UNIVERSITY OF CALIFORNIA SAN DIEGO

A high-throughput microfluidic platform for genome-scale transcriptional dynamics and environmental sensing

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

in

Bioengineering

by

Grégoire Thouvenin

Committee in charge:

Professor Jeff Hasty, Chair
Professor Stephanie Fraley
Professor Xiaohua Huang
Professor Nathan Lewis
Professor Milton Saier

2021

The dissertation of Grégoire Thouvenin is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my loving parents

EPIGRAPH

La Nature est un temple où de vivants piliers
Laissent parfois sortir de confuses paroles;
L'homme y passe à travers des forêts de symboles
Qui l'observent avec des regards familiers.
—Charles Baudelaire

Nature is a temple in which living pillars
Sometimes give voice to confused words;
Man passes there through forests of symbols
Which look at him with understanding eyes.
— Translated by William Aggeler, 1954

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

The Ph.D. is a long journey, and long is the list of the people whose support I am deeply thankful for. I am grateful to my parents for the love and steadfast support that I felt intensely despite the many miles separating California's coast from the shores of Lake Leman. In planting early in their children the seeds of curiosity, perseverance, and conscientiousness, they gave me the richest of parental gifts. I feel deep gratitude towards my siblings, whose continued tolerance of a very microbially-obsessed older brother is as strong a mark of fraternal love as I can imagine. After the long days (and occasional nights) in the lab, I was fortunate to feel around me a support system that spanned multiple time zones. Starting with my amazing girlfriend Laura, whose loving support and optimism were a treasure that I carried with me as I pushed forward and grew as a person. I was also lucky to consistently come home to a dear friend, Pietro, whose presence was synonymous with a delicious dinner and, more importantly, soul-nourishing conversation. I also have in mind all the beautiful humans with whom I could squeeze in a couple games of tennis while my bacteria were growing, an overseas phone call while my PDMS was curing, and deep conversations about science and the meaning of it all while the growth media was stirring. The warmth of all these friendships brings to mind some of the bonfires and campfires we shared on San Diego's beaches and in the beautiful wilderness of the American West. A warmth that acted as a constant reminder that science is a long march, and along the way, one must look for "the little daily miracles, illuminations, matches struck unexpectedly in the dark."

The transformative nature of the years I spent in the Hasty lab is a testament to the quality of the people I worked with. I was blessed to have remarkable scientists as teammates for the Dynomics project. Teammates who I am lucky to count as friends and without whom this thesis wouldn't exist: Garrett, Lizzy, and Nick. I discovered in them beautiful and complementary qualities that I could only strive to match: an insatiable appetite for science, an unfaltering can-do attitude, meticulous experimental patience, and the desire to share discoveries both big and small. I flowered as a scientist and as a human being in their presence. Thank you also to all of the Hasty

Bunch - you consistently put an (admittedly well concealed) smile on my face as I walked into the lab every morning.

Such a unique environment does not form by chance. For bringing together a lab rich in stimulating personalities and scientific talents, I have to thank Jeff Hasty, an incredibly supportive advisor whose passion for synthetic biology is matched by a remarkable ability to spur creativity in a research group. I owe him a deep thank you for giving me the space to grow as a scientist. Working with Jeff was also a great school for unorthodox problem-solving and lateral thinking. Knowing how and when to mix intuition with scientific methodicity to push the boundaries of innovation is a lesson that I will keep with me for a lifetime.

Chapter 3 contains material published in 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 2 and 5 contain material being prepared for submission as Multiplexed microfluidic platforms for real-time heavy metal sensing and dynamic phenotype screening. Thouvenin Gregoire*, Csicsery Nicholas*, O'Laughlin Richard, Stasiowski Elizabeth, Lezia Andrew, and Hasty, Jeff. (*equal contribution). The dissertation author was the primary author and researcher of this material.

Chapter 4: *A microfluidic biosensor for the detection of contaminants in seawater* was co-authored with Andrew Lezia, Alyssia Chiang, Bryan Thai, Michael Ferry, and Jeff Hasty. The dissertation author was the primary author and researcher of this material.

VITA

2011-2014	B. S. in Life Sciences Engineering École polytechnique fédérale de Lausanne
2015-2021	Ph. D. in Bioengineering University of California San Diego

PUBLICATIONS

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. Beal, Jacob, Traci Haddock-Angelli,

Markus Gershater, Kim De Mora, Meagan Lizarazo, Jim Hollenhorst, Randy Rettberg, and iGEM Interlab Study Contributors. "Reproducibility of fluorescent expression from engineered biological constructs in E. coli." PloS one 11, no. 3 (2016): e0150182. The dissertation author was a consortium co-author of this material.

ABSTRACT OF THE DISSERTATION

A high-throughput microfluidic platform for genome-scale transcriptional dynamics and environmental sensing

by

Grégoire Thouvenin

Doctor of Philosophy in Bioengineering

University of California San Diego, 2021

Professor Jeff Hasty, Chair

Genome-scale technologies have transformed our understanding of the biomolecular signaling networks that underpin cellular function and adaptation. 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. In model organisms such as *E. coli*, biomolecular networks are often elucidated by observing how gene expression patterns change in reaction to experimentally-induced perturbations. However, the high-throughput experimental techniques traditionally used for this purpose are inherently destructive and only offer snapshots of a cell's state. As such, these technologies do not fully capture the information encoded in the dynamics of biomolecular networks, which are complex, time-dependent signals.

In the past twenty years, microfluidic technology combined with fluorescence microscopy has established itself as a powerful tool to study time-dependent biological processes while precisely controlling the cellular environment. This thesis focuses on bridging the gap between genome-wide assays and microfluidics-based dynamic perturbation experiments. Here I report the development of a high-throughput microfluidic platform capable of culturing 2176 unique microbial microcolonies in parallel and monitoring the changes in expression of fluorescent proteins in each strain. By loading the platform with some of the readily available libraries of fluorescent transcriptional reporters and dynamically tuning the growth media, I show that we can measure microbial gene expression dynamics in response to environmental inputs *in vivo* and genome-wide.

Chapter 1 provides an overview of the role of high-throughput microfluidics in systems and synthetic biology research. Chapter 2 describes the design of a highly multiplexed microfluidic platform for monitoring gene expression in GFP-tagged *E. coli* with both industrial and research applications. Chapter 3 illustrates the platform's applicability as an environmental biosensor that uses the dynamics of 2000 *E. coli* GFP-promoter strains coupled with machine learning algorithms to detect the presence of heavy metals in drinking water in real-time. Chapter 4 further demonstrates the potential of microfluidics-based biosensing by reporting the use of devices loaded with diverse engineered microbes to detect pollutants in seawater. Finally, in Chapter 5, I use the platform to probe the dynamics of the *S. cerevisiae* proteome in response to the drug metformin and lay the foundations for a new type of dynamics-based chemogenetic screen.

The overarching aim of this research is the capture of microbial gene expression dynamics in response to environmental stimuli on a genome-wide scale with applications in biosensing and the characterization of drug targets.

Chapter 1

Introduction

1.1 Personal perspective

Reading a cellular and molecular biology textbook has always filled me with wonder. Wonder in the face of the networks of biochemical reactions orchestrated by the cell to continually adapt to a changing environment. Of the thousands upon thousands of molecules that act swiftly and with precision in chaotic, crowded conditions to transform matter, transfer energy and transmit information. Of the beautifully coordinated behavior that emerges from this intracellular symphony. As an engineer flipping through the pages of a said textbook and discovering the well-tuned functions accomplished by this "molecular soup", I was rapt with excitement. And full of questions. How could we map out these networks of reactions? Find the design principles that underlie them? Or more intriguingly yet, work with those systems and create new functions inside cells? It was a special moment to realize as a student that we had entered an age in which we could tangle with the cell's inner workings, build **with** biology, and find new ways to interact with living matter. I knew then that I wanted to become a scientist.

Quantitative Biology

Born a descriptive science, biology has morphed into a quantitative discipline. Rapid progress in experimental technology, imaging tools, and computational power has fuelled this transformation and propelled biology into a new era [1]. This thesis is a small contribution to a series of technological developments whose harbinger was the sequencing of the human genome at the turn of the millennium. The post-genomic age has radically changed how we conduct biological research. Endowed with new tools, biologists now face an ever-increasing volume of "omics" data that describes the genes (genomics), mRNA (transcriptomics), proteins (proteomics), and metabolites (metabolomics) inside our cells. Yet this avalanche of data does not always equate better mechanistic understanding [2]. "The map is not the territory" [3], particularly when it comes to predicting the function of biological networks and untangling the "technical debt" of evolution [4].

Numerous scientific disciplines have rallied to help link biological maps to territories of understanding. Mathematics, physics, chemistry, computer science, but also engineering have brought new tools and conceptual approaches to the life sciences. They have bolstered the effort to uncover generalizable principles that govern the molecular underpinnings of life [5]. This movement to make biology quantitative can be traced back to Schrödinger's "What is life?". In his 1944 book, [6], the physicist of feline infamy initiated an approach to biological systems that relies more heavily on the laws of the physical sciences that govern the rest of the material world [7]. In this vein, quantitative biology strives to bridge components with systems, formulate predictions via quantitative models, guide the design of new experiments and technology, and power the generality of (falsifiable) ideas [8].

1.2 Systems and synthetic biology

Two of quantitative biology's offsprings are systems biology and synthetic biology. Both are interdisciplinary approaches that aim to understand, predict, and control living systems [9]. Both have directly benefited from the modern tools of molecular biology, genomic analyses, and high-throughput measurement techniques [10]. While the exact definition of these young fields is difficult to pin down [11], systems biology is often characterized as a top-down knowledge-driven approach, and synthetic biology as a bottom-up, application-driven approach with roots in engineering [12, 13]. Systems biology includes the reconstruction of genome-scale biochemical reaction networks *in silico* [8, 14], and focuses not on the components of the network as much as the relationship between them. This approach links the states of intracellular biological networks to physiological states [15, 16] and enables holistic modeling of entire cells [9]. Systems biology has numerous applications in bioprocess development [17], metabolic engineering or the modeling of complex diseases [18]. Additionally, systems biology presents biologists with an interesting epistemological challenge. The analysis of large omics data sets now goes beyond complementing hypothesis-driven experimentation. It is a source of new hypotheses and knowledge [19].

Synthetic biologists often quote the physicist Richard Feynman to explain their approach to biology: "What I cannot create, I do not understand." Since its inception in the early two-thousands, the field's main focus has been creating synthetic gene circuits by assembling modular genetic elements [20], mostly in microbial organisms. By exploiting the diversity of genetic parts found in nature, synthetic biologists have re-created numerous functions in cells. Toggle switches [21, 22], logic gates [23, 24] oscillators [25–27] and sensors [22, 28, 29] are some salient examples. Additional offshoots of the field include the construction of entirely synthetic genomes [30–32], the expansion of the genetic code [33] and cell-free synthetic biology [34]. Underlying these efforts is an attempt to bring principles of engineering (abstraction, standardization, characterization, modularity, decoupling) and foundational technologies to biology [35].

The acts of designing, modeling, and studying synthetic gene circuits have a two-fold repercussion. First, insight is gained into how evolution shaped biological networks to perform specific tasks in the cell. Second, we learn to create genetic circuits that have a multitude of societal applications. Genetically engineered "living factories [36] can produce chemicals [37, 38], biofuels [39], biomaterials [40], detect chemical cues in their environment [41–44] or autonomously treat diseases [36, 45–47]. Living factories leverage the richness of the biological world to build a more healthy and sustainable future. The desire to engineer biological machines and build better tools to study them is what drove me to pursue a Ph.D.

There are limits to both holistic and engineering approaches to understanding gene networks. On the one hand, the search for design principles using a holistic approach can be made difficult by low signal-to-noise datasets ladened with confounding factors, the existence of nonlinear interactions between components, and genotype-phenotype disparity [2]. Performed without the appropriate experimental grounding, systems biology describes phenomena in a vacuum and creates *in silico* models with little real-world significance. On the other hand, the engineering approach faces the simple reality that organisms are far from machines [48, 49]. "Biology is messy" is a term you will often hear in the lab when an experiment does not produce the expected results. While this may be simply the reflection that one's experiments were "messy" (!), it also reveals the context-sensitivity and unpredictability of biology that separate it from engineering disciplines. Francois Jacob, one of the fathers of molecular biology, famously called evolution "a tinkerer" [50] to describe how it has shaped and continues to shape living organisms. Biological networks were forged by evolution through the random repurposing of pre-existing forms [7]. They are not, like electronic networks, designed by an engineer. Hence the engineering approach is bound to remain a metaphor.

It is essential to connect models of gene circuits with experimental truth in both top-down and bottom-up approaches. Both synthetic and systems biology benefit from experiments that characterize the relationships between the molecules that compose genetic network. These

experiments are particularly valuable to those seeking to understand the genetic program when they include the spatial dimension, temporal dimension and place the cell in different contexts [51].

1.3 The dynamic phenotype

To uncover generalizable principles about the function of the cell, it is necessary to consider the environment in which cells evolve. That environment is not the artificial, nutrient-plenty one offered by conventional *in vitro* molecular biology assays. Take the life of an ordinary *E. coli*. We initially stumble upon the bacillus growing slowly in the environment. Suddenly, it is ingested by a mammal and undergoes a series of brutal changes. Its temperature rises dramatically from 20°C to 37°C. Oxygen levels drop. Acidity rises strongly as it reaches the mammal's stomach, where it is assaulted by an army of enzymes. As it progresses down the intestinal tract, pH rises again, and it finds itself dividing at a much faster rate and competing for nutrients with other species of bacteria. Finally, it finds itself out in the open again and is exposed to oxygen, lower temperatures and the absence of nutrients [51]. To survive this series of unfortunate events, *E. coli* displays a dynamic phenotype and uses many layers of regulatory feedback to assess and precisely react to its environment.

Dynamic phenotypes are pervasive throughout biology and govern many important biological processes such as cellular differentiation [52, 53], stress response [54, 55] and epigenetic regulation [56]. An increasing body of evidence points to dynamics as a key strategy that biological systems use to encode information [57] and dysregulation of dynamics as a hallmark of disease [58, 59] and aging [60]. Thus characterizing the dynamic nature of changes that occur upon a cell's exposure to stimuli at the omic level is necessary to understand the mechanisms of adaptation that pervade through life [61].

1.4 Capturing gene expression dynamics

The importance of complex feedback in controlling the real-time response to external stimuli has created a need for a new generation of cell-based technologies that enable both the collection and analysis of long time, high-throughput temporal data [62]. This is the main motivation for the work I accomplished in graduate school and present in this thesis. These novel technologies will map out with greater certainty the expression dynamics that genes display after reacting to a stimuli [63, 64]. And will thus allow us to differentiate transient from sustained expression patterns, fast from slow response kinetics and drive towards an understanding of the causal relationships between genes that belong to the same networks [65].

Despite the importance of capturing time-series data, measurement techniques used to study biological dynamics suffer from a variety of drawbacks [14] (see Table 1.1). High-throughput, omics techniques (including RNA-seq [66, 67] , ribosome profiling [68], mass spectrometry [69], and microarray-based expression profiling [70, 71]) are destructive in nature and only offer snapshots of a cell's state [14]. Other techniques such as qPCR [72] , and flow cytometry [73] also require multiple experiments to capture more than one timepoint and track more than a handful of genes. Fluorimetric plate reader and batch culture experiments, while better positioned to capture time series data, are also limited in the number of genes they can track. In addition, these techniques rarely give scientists full control over the type and duration of perturbation that can be applied to biological signaling networks. Finally, microfluidic technologies, especially when combined with fluorescence microscopy, offer a method to study cell phenotype at a high resolution within a chemostatic environment; however, these approaches are notoriously low-throughput and limited to very few genes-of-interest.

Our ability to measure large-scale dynamics has not kept pace with our rapidly growing ability to take information-rich single-time point measurements. This paucity of data describing biological signaling dynamics helps explain why bridging network structure with function remains

Table 1.1: Methods to assay gene expression dynamics. Adapted from [14]

Method	Time course	Scope	Throughput	Live vs. Dead	Ref.
Flow cytometry	Snapshot	Single cells	Medium	Either	[73]
RT - qPCR	Snapshot	Population	Medium	Destructive	[72]
RNA-seq	Snapshot	Population	High	Destructive	[74]
Microarray	Snapshot	Population	High	Destructive	[75]
Fluorescence Microscopy	Dynamic	Single cells	Low	Live	[76]
Fermenter	Dynamic	Population	Low	Live	[77]
Plate reader	Dynamic	Population	Medium	Live	[78]
Dynomics	Dynamic	Population	High	Live	[79]

a challenge despite the accumulation of functional genomics data [80–82] and the progress of computational methods and analysis [62, 83]. To bridge this gap, I develop in this thesis a broadly applicable research platform (Dynomics) combining large-scale microfluidic technology with fluorescence microscopy. This platform will capture high-quality time series data simultaneously for thousands of signaling network constituents in live cells [84, 85].

1.5 Microfluidics and time-lapse fluorescence microscopy for quantitative biology

In the past twenty years, microfluidics has emerged as a powerful tool in biology dynamics. Microfluidic platforms enable scientists to precisely control cells' environment and capture their behavior using high-resolution microscopy [86]. The micro-scale features of microfluidics offer control over cell location and growth [87] while the laminar flow conditions enable precise delivery of media [88, 89]. Finally, the small scale of those devices minimizes resource use : reagents, consumables, and time [90]. As a result, microfluidic devices have been engineered to serve a multitude of research purposes in both synthetic and systems biology [91]. These include studying single-cell gene expression dynamics, tissue growth, biofilms, the effects of

temperature and chemical gradients or accomplishing cell sorting, on-chip DNA amplification and sequencing [92–105].

However, the use of microfluidics has principally been low-throughput, allowing users to track the behavior of a few nodes to answer narrow biological questions [106–108]. While individual devices have been developed for loading, culturing, stimulating or tracking cells, few systems attempt to bring together these capabilities necessary for the study of gene expression dynamics at the ”omics” scale. To further increase the throughput of microfluidics, novel parallelized platforms with automated data extraction pipelines are required [109].

1.6 High-throughput microfluidics and the Dynomics platform

Recently, several studies have demonstrated how microfluidic parallelization and automation of cell handling permit the simultaneous tracking of hundreds to thousands of strains in parallel. These studies represented significant steps towards increasing the upper bound of throughput in microfluidics-based experiments. Common to those studies is the use of fluorescent libraries for model organisms such as *S. cerevisiae* [110,111], *E. coli* [112,113] or mammalian cell lines [114]. In these existing libraries, the signals associated with the production of a fluorescent protein are measured for each strain to monitor gene expression and protein localization. This combination of microfluidics and genome-scale fluorescent reporter strain libraries has facilitated the study of genomic transcriptional dynamics.

These approaches, however, have been hampered by the “Pareto Frontier”‘ of microfluidics: the inevitable trade-off between device complexity and ease-of-use [115]. This resulted in short experimental lifespans, limited temporal resolution, static environmental conditions, and the development of single-purpose devices. For the few high-throughput microfluidic devices able to handle entire libraries, experimental protocols require complicated fluidic connections and valves that can take hours to set up [88, 110, 113]. Such devices are also not compatible with the high

temporal resolution needed to study signaling networks. For example, in the oxidative response network in *S. cerevisiae*, the time scale of protein expression change is on the order of minutes, yet current large-scale screening devices image with a period of over 20 minutes [110]. Finally, these platforms are not all capable of finely and continually controlling environmental conditions, despite such control being one of the most salient advantages of microfluidics [90]. All the above prevent the democratization of potentially transformative technology and its application towards a systems-level, fundamental understanding of the rules of signaling pathway dynamics [111].

1.7 Thesis overview

Dynomics is a broadly applicable research platform that bridges the gap between microfluidic dynamic perturbation experiments and genome-wide assays.

The Dynomics microfluidic platform enables continuous growth, precise environmental control, and simultaneous optical monitoring of up to 2176 unique microcolonies of *E. coli* or *S. cerevisiae* for up to 14 days. The platform meshes hardware (a large-scale microfluidic device and associated custom optical enclosure) and software (an automated data extraction pipeline and downstream machine-learning algorithms) to achieve this aim. In the Dynomics microfluidic device, all strains grow in physically separate microchambers but are exposed to the same dynamically tunable media. This unique setup allows for tracking genetic markers such as the expression levels of fluorescent protein over time for each strain. By loading the device with some of the readily available genome-scale libraries of fluorescent transcriptional reporters, Dynomics can be used to resolve genome-scale transcriptional dynamics.

Chapter 2 of this thesis details the design and development of the Dynomics microfluidic platform and its use to generate data from fluorescent *E. coli* subjected to dynamical environmental perturbations. Chapter 3 demonstrates the use of Dynomics as a versatile biosensing platform capable of detecting time-varying concentrations of toxins using a library of fluorescent *E. coli* and

machine learning algorithms. In this work, 2000 fluorescent *E. coli* reporter strains were screened against nine heavy metals at levels close to or below EPA drinking water limits, identifying naturally responsive genes and highlighting their transcriptional dynamics. Machine learning techniques were used to capture the most salient and discernible patterns in the genome-scale transcriptional response to each heavy metal. Chapter 4 further demonstrates the applicability of Dynomics as a biosensing platform by detecting pollutants in challenging environment of seawater. To this end, I showed that different types of genetically engineered *E. coli* can be used in a microfluidic environment to sense chemical inputs in seawater continuously. Finally, in Chapter 5, I leverage Dynomics to monitor the time-resolved response of the *S. Cerevisiae* proteome to the drug Metformin, illustrating the platform's use for dynamics-based chemogenetic screens and the elucidation of drug targets.

In the broad context of biological research, the significance of this research is the development and application of a technology that can reveal insights embedded in temporal responses of microbial biological networks and elucidate gene regulatory phenomena that are fundamentally dynamical in nature.

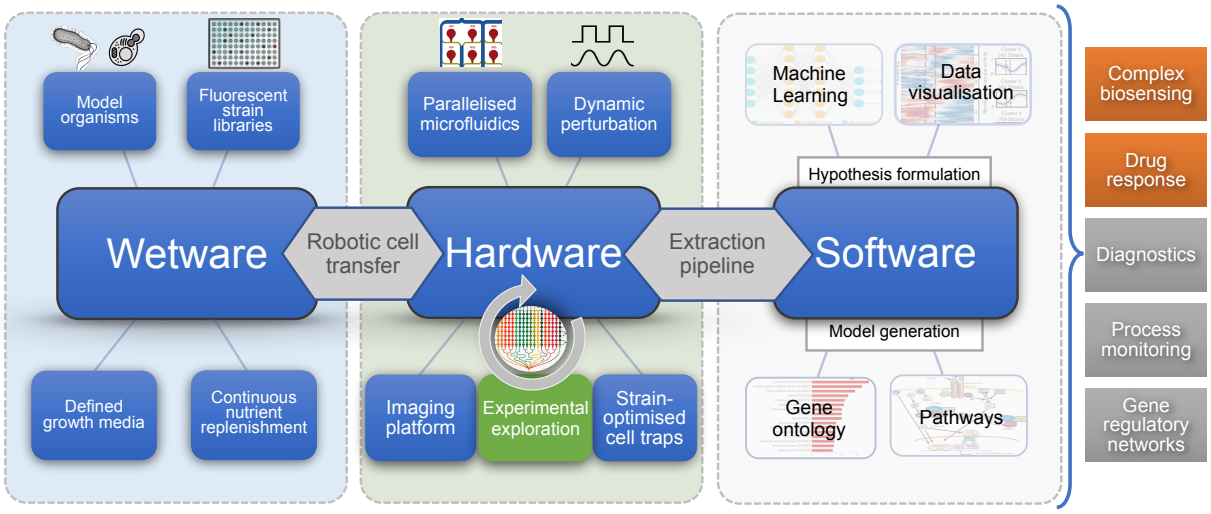

Figure 1.1: The Dynomics research platform combines wetware tools facilitating the handling of libraries of fluorescent strains, hardware tools centered around high-throughput microfluidics and fluorescence microscopy, and a suite of associated dynamic data-oriented software tools. On the right are some of the potential applications of Dynomics, with the ones explored in this thesis highlighted in orange

Chapter 2

Development of a microfluidic platform for high-throughput and continuous culturing of fluorescent *E. coli*

2.1 Introduction

In this chapter, I describe the design, development and validation of a microfluidic chemostat array capable of continuously culturing and monitoring up to 2,176 unique strains of *E. coli*. Coupled to time-lapse fluorescence microscopy via a custom optical system, this high-throughput multiplexed microfluidic device is designed to bridge omics-scale throughput with a precisely controlled dynamic environment, resulting in high-quality time-series data.

To obtain a functional, highly parallelized microfluidic platform, the geometry of the device was optimized for cell loading, continuous cell growth, and the collection of fluorescence responses upon exposure of cells to experimentally-induced environmental stimuli.

2.2 Microfluidic device design

2.2.1 Design endpoints

The design goal was to produce a 62 mm x 88 mm microfluidic device that allows for high-throughput, long-term (a week or more) culturing of up to two thousand microbial microcolonies. The device was designed to be housed in an optical enclosure that includes the hardware and software to capture transmitted light and GFP fluorescence images at high temporal resolution with a 70 mm x 70 mm field of view. To improve experimental tractability and dissemination of the platform, the device was conceived with a single layer design that requires minimal fluidic connections to the outside and contains no valves. In addition, the device was designed to be compatible with high-throughput cell loading using a Singer ROTOR colony-picking robot. This added the following design requirements: a cell trap region at least 400 μm in diameter where each initial cell spot would be placed and a 1.125 mm spacing between each cell trap in both the x and the y dimension to match the 6144-format of the Society for Biomolecular Sciences (SBS).

2.2.2 Cell trap design

Development

The Dynomics microfluidic device cell traps are the fruit of an iterated design process and were optimized for reliable trap filling, cell retention, and strength of the fluorescent signal obtained from each spotted microcolony. The goal was also to minimize the variability between any two positions on the device. Finally, we sought to minimize excess cell mass streaming out of each cell trap which leads to clogging of media channels and shortens experimental lifetime.

Figure 2.1a shows the 'gill' cell traps initially designed for this purpose. Gill traps had a high aspect ratio which resulted in a high fluorescence signal. However, we noticed convective flow into the reservoir that entered the first gill and left through the last gill leading to poor

cell retention. The Gill-Shunt design (Figure 2.1b) was designed to reduce convective flow in the gill traps by providing a path of lesser resistance for flow to follow before cell traps are filled. Although this design helped redirect flow, cell retention was still variable across the gills. The cell traps in Figure 2.1c were designed to increase retention by constraining the cells in a "mushroom" shaped cell trap. Although this solved the cell retention issue, the growth state of cells in the mushrooms was variable, leading to inconsistent responses across mushrooms. Further investigation revealed that the variability was often due to either the rapid emptying of cells from a mushroom due to fluid flow instabilities, or to heterogeneous cell compaction and growth states within each mushroom. Figure 2.1d 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 2.1e shows a design similar to 2.1d with an added spotting post to limit biomass while providing a support pillar for the spotting process. This design also resulted in excessive biomass and device clogging. Figure 2.1f 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 2.1g-j). Figure 2.1g has a 4 μm spotting region and 50 μm funneled cone. Figure 2.1h 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 2.1g. The narrow funneled cone prevented clogging and large biomass, but the 50 μm height still resulted in variability in the growth state. Similarly, Figure 2.1i 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 2.1j has a 4 μm spotting region and 4 μm cone. This design limited biofilm formation, decreased growth state variability, and had sufficient fluorescence signal. This trap design formed the basis for the final iteration of our cell trap design. It was successfully adapted to work with the manifold channel design (Figure 2.1k), successfully retaining cells in a homogeneous growth state with detectable fluorescence signal.

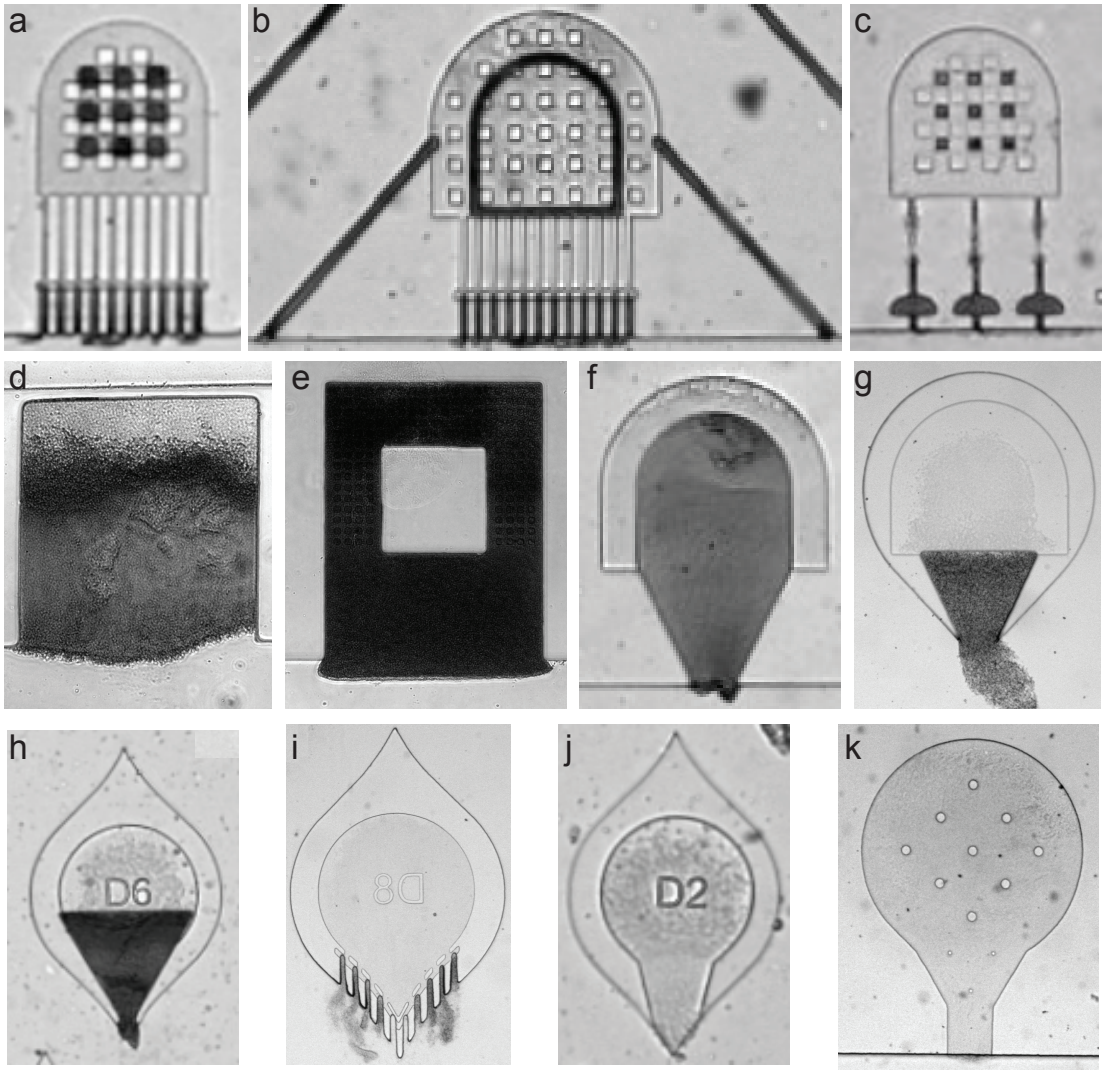

Figure 2.1: *E. coli* growing in different cell traps designed for multi-strain microfluidic devices. Cell trap designs tested to maximize *E. coli* cell retention, homogeneous growth states, and fluorescence signal. Gill traps with (a) and without (b) shunts inconsistently filled traps. Dead end gills (c) and side traps (d,e) lead to heterogeneous growth states. Channel-free traps (f-j) were tested with 4 μm (f,j) and 50 μm (g-i) openings with various opening sizes for cell retention. k) The final trap design used with the manifold channel system

Final cell trap design

Dynomics device cell traps possess a spotting area large enough to be seeded with a microcolony grown on an agar plate at 6144 density. The trap has a "light bulb" shape, with the area of the trap furthest away from the minor channel possessing a disk-shaped geometry with a large 500 μm diameter to receive the microcolony transferred from the 6144-density agar plate

(Figure 2.15). As cells grow out of the spotting area, the trap geometry narrows and eventually forms a rectangular area of 100 μm by 110 μm (named the bulb).

