Title

Tools for engineering and modeling bacterial populations dynamics

Permalink

https://escholarship.org/uc/item/7s95q3r1

Author

Miano, Arianna

Publication Date

2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Tools for engineering and modeling bacterial populations dynamics

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

in

Bioengineering

by

Arianna Miano

Committee in charge:

Professor Jeff Hasty, Chair
Professor Lin Chao, Co-Chair
Professor Bernhard Palsson, Co-Chair
Professor Terence Hwa
Professor Prashant Mali

2022

The Dissertation of Arianna Miano is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2022

DEDICATION

To my family for always believing in me.

EPIGRAPH

Basic research is what I am doing when I don't know what I am doing.

Wernher von Braun

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are many people that I am deeply grateful for, they helped me succeed as a scientist and they cheered me up during the tougher times. First, I would like to thank my lab partner Michael Liao. He was the best mentor I could ask for, he showed me how to be a great scientist and a big picture thinker. He was a great partner, a great friend and an inspirational co-author. I also want to acknowledge all the people in the Hasty lab that over the years made me feel welcomed and at home. Our long lunches at the table outside made the days spent in lab much more fun. I would like to thank espectially Aida Martin, Omar Din and Andrew Lezia for always being there for me when I wanted to exchange ideas or just chat about life. Most importantly I want to thank my advisor Jeff Hasty who made all of this possible. Jeff welcomed me in his lab when I was just an undergraduate student in exchange from Europe. He gave me a huge amount of trust and freedom which I will be forever grateful for. I want to thank him for always believing in me and supporting any idea and decisions I wanted to pursue. Finally, I want to thank my parents Salvo and Dora, my friends Flavia Roberta and Paola, my roommate Margot for being my rocks during all this years and Edo for being my support while I was deciding on what to do next.

Chapter 1 contains material originally published as Miano, A., Liao, M.J. Hasty, J. Inducible cell-to-cell signaling for tunable dynamics in microbial communities. **Nature Communications** 11, 1193 (2020). The dissertation author was one of the primary researchers and is the first author of this paper.

Chapter 2 contains material originally published as Liao, M.J., Miano, A., Chao, L.,and Hasty, J., 2020. Survival of the weakest in non-transitive asymmetric ecologies. **Nature Communications**, 11 (1), 1-8. The dissertation author was one of the primary researchers and is the first author of this paper.

Chapter 3, in part is currently being prepared for submission for publication of the material. Arianna Miano, Anand Sastry, Kevin Rychel, Bernard Palsson Jeff Hasty. The dissertation author was the primary investigator and author of this material.

2018   Master of Engineering in Bioengineering , Imperial College London

2022   Doctor of Philosophy, University of California San Diego

PUBLICATIONS

Miano, A., Liao, M.J. Hasty, J. Inducible cell-to-cell signaling for tunable dynamics in microbial communities. Nat Commun 11, 1193 (2020).

Liao, M. J., **Miano, A.**, Nguyen, C. B., Chao, L., Hasty, J. (2020). Survival of the weakest in non-transitive asymmetric interactions among strains of e. coli. Nature communications, 11(1), 1-8.

Lezia, Andrew, **Arianna Miano**, and Jeff Hasty. ”Synthetic Gene Circuits: Design, Implement, and Apply.” Proceedings of the IEEE (2021).

ABSTRACT OF THE DISSERTATION

Tools for engineering and modeling bacterial populations dynamics

by

Arianna Miano

Doctor of Philosophy in Bioengineering

University of California San Diego, 2022

Professor Jeff Hasty, Chair
Professor Lin Chao, Co-Chair
Professor Bernhard Palsson, Co-Chair

The last decade has seen bacteria at the forefront of biotechnological innovation, with applications including biomolecular computing, living therapeutics, microbiome engineering and microbial factories. These emerging applications are all united by the need to precisely control complex microbial dynamics in spatially extended environments, requiring tools that can bridge the gap between intracellular and population-level coordination. Therefore, in this thesis we propose genetic circuits for predictable control of microbial communities with potential applications in both healthcare and industry, ranging from therapeutics and drug delivery, to bioproduction and bioremediation. In Chapter 1 we focus on intra-population dynamics control

by designing a genetic circuit which merges the benefits of the synchronization capabilities of quorum sensing to the advantages of external control of chemical inducers. This circuit design can provide more fine-tuned control of cargo release from bacterial carriers by enabling time and space regulation. Subsequently, in Chapter 2 we engineer and characterize inter-species dynamics of a three-strain bacterial community which interacts through cyclic competition, demonstrating the emergence of non-trivial dynamics. In particular, we found that intrinsic differences in the bacterial competitive strength lead to an unbalanced community that was, counterintuitively, dominated by the weakest strain. We used computational tools to model the three strains dynamics in 2D, providing a tool which accurately simulates the competition outcome of the community for any chosen parameter set. Finally, in Chapter 3 we focus on single cell dynamics and developed a bioinformatics pipeline to analyze dynamic transcriptomics data obtained using an innovative high throughput microfluidic device developed in our lab.

Chapter 1

Design of an inducible quorum sensing system

1.1   Introduction

Synthetic biology has the potential to revolutionize both healthcare and industry, with applications ranging from therapeutics[51, 76, 52, 98] and drug delivery[29, 89, 22], to bio-production and bioremediation[91]. These emerging applications have uncovered the need to engineer spatially extended complex multi-cellular populations, requiring new tools that can bridge the gap between single cell, population, and community level engineering[13, 47, 11, 71]. To achieve this, significant research efforts have been focused on engineering and characterizing a variety of cell-to-cell communication systems, with a particular focus on bacterial quorum sensing[110, 20, 109, 64]. Currently, the majority of quorum sensing systems used in synthetic biology rely on self-produced small molecules that result in spatially and temporally self-organized systems, which can not be easily externally regulated[128, 20, 27, 5, 108]. In this study, we propose a tool which combines two pillars of population control: inducibility and cell-to-cell communication. To design this inducible quorum sensing system (iQS), we took inspiration from the native components of the photosynthetic bacterium *Rhodopseudomonas palustris*, which relies on a plant derived organic compound for the production of its signaling molecule[106]. This inducer, p-coumaric acid, is a ubiquitous molecule present in most fruits and vegetables[12] and has proven to be safe for both bacteria[62, 81] and human cells[90, 17]

at relevant concentrations.

The iQS can be coupled with any gene of interest to enable tunable population density-dependent gene expression. We first demonstrated this principle by coupling the iQS system to the production of a fluorescent reporter protein in order to characterize the inducible circuit dynamics. Next, as a proof of concept, we coupled the iQS to a lysis gene, creating a tunable platform for cargo release. In direct comparison to non-inducible quorum sensing system, we demonstrate that the iQS significantly expands the range of population dynamics, allowing for temporal and spatial control of cargo release and population death. Finally, we exploit the orthogonality properties of the iQS system to demonstrate the ability to scale up inducibility from the population to the community level. In fact, it has been shown that the quorum sensing molecule produced by the bacterium *R. palustris* is orthogonal to the majority of well-characterized quorum sensing systems (Lux, Las,Tra, Rhl, Cin), providing a communication channel that can propagate information with minimal signal interference[109, 64]. By co-culturing a two strain community, we demonstrate the ability to control population composition and dynamics by varying inducer concentrations. Overall, the iQS system combines many desirable characteristics into a single genetic circuit: inducibility, tunability, population-level coordination, inducer safety and orthogonality.

1.2 Results

1.2.1 Design and characterization of the iQS system

To design the iQS system, we genetically reconstructed the two-step pathway which converts pCA into p-coumaroyl-HSL (pC-HSL) through the production of the intermediate molecule p-coumaroyl-CoA[57] (Fig. 1.1a) . The first conversion is catalyzed by the p-coumaric acid-CoA ligase encoded by the *4CL2nt* gene from the plant *Nicotiana tabacum*[66] while the second step is catalyzed by the RpaI synthase. Additionally, we added superfolder green fluorescent protein (sfGFP) to be able to monitor the inducible dynamics. All genes, with
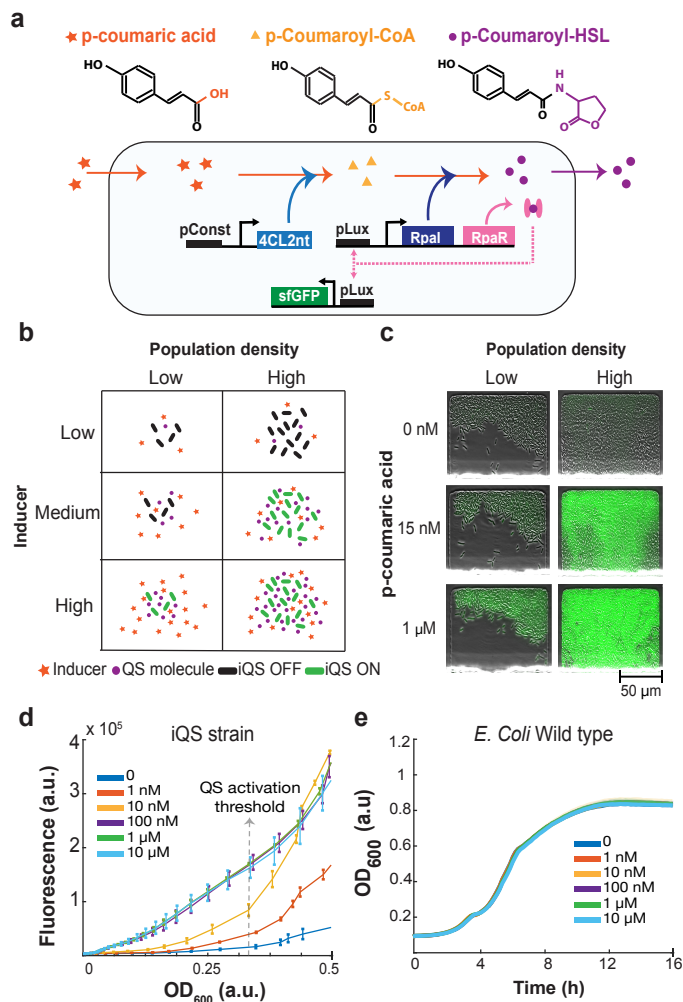
Figure 1.1. Design and characterization of the p-coumaric acid mediated iQS strain. **a** Diagram of the iQS genetic circuit. The chemical structures of the molecules involved in the synthesis of the QS molecule are shown at the top. **b** Diagram to illustrate predicted dynamics associated to the inducible quorum sensing system as a function of population density and external inducer concentration. **c** Fluorescence microscopy images showing a composite of phase-contrast and GFP fluorescence in microfluidic traps. The data from the raw fluorescence values reflect the iQS dynamics predicted in part b. **d** Data from microplate reader experiment obtained by culturing the iQS strain in different p-coumaric acid concentrations. All data points represent mean $\pm$ standard deviation of three independent replicates. **e** Microplate reader experiment data obtained by culturing the wild type *E. Coli* strain in a range p-coumaric acid concentrations. All data points represent mean (solid line) $\pm$ standard deviation of three independent replicates (shaded areas). Source data are provided as a Source Data file.

the exception of *4CL2nt*, are driven by the pLux promoter which has been shown to perform better than the native promoter of *R. palustris* when heterologously expressed[109]. We chose *Escherichia Coli* as the host chassis for our double plasmid circuit (Table 1.1).

We predicted that inducibility would expand the range of population dynamics by including an 'OFF' state at low inducer concentration and an 'ON' state at high inducer concentrations which are both independent of population density. On the other hand, for a range of intermediate concentrations the iQS would behave like a standard QS system by exhibiting the typical population density dependent activation (Fig. 1.1b). To investigate how the engineered iQS strain would relate to these expected dynamics, we used microfluidic devices which enabled the simultaneous exposure of varying pCA concentrations to different subgroups of cells. Time-lapse fluorescence microscopy was used to observe fluorescence expression as a function of population size and inducer concentration. The data matched the expected dynamics of an inducible quorum sensing system when exposed to zero, medium (15 nM) or high (1 μM) inducer concentrations (Fig. 1.1c). Additionally, we further characterized the iQS response over a broader range of pCA concentrations using a microwell plate reader. Zero inducer resulted in the circuit being OFF except for a baseline of leaky expression. Intermediate levels (1 nM and 10 nM) induced the typical population-dependent switch-like behavior with a steep increase in fluorescence signal at an optical density threshold of around 0.3. Finally, high concentrations (100 nM to 10 μM) resulted in a linear relationship between fluorescence expression and population density, confirming the assumption of a population independent 'ON' state (Fig. 1d). As a control, we repeated the same experiment with a wild type *E. Coli* strain and observed that the growth curves were unaffected, confirming pCA is not toxic within the characterized range[62, 81] (Fig. 1.1e).

