponential growth begins, the Crabtree effect is observed as the glucose concentration drops to 0% while the ethanol concentration rises to about 1.5%. Cells enter lag phase while adjusting machinery and metabolism for ethanol consumption in the diauxic shift phase. Post diauxic growth is then observed with cells using ethanol as their carbon source until the ethanol is consumed and cells enter stationary phase.

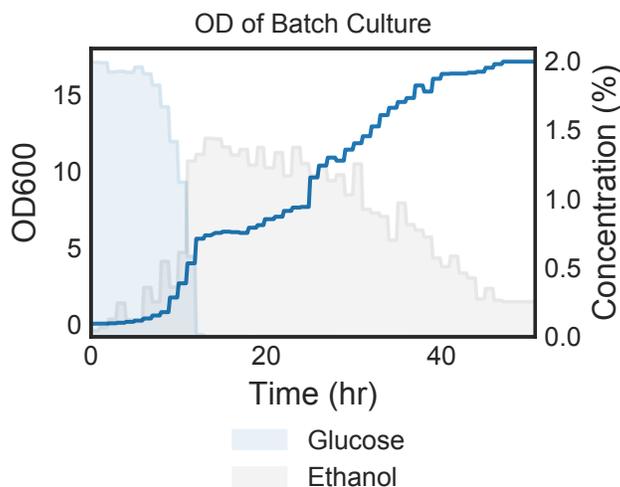


Figure 4.4: Batch culture OD (blue line) and measured glucose (blue shading) and ethanol (gray shading) concentrations over time.

The supernatant from the batch culture with the glucose and ethanol concentrations shown in Figure 4.4 was pipetted hourly onto the two microfluidic devices with 4,156 GFP-fusion strains. After excluding clogged strains and strains with gene expression below the detection limit, 3,761 strains were analyzed. Measured gene expression is reported as relative intensity, where the value at each time point is calculated as the percentage of total expression for each strain.

The relative intensity profile of the 3,761 strains grown on the supernatant of the batch culture in real-time is seen in Figure 4.5. While some strains have drastic changes in relative

intensity, a large bulk of the strains have much more subtle changes. To further characterize these changes, agglomerative clustering with Gene Ontology (GO) enrichment was performed on this dataset and can be seen in Figure 4.6.

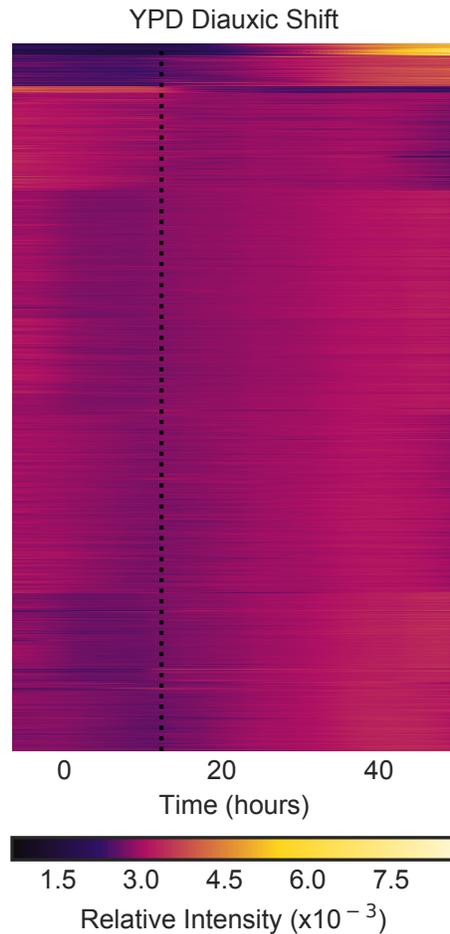


Figure 4.5: Relative intensity of 3,761 strains grown on the supernatant from a batch culture undergoing a diauxic shift. The microfluidic device was induced with the first batch culture sample at time = 0 hours. The dotted black line shows where the glucose concentration drops to 0%.

Six unique clusters of gene expression dynamics were observed in this dataset. Clusters 1 and 2 both include up-regulated genes with similar GO terms that are significant. Though the dynamics are similar in the two clusters, the fold change is higher in cluster 1 and cluster two has quicker response time upon the depletion of glucose. Both clusters include genes that

have been previously reported [95, 96] as genes that are upregulated during a diauxic shift such as ACS1, an acetyl-coA synthetase, ALD4, an aldehyde dehydrogenase required for growth on ethanol, and ADH2, a glucose repressible alcohol dehydrogenase that converts ethanol into acetylaldehyde. Similar GO terms were enriched between clusters 1 and 2, including the TCA cycle, the oxidation-reduction process, oxidative phosphorylation, cellular response to oxidative stress, and negative regulation of ATPase activity. These results are not surprising because glucose is converted to ethanol during glucose metabolism and the TCA cycle and oxidative phosphorylation does not occur until ethanol is metabolized and converted into acetyl-coA. A full list of significantly enriched GO terms for each cluster can be found in Tables 4.1-4.3. These clusters alone depict that the cells indeed convert glucose to ethanol and then metabolize the ethanol.

Clusters 3 and 4 both include down-regulated strains, but with different dynamics. Cluster 3 only includes 36 strains. Expression is relatively stable during growth on glucose and growth on ethanol, but has a very sharp decrease when glucose is depleted. Genes in this cluster include HXT2, HXT3, and HXT6, all of which are high-affinity glucose transporters that are not needed by the cell when glucose is depleted. Ribosomal proteins are also in this category, which too have been reported as down-regulated during a diauxic shift [95, 96]. Cluster 4 also includes several ribosomal proteins, with translation and ribosome being two of the significantly enriched GO terms. This cluster has a more subtle fold change and has decreasing expression during growth on glucose and growth on ethanol, but stable expression during the diauxic lag. Other genes in this cluster include TAT1, an amino acid transporter, NOG1, a GTPase that associates with ribosomes, and LYP1, a lysine permease that uptakes amino acids. As nutrients are depleted and cells enter stationary phase, it is expected that translation and ribosomal proteins would decrease in expression.