While previous designs in the Hasty lab used "monolayer" (1 μm deep [116]) or "pseudo-monolayer" (1.7 μm deep [26]) cell traps for the culturing of single sheets of *E. coli*, the Dynomics traps are 4 μm tall. This increases the number of cells in the imaged area and hence the signal-to-noise ratio. Increased depth of the trap is also necessary to contain the large amount of cells deposited into the trap during a spotting event. In our final design (Fig. 2.1k), the minor channel intersects the cell trap at a right angle, which allows for the rapid clearance of cells growing out of the cell trap. Despite cell trap and minor channel being orthogonal, the height difference between them shields the bacteria from shear stress. Indeed, in microfluidic devices with low aspect ratio channels, the resistance of a channel scales with the cube of the height [90]. Given that the cell trap has a 4 μm height, the fluid velocity into the cell trap is only a fraction of the one in the adjacent minor channel, which is 50 μm tall and experiences 200 μm/s flow on average. Although flow of media into the trap is minimal, the small scale of the bulb area allows the rapid inward diffusion of nutrients and outwards diffusion of waste between the channel and cell trap relative to the timescale of bacterial growth. Consequently, cells in the part of the trap that is closest to the fresh media supplied by the channel grow exponentially and form a homogeneous layer of cells. Cells further away from the channel and its flow of fresh nutrients remain in stationary or decline phase. While we image all the areas of the trap (numbered 1 through 4 in Figure 2.8 A), we have found the data collected from the front of the bulb (labelled "bulb 1") consistently produced data with the highest signal to noise ratio.

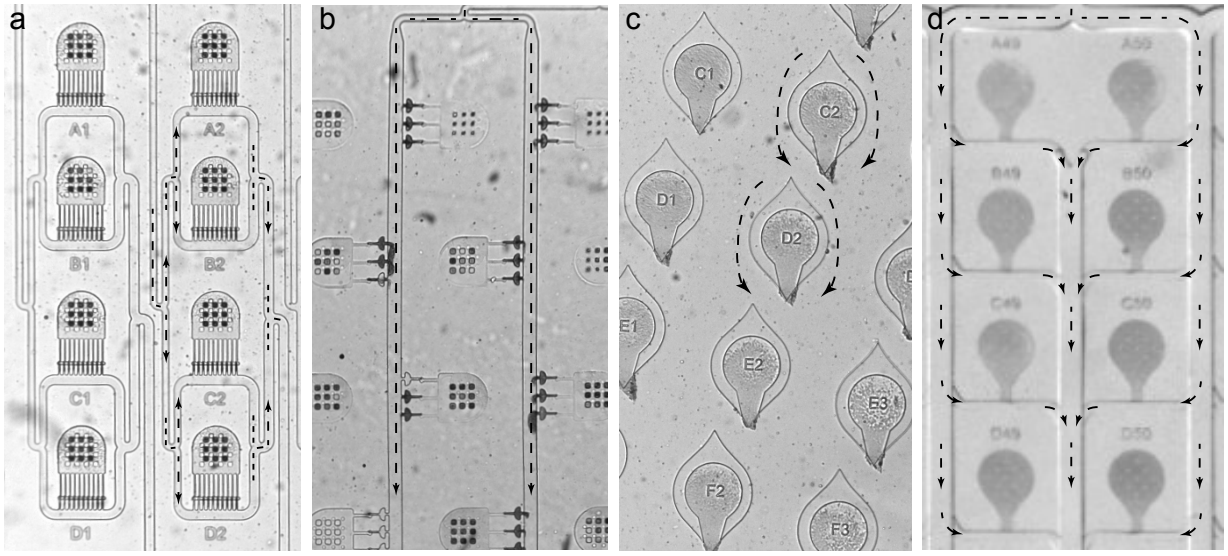

Figure 2.2: Different channel systems tested during the development of a multi-strain microfluidic device constrained to 1.125 mm spacing between traps for compatibility with 6144-SBS format. a) Binary splitting channel system traditionally used in microfluidics. b) Shared channel system. c) Channel-free system. d) Manifold channel system which was selected for used in the 2,176-strain device. Black arrows denote flow direction.

2.2.3 Channel system design

Development

Adequate channel design complements cell trap design to limit flow rate variability across the device and to prevent clog formation that limits device experimental lifespan. Channel design was also the fruit of a design-build-test-learn process, which is described below.

Figure 2.2a illustrates a binary splitting channel system design often seen in microfluidic devices. While it has the advantage of offering symmetrical splitting of flow channels leading to identical flow rates across the device, this channel system requires large amounts of space and is not scalable to the level of a two-thousand strain device. In addition, it increases the number of zones with low fluid velocity in the device. A critical factor in the prevention of clog formation in microfluidic devices is the maintenance of a constant shear flow to remove excess cells from accumulating along the media channel walls. Regions with low flow rates

are potential areas where *E. coli* cells can settle, proliferate, and begin to form a biofilm. The shared channel system (Figure 2.2b) is designed with of a single inlet channel splitting into multiple channels of equal length that feed cell trap positioned in series. This design helps with even flow rate across the device and limits clogging. However, it also increases the chances of cross-contamination of strains due to convective flow into the traps and multiple cell traps sharing the same inlet channel. To eliminate this, we also tested an open channel design (Figure 2.2c) where flow comes in from the top of the traps, with no channels present other than inlet and outlet plumbing (Figure 9c). The traps were rotated 15° to avoid flow paths from the mouth of one cell trap from feeding into another, thus avoiding cross-contamination, and maximizing nutrient availability. 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.

2.2.4 Media flow considerations

The final channel design is a manifold system depicted in Figure 2.2d. It consists of an inlet channel that splits into major media channels with a height of 230 μms. Major channels, in turn, split into parallel minor channels 50 μms tall. Finally, each minor channel feeds a unique cell trap described in 2.2.2. During outflow, minor channels combine again into major outlet media channels. Of all the channel systems tested, the manifold channel system is the only one to meet the design endpoints, allowing consistent replenishing of media at each cell trap and limiting cross-contamination and variability in flow rates across the device. It also enables 1.125 mm spacing between cell traps, a constraint tied to our robotic loading method.

Flow rate

To further characterize the Dynomics device and validate our design choices, we sought to measure the average flow rate across the device. To this effect, we used fluorescent micro bead flow analysis, a technique previously used in the Hasty lab to characterize microfluidic devices [116]. Fluorescent beads of 0.5 μm flowing in the channels were imaged upon 200ms exposure to fluorescent light. The resulting bead trace lengths were averaged to produce an estimate of flow velocity in the channel at the level of the cell trap. We averaged at least 100 bead traces for 30 positions spread out over the device to produce Figure 2.3. This bar plot shows that some variation exists in the flow rates across the device. This is partly due to the manifold geometry of the chip and the small imperfections that can arise in the microfluidic wafer microfabrication process. Overall, variation in flow is limited given the scale of the device (RSD $<30\%$) and did not seem to affect cell response to stimuli (see 2.2.5). We note here that further mitigation of flow rate variability could be done via the use of COMSOL modeling or fluid flow circuit analysis.

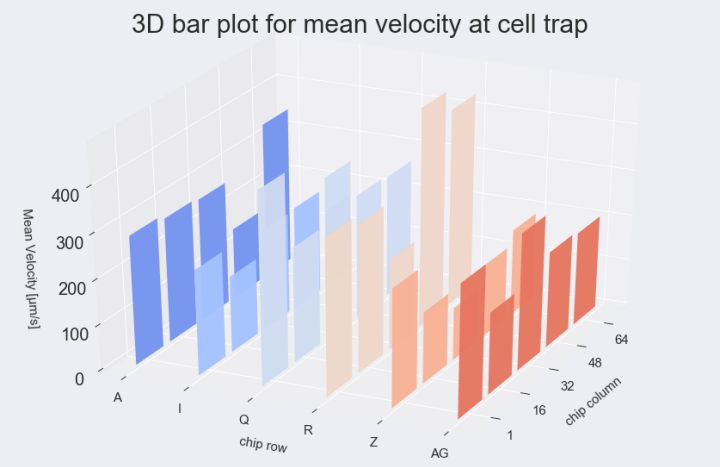

Figure 2.3: Overview of flow rate across 2176-strain device. Average flow rate for 30 cell trap positions spread across 6 columns and 5 rows of the Dynomics chip. Some variability exists between positions, with cell traps towards the center of the chip experiencing slower flow rates. Further characterization of the impact of this variability in flow rates showed that impact on cell growth was limited.

Pump-driven flow

In the initial iterations of our device, we used small-scale peristaltic pumps to drive flow through our large-scale microfluidic devices. While we eventually reverted to using the true and tried method of hydraulic head difference-driven flow which provides more stable flow rates, we solved an interesting problem tied to the use of peristaltic pumps with microfluidic devices along the way which is presented below.

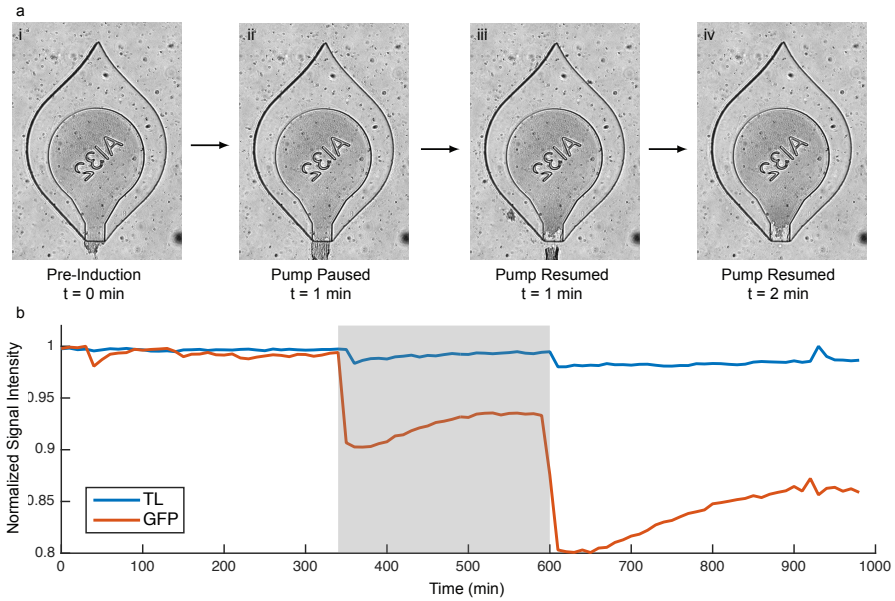

Figure 2.4: Cell washout caused by pump-driven pressure fluctuations. a) i. Cells grow to fill the trap. ii. When pumps are briefly paused for an induction, the pressure change causes cells to surge out of the traps. iii-iv. When pumps resume normal operation, cells break off and exit the trap, affecting the density of cells in the mouth of the cone where data is extracted. b) Representative trace from a trap with washout. Pumps are paused at the beginning and end of the induction (gray bar). Washout causes a significant drop in fluorescence and a weak induction response.

When using a peristaltic pump to drive flow on a microfluidic device, a user is required to pause the pump to introduce novel experimental conditions (e.g. introducing an inducer) to the cells growing on the chip. This is necessary to prevent air bubbles from entering the chip and obscuring flow. However, doing so causes pressure on the chip to drop during the pause and yo rapidly increase when pumps resume motion. This pressure fluctuation causes cells to surge out of the traps as seen in Figure 2.4a. Additional pressure waves are caused by the peristaltic

pumps successively pinching and releasing tubing to deliver media. Cell washout causes a sharp change in signal at the beginning and end of each induction as seen in Figure 2.4b. To prevent such deviations in pressure, we implemented a passive flow stabilization technique [117], whose effects are akin to a capacitor. Our hydraulic capacitor takes advantage of air's higher compliance relative to the media flown into the chip. The operational setup involves a syringe (5mL, Becton Dickinson) filled with a predetermined ratio of air to media and fitted in-line between the pump and chip (Figure 2.3a.). When a pressure wave travels to the air/media interface in the closed volume of the syringe, the air absorbs the pressure wave. Due to its greater compressibility, the air will restitute a pressure variation more slowly and evenly. This setup both smoothes out the pressure drop due to pause-pumps events, and dampens the naturally pulsatile outflow from the peristaltic pumps. We tested the capacitor on Dynomics devices imaged at 4x magnification on an Olympus IX81 microscope and on the Dynomics optical enclosures (see Fig. 2.14). These experiments showed improved filling of the cell traps and reduced washout and clogging (Figure 2.5b.). When comparing transmittance traces between experiments with and without capacitors, we observe higher overall stability in capacitor runs (Figure 2.5c.).

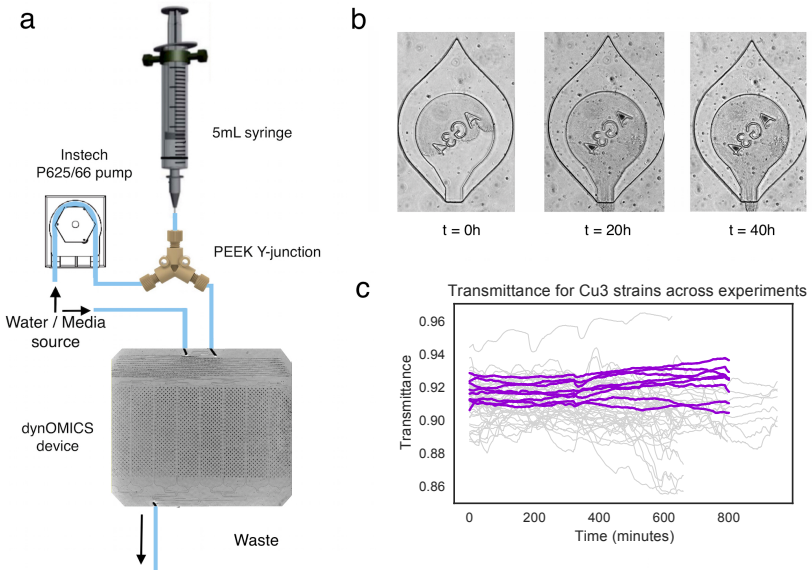

Figure 2.5: Use of an in-line capacitor with a Dynomics microfluidic device. a) Simplified diagram of the setup of our in-line capacitor b) Improved filling of cell traps when capacitor is implemented c) Transmittance traces for an *E. coli* growing in our device strain across five experiments without capacitors (gray) and one with a capacitor (purple).

2.2.5 Final design

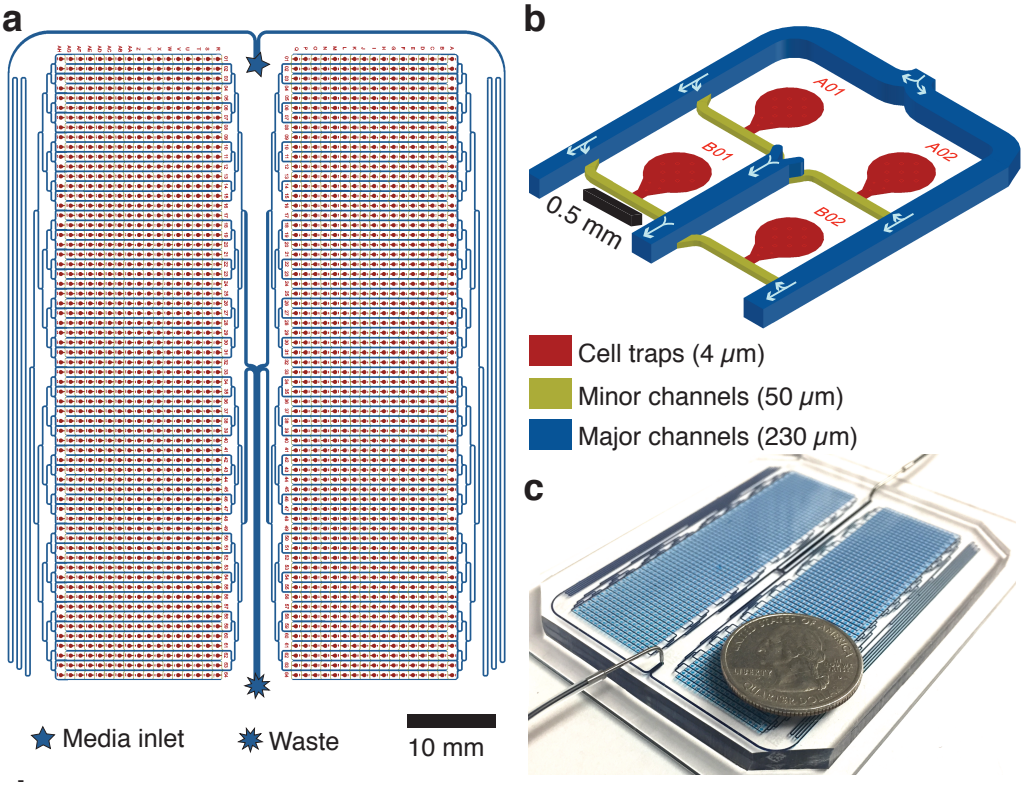

Figure 2.6: The Dynomics microfluidic device. a) Design of the Dynomics 2,176-strain microfluidic device with cell traps in red and media channels in blue and yellow. b) Detailed schematic of four strain banks with arrows showing direction of media flow.

The microfluidic design of Dynomics is simple in nature, consisting of a single-layer with a single media inlet and outlet, greatly reducing the complexity typically associated with multi-strain devices [110, 111, 113]. The single media inlet feeds two halves of the device, each containing 17 rows and 64 columns of cell traps, for a total of 2,176 cell traps (Fig. 2.6a). Cell traps are 4 μm tall and receive fresh media from 230μm tall major channels that split off into 50 μm tall minor channels. Waste from each cell trap flows into a common major channel (Fig. 2.6b). This device geometry ensures each cell chamber region is fluidically isolated to prevent strain cross-contamination. Cells grown on agar plates are transferred to the PDMS device, with a unique microcolony of isogenic *E. coli* directly spotted into the center of each cell trap

before bonding of the device to a glass slide. Spacing between the reservoir regions is 1.125 mm for compatibility with standard SBS-format 6144-density spacing. Once the device is bonded using plasma bonding, growth media is flown onto the device via gravity-driven flow induced by hydraulic-head difference between inlet and outlet syringes connected to the microfluidic device. Spotted cells grow to fill each trap within 24 hours.

The Dynomics enclosure shown in Fig 2.14 serves as a stand-alone microscope for imaging the microfluidic device. The optical enclosure images in both bright-field and fluorescence channels with a 36 μm optical resolution, comparable to a lab-grade microscope at 4x magnification. With a nominal imaging frequency of 2 minutes, Dynomics allows us to adjust environmental conditions and measure the resulting changes in gene expression at a high temporal resolution.

Characterization of cell growth in Dynomics

To further characterize the cell state of *E. coli* growing on our device, we took higher magnification time-series images of the bulb1/bulb2 areas (see Fig.2.8) using a Olympus IX81 inverted microscope at 6X magnification. While precise growth rate can be estimated in monolayer traps by tracking single cells, the size and 4μm depth of Dynomics cell traps does not allow for single cell tracking [118]. Hence we obtained a proxy for growth rate by approximating "time-to-exit" of cells: the amount of time cells take to grow from the boundary of non-growing cells to the mouth of the trap where they are washed away. Time-series images were used to produce kymographs of time-to-exit, such as the one depicted in Figure 2.7. We recorded an average time-to-exit of 18h, corresponding to over 50 doublings of *E. coli* growing in exponential phase. This result, helps explain the long relaxation time observed post-induction in Figures 2.8 and 2.11. Indeed, the fast-folding GFPmut2 produced by the strains loaded to the device is a stable protein. Most proteins in *E. coli* are not actively degraded during exponential growth [119]. Therefore dilution by cell growth governs the off kinetics of the observed GFP signal, which is dictated by cells dividing and growing out of the trap.

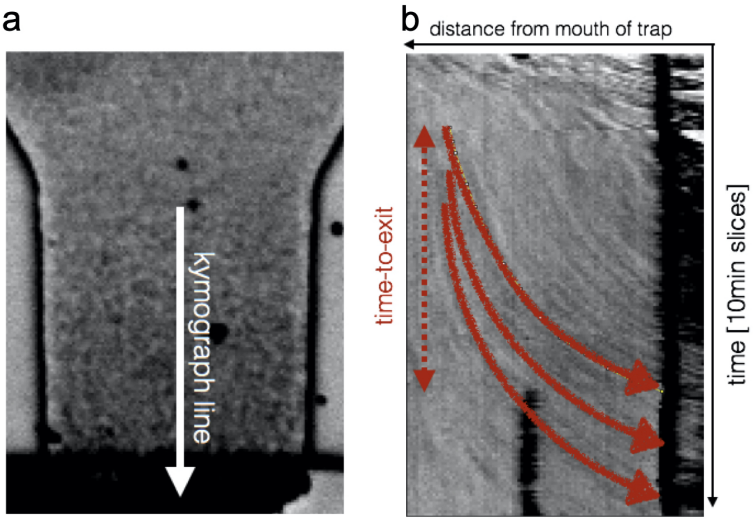

Figure 2.7: *E. coli* growing in Dynomics traps. Higher magnification images are taken on a research-grade microscope at 6X to investigate the long relaxation time of the GFP signal following an induction response. Left: image of a dynomics cell trap from which the kymograph was extracted. Right: 2D kymograph from a cell trap with estimated cell exit trajectories in red ranging from 16 to 20 hours. Kymographs are x vs. t scans, where the intensity along a given line is plotted for all images of a time-series

Spatial dynamics and depth-dependent cell phenotype

Data collected from Dynomics experiments is extracted from the area of the cell trap containing well-fed and actively growing cells (labelled ”bulb 1”, see Fig.2.8A). Cells further up the trap experience diffusion-limited access to nutrients [120]. Plotting gene response across the difference bulb areas shows the depth-dependent response of strains from the *E. coli* GFP-promoter library exposed to heavy metals and carbon source shifts and is shown in Fig. 2.8. We observe that areas further up the cell trap exhibit a stronger change in GFP expression when the stimuli is a changing carbon source rather than exposure to elevated levels of heavy metals. This may hint at stronger depth-dependency of transcriptional response for genes directly tied to carbon metabolism in our device. We note here that the fast-growing cells present in bulb 1 (the area from which data is extracted) are not physically separated from the stationary phase cells present deeper in the trap (bulb 2,3 & 4 areas). Hence bulb 1 area is not entirely shielded from phenotypically

distinct cells that exist further up the trap. This source of experimental noise will likely be added to the intrinsic noise that naturally arises in gene expression [121] and metabolomic profiles [122] across clonal populations growing in strictly identical conditions. Work from the Hasty lab and other groups [123, 124] has further investigated this source of extrinsic noise in microfluidic devices and show that phenotypic differences emerge when the colony consumes nutrients at a faster rate than the diffusion-based replenishment of those nutrients from the channel, thereby creating a metabolite gradient across the cell pack. To reduce the effect of metabolite gradients in our cell traps, we increased the concentration of Glucose present in the minimal media used for Dynomics experiments from 0.4% to 2% (see Table 3.3) and limited our analysis to the data extracted from the "bulb 1" area.

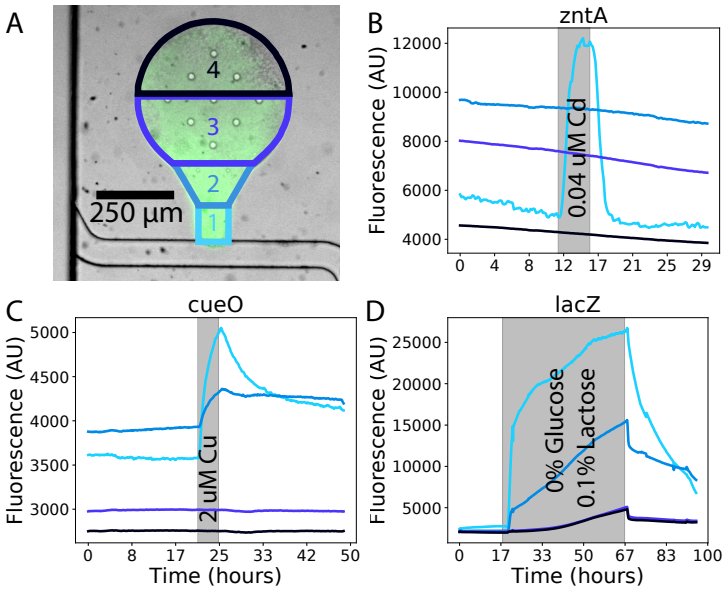

Figure 2.8: Spatial dynamics in Dynomics: A) A Dynomics cell trap loaded with fluorescent *E. coli*, with labeled regions representing the different depths of fluorescence time-series extraction. B) The *zntA* promoter strain from the Alon library [112] only shows a response to Cd(II) at the front of the cell trap region. C) The *cueO* promoter strain responds to Cu(III) in the first two extraction regions. D) The *lacZ* promoter strain response throughout the entire trap when the carbon source is changed from glucose to lactose, then back to glucose.

Reproducibility

Given the novelty of using robotic spotting to load multiple strains onto a microfluidic device to capture gene expression dynamics, we sought to characterize the variability of the data collected both within and across Dynomics experiments.

To that effect, we initially loaded the device with a set of seven strains of the Uri Alon *E. coli* GFP-promoter library [112] using over 300 replicates for each of the seven strains. The GFP-promoter library contains approximately 2,000 unique *E. coli* strains in which 50-150 bp "promoter regions" were inserted into low copy (pSC101) plasmids upstream of a strong ribosome binding site and GFP , each strain drives the production of a fluorescent protein using a unique promoter from the *E. coli* genome [112]. Previous work indicated that this library can serve as an accurate tool to measure promoter activities [112, 125]

The seven strains were chosen for their relevance in *E. coli* metabolism or ability to be selectively activated via external inducers. We chose three strains containing the promoter region for sigma factors (*rpoS*, *rpoD* and *rpoH*) and two other strains with previously established promoter responsiveness to environmental stressors (*cueO*, a copper-responsive gene, and *zntA* a zinc- and cadmium-responsive gene). We also selected two negative control strains whose expression was expected to remain unchanged across conditions (*fimD* and *U139*). To test our device, we monitored the response of the seven strains to cell state stressors. We call these step-function perturbations "inductions", corresponding to a period of four hours during which the media being flown into the device is switched to a media source containing the stressor. After a 4h induction, the cells are given sufficient time for the change in GFP expression to decay back to the uninduced steady state before a new induction is performed (approximately 20h). An induction with 0.5 μm Cd(II) is plotted for an experiment in which 2176 replicates of the *zntA* strain were loaded in Dynomics in Figure 2.11. In performing these inductions, we noted good agreement in the overall shape of induction response (including relaxation time) both across strains and across replicate inductions.

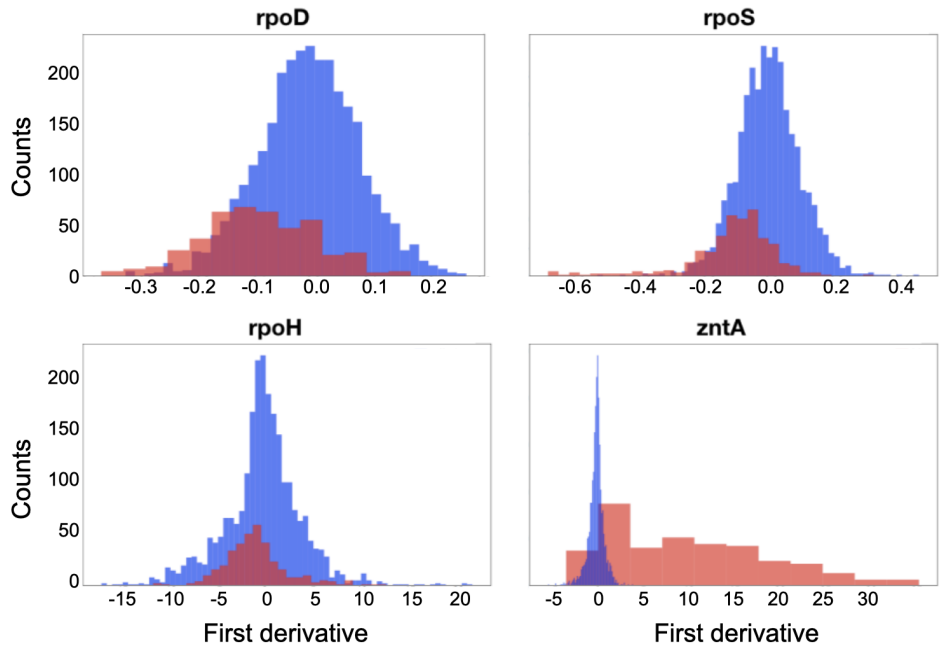

Figure 2.9: Distribution of first derivatives in baseline vs induced state for control strains loaded on Dynomics. Finite difference of background subtracted, smoothed and normalized GFP data is plotted for 4 control strains. The baseline (uninduced) state is plotted in blue while the induced state, corresponding to exposure to 0.04µm Cd(II) is plotted in red. *rpoD* is the exponential growth σ factor, *rpoS* the stationary phase and general stress response σ factor and *rpoH* the heat shock σ factor. *zntA* is a zinc/cadmium/lead-transporting P-type ATPase specifically upregulated when exposure to Cd(II).

Figure 2.9 shows the first derivative value of the seven control over five separate experiments during which the control strains were exposed to the same Cd(II) induction. It enables a further assessment of the variability of gene expression data collected across Dynomics experiments. In Fig 2.9, strains that do not specifically respond to Cd(II), such as *rpoH* (the heat shock σ factor), *rpoD* (the exponential growth σ factor) and *rpoS* (the stationary phase σ factor) display only small deviations from their basal state (blue histogram) when Cd(II) is introduced in the media (red histogram). This contrasts with the strong bimodality of the Cadmium-responsive promoter *zntA*, whose induced state exhibits much higher first derivative features. We note here that the leftward shift of the induced state histograms for *rpoD* and *rpoS* may be indicative of some change in growth rate or metabolic state induced by the exposure to Cd(II). It is well

documented that in *E. coli*, exposure to sources of stress affects metabolism, growth rate, and gene expression in a global, non-specific fashion [126] as those properties are interrelated [127].