1.2.2   Characterization of the iSLC in microfluidics

To demonstrate the usefulness of the iQS over non-inducible quorum sensing based circuits, we coupled the iQS with the expression of a lysis gene[29, 15, 53, 22] to create a
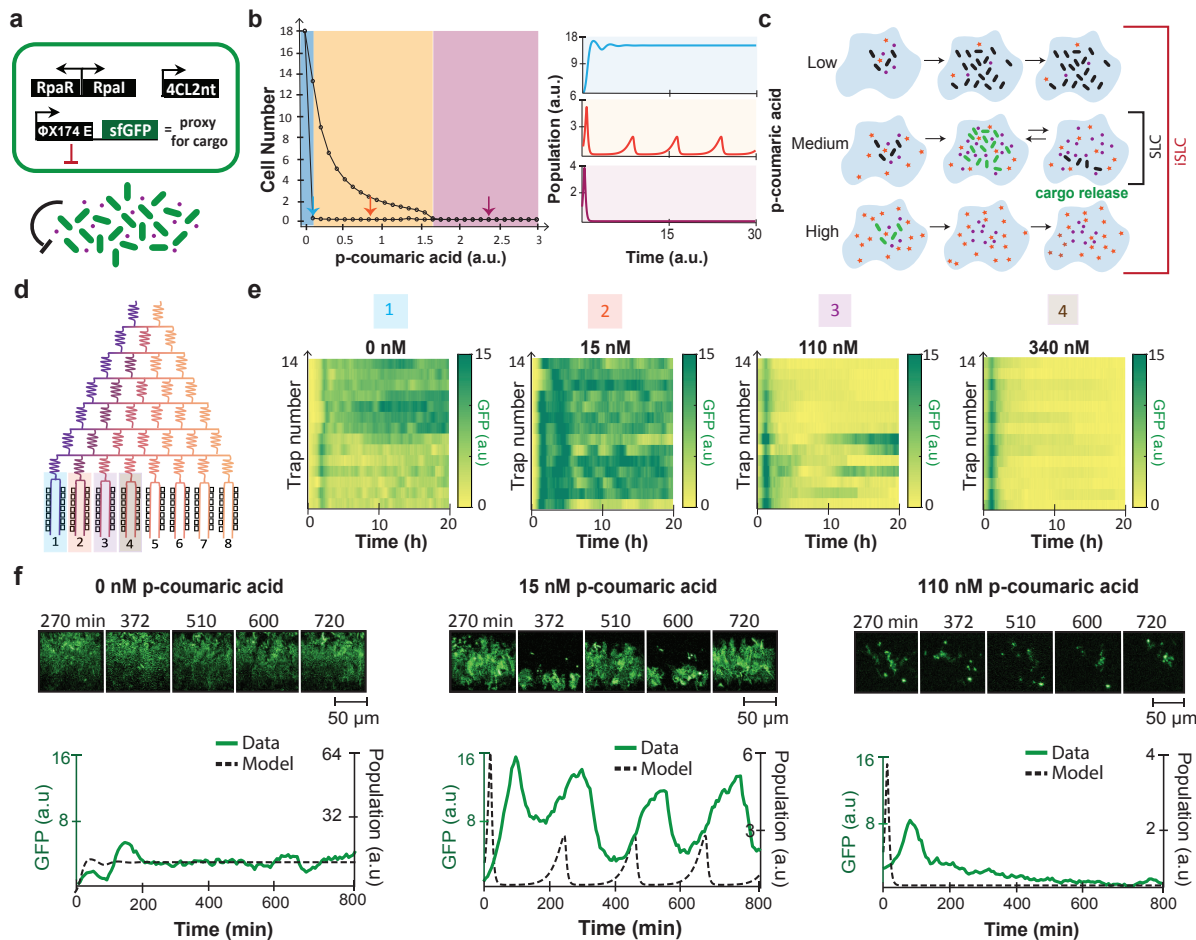
Figure 1.2. Characterization of the inducible syncronized lysis circuit (iSLC). **a** Genetic diagram of the iSLC strain. **b** Simulations of the mathematical model showing three different dynamics at low (blue), medium (orange) and high (purple) inducer values. Medium values are predicted to result in sustained oscillations. Left: steady state maximum and minimum cell population values are plotted for a range of inducer concentrations. Right: simulated time traces for three representative p-coumaric acid values predict three emergent population dynamics. **c** Comparison between the iSLC and SLC dynamics based upon the model simulations in part c. **d** Diagram of the microfluidic device used to generate the inducer gradient. **e** Heatmaps representing the fluorescence time traces of all fourteen traps present per column of the device. GFP signal is used as a proxy for population density. Four different inducer conditions are shown: low, medium, high, extra high respectively. **f** Top: representative time series images from the fluorescence channel with three different inducer concentrations: low, medium, high respectively. Bottom: fluorescence time traces plotted together with computer simulations of the mathematical model. For the simulation (dashed) time units are arbitrary, therefore the correspondence is strictly qualitative. Source data are provided as a Source Data file.

platform for inducible population dependant bacterial cargo delivery (Fig. 1.2a). To qualitatively predict the dynamics of this iSLC strain (induced synchronized lysis circuit), we developed a deterministic model (see details in the section "Methods"). A bifurcation plot was obtained by simulating the steady state values of cell population as a function of pCA concentration, predicting the emergence of three main population regimes (Fig. 1.2b). Small amplitude lysis events followed by steady growth are predicted at low inducer concentrations. Intermediate pCA concentrations result in sustained oscillations in population density. Finally, high inducer concentrations lead to a single lysis event with zero survivors. Therefore, we expected the iSLC to considerably expand the range of possible population dynamics by adding a quiescent state of circuit inactivation and a termination state in which all cells undergo lysis, regardless of population density (Fig. 1.2c).

A preliminary test using a microwell plate reader showed an inverse correlation between the population OD (optical density) at lysis and the inducer concentration (Figure 1.5). Interestingly, we noticed the presence of a lysis event at zero pCA concentration which we proved to be caused by leaky expression of the pLux promoter in the sole presence of RpaR (Figure 1.6). To fully visualize the dynamics of the iSLC strain over time, we used microfluidic devices with an upstream serpentine of branching channels to generate a gradient of eight different inducer concentrations (Fig. 1.2d and Fig. 1.7).

Using fluorescence microscopy to monitor population dynamics, we observed the same three emerging population behaviors predicted by the mathematical model (Fig. 1.2e, f and Supplementary Video 1, 2). With zero inducer, we observed constant cell growth with sporadic asynchronous lysis due to promoter leakiness. Intermediate inducer concentrations (15 nM) resulted in sustained synchronized oscillations with an average period of approximately 200 minutes. Finally, high pCA concentrations (110 nM to 1 μM) caused universal cell lysis regardless of population density, with only a small fraction ($< 3\%$ of all microfluidics traps) able to survive. (Fig. 1.8).

6
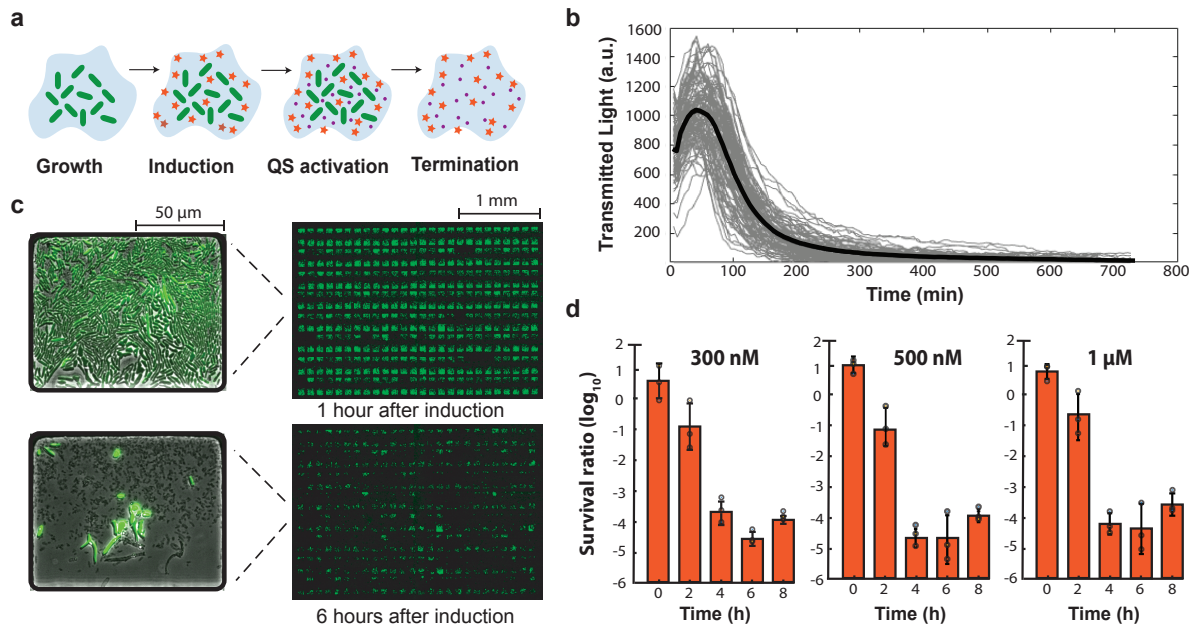
Figure 1.3. Characterization of the iSLC kill switch properties in microfluidics and liquid culture. **a** Illustrated iSLC strain kill switch mechanism. **b** Example time traces (n = 104) extracted from the transmitted light channel (grey). Solid black line represent the mean. At time zero the cells were induced with 500 nM p-coumaric acid. **c** Video stills (4x) of the microfluidic chip before (top) and 6 hours after (bottom) induction with 500 nM p-coumaric acid. Left side shows magnified images (30x) of a single representative trap. **d** Cell viability measured by CFU count following addition of the killing signal (p-coumaric acid) in liquid culture. Individual data points are represented by circles, the bars represent mean ± standard deviation of the three independent replicates. Source data are provided as a Source Data file.

1.2.3 Characterization of the iSLC kill switch

Next, we hypothesized that modulating the inducer concentration in time would enable switching between states of constant growth (circuit quiescence), synchronized oscillations in population density (cyclic cargo release) and inducible population death (kill switch). To confirm this principle, we investigated the ability to drive the population dynamics from state one (population growth) to state three (population death) through pCA induction (Fig. 1.3a). We tested this hypothesis in microfluidic devices by initially growing the cells with zero inducer concentration (circuit quiescence). Following complete saturation of all traps, we induced with

500 nM pCA. We observed a synchronized lysis event throughout the entire device which resulted in cell death in more than 97% of the 406 microfluidic traps present (Fig. 1.3b, c). To test the reproducibility of this property in larger volumes, we grew the iSLC strain in 3 mL culture tubes with and without inducer for a period of eight hours. In the presence of p-coumaric acid, we observed a several order of magnitude decrease in cell survival which was independent of inducer concentration (Fig. 3d and Supplementary Video 3). This ability to bypass population dependency and intentionally decimate the population, regardless of its density, may serve as an integrated kill switch to regulate strain removal in space and time.
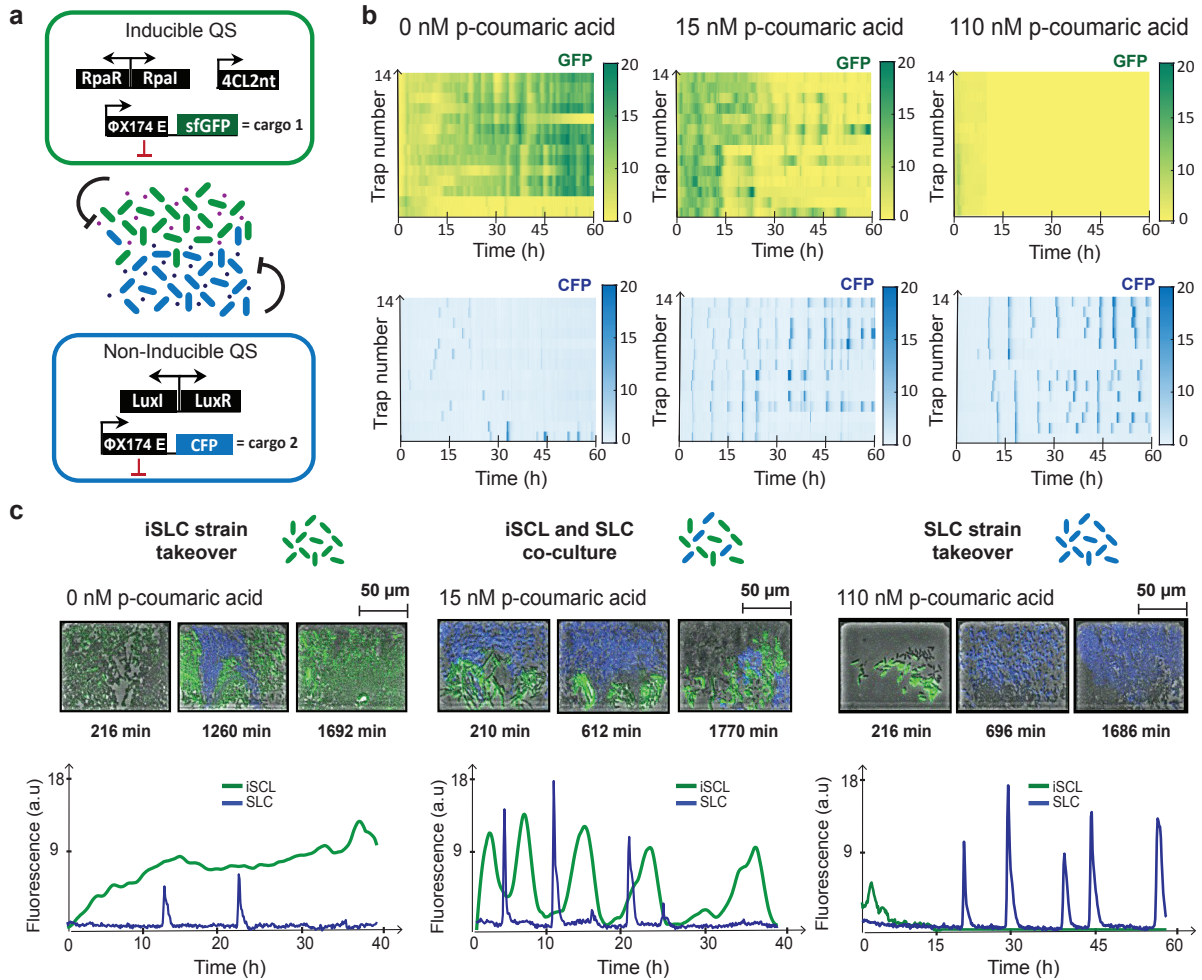
Figure 1.4. iQS enabled modulation of orthogonal multi-strain dynamics. **a** Schematics representing co-culture of the two strains used with orthogonal inducible (iSLC) and non-inducible (SLC) quorum sensing respectively. **b** Heatmaps representing the fluorescence time traces of all fourteen traps present per column of the device. Top rows show the GFP values and bottom rows the CFP values. Fluorescence signals are used as a proxy for population density. **c** Shown at the top are video stills of the co-culture for three inducer concentrations at multiple time points. The corresponding fluorescence time traces for GFP and CFP are plotted at the bottom. Source data are provided as a Source Data file.