Clusters 5 and 6 include the bulk of the strains with 2,140 strains and 838 strains, respectively. The dynamics for both clusters are similar, with a decrease in expression during growth on

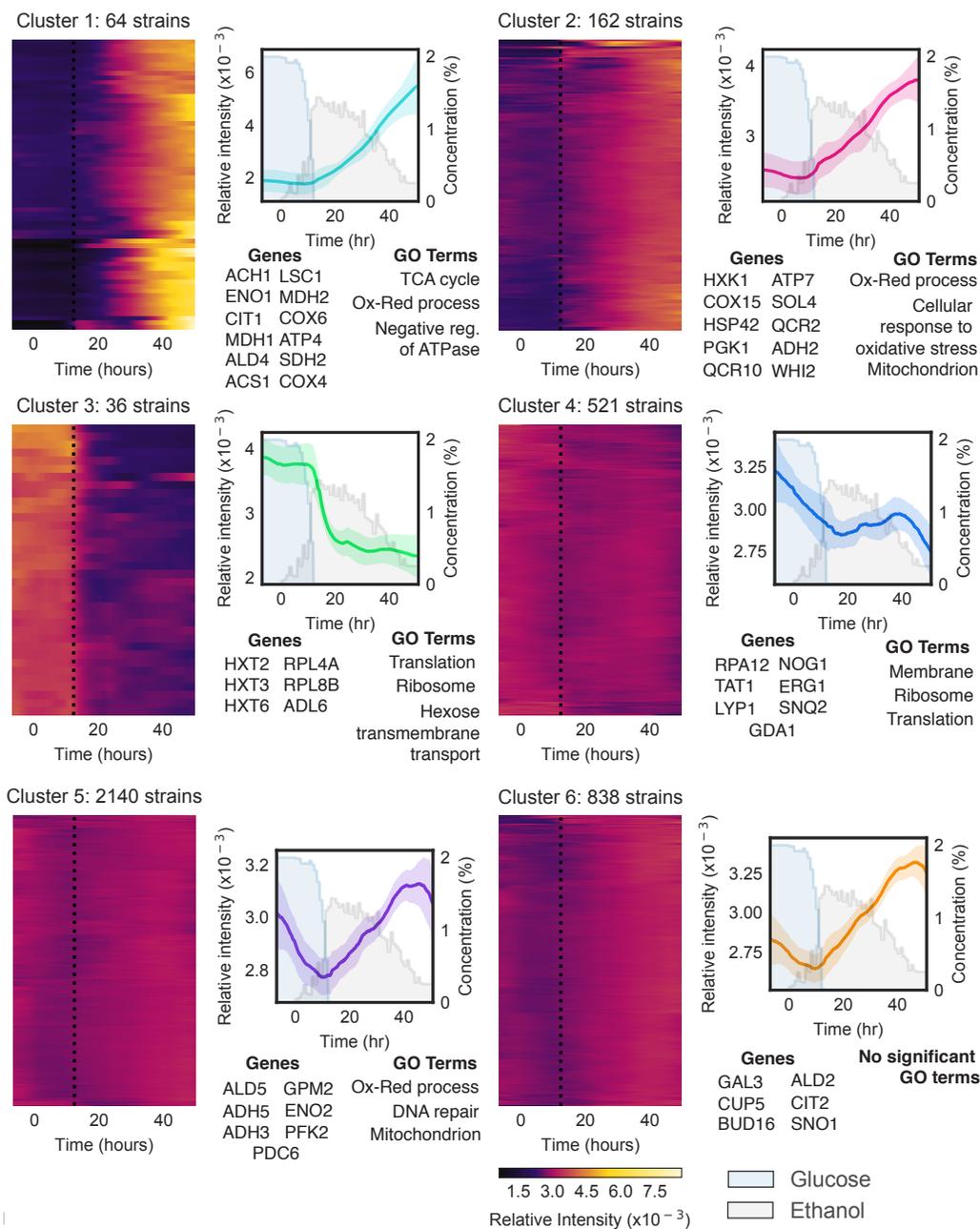


Figure 4.6: Agglomerative clustering and GO enrichment of the relative intensity of 3,761 strains grown on the supernatant from a batch culture undergoing a diauxic shift. The relative intensity of each gene in a cluster is plotted as a row on the heatmaps. The microfluidic device was induced with the first batch culture sample at time = 0 hours. The dotted black line shows where the glucose concentration drops to 0%. The average relative intensity (solid line) and standard deviation (shaded region) for all strains in a cluster is plotted on the line graph. Glucose (blue shading) and ethanol (gray shading) concentrations are also plotted. Representative genes and three significant GO terms from each cluster are listed. The heat map color bar and legend for all plots is shown.

glucose, an increase in expression during the diauxic shift and post-diauxic growth, and stabilized or decreasing expression when entering stationary phase. Overall the fold change is much smaller than the other clusters and there is little change in expression. While cluster 5 has significant GO terms, it encompasses 2,140 strains, so many of these GO terms are likely significant because of the large amount of strains in the cluster. Cluster 6 does not have any significant GO terms. Both clusters 5 and 6 include genes that should have little to no change in expression such as ALD5, an aldehyde dehydrogenase that is constitutive, CUP5, a proteolipid subunit important for copper binding, and GAL3, a galactose transcriptional regulator. However, both clusters also include strains that are expected to be up or down regulated such as ADH3, an aldehyde dehydrogenase that should be up-regulated, and CIT2, a citrate synthase that should be up-regulated. It is possible that the expression of these genes was too low to be detected by the fluorescence optics or that the fluorescence was lower than that of the YPD media or auto-fluorescence of the cells. To more accurately determine which of these 3,000 strains do not respond to the shift or are below the detection limit, further testing would need to be done such as determining the linear range of the optics and testing expression of these strains on lower fluorescent media.

By observing all six clusters together, one can determine not only that the batch culture underwent a diauxic shift, but also when the shift occurred based on the change in slope of expression. Several of the clusters depict exactly when glucose was depleted. Cluster 4 shows when growth resumes on ethanol, and clusters 5 and 6 show when cells enter stationary phase. While the diauxic shift is relatively simple and the phases of the shift can be determined simply by measuring the OD of the batch culture, this device could likely monitor more complex fermentation processes with multiple sugars, showing the time and rate of consumption of each sugar in real-time.

In addition to looking at global response dynamics and characteristics, individual traces can also inform about the fermentation process. Figure 4.7 shows several individual traces from the diauxic shift. The high-affinity glucose transporters (Figure 4.7a) have similar dynamics to one

another, but a decrease in gene expression starts at different glucose concentrations. With further characterization, the gene expression of these strains could be used to determine the concentration of glucose during fermentation processes. These proteins are also expressed higher at low levels of glucose and the rate of glucose depletion affects the slope of that increase in gene expression. At the rate of depletion in this batch culture experiment, HXT3 expression increases slightly when glucose is low, however HXT2 does not. At slower rates of depletion such as the experiment in Figure 4.3, HXT2 has a much greater increase in expression. This information could lead to the calculation of the rate of glucose depletion. The aldehyde dehydrogenases (Figure 4.7b) also offer additional information and can better pinpoint the start of ethanol metabolism after the diauxic shift lag. Understanding when cells halt metabolism or start growth again is key to troubleshooting fermentation batches, especially when testing new strains with unpredicted shifts in growth.