Having extensively characterized the Dynomics device and the initial data produced by step function inductions, we progressed to performing experiments with the entire *E. coli* GFP-promoter library [112] which we report in Chapter 3. We also developed an induction protocol that enabled repeated inductions to be carried out over the course of many days using Dynomics as shown in Figure 2.10. Finally, we demonstrated our ability to perform periodic induction scenarios. Periodic inductions combined with systems-engineering approaches have been used to study the properties of signal transduction cascades [128], and are one of the salient advantages of microfluidic platforms such as Dynomics. Fig 2.10 demonstrates frequency-dependent changes in gene expression upon periodic 0.04 μM Cd(II) inductions of increasing period.

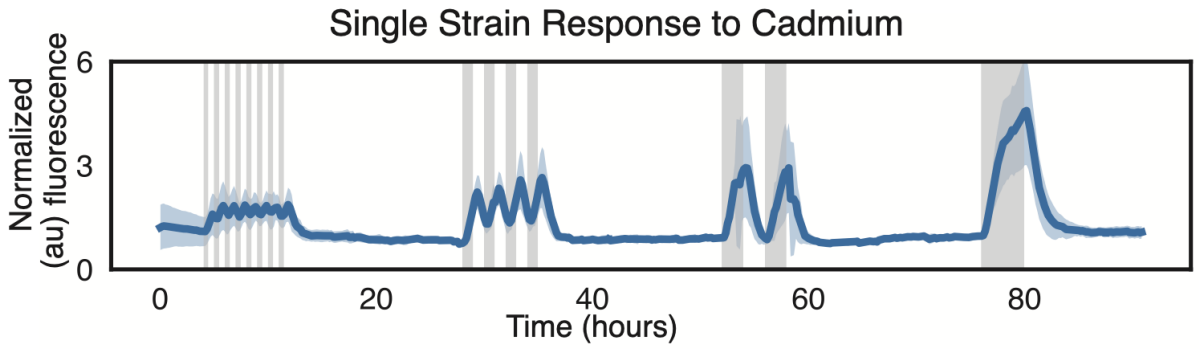

Figure 2.10: Response of *zntA* to frequency-encoded signals on Dynomics. Mean fluorescence (solid blue) and SD (shaded blue) of the *E. coli* zntA promoter driving GFP to repeated cadmium inductions (gray bars) with periods increasing from left to right (30 min, 2 h, 4 h, and 8 h)

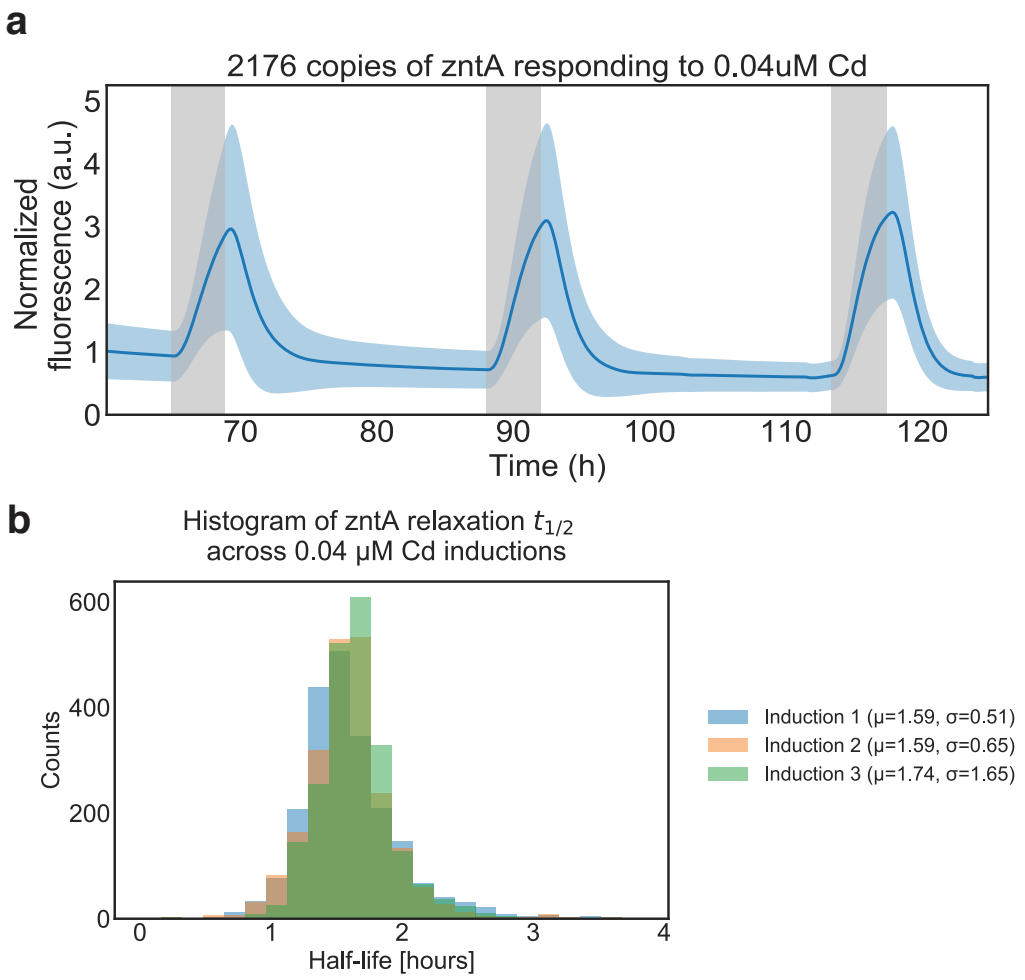

Figure 2.11: Device validation with a Dynomics chip with 2,176 replicates of the *zntA* strain from the *E. coli* GFP-promoter library [112]. a) Mean (dark blue) plus/minus one standard deviation (light blue) for the 2,176 strain responding to three consecutive four-hour 0.04 µM Cadmium inductions. b) Histograms of the decay half-life of post-induction responses (i.e the time taken for a strain's fluorescence levels to decay back down to its half-peak value) are overlaid for the three inductions plotted in a).

2.2.6 HD biopixel design

While the design presented in 2.2.5 is a functional platform for the culturing of *E. coli* on the Dynomics platform, it suffers from the disadvantage of imaging cells that are connected to the area where they are initially seeded. As seen in section 2.2.5, this creates some depth-dependent phenotype in the bulb area (see Figure 2.8), which may be a confounding effect for studies that target metabolism and growth-state dependent gene expression patterns. In addition, the bulb design limits the use of the Dynomics platform to *E. coli*, as *S. cerevisiae* strains were observed to not robustly grow in the bulb area. Elizabeth Stasiowski and Richard O'Laughlin developed a device [129] which solves both problems by decoupling the spotting area from the cell growth area. This novel device is capable of hydrodynamic trapping - a technique that enables cells spotted from a bulb area to seed a set of four downstream cell traps named HD (hydrodynamic) "biopixels" [130]. The HD biopixels are cone-shaped traps 70 µm by 70 µm with a 4µm height. 2µm-tall "conduits" at the bottom of the trap enable cells to be naturally sucked into the trap for rapid and consistent seeding of the biopixels (see Fig. 2.12).

This HD biopixel Dynomics device was designed to host 2176 unique *S. cerevisiae* microcolonies in a layout that matches the previously published chip for *E. coli* [79] presented in (Fig. 2.6a). I demonstrated that this design could alternatively be used to culture *E. coli*. The advantages of using this device are multiple. First, four biological replicates (corresponding to four HD biopixel traps) are collected for each strain instead of one previously (one "bulb" trap, see Fig. 2.1). Second, cells receive media from both ends of the trap and therefore are less likely to experience diffusion-limited access to nutrients, and diffusion-limited depletion of waste. Figure 2.13 shows the difference in the response of fluorescent *E. coli* cells growing either in the "bulb" area from the original design or the "biopixel" area from this new design. The advantages of the new design are clear, with higher response amplitude and faster signal decay when observing a *lacZ* reporter strain from the Alon library [112] responding to IPTG.

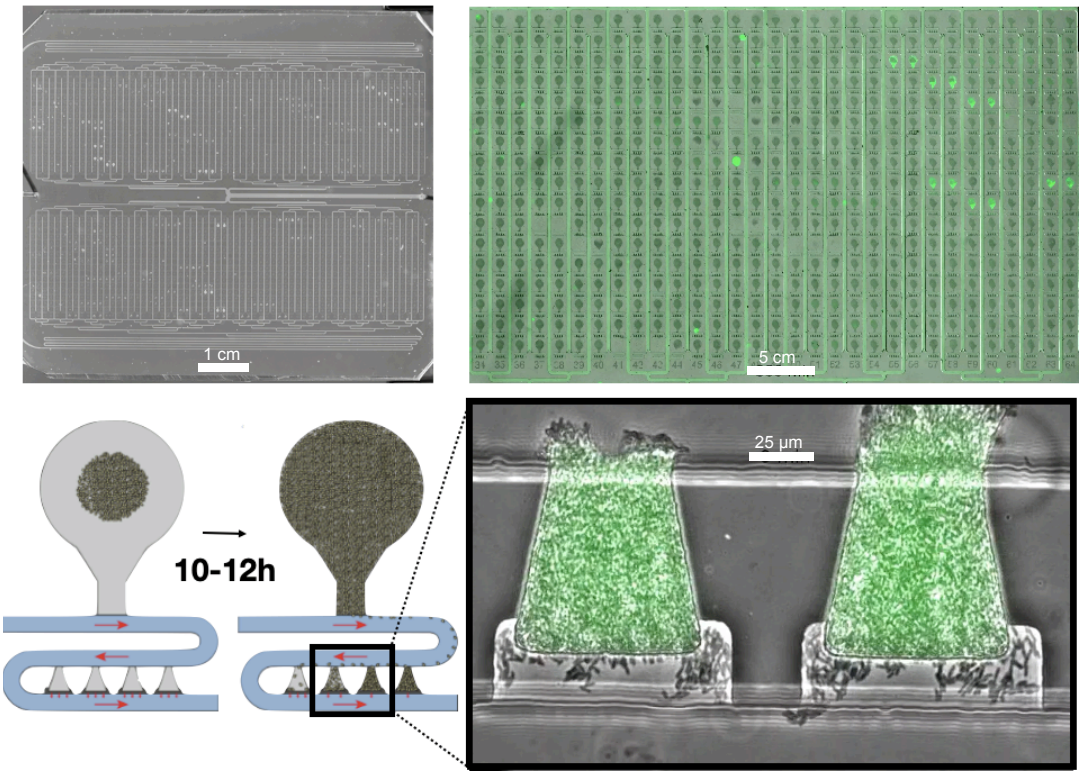

Figure 2.12: *E. coli* growing on 2176-strain HD biopixel device Top left: view of 2176-strain microfluidic device. Top right : combined view of GFP (green) and brightfield (gray) image of a subset of 500 *E. coli* strains growing on the device. Each strain has a different baseline GFP expression level. Bottom left: View of one of the 2176 device positions, with upstream spotting area and downstream array of 4 biopixels where cells are monitored for their changes in fluorescence. Biopixel cell traps reach confluence 10-12h after initial seeding of the device. Bottom right: View at 20X of two out of the four biopixel traps present for each spotted position on the device. The bottom left panel is adapted from [129]

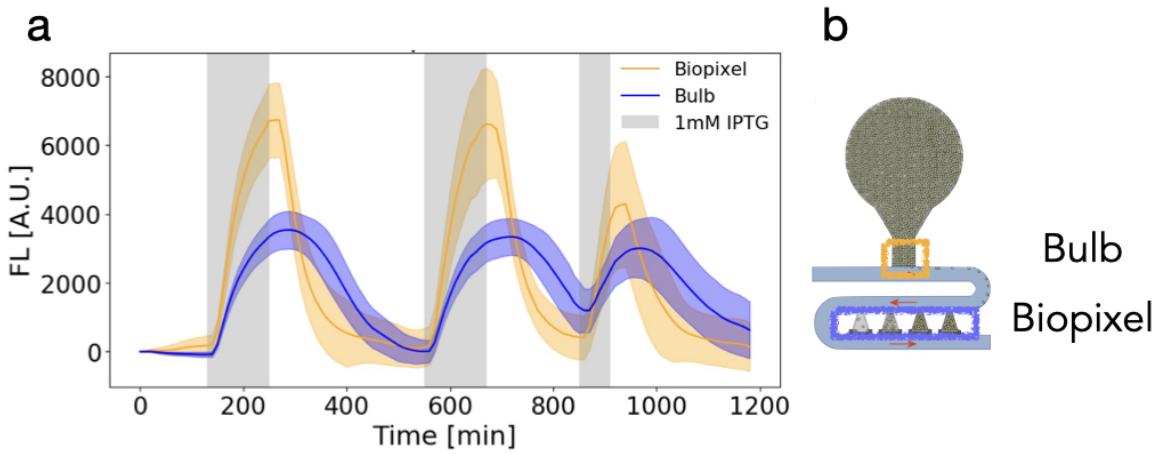

Figure 2.13: *lacZ* responding to IPTG in the HD biopixel dynomics device. a) The mean (dark line) ±1 standard deviation (shaded) of 544 copies of the *lacZ* reporter strain from the Alon GFP library [112] is plotted as it responds to three 1mM IPTG inductions of 2 hrs, 2 hrs and 1hr duration when measured in either the bulb area (orange) or biopixel area (blue) of the HD biopixel Dynomics device. Response amplitude is higher and decay time is lower for cells growing in the HD biopixels. The grey window denotes when IPTG was present. b) Schematic of a trap on the HD biopixel Dynomics device with Bulb and Biopixel areas highlighted. Adapted from [129]

2.3 Methods

2.3.1 Dynomics optical enclosure

The Dynomics optical enclosure is a stand-alone microscope for imaging the microfluidic device with a 70 mm x 70 mm field-of-view (Figure 2.14a). 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. 2.14c-e). With a nominal imaging frequency of 2 minutes, Dynomics allows us to adjust environmental conditions and measure the resulting changes in gene expression at a high temporal resolution.

2.3.2 Robotic loading of Dynomics microfluidic devices

To enable the loading of our 2176-strain microfluidic device, the first step was to arrange cells in the appropriate grid format at a 6144-density format. Fluorescent strain libraries are often stored in 96-well glycerol. We use a Singer ROTOR strain handling robot (Fig. 2.15a) to sequentially condense strains from 96- to 384-density using pin pad-based bulk strain movement (Fig. 2.15b). After this initial step, single colony picking was required to pattern the strains in the desired grid on 1536-density agar plates for loading of the device. We used a Singer Stinger (Fig. 2.15c) to create four cell grids on 1536-density agar plates for this step (Figure 2.15d). The penultimate step, done immediately prior to spotting to the PDMS device is to condense the four

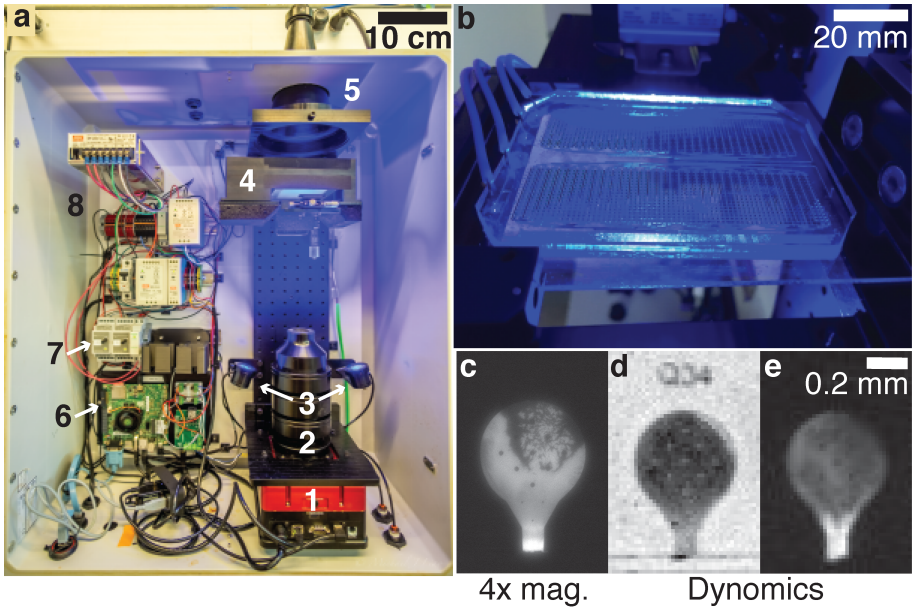

Figure 2.14: The Dynomics custom optical enclosure. a) Components include (1) a SBIG STX-16803 CCD camera, (2) a custom lens stack assembly, (3) blue excitation LEDs for GFP fluorescence imaging, (4) a temperature-controlled enclosure where the microfluidic device is mounted, (5) a green LED and associated 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 Dynomics microfluidic device imaged on a standard research grade scope with 4x magnification, and on the Dynomics enclosure for both transmitted light (d) and fluorescence (c,e) channels. The Dynomics enclosure has 36 µm resolution.

1536-density plates into one final 6144-density plate. In the final step, cells grown on agar in the 6144-format are transferred to the PDMS device (pre-exposed to Oxygen plasma, see 2.3.9) in a single step. Once cells are spotted, the device is completed by bonding to a glass slide.

The ROTOR loading protocol is a reliable way to load Dynomics devices, with the caveat of being a slow process in the initial arrangement of cells (the transfer from 384- to 1563- density using the Stinger). In parallel to the ROTOR-based loading protocol, we developed a new hybrid protocol combining two strain-handling robots: the Singer ROTOR HDA plating robot (that transfer many strains at a time, albeit in a predetermined geometrical pattern) and the acoustic droplet ejection Labcyte ECHO 550 (which can spot individual stains onto any position on target plat, see Fig. 2.16). We used the ECHO to transfer strains from a 384 well plate to 1536-density

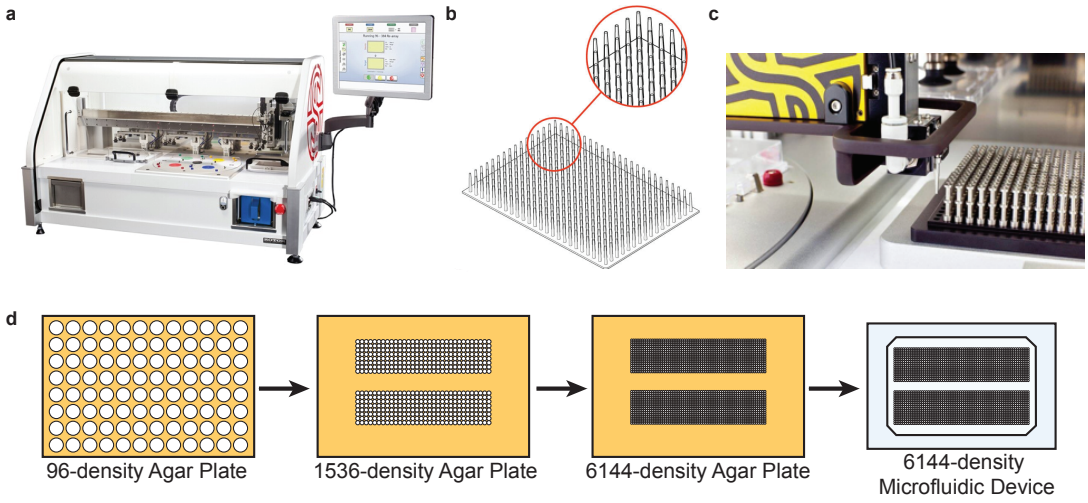

Figure 2.15: Cell arraying equipment and workflow for loading high-throughput microfluidic devices. a) Singer ROTOR HDA. 3) A 384-density long-pin RePad, used for bulk movement of strains. c) Singer Stinger, an attachment used for single colony movement. d) Workflow used to load strains onto a 2176-strain large-scale microfluidic device. Cells are re-arrayed from 96-density to 1536-density using the Stinger. 1536- density plates are combined into a 6144-density format, and then spotted to the microfluidic device

agar plates in an array whose shape and inter-strain spacing (each strain distant from one another by 1.125mm in x- and y-axis) corresponded to the layout of our microfluidic device. Once the strains and their replicates are re-arrayed to agar plates, we use the ROTOR HDA to transfer all 2176 strains at once onto a PDMS microfluidic device, as described in the original protocol. The advantage of the hybrid protocol is that different patternings of strains can be rapidly obtained from the source 384 plates, which is useful if there is a need for rapid iteration regarding which strains should be present on the device. The ECHO 550 can also be used to transfer cells directly into the 6144-format using a protocol developed by Nicholas Csicsery from the Hasty Lab. However, the limited accuracy and precision of the spotting at that scale limits its use to the loading of smaller devices, such as the 48-strain devices used in Chapter 4. We believe this loading protocol can be of great usefulness to members of the synthetic biology community who wish to load microfluidic devices flexibly and rapidly with multiple strains.

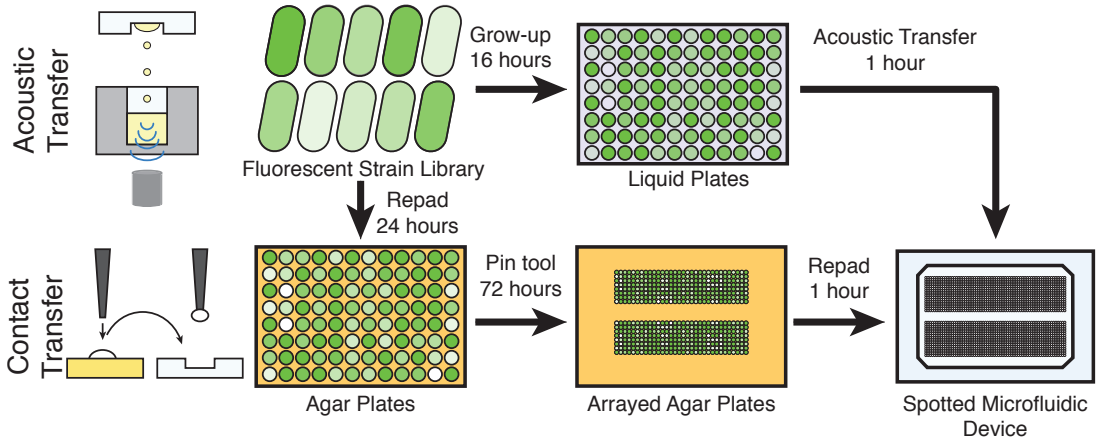

Figure 2.16: Individual *E. coli* strains can be loaded onto PDMS microfluidic devices using two different techniques. Contact transfer involves spotting cells from agar to PDMS using a Singer ROTOR colony-picking robot, but this technology is not as high-throughput or adaptable as liquid-handling methods. Acoustic transfer using the Labcyte Echo reduces the strain-to-chip loading time approximately four-fold and allows flexible selection and arrangement of strains to be loaded on chip.

Below is the step-by-step loading and bonding protocol for mid-scale and large-scale devices, for both *E. coli* and *S. cerevisiae* and was reproduced with permission from Graham *et al.* [79] and Elizabeth Stasiowski's PhD thesis [129]

2.3.3 Wafer fabrication

Our group has previously described the steps to design and build a silicon wafer patterned with the features of the microfluidic device [90]. 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.

2.3.4 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.

2.3.5 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.

2.3.6 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 of 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.

2.3.7 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-denstiy 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	**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 *E. coli*	55	70	10	10	0.6	0.6
6144 agar to microfluidic device for *S. cerevisiae*	50	64	10	10	0.6	0.6

2.3.8 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.

2.3.9 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.

2.3.10 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.

2.3.11 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 [90].

3. Prepare and outlet syringe (mid-scale) using methods previously described by our group [90] 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.

2.4 Acknowledgements

This chapter contains material published in 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.

This chapter also contains material being prepared for submission as Multiplexed microfluidic platforms for real-time heavy metal sensing and dynamic phenotype screening. Thouvenin Gregoire, Csicsery, Nicholas, O'Laughlin, Richard, Stasiowski, Elizabeth and Hasty, Jeff. (*equal contribution) The dissertation author was the primary author and researchers of this material.

Chapter 3

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

3.1 Introduction

Progress in DNA sequencing, synthesis, and assembly has fostered the use of microbes for a variety of medical, industrial, and environmental purposes. One of the applications to emerge from synthetic biology research in recent years is the use of microbes as biosensors to detect a variety of environmental toxins, including heavy metals. The detection of heavy metals in human, environmental and industrial settings is important given the danger they represent for ecosystems and human populations [131]. To this effect, various groups have worked on developing microbial biosensors by coupling metalloregulatory proteins (such as metal-sensing transcription factors) [132, 133] or two-component systems [134] to various output modules (bioluminescent [135], fluorescent [133], enzymatic [136] or even bioelectrochemical [137]) within eukaryotic microorganisms and prokaryotes. Light-based sensing offers the advantage of

being both a non-invasive and a sensitive form of measurement. Amongst light-based reporters, fluorescent proteins offers the further advantage of stability and do not impose too high a metabolic burden on the cell. Contrarily to other bio luminescent outputs, fluorescent proteins does not require consumption of energy sources such as ATP for their signal to be produced.

Compared to classical analytical techniques, whole-cell microbial biosensors offer the advantage of rapid, *in situ* detection and produce both qualitative and quantitative outputs in response to heavy metal ions [138]. Their use can increase the flexibility of the environmental monitoring process, limit costs, and improve the temporal and spatial resolution of heavy metal detection.

Microorganisms such as *E. coli* have evolved complex networks to ensure metal homeostasis in fluctuating environments. The *E. coli* genome contains genes that code for proteins specifically tailored to the detection, uptake, storage, and efflux of heavy metals [139, 140]. While previous approaches to microorganism-based heavy metal sensing have mostly relied on engineering a small number of biosensors that are specific to one metal [140, 141], few approaches have leveraged a microorganism's genome-wide response to detect environmental stressors.

Heavy metal detection

Heavy metal detection is typically performed by collecting grab-samples at the water source and performing highly sensitive and quantitative testing in centralized laboratories using spectroscopy-based techniques (such as atomic absorption spectroscopy or inductively coupled plasma mass spectrometry [142]). This process is not well suited for frequent field application due to its time-consuming nature (transport, sample pre-treatment) and cost (complex instruments and trained technicians are required) . This can result in days separating sampling from results and limit their application in developing countries [133, 143].

3.2 Approach

We used the device described in Chapter 2 as a high-throughput microfluidic biosensor to simultaneously co-culture and monitor 1,807 strains of a promoter-GFP *E. coli* library [144] 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 In this chapter, we loaded the platform Dynomics with the *E. coli* GFP-promoter library (which contains more than half the genes in the *E. coli* genome) to monitor the dynamic response of the many genes (including metalloregulatory) that are involved in heavy metal stress response. Our device will serve as a "biosensor array" and provide a comprehensive view of *E. coli*'s transcriptional response to heavy-metal exposure.

Our biosensing approach differs significantly from the rational engineering of orthogonal sensors. Since we use a library of "agnostic" transcriptional reporters initially developed as a tool to measure general gene expression dynamics in *E .coli*, these strains display the cross talk that naturally exists in *E. coli* gene regulatory networks. The question for our biosensor thus shifts from avoiding cross-talk to discerning unique fingerprints in the genome-wide transcriptional response induced by each heavy metal. By recording the time-series response of the 1,807 promoters of the Alon library with Dynomics, we effectively obtain a time-varying "barcode" response to each metal. Conceptually, while a cross-talking strain may perturb classification in the context of single-strain biosensors, the likelihood that thousands of different strains exhibiting the same reaction when exposed to different heavy metals becomes very slim and allows for accurate machine-learning based classification. We believe robustness in the face of cross-talk is, in fact, one of the differentiating properties of our sensor.

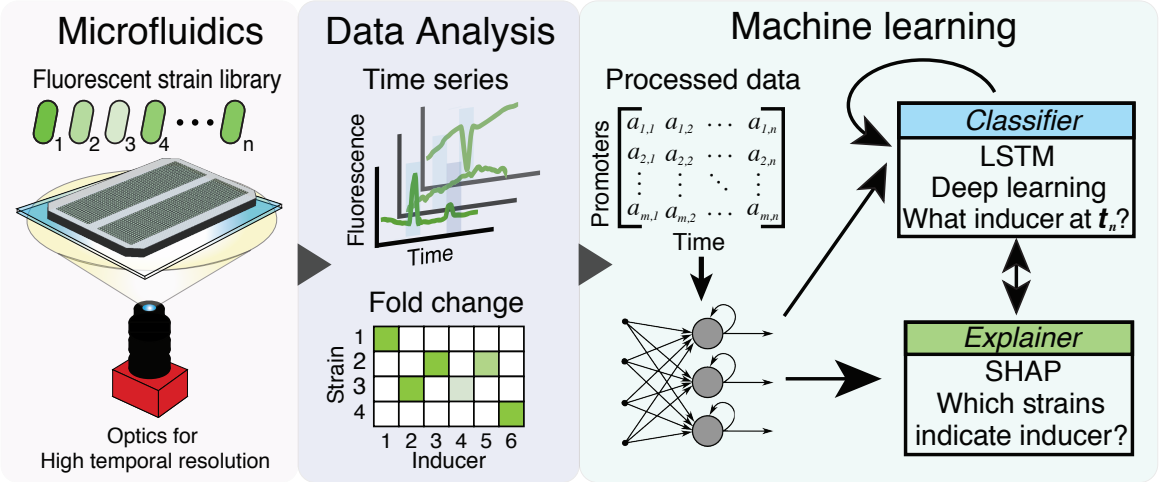

Figure 3.1: The Dynomics platform for biosensing. Fluorescent strain libraries are loaded onto large-scale microfluidic devices that can be fully captured in a single image using custom optics. Parallel cultures of *E. coli* are subjected to multiple exposures of different stimuli with time series and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

3.3 Screening for responsive promoters to heavy metals

Using the Dynomics platform with a previously developed GFP *E. coli* promoter library [112], 1,807 unique *E. coli* promoters were screened against nine heavy metals (Cu(II), Zn(II), 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 3.2).

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 3.3a) and subsequent clustering (Figure 3.3b) reveal distinct classes of transcriptional responses to a single four hour zinc exposure. In (Figure 3.3b), clusters 1 and 2 include promoters that are up- and down-regulated, respectively, in the presence of zinc, but return to baseline expression

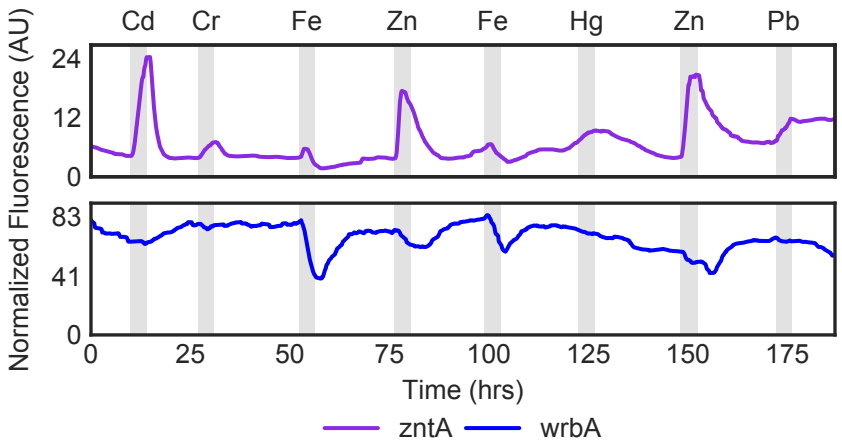

Figure 3.2: Responsive strains over the duration of a Dynomics experiment. Normalized fluorescence for two strains is plotted over the duration of one experiment, with four-hour heavy metal inductions (gray bars) occurring once daily.