1.2.4 Characterization of tunable multi-strain dynamics

Finally, we investigated the orthogonal properties[109, 64] of the iQS circuit by co-culturing the iSLC with a non-inducible SLC strain based on the Lux QS system, one of the most commonly used and well understood QS systems in synthetic biology [32, 27](Fig. 1.4a). First, we demonstrated the absence of crosstalk between p-coumaric acid and the Lux QS system

by growing the iSLC strain in a range of different pCA concentrations (Fig. 1.9). Next, we used the previously described microfluidic chip (Fig. 1.4d and Fig. 1.7) to co-culture the iSLC and SLC strains (starting at 1:1 ratio) under a gradient of inducer concentrations. Interestingly, we observed that by varying pCA concentration, we could precisely control the community composition of the microfluidic traps. Without inducer, the quiescent iSLC strain was able to displace the SLC strain due to an advantage in population growth rate. At intermediate concentrations (15 nM), we were able to maintain stable co-culture of both strains with independently sustained oscillations for more than 40 hours. High pCA concentrations (110 nM) resulted in the death of all iSLC cells, allowing the SLC strain to fully take over (Fig. 1.4b, c and Supplementary Video 4). For inducer values higher than 340 nM, the dynamics were unanimously characterized by iSLC death and SLC takeover (Fig. 1.10). As demonstrated, the absence of cross-talk between the pCA derived signaling molecule and the majority of well-characterized quorum sensing systems, enables the simultaneous use of multiple quorum sensing in co-culture, providing exciting new possibilities for microbial community engineering.

1.3   Discussion

Given the importance of dynamic gene expression in nature[97, 42] and the increasing availability of tools for modular and robust design of genetic circuits[11, 64], synthetic biologists have been attracted towards systems that can achieve tunable complex behaviors, such as the iQS, as opposed to simple steady state dynamics. Our results demonstrate the broad potential of the iQS platform to enable a wide range of new functionality for synthetic circuits that rely on cell-to-cell communication systems for population-level coordination. In particular, we show how inducibility broadly expands the functionality of quorum sensing based circuits by enabling switching between states of circuit quiescence, quorum-dependent circuit activation and population wide constitutive expression. As a proof-of-principle, we created the inducible

10

synchronized lysis circuit (iSLC) which enables spatial and temporal modulation of inducer-dependent cargo release for single or multi-strain communities. In particular, we demonstrate the ability to temporally and spatially control transitions between inactivated population growth, quorum enabled cyclic cargo release and global population death (kill switch). This precise control of circuit activation, the ability for timed strain elimination and the non-toxic nature of p-coumaric acid make this system particularly attractive for potential therapeutic applications *in vivo*. Finally, we exploited the orthogonality of the iQS system to demonstrate precise control of multi-strain dynamics which has potential to become a key tool for engineering synthetic bacterial communities. Although the circuit functionalities of the iQS were demonstrated in the *E.Coli* strain MG1655, we believe that the circuit could potentially be extended to other bacterial species. We expect the main challenge to be the functional expression of the p-coumaric acid-CoA ligase enzyme (encoded by the *4CL2nt* gene) due to its heterologous plant origins. Possible solutions to this challenge might include species specific codon-optimization techniques or the use of homologous proteins.

Looking ahead, the modularity and simplicity of the iQS genetic parts make it straightforward for coupling to any gene of choice, enabling precise spatial and temporal control over population level gene expression. This may open up new possibilities for sophisticated, and safe, biotechnological applications. Overall, this work demonstrates the translation of more than a decade of circuit design into microbiological organization, from molecular regulatory mechanisms (synthetic promoters), to single cell protein expression (enzyme catalysis), multi-cellular population coordination (cell-to-cell communication) and multi-species interaction (orthogonal quorum sensing).

Chapter 1 contains material originally published as Miano, A., Liao, M.J. Hasty, J. Inducible cell-to-cell signaling for tunable dynamics in microbial communities. **Nature Communications** 11, 1193 (2020). The dissertation author was one of the primary researchers and is the first author of this paper.

1.4 Methods

1.4.1 Plasmids and strains

Our iSLC and SLC strains were both cultured in lysogeny broth (LB) media with 50 μg ml-1 kanamycin, 34 μg ml-1 chloramphenicol and 0.2% glucose for strains containing ColE1 origin and p15A origin plasmids in a 37° shaking incubator. Plasmid pTD103 RpaR-RpaI-LAA-sfGFP[108], ptD103-CFP[27] and pZA35-X174E(+LuxR)[29] were constructed by our group in previous studies (Fig. 1.11). Both plasmids pAM014 and pAM021 were obtained by inserting the gene *4CL2nt*[57] under the constitutive promoter J23106 from the Anderson promoter library. The *4CL2nt* gene and the promoter were synthesized with overlapping PCR of long oligos (IDT). All plasmids were constructed by Gibson assembly followed by transformation into DH5α (Thermofisher) chemically competent *E.coli*. All plasmids were verified by Sanger sequencing before transformation into *E.coli* strain MG1655.

1.4.2 Microfluidics and Microscopy

The microscopy and microfluidics techniques used in this study are similar to those previously reported by our group[92]. Our microfluidic devices were constructed from PDMS (poly-dimethylsiloxane) which was molded and baked on a silicon wafer with micron-scale features formed by cross-linked photoresist. Once the PDMS hardened, it was peeled off, and individual devices were cut out. In order to connect fluid lines to the device, holes where punched in correspondence of the inlets and outlets of the device. Afterwards, the devices were bonded onto glass slides using plasma activation. Before each experiment, the devices were left for 30

minutes in a vacuum chamber. Meanwhile, 1 ml of overnight cell culture was spun down by centrifugation and re-suspended in 10 μl of fresh media with appropriate antibiotics. After taking the device out of the vacuum chamber, a single droplet of resuspended cells was positioned in correspondence of the outlet opening. Similarly, droplets of sterile fresh media were placed in correspondence of the inlets openings[33]. In all cases 0.075% Tween20 was added to the medium to prevent cells from sticking to the PDMS walls. After all chip features were wetted, the fluids lines were plugged in and the height of the inlet was raised 10 to 40 cm above the device. The outlet syringe was instead placed at the same height of the device. For co-culturing experiments in Fig. 4, cells were cultured individually overnight and eventually spun down and re-suspended together (1:1 ratio) allowing for a single droplet to be loaded on the device. All experiments shown in Fig. 2 and Fig. 4 were performed in a side-trap array device with bacteria growth chambers approximately 100 x 80 μm in area and approximately 1.2 μm in height. The upstream channels consists of a series of dividing serpentine branches which allow for sequential dilutions of the two input media, generating a gradient of eight different inducer concentrations (Fig. 1.7). On the other hand, for the kill switch experiments, reported in Fig. 3, we used a simpler device with a single input and an ordered array of traps[92]. The dimensions of the traps are the same as the ones described for the gradient device. Experiments in Fig. 2 and 4 where carried out by connecting a syringe with LB + antibiotics + 0.075% Tween20 as inlet 1 and and a syringe with LB + antibiotics + 0.075% Tween20 + 1μM of p-coumaric acid as inlet 2. P-coumaric acid inductions for microfluidic experiments in Fig. 3b were performed by unplugging the syringe with pure media and substituting it with a second syringe containing media plus the appropriate acid concentration. For microscopy we used the same system as described in our previous work[92]. In brief, images were acquired with a Nikon TI2 using a Photometrics CoolSnap cooled CCD camera. The scope and accessories were programmed using the Nikon Elements software. The microscope was housed in a plexiglass incubation chamber maintained at 37° by a heating unit. Phase-contrast images were taken at 4x and 10x magnification at 50-100 μs exposure times. At 4x magnification fluorescence exposure times

were 2s at 30% intensity for both gfp and cfp while at 10x magnification they were 200 μs at 30% intensity for both gfp and cfp. Images were taken every 6 minutes for each experiment. For induction experiments, imaging was paused while syringes were swapped.

1.4.3   Data Analysis

Fluorescence intensity profiles were obtained by analyzing frames from the fluorescent channels. The mean fluorescence values were calculated by drawing a rectangle surrounding each trap individually and extracting the z-axis profile on ImageJ. Fluorescence values shown in Figure 2 were normalized by dividing all data by a constant factor. In addition, the subset of traps for each column was normalized by subtracting the minimum value among the traps within the subset. Transmitted light data from Fig. 3b was normalized by subtracting the minimum value for each time trace. Fluorescence data shown in Figure 4 was normalized by first dividing the subsets of traps by their overall maximum values and subsequently subtracting the respective minimum values. In addition, the data in Figure 4 was smoothed with the command *smoothdata()* in Matlab. Heatmaps were generated in Matlab using the function *heatmap()*. For the co-culture experiments, when overlap between the gfp and cfp channels was observed, values were corrected taking into consideration the image frames in order to subtract overlapping signal.

1.4.4   Plate reader experiments

For plate reader experiments, the appropriate strains were seeded from a -80° glycerol stock into 3 ml LB with 0.2% glucose and appropriate antibiotics and incubated in a 37° shaking incubator. The following day, 2 μl of overnight culture were added to 200μl of fresh media with appropriate antibiotics in a standard Falcon tissue culture 96-well flat bottom plate. Cells were

14

incubated at 37° shaking in a Tecan Infinite M200 Pro. Cells were grown for about 12 hours. The optical density at 600 nm absorbance was measured every 10 min.

1.4.5 Cells survival assay

Cell viability assay to test the efficacy of iQS as a kill switch was done measuring colony forming units (CFUs), following a protocol found in the literature[18]. Cells were grown under survival conditions in LB with 0.2% glucose which inhibits the LuxI promoter thanks to the presence of a binding site for the CAP-cAMP activating complex[78]. In the morning, they were transferred into four liquid cultures of fresh LB medium with 0.2% glucose, 300 nM p-coumaric acid, 500 nM p-coumaric acid and 1 μM p-coumaric acid respectively. Samples were collected every 2 hours and serially diluted in PBS over a 7-log range and spotted (2μL) onto LB agar plates with 0.2% glucose. The equations used are: CFU/ml = (number of colonies) x (dilution factor)/0.002 mL, survival ratio (log_{10}) = log(CFU/ml with glucose)/(CFU/ml with p-coumaric acid).