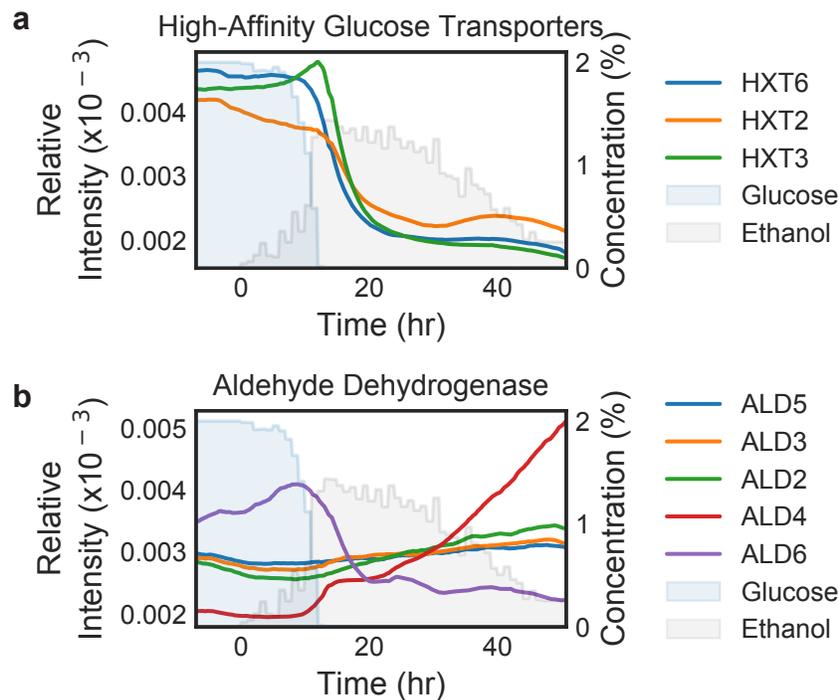


Figure 4.7: Relative intensity of high-affinity glucose transporters (a) and aldehyde dehydrogenases (b) grown on supernatant from a batch culture undergoing a diauxic shift. The microfluidic device was induced with the first batch culture sample at time = 0 hours.

4.3 Conclusion

The fermentation industry lacks real-time monitoring that would greatly decrease cost and tedious sampling, and increase understanding of real-time metabolism to improve the product, yield, and overall performance of the process. Here we have developed a 2,176-strain microfluidic device spotted with a GFP-fusion library of *S. cerevisiae*, that coupled with fluorescence imaging can monitor the metabolism of a batch culture. In this work we performed a proof-of-concept experiment where we manually took supernatant off of a batch culture of *S. cerevisiae*, and pipetted the supernatant onto two microfluidic devices with a 4,156-strain GFP-fusion library. We measured gene expression across all strains and used agglomerative clustering and gene ontology enrichment to accurately depict what carbon source the batch culture was utilizing and when the batch culture went through the different phases of a diauxic shift.

Although a diauxic shift is a simple fermentation process, many fermentation processes involve multiple carbon sources with sequential metabolism. This genome-scale device could be characterized to inform what carbon sources are being metabolized, in what order, and at what rate. The immense amount of information in over 4,000 dynamic *S. cerevisiae* responses also likely means that the concentration of different metabolites could be determined at different time points. For simpler fermentation processes, this device could be decreased in size to have on the order of tens of strains, greatly decreasing the flow rate and amount of supernatant consumed off of a batch culture, ultimately increasing yield of the product.

While this device provides far more information for understanding a fermentation process than existing technologies, the device would benefit from being continuous. This version of the device requires an operator to sample the batch culture and add it to the device, however, a simple peristaltic pump could be included with a microfluidic cell filter [97] in order to continuously sample from a batch culture without operator input. Another limitation of the device is that the fluorescence of the sample can mask the fluorescence of the proteins produced by the cells. There

Table 4.1: Significant GO terms from up-regulated genes in Clusters 1 and 2 from Figure 4.5 with a p-value less than 0.05.

Cluster	Term	Term Name	P-Value
1	GO:0006099	tricarboxylic acid cycle	7.681185e-11
1	GO:0005739	mitochondrion	7.681185e-11
1	GO:0055114	oxidation-reduction process	1.042798e-09
1	GO:0016491	oxidoreductase activity	8.751906e-07
1	GO:0006103	2-oxoglutarate metabolic process	3.135549e-03
1	GO:0032780	negative regulation of ATPase activity	3.135549e-03
1	GO:0009353	mitochondrial oxoglutarate dehydrogenase complex	3.135549e-03
1	GO:0004591	oxoglutarate dehydrogenase (succinyl-transferr...	3.135549e-03
1	GO:0042645	mitochondrial nucleoid	3.849536e-03
1	GO:0005759	mitochondrial matrix	5.223755e-03
1	GO:0046872	metal ion binding	6.496136e-03
1	GO:0006119	oxidative phosphorylation	6.607367e-03
1	GO:0006546	glycine catabolic process	6.607367e-03
1	GO:0005960	glycine cleavage complex	6.607367e-03
1	GO:0004375	glycine dehydrogenase (decarboxylating) activity	6.607367e-03
1	GO:0006090	pyruvate metabolic process	1.074172e-02
1	GO:0003824	catalytic activity	1.074172e-02
1	GO:0015986	ATP synthesis coupled proton transport	1.370487e-02
1	GO:0046933	proton-transporting ATP synthase activity, rot...	1.370487e-02
1	GO:0005743	mitochondrial inner membrane	1.686306e-02
1	GO:0006094	gluconeogenesis	1.713736e-02
2	GO:0005739	mitochondrion	1.726741e-08
2	GO:0055114	oxidation-reduction process	2.804730e-08
2	GO:0016491	oxidoreductase activity	1.939967e-05
2	GO:0006096	glycolytic process	1.719811e-04
2	GO:0005758	mitochondrial intermembrane space	1.911379e-04
2	GO:0005737	cytoplasm	4.148816e-04
2	GO:0003779	actin binding	1.148191e-03
2	GO:0003824	catalytic activity	1.148191e-03
2	GO:0051015	actin filament binding	4.852143e-03
2	GO:0006094	gluconeogenesis	7.088189e-03
2	GO:0034599	cellular response to oxidative stress	1.901500e-02
2	GO:0046323	glucose import	3.221135e-02
2	GO:0000340	RNA 7-methylguanosine cap binding	3.221135e-02
2	GO:0051016	barbed-end actin filament capping	4.013873e-02

Table 4.2: Significant GO terms from down-regulated genes in Clusters 3 and 4 from Figure 4.5 with a p-value less than 0.05.