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 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 3.4a) 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 3.4b) quantitatively elucidates dynamics, such as amplitude, relaxation time, and response rate, all of which important factors for their use in the study of gene expression regulation and continuous biosensing applications.

There is a growing body of knowledge concerning genes in the *E. coli* genome that code for proteins specifically tailored to the detection, uptake, storage, and efflux of heavy metals [139].

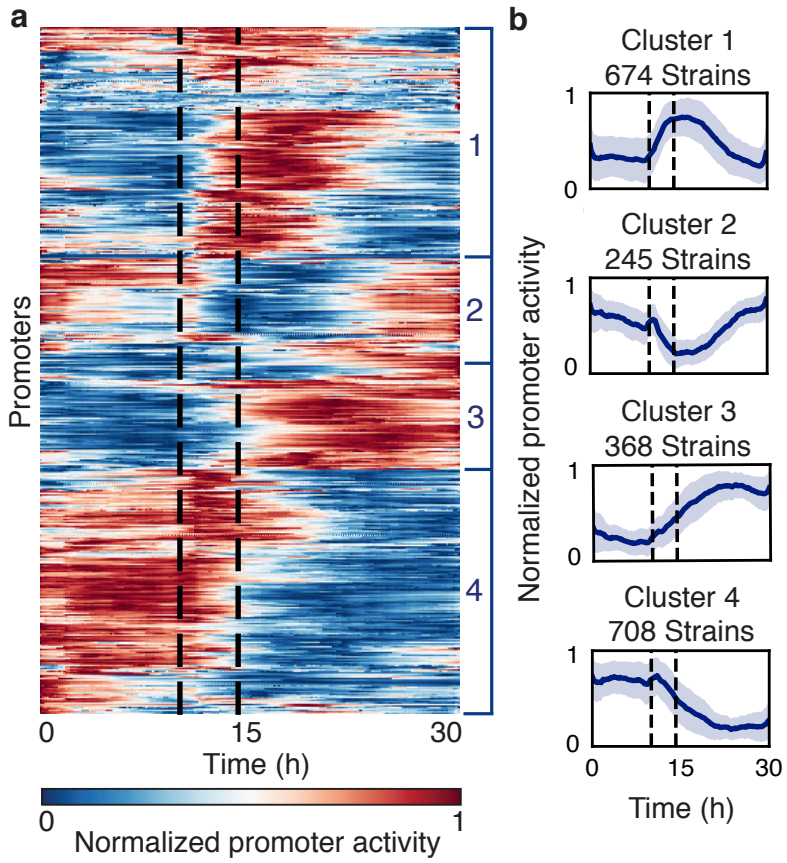

Figure 3.3: Dynomics as a screening tool for heavy metal responsive promoters in *E. coli* . a) Fluorescence response of an *E. coli* promoter library during a four hour 50 ppb Zn induction (dotted window). Each row represents the promoter activity, normalized between 0 and 1, of a single strain, with 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.

One of the advantages of collecting genome-wide transcriptional responses to heavy metal stressors with our platform is that we can observe responses from both these metalloregulatory genes and from a larger group of genes not directly tied to heavy metals homeostasis, but whose modified activity reflect the overall stress imposed to the cell.

Indeed, many of the strongly-responding genes in our dataset, both in terms of response fold-change (Fig. 3.4) or SHAP values (Fig. 3.9) are well-characterized, metal-specific regulatory proteins. Such proteins play different roles within the cell such as metal detoxification (*cueO*) ,

Table 3.1: Gene ontology enrichment analysis of the clusters obtained by agglomerative clustering of Figure 3.3.

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

metal-transport (*zntA*, *znuA* and *znuC*) or metal-related transcription factor activity (*iscR*). Within those metalloregulatory proteins, some response specifically to one metal (*cueO*, for example,

responds specifically to Copper [145], see Fig. 3.4), while others experience cross-talk (*zntA* responds to Cadmium, Lead and Zinc, a behavior that is well documented [146]).

As mentioned in the discussion, the up-regulated gene clusters identified from genome-wide response time series (Fig. 3.3) are enriched in gene ontology terms for cellular detoxification and response to toxic substances. This is reflective of the non-specific stress response concurrent with heavy metal exposure. This is also reflected in our fold-change and SHAP analysis, with genes such as *codB* (involved in cytosine transport), *ilvC* (involved in biosynthesis), *yeeE* (membrane protein) and *ybiN* (Ribosomal RNA large subunit methyltransferase) experiencing non-specific upregulation in response to Iron induction, for example.

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

3.4 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 Dynomics data. The first model, known as extreme gradient boosted trees (XGBoost), is a popular decision tree ensemble-based classifier known for its ability to learn nonlinear models [147]. The second, known as a long short-term memory recurrent neural network (LSTM-RNN), is a DNN [148] 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

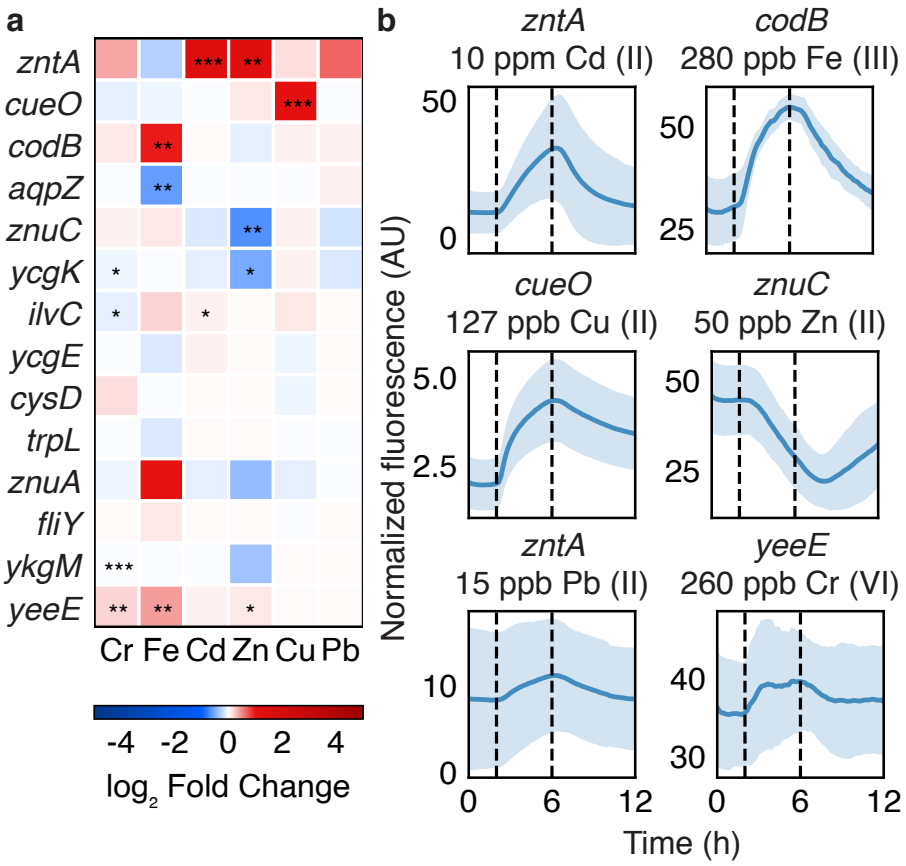

Figure 3.4: a) Fold change for top responding strains to all metals. Log2 of the average fold change is shown for the top responding strains to each heavy metal. *, **, *** indicate p-values of 0.05, 0.01, 0.001, respectively. b) Significant single strain normalized fluorescence response (blue line) and standard deviation (blue shading) across all inductions for a given metal (dotted window).

overall. As seen by examining the diagonal elements in the confusion matrix in Fig.3.5, the LSTM-RNN was able distinguish both biotic and xenobiotic metal-spiked water from pure water with a high level of reliability.

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 [145, 149]. 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 [146, 150]. Most classification errors occurred during the 10-40 minutes at the start or end of the induction periods, when the LSTM-RNN occasionally had difficulty determining the exact time that each metal was added or removed from the media (Fig. 3.7B). This is most pronounced with the prediction of lead, for which the classifier incorrectly predicted no toxin for 48% of time points where lead was present. This is largely due to the weak promoter responses induced by 0.03 ppm lead, which is only double the Environmental Protection Agency (EPA) maximum contaminant level. In lead exposures with poor prediction, time points at the start of the four hour induction window are misclassified as no toxin, while lead is accurately predicted near the end of this window.

While past studies have used machine learning frameworks to assign cells to chronologically distinct phenotypes based on their transcriptomes [151], we believe this is the first instance of a multiclass classifier successfully leveraging genome-wide transcriptional dynamics in live cells to predict exposure of a biological organism to an environmental stressor.

3.5 Explainable artificial intelligence

Over the last two decades, computer science has experienced a massive increase in computing and algorithmic power in the field of artificial intelligence (AI) [152]. Concurrently, the rise of -omics technologies has lead to exponential increase in the volume of experimental data [153–155]. Machine learning has not yet fully delivered on its potential to facilitate scientific discovery because of 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. While AI has demonstrated proficiency in discovering the ”what” of scientific phenomena, researchers are still left struggling to understand the corresponding ”why”. Hence, the rise of deep neural networks, which currently are the most extreme algorithmic embodiment of the black-box problem, has motivated a search for explainable artificial intelligence (XAI) techniques [156].

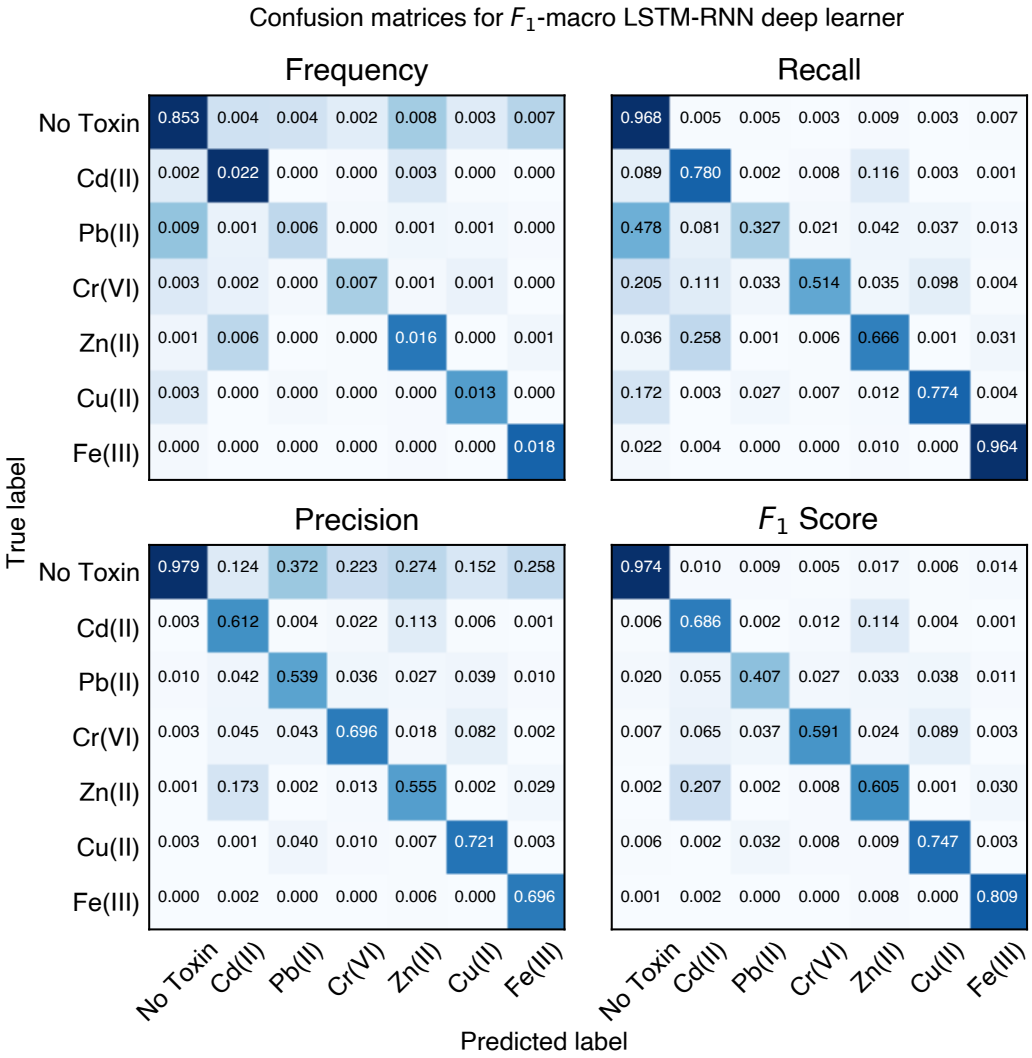

Figure 3.5: Confusion matrices showing the frequency, recall, precision, and F1 score of the LSTM-RNN classifier in predicting six metals across all experimental data (14,332 time points).

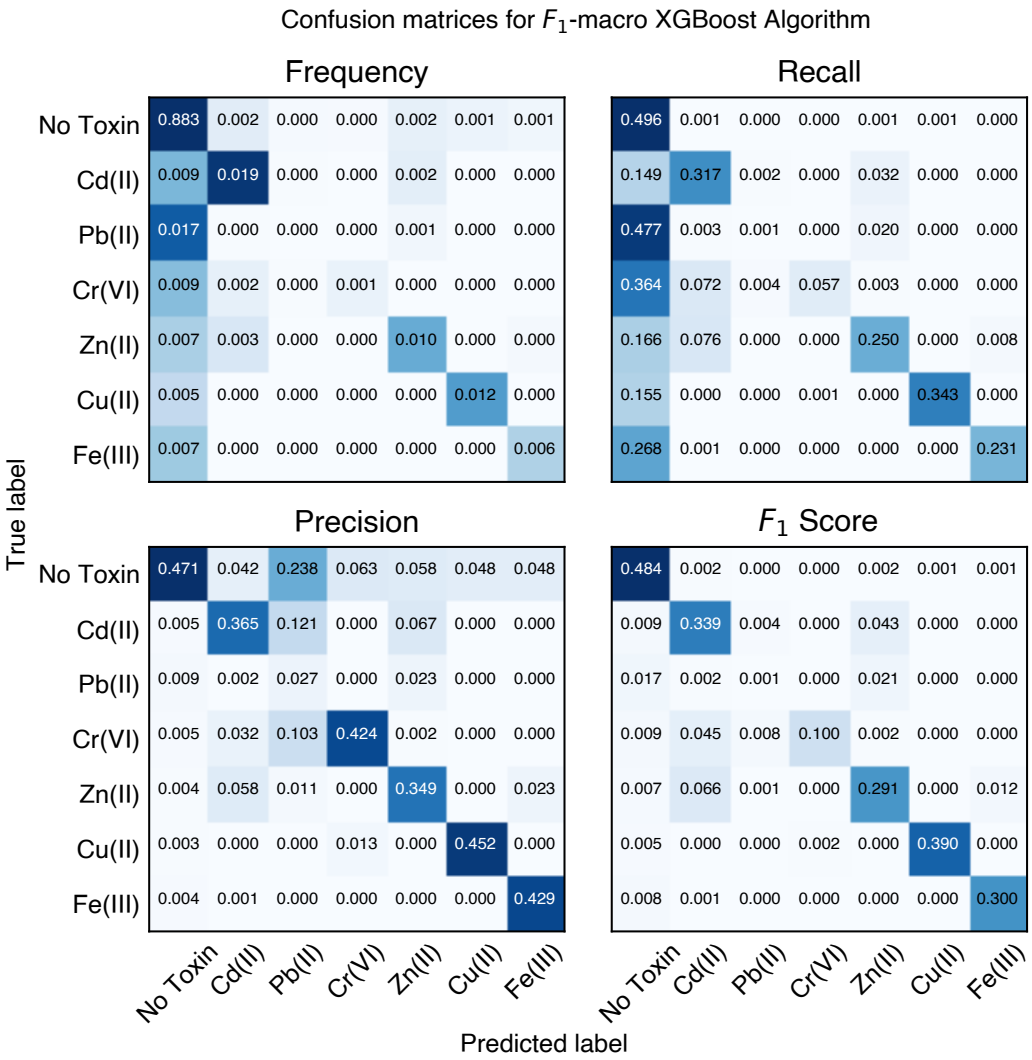

Figure 3.6: Confusion matrices showing the frequency, recall, precision, and F1 score of the XGBoost classifier in predicting six metals across all experimental data (14,332 time points).

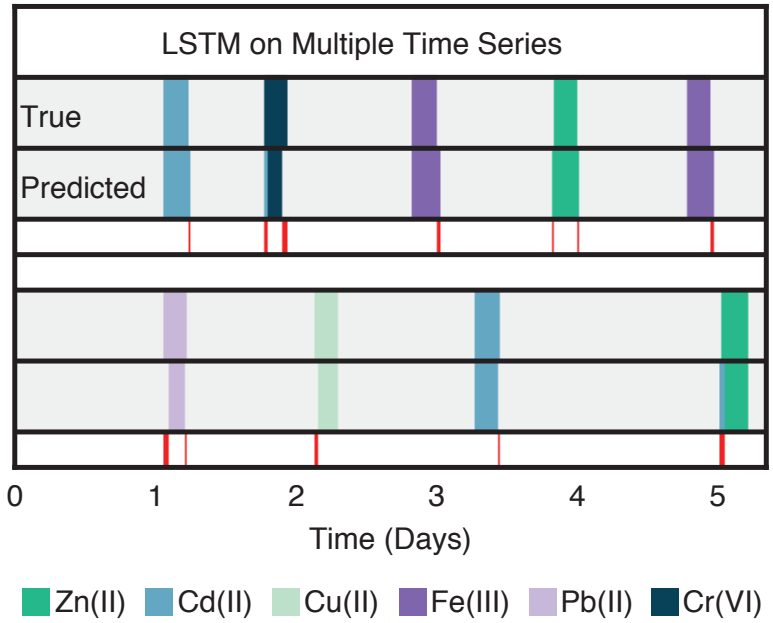

Figure 3.7: LSTM-RNN classifier applied to time series data for all six detectable metals in two different experiments. Misclassified time points are shown in red.

The quantitative life sciences have seen some excellent efforts to create research-focused XAI, but these techniques are either model-specific [157] or limited in their applicability because of non-uniqueness or computational complexities [158, 159]. Recently, however, 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 [160, 161].

Taking advantage of these recent XAI advances, we trained a Shapley additive explanations (SHAP) XAI on both our XGBoost and LSTM learners [160, 162]. 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 3.8). 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 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.

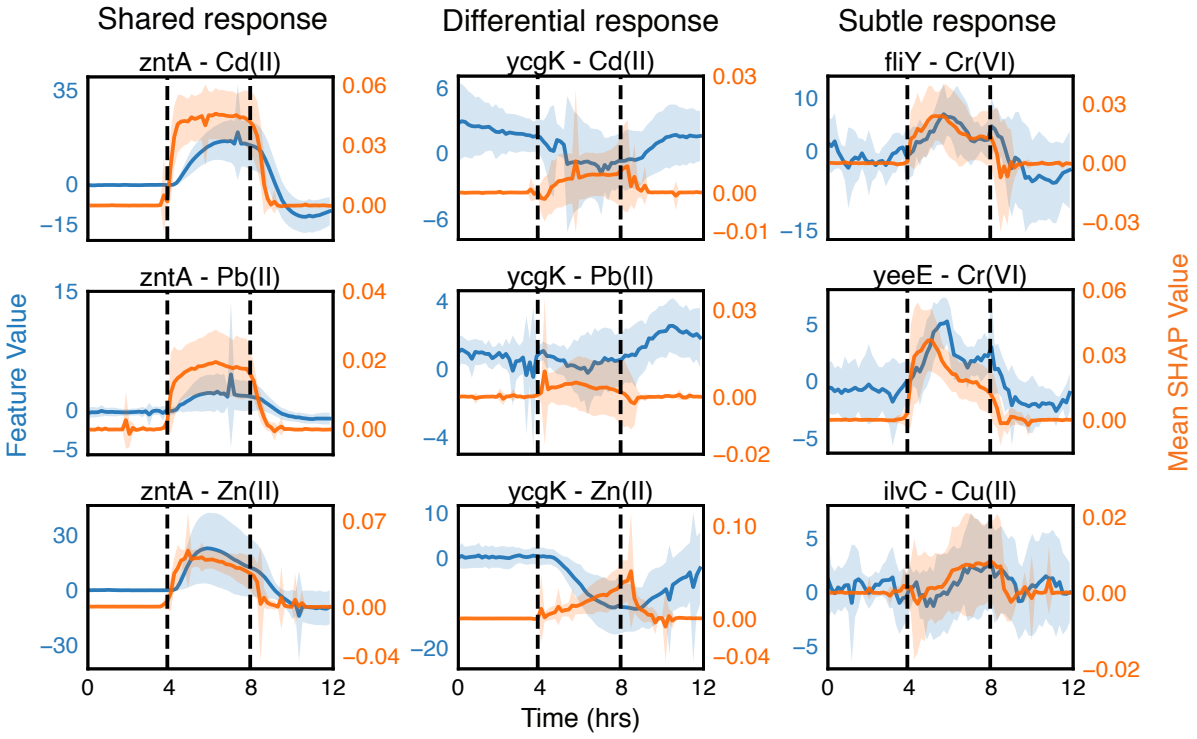

Figure 3.8: Dynamic SHAP and feature values during metal exposures. Feature (blue) and SHAP (orange) time trajectories for individual promoters during metal exposures. Solid lines show the mean value over all inductions for that metal and shaded regions around lines represents standard deviation. Dashed 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.

The SHAP-XAI highlights the similarities and differences between how the LSTM-RNN and XGBoost make decisions. Figure 3.9 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 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 [163].

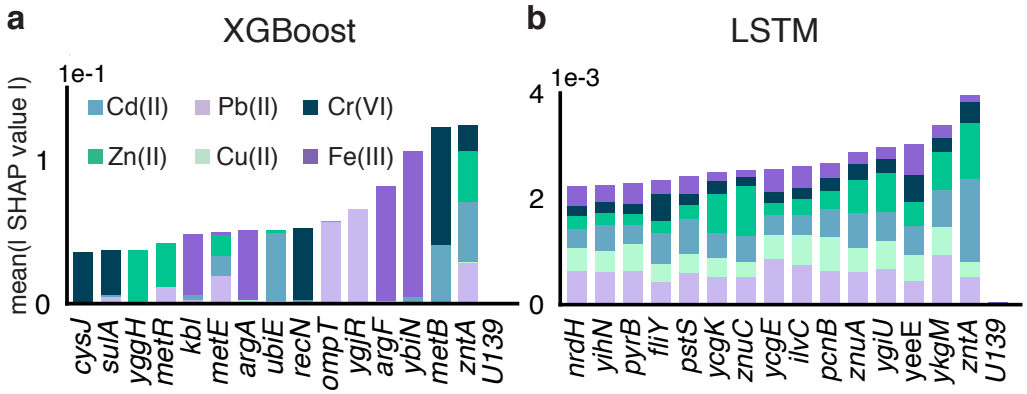

Figure 3.9: Explainable machine learning reveals *E. coli* transcriptional dynamics contributing to metal classification. The bar plots show the cumulative contribution based on the Shapley Additive Explanations (SHAP) values of 15 top promoters and a negative control (promoterless strain U139) to the 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.

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

and iron inductions (Figure 3.10). 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 argenine synthesis, *argA* and *argC*, while the LSTM-RNN relies on different promoters that are involved in other metabolic or biosynthetic processes.

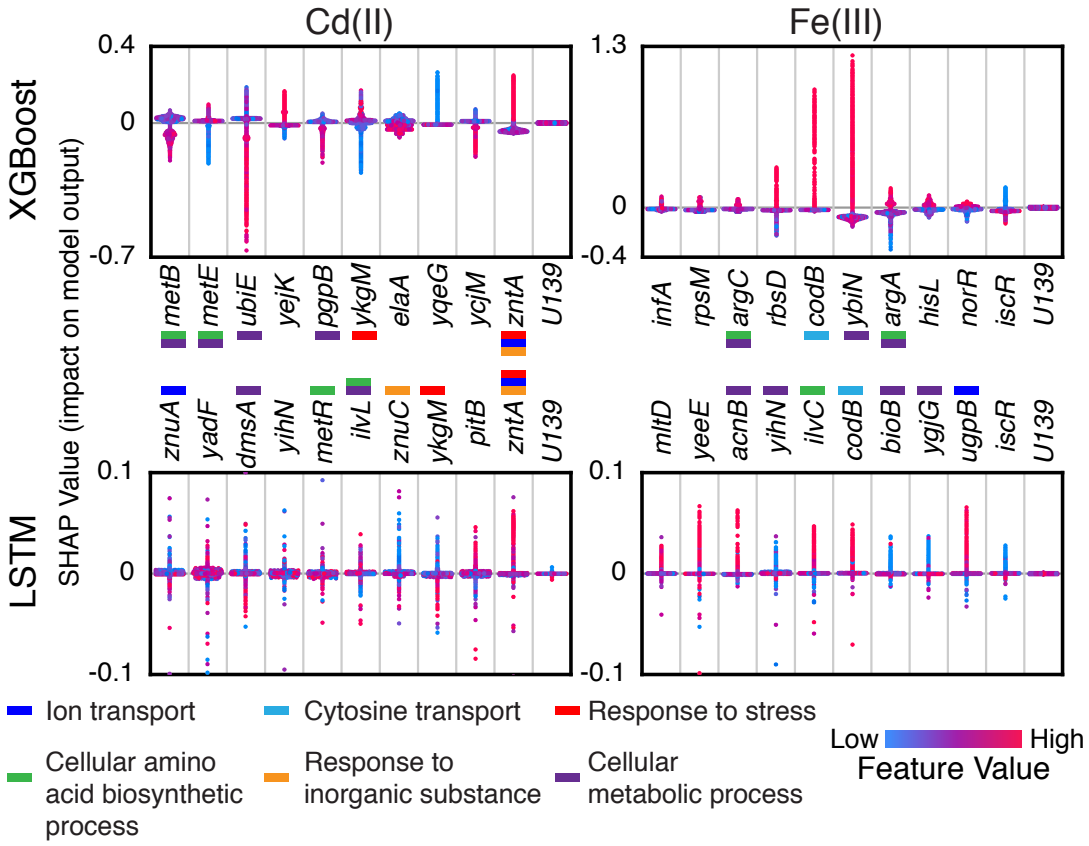

Figure 3.10: SHAP values are shown for 10 top promoters and a negative control (promoterless strain U139) for Cd(II) and Fe(III) for XGBoost and LSTM-RNN. Each point represents the feature value (normalized first derivative) at a time point. Positive SHAP values suggest that a given metal is present 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.

3.6 Biosensor validation

Given the severe impact of heavy metals on human health [164] and the persistence of water quality issues in the US [165] we sought to deploy the Dynomics 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 3.11a 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.

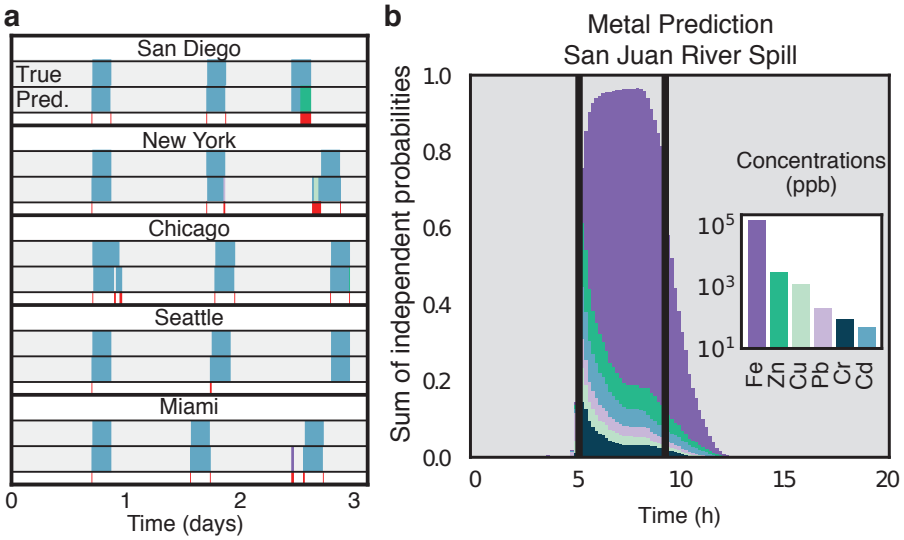

Figure 3.11: Dynomics and machine learning on environmental samples. a) LSTM-RNN classification of cadmium contamination added to five different urban water sources. The colors correspond to the metals in the inset in (b). b) Multi-class, multi-label classification of water samples from the San Juan 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.

The Dynomics device was also exposed to samples collected from the Gold King Mine Spill in August 2015. Figure 3.11b 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.

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

3.7 Discussion

Adaptation to rapidly changing environments is a hallmark of bacterial life and requires critical dynamic properties, many of which can be traced down to rapid and specific control of gene expression [166]. Transcriptomic technologies [167–169] have identified genes involved in environmental stress response and 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. The Dynomics platform offers a novel experimental approach through its high temporal resolution, degree of multiplexing, and precise experimental control. In a high-throughput screen using Dynomics, we simultaneously exposed 1,807 strains of the promoter-based *E. coli* GFP library to nine different heavy metals. The fine-grained temporal gene expression data it produced highlighted the unique dynamics of stimuli-specific genes previously reported as heavy metal responsive [168], and identified gene clusters that shared similar response dynamics.

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 such as nonlinear interaction patterns and feedback loops. These phenomena are ubiquitous in biology, yet difficult to discern using static gene expression data [14].

We illustrate our platform's potential for exploring the dynamics of transcriptional networks by applying machine learning techniques to examine heavy metal stress responses in *E. coli*. Here we demonstrate that supervised machine learning can infer exposure to environmental stressors from real-time observation of transcriptional activity at the genome-scale. Time series from 1,807 strains were used to differentiate between multiple biotic and xenobiotic heavy metals. We believe this study is the first instance of dynamic mapping between transcriptomic changes captured in live microorganisms on the one hand, and their surrounding environment in the other. These data, with genome-scale coverage and high sampling frequency, could be used in

future studies to screen large strain libraries for common motifs, such as nonlinear interaction patterns and feedback loops, which are difficult to discern using static gene expression data [14]. Our findings suggest that *E. coli*'s internal models of its environment can be instantiated in the dynamics of its adaptive response to heavy metal stress [170].