1.4.6 Modeling

We constructed a deterministic model to qualitatively describe the dynamic behavior of the iSLC strain. The model is based on a set of six ODEs which track the evolution of the following six variables: the cell number into a single microfluidic trap (N), the external concentration of p-coumaroyl-HSL (pH), the intracellular concentration of lysis protein (L), the intracellular concentration of the pC-HSL synthase RpaI (I), the intracellular concentration of p-coumaric acid-CoA ligase (E) and the intracellular concentration of the intermediate compound p-coumaroyl-CoA (pA). The inducer concentration is defined as a fixed parameter.
A non-zero p-coumaric acid concentration induces the production of the intermediate molecule

p-coumaroyl-CoA (pA) through the conversion mediated by the p-coumaric acid-CoA ligase (encoded by gene *4CL2nt*) (E). The production term of the latter is defined as a constant variable according to the constitutive promoter which drives it. The intermediate product (pA) is eventually transformed into the quorum sensing molecule pC-HSL through the RpaI synthase enzyme (I). Once a threshold value of extracellular pC-HSL is reached, the intracellular production of the pLux promoter driven genes (RpaI, RpaR and E) are brought to the ON state. The same promoter also drives the lysis gene, therefore positive feedback also results in cell lysis. We assume that the quorum sensing molecule diffuses quickly through the membrane, therefore we do not distinguish between intracellular and extracellular HSL concentration. In addition, we assume that the pC-HSL-RpaR complex binding is instantaneous, so that the model can be simplified by ignoring the dynamics of the binding complex. Degradation of all proteins (L, I, E) is associated to dilution due to cell growth (μ_G) as well as basal intracellular degradation (γ_L and γ_I). In addition to those terms, RpaI (I) is also actively degraded by ClpXP proteases (γ_C). Overall, our model can accurately predict the three main dynamics of the bacterial population as the inducer concentration is varied. With zero inducer concentration, the population grows reaching a steady state value. Similarly, at very small inducer concentrations, the population undergoes small amplitude lysis events followed by steady state. We observed a finite range of intermediate inducer values which resulted in sustained oscillations of population density. Finally, we observed total population death with no survivors for high concentrations (Fig. 1.12). We can visualize the non-linear dynamics of this system using phase portraits obtained by plotting N (cell number) against L (lysis protein) (Fig. 1.13). As the p-coumaric acid concentration is increased, the simulations show a first transition from a stable spiral to a limit cycle which indicates sustained oscillations. A further increase in the inducer parameter causes the limit cycle to disappear in favor of a stable fixed point. All plots were generated in MATLAB.

$$\frac{dN}{dt} = \mu_G * N * (N_0 - N) - N * \frac{k * L^n}{(L_0)^n + L^n} \tag{1.1}$$

$$\frac{dpA}{dt} = \mu_4 * inducer * E - \gamma_{CoA} * pA - \mu_G * pA \tag{1.2}$$

16

$$\frac{dpH}{dt} = \mu_H * N * I * pA - \frac{u * pH}{1 + \frac{N}{N_0}} \tag{1.3}$$

$$\frac{dL}{dt} = C_l * \left(\alpha_0 + \frac{\alpha_H * (\frac{pH}{H_0})^4}{1 + (\frac{pH}{H_0})^4} \right) - \gamma_L * L - \mu_G * L \tag{1.4}$$

$$\frac{dI}{dt} = C_i * \left(\alpha_0 + \frac{\alpha_H * (\frac{pH}{H_0})^4}{1 + (\frac{pH}{H_0})^4} \right) - \gamma_I * I - \mu_G * I \tag{1.5}$$

$$\frac{dE}{dt} = C_l * P_{const} - \gamma_4 * E - \mu_G * E - \gamma_C * E \tag{1.6}$$

We chose model parameters based on a similar model previously published [29]. Compared to the Lux quorum sensing system, the iSLC showed higher promoter leakiness which was taken into account by increasing the basal production term (α_0). The parameter values used in the model are $\mu_G = 0.2$ (dilution due to cell growth), $N_0 = 20$ (cell capacity of a single trap), $k = 10$ (maximum rate of cell lysis), $L_0 = 1$ (concentration of lysis protein resulting in half maximum lysis), $n = 2$ (Hill's coefficient), $C_l = 0.5$ (copy number of the lysis gene), $C_i = 1$ (RpaI gene copy number), $P_{const} = 20$ (strength constitutive promoter driving gene *4CL2nt*), $\gamma_4 = 2$ (degradation of enzyme p-coumaric acid-CoA ligase), $\alpha_0 = 1$ (pLux basal leakiness), $\alpha_H = 20$ (Lux promoter pC-HSL-RpaR induced production), $H_0 = 4.5$ (pC-HSL-RpaR binding affinity to pLux), $\gamma_L = 2$ (lysis protein basal degradation), $\gamma_I = 2$ (RpaI protein basal degradation), $\gamma_C = 12$ (RpaI protein degradation due to ClpXP), $\gamma_{CoA} = 2$ (p-coumaroyl-CoA basal degradation), $\mu_4 = 30$ (conversion rate of p-coumaric acid into p-coumaroyl-CoA), $\mu_H = 15$ (pC-HSL production rate) and $\mu = 12$ (maximum AHL clearance rate due to flow).

1.5   Supplementary Videos

Supplementary Movie 1

Description: iSLC strain subject to multiple concentrations of pcoumaric acid in the gradient mi-

crofluidic device with traps of dimensions 100 x 80 x 1.2 μm. Images taken at 10x magnification. Multiple inducer concentrations results in different population dynamics.

Supplementary Movie 2

Description: SLC strain subject to three concentrations of pcoumaric acid in the gradient microfluidic device with traps of dimensions 100 x 80 x 1.2 μm. Images taken at 60x magnification. Zero p-coumaric acid results in circuit quiescence with low level expression due to promoter leakiness. Intermediate inducer concentration (15 nM) corresponds to circuit activation with sustained oscillations in population density. High concentration (110 nM) results in total strain killing with no survivors left.

Supplementary Movie 3

Description: iSLC strain kill switch response to 500 nM pcoumaric induction at time zero in microfluidic device with traps of dimensions 100 x 80 x 1.2 μm. Images taken at 4x magnification. Number of culture regions in the chip is 406.

Supplementary Movie 4

Description: Co-culture of iSLC (green) and SLC (blue) strain (seeded at 1:1 ratio) subject to multiple concentrations of p-coumaric acid in the gradient microfluidic device with traps of dimensions 100 x 80 x 1.2 μm. Images taken at 10x magnification. Different inducer concentrations result in three emerging multi-population dynamics.

1.6 Supplementary Data
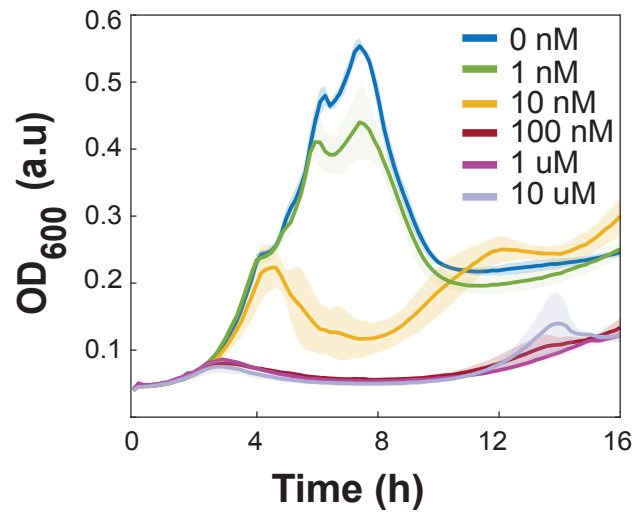
Figure 1.5. iSLC growth curves from plate reader experiments obtained with varying concentrations of p-coumaric acid. Lines and shaded areas represent the mean and standard deviation (n = 3) respectively.
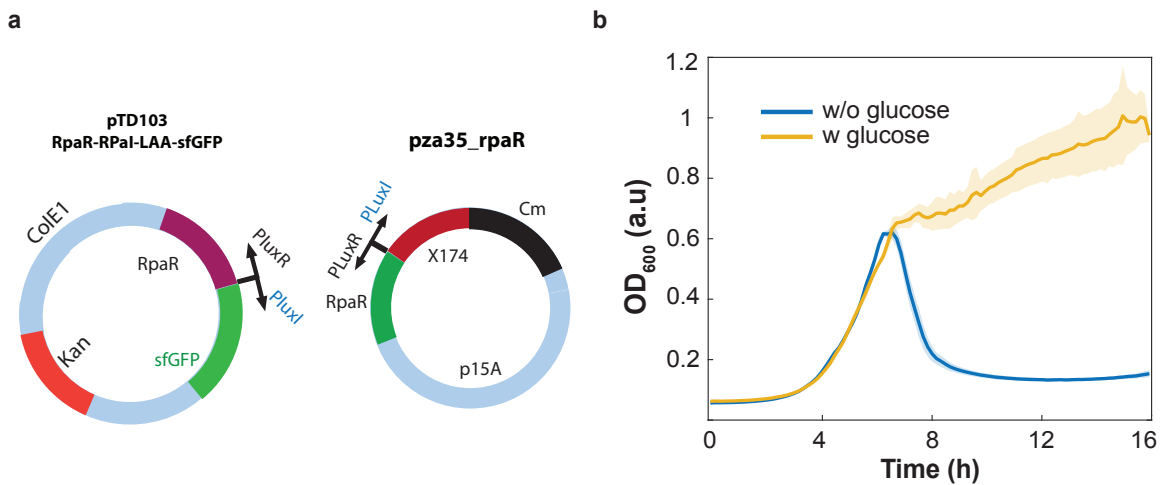
Figure 1.6. a, Plasmids maps from strain LEAKY which lacks the Rpa cassette (RpaI - Plux - RpaR). **b,** Plate reader data obtained by growing strain LEAKY in presence and absence of glucose respectively. High glucose levels in the cell correspond to low cAMP levels which is involved into the luxI promoter activation due to the presence of a binding site for the CAP-cAMP activating complex[78].
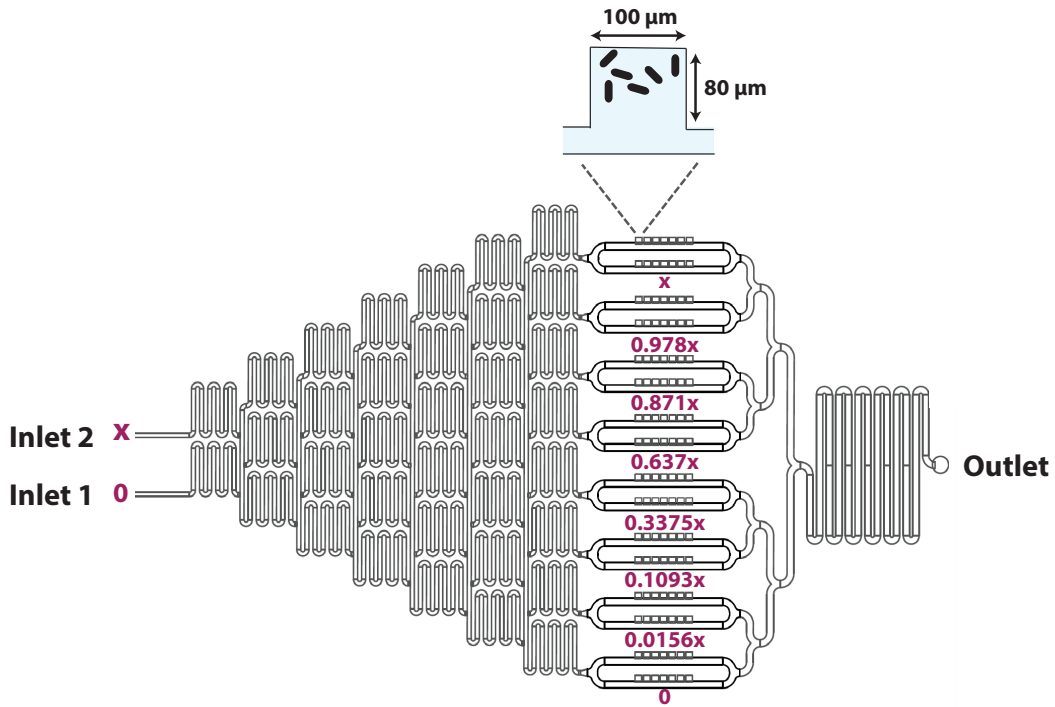
Figure 1.7. Microfluidic gradient device used for experiments in Fig. 2 and Fig. 4. Inlet 1 is set to zero inducer concentration, while Inlet 2 is set to concentration X to generalize dilution factors.
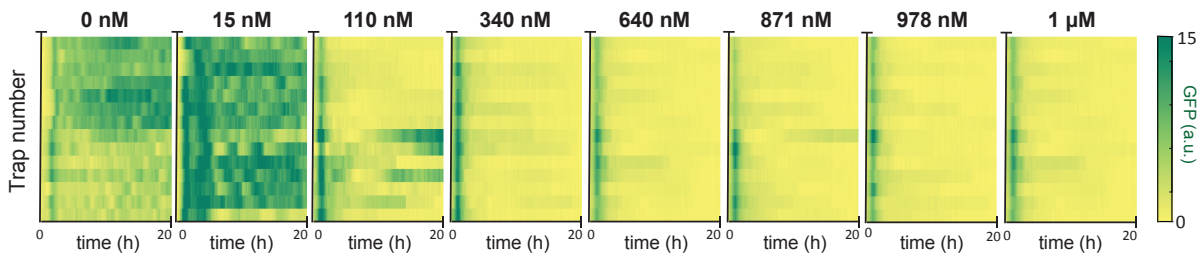
Figure 1.8. Complete heatmaps of fluorescence data for experiment described in Fig. 2. The approximate p-coumaric acid concentration present in each column is reported at the top. For each column, all fourteen traps are reported.
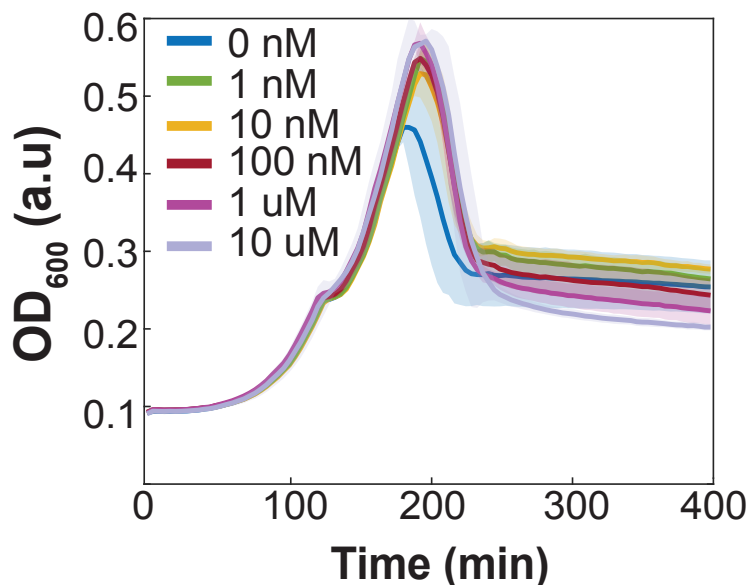
Figure 1.9. Growth curves from plate reader experiments for the iSLC strain. For all p-coumaric acid concentrations tested the growth curves are unaffected, confirming complete orthogonality to the compound. Lines and shaded areas represent the mean and s.d. (n = 3) respectively.