Cluster	Term	Term Name	P-Value
3	GO:0002181	cytoplasmic translation	2.328590e-06
3	GO:0005840	ribosome	1.649079e-05
3	GO:0003735	structural constituent of ribosome	4.736327e-05
3	GO:0022625	cytosolic large ribosomal subunit	1.838265e-04
3	GO:0006412	translation	2.055458e-04
3	GO:1904659	glucose transmembrane transport	3.070526e-04
3	GO:0015761	mannose transmembrane transport	3.070526e-04
3	GO:0015755	fructose transmembrane transport	3.070526e-04
3	GO:0098704	carbohydrate import across plasma membrane	3.070526e-04
3	GO:0005351	carbohydrate:proton symporter activity	3.070526e-04
3	GO:0005353	fructose transmembrane transporter activity	3.070526e-04
3	GO:0015149	hexose transmembrane transporter activity	3.070526e-04
3	GO:0015578	mannose transmembrane transporter activity	3.070526e-04
3	GO:0005355	glucose transmembrane transporter activity	3.070526e-04
3	GO:0055085	transmembrane transport	5.564942e-04
3	GO:0005886	plasma membrane	8.872409e-04
3	GO:0008643	carbohydrate transport	2.495324e-03
3	GO:0022857	transmembrane transporter activity	3.810879e-03
3	GO:0071944	cell periphery	4.946921e-03
3	GO:0005215	transporter activity	5.270063e-03
3	GO:0005634	nucleus	1.365578e-02
3	GO:0008645	hexose transmembrane transport	2.150518e-02
3	GO:0003746	translation elongation factor activity	2.924513e-02
3	GO:1902600	proton transmembrane transport	3.622409e-02
4	GO:0016021	integral component of membrane	5.195558e-10
4	GO:0016020	membrane	7.699885e-07
4	GO:0002181	cytoplasmic translation	3.090864e-06
4	GO:0022625	cytosolic large ribosomal subunit	1.547123e-05
4	GO:0003735	structural constituent of ribosome	1.683960e-04
4	GO:0005840	ribosome	9.070519e-04
4	GO:0005737	cytoplasm	9.185954e-04
4	GO:0000139	Golgi membrane	1.646644e-03
4	GO:0005783	endoplasmic reticulum	2.175216e-03
4	GO:0005524	ATP binding	3.952228e-03
4	GO:0055085	transmembrane transport	7.190739e-03
4	GO:0006412	translation	4.189781e-02

Table 4.3: Significant GO terms from genes in Cluster 5 from Figure 4.5 with a p-value less than 0.05.

Cluster	Term	Term Name	P-Value
5	GO:0005739	mitochondrion	1.784191e-07
5	GO:0055114	oxidation-reduction process	2.820886e-05
5	GO:0003735	structural constituent of ribosome	3.471530e-04
5	GO:0005840	ribosome	3.807108e-04
5	GO:0005634	nucleus	1.005850e-03
5	GO:0016491	oxidoreductase activity	1.760415e-03
5	GO:0006281	DNA repair	4.826727e-03
5	GO:0006412	translation	5.203735e-03
5	GO:0006974	cellular response to DNA damage stimulus	7.110102e-03
5	GO:0002181	cytoplasmic translation	7.110102e-03
5	GO:0003677	DNA binding	7.110102e-03
5	GO:0007049	cell cycle	1.354622e-02
5	GO:0003824	catalytic activity	1.423408e-02
5	GO:0051301	cell division	4.409527e-02

is likely an upper limit to a sample's fluorescence for proper monitoring, though strains of *S. cerevisiae* could be engineering to have a greater fluorescence.

Finally, not only does this device serve as a tool for fermentation monitoring, this device can also be used to monitor gene expression of *S. cerevisiae* in response to any environmental perturbations. This high-throughput device can house twice as many strains as the other high-throughput *S. cerevisiae* device and is simpler to operate. This device can be used to better understand genetic pathways in yeast and advance research in aging, cancer, drug discovery and other related fields.

4.4 Materials and methods

4.4.1 Wafer Fabrication

The silicon wafer was fabricated using standard photolithography techniques previously described by our group [15]. The conduit layer was fabricated using 2002 SU-8 photoresist with

a spin speed of 1000 rpm and had a resulting height of 2.3 μms , the spotting region layer was fabricated using 2005 SU-8 photoresist with a spin speed of 2600 rpm and had a resulting height of 5.3 μms , the HD biopixel layer was fabricated using 2007 SU-8 photoresist with a spin speed of 1000 rpm and had a resulting height of 13.1 μms , the minor channel layer was fabricated using 2075 SU-8 photoresist with a spin speed of 4000 rpm and had a resulting height of 55-65 μms , and the major channel layer was fabricated using 2075 SU-8 photoresist with a spin speed of 2400 on top of an undeveloped major channel layer, resulting in a final height of 145-170 μms .

4.4.2 Microfluidic device fabrication

The microfluidic devices were fabricated, loaded, and bonded using the protocol in section 2.5.

4.4.3 Batch culture diauxic shift experimental methods

The Yeast GFP Clone Collection [24] library was spotted across two microfluidic devices per the protocol in section 2.5. Cells were grown on YPD + 0.04% tween media with a flow rate of 5 mL/hr for about 24 hours until traps were filled to confluence.

Simultaneously, 4 x 400 mL of YPD + 0.04% tween in 2L flasks were inoculated with DBY12020 at an OD of 0.02. The flasks were shaken at 270 RPM at 30°C.

Each hour, 5 mL from each of the four 2L flasks was pipetted off and mixed. The OD was measured, cells were filtered out, and the supernatant was introduced on each microfluidic device. 2 mL of each supernatant sample was frozen for glucose and ethanol concentration measurements. The microfluidic devices were imaged using the Dynamics optical enclosure described in Section 3.1.5 with a temporal resolution of ten minutes at a 1 second and 60 second exposure for transmitted light and fluorescence channels, respectively. Glucose concentrations were measured using a YSI 2950 metabolite analyzer at the Cancer Metabolism group at the

Sanford Burnham Prebys Medical Discovery Institute. Ethanol concentrations were measured using an Ethanol Assay Kit (Millipore Sigma MAK076-1KT).

4.4.4 Data processing

Images were flat-field corrected and registered to an extraction mask to account for drifting and thermal expansion as described in section 3.3.5.

The mean pixel value for each biopixel was extracted and normalized by the local background as seen in Figure 4.8 using the following equation:

$$GFP_{Raw_n} = \frac{GFP_{HDBP_n} - GFP_{Background}}{GFP_{Background}} \quad (4.1)$$

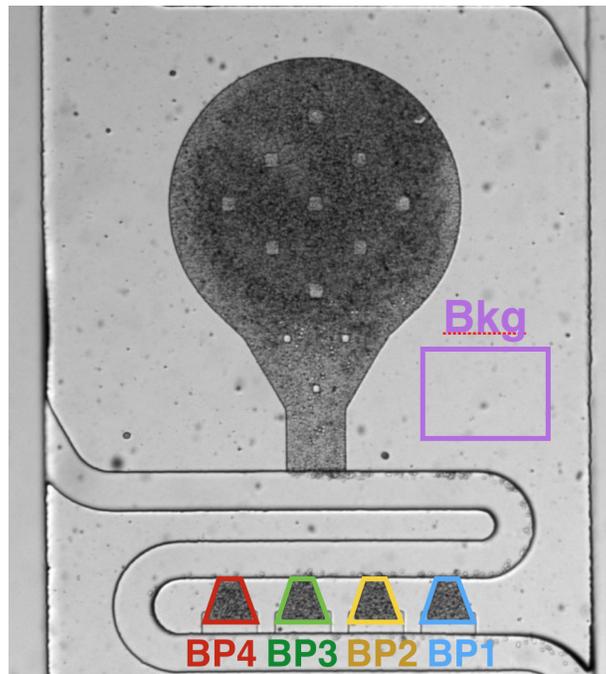


Figure 4.8: In the strain bank shown, data is extracted from each of the colored regions for both transmitted light and fluorescence images. The data for each biopixel is normalized by the background as in equation 4.1 and then the four biopixels are averaged together. 2,176 sets of this data are extracted from every image.