While machine learning is well poised to address the complexity of large, multi-dimensional biological data sets such as the ones generated by our platform, more powerful algorithms often lose in interpretability what they gain in predictive powers. We addressed this shortcoming by using recent XAI techniques to gain insight into the features used by the predictive algorithms trained on our data. The SHAP XAI revealed that formally different algorithms rely on different biological features to classify transcriptomic adaptation to stress. While a decision-tree model relied heavily on a small number of strains, a more performant deep learning algorithm based its prediction on many strains of moderate influence (Fig. 3.10). These findings reveal that there are different ways to segregate the high-dimensional space explored by an organism's transcriptome during sensory response.

Finally, we show the real-world applicability of our platform for the detection of heavy metals in both urban water sources and field samples from a recent environmental catastrophe. As compared to conventional methods of metal quantification, such as atomic absorption spectroscopy or ICP-MS, the Dynomics platform sacrifices detection sensitivity for the ability to report continuous measurements, eliminating the need to take discrete samples. Although the Dynomics platform sometimes experiences a slight lag in the detection of metals when they are first introduced or removed, the platform is still a significant improvement over grab-sampling. While previous approaches to microorganism-based heavy metal sensing have relied on engineering a small number of biosensors that are specific to one metal [171], here we use *E. coli*'s transcriptomic response at the genome-scale to detect environmental stressors. Our biosensor was robust to the differences in ionic composition of five urban water sources and consistently detected cadmium in those samples. In addition, it was able to simultaneously detect multiple target metals

in mine spill samples, despite not being trained to perform this type of multi-class, multi-label classification. This result suggests our approach may outperform single purpose biosensors in accuracy and robustness, and may be adaptable to more varied sensing tasks via optimization through testing combinations of metals and different concentrations of metals. In summary, combining high-throughput microfluidics and machine learning can produce new insights into the coordination of cellular processes at a system-level and this new type of data can be leveraged for environmental monitoring.

3.8 Materials and methods

3.8.1 Wafer Fabrication

The silicon wafer was fabricated using standard photolithography techniques previously described by our group [90]. The cell trap layer was fabricated using 2005 SU-8 photoresist with a spin speed of 5000 rpm and had a resulting height of 3.85-4.05 μm, the minor channel layer was fabricated using 2075 SU-8 photoresist with a spin speed of 4250 rpm and had a resulting height of 45-50 μm, 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 μm.

3.8.2 Microfluidic device fabrication

The microfluidic devices were fabricated, loaded, and bonded with the *E. coli* promoter library [112] using the protocol in section 2.3.2.

3.8.3 Experimental protocol

Microfludic experiments were performed on the Dynomics 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 PTFE#24 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* [43, 90]. Methyl α-D-mannopyranoside inhibits normal surface attachment of type I pili but cannot be metabolized by *E. coli* [172]. 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 [173] 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 Streptomyces kanamyceticus	0.086
Tween 20	0.611
Methyl α-D-mannopyranoside	50

3.8.4 Live-cell imaging and data

Microfluidic devices were imaged using the Dynomics 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 optical vignetting using the following equation:

$$C = m * \frac{R - D}{F' - D'} \tag{3.1}$$

where R is the raw image to be flat-field corrected (Figure 3.12a), 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 3.12b).

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 3.13).

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
		298.4	0	1
Mercury	2	2	6	6
Zinc	5000	50	5	5
		500	0	3
		1310	0	1

Raw data was initially processed as:

$$GFP_{Raw} = \frac{GFP_{Bulb} - GFP_{Background}}{GFP_{Background}} \tag{3.2}$$

For fold change analysis, fluorescence values were passed through a median filter (scipy.signal.medfilt, kernel_size=11) and normalized by promoterless strains (Figure 3.14). Promoterless strains U139 and U66 from the GFP promoter library [112] were spotted at various locations across each device, with 20 device positions of each, for 40 in total. Normalized fluorescence for Figure 3.4 was thus determined as:

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

Fold change in Figure 3.4 was calculated as the quotient of the normalized fluorescence at the first

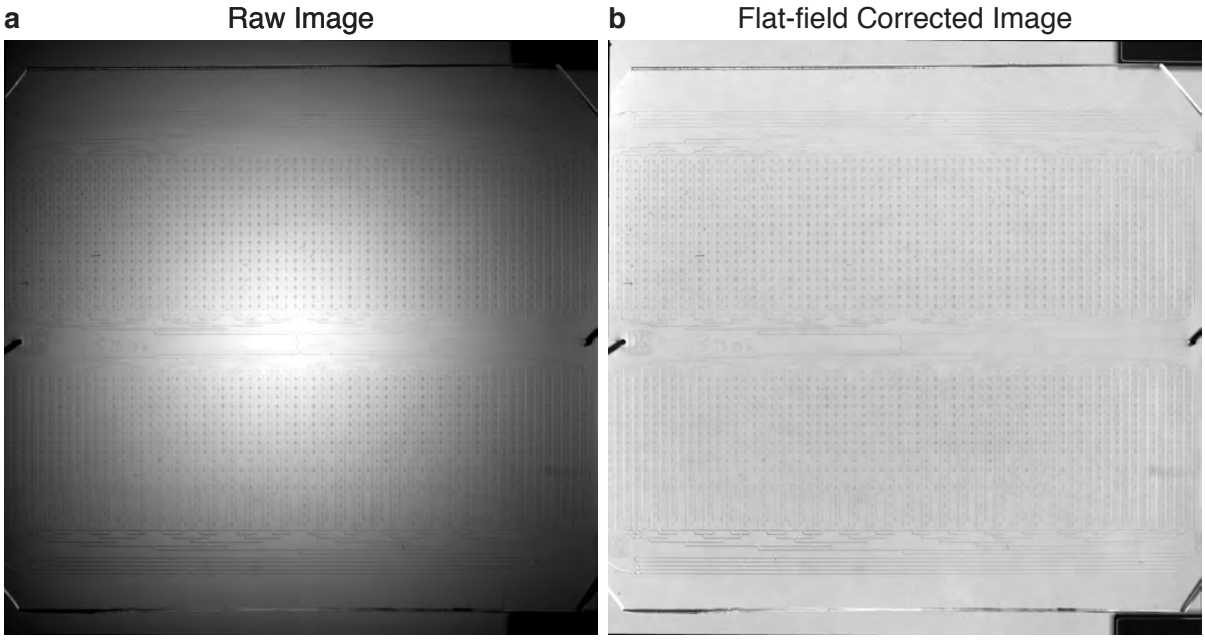

Figure 3.12: A full Dynomics image taken on custom optics before (a) and after (b) flat-field correction.

and last time point of each metal exposure. P-values were determined by a dependent two-sided t-test on the log_2 fluorescent values at the start and end of each metal exposure (scipy.stats.ttest_rel).

3.8.5 Gene ontology enrichment analysis

We performed enrichment analysis [174] 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 3.3). 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 [112].

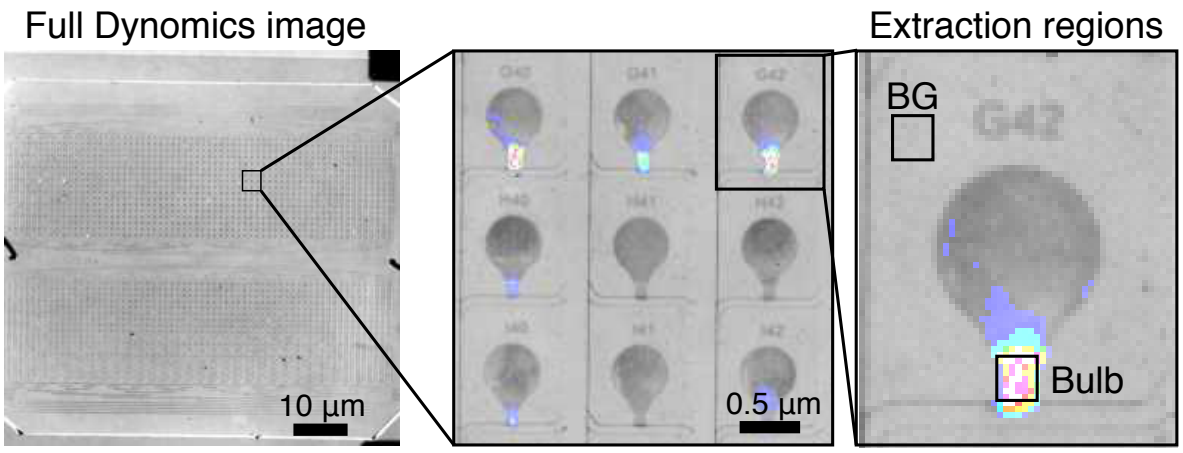

Figure 3.13: Data extraction from a Dynomics image. 2,176 device positions are simultaneously imaged in both transmitted light (gray) and green fluorescence (16 color) channels. At each time point for each device position, the mean fluorescence of the boxed bulb and background (BG) regions are extracted.

3.8.6 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 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 3.15 illustrates the results of the feature engineering process.

In order to optimize the classifiers, extensive Bayesian Optimization searches were used to find optimal hyperparameter combinations [175]. Throughout our hyperparameter searches,

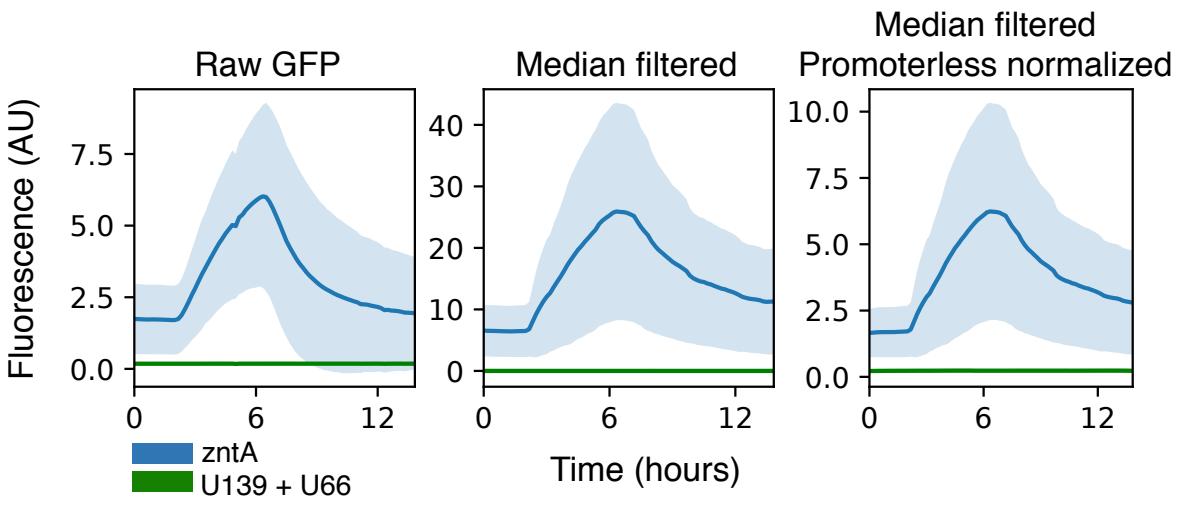

Figure 3.14: Normalization process for data presented in Figure 3.4. Raw data was median filtered, then divided by the mean traces of promoterless strains U139 and U66. Dark lines show the mean of all time series while shaded regions represent the standard deviation.

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 [176]. 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

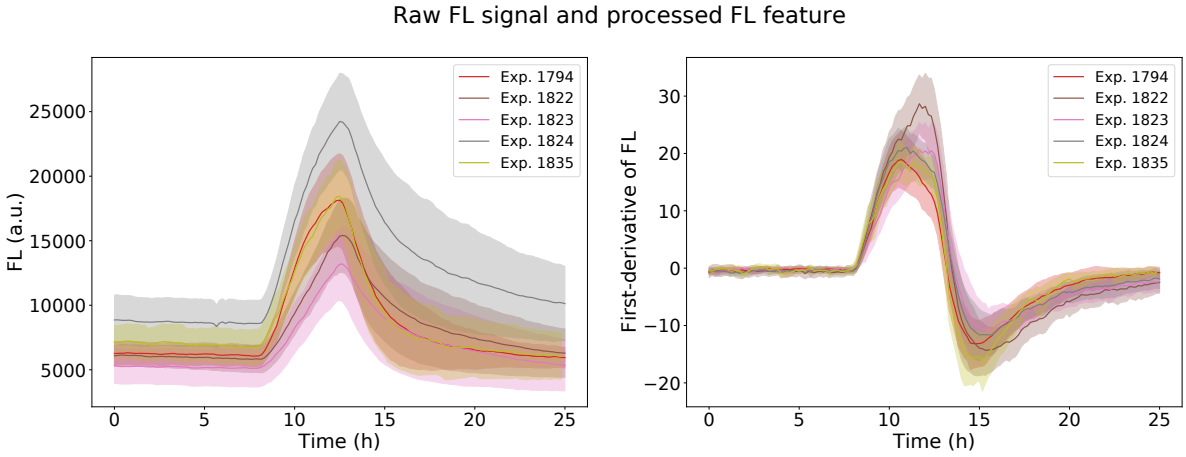

Figure 3.15: Raw and processed fluorescent signals: Our feature engineering eliminates a significant amount of intra-experiment variability by rendering the raw signal into a first derivative-like feature. This variability is due to differing hardware between Dynomics devices, among other sources.

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 [125]. These operons are known to be highly efficient at sensing and exporting arsenic and mercury, respectively, from their cells [177, 178]. 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* [179]. To focus classification on detectable metals, features during these metal inductions were relabeled as No Toxin.

3.8.7 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 3.8.3.

3.8.8 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 3.8.3.

3.9 Acknowledgements

This chapter contains material published in 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

A microfluidic biosensor for the detection of contaminants in seawater

4.1 Introduction

As discussed in Chapters 1 and 3, the use of biological systems and especially bacteria as biosensors for *in situ* analysis of water samples offers multiple advantages. Microbes naturally operate a vast range of niches, often residing in environments that are toxic to other organisms. In those niches, they withstand variations in temperature, pH, and salinity. As such, biosensor-based analysis seldom requires complex sample pre-treatment and can function with low-volume samples. Biosensors also benefit from the improvements and cost-effectiveness of microelectronics, optical devices, and wireless communication modules to which they are interfaced. As such, biosensors offer both low operational costs and relative ease of use. This is an advantage in the development of sensing networks deployed over large areas [180].

In parallel, the discovery, characterization, and modularization of genetic parts brought about by synthetic biology have fostered the development of a new generation of whole-cell biosensors. Scientists are developing a growing catalog of genetically-encoded sensor modules

that can respond to various environmental stimuli. The sensors modules often stem from naturally occurring microbial stress responses or metabolic pathways, a logical reflection of the diverse and unpredictable stimuli to which bacteria are naturally exposed to [181, 182]. The diversity of sensor modules is particularly important in the context of seawater pollution, where the source of pollutants is diverse (sewage, agricultural run-off, industrial discharge), as is the nature of the pollutants: both inorganic compounds (e.g. heavy metals) and organic compounds (e.g hydrocarbons) can be present [180]. The use of whole-cell biosensors also offers the promise of rapid detection: in microbial cells, sensor modules are rewired to transducing modules to produce measurable colorimetric, fluorescent, or enzymatic outputs, often within 5 minutes of exposure to the input [183].

Microbial biosensors are thus poised for use in the sensing of complex targets in marine environments [184]. However, a number of challenges prevent widespread use in the field. Firstly, biosensors developed in the lab often under-perform when exposed to their "working environment", Secondly, biosensors need to be combined with an adequate platform for cell culturing, and multiplexed sample handling. Thirdly, there are biosafety risks associated with using genetically engineered microbes in the field [133, 185]. In this chapter, I present work that directly addresses the first two of those challenges.

4.2 Approach

In line with the application of the dynomics platform as a multiplexed biosensor in Chapter 3, here I describe the extension of the platform's biosensing capabilities to the detection of a wider range of chemical and biological contaminants in the context of real-time seawater monitoring.

For a proof of principle application of our seawater biosensing platform, we decided to focus on a set of contaminants of concern (COCs) described by Kara Sorenson and co-workers ([186]) as being of particular relevance to Navy divers, and aimed to detect those chemicals at

concentrations up to a hundred fold below those at which they become a health hazard during frequent dives in contaminated waters (see Table 4.1). To this end, we used a medium-scale multiplexed microfluidic device capable of handling up to 48 separate strains loaded with *E. coli* sensor strains using the robotic loading techniques described in chapter 2 (see Fig. 4.8. Once cells are loaded into the device, they grow on seawater-containing media and are repetitively exposed to various dilutions of the COCs. Their fluorescent responses are measured over time to ensure their sensing abilities are not hindered by prolonged growth in seawater.

While *E. coli* is commonly found in the gastrointestinal tract of mammals, prior research has shown that some strains of *E. coli* are capable of surviving prolonged exposure to seawater, an environment they may be exposed to after a transit through sewage or coastal animals [187]. Growth of the enteric bacterium is reportedly more robust in the face of the stress-inducing pH, temperature, and salinity of seawater when the exposure to light radiation is limited and organic nutrients availability is sufficient [188]. Some unknowns remained regarding the viability of the *E. coli* laboratory strain (and synthetic biology workhorse) MG1655 grown in seawater in a microfluidic device for prolonged periods of time, as most of those studies used different *E. coli* strains for their viability studies, and most observed the presence of cells in a viable but non culturable state [189, 190]. Although not a primary object of this study, we note here that the ability of microfluidics-based assays to reliably create controlled microenvironments is a boon to the study of microbial ecologies in their natural contexts, with applications in the study of biofilm formation and antibiotic resistance [191].

Of contaminants of concern presented in Table 4.1, we note the presence of some of the heavy metals discussed in chapter 3 such as chromium (Cr), cadmium (Cd), lead (Pb), arsenic (As), mercury (Hg), Selenium (Se), and zinc (Zn). In addition to those metals, we also note the presence in our target compounds of monoaromatic hydrocarbon (notably Benzene and Styrene), which present a whole new class of chemicals for which to validate our seawater-compatible microfluidic sensor. Faced with the added challenge of sensing a more varied set of inputs we

aimed to augment our sensing technology by combining our microfluidic platform with two types of bacterial sensors: one-component (OCS) and two-component systems (TCS) (see Figure 4.1)

The first major family of bacterial sensors are one-component systems (OCSs) [182] and are illustrated in Figure 4.1. In OCS sensors, a cytoplasmic transcription factor directly binds the input, and responds by activating one or more output promoters. The metal-responsive strains of the *E. coli* GFP library used in chapters 2 and 3 are of the OCS type and require inputs to diffuse or be transported across the cell membrane in order for them to bind to a cytoplasmic transcription factor and elicit a response. In this chapter, we describe the use of a small set of eight engineered *E. coli* to detect heavy metals in seawater (see Section 4.5.1).

The second family of bacterial sensors, TCSs, are one of the predominant channels by which bacteria senses extracellular signals, including phosphate, nitrate, and carbon sources [192]. They are composed of a tandem of proteins: a membrane-bound histidine kinase sensor and a cytoplasmic response regulator, with the first activating the latter upon external stimulus. TCSs can sense inputs outside the cell via their membrane-bound sensors. TCSs also hold the advantage of being modular, as signal input, (phospho)transfer and output response are separate processes [29, 193, 194]. In this chapter, we describe the use of a library of over 500 TCS *E. coli* sensing strains engineered by the Jeff Tabor lab [134] for the purpose of detecting both organic and nonorganic toxic compounds (Figure 4.1).

Our hypothesis for this work was that the inclusion of both OCS and TCS-based strains in our microfluidic biosensor array sensing array would allow for the continuous monitoring of seawater samples with broad sensing applications in a variety of fields.

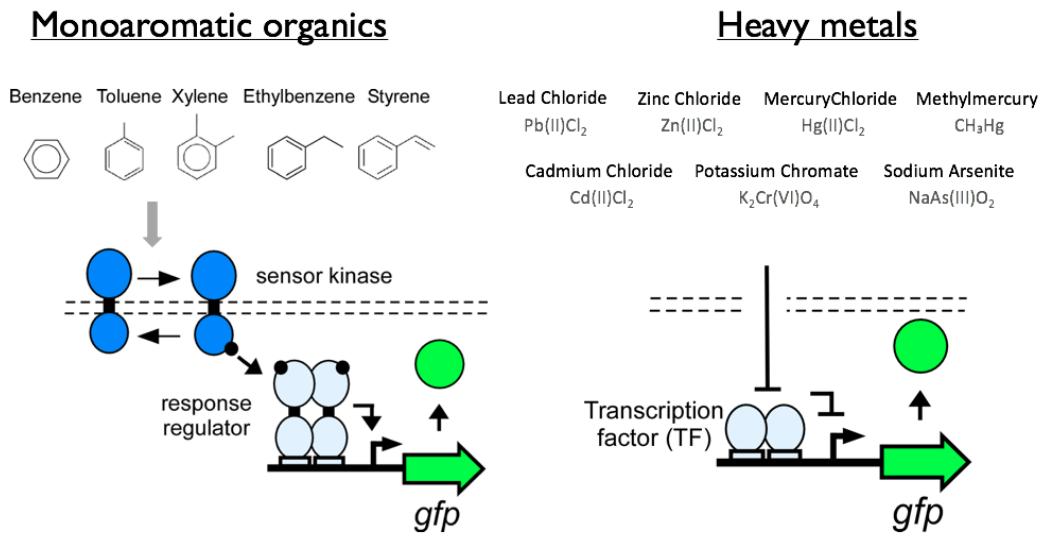

Figure 4.1: Schematic of sensing system for different types of contaminants of concern to Navy divers. Left : monoaromatic contaminants can be sensed at the extracellular level using TCSs. Right : heavy metals can be sensed using the OCSs due to the presence of metal importers and diffusion of metals across the *E.coli* cell membrane.

4.3 Results

4.3.1 Sensing heavy metals in seawater

We validated that *E. coli* sensor strains can grow uninterruptedly in seawater-containing media in our multi-strain microfluidic device (see Methods Figure 4.8). We grew up to 6 different *E. coli* biosensing strains (see Table 4.1) in either lysogeny broth or M9 minimal media containing 25% (filtered) seawater for up to 7 days. *E. coli* produced a noticeable physiological response to media augmented with 25% seawater: cell growth was impacted and cells tended to aggregate, which led to frequent clogging of our microfluidic devices Hence the upper limit of 7 days was not due to a decrease in cell viability, but to the clogging or our devices. Clogging occurs when large amounts of cells grew out of the cell traps and accumulate in inlet and outlet channels that are designed to be cell-free (see Figure 4.8). We chose to perform all subsequent experiments in a 10% seawater background to limit the effects of osmotic stress on cell growth. Although osmotic

Table 4.1: Toxins of interest for seawater biosensing, including inorganic and organic contaminants of concern, and the concentration at which chronic exposure is toxic to Navy divers. Table adapted from Sorensen, 2009 [186]

Toxin	Detection target
Arsenic (As)	0.01 mM
Cadmium (Cd)	0.02 mM
Mercury (Hg)	0.17 mM
Chromium (Cr)	0.89 mM
Zinc (Zn)	275.3 mM
Lead (Pb)	0.07 mM
Selenium (Se)	2.03 mM
Benzene	0.06 mM
Styrene	9.03 mM

stress at 10% seawater still had the effect of increasing the adherence of cells growing out of the trap to the PDMS channel walls, we were able to reliably perform experiments lasting over 10-25 days with cells growing in 10% seawater. These results indicate that seawater-containing media can be used for long-term culturing of MG1655 *E. coli* , with toxins added to the media during induction periods.

We performed quintuplicate inductions for 1/10, 1/50, 1/250, and 1/1000 dilutions of the target concentration for the metals (As, Cd, Hg, Cr, Zn, and Pb, see table 4.1) in three background media environments: regular HM9 (with Milli-Q water, see Section 3.3), media with 10% natural seawater, and media with 10% synthetic seawater, totaling over 600h of microfluidic induction data. Our OCS metal sensing sensor strains produced detectable responses for all metal COCs in all media conditions at 100 fold dilution of the target levels except for Zn(II). Zn(II) inductions proved to be toxic for the cells at any level above 1:10'000 dilution of target levels. Hg(II) and Cd(II) were detectable up to a 1000 fold dilution of EC concentrations. Table 4.2 gives the observed detection limit in 10% seawater.

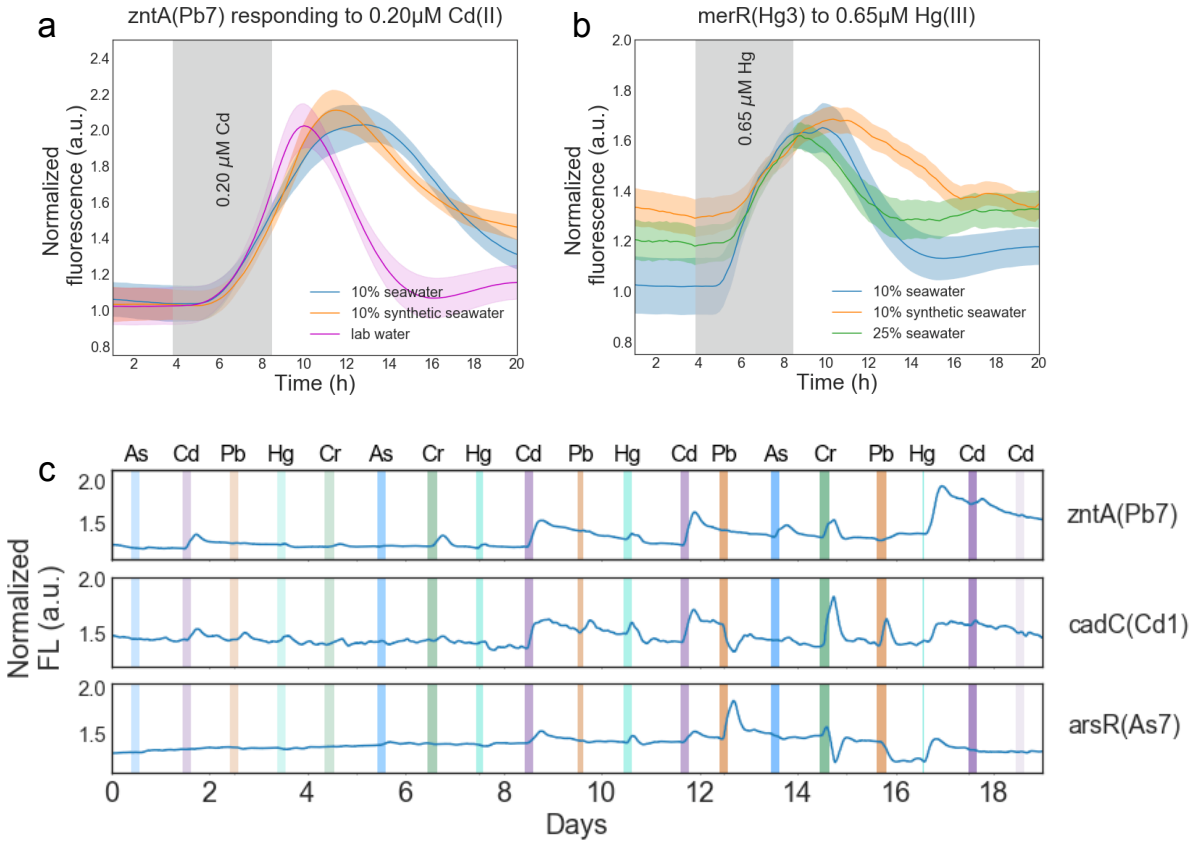

Figure 4.2: a: A *zntA*-based biosensing strain, Pb7, responding to 2μM Cd(II) (a 1:10 dilution of the target level) in media with either 10% seawater, 10% synthetic seawater or lab water (Milli-Q). b) Hg-sensing strain Hg3 responding to Hg at a 1:1000 dilution of the target concentration. GFP signal was background subtracted, normalized and smoothed. For each condition in a and b eight biological replicates are plotted, with the mean in dark and standard deviation above and below the mean in a shade of the same color. c) Three sensor strains (Pb7, Cd1 and As7) responding to a 4h heavy metal induction every 24 hours for a period of 19 days of continuous growth in microfluidics. Inductions are color-coded according to the heavy metal and shaded according to the dilution of the target level (lightest, medium and darkest shade corresponding to a 1:250, 1:50 and 1:10 dilution of target sensing levels for each metal respectively

OCS-based sensing strains produced similar responses to heavy metal inductions performed in synthetic seawater and natural seawater. Differences were more pronounced between the inductions performed in seawater and Milli-Q, which may be attributed to the effects of osmotic stress on the engineered strains' response. Osmotic stress has previously been reported to affect cell volume [195], cytoplasm and cell membrane properties [188] and gene expression [196].

Table 4.2: Detection limits for seawater contaminants and the sensing strain with which it was achieved. The detection limit corresponds to the lowest concentration tested for a given COC that produced a significant variation in the fluorescence signal of the apposite sensing strain. The fold change is the ratio of the detection target to the detection limit. A high fold change indicates a signal was detected at levels below the target concentration.

COC	*E. coli* sensor strain	Detection limit in seawater	Fold-change below target
Arsenic (As)	As7	2.0 μM	5
Cadmium (Cd)	Pb7	0.2 μM	100
Mercury (Hg)	Hg3	7.0 μM	23
Chromium (Cr)	Cr14	88.5 μM	10
Zinc (Zn)	Pb7	2750.0 μM	100
Lead (Pb)	Pb7	0.3 μM	248
Selenium	TCS Library	400.0 μM	5
Benzene	StySR	N/A	N/A
Styrene	StySR	76.0 μM	119

An overview of the data collected for the seawater experiments is given in Figure 4.2 which examples of single-strain response to 4 hours inductions (panels a and b), and repeated inductions of three strains are shown during 19 days of continuous sensing with one heavy metal induction per day (panel c). The long duration over which we cultured MG1655 supersedes those previously reported in the literature. While *E. coli* has previously been used to sense contaminants in 25% seawater with no negative impacts on growth during 2-4 hour experiment, [190] we believe this is the first instance of continuously exponentially growing microbes housed in microfluidic devices being used to detect multiple contaminants in seawater continuously and for days.