Table 1.1. The strains used in this study.

Strain name	Strain #	Host Strain	Plasmids
AM15	iQS	MG1655	pTD103-RpaR-sfGFP + pAM21
AM09	LEAKY	MG1655	pTD103-RpaR-sfGFP + pZA35 X714E (+RpaR)
AM010	iSLC	MG1655	pTD103-RpaR-RpaI-LAA-sfGFP + pAm014
AM013	SLC	MG1655	pTD103-LuxR-LuxR-LAA-CFP + pZA35 X714E (+LuxR)
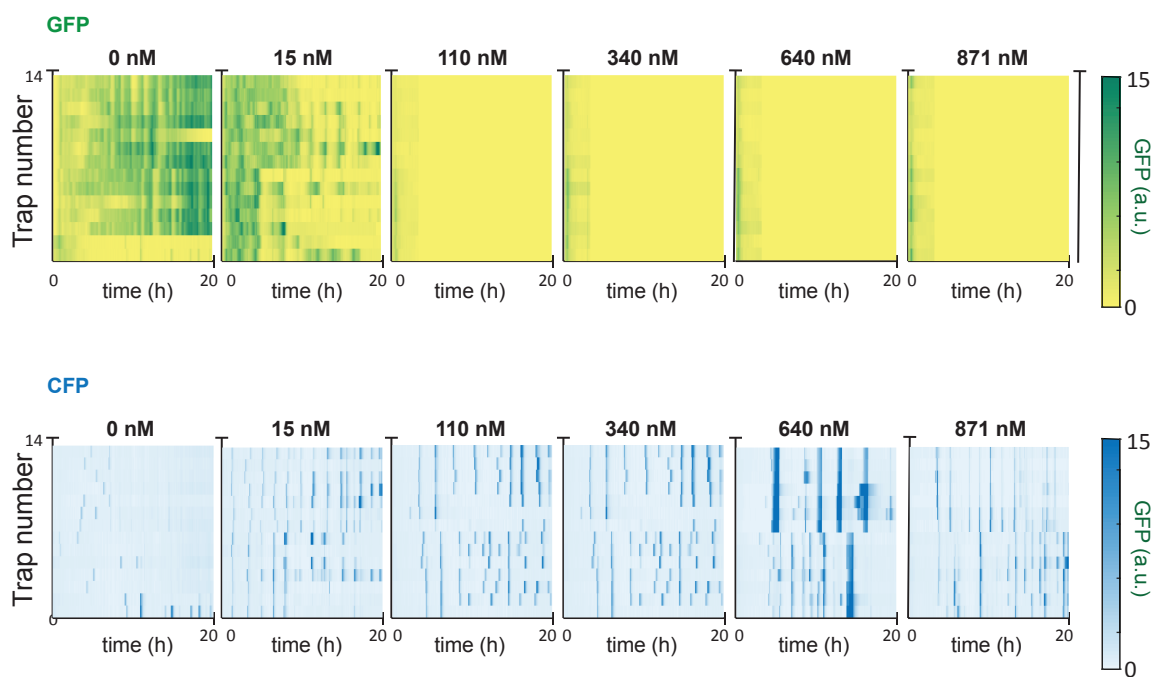
Figure 1.10. Complete heatmaps of fluorescence data for the experiment described in Fig. 4. The approximate p-coumaric acid concentration present in each column is reported at the top. For each column, all fourteen traps are reported. GFP is reported in the top row. CFP is reported in the bottom row.
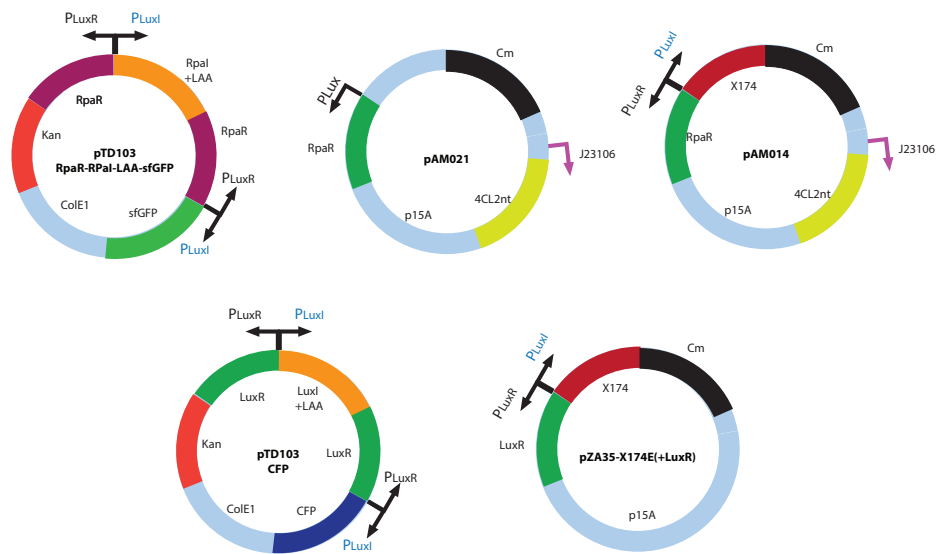
Figure 1.11. Plasmids used in this study. LAA stands for the LAA ClpXP mediated degradation tag.
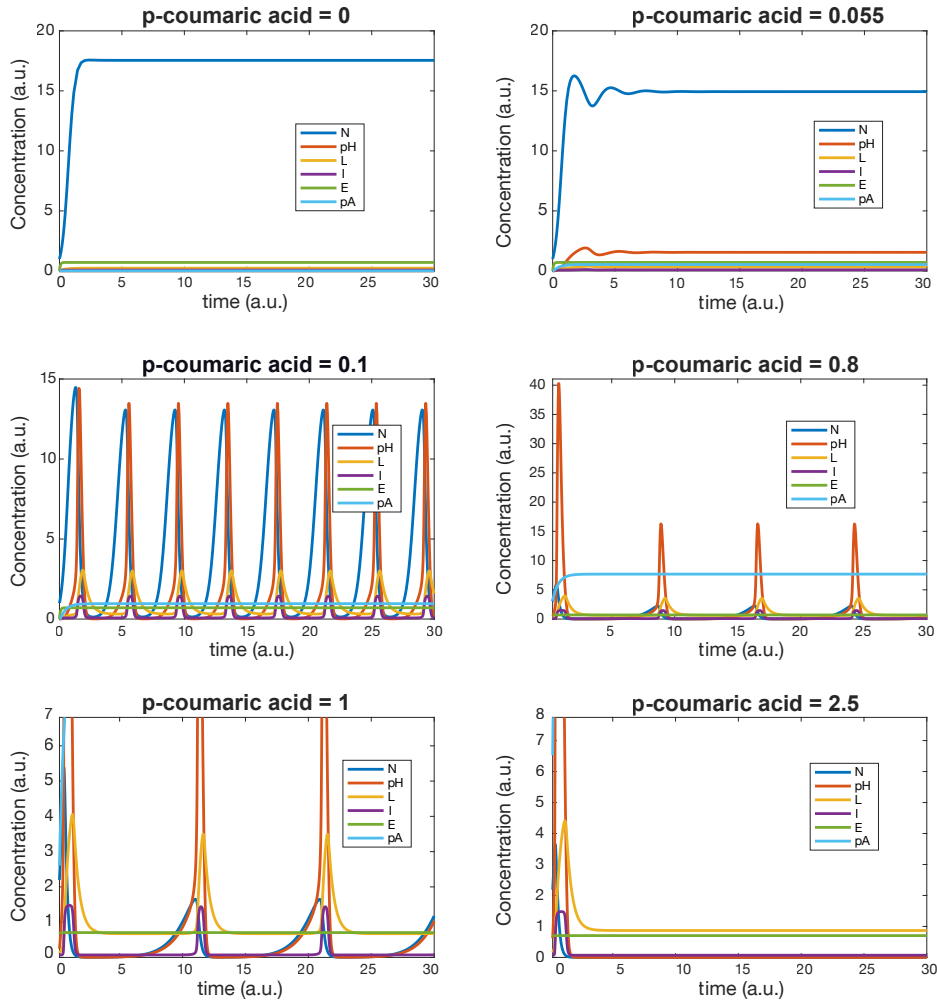
Figure 1.12. Simulation of the evolution of all six variables over time. Six p-coumaric acid concentrations are considered to span the emerging population dynamics. Initally, all the variables are set to zero except cell number (N) which is set to 1.
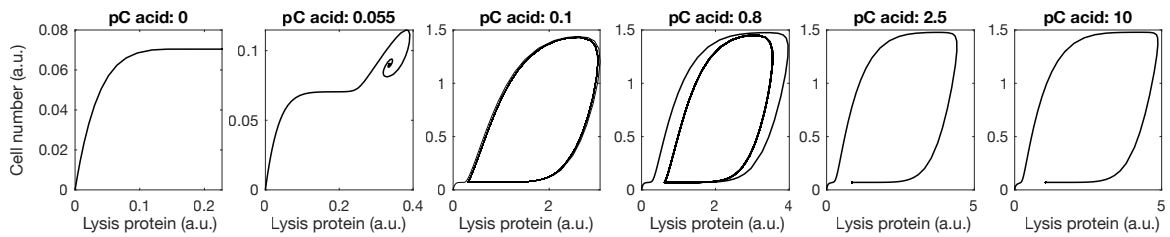
Figure 1.13. Phase portraits of variables N (cell number) and L (lysis protein) showing a Hopf bifurcation caused by the appearance and disappearance of a limit cycle.

Chapter 2

Characterisation of a three-strain non-transitive asymmetric ecology

2.1   Introduction

Inter-species interactions form a complex web that drives ecological dynamics[67, 73, 123, 21, 113]. Competition, in particular, has been hypothesized as a driving force for the evolution and maintenance of biodiversity within various ecosystems[19, 49, 4]. As opposed to a hierarchical competitive structure, previous theoretical studies have shown that species diversity may be promoted by cyclical non-transitive interactions, which describe interactions where there is no single best competitor, but rather the network of species competition resembles a loop[45, 96, 26, 111, 54, 80, 100]. The most simplified example of this system can be described as a basic game of rock-paper-scissors. In this system, rock beats scissors, scissors cuts paper, and paper beats rock, resulting in cyclic competition with no hierarchical organization. Ecologies based on this type of interaction have been observed in various natural settings, ranging from desert lizards[111], and costal reef invertebrates[14] to bacterial communities[59, 26, 60].

Of particular interest within this field of ecology is to better understand the underlying mechanisms that contribute to the stability of these ecosystems[59, 60, 48, 9]. While the application of theory in ecology is still limited, the principles of a non-transitive ecology were first demonstrated experimentally using three isolated bacterial strains consisting of a toxin

producing, toxin sensitive, and toxin resistant strain[59]. In this natural rock-paper-scissors ecology, the toxin producer could kill the toxin sensitive strain, the toxin sensitive strain could outgrow the toxin resistant strain, likewise the toxin resistant strain could outgrow the toxin producing strain. This simple, non-transitive triplet demonstrated that when interactions among the strains remained local, cyclic competition could maintain species diversity. However, because this study only focused on a short observation duration of seven days, the ability of non-transitive competition to maintain biodiversity over a ecologically relevant duration remains unclear. Additionally, because this previous study did not provide a characterization of the relative competitive advantages of each strain, it is difficult to relate this system with other non-transitive ecologies. Intuitively, one would expect competitive asymmetry within such an ecology. For example, a toxin producing strain might outcompete a toxin-sensitive strain at a faster rate than a toxin sensitive strain could outgrow a toxin resistant strain. Due to this potential asymmetry of competitive advantages within this system, it is unlikely that coexistence would be maintained over a longer experimental duration.