To account for change in cellular autofluorescence and media fluorescence over the

duration of an experiment, several copies of a BY4741 control strain were spotted onto the device with the GFP-Fusion library [24]. For the construction of the control strain, SpHIS5 from Addgene plasmid 44836 was inserted into the his3 region of the BY4741 parent strain, BY4741 his3::SpHIS5. The mean fluorescence of control strain biopixel replicates was subtracted from each strain on the device. Replicate biopixels were then averaged for each strain and the average trace was smoothed.

Due to imperfections in local background correction, some traces resulted in negative values post-control-strain correction. To eliminate negative values and values close to zero, the minimum value across all strains plus 0.1 was added to each strain. Relative intensity at each time point was calculated as the percentage of total expression for each strain.

Excluded Data

All experiments included transient periods over the first several days of the experiment. These transients were caused by the colonies' recoveries post-spotting shock and their growth-to-effluence within their individual traps. Therefore, time traces more than eight hours before the first induction were excluded from analysis. Clogged or empty biopixels were also excluded from analysis.

Strains below the detection limit of the device were also excluded. The detection limit was calculated using raw fluorescence traces that were extracted from flat-field corrected images. The total area under the curve was calculated for each strain. Strains with areas below that of the lowest control strain were deemed not detectable and excluded from analysis.

4.4.5 Data analysis

Data was clustered using agglomerative clustering in the scikit-learn python clustering package. Gene ontology enrichment was done using GOATOOLS in Python [98]. The p-value used for significance was 0.05 and the False Discovery Rate test used was the Benjamini/Hochberg

test.

4.5 Acknowledgements

This chapter contains material being prepared for submission with the following individuals: Stasiowski, Elizabeth*, O’Laughlin, Richard*, and Hasty, Jeff. (*equal contribution) The dissertation author was one of the primary authors and researchers of this material.

Chapter 5

Summary

Systems and synthetic biology have evolved over the past several decades and have led to a deeper understanding of biological networks, an emergence of multiplexed genome-scale technology, and an application of this information and technology in various industries. While microfluidic technology offers many advantages to studying networks in systems and synthetic biology, current technologies are either low-throughput or multi-layered complex devices that limit them to lab-based applications. In this work, we engineered a simpler, one-layer, microfluidic device that can continuously monitor 2,176 strains of *E. coli* or *S. cerevisiae* for up to 14 days at a time.

We have provided a detailed protocol in Chapter 2 for the manufacturing of these devices so that they can be replicated and used in other laboratories. In Chapter 3 we optimize this device for measuring *E. coli* gene expression and coupled with XAI and LSTM-RNN algorithms, are able to detect copper, lead, chromium, cadmium, iron, and zinc in a laboratory setting, as well as in municipal water from around the United States and in the Gold King Mine Spill that occurred in August 2015. The ability to detect multiple heavy metals in real-time fills a need in an industry where metal presence is discontinuously tested in laboratory settings on specialized equipment. In Chapter 4 we optimize this device for measuring gene expression in *S. cerevisiae* and use a

GFP-fusion library to monitor the gene expression and metabolic profile of a diauxic shift in batch in real-time through agglomerative clustering and GO enrichment. Current fermentation technology is limited in what analytes can be measured such as overall sugar content but not the specific carbohydrate profile, and requires specialized equipment that many smaller companies cannot afford. This technology would not only increase information gathered such as specifically which carbohydrates are being metabolized and at what rate, but would also be more accessible to smaller companies. In both of these applications, the large number of strains allows for the device to be versatile and likely able to detect much more than the analytes tested in this work. These devices can be used for a variety of applications, including but not limited to, studying libraries of strains to understand dynamic gene expression in response to external perturbations, multiplexed drug screens, monitoring the fermentation process of a batch culture, screening preliminary strains to understand their metabolic profile through a fermentation process before production, and sensing of various analytes.

Microfluidics is a powerful technology that currently has little application in the biotechnology industry due to its complicated protocols and non-robust nature. The work in this dissertation reduces these complications and will benefit both the industrial and academic communities, making multiplexed microfluidics more accessible.

Bibliography

- [1] Markus Krupp, Jens U. Marquardt, Ugur Sahin, Peter R. Galle, John Castle, and Andreas Teufel. RNA-Seq Atlas—a reference database for gene expression profiling in normal tissue by next-generation sequencing. *Bioinformatics*, 28(8):1184–1185, apr 2012.
- [2] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastri, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells, 2018.
- [3] Nicholas T. Ingolia, Sina Ghaemmaghami, John R.S. Newman, and Jonathan S. Weissman. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*, 2009.
- [4] Yuen Ho, Albrecht Gruhler, Adrian Heilbut, Gary D. Bader, Lynda Moore, Sally Lin Adams, Anna Millar, Paul Taylor, Keiryn Bennett, Kelly Boutilier, Lingyun Yang, Cheryl Wolting, Ian Donaldson, Søren Schandorff, Juanita Shewnarane, Mai Vo, Joanne Taggart, Marilyn Goudreault, Brenda Muskat, Cris Alfarano, Danielle Dewar, Zhen Lin, Katerina Michalickova, Andrew R. Willems, Holly Sassi, Peter A. Nielsen, Karina J. Rasmussen, Jens R. Andersen, Lene E. Johansen, Lykke H. Hansen, Hans Jespersen, Alexandre Podtelejnikov, Eva Nielsen, Janne Crawford, Vibeke Poulsen, Birgitte D. Sørensen, Jesper Matthiesen, Ronald C. Hendrickson, Frank Gleeson, Tony Pawson, Michael F. Moran, Daniel Durocher, Matthias Mann, Christopher W.V. Hogue, Daniel Figeys, and Mike Tyers. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry, 2002.
- [5] D A Lashkari, J L DeRisi, J H McCusker, A F Namath, C Gentile, S Y Hwang, P O Brown, and R W Davis. Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl. Acad. Sci.*, 1997.
- [6] Teresa M. Przytycka, Mona Singh, and Donna K. Slonim. Toward the dynamic interactome: It’s about time. *Briefings in Bioinformatics*, 11(1):15–29, 2010.
- [7] Roger P. Alexander, Philip M. Kim, Thierry Emonet, and Mark B. Gerstein. Understanding modularity in molecular networks requires dynamics. *Science Signaling*, 2(81), 2009.