Our results are further differentiated from previous seawater-sensing efforts by the multiplexing abilities offered by our microfluidic setup on which are up to 48 strains. Exposure to each toxin produced a unique "fingerprint" of responses by the array of 48 fluorescent bacteria, allowing for finer discrimination between contaminants. We illustrate the breadth of our sensor's coverage in Figure 4.3 displaying a heat-map of five different strains responding to the 10X dilution of the target sensing levels for five metals. Fig4.3 reveals the existence of cross-talk

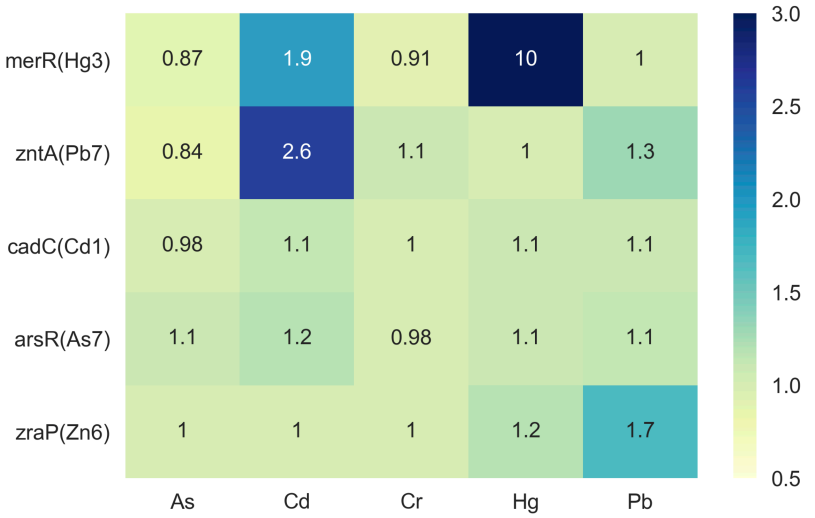

Figure 4.3: c) Heat-map of fold change over pre-induction baseline for 5 strains (rows) responding to 5 heavy metals (columns) at levels 10 fold below those deemed toxic to Navy divers by the Sorensen toxicology group. Metals can be distinguished based on the pattern of responses they produce in the five sensing strains. Fold-change is computed by dividing the peak of a strain's fluorescent signal by its fluorescence level prior to the induction

between metals. For example, the *merR* gene (present in strain Hg3 c.f Table 4.2) responds to both Hg(III) and Cd(II). This is not surprising in itself - the literature shows that metalloregulatory protein often responds to a subset of metal ions [140], but illustrates the limitations of single-strain biosensors. Given cross talk, a comprehensive fingerprint of gene expression change across multiple sensor strains allows for accurate discrimination between metals.

4.3.2 Sensing styrene using a TCS biosensor

We cultured StySR, a TCS-based sensor strain on our microfluidic platform with the aim of detecting monoaromatic compounds (see Table 4.1). StySR is a sensing strain engineered by the Tabor Lab at Rice University to detect Styrene and Benzene that uses the sty promoter of *Pseudomonas sp*. [197] (see Figure 4.7).

We used a syringe-pump-based set-up to ensure all the media-containing syringes were

hermetically closed to avoid any volatilization of monoaromatics from the media which would generate both safety concerns and prevent StySR from detecting the target compounds. We performed triplicate 4h styrene inductions in M9 media at a concentration of 76 µM (8mg/ml), resulting in a quantifiable and reproducible fluorescent response (Fig. 4.4). Time-trace analysis of the styrene induction on our microfluidic platform showed that StySR produced a fluorescence fold-change that was comparable to results produced by the Tabor lab's flow cytometry characterization of the strain's styrene response. These results validate our platform's ability to use TCS-based sensors to detect volatile organic compounds in addition to one-component systems for the detection of heavy metals.

Compared to the detection of heavy metals, the detection of monoaromatics in our microfluidic proved to be difficult due to the cumbersome setup needed to pressurize the fluidic circuit and avoid any volatilization of the targeted compounds. This significantly increased the time to set up each experiment and increased their failure rate. While microfluidics represents a powerful tool way to perform biosensing, the use of microfluidics to sense volatile targets in a high-throughput fashion presents itself with a unique set of challenges that may prevent widespread use without further engineering efforts.

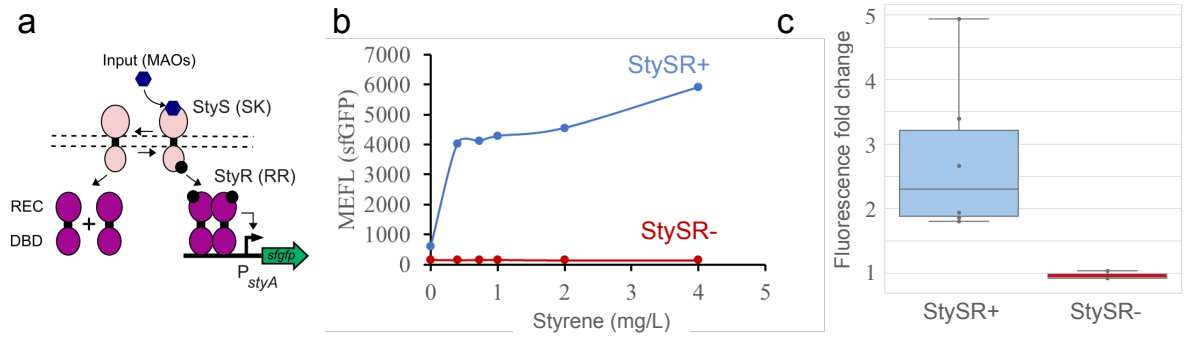

Figure 4.4: Detection of styrene using the StySR system. a) TCS system used for Styrene detection. b) Response curve of the SySR strain measured by fluorescence cytometry. c) GFP response of StySR over time in a microfluidic setup. Fold change compared to control is calculated using data acquired from a 76 µM (8mg/ml) styrene inductions across 3 replicates and compared to the negative control strain.

4.3.3 Sensing using a library of engineered TCS biosensor strains

We interfaced a library of over 500 TCS sensors with our multiplexed microfluidic platform to leverage the sensing potential of microbial TCS systems and generate rapid, high-throughput biosensing data for a wide range of contaminant. The goal of this approach was to identify contaminants by interpreting the pattern created by the responses of the entire library exposed to a sample. In this section, we present the work accomplished to integrate the TCS library with a two-thousand strain microfluidic device and the detection of two contaminants that serve as a proof of concept for this "synthetic olfaction" approach.

We received a library of TCS strains which were engineered by the Tabor lab according to the methods described in Schmidt et al. [134]. The library contains 521 genetically engineered *E. coli*, each containing an uncharacterized TCS identified by the Tabor lab by mining gut bacterial genomes available on the NIH Human Microbiome Project. In each sensor strain, the TCS is rewired to synthetic output promoters driving *sfGFP* via a DNA-binding domain. Thus, each strain in the library is capable of producing *sfGFP* in the presence of inputs that activate their TCS system.

Contrarily to the OCS-based heavy metal sensing strains whose behavior had been previously characterized in our lab (see Section 3.1), the Tabor library strains had not been tested in microfluidic devices. We used a biopixel-type cell trap with hydrodynamic-trapping described in Chapter 2 (see 2.12) to culture the TCS library strains in either 48- or 2176-strain microfluidic devices. We loaded cells onto our microfluidic devices using either the acoustic droplet ejection or pinpad-based robotic spotting techniques described in Chapter 2.3. While the growth of the strains was more robust on LB media, we found that M9 minimal media improved the detection of the strains' fluorescent response due to its lower autofluorescence. However, we also found that M9 did not robustly support the exponential growth of the entire population of cells in each microfluidic trap, creating variability in the observed response profiles. This can likely be explained by the diffusion-limited access to nutrients for cells residing deep in the biopixel

trap and far away from the media-supplying channels [120]. Hence we pivoted towards using EZ rich media for TCS library experiments, a media which is defined, rich, and possesses low autofluorescence [198].

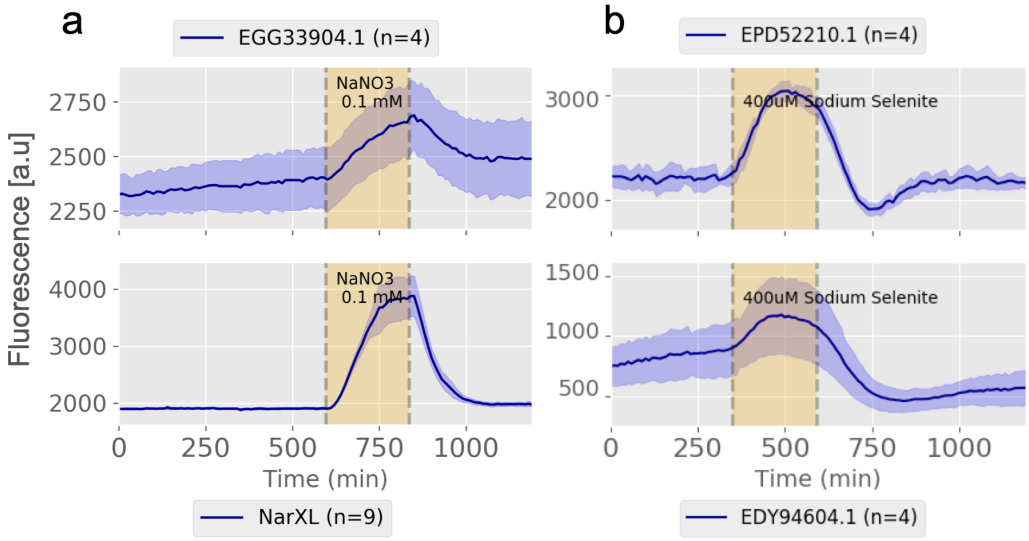

Figure 4.5: TCS library strains responding to Nitrate and Selenite. Time-series response of the three uncharacterized TCS strain and one positive control strain (NarXL) when induced with a) 0.1 mM Sodium Nitrate for 4h (yellow shaded region) or b) 400 µM Sodium Selenite. The mean response across 4 to 9 replicates is shown in dark blue, with the shaded blue area representing plus/minus one standard deviation

In order to validate our TCS library pipeline, we relied on the existence of the library's internal positive control strains (NarXL), which can detect Nitrate [29]. We induced the entire TCS library with 0.1 mM Sodium Nitrate and observed a clear response of NarXL. Moreover, we also detected another nitrate-sensing strain, EFF33904.1 (see Fig. 4.5). In Figure 4.5 we also report clear responses from two strains in response to Sodium Selenite, one of the toxic compounds initially identified as being of particular interest to seawater biosensing (see Table 4.1). In addition to using single strains to detect COCs, the response profiles of multiple strains can be leveraged using machine learning algorithms, as previously demonstrated in Chapter 3 [79]. An example of an approach that uses unsupervised machine learning techniques to visualize

the library's response to an inducer is displayed in Fig. 4.6, which shows a PCA biplot of the 521 TCS library strains exposed to 0.1 mM Sodium Nitrate. PCA is a commonly used method of dimensionality reduction that has been extensively applied to genomic and transcriptomic data [199]. It uses linear combinations of variables to generate orthogonal axes that efficiently capture the variation present in the data with fewer variables. In this plot, each point represents an aggregate view of the library as it progresses through the different stages of the response to the induction and shows that multiple strains contribute to the overall change in the library's time-dependent GFP expression profile.

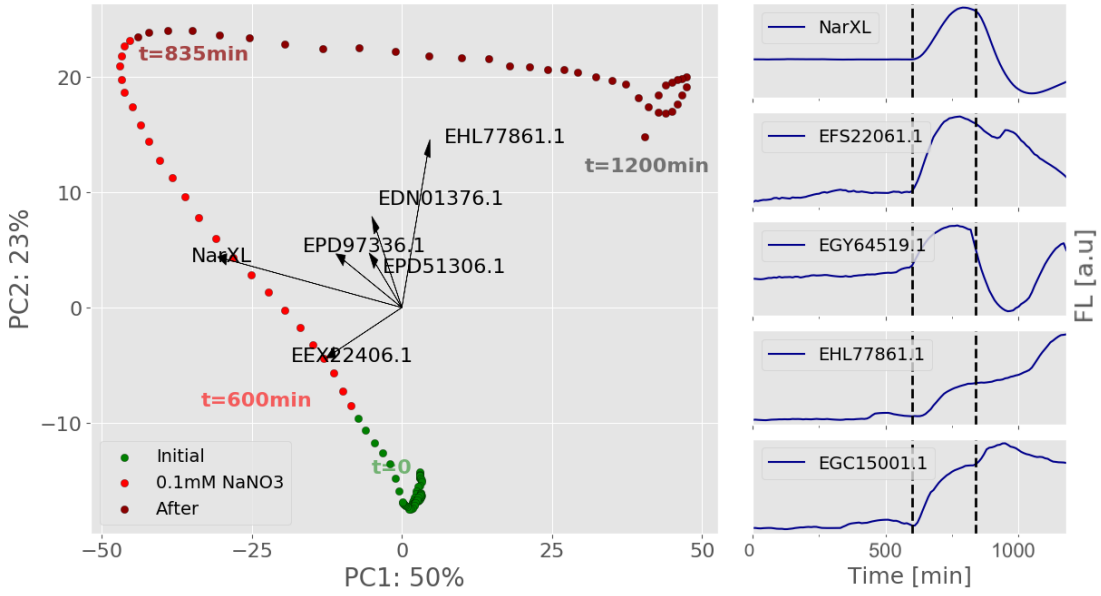

Figure 4.6: PCA of TCS library response to Nitrate Left: PCA bi-plot of TCS library responding to 0.1mM NaNO3. The bi-plot represents the scores of 119-time points in the PC1/PC2 plane. The arrows represent the loadings for the strains with the highest loadings in PC1/PC2 (i.e which strain contributes the most to the first two PCA axis, with the arrows showing in which "direction" a strain exerts traction in the PC1/2 plane). We used the first derivative of the smoothed, mean-centered GFP expression data, with each strain representing the mean of 4-10 replicates of each strain. Right: Time-series response of the five strains with the highest loadings in PC1/PC2

4.4 Discussion

Field-deployable technologies that can detect toxic compounds in maritime environments are necessary to adequately monitor seawater quality for environmental and human health purposes. We have developed a novel biosensing platform based on engineered sensor bacteria, microfluidics, and "synthetic olfaction" to detect chemical contaminants of concern in seawater. We have used a suite of engineered bacterial strains capable of producing a fluorescent response to specific chemicals (i.e. heavy metals and monoaromatic compounds) while being exposed to seawater. Multiplexed microfluidic and optical technologies were combined to culture the bacteria and monitor their fluorescence response over long periods of time. Live bacteria and the accompanying hardware were combined in a field-deployable "sensor box" capable of analyzing real seawater samples in the field.

4.5 Materials and Methods

4.5.1 Strains and plasmids construction

The heavy metal bisoensing strains used in this study (see Table 4.3) were previously developed by our group for heavy metal sensing in drinking water [200], except for Cr14 which was created for this study. All strains were engineered by using known response pathways that mediate resistance to heavy metal in various microorganisms [201–205]. As7 uses *arsRp* from the *E. coli* R773 plasmid to drive a synthetic operon encoding both *arsR* and sfGFP [206]. To further increase signal we used the strong Lutz RBS to enhance sfGFP [207]. To detect cadmium, we use two strains, named Pb7 and Cd1. The construct in Pb7 is composed of the *E. coli zntA* promoter (*zntAp* which was identified using RNA-Seq) driving Lutz-sfGFP. The Cd1 construct uses *cadCp* to drive a synthetic operon of *cadC* and sfGFP using a native *E. coli* RBS. The *cadC* components are derived from the Staphylococcus aureus pi258 plasmid [208]. To detect lead, we

again use Pb7 along with another strain called Zn6. Zn6 has *zraPp* (identified using RNA-Seq) driving Lutz-sfGFP. For mercury, we use the Hg3 strain, which contains *merRp* driving *merR* and Lutz-sfGFP with the *mer* components originally derived from Shigella flexneri 2b [201, 204]. For chromium, we use a strain known as Cr14. This strain has sulAp (identified using RNA-Seq) driving Lutz-sfGFP.

Table 4.3: *E. coli* biosensing strains for the detection of heavy metals in seawater

Strain name	Gene description	Sensing target	Secondary sensing targets
As7	arsenate resistance	Arsenic (As)	Cd, Hg
Cd1	cadmium resistance	Cadmium (Cd)	Hg
Hg3	mercury resistance	Mercury (Hg)	Cd
Pb7	zinc/cobalt/lead efflux system	Lead (Pb)	Cd, Hg
Zn6	zinc homeostasis protein	Zinc (Zn)	Hg
Cr14	cell division inhibitor	Chromium (Cr)	N/A

The StySR TCS biosensor strain engineered by Moshe Baruch from the Tabor lab. The StyS/StyR TCS is ported from *Pseudomonas sp.*, with the sensor kinase placed on one plasmid and the response regulator (RR) on the other. The RR drives the expression of GFP,

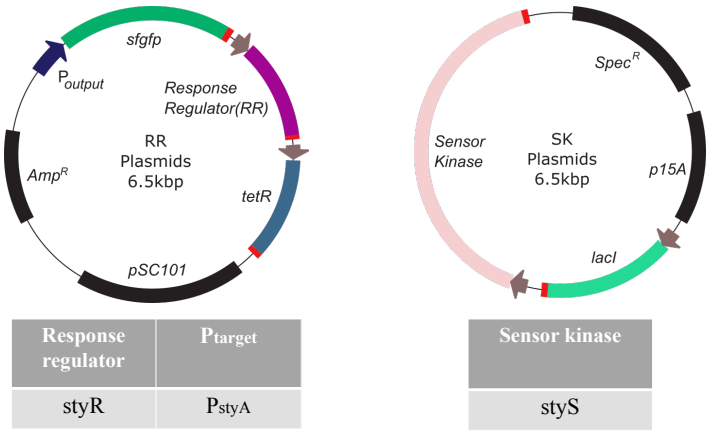

Figure 4.7: Design of the expression plasmids for the TCS StySR system

4.5.2 Microfluidic sensor

Sensor design: Our biosensor device, capable of culturing *E. coli* continuously in seawater-containing media, is an evolution of microfluidic devices previously designed for urban water biosensing (see Chapters 2 and 3). Our device consists of a single media inlet and outlet, and a manifold geometry to mitigate the risk of clogging, a potential limiting factor when using environmental samples. In this design, the single media inlet feeds 8 columns and 6 rows of cell traps, which are 4μm tall and receive fresh media from 150 μm tall major channels that split off into 50µm tall minor channels. Waste from each cell trap flows into a common major channel (the manifolds, see Fig.1). This device geometry ensures each cell chamber region is fluidically isolated to prevent strain cross-contamination. Spacing between the reservoir regions is 1.125 mm for compatibility with standard SBS-format 6144-density spacing for robotic loading. The trap has a "light bulb" shape, with the area of the trap furthest away from the minor channel possessing a disk-shaped geometry with a large 500 μm diameter to receive the cells loaded by the cell loading robot. As cells grow out of the spotting area, they reach four 100µm by 90 µm regions named the "biopixels" which are 4μm tall. These are the areas in which cells are constantly supplied with fresh nutrients from the minor channels and from which fluorescence data is extracted. Having four biopixels associated with each strain loading region allows us to collect more data for each strain. Results: We have evolved a 48-strain microfluidic device that can be loaded with robotic loading techniques which builds on previous designs that have shown robust biosensing performance with urban water samples. Our device allows engineered *E. coli* to grow in a controlled microenvironment for long periods of time in seawater-containing media. The device geometry has been optimized to limit clogging and facilitate operator use.

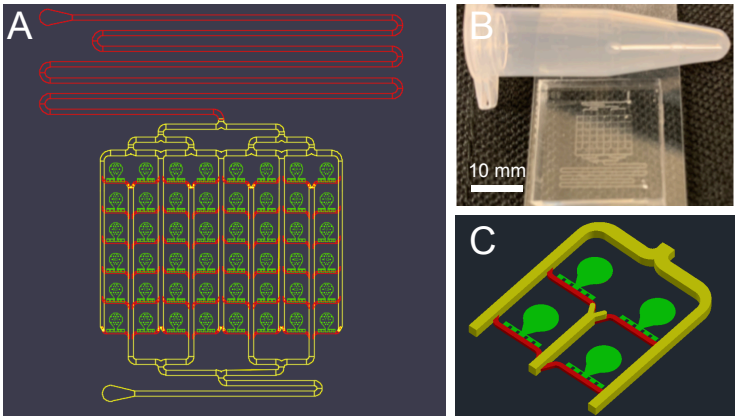

Figure 4.8: A) CAD drawing of 48-strain microfluidic device. B) PDMS microfluidic device with microcentrifuge tube for reference. C) Closeup of 4 cell traps. The major channels supplying nutrients across the device are in yellow (150 μm tall). Minor channels feed (red, 50μ m tall) individual cell traps (green 4μ m tall). Cell traps possess a round geometry to accommodate cell loading and four biopixel rectangles for cell growth and imaging.

4.5.3 Setting up a microfluidics experiment

Sensing of heavy metals in seawater, synthetic seawater and Milli-Q water at various concentrations. We loaded our biosensor device with 8 replicates of 6 engineered strains responsive to heavy metal COCs presented in Table 4.3.

Each microfluidic device was checked carefully prior to loading the media to ensure no dust or broken features were present. Before setting up a microfluidics experiment, devices are placed in a vacuum for 20 min to facilitate wetting. Loaded devices are wetted with Lysogeny Broth (LB) growth media, which is continuously flown through the device via hydraulic head pressure difference until the cell traps (the regions of the device where the cells are imaged) become full. Cells are grown on LB media until they reach confluence in their parallelized microchambers. This process takes about 12h (see Fig. 2.12). The media is then switched to HM9 minimal media (see Tab. 3.3) containing 10-25% of filtered water. Two media ports were connected to plastic PTFE #24 (Cole Parmer) tubing, which connected to 60 ml syringes with the inlet syringe containing fresh medium at a height of 30" at the outlet syringe at 5" above chip level. Flow-rate was about 1ml/hour for 48-strain devices and 10ml/hour for 2176-strain devices.

Devices were imaged in our sensor boxes and exposed daily to COCs for 4 hours before returning to uncontaminated media for 20 hours. The un-induction period allows the strains whose fluorescence output increases in response to toxins to return to their pre-induction baseline. Using this method, we conducted 5 replicate toxin inductions at 4 concentration levels for 5 out of the 6 target heavy metals (As, Cd, Cr, Hg and Pb). Zn inductions were found to lead to cell death at the tested levels. The concentration levels chosen for inductions corresponding to 1X, 5X, 25X, 100X dilutions of the contaminant levels considered to be toxic for chronically exposed divers by the Kara Sorensen's Navy toxicology group (Table 4.1). Since we supplemented our growth media with 10% seawater to limit the levels of osmotic stress, the actual levels the biosensor strains were exposed to were 10X, 50X, 250X and 1000X dilutions of the Sorensen concentrations. Seawater was obtained from the seawater tap at the Scripps Institute of Oceanography (La Jolla, California) and filtered using Nalgene sterile disposable bottle filters with 0.2 μm SFCA membrane prior to being mixed with concentrated media (either HM9 as described in Table 3.3 Or EZ rich media [198]. In addition, we performed duplicates of the same inductions in synthetic seawater (Ricca Substitute Ocean Water without Heavy Metals ASTMD1141) and single inductions in reference lab grade-purified water (Milli-Q). Over 150 inductions were performed in total.

4.5.4 Data extraction pipeline and data processing.

One of the advantages of using engineered bacteria for biosensing purposes is the rapidity of their response to environmental toxins (in the order of 10min). To capture this rapid response we acquire fluorescence images of the bacteria growing in our microfluidic device every 5 minutes. To translate the pixel value of the bacteria's fluorescence levels in the biopixels into a meaningful response trajectories, we use a previously developed automated software pipeline [79]. This pipeline is composed of an image processing and data extraction module, and a large MySQL database to store the processed data. Wireless communication between the sensor box and the database enables data to be stored and accessed remotely.

As with many biological measurement methods, one of the important aspects of microfluidics-based experimentation is finding appropriate techniques to process the raw data. In the case of experiments carried out in parallel on multiple sensor boxes, there is the added challenge of different boxes having slightly different background fluorescent values. In addition, most sensor experiments experience some global fluorescent drift as a result of the maturing of the strains and their adjustment to growth or induction media, especially when sample-to-sample variation exists between the seawater samples added to the media. To mitigate those effects we have explored different data normalization techniques and have developed a three-step processing technique depicted. First, we subtract the local background signal to cancel out local fluctuations across the PDMS device. Second, we standardize each cell trap's fluorescent response with respect to itself. Finally, we smooth the signal using Holt-Winters exponential smoothing. This process helps standardize the data within experiments (across biological replicates on-chip) and across experiments (same induction protocol applied across experiments and sensor boxes).

4.5.5 Syringe system used for styrene inductions

Eliminating the headspace in media syringes is crucial for performing inductions with monoaromatics (MAOs) such as styrene as they are highly volatile and will diffuse into the surrounding atmosphere if they are not trapped. Diffusion of MAOs out of solution not only poses a potential safety hazard, but also impacts the reliability of the device as diffusion of MAOs would affect the concentration sensed by the cells during the course of an experiment or field test. In our protocol, we use a syringe pump to eliminate headspace in syringes containing MAO inducer. Specifically, we prepare syringes with defined concentrations of MAO in M9 media by mixing the MAO with M9 media and drawing the solution into a syringe in a fume hood. The capped syringe is then affixed to a syringe pump and connected to a valve that feeds to the microfluidic device using plastic tubing. At the start of an experiment, during the cell grow-up phase, the valve is positioned so that regular LB or M9 media flows through the device from an uncapped

syringe via gravity-driven flow. Once cells have filled the trapping regions and the fluorescent signal from the cells has stabilized, an induction is performed by turning on the syringe pump and switching the valve so that media from the MAO-spiked syringe can flow into the microfluidic device. The flow-rate of the syringe pump is set to match the flow-rate used during the cell-grow up phase where hydrostatic pressure (rather than a pump) is used to maintain flow. Additionally, the syringe connected to the waste port of the device is capped with a plunger to prevent MAO contamination into the surrounding air for safety purposes. After an induction is complete, the valve is switched back to the syringes containing normal media, and the syringe pump is turned off. Using this system, we were able to introduce relevant concentrations of styrene in M9 media into our microfluidic platform and measure a fluorescent response from the *stySR* sensing strain. Although this protocol allowed us to safely introduce styrene containing media into our device while essentially eliminating any headspace for volatilization in the source syringe, the number of fluidic connections and moving parts in this system made it somewhat prone to the formation of air bubbles in the microfluidic device.

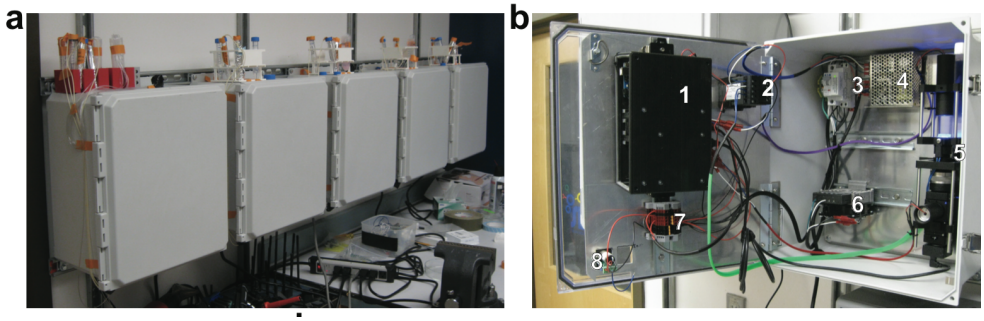

Figure 4.9: Custom sensor box used for biosensing in the field: a) Device enclosure, note temperature controller (upper left) and peristaltic pump (lower right) mounted to the aluminum front panel. b) Enclosure with front panel opened to expose internal components. b1: Electronics sub-enclosure, b2: temperature controller, b3: AC power distribution devices, b4: tri-output DC power supply, b5: Ziva optical assembly, b6: fan/heater, b7: DC power distribution block, b8: peristaltic pump.

4.5.6 Sensor box

We used previously-developed custom ”sensor boxes” to image our multi-strain microfluidic devices. Senor boxes contains the hardware and software needed to sustain maintain a constant temperature for cell growth, handle image acquisition and perform data processing. The custom optical system using low-cost, off-the-shelf components to image 48 separate microbial strains within a single field of view in both transmitted light and GFP fluorescence channels (this number can fluctuate based on the MFD design). The enclosure's environmental control is handled by a fan-heater capable of maintaining optimal *E. coli* growth temperature of 37°C over an ambient range of at approximately 25°C – 33°C.

4.6 Acknowledgements

This chapter was co-authored with Alyssia Chiang, Andrew Lezia, Bryan Thai, Michael Ferry, Moshe Baruch and Jeff Hasty. The dissertation author was the primary author and researcher of this material.

Chapter 5

Investigating the dynamics of *S. cerevisiae* proteome dynamics in response to the anti-aging drug metformin

5.1 Introduction

5.1.1 *S. cerevisiae* as a tool for chemogenomic screens

S. cerevisiae, commonly known as brewer's yeast, is a key model organism for eukaryotic systems and synthetic biology. It offers great experimental tractability with a short generation time, economic culture requirements, and the availability of numerous mutant libraries [209]. *S. cerevisiae*'s genome is well-characterized: a large number of its pathways have been mapped, and many of those are conserved in higher eukaryotes. *S. cerevisiae*'s pathways are therefore relevant for studying processes tied to aging and disease (e.g. DNA repair mechanisms, mitochondrial homeostasis, protein folding, and stress response) [209]. As a result, *S. cerevisiae* has been extensively used in studies that map genome-wide response to environmental, nutrient, and chemical perturbations, including anti-aging drugs [210]. By identifying the main effects of

these perturbations on a genome scale, these studies have furthered our understanding of cellular metabolism, signaling pathways, and disease mechanisms [61, 211]. In this space, cell-based assays that have both genome-wide coverage and the capacity to capture the dynamics of pathways would offer an advantage of existing screens [212].

We present here an experimental strategy to map out target pathways of drugs, based on gene expression dynamics of over 2500 different proteins in yeast.

5.1.2 The drug metformin and its effect on *S. Cerevisiae*

Metformin is an antiglycemic diabetes drug initially discovered in 1921 and used in patients with type II diabetes to decrease blood glucose levels [213]. Recent studies have pointed at additional health benefits of metformin treatments, including anti-cancer and anti-aging effects. These studies span multiple organisms: metformin treatment improved the healthspan of *S. Cerevisiae* [213], *C. elegans* [214], rats [215] and humans [210, 216, 217]. Yet, we possess an incomplete understanding of the targets and mechanisms through which metformin acts [218].

A couple of studies have been conducted on the *S. Cerevisiae* response to metformin, but all of them were performed using either static assays (microarrays [213] and agar plate-based protein-fragment complementation assay [212]) or using low-throughput fluorometric plate reader assays [219]. Metformin response in yeast has some important features, which are conserved across these experimental paradigms and show a broad effect of the drug on cellular metabolism. I summarize these effects below.

Energy metabolism

According to previous *S. Cerevisiae* screens, Metformin exerts an important effect on energy metabolism and glucose metabolism in particular. Metformin was observed to produce a transcriptional response similar to glucose deprivation [213], with cells exposed to the drug relying more heavily on aerobic respiration and ethanol fermentation, while simultaneously stimulating

glucose transport and the AMPK pathway [212, 213]. Additional observations have show that metformin treatment leads to discrepancies in electron transport, indicating that metformin targets mitochondrial function [220].