In this work, we investigate how asymmetric competitive advantages affect a three strain non-transitive ecology, expanding upon the previous experimental work using rationally engineered and well characterized strains of *E. coli*. Specifically, we engineer three strains of *E. coli* to produce and release three different toxins that kill other members of the same species that lack the production of the corresponding immunity protein. In order to create a one directional rock-paper-scissors competition dynamic, each strain is engineered to produce an additional second immunity protein, providing immunity to one other strain in the ecology. The competitive relationships between each of the strain pairs are characterized in pairwise competition in order to establish relative competitive advantages and identify competitive asymmetry within the system. Using this characterized model microbial system, we explore the outcome of non-transitive competition in a solid growth medium environment (agar) where competitive interactions remain local. We find that given equal starting fractions among the three strains, the strain with the

26
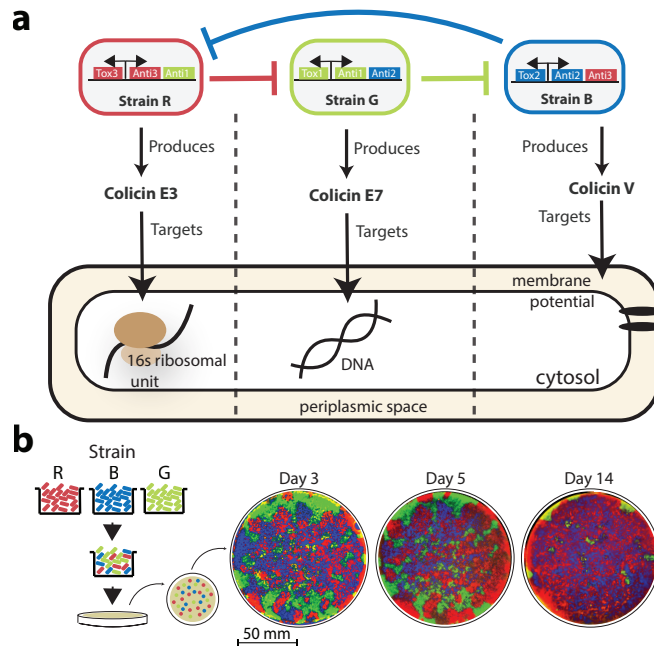
Figure 2.1. Cyclic non-transitive dynamics enable biodiversity **a)** Diagram of the engineered ecology strains including one toxin and two immunity genes. Each toxin produced targets a different essential biological component of *E. Coli* cells. **b)** The three strains were combined at an equal ratio and immediately plated on agar. Each day corresponds to a passage obtained through replica plating. Image stills were obtained by fluorescence imaging. This experiment was executed one time at this density, but it was repeated at multiple densities as shown in Supplementary Figure 2.

weakest competitive advantage consistently dominates the ecosystem. Additionally, we develop a computational model to investigate the relationship between asymmetric competition and parameter space. In particular, we examine the effect of varying the relative competitive strengths among the three strains and find that within a certain parameter space, a steady state three strain equilibrium can be reached by the ecology, albeit at different fractional occupancies. Finally, we demonstrate that there are initial conditions such as different patterns of initial distributions that can contribute to, or prevent, the establishment of steady state coexistence.

2.2   Results

We created a synthetic ecology of three bacterial populations which compete with each other through cyclic non-transitive interactions[71]. In order to create this three-strain ecology

where each *E. coli* strain possesses a competitive advantage over another, we used a class of naturally occurring peptide toxins called colicins. These peptides are lethal to certain *E. coli* strains and provide the producing strain with the ability to inhibit growth of competing strains[16]. Among colicins, the mechanisms of action fall within three major categories: disruption of protein production, degradation of DNA, and disruption of the cell membrane[16]. In order to mimic this natural competition diversity, we designed each strain to produce a different colicin that acts through each of the three mechanisms. Additionally, in order to establish a cyclic rock-paper-scissors competitive relationship, each strain was given a second immunity to the toxin produced by the targeted strain (**Fig. 2.1a**). In this system, Strain Red (R), contains a plasmid producing colicin E3 and E3 immunity protein, a fluorescence reporter protein, a lysis protein (for toxin release), and a secondary E7 immunity protein. Strain Green (G) contains the same plasmid structure expressing colicin E7, and the E7 immunity protein, and a secondary Col V immunity protein. Strain Blue (B) contains the wild-type Col V operon consisting of the toxin (*cvaC*), immunity (*cvi*), as well as export proteins (*cvaA* and *cvaB*)[38], and a secondary E3 immunity protein (**Fig. 2.5**). Due to their different mechanisms of action, inadvertent cross immunity due to structural or functional homology between related colicins was minimized.

The first experiment we carried out to investigate our three strain ecology consisted of mixing the strains, from liquid cultures, at an equal ratio and plating them on a static agar plate environment. The choice to carry out all our experiments on solid media was based on the previously reported observation that local interactions and dispersal promote diversity within a community[59]. After incubation, the bacteria lawn on the agar plate was passaged by replica plating every 24 hours to guarantee a constant source of fresh nutrients and uncolonized space to invade (**Fig. 2.1b**). We tested a range of initial plating densities spanning a 10^4 fold change (**Fig. 2.5a**). We found that the duration of coexistence was inversely proportional to the initial plating densities (**Fig. 2.5b**). This outcome is attributed to the ability of the lower density cells to establish larger colonies due to the higher availability of free space surrounding them. In this scenario, cells at the colony boundary are able to shield cells in the interior from toxin exposure.
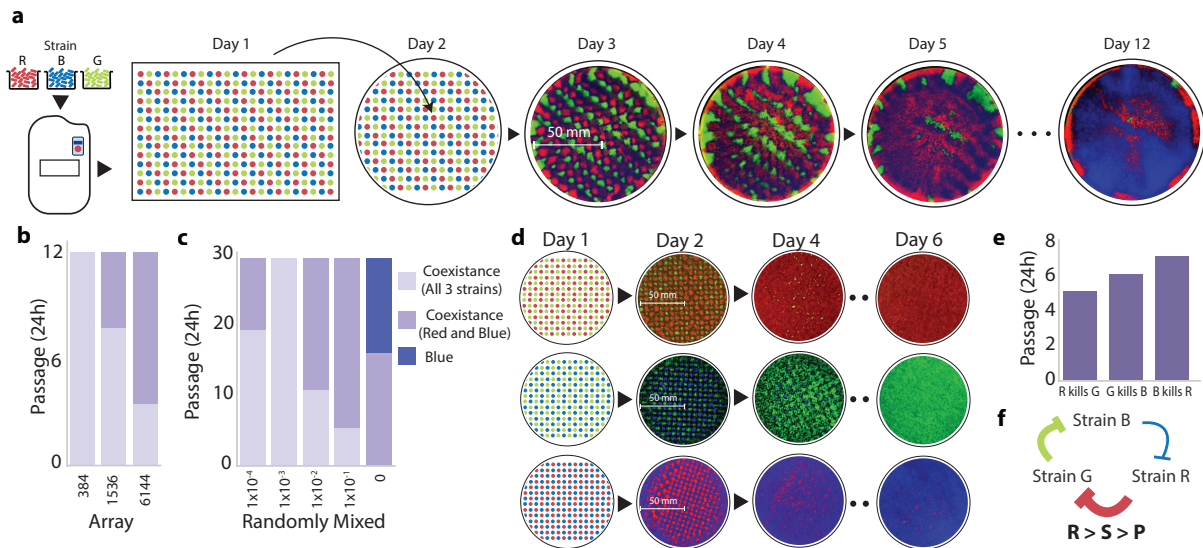
Figure 2.2. Characterization of the ecology and causes of asymmetry. **a)** Schematic diagram of the method used to create an ordered grid of RGB colonies on agar plate through the use of a liquid handling robot. On the right, images corresponds to passages of the lowest density grid (384). Passaging of plates was done by replica plating. This experiment was executed one time at this density and it was repeated at a higher density as shown in Figure 2.7. **b)** Quantification of strain coexistence as a function of time for three different initial densities plated in grid format. **c)** Quantification of strain coexistence as a function of time for five different initial densities plated in a random distribution. **d)** Paired competition on agar plates starting from an initial grid at 1536 density. **e)** Quantification of time to takeover for each competing pair. **f)** Diagram illustrating asymmetry in toxin strength in the RPS ecology. Strain R produces strong inhibition of Strain G, Strain G produces intermediate inhibition of Strain B, and Strain B produces weak inhibition of Strain R.

Interestingly, we observed one scenario (1 to 1000 dilution) in which all three strains coexisted for the entire duration of the 30 day experiment (**Fig. 2.5c and Supplementary Video 1**). In this scenario, Strain B was nearly able to completely colonize the agar plate, however a small patch of Strain G cells managed to survive the initial attack from Strain R. As a result, Strain G was able to invade into space colonized by Strain B, and Strain R could invade into the newly acquired area of Strain G. Nevertheless, in the four other scenarios we tested, the system always converged to a single winner, Strain B (**Supplementary Video 2**). We wondered whether the survival of Strain G in the 1 to 1000 dilution was an outlier that occurred because of random distributions in the initial plating, or whether coexistence of this system of 3 strains was an
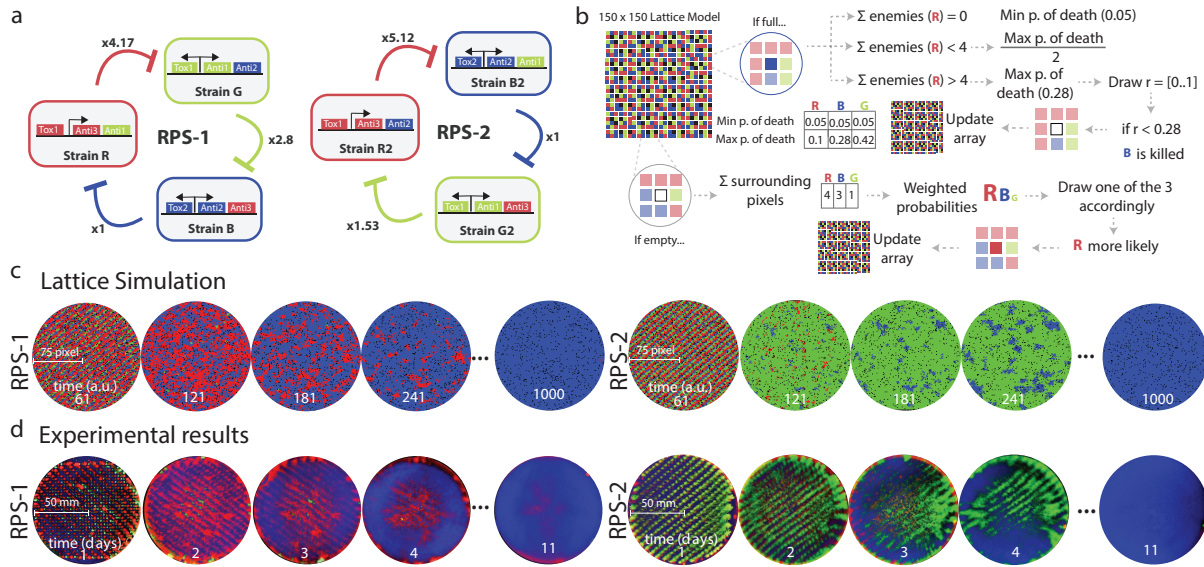
Figure 2.3. Characterization of reversed RPS ecology and computational model. **a)** Diagrams of RPS (RPS-1) and RPS reversed (RPS-2) in comparison. The numbers represent the relative toxin strength obtained from kill curve experiments. **b)** Pseudocode illustrates the rules behind the lattice model simulations. **c)** Comparison of model simulations of RPS-1 (left) to model simulations of RPS-2 (right) with a grid of density corresponding to 1536. **d)** Comparison of experimental results of RPS-1 (left) to experimental results of RPS-2 (right) with density 1536. All lattice simulations plots were simulated in a square lattice grid and subsequently trimmed into circular shapes to ease the comparison with experimental results. This experiment was executed one time at this density and it was repeated at a lower density as shown in Figure 2.7 and 2.12.

expected outcome under these conditions. Therefore, we designed an experimental protocol which could guarantee the sequential placement of the three strains in individual colonies which are equally spaced from each other like the intersections of a grid.

2.2.1   Enemy of the strongest or predominance of the weakest?

In order to achieve precise control of our experimental conditions we used a liquid handling robot to array the three strains into equally spaced grids of varying initial densities. After arraying the cells into grid format on agar plates, the strains were grown for 24 hours, then passaged every subsequent 24 hours by replica plating (**Fig. 2.1a**). The results of this experiment were in agreement with our previous observations. As before, the initial density was inversely related to the duration of coexistence of the three strains, and Strain B was able to take over the

agar plate (**Fig. 2.1b and Fig. 2.7**). Because competition within microbial communities may be affected by differences in growth rate, we first measured the growth rate of each RPS strain and found that the strains had similar growth rates (**Supplementary Fig. 2.8a**). Therefore, we hypothesized that the reason Strain B was consistently winning was due to asymmetry among the potency of the toxins produced by the three competing strains. In order to determine the source of this asymmetry, we characterized the interactions between each pair of competing strains using the same grid pattern arrangement (**Fig. 2.1d**). We found that Strain R produced the most lethal toxin, as it was able to completely eliminate Strain G prior to day 5. The next strongest toxin produced by Strain G was able to completely eliminate the Strain B by day 6. Finally, we found Strain B to be the weakest toxin producer, taking 7 days to completely eliminate Strain R (**Fig. 2.1e**). The resulting ranking of relative toxin strengths was also demonstrated in kill curve experiments conducted in liquid culture (**Fig. 2.8b, c**). Therefore, we confirmed the asymmetry in the system, whereby Strain R was shown to be the strongest and Strain B the weakest (**Fig. 2.1f**).