- [8] Gurkan Bebek, Mehmet Koyutürk, Nathan D. Price, and Mark R. Chance. Network biology methods integrating biological data for translational science. *Briefings in Bioinformatics*, 2012.
- [9] Maria Secrier and Reinhard Schneider. Visualizing time-related data in biology, a review. *Briefings in Bioinformatics*, 2013.
- [10] M. Omar Din, Tal Danino, Arthur Prindle, Matt Skalak, Jangir Selimkhanov, Kaitlin Allen, Ellixis Julio, Eta Atolia, Lev S. Tsimring, Sangeeta N. Bhatia, and Jeff Hasty. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature*, 536(7614):81–85, 2016.
- [11] Hyun Ju Kim, Haeyoung Jeong, and Sang Jun Lee. Synthetic biology for microbial heavy metal biosensors. *Analytical and Bioanalytical Chemistry*, 410(4):1191–1203, feb 2018.
- [12] Neelam Verma and Minni Singh. Biosensors for heavy metals. *BioMetals*, 18(2):121–129, 2005.
- [13] C. J. Paddon, P. J. Westfall, D. J. Pitera, K. Benjamin, K. Fisher, D. McPhee, M. D. Leavell, A. Tai, A. Main, D. Eng, D. R. Polichuk, K. H. Teoh, D. W. Reed, T. Treynor, J. Lenihan, H. Jiang, M. Fleck, S. Bajad, G. Dang, D. Dengrove, D. Diola, G. Dorin, K. W. Ellens, S. Fickes, J. Galazzo, S. P. Gaucher, T. Geistlinger, R. Henry, M. Hepp, T. Horning, T. Iqbal, L. Kizer, B. Lieu, D. Melis, N. Moss, R. Regentin, S. Secrest, H. Tsuruta, R. Vazquez, L. F. Westblade, L. Xu, M. Yu, Y. Zhang, L. Zhao, J. Lievense, P. S. Covello, J. D. Keasling, K. K. Reiling, N. S. Renninger, and J. D. Newman. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446):528–532, 2013.
- [14] Harry Yim, Robert Haselbeck, Wei Niu, Catherine Pujol-Baxley, Anthony Burgard, Jeff Boldt, Julia Khandurina, John D. Trawick, Robin E. Osterhout, Rosary Stephen, Jazell Estadilla, Sy Teisan, H. Brett Schreyer, Stefan Andrae, Tae Hoon Yang, Sang Yup Lee, Mark J. Burk, and Stephen Van Dien. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chemical Biology*, 7(7):445–452, 2011.
- [15] M. S. Ferry, I. A. Razinkov, and J. Hasty. *Microfluidics for synthetic biology: From design to execution*, volume 497. 2011.
- [16] Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, 2008.
- [17] Yang Li, Meng Jin, Richard O’Laughlin, Philip Bittihn, Lev S. Tsimring, Lorraine Pillus, Jeff Hasty, and Nan Hao. Multigenerational silencing dynamics control cell aging. *Proceedings of the National Academy of Sciences*, 114(42):11253–11258, 2017.
- [18] Arthur Prindle, Phillip Samayoa, Ivan Razinkov, Tal Danino, Lev S. Tsimring, and Jeff Hasty. A sensing array of radically coupled genetic ‘biopixels’. *Nature*, 481(7379):39–44, 2012.

- [19] Matthew R. Bennett, Wyming Lee Pang, Natalie A. Ostroff, Bridget L. Baumgartner, Sujata Nayak, Lev S. Tsimring, and Jeff Hasty. Metabolic gene regulation in a dynamically changing environment. *Nature*, 454(7208):1119–1122, 2008.
- [20] J. Uhlerdorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences*, 109(35):14271–14276, 2012.
- [21] Jerome T. Mettetal, Dale Muzzey, Carlos Gomez-Urbe, and Alexander van Oudenaarden. The Frequency Dependence of Osmo-adaptation in *Saccharomyces cerevisiae*. *Science*, 319:482–484, 2008.
- [22] Nicolas Dénervaud, Johannes Becker, Ricard Delgado-gonzalo, Pascal Damay, Arun S Rajkumar, and Michael Unser. A chemostat array enables the spatio-temporal analysis of the yeast proteome. *Proceedings of the National Academy of Sciences*, 110(39):15842–15847, 2013.
- [23] Rongfei Zhang, Haiyu Yuan, Shujing Wang, Qi Ouyang, Yong Chen, Nan Hao, and Chunxiong Luo. High-Throughput single-cell analysis for the proteomic dynamics study of the yeast osmotic stress response. *Scientific Reports*, 7(February), 2017.
- [24] Won-ki Huh, James V Falvo, Luke C Gerke, Adam S Carroll, Russell W Howson, Jonathan S Weissman, and Erin K O Shea. Global analysis of protein localization in budding yeast. *Nature*, 425:686–691, 2003.
- [25] Alon Zaslaver, Anat Bren, Michal Ronen, Shalev Itzkovitz, Ilya Kikoin, Seagull Shavit, Wolfram Liebermeister, Michael G. Surette, and Uri Alon. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature Methods*, 3(8):623–628, 2006.
- [26] Yuichi Taniguchi, Paul J Choi, Gene-wei Li, Huiyi Chen, Mohan Babu, Jeremy Hearn, Andrew Emili, and X Sunney Xie. Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*, 329:533–539, 2010.
- [27] Francesca Volpetti, Ekaterina Petrova, and Sebastian J. Maerkl. A Microfluidic Biodisplay. *ACS Synthetic Biology*, 6(11):1979–1987, 2017.
- [28] Ce Zhang, Hsiung-Lin Tu, Gengjie Jia, Tanzila Mukhtar, Verdon Taylor, Andrey Rzhetsky, and Sava Tay. Ultra-multiplexed analysis of single-cell dynamics reveals logic rules in differentiation. *Science Advances*, 5(4):eaav7959, apr 2019.
- [29] Boris Kholodenko, Michael B Yaffe, and Walter Kolch. Computational approaches for analyzing information flow in biological networks. *Science signaling*, 5(220):re1, apr 2012.
- [30] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 2002.

- [31] Galit Lahav, Nitzan Rosenfeld, Alex Sigal, Naama Geva-Zatorsky, Arnold J. Levine, Michael B. Elowitz, and Uri Alon. Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nature Genetics*, 2004.
- [32] S. Hooshangi, S. Thiberge, and R. Weiss. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences*, 2005.
- [33] Jonathan M. Raser and Erin K. O’Shea. Control of stochasticity in eukaryotic gene expression. *Science*, 2004.
- [34] François Jacob and Jacques Monod. Genetic regulatory mechanisms in the synthesis of proteins, 1961.
- [35] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 2000.
- [36] Michael J. Heller. DNA Microarray Technology: Devices, Systems, and Applications. *Annu Rev Biomed Eng.*, 4:129–153, 2002.
- [37] David L Shis, Matthew R Bennett, and Oleg A Igoshin. Dynamics of Bacterial Gene Regulatory Networks. *Annu. Rev. Biophys.*, 47:21(21):1–21, 2018.
- [38] Yolanda T. Chong, Judice L. Y. Koh, Helena Friesen, Supipi Kaluarachchi Duffy, Michael J. Cox, Alan Moses, Jason Moffat, Charles Boone, and Brenda J. Andrews. Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell*, 161:1413–1424, 2015.
- [39] Jeremy E. Purvis and Galit Lahav. Encoding and decoding cellular information through signaling dynamics. *Cell*, 152(5):945–956, feb 2013.
- [40] Todd Thorsen, Sebastian J. Maerkl, and Stephen R. Quake. Microfluidic large-scale integration. *Science*, 298(5593):580–584, 2002.
- [41] Wei Sha, Ana M. Martins, Reinhard Laubenbacher, Pedro Mendes, and Vladimir Shulaev. The Genome-Wide Early Temporal Response of *Saccharomyces cerevisiae* to Oxidative Stress Induced by Cumene Hydroperoxide. *PLoS ONE*, 8(9):1–15, 2013.
- [42] Tong Zhang and Herbert H.P. Fang. Quantification of *Saccharomyces cerevisiae* viability using BacLight. *Biotechnology letters*, 26(12):989–992, 2004.
- [43] Reiko HIRASAWA, Kumio YOKOIGAWA, Yuka ISOBE, and Hiroyasu KAWAI. Improving the Freeze Tolerance of Bakers’ Yeast by Loading with Trehalose. *Bioscience, Biotechnology, and Biochemistry*, 65(3):522–526, 2001.
- [44] Charlemagne Gilles Hounsa, E. Vincent Brandt, Johan Thevelein, Stefan Hohmann, and Bernard A. Prior. Role of trehalose in survival of *Saccharomyces cerevisiae* under osmotic stress. *Microbiology*, 144(3):671–680, 1998.