Iron-binding proteins

It has been proposed that the metformin-induced disruption in mitochondrial activity mentioned above may stem from the metal binding property of the compound [221]. Additional studies have determined that metformin interferes with iron distribution in the cell, leading to an iron-deficiency-like state. Stynen *et al.* [212] found that iron-binding proteins (e.g those including iron–sulfur clusters) showed a reduced activity while proteins that counteract iron depletion (e.g ion transport machinery) have an increased activity [220]. Li et al. found that metformin suppressed the production or iron-binding heme [220]. The effect of metformin on intracellular iron levels and iron-related proteins is coherent with its effect on central metabolism and may play a direct role in its life-span increasing effects [212, 222] via a hypoxia-type response and respiratory uncoupling [213].

Additional effects

Additional effects of metformin on *S. cerevisiae* metabolism and signaling pathways have been reported across the literature.

- Metformin was shown to impact the protein synthesis machinery (including ribosome biogenesis and translational processes), an effect that is generally attributed to deficiencies in the TOR signaling pathway [212, 213].

- Metformin leads to a transcriptional down-regulation of sulfur-containing amino acid, especially methionine. It is worth noting that a decrease in cellular methionine levels was proved to delay aging in multiple organisms [223]

- Metformin seems to induce filamentous growth signaling pathways, lipid and phosphate metabolisms [213], purine biosynthesis and DNA repair [212, 213], the latter being of particular interest for the use of metformin in cancer therapies.

Overall, the literature on metformin effects in yeast highlights the broad effects of the drug and the interconnectedness of the cellular functions that it affects, emphasizing the need for high-throughput analysis.

5.1.3 Approach

Given the development of an *S. Cerevisiae* - compatible Dynomics device (see 2.2.6), there was an opportunity to apply it towards chemogenetic screens by using the existing *S. Cerevisiae* GPF library [144], a counterpart to the *E. Coli* library used in Chapters 2 and 3 . This library has been used in a variety of studies to characterize protein function, abundance, and localization using high-throughput microscopy [84] and a microchemostat array [110]. Each member of the *S. Cerevisiae* library has the GFP gene fused to a known ORF. Expression driven by the endogenous ORF promoter generates a full-length protein with a COOH-terminus GFP fusion. The library covers 75% of the *S. Cerevisiae* proteome (4100 proteins represented out of 5797 identified) in that manner [144], making it one of the most comprehensive libraries available for proteome-wide screens.

We are able to load the library on the Dynomics platform by splitting it into two 2176-strain devices. We can then subject each half to the same environmental conditions, generating proteome response dynamics with a 10 min resolution in real-time. This approach can greatly complement existing transcriptomic and proteomic methods used to characterize the *S. Cerevisiae* response to environmental stimuli, including the global cellular to drugs such as metformin. We used our platform to characterize the monitor the proteome response to 10h-long metformin inductions and to compare it with iron deprivation inductions of the same duration. This approach yields high temporal resolution data of the protein expression dynamics that occur during these

environmental stimuli. While existing studies have shed some light on the intracellular processes affected by metformin, these previous approaches did not yield rich dynamics-based information on the effect of the drug at a genome-scale. High-frequency time-course sampling of gene expression of the metformin response may help distinguish between transient and sustained expression patterns and, ultimately, lead to a better understanding of the gene regulatory networks underpinning the global cellular response to the drug [125].

5.2 Results

5.2.1 Quantitative real-time observation of response to metformin

The Dynamics-based metformin screen revealed 1115 and 1094 proteins with an increased or reduced signal, respectively, out of the 4149 unique strains in the *S. cerevisiae* GFP library. The response of the 2209 strains is displayed as a heat map in Figure 5.1. The rest of the strains produced a signal that was equivalent to the non-fluorescent control strain and were discarded from the analysis (see Methods section 5.4.2). There was no correlation between the abundance of the protein and the direction (up vs. down-regulation) in the metformin screen. Signal upregulation and downregulation was contained within +40% to -25% of the mean strain behavior, a reminder that measured fluorescence in library strains is a proxy for protein number, and that our measurements are principally useful to measure the change in protein number rather than quantification of protein number.

The response from the library is also displayed as a volcano plot in Figure 5.2 which displays the significance vs. fold-change for each gene and is extensively used in omics experiments. Although volcano plots provide an overview of the responses to metformin and iron deprivation, a more detailed analysis is required to understand the cellular response.

A large number of the key pathways previously reported being affected by metformin (glucose metabolism, amino acid synthesis, lipid biosynthesis - see 5.1.2) show modified activity

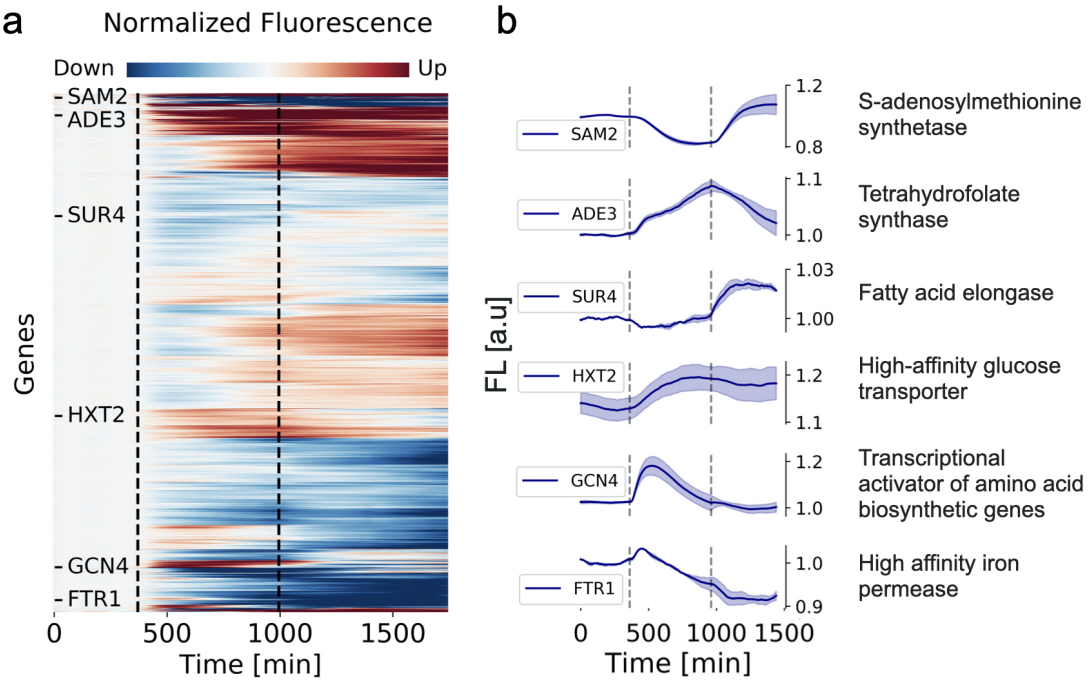

Figure 5.1: Time-series response of *S. cerevisiae* GFP library to metformin exposure. Relative intensity of 2720 strains exposed to metformin hydrochloride. a) 2720 strains loaded on two dynomics microfluidics platform are induced with 100mM Metformin for a period of 10 hours indicated by the dotted black line. Each row represents the normalized fluorescence of a single. b) Six genes with noticeable responses from panel a) are highlighted, with each corresponding strain's mean fluorescence plotted in dark blue. One standard deviation above and below the mean is indicated in a lighter shade of blue. The role of the protein encoded by each gene is indicated on the right

.

upon metformin induction on Dynomics. There are, however, also important departures from what was previously reported in the effect of metformin on iron-binding proteins and purine biosynthesis. These differences may be the fruit of the different experimental paradigms used to study the effect of metformin, but also due to the dynamics-based information that we report here for the first time. We summarize the overlap and discrepancies between our data and what was previously reported in the literature below.

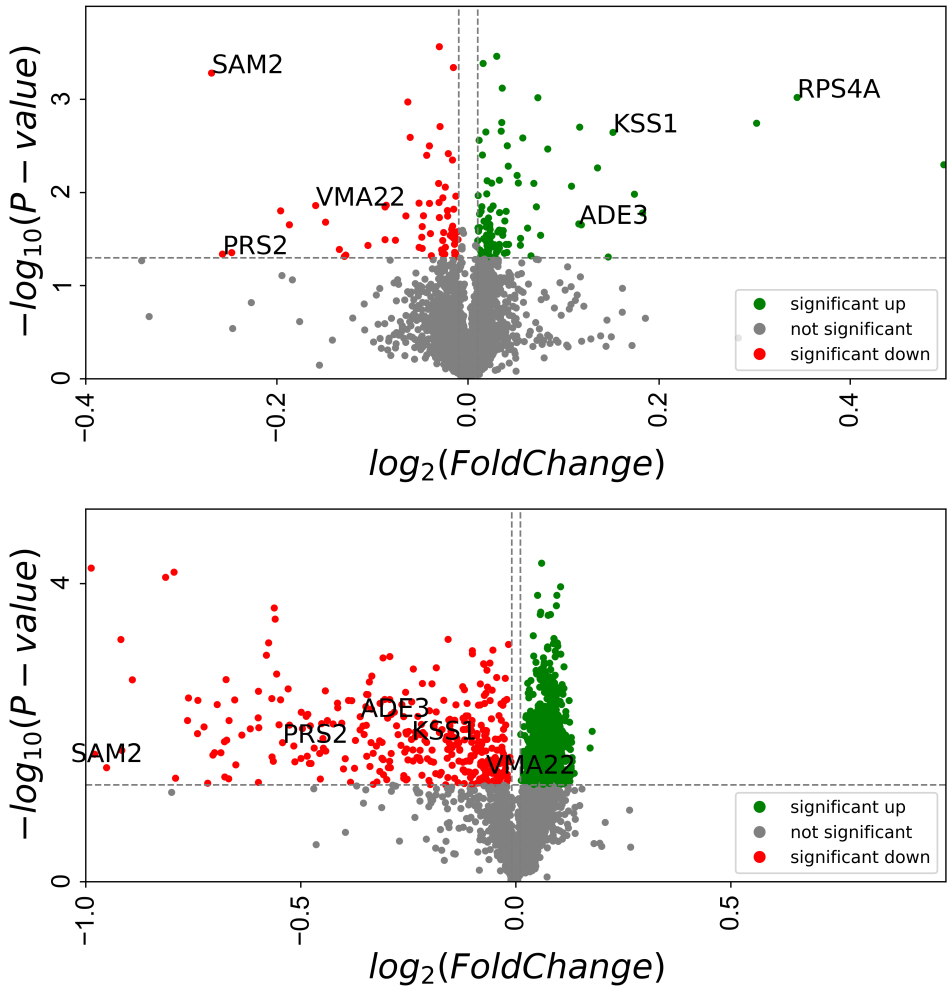

Figure 5.2: Volcano plot of the *S. cerevisiae* GFP library exposed to Metformin

.

Energy metabolism

The response of the *S. cerevisiae* GFP library growing on Dynomics to drug-supplemented mediums shows that central glucose metabolism is impacted, a well-documented effect [212,213]. Glucose transporters are upregulated (HXT2 is shown in Fig. 5.1) as are TCA cycle and respiration enzymes (the fumarate hydratase FUM1 is amongst the top 20 most upregulated proteins, as measured by fold-change). We also observed disruption of mitochondria-associated proteins with MMR1 and YTA12 amongst the 20 most significantly down-regulated proteins in the assay while

ACN9 (a mitochondrial protein involved in the assembly of succinate dehydrogenase) is amongst the 20 most strongly upregulated proteins. This confirms the impact of metformin on the electron transport chain and ATP synthesis [220].

Iron-related genes

The effect of metformin on iron-binding proteins in our assay did not map as closely to what was previously reported in the literature. While iron transporters FET3, FTH1, and FTR1 showed increase activity in response to metformin in the protein-fragment complementation screen [212] and microarray assays [213], we observed little effect of metformin on FET3 and FTH1, and down-regulation of FTR1 on Dynomics. Interestingly, FTR1 experiences an initial burst of upregulation in the hour following metformin exposure in our assay (Fig. 5.1 bottom right), before decaying to lower-than-baseline levels.

Proteins with iron-sulfur complexes exhibited behavior that overlapped much more closely with what was reported using classical screening techniques. ISA2, which acts in the mitochondria, was upregulated, while RLI1, which lives in the cytoplasm and is required for ribosome biogenesis, was down-regulated. We note that other protein synthesis machinery genes (RPL5, YTM1) were also strongly downregulated, hinting at an important role of metformin on translation. Suppression of translation has been reported to be a common trait of quiescence, which lengthens chronological life span [213].

Amino acid and Lipid biosynthesis

Previous studies have demonstrated that metformin treatment leads to transcriptional suppression of sulfur-containing amino acids, especially metformin [213]. In addition, phenotypes involving decreased methionine levels have also been shown to delay aging in yeast, mouse and human cells [224]. We observed significant downregulation of methionine-associated genes in our assay. SAM1 and SAM2 were strongly down-regulated genes (see 5.1 top right), concurrent with

their down-regulation in the protein-fragment complementation screen from Stynen et al. [212]. We also observed downregulation of the arginine biosynthesis pathway, notably ARG1.

Purine nucleotide biosynthesis

A strong signal from our data was the upregulation of purine nucleotide biosynthesis. ADE2, ADE3, ADE4, ADE5-7, ADE6, ADE12, and ADE17 were all upregulated, with ADE17 the most-strongly upregulated across our data. Surprisingly, both the microarray-based and protein-fragment complementation studies of yeast response to metformin measured repressed purine metabolic pathways. This may warrant further studies, as mutations in de novo purine biosynthesis pathways have been observed to extend the yeast chronological life span [225].

Lipid biosynthesis

We observed most proteins with a role in lipid biosynthesis to be downregulated (see SUR4 in Fig. 5.1), with the important exception of FAA4, which also participates in stress response, and ACC1, which catalyzes carboxylation of acetyl-CoA and into fatty acid precursors. Lipid metabolism was previously shown to be affected by metformin [213].

Signaling pathways

Metformin has been observed to affect TOR signaling in yeast. This effect may be upstream of the protein translation inhibitory effects of the drug, and downstream of the decrease in amino acid levels that it also causes [213]. We observed the transcription factor GCN4, one of the master regulators for gene expression in *S. cerevisiae* to be rapidly and strongly activated in our assay (see Figure 5.1). We also observed KSS1, a mitogen-activated protein kinase that controls filamentous growth to be activated, which is in line with previous findings from microarray data.

5.2.2 Comparing the meformin and iron-deprivation response

Multiple studies have advanced the notion that metformin treatment induced an iron deprivation-like state. We also observed multiple iron-binding proteins to be affected by metformin treatment. To explore this phenomenon further, we decided to expose the *S. cerevisiae* GFP library to iron-depleted media for a period of 10 hours, corresponding to the duration to which the library was exposed to metformin. We performed these experiments in triplicate.

A volcano plot of the iron-deprivation response is also displayed in 5.2. Qualitatively, iron deprivation leads to more down-regulation than upregulation, which differs from the more balanced effect of metformin. We observed a strong signal for proteins listed as possessing iron ion binding properties on the Saccharomyces Genome Database (FET3, FTH1, FTR1, CCC1, MTD1, ISA2, and RLI1) which is in line with previous studies [226].

Fold-change analysis revealed that 777 and 375 were up- and down-regulated in both conditions, respectively. To narrow down the analysis, we focused on the 20 genes that had the most similar fold change across the two conditions, as shown in Figure 5.3. Of those 20 similarly-regulated genes, several of those that were upregulated were linked to iron homeostasis, such as AFT2, an Iron-regulated transcriptional activator, and YPK9, which has a role in sequestering heavy metals. While we did not observe a strong effect of metformin on iron transport proteins, iron-binding proteins saw their activity upregulated, which partially reconciles our data with that of previous studies.

One of the strongly upregulated genes across both conditions was MTD1, which plays a one-carbon metabolism and was observed to respond to metformin in prior studies [212]. DNA replication stress proteins are activated in Figure 5.3 (PHM7, ZRP1) while amino acid synthesis proteins (ARG5,6 /ASN11 /MEU1) are downregulated. Both trends conform to previous analyses performed using metformin in yeast.

While fold change analysis highlighted genes that are likely to play a role in the *S. Cerevisiae* response to both metformin and iron deprivation, we note that grouping genes by

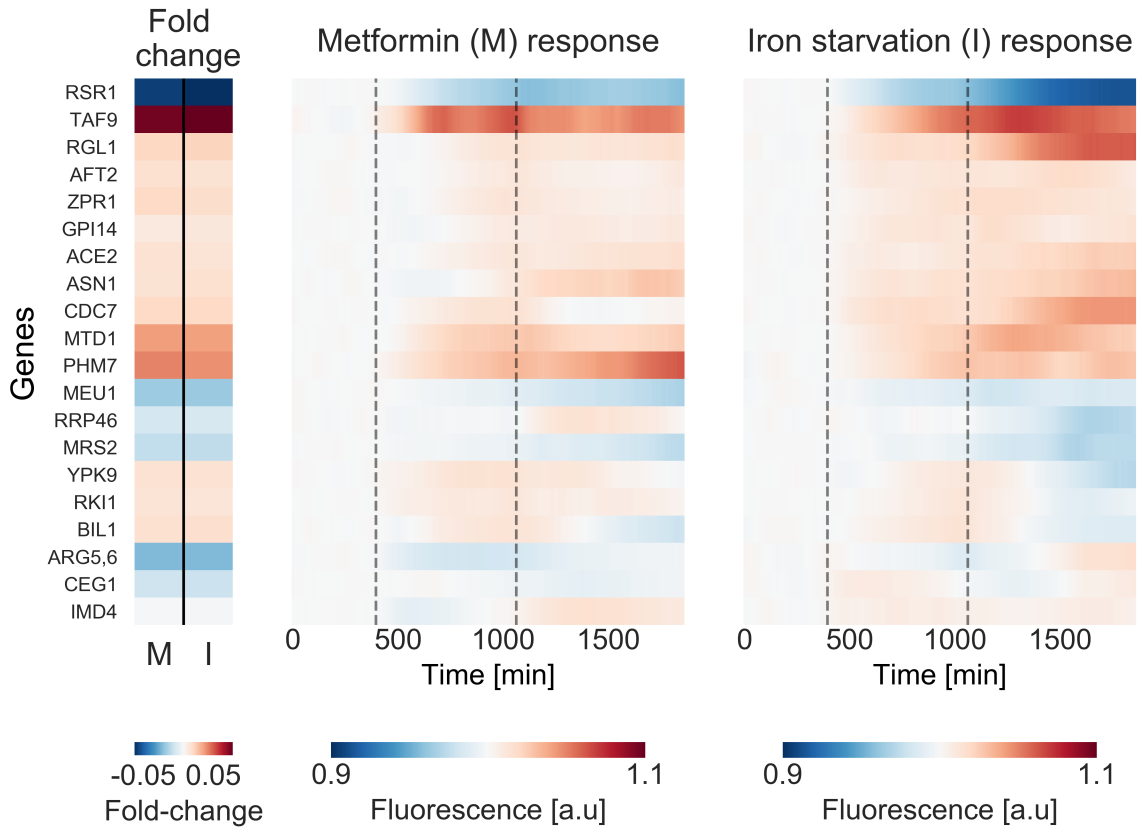

Figure 5.3: Genes with similar fold change-response to metformin and iron starvation. The top 20 genes (across 2720) with the most similar fold change are displayed. Left: Log2 of the average fold change across three replicates in response to either 100mM Metformin (M) or iron deprivation (I). Center: time-series response of those genes when exposed to metformin. The period of 10 hours during which cells are exposed to metformin is indicated by the dotted black line. Center: time-series response of those genes when exposed to iron deprivation

similar fold-change does not necessarily match genes with similar dynamics. This highlights the importance of capturing dynamics-based comparisons between expression signatures for more in-depth analysis.

5.3 Discussion

Our dynamics-based metformin screen is a proof of concept that *S. cerevisiae* drug response can be measured in real-time and over long periods of time using the Dynomics platform

adapted. Our screen showed good overlap with existing studies on the effect of metformin on the *S. cerevisaie* proteome, notably in the upregulation of energy metabolism, disruption of mitochondrial function and lipid biosynthesis, and downregulation of iron-sulfur-containing proteins. We also saw a strong signal that departed from previous studies in the upregulation of purine biosynthesis and the lack of upregulation of iron transport proteins. Taken together, these results confirm that the cellular machinery impacted by metformin and that are relevant to life span are deeply inter-connected. In turn, this stresses the importance of performing high-throughput assays to uncover these interactions and drive towards a better understanding of drug mechanism of action.

While this preliminary analysis confirms the utility of our dynamics-based chemogenomic screen, additional analysis using *ad hoc* dynamics-based modeling tools are needed to fully leverage the data produced by Dynomics and validate the hits produced by our screen. This is particularly important given the lack of a standard for processing time-course microfluidics-based genomics data and the risk of false positives in datasets possessing high dimensionality. The integration of this data with other omics datasets, omics computational tools, and causal modeling approaches would be a step forward in the construction of system-based models that predict cellular response to perturbations [63].

In this chapter, we established that the Dynomics platform can be used to profile global cellular responses to the action of therapeutic molecules with potential applications in the diagnosis and treatment of aging-related diseases.

5.4 Methods

5.4.1 Experimental set up

The *S. cerevisiae* GFP library [144] library was spotted across two 2176-strain microfluidic devices with HD biopixels (see 2.2.6) using the protocol described in section 2.3.2. The silicon wafer was fabricated using standard photolithography techniques previously described by our group [90, 129]. Cells were grown in Synthetic Defined media (using Sunrise components) with 2% Glucose (Difco) using a flow rate of 5 mL/hr on each device. Cells took about 24hours to reach confluence in the HD biopixel microchambers. Once the microchambers are filled, we monitored library response by taking fluorescent images every ten minutes with 1 second exposure time for transmitted light images and 1 minute exposure time for GFP fluorescence images. 100mM Metformin Hydrochloride (Sigma D150959) and iron deprivation inductions were performed by swapping the media in the inlet syringe with media premixed with the inducer. Inductions were carried out in triplicate across at least two experiments. We note here that concentrations of the Metformin induction (100mM) are high compared to clinically relevant ranges of 10 μM, but chosen consistent with previous *S. Cerevisiae*-based screens of the drug due to yeast's high resistance to a variety of drugs because of their efficient multidrug efflux pumps [212].

5.4.2 Data processing

Images were flat-field corrected and registered to an extraction mask to account for drifting and thermal expansion as described in section 3.8.4.

The mean pixel value for each biopixel was extracted and normalized by the inverted transmitted light signal to account for variation in the number of cells per trap over the course of an induction. We smooth the signal using Holt-Winters exponential smoothing. A strain containing only a histidine selection marker without GFP was present in each microfluidic device. This strain was yRO136 (BY4741 his3::SpHIS5 (HISMX6)). The histidine selection marker

inserted into this strain was from Addgene plasmid 44836. We divided the mean HD biopixel fluorescence values from these strains on each chip from all other strain fluorescence values on that chip and discarded all strains that were below 1. Finally, each strain's signal is divided by the mean of all strains to obtain the differentiating signal for each strain compared to the mean.

Fold change in Figure 5.2 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 two-sided t-test on the log_2 fluorescent values at the start and end of each metal exposure (scipy.stats.ttest_rel).

5.5 Acknowledgements

This chapter contains material being prepared for submission as Multiplexed microfluidic platforms for real-time heavy metal sensing and dynamic phenotype screening. Thouvenin Gregoire*, Csicsery, Nicholas*, Stasiowski, Elizabeth*, Graham, Garrett, O'Laughlin, Richard and Hasty, Jeff. (*equal contribution). The dissertation author was one of the primary authors and researchers of this material.

Bibliography

[1] Graham Joyce. Physical biology of the cell. *Journal of Biological Education*, 2014.

[2] Ferric C. Fang and Arturo Casadevall. Reductionistic and holistic science, 2011.

[3] Howard Becker and Alfred Korzybski. Science and Sanity: An Introduction to Non-Aristotelian Systems and General Semantics. *American Sociological Review*, 1942.

[4] Vijay Pande. https://a16z.com/2018/01/07/bio-as-tech-debt/, Jan 2018.

[5] Uri Alon. *An introduction to systems biology: Design principles of biological circuits*. 2006.

[6] Erwin Schrödinger. What is life?: The physical aspect of the living cell ; with Mind and matter ; & Autobiographical sketches. *Mind and Matter*, 1992.

[7] Víctor De Lorenzo. Evolutionary tinkering vs. rational engineering in the times of synthetic biology Carmen McLeod, Brigitte Nerlich. *Life Sciences, Society and Policy*, 2018.

[8] Hans V. Westerhoff and Bernhard O. Palsson. The evolution of molecular biology into systems biology, 2004.

[9] Markus W. Covert. *Fundamentals of Systems Biology*. 2017.

[10] Hiroaki Kitano. Systems biology: A brief overview, 2002.

[11] Steven A. Benner and A. Michael Sismour. Synthetic biology, 2005.

[12] Rainer Breitling. What is systems biology? *Frontiers in Physiology*, 2010.

[13] Alec A.K. Nielsen, Bryan S. Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A. Strychalski, David Ross, Douglas Densmore, and Christopher A. Voigt. Genetic circuit design automation. *Science*, 2016.

[14] David L Shis, Matthew R Bennett, and Oleg A Igoshin. Dynamics of Bacterial Gene Regulatory Networks. *Annu. Rev. Biophys*, 4721(21):1–21, 2018.

[15] Bernhard O. Palsson. *Systems biology: Constraint-based reconstruction and analysis*. 2015.

[16] Adam M. Feist, Markus J. Herrgård, Ines Thiele, Jennie L. Reed, and Bernhard Palsson. Reconstruction of biochemical networks in microorganisms, 2009.

[17] Jin Hwan Park, Sang Yup Lee, Tae Yong Kim, and Hyun Uk Kim. Application of systems biology for bioprocess development, 2008.

[18] Leroy Hood, James R. Heath, Michael E. Phelps, and Biaoyang Lin. Systems biology and new technologies enable predictive and preventative medicine, 2004.

[19] Douglas B. Kell and Stephen G. Oliver. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era, 2004.

[20] D. Ewen Cameron, Caleb J. Bashor, and James J. Collins. A brief history of synthetic biology, 2014.

[21] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 2000.

[22] Mariette R. Atkinson, Michael A. Savageau, Jesse T. Myers, and Alexander J. Ninfa. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell*, 2003.

[23] Alvin Tamsir, Jeffrey J. Tabor, and Christopher A. Voigt. Robust multicellular computing using genetically encoded NOR gates and chemical 'wiresg'. *Nature*, 2011.

[24] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C. Stanton, and Christopher A. Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 2012.

[25] 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*, 2008.

[26] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. A synchronized quorum of genetic clocks. *Nature*, 2010.

[27] Elowitz. A synthetic oscillatory network repressilator. *Nature*, 2000.

[28] Jeffrey J. Tabor, Anselm Levskaya, and Christopher A. Voigt. Multichromatic control of gene expression in escherichia coli. *Journal of Molecular Biology*, 2011.

[29] Brian P. Landry, Rohan Palanki, Nikola Dyulgyarov, Lucas A. Hartsough, and Jeffrey J. Tabor. Phosphatase activity tunes two-component system sensor detection threshold. *Nature Communications*, 2018.

[30] Daniel G. Gibson, John I. Glass, Carole Lartigue, Vladimir N. Noskov, Ray Yuan Chuang, Mikkel A. Algire, Gwynedd A. Benders, Michael G. Montague, Li Ma, Monzia M. Moodie, Chuck Merryman, Sanjay Vashee, Radha Krishnakumar, Nacyra Assad-Garcia, Cynthia Andrews-Pfannkoch, Evgeniya A. Denisova, Lei Young, Zhi Ng Qi, Thomas H. Segall-Shapiro, Christopher H. Calvey, Prashanth P. Parmar, Clyde A. Hutchison, Hamilton O. Smith, and J. Craig Venter. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 2010.

[31] Clyde A. Hutchison, Ray Yuan Chuang, Vladimir N. Noskov, Nacyra Assad-Garcia, Thomas J. Deerinck, Mark H. Ellisman, John Gill, Krishna Kannan, Bogumil J. Karas, Li Ma, James F. Pelletier, Zhi Qing Qi, R. Alexander Richter, Elizabeth A. Strychalski, Lijie Sun, Yo Suzuki, Billyana Tsvetanova, Kim S. Wise, Hamilton O. Smith, John I. Glass, Chuck Merryman, Daniel G. Gibson, and J. Craig Venter. Design and synthesis of a minimal bacterial genome. *Science*, 2016.

[32] Sarah M. Richardson, Leslie A. Mitchell, Giovanni Stracquadanio, Kun Yang, Jessica S. Dymond, James E. DiCarlo, Dongwon Lee, Cheng Lai Victor Huang, Srinivasan Chandrasegaran, Yizhi Cai, Jef D. Boeke, and Joel S. Bader. Design of a synthetic yeast genome. *Science*, 2017.

[33] Julius Fredens, Kaihang Wang, Daniel de la Torre, Louise F.H. Funke, Wesley E. Robertson, Yonka Christova, Tiongsun Chia, Wolfgang H. Schmied, Daniel L. Dunkelmann, Václav Beránek, Chayasith Uttamapinant, Andres Gonzalez Llamazares, Thomas S. Elliott, and Jason W. Chin. Total synthesis of Escherichia coli with a recoded genome. *Nature*, 2019.

[34] Jessica G. Perez, Jessica C. Stark, and Michael C. Jewett. Cell-free synthetic biology: Engineering beyond the cell. *Cold Spring Harbor Perspectives in Biology*, 2016.

[35] Drew Endy. Foundations for engineering biology, 2005.

[36] Michael Eisenstein. Living factories of the future. *Nature*, 2016.

[37] Michael J. Smanski, Hui Zhou, Jan Claesen, Ben Shen, Michael A. Fischbach, and Christopher A. Voigt. Synthetic biology to access and expand nature's chemical diversity, 2016.

[38] Jay D. Keasling. Manufacturing molecules through metabolic engineering, 2010.

[39] Pamela P Peralta-Yahya, Fuzhong Zhang, Stephen B del Cardayre, and Jay D Keasling. Microbial engineering for the production of advanced biofuels. *Nature*, 488(7411):320–8, aug 2012.

[40] Lisa A. Anderson, M. Ahsanul Islam, and Kristala L.J. Prather. Synthetic biology strategies for improving microbial synthesis of "green" biopolymers, 2018.

[41] Leo Scheller, Tobias Strittmatter, David Fuchs, Daniel Bojar, and Martin Fussenegger. Generalized extracellular molecule sensor platform for programming cellular behavior article. *Nature Chemical Biology*, 2018.

[42] Fuzhong Zhang and Jay Keasling. Biosensors and their applications in microbial metabolic engineering, 2011.

[43] 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.

[44] Tal Danino, Arthur Prindle, Gabriel A. Kwong, Matthew Skalak, Howard Li, Kaitlin Allen, Jeff Hasty, and Sangeeta N. Bhatia. Programmable probiotics for detection of cancer in urine. *Science Translational Medicine*, 2015.

[45] Jang Hwan Cho, James J. Collins, and Wilson W. Wong. Universal Chimeric Antigen Receptors for Multiplexed and Logical Control of T Cell Responses. *Cell*, 2018.

[46] 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*, 2016.