Following this observation, our hypothesis was that since Strain R was the strongest toxin producing strain, it was quickly, and completely, eliminating Strain G. As a result, when Strain B and Strain R were the only two left competing, Strain B was eventually able to win due to its immunity and toxicity to Strain R. In this first hypothesis, Strain B would win due to its position as the enemy of the strongest strain (**Fig. 2.9a**). However, we also considered the counterintuitive possibility that Strain B was consistently winning because it was producing the weakest toxin (**Fig. 2.9b**). In order to test this theory, we created a new three strain ecology (RPS-2) in which we reversed the order of the rock-paper-scissors competitive relationship by swapping the secondary immunity genes for each strain. With these new constructs, Strain R2 now attacked Strain B2, Strain B2 attacked Strain G2, and Strain G2 attacked Strain R2. Although we reversed the order of competition by reversing the secondary immunity, the relative toxin strengths remained the same (**Fig. 2.3a and Fig. 2.10**). Therefore, while Strain B2 remained the weakest strain, the enemy of the strongest then became Strain G2. If our initial hypothesis of the winner being the

31

enemy of the strongest were correct, we would expect Strain G2 to eventually take over the plate. Alternatively, if "predominance of the weakest" were the driving force, then we would expect Strain B2 to win in RPS-2 as well (**Fig. 2.9b**). Simultaneously, we developed a lattice-based computational model to simulate competition of the three strains in a 150x150 square grid (**Fig. 2.3b**). With the parameter values established in our kill curve experiments, simulations of our computational model agreed with our previous experimental results demonstrating that Strain B should outcompete the other strains (**Supplementary Video 3**).

The simulations also confirmed our observation regarding the relationship between duration of coexistence and initial density (**Fig. 2.11**). Furthermore, when simulating the RPS-2 system, the model predicted that the eventual winner of the RPS ecology would still be Strain B, confirming the idea of "predominance of the weakest" (**Fig. 2.3c and Supplementary Video 4**), a theory that had been hypothesized by previous theoretical works[34, 46, 75, 114, 2, 79]. As predicted, we found experimentally that the final winner of the RPS-2 ecology was indeed Strain B2 (**Fig. 2.3d, Fig. 2.12 and Supplementary Video 5**). This outcome occurs because the weakest strain allows its target (Strain R in RPS-1 and Strain G2 in RPS-2) to rapidly expand and eliminate the third strain in the system (Strain G in RPS-1 and Strain R in RPS-2). When only two strains are left, the "weakest" strain is then able to slowly take over due to the rock-paper-scissors dynamics (**Fig. 2.13**). While we were enthusiastic to experimentally demonstrate the theory of "predominance of the weakest", we still wondered about the scenario we observed in which all three strains were not fully eliminated, and seemed to have developed stable existence. We therefore hypothesized that there must be alternate possible outcomes that arise based on probability for a given set of toxin parameters.

2.2.2 Spatial patterns and parameter space

To further explore this concept, we investigated the relationship between toxin strength and the system steady state. We used our computational model to do a parameter sweep of toxin production strengths. For each combination of parameters, we ran 100 simulations. We
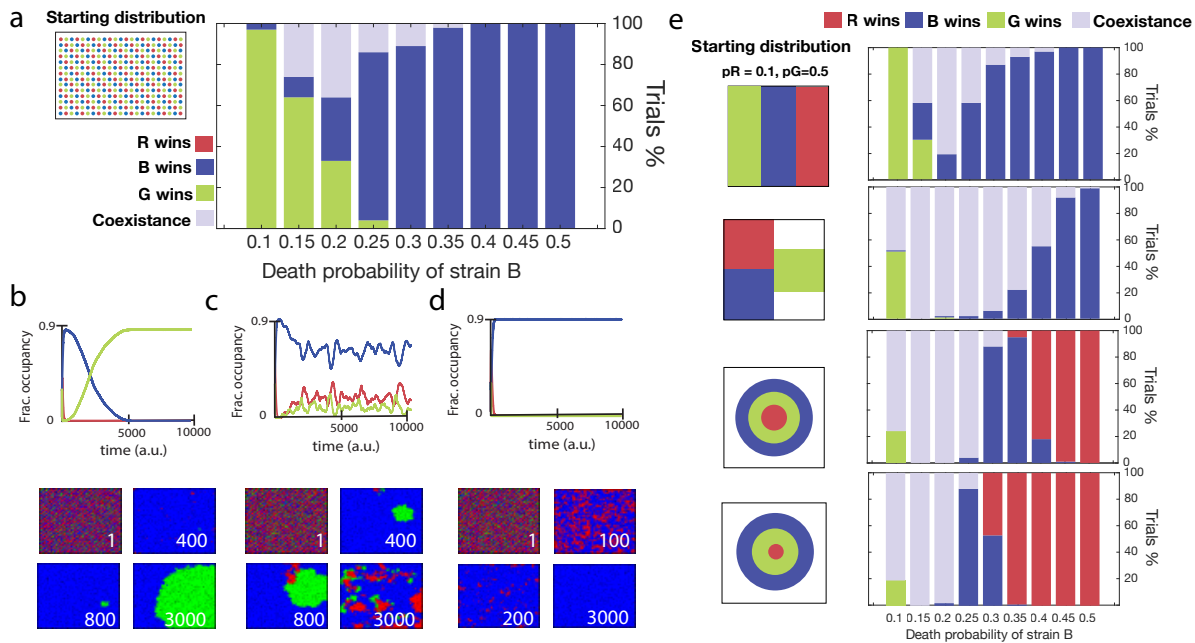
Figure 2.4. Model simulations exploring spatial patterns and parameter space. **a)** In this simulation the starting condition is an ordered array of alternating strains. The bar plots shows the outcome of 100 trials for multiple parameters of Pb (Probability of death of strain B). For all simulations, the probability of death of strain R (0.1) and strain G (0.5) are kept constant. **b)** Time series representing the possible scenario where strain G wins (left-hand side of bar plot in part a). Below, the corresponding image stills at multiple time points. **c)** Time series representing the possible scenario where strain R B G coexist (left-hand side of bar plot in part a). Below, the corresponding image stills at multiple time points. **d)** Time series representing the possible scenario where strain B wins (right-hand side of bar plot in part a). Below, the corresponding image stills at multiple time points. **e)** We generated the same bar plots described in part a with different geometries as initial conditions. The bar plots shows the outcome of 100 trials for multiple parameters of Pb (Probability of death of strain B). For all simulations, the probability of death of strain R (0.1) and strain G (0.5) are kept constant.

found that "predominance of the weakest" was not always the most probable outcome. When keeping the toxin strength of both the weakest and strongest toxins fixed while varying the toxin strength of the intermediate strain, the most probable steady state outcome was dependent on the intermediate toxin strength (**Fig. 2.4a**). The model shows that if the toxin strength of the intermediate strain (Strain G) is close in value to the toxin production strength of the weakest toxin producer (Strain B), the most probable outcome is that the intermediate wins (strain G) (**Fig. 2.4b**). On the other hand, when the toxin production strength of the intermediate strain falls

33

within an intermediate range, the model predicts the possibility of stable coexistence between the three strains at different fractional occupancies (**Fig. 2.4c and Supplementary Video 6**). Finally, if the toxin production strength of the intermediate strain is close in value to the strength of the strongest toxin producer (Strain R), then the weakest strain (Strain B) always wins as predicted by the hypothesis of "predominance of the weakest" (**Fig. 2.4d**). Interestingly, we did not observe any case in which the strongest strain (Strain R) won.

When compared to our previous experimental results, our experimental parameter values fall within the parameter space that predicts two possible outcomes. The most probable scenario is that Strain B wins, but there is also a small probability of the establishment of coexistence. In agreement with the model, we observed that on the plate that established coexistence, the three strains maintained different fractional occupancies, with Strain B having the highest fractional occupancy. These results highlight the value of the computational model for examining a wide variety of scenarios that would be too difficult or time consuming to cover experimentally. To this end, we wanted to investigate how different initial fractional occupancies would effect the steady state outcome. We found that, in grid format, the final outcome is highly influenced by the abundance of the weak strain, where an inverse relationship exists between the starting fractional occupancy of the weakest strain and its probability to takeover (**Fig. 2.14**). We also explored how different patterns of initial strain distribution would affect the steady state outcome of the competition in relation to toxin strength. Using the computational model, we simulated multiple scenarios with common spatial patterns such as stripes, isolated clusters, and concentric circles (**Fig. 2.4e**).

We found that stripes led to similar results compared to the previously discussed case with the grid pattern in which the weakest strain is able to mostly dominate the system (**Supplementary Video 7**). On the contrary, having the strains separated in three different clusters resulted in a significant increase in the parameter ranges that could lead to coexistence (**Supplementary Video 8**). As before, in both scenarios we did not observe any case in which the strongest strain (Strain R) could win. On the other hand, when initially distributed in the

pattern of concentric rings, it was possible for the strongest strain (Strain R) to take over the plate (**Supplementary Video 9**). This only occurs when the strongest strain is placed in the inner circle, shielded from the weakest strain (Strain B). This way, if the intermediate strain in the middle manages to kill off the outer strain before Strain R reaches it, then the strongest strain can win. As expected, we also demonstrate that this outcome is dependant upon the ratio between the ring dimensions and the rate of takeover of each strain (**Supplementary Video 10**).

2.3 Discussion

Bacterial communities occupy a myriad of diverse niche environments, playing important roles in processes ranging from nutrient recycling[118, 74] to the regulation of human health[25, 24, 35]. While the ability to engineer robust bacterial communities could lead to major advancements in fields such as recycling, sustainability, and healthcare, the mechanisms underlying species diversity and stability are still not yet well understood. Although cooperative interactions between the species comprising such communities contribute to ecological stability[122, 31, 95], it is generally accepted that competition is the guiding force[99, 25, 48, 69, 119]. However, due to the complexity of natural ecologies, these types of competitive interactions are difficult to isolate or quantify in nature. Our study demonstrates the feasibility of using an engineered synthetic ecology to simplify complex community relationships in order to study underlying mechanisms that may lead to community stability and the maintenance of diversity.

Due to the immense species diversity and wide range of competitive strategies organisms employ in nature[48, 129, 119], we hypothesized that natural competitive dynamics are likely to be unbalanced[112, 28]. Unlike a perfectly balanced game of rock-paper-scissors in which each of the three species could kill each other at an equal rate, we focus on characterizing an asymmetric system in which the relative competitive advantages among each predator-prey pair varies. While previous studies focused on the coexistence of non-transitive communities over

a relatively short timeframe[59], they do not allow the system to reach steady state, in which case it is likely that only one species would remain. Here, we demonstrate that intransitivity fails to promote biodiversity over a long time horizon when the relative competitive advantages are imbalanced [39]. Therefore, we believe that an asymmetric non-transitive ecology is a useful base model to study complex interactions among competing bacterial species.

Using our three-strain ecology we experimentally demonstrate that a uniformly distributed asymmetric game of rock-paper-scissor is most likely to be won by the weakest species[34, 46, 75, 114, 2, 79]. Interestingly, we show that an asymmetric ecology can develop steady state coexistence, and that the relative toxin strengths among the three species dictate the extent of the coexistence space. Counterintuitively, under the same initial conditions of uniform distribution, our models predict that the producer of the most lethal toxin never wins. As opposed to pairwise competition, where the producer of the strongest toxin has a competitive advantage[19], the producer of the strongest toxin in non-transitive communities are at an evolutionary disadvantage. This could be an important selective force against continuous evolution of ever-more lethal warfare chemicals in microorganisms, resulting in increased diversity of chemicals that are constrained to a specific toxin strength parameter space. The role that toxin mediated competition plays in community stability may also explain the observed relative abundance of membrane targeting, DNA, or ribosome targeting bacterial toxins among bacterial communities[72].

Finally, we also observe that the steady-state outcomes of the system can be altered by changing the initial strain distribution patterns. For example, we find that separate blocks in a triangular conformation can greatly expand the parameter space for coexistence, supporting the idea that spatially separated niches are more likely to sustain biodiversity[68, 50, 6]. On the other hand, strains initially distributed in concentric circles can enable the strongest toxin producer to win. These results demonstrate that many factors need to be considered if the goal is to design stable synthetic ecologies in an environment where interactions are local[19]. Overall, this study provides a mathematical model and engineering framework to study competitive interactions, gain mechanistic insight, and ultimately, predictive power that can be used as a guide to design

stable communities.

Chapter 2 contains material originally published as Liao, M.J., Miano, A., Chao, L.,and Hasty, J., 2020. Survival of the weakest in non-transitive asymmetric ecologies. **Nature Communications**, 11 (1), 1-8. The dissertation author was one of the primary researchers and is the first author of this paper.