- [45] Ahmad S Khalil and James J Collins. Synthetic biology: applications come of age. *Nature reviews. Genetics*, 2010.
- [46] Davide Merulla and Jan Roelof Van Der Meer. Regulatable and Modulable Background Expression Control in Prokaryotic Synthetic Circuits by Auxiliary Repressor Binding Sites. *ACS Synthetic Biology*, 2016.
- [47] Jameson K. Rogers and George M. Church. Genetically encoded sensors enable real-time observation of metabolite production. *Proceedings of the National Academy of Sciences*, 113(9):2388–2393, 2016.
- [48] Sambandam Ravikumar, Mary Grace Baylon, Si Jae Park, and Jong il Choi. Engineered microbial biosensors based on bacterial two-component systems as synthetic biotechnology platforms in bioremediation and biorefinery. *Microbial Cell Factories*, 16(1):1–10, 2017.
- [49] Irisappan Ganesh, Sambandam Ravikumar, Seung Hwan Lee, Si Jae Park, and Soon Ho Hong. Engineered fumarate sensing *Escherichia coli* based on novel chimeric two-component system. *Journal of Biotechnology*, 168(4):560–566, 2013.
- [50] Liangpo Li, Ran Tu, Guotian Song, Jie Cheng, Wujiu Chen, Lin Li, Lixian Wang, and Qinhong Wang. Development of a Synthetic 3-Dehydroshikimate Biosensor in *Escherichia coli* for Metabolite Monitoring and Genetic Screening. *ACS synthetic biology*, 8(2):297–306, 2019.
- [51] Jonathan W Kotula, S Jordan Kerns, Lev A Shaket, Layla Siraj, James J Collins, Jeffrey C Way, and Pamela A Silver. Programmable bacteria detect and record an environmental signal in the mammalian gut. *Proceedings of the National Academy of Sciences of the United States of America*, 111(13):4838–43, apr 2014.
- [52] Tal Danino, Kaitlin Allen, Matthew Skalak, Sangeeta N Bhatia, Arthur Prindle, Howard Li, Jeff Hasty, and Gabriel A Kwong. Programmable probiotics for detection of cancer in urine. *Science Translational Medicine*, 7(289):289ra84–289ra84, 2015.
- [53] Carlos Piñero-Lambea, Gustavo Bodelón, Rodrigo Fernández-Periáñez, Angel M. Cuesta, Luis Álvarez-Vallina, and Luis Ángel Fernández. Programming controlled adhesion of *E. coli* to target surfaces, cells, and tumors with synthetic adhesins. *ACS Synthetic Biology*, 4(4):463–473, 2015.
- [54] James A. Imlay. The mismetallation of enzymes during oxidative stress. *Journal of Biological Chemistry*, 289(41):28121–28128, 2014.
- [55] Pete Chandrangsu, Christopher Rensing, and John D. Helmann. Metal homeostasis and resistance in bacteria. *Nature Reviews Microbiology*, 15(6):338–350, mar 2017.
- [56] Kanwal Rehman, Fiza Fatima, Iqra Waheed, and Muhammad Sajid Hamid Akash. Prevalence of exposure of heavy metals and their impact on health consequences. *Journal of Cellular Biochemistry*, 119(1):157–184, 2018.

- [57] Mona Hanna-Attisha, Jenny LaChance, Richard Casey Sadler, and Allison Champney Schnepf. Elevated blood lead levels in children associated with the flint drinking water crisis: A spatial analysis of risk and public health response. *American Journal of Public Health*, 106(2):283–290, 2016.
- [58] Dustin Renwick. Five years on, the Flint water crisis is nowhere near over, 2019.
- [59] M.B. Pell and Joshua Schneyer. The thousands of U.S. locales where lead poisoning is worse than in Flint, 2016.
- [60] Karletta Chief, Janick F Artiola, Paloma Beamer, Sarah T Wilkinson, and Raina M Maier. Understanding the Gold King Mine Spill. *Superfund Research Program*, (April):1–7, 2016.
- [61] Alysa Landry. Navajo Crops Drying Out as San Juan River Remains Closed After Toxic Spill, 2015.
- [62] Luca Pujol, David Evrard, Karine Groenen-Serrano, Mathilde Freyssinier, Audrey Ruffien-Cizsak, and Pierre Gros. Electrochemical sensors and devices for heavy metals assay in water: the French groups’ contribution. *Frontiers in Chemistry*, 2(April):1–24, 2014.
- [63] Rakesh Sharma, Christopher Rensing, P Rosen, Bharati Mitra, and Barry P Rosen. The ATP Hydrolytic Activity of Purified ZntA , a Pb(II)/Cd(II)/Zn(II)-translocating ATPase from Escherichia coli. *The journal of biological chemistry*, 275(Ii):3873–3878, 2000.
- [64] G. Grass and C. Rensing. CueO is a multi-copper oxidase that confers copper tolerance in Escherichia coli. *Biochemical and Biophysical Research Communications*, 286(5):902–908, 2001.
- [65] Sankha Ghatak, Zachary A. King, Anand Sastry, and Bernhard O. Palsson. The y-ome defines the 35% of Escherichia coli genes that lack experimental evidence of function. *Nucleic acids research*, 47(5):2446–2454, 2019.
- [66] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. *ArXiv e-prints*, 2016.
- [67] Sepp Hochreiter and Jurgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1–32, 1997.
- [68] Jonathan P. McHugh, Francisco Rodríguez-Quiñones, Hossein Abdul-Tehrani, Dimitri A. Svistunenko, Robert K. Poole, Chris E. Cooper, and Simon C. Andrews. Global Iron-dependent Gene Regulation in Escherichia coli. *Journal of Biological Chemistry*, 278(32):29478–29486, 2003.
- [69] Christopher Rensing, Bharati Mitra, and Barry P Rosen. The zntA gene of Escherichia coli encodes a Zn(II)-translocating P-type ATPase. *Biochemistry*, 94(December):14326–14331, 1997.