[47] Sreyan Chowdhury, Samuel Castro, Courtney Coker, Taylor E. Hinchliffe, Nicholas Arpaia, and Tal Danino. Programmable bacteria induce durable tumor regression and systemic antitumor immunity. *Nature Medicine*, 2019.

[48] Maarten Boudry and Massimo Pigliucci. The mismeasure of machine: Synthetic biology and the trouble with engineering metaphors. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, 2013.

[49] Eleonore Pauwels. Mind the metaphor. *Nature*, 2013.

[50] François Jacob. Evolution and tinkering. *Science*, 1977.

[51] A. Danchin. The Delphic boat or what the genomic texts tell us, 1998.

[52] Gürol M Süel, Jordi Garcia-Ojalvo, Louisa M Liberman, and Michael B Elowitz. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(7083):545–550, 2006.

[53] Gürol M Süel, Rajan P Kulkarni, Jonathan Dworkin, Jordi Garcia-Ojalvo, and Michael B Elowitz. Tunability and noise dependence in differentiation dynamics. *Science*, 315(5819):1716–1719, 2007.

[54] Nan Hao and Erin K O'shea. Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nature structural & molecular biology*, 19(1):31, 2012.

[55] Zohreh AkhavanAghdam, Joydeb Sinha, Omar P Tabbaa, and Nan Hao. Dynamic control of gene regulatory logic by seemingly redundant transcription factors. *Elife*, 5:e18458, 2016.

[56] Lacramioara Bintu, John Yong, Yaron E Antebi, Kayla McCue, Yasuhiro Kazuki, Narumi Uno, Mitsuo Oshimura, and Michael B Elowitz. Dynamics of epigenetic regulation at the single-cell level. *Science*, 351(6274):720–724, 2016.

[57] Nan Hao, Bogdan a Budnik, Jeremy Gunawardena, and Erin K. O'Shea. Tunable signal processing through modular control of transcription factor translocation. *Science (New York, N.Y.)*, 339(6118):460–4, jan 2013.

[58] Albert László Barabási, Natali Gulbahce, and Joseph Loscalzo. Network medicine: A network-based approach to human disease. *Nature Reviews Genetics*, 12(1):56–68, 2011.

[59] L. J. Bugaj, A. J. Sabnis, A. Mitchell, J. E. Garbarino, J. E. Toettcher, T. G. Bivona, and W. A. Lim. Cancer mutations and targeted drugs can disrupt dynamic signal encoding by the Ras-Erk pathway. *Science*, 361(6405), 2018.

[60] 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.

[61] Hilal Taymaz-Nikerel, Ayca Cankorur-Cetinkaya, and Betul Kirdar. Genome-wide transcriptional response of Saccharomyces cerevisiae to stress-induced perturbations, 2016.

[62] Teresa M. Przytycka, Mona Singh, and Donna K. Slonim. Toward the dynamic interactome: It's about time. *Briefings in Bioinformatics*, 2010.

[63] Ziv Bar-Joseph, Anthony Gitter, and Itamar Simon. Studying and modelling dynamic biological processes using time-series gene expression data, 2012.

[64] Daniel Spies and Constance Ciaudo. Dynamics in Transcriptomics: Advancements in RNA-seq Time Course and Downstream Analysis, 2015.

[65] Nitai D. Mukhopadhyay and Snigdhansu Chatterjee. Causality and pathway search in microarray time series experiment. *Bioinformatics*, 2007.

[66] 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*, 2012.

[67] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, 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.

[68] 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.

[69] Yuen Ho, Albrecht Gruhler, Adrian Heilbut, Gary D. Bader, Lynda Moore, Sally Lin Adams, Anna Millar, and Taylor et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry, 2002.

[70] 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.

[71] Michael J. Heller. DNA Microarray Technology: Devices, Systems, and Applications. *Annu Rev Biomed Eng.*, 4:129–153, 2002.

[72] Leo Baumgart, William Mather, and Jeff Hasty. Synchronized DNA cycling across a bacterial population. *Nature Genetics*, 2017.

[73] Bryan P. Tracy, Stefan M. Gaida, and Eleftherios T. Papoutsakis. Flow cytometry for bacteria: Enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes, 2010.

[74] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, 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. *Nature*, 2018.

[75] M. Zheng, X. Wang, L. J. Templeton, D. R. Smulski, R. A. LaRossa, and G. Storz. DNA microarray-mediated transcriptional profiling of the Escherichia coli response to hydrogen peroxide. *Journal of Bacteriology*, 2001.

[76] 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.

[77] Isaac Nuñez, Tamara Matute, Roberto Herrera, Juan Keymer, Timothy Marzullo, Timothy Rudge, and Fernán Federici. Low cost and open source multi-fluorescence imaging system for teaching and research in biology and bioengineering. *PLoS ONE*, 2017.

[78] Jacob Beal, Traci Haddock-Angelli, Markus Gershater, and Kim et al. De Mora. Reproducibility of fluorescent expression from engineered biological constructs in E. coli. *PLoS ONE*, 2016.

[79] Garrett Graham, Nicholas Csicsery, Elizabeth Stasiowski, Gregoire Thouvenin, William H. Mather, Michael Ferry, Scott Cookson, and Jeff Hasty. Genome-scale transcriptional

dynamics and environmental biosensing. *Proceedings of the National Academy of Sciences of the United States of America*, 2020.

[80] Roger P. Alexander, Philip M. Kim, Thierry Emonet, and Mark B. Gerstein. Understanding modularity in molecular networks requires dynamics. *Science Signaling*, 2(81), 2009.

[81] Heather A. Piwowar, Michael J. Becich, Howard Bilofsky, and Rebecca S. Crowley. Towards a data sharing culture: Recommendations for leadership from academic health centers, 2008.

[82] Jeremy P. Birnholtz and Matthew J. Bietz. Data at work: Supporting sharing in science and engineering. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work (GROUP '03)*, 2003.

[83] Albert László Barabási and Zoltán N. Oltvai. Network biology: Understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[84] Yolanda T. Chong, Judice L.Y. Koh, Helena Friesen, 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*, 2015.

[85] Jeremy E. Purvis and Galit Lahav. Encoding and decoding cellular information through signaling dynamics. *Cell*, 152(5):945–956, 2013.

[86] Chunxiong Luo, Xuejun Zhu, Tao Yu, Xianjia Luo, Qi Ouyang, Hang Ji, and Yong Chen. A fast cell loading and high-throughput microfluidic system for long-term cell culture in zero-flow environments. *Biotechnology and Bioengineering*, 101(1):190–195, 2008.

[87] Yuan Tian, Chunxiong Luo, and Qi Ouyang. A microfluidic synchronizer for fission yeast cells. *Lab on a chip*, 13:4071–4077, 2013.

[88] Todd Thorsen, Sebastian J. Maerkl, and Stephen R. Quake. Microfluidic large-scale integration. *Science*, 2002.

[89] George M. Whitesides. The origins and the future of microfluidics, 2006.

[90] M. S. Ferry, I. A. Razinkov, and J. Hasty. *Microfluidics for synthetic biology: From design to execution*, volume 497. 2011.

[91] Yi Liu and Hang Lu. Microfluidics in systems biology - hype or truly useful?, 2016.

[92] Helene Andersson and Albert Van Den Berg. Microfabrication and microfluidics for tissue engineering: State of the art and future opportunities, 2004.

[93] Frederick K. Balagaddé, Lingchong You, Carl L. Hansen, Frances H. Arnold, and Stephen R. Quake. Microbiology: Long-term monitoring of bacteria undergoing programmed population control in a microchemostat. *Science*, 2005.

[94] Matthew R. Bennett and Jeff Hasty. Microfluidic devices for measuring gene network dynamics in single cells, 2009.

[95] Gilles Charvin, Frederick R. Cross, and Eric D. Siggia. A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells. *PLoS ONE*, 2008.

[96] S. K.W. Dertinger, D. T. Chiu, Noo Li Jeon, and G. M. Whitesides. Generation of gradients having complex shapes using microfluidic networks. *Analytical Chemistry*, 2001.

[97] Anne Y. Fu, Hou Pu Chou, Charles Spence, Frances H. Arnold, and Stephen R. Quake. An integrated microfabricated cell sorter. *Analytical Chemistry*, 2002.

[98] Alex Groisman, Markus Enzelberger, and Stephen R. Quake. Microfluidic memory and control devices. *Science*, 2003.

[99] Jong Wook Hong, Vincent Studer, Giao Hang, W. French Anderson, and Stephen R. Quake. A nanoliter-scale nucleic acid processor with parallel architecture. *Nature Biotechnology*, 2004.

[100] Dongeun Huh, Wei Gu, Yoko Kamotani, James B. Grotberg, and Shuichi Takayama. Microfluidics for flow cytometric analysis of cells and particles, 2005.

[101] Hanbin Mao, Tinglu Yang, and Paul S. Cremer. A microfluidic device with a linear temperature gradient for parallel and combinatorial measurements. *Journal of the American Chemical Society*, 2002.

[102] J. Cooper McDonald and George M. Whitesides. Poly(dimethylsiloxane) as a material for fabricating microfluidic devices. *Accounts of Chemical Research*, 2002.

[103] Brian M. Paegel, Robert G. Blazej, and Richard A. Mathies. Microfluidic devices for DNA sequencing: Sample preparation and electrophoretic analysis, 2003.

[104] Marc A. Unger, Hou Pu Chou, Todd Thorsen, Axel Scherer, and Stephen R. Quake. Monolithic microfabricated valves and pumps by multilayer soft lithography. *Science*, 2000.

[105] Francis Lin, Wajeeh Saadi, Seog Woo Rhee, Shur Jen Wang, Sukant Mittal, and Noo Li Jeon. Generation of dynamic temporal and spatial concentration gradients using microfluidic devices. *Lab on a Chip*, 2004.

[106] 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.

[107] J. Uhlendorf, A. Miermont, T. Delaveau, G. Charvin, F. Fages, S. Bottani, G. Batt, and P. Hersen. Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences*, 109(35):14271–14276, 2012.

[108] Jerome T. Mettetal, Dale Muzzey, Carlos Gomez-Uribe, and Alexander van Oudenaarden. The Frequency Dependence of Osmo-adaptation in Saccharomyces cerevisiae. *Science*, 319:482–484, 2008.

[109] James C.W. Locke and Michael B. Elowitz. Using movies to analyse gene circuit dynamics in single cells, 2009.

[110] 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.

[111] 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.

[112] 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.

[113] 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.

[114] Ce Zhang, Hsiung-Lin Tu, Gengjie Jia, Tanzila Mukhtar, Verdon Taylor, Andrey Rzhetsky, and Savaş Tay. Ultra-multiplexed analysis of single-cell dynamics reveals logic rules in differentiation. *Science Advances*, 2019.

[115] Anders S. Hansen, Nan Hao, and Erin K. OShea. High-throughput microfluidics to control and measure signaling dynamics in single yeast cells. *Nature Protocols*, 2015.

[116] Octavio Mondragón-Palomino, Tal Danino, Jangir Selimkhanov, Lev Tsimring, and Jeff Hasty. Entrainment of a population of synthetic genetic oscillators. *Science*, 333(6047):1315–1319, 2011.

[117] Yang Jun Kang and Sung Yang. Fluidic low pass filter for hydrodynamic flow stabilization in microfluidic environments. *Lab on a Chip*, 2012.

[118] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of escherichia coli. *Current Biology*, 2010.

[119] K. Nath and A. L. Koch. Protein degradation in Escherichia coli. I. Measurement of rapidly and slowly decaying components. *Journal of Biological Chemistry*, 1970.

[120] William Mather, Octavio Mondragón-Palomino, Tal Danino, Jeff Hasty, and Lev S. Tsimring. Streaming instability in growing cell populations. *Physical Review Letters*, 2010.

[121] Michael B. Elowitz, Arnold J. Levine, Eric D. Siggia, and Peter S. Swain. Stochastic gene expression in a single cell. *Science*, 2002.

[122] Nela Nikolic, Frank Schreiber, Alma Dal Co, Daniel J. Kiviet, Tobias Bergmiller, Sten Littmann, Marcel M.M. Kuypers, and Martin Ackermann. Cell-to-cell variation and specialization in sugar metabolism in clonal bacterial populations. *PLoS Genetics*, 2017.

[123] Philip Bittihn, Andriy Didovyk, Lev S. Tsimring, and Jeff Hasty. Genetically engineered control of phenotypic structure in microbial colonies. *Nature Microbiology*, 2020.

[124] Alma Dal Co, Simon Van Vliet, and Martin Ackermann. Emergent microscale gradients give rise to metabolic cross-feeding and antibiotic tolerance in clonal bacterial populations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2019.

[125] 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.

[126] Leeat Keren, Ora Zackay, Maya Lotan-Pompan, Uri Barenholz, Erez Dekel, Vered Sasson, Guy Aidelberg, Anat Bren, Danny Zeevi, Adina Weinberger, Uri Alon, Ron Milo, and Eran Segal. Promoters maintain their relative activity levels under different growth conditions. *Molecular Systems Biology*, 2013.

[127] Alexander Schmidt, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer, Luciano Callipo, Kèvin Knoops, Manuel Bauer, Ruedi Aebersold, and Matthias Heinemann. The quantitative and condition-dependent Escherichia coli proteome. *Nature Biotechnology*, 2016.

[128] Jerome T. Mettetal, Dale Muzzey, Carlos Gómez-Uribe, and Alexander Van Oudenaarden. The frequency dependence of osmo-adaptation in Saccharomyces cerevisiae. *Science*, 2008.

[129] E. Stasiowski. Multiplexed microfluidics utilizing genome-scale dynamics for biosensing and fermentation monitoring, 2019.

[130] Arthur Prindle, Phillip Samayoa, Ivan Razinkov, Tal Danino, Lev S. Tsimring, and Jeff Hasty. A sensing array of radically coupled genetic 'biopixels'. *Nature*, 2012.

[131] N. Rajmohan and S. A. Prathapar. Extent of arsenic contamination and its impact on the food chain and human health in the Eastern Ganges Basin: A review. *IWMI Working Papers*, 2014.

[132] Regina Mahr and Julia Frunzke. Transcription factor-based biosensors in biotechnology: current state and future prospects, 2016.

[133] Lara Tess Bereza-Malcolm, Gülay Mann, and Ashley Edwin Franks. Environmental Sensing of Heavy Metals Through Whole Cell Microbial Biosensors: A Synthetic Biology Approach. *ACS Synthetic Biology*, 2015.

[134] Sebastian R. Schmidl, Felix Ekness, Katri Sofjan, Kristina N.M. Daeffler, Kathryn R. Brink, Brian P. Landry, Karl P. Gerhardt, Nikola Dyulgyarov, Ravi U. Sheth, and Jeffrey J. Tabor. Rewiring bacterial two-component systems by modular DNA-binding domain swapping. *Nature Chemical Biology*, 2019.

[135] Kiyohito Yagi. Applications of whole-cell bacterial sensors in biotechnology and environmental science, 2007.

[136] Chi Wei Huang, Shih Hung Yang, Man Wai Sun, and Vivian Hsiu Chuan Liao. Development of a set of bacterial biosensors for simultaneously detecting arsenic and mercury in groundwater. *Environmental Science and Pollution Research*, 2015.

[137] Dylan P. Webster, Michaela A. TerAvest, Devin F.R. Doud, Arun Chakravorty, Eric C. Holmes, Caleb M. Radens, Swati Sureka, Jeffrey A. Gralnick, and Largus T. Angenent. An arsenic-specific biosensor with genetically engineered Shewanella oneidensis in a bioelectrochemical system. *Biosensors and Bioelectronics*, 2014.

[138] Lara Tess Bereza-Malcolm, Gülay Mann, and Ashley Edwin Franks. Environmental Sensing of Heavy Metals Through Whole Cell Microbial Biosensors: A Synthetic Biology Approach. *ACS Synthetic Biology*, 2015.

[139] Pete Chandrangsu, Christopher Rensing, and John D. Helmann. Metal homeostasis and resistance in bacteria, 2017.

[140] Zhen Ma, Faith E. Jacobsen, and David P. Giedroc. Coordination chemistry of bacterial metal transport and sensing. *Chemical Reviews*, 2009.

[141] Xinyi Wan, Francesca Volpetti, Ekaterina Petrova, Chris French, Sebastian J. Maerkl, and Baojun Wang. Cascaded amplifying circuits enable ultrasensitive cellular sensors for toxic metals. *Nature Chemical Biology*, 2019.

[142] Sayantani Sikdar and Madhusree Kundu. A Review on Detection and Abatement of Heavy Metals. *ChemBioEng Reviews*, 2018.

[143] Zhiqiang Wang, Xia Sun, Caihong Li, Xianquan He, and Gang Liu. On-site detection of heavy metals in agriculture land by a disposable sensor based virtual instrument. *Computers and Electronics in Agriculture*, 2016.

[144] 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.

[145] 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.

[146] 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.

[147] Tianqi Chen and Carlos Guestrin. XGBoost: A Scalable Tree Boosting System. *ArXiv e-prints*, 2016.

[148] Sepp Hochreiter and Jurgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1–32, 1997.

[149] 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.

[150] 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.

[151] Sumeet Pal Singh, Sharan Janjuha, Samata Chaudhuri, Susanne Reinhardt, Annekathrin Kränkel, Sevina Dietz, Anne Eugster, Halil Bilgin, Selçuk Korkmaz, Gökmen Zararsız, Nikolay Ninov, and John E. Reid. Machine learning based classification of cells into chronological stages using single-cell transcriptomics. *Scientific Reports*, 2018.

[152] Max Tegmark. *Life 3.0: Being Human in the Age of Artificial Intelligence*. Knopf Publishing Group, 2017.

[153] Zachary D. Stephens, Skylar Y. Lee, Faraz Faghri, Roy H. Campbell, Chengxiang Zhai, Miles J. Efron, Ravishankar Iyer, Michael C. Schatz, Saurabh Sinha, and Gene E. Robinson. Big Data: Astronomical or Genomical? *PLOS Biology*, 13(7):e1002195, 7 2015.

[154] Oswaldo Trelles, Pjotr Prins, Marc Snir, and Ritsert C. Jansen. Big data, but are we ready? *Nature Reviews Genetics*, 12(3):224–224, 3 2011.

[155] Casey S. Greene, Jie Tan, Matthew Ung, Jason H. Moore, and Chao Cheng. Big Data Bioinformatics. *Journal of Cellular Physiology*, 229(12):1896–1900, 12 2014.

[156] Interpretable ML Symposium - NIPS 2017, 2017.

[157] 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.

[158] 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.

[159] Jian Zhou and Olga G Troyanskaya. Predicting effects of noncoding variants with deep learning-based sequence model. *Nature methods*, 12(10):931–4, 2015.

[160] Scott Lundberg and Su-In Lee. A Unified Approach to Interpreting Model Predictions. *ArXiv e-prints*, (Section 2):1–10, 2017.

[161] Scott M. Lundberg, Gabriel G. Erion, and Su-In Lee. Consistent Individualized Feature Attribution for Tree Ensembles. *ArXiv e-prints*, (2), 2018.

[162] 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.

[163] 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.

[164] 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.

[165] 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.

[166] Frans J. De Bruijn. *Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria*. 2016.

[167] 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.

[168] Stephen P. LaVoie and Anne O. Summers. Transcriptional responses of Escherichia coli during recovery from inorganic or organic mercury exposure. *BMC Genomics*, 2018.

[169] K. R. Brocklehurst and A. P. Morby. Metal-ion tolerance in Escherichia coli: Analysis of transcriptional profiles by gene-array technology. *Microbiology*, 2000.

[170] Ilias Tagkopoulos, Yir Chung Liu, and Saeed Tavazoie. Predictive behavior within microbial genetic networks. *Science*, 2008.

[171] 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.

[172] 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.

[173] 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.

[174] 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.

[175] 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.

[176] Zachary Chase Lipton, Charles Elkan, and Balakrishnan Narayanaswamy. Thresholding Classifiers to Maximize F1 Score. *ArXiv e-prints*, 2014.

[177] 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.

[178] 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.

[179] Joe J. Harrison, Howard Ceri, and Raymond J. Turner. Multimetal resistance and tolerance in microbial biofilms. *Nature Reviews Microbiology*, 5(12):928–938, 2007.

[180] Celine I.L. Justino, Ana C. Freitas, Armando C. Duarte, and Teresa A.P.Rocha Santos. Sensors and biosensors for monitoring marine contaminants, 2015.

[181] Rong Gao and Ann M. Stock. Biological insights from structures of two-component proteins, 2009.

[182] Luke E. Ulrich, Eugene V. Koonin, and Igor B. Zhulin. One-component systems dominate signal transduction in prokaryotes, 2005.

[183] Evan J. Olson, Lucas A. Hartsough, Brian P. Landry, Raghav Shroff, and Jeffrey J. Tabor. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nature Methods*, 2014.

[184] Qingyuan Gui, Tom Lawson, Suyan Shan, Lu Yan, and Yong Liu. The application of whole cell-based biosensors for use in environmental analysis and in medical diagnostics, 2017.

[185] Xinyi Wan, Trevor Y. H. Ho, and Baojun Wang. Engineering Prokaryote Synthetic Biology Biosensors. In *Handbook of Cell Biosensors*. 2019.

[186] K. Sorenson. Contaminated Water Diving: Effects Criteria, Categorization, and Tools for Pre-Dive Assessment. Modified Document for Purposes of Providing to LCDR Matthew Swiergosz, MC USN, Program Manager for SBIR N091-084 (real-Time Assessment of In-Water Contaminants). (2-3):355–384, 2009.

[187] Marta S. Alves, Anabela Pereira, Susana M. Araújo, Bruno B. Castro, António C.M. Correia, and Isabel Henriques. Seawater is a reservoir of multi-resistant Escherichia coli, including strains hosting plasmid-mediated quinolones resistance and extended-spectrum beta-lactamases genes. *Frontiers in Microbiology*, 2014.

[188] Yael Rozen and Shimshon Belkin. Survival of enteric bacteria in seawater, 2001.

[189] Corina P.D. Brussaard, Louis Peperzak, Siham Beggah, Lukas Y. Wick, Birgit Wuerz, Jan Weber, J. Samuel Arey, Bart Van Der Burg, Arjen Jonas, Johannes Huisman, and Jan Roelof Van Der Meer. Immediate ecotoxicological effects of short-lived oil spills on marine biota. *Nature Communications*, 2016.

[190] Lara Bereza-Malcolm, Sanja Aracic, Ruban Kannan, Gülay Mann, and Ashley E. Franks. Functional characterization of Gram-negative bacteria from different genera as multiplex cadmium biosensors. *Biosensors and Bioelectronics*, 2017.

[191] Roberto Rusconi, Melissa Garren, and Roman Stocker. Microfluidics expanding the frontiers of microbial ecology. *Annual Review of Biophysics*, 2014.

[192] Victor Chubukov, Luca Gerosa, Karl Kochanowski, and Uwe Sauer. Coordination of microbial metabolism, 2014.

[193] Emily J. Capra and Michael T. Laub. Evolution of two-component signal transduction systems, 2012.

[194] Kristina N-M Daeffler, Jeffrey D Galley, Ravi U Sheth, Laura C Ortiz-Velez, Christopher O Bibb, Noah F Shroyer, Robert A Britton, and Jeffrey J Tabor. Engineering bacterial thiosulfate and tetrathionate sensors for detecting gut inflammation. *Molecular Systems Biology*, 2017.

[195] Janet M. Wood. Bacterial responses to osmotic challenges, 2015.

[196] Benjamin Withman, Thusitha S. Gunasekera, Pavani Beesetty, Richard Agans, and Oleg Paliy. Transcriptional responses of uropathogenic Escherichia coli to increased environmental osmolality caused by salt or urea. *Infection and Immunity*, 2013.

[197] Sergio Alonso, Juana María Navarro-Llorens, Antonio Tormo, and Julián Perera. Construction of a bacterial biosensor for styrene. *Journal of Biotechnology*, 2003.

[198] F. C. Neidhardt, P. L. Bloch, and D. F. Smith. Culture medium for enterobacteria. *Journal of Bacteriology*, 1974.

[199] Genevieve L. Stein-O'Brien, Raman Arora, Aedin C. Culhane, Alexander V. Favorov, Lana X. Garmire, Casey S. Greene, Loyal A. Goff, Yifeng Li, Aloune Ngom, Michael F. Ochs, Yanxun Xu, and Elana J. Fertig. Enter the Matrix: Factorization Uncovers Knowledge from Omics, 2018.

[200] Leo Baumgart. *Gene Regulation in Synthetic Biology: Biosensing and Novel Tools for The Construction of ComplexGenetic Circuits*. PhD thesis, 2017.

[201] Rintaro Nakaya, Akiko Nakamura, and Yukio Murata. Resistance transfer agents in Shigella. *Biochemical and Biophysical Research Communications*, 1960.

[202] Jon L. Hobman, John Wilkie, and Nigel L. Brown. A design for life: Prokaryotic metal-binding MerR family regulators. In *BioMetals*, 2005.

[203] Simon Silver and Le T. Phung. Bacterial heavy metal resistance: New surprises, 1996.

[204] B. Diane Gambill and Anne O. Summers. Synthesis and degradation of the mRNA of the Tn21 mer operon. *Journal of Molecular Biology*, 1992.

[205] L. Silver, M. Chandler, E. Boy De La Tour, and L. Caro. Origin and direction of replication of the drug resistance plasmid R100.1 and of a resistance transfer factor derivative in synchronized cultures. *Journal of Bacteriology*, 1977.

[206] J. Wu and B. P. Rosen. Metalloregulated expression of the ars operon. *Journal of Biological Chemistry*, 1993.

[207] Rolf Lutz and Hermann Bujard. Independent and tight regulation of transcriptional units in escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, 1997.

[208] G. Endo and S. Silver. CadC, the transcriptional regulatory protein of the cadmium resistance system of Staphylococcus aureus plasmid pI258. *Journal of Bacteriology*, 1995.

[209] Andreas Zimmermann, Sebastian Hofer, Tobias Pendl, Katharina Kainz, Frank Madeo, and Didac Carmona-Gutierrez. Yeast as a tool to identify anti-aging compounds, 2018.

[210] Nir Barzilai, Jill P. Crandall, Stephen B. Kritchevsky, and Mark A. Espeland. Metformin as a Tool to Target Aging, 2016.

[211] Ainslie B. Parsons, Andres Lopez, Inmar E. Givoni, David E. Williams, Christopher A. Gray, Justin Porter, Gordon Chua, and Sopko et al. Exploring the Mode-of-Action of Bioactive Compounds by Chemical-Genetic Profiling in Yeast. *Cell*, 2006.

[212] Bram Stynen, Diala Abd-Rabbo, Jacqueline Kowarzyk, Leonor Miller-Fleming, Simran Kaur Aulakh, Philippe Garneau, Markus Ralser, and Stephen W. Michnick. Changes of Cell Biochemical States Are Revealed in Protein Homomeric Complex Dynamics. *Cell*, 2018.

[213] Esra Borklu-Yucel, Serpil Eraslan, and Kutlu O. Ulgen. Transcriptional remodeling in response to transfer upon carbon-limited or metformin-supplemented media in S. cerevisiae and its effect on chronological life span. *Applied Microbiology and Biotechnology*, 2015.

[214] Brian Onken and Monica Driscoll. Metformin induces a dietary restriction-like state and the oxidative stress response to extend C. elegans healthspan via AMPK, LKB1, and SKN-1. *PLoS ONE*, 2010.

[215] Daniel L. Smith, Calvin F. Elam, Julie A. Mattison, Mark A. Lane, George S. Roth, Donald K. Ingram, and David B. Allison. Metformin supplementation and life span in fischer-344 rats. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences*, 2010.

[216] Rosina Pryor and Filipe Cabreiro. Repurposing metformin: An old drug with new tricks in its binding pockets, 2015.

[217] Alejandro Martin-Montalvo, Evi M. Mercken, Sarah J. Mitchell, Hector H. Palacios, Patricia L. Mote, Morten Scheibye-Knudsen, Ana P. Gomes, Theresa M. Ward, Robin K. Minor, Marie José Blouin, Matthias Schwab, Michael Pollak, Yongqing Zhang, Yinbing Yu, Kevin G. Becker, Vilhelm A. Bohr, Donald K. Ingram, David A. Sinclair, Norman S. Wolf, Stephen R. Spindler, Michel Bernier, and Rafael De Cabo. Metformin improves healthspan and lifespan in mice. *Nature Communications*, 2013.

[218] Anila K. Madiraju, Derek M. Erion, Yasmeen Rahimi, Xian Man Zhang, Demetrios T. Braddock, Ronald A. Albright, Brett J. Prigaro, John L. Wood, Sanjay Bhanot, Michael J. MacDonald, Michael J. Jurczak, Joao Paulo Camporez, Hui Young Lee, Gary W. Cline, Varman T. Samuel, Richard G. Kibbey, and Gerald I. Shulman. Metformin suppresses gluconeogenesis by inhibiting mitochondrial glycerophosphate dehydrogenase. *Nature*, 2014.

[219] J. Abraham Avelar-Rivas, Michelle Munguía-Figueroa, Alejandro Juárez-Reyes, Erika Garay, Sergio E. Campos, Noam Shoresh, and Alexander DeLuna. An Optimized Competitive-Aging Method Reveals Gene-Drug Interactions Underlying the Chronological Lifespan of Saccharomyces cerevisiae. *Frontiers in Genetics*, 2020.

[220] Xiyan Li, Xin Wang, and Michael P. Snyder. Metformin affects heme function as a possible mechanism of action. *G3: Genes, Genomes, Genetics*, 2019.

[221] Lisa Logie, Jean Harthill, Kashyap Patel, Sandra Bacon, D. Lee Hamilton, Katherine Macrae, Gordon McDougall, Huan Huan Wang, Lin Xue, Hua Jiang, Kei Sakamoto, Alan R. Prescott, and Graham Rena. Cellular responses to the metal-binding properties of metformin. *Diabetes*, 61(6), 2012.

[222] William J. Jo, Jeung Hyoun Hyoun, Eric Oh, Daniel Jaramillo, Patricia Holman, Alex V. Loguinov, Adam P. Arkin, Corey Nislow, Guri Giaever, and Chris D. Vulpe. Novel insights into iron metabolism by integrating deletome and transcriptome analysis in an iron deficiency model of the yeast Saccharomyces cerevisiae. *BMC Genomics*, 2009.

[223] Jay E. Johnson and F. Brad Johnson. Methionine restriction activates the retrograde response and confers both stress tolerance and lifespan extension to yeast, mouse and human cells. *PLoS ONE*, 2014.

[224] Jay E. Johnson and F. Brad Johnson. Methionine restriction activates the retrograde response and confers both stress tolerance and lifespan extension to yeast, mouse and human cells. *PLoS ONE*, 2014.

[225] Mirela Matecic, Daniel L. Smith, Xuewen Pan, Nazif Maqani, Stefan Bekiranov, Jef D. Boeke, and Jeffrey S. Smith. A microarray-based genetic screen for yeast chronological aging factors. *PLoS Genetics*, 2010.

[226] Lucía Ramos-Alonso, Antonia María Romero, María Teresa Martínez-Pastor, and Sergi Puig. Iron Regulatory Mechanisms in Saccharomyces cerevisiae, 2020.