2.4 Methods

2.4.1 Strains and plasmids

Our strains were cultured in lysogeny broth (LB) media with 50μg ml-1 spectinomycin for the WT and TP strains, in a 37° shaking incubator. The plasmids used in this study are described in (**Supplementary Fig. 1**). The colicin E3, Im3, colicin E7, Im E7, and E1 lysis genes were taken from previously used plasmids[71] using PCR and assembled with Gibson assembly[37]. The 4.5kb colicin V expression cassette comprising *CvaA*, *CvaB*, *CvaC* and *cvi* was isolated from the pHK11 plasmid from the wild type colicin V strain ZK503 by PCR[38]. All plasmids were transformation into DH5α (Thermofisher) chemically competent E.coli and verified by Sanger sequencing before transformation into E.coli strain MG1655. The strains used in this study are described in (**Supplementary Table 1**). The gene sequences used in this study are described in (**Supplementary Table 2**).

2.4.2 Growth rate

For growth rate experiments, the appropriate E.coli strains were seeded from a -80° glycerol stock into 2ml LB and the appropriate antibiotics and incubated in a 37°C shaking incubator. After cells reached an OD600 of 0.1, 1ml culture was added to a 125ml erlenmeyer

flask containing 25ml fresh media with appropriate antibiotics and left shaking at 270 rpm. Once the samples reached an OD600 of 0.1 samples were taken every 20 minutes and measured at OD600 using a DU 740 Life Science Uv/vis spectrophotometer.

2.4.3 Toxin validation

To prepare colicin lysate, the colicin E3 *E.coli* strain was seeded from a -80° glycerol stock into 2ml LB and incubated in a 37°C shaking incubator. After cells reached an OD600 between 0.4-0.6, 1ml of the grown culture was collected in a 2 ml eppendorf tube and two cycles of incubation at 98°C for 5 minutes followed by 10 minutes at -80°C were performed. The resulting media was then filtered and collected using a 0.22μm syringe filter. For toxin co-culture experiments, wild type MG1655 *E.coli* strains were seeded from a -80° glycerol stock into 2ml LB and incubated in a 37°C shaking incubator. After cells reached an OD600 between 0.2-0.4, 5μl culture was added to 200μl fresh media in a standard Falcon tissue culture 96-well flat bottom plate. Additionally, 5μl of the purified colicin lysate was added to each well. Cultures were grown at 37° shaking for 19 hours and their optical density at 600nm absorbance was measured every 5 min with a Tecan Infinite M200 Pro.

2.4.4 Plate passage experiments

Each *E.coli* strain was seeded from a -80° glycerol stock into 5ml LB with 50μg ml-1 spectinomycin. After growth for 8-12 h at 37° in a shaking incubator, the culture was diluted 100-fold into 25ml of the same medium in a 50ml erlenmeyer flask and grown until reaching an OD of 0.4 (Plastibrand 1.5 ml cuvettes were used). Strains were then diluted 1:10, 1:100, 1:1000, and 1:10000. 20uL of each strain was then plated into separate regions of 100x15mm

petri dishes containing LB agar and 50μg ml-1 spectinomycin. The strains were then mixed using 10 glass beads and incubated overnight at 37° for 24 hours. Plates were passaged every 24 hours by replica plating onto a fresh agar plate containing the appropriate antibiotics.

2.4.5   Liquid handling robot

As before, each *E.coli* strain was seeded from a -80° glycerol stock into 5ml LB with 50μg ml-1 spectinomycin. After growth for 8-12 h at 37° in a shaking incubator, the culture was diluted 100-fold into 25ml of the same medium in a 50ml erlenmeyer flask and grown until reaching an OD of 0.75. Upon reaching this OD, 5% glycerol was added to each culture and 45μl of each prepared culture was then transferred into a single 384 well labcyte plate. To plate the grid array we used a Labcyte Echo liquid handling robot to transfer 2.5 nL volume from the source well on the labcyte well plate onto SBS-format PlusPlates containing 42mL LB agar and 50μg ml-1 spectinomycin. The transferred cells were arrayed according to 384, 1536 and 6144 well plate formats in order to create varying densities. The cells were then incubated overnight at 37° for 24 hours. Colonies were transferred by replica plating from the SBS-format PlusPlates onto a standard 100x15mm petri dish containing LB agar and 50μg ml-1 spectinomycin. Plates were passaged every 24 hours by replica plating onto a fresh agar plate containing the appropriate antibiotics.

2.4.6   Kill Curve

To prepare colicin lysate for each strain, the appropriate strains were seeded from a -80° glycerol stock into 5ml LB and incubated in a 37°C shaking incubator overnight. The overnight cultures were then passaged 1:100 into 2mL LB and incubated in a 37°C shaking incubator until

reaching an OD600 of 1.0. 1mL of the culture was then collected in a 2 ml eppendorf tube, centrifuged at 21130 rcf for 10 minutes and the supernatant was passed through a 0.22um syringe filter. For the kill curve assay, the appropriate strains were grown in a 37°C shaking incubator to an OD600 between 0.3 and 0.4 in 5mL LB in a 25mL flask. 500uL of the corresponding colicin lysate was then added to the flask. CFU measurements were taken by plating serial dilutions (n=3) and counting colonies. The first time point was taken immediately before each colicin was added, then every 10 minutes afterwards. For analysis, we took the difference of 5 time points (40 minutes) that corresponded to the maximum change in total alive cell count divided by the initial CFU count of the initial time point in order to calculate the percent cell death. The relative toxin strengths were then found by taking the ratios between the percent cell deaths for each strain pair.

2.4.7   Image Processing

For plate imaging a Syngene PXi fluorescent imager was used. Strains producing sfGFP were captured using the Blue LED Module for excitation and SW032 emission filter, and strains producing mKate2 were captured using the Red LED Module excitation source and Filt 705 emission filter. Images were processed using Image J. Images were converted to 8-bit and background subtracted. In order to assign the false color blue, we inverted the fluorescence values of the sfGFP image using the math function v=-v+v(mean). We then took the difference from the mKate2 image in order to create a mask of the negative space from both GFP and RFP. This image was assigned as the "Blue" channel for composite images.

2.4.8   Modeling

We developed a lattice-based model in Matlab to simulate the competition dynamics between the three strains RGB. The model was based on similar principles previously described in the literature [59]. The lattice is a 150 x 150 regular square lattice with zero boundary conditions. This means that the edges of the lattice are set to zero (absorbing boundary conditions), simulating the physical boundary of the petri dish that prevents cells from expanding beyond it as well as the effect of disregarding cells beyond the boundary of the replica plating. Therefore, the four edges of the square are kept at a value of zero (no cells can grow/expand in that direction) and the grid is updated only for the internal pixel (pixel 2 to N-1 on all sides).The simulations in Figure 2 and 3 were obtained by starting the lattice with a grid array of alternating strains. The remaining lattice points are classified as "empty space". The probability of death for each competing strain is associated to the relative potency of its enemy's toxin. pR, pG and pB refer to the maximum probability of death of strain R, G and B respectively. For each time loop, the lattice array is scanned pixel by pixel (ignoring boundary pixels which have a fixed value of 0) and is updated according to two main rules as shown in Figure 3b. If the pixel considered is empty, the algorithm takes into account the relative occupancy of the 8 neighboring pixels for the three strains R,G, or B. Three probabilities are calculates as the sum of locations occupied by each strain divided by the total neighboring locations (8). Finally, the Matlab function *randsample* is used to choose how to fill the spot according to the previously calculated probabilities. This process simulates expansion due to growth. On the other hand, if a given location is full, the strain is killed with a thresholded probability that is dependent on the number of surrounding enemies present. If the number of enemies is below 4 (corresponding to being $\leq$ half surrounded), the probability of death is capped at half the maximum probability of death for the given strain. On the other hand, if the number of surrounding enemies is greater than 4, the probability of death corresponds to the maximum probability associated to the given strain. In addition, we set a baseline probability of death equal to 0.05 which cumulatively represents the

random death of cells and their removal due to the replica plate passaging. For the simulations shown in Figure 4, each simulation was run for t=10000 and each parameter set was simulated 100 times. Each time point corresponds to a reproduction event (around 25 minutes). Therefore, the time chosen to investigate steady state dynamics corresponds to about 170 days, which we established to be a long enough interval both computationally and biologically. In terms of spatial parameters, the distance between two consecutive pixels represents roughly 1 mm on the agar plate. Therefore the entire grid represents a square with a side of about 10 cm. The relationship between grid densities on the agar plates compared to the model are illustrated in Supplementary Fig. 7.

2.5 Supplementary Videos

Supplementary Movie 1

Image sequence of fluorescent agar plate images for RPS-1 at a 1:1000 dilution. Strains were initially seeded at a 1:1:1 ratio of Strain R, G, B randomly mixed. Images were taken every 24 hours.

Supplementary Movie 2

Image sequence of fluorescent agar plate images for RPS-1 at a 1:10000 dilution. Strains were initially seeded at a 1:1:1 ratio of Strain R, G, B randomly mixed. Images were taken every 24 hours.

Supplementary Movie 3

Lattice model simulation of RPS-1 in a medium density grid. Toxin strength parameters derived from liquid kill curve experimental results. Probability of death of Strain R, B, G was 0.1, 0.28, and 0.417 respectively.

Supplementary Movie 4

Lattice model simulation of RPS-2 in a medium density grid. Toxin strength parameters derived

42

from liquid kill curve experimental results. Probability of death of Strain R, B, G was 0.153, 0.512, and 0.1 respectively.

Supplementary Movie 5

Image sequence of fluorescent agar plate images for RPS-1 (left) and RPS-2 (right) initially spotted in a grid format. The top row shows low density initial seeding. The bottom row shows medium density initial seeding. Images were taken every 24 hours.

Supplementary Movie 6

Lattice model simulation of RPS-1 in a medium density grid. Toxin strength parameters derived from liquid kill curve experimental results. Probability of death of Strain R, B, G was 0.1, 0.2, and 0.5 respectively.

Supplementary Movie 7

Lattice model simulation demonstrating one of the possible scenarios with RPS-1 initially distributed in vertical stripes. Toxin strength parameters derived from liquid kill curve experimental results. Probability of death of Strain R, B, G was 0.1, 0.28, and 0.417 respectively.

Supplementary Movie 8

Lattice model simulation demonstrating one of the possible scenarios with RPS-1 initially distributed in separate blocks. Toxin strength parameters derived from liquid kill curve experimental results. Probability of death of Strain R, B, G was 0.1, 0.28, and 0.417 respectively.

Supplementary Movie 9

Lattice model simulation demonstrating one of the possible scenarios with RPS-1 initially distributed in concentric circles with the strongest strain in the middle. In this example, the central circle has a smaller radius compared to the thickness of the other rings. Toxin strength parameters derived from liquid kill curve experimental results. Probability of death of Strain R, B, G was 0.1, 0.28, and 0.417 respectively.

Supplementary Movie 10

Lattice model simulation demonstrating one of the possible scenarios with RPS-1 initially distributed in concentric circles with the strongest strain in the middle. Toxin strength parameters

derived from liquid kill curve experimental results. Probability of death of Strain R, B, G was 0.1, 0.28, and 0.417 respectively.

2.6 Supplementary Data
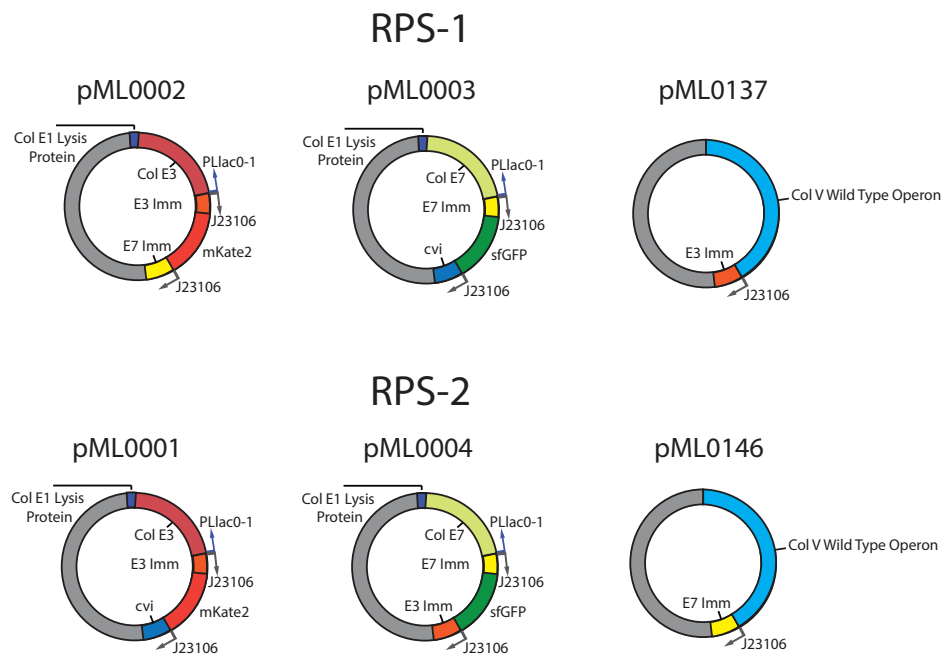