- [70] Jianzhu Ma, Michael Ku Yu, Samson Fong, Keiichiro Ono, Eric Sage, Barry Demchak, Roded Sharan, and Trey Ideker. Using deep learning to model the hierarchical structure and function of a cell. *Nature Methods*, 15(4):290–298, 2018.
- [71] Jason H. Yang, Sarah N. Wright, Meagan Hamblin, Douglas McCloskey, Miguel A. Alcantar, Lars Schrübbers, Allison J. Lopatkin, Sangeeta Satish, Amir Nili, Bernhard O. Palsson, Graham C. Walker, and James J. Collins. A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell*, pages 1649–1661, 2019.
- [72] Jian Zhou and Olga G Troyanskaya. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature methods*, 12(10):931–4, 2015.
- [73] Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. *ArXiv e-prints*, (Section 2):1–10, 2017.
- [74] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. Consistent Individualized Feature Attribution for Tree Ensembles. *ArXiv e-prints*, (2), 2018.
- [75] Scott M Lundberg, Gabriel Erion, Hugh Chen, Alex DeGrave, Jordan M Prutkin, Bala Nair, Ronit Katz, Jonathan Himmelfarb, Nisha Bansal, and Su-In Lee. Explainable AI for Trees: From Local Explanations to Global Understanding. *arXiv preprint arXiv:1905.04610*, 2019.
- [76] Samuel Dodge and Lina Karam. A study and comparison of human and deep learning recognition performance under visual distortions. *2017 26th International Conference on Computer Communications and Networks, ICCCN 2017*, 2017.
- [77] Paul B Tchounwou, Clement G Yedjou, Anita K Patlolla, and Dwayne J Sutton. Molecular, clinical and environmental toxicology: v.2: Clinical toxicology. *Choice Reviews Online*, 47(10):47–5683–47–5683, 2013.
- [78] Maura Allaire, Haowei Wu, and Upmanu Lall. National trends in drinking water quality violations. *Proceedings of the National Academy of Sciences*, 115(9):2078–2083, 2018.
- [79] Anyou Wang and David E Crowley. Global gene expression responses to cadmium toxicity in *Escherichia coli*. *Journal of bacteriology*, 187(9):3259–66, may 2005.
- [80] Stephen P. LaVoie and Anne O. Summers. Transcriptional responses of *Escherichia coli* during recovery from inorganic or organic mercury exposure. *BMC Genomics*, 2018.
- [81] K. R. Brocklehurst and A. P. Morby. Metal-ion tolerance in *Escherichia coli*: Analysis of transcriptional profiles by gene-array technology. *Microbiology*, 2000.
- [82] Leslie A. Pratt and Roberto Kolter. Genetic analysis of *Escherichia coli* biofilm formation: Roles of flagella, motility, chemotaxis and type I pili. *Molecular Microbiology*, 30(2):285–293, 1998.

- [83] Robert A. LaRossa, Dana R. Smulski, and Tina K. Van Dyk. Interaction of lead nitrate and cadmium chloride with *Escherichia coli* K-12 and *Salmonella typhimurium* global regulatory mutants. *Journal of Industrial Microbiology*, 14(3-4):252–258, 1995.
- [84] Huaiyu Mi, Anushya Muruganujan, Dustin Ebert, Xiaosong Huang, and Paul D Thomas. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic acids research*, 47(D1):D419–D426, jan 2019.
- [85] Bobak Shahriari, Kevin Swersky, Ziyu Wang, Ryan P. Adams, and Nando De Freitas. Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175, 2016.
- [86] Zachary Chase Lipton, Charles Elkan, and Balakrishnan Narayanaswamy. Thresholding Classifiers to Maximize F1 Score. *ArXiv e-prints*, 2014.
- [87] Alon Zaslaver, Avi E. Mayo, Revital Rosenberg, Pnina Bashkin, Hila Sberro, Miri Tsalyuk, Michael G. Surette, and Uri Alon. Just-in-time transcription program in metabolic pathways. *Nature Genetics*, 2004.
- [88] Laura S. Busenlehner, Mario A. Pennella, and David P. Giedroc. The SmtB/ArsR family of metalloregulatory transcriptional repressors: Structural insights into prokaryotic metal resistance. *FEMS Microbiology Reviews*, 27(2-3):131–143, 2003.
- [89] Tamar Barkay, Susan M. Miller, and Anne O. Summers. Bacterial mercury resistance from atoms to ecosystems. *FEMS Microbiology Reviews*, 27(2-3):355–384, 2003.
- [90] Joe J. Harrison, Howard Ceri, and Raymond J. Turner. Multimetal resistance and tolerance in microbial biofilms. *Nature Reviews Microbiology*, 5(12):928–938, 2007.
- [91] Fermentation Chemicals Market Size, Share & Trends Analysis Report By Product (Alcohols, Enzymes), By Application (Industrial, Plastic, Pharmaceutical), By Region, And Segment Forecasts, 2019 - 2025. Technical report, Grand View Research, 2019.
- [92] Arne Hagman, Torbjörn Säll, and Jure Piškur. Analysis of the yeast short-term Crabtree effect and its origin. *FEBS Journal*, 281(21):4805–4814, 2014.
- [93] Amelia Ahmad Khalili, Mohd Ahmad, Masaru Takeuchi, Masahiro Nakajima, Yasuhisa Hasegawa, and Razauden Mohamed Zulkifli. A Microfluidic Device for Hydrodynamic Trapping and Manipulation Platform of a Single Biological Cell. *Applied Sciences*, 6(2):40, 2016.
- [94] Amparo Ruiz, Raquel Serrano, and Joaquín Ariño. Direct regulation of genes involved in glucose utilization by the calcium/calcineurin pathway. *Journal of Biological Chemistry*, 283(20):13923–13933, 2008.
- [95] Paola Picotti, Bernd Bodenmiller, Lukas N Mueller, Bruno Domon, and Ruedi Aebersold. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. 71(2):233–236, 2013.

- [96] J Patrick Murphy, Ekaterina Stepanova, Robert a Everley, Joao a Paulo, and Steven P Gygi. Comprehensive Temporal Protein Dynamics during the Diauxic Shift in *Saccharomyces cerevisiae*. *Molecular & Cellular Proteomics*, 14:2454–2465, 2015.
- [97] Majid Ebrahimi Warkiani, Andy Kah Ping Tay, Guofeng Guan, and Jongyoon Han. Membrane-less microfiltration using inertial microfluidics. *Scientific Reports*, 5:1–10, 2015.
- [98] D. V. Klopfenstein, Liangsheng Zhang, Brent S. Pedersen, Fidel Ramírez, Alex Warwick Vesztrocy, Aurélien Naldi, Christopher J. Mungall, Jeffrey M. Yunes, Olga Botvinnik, Mark Weigel, Will Dampier, Christophe Dessimoz, Patrick Flick, and Haibao Tang. GOATOOLS: A Python library for Gene Ontology analyses. *Scientific Reports*, 8(1):1–17, 2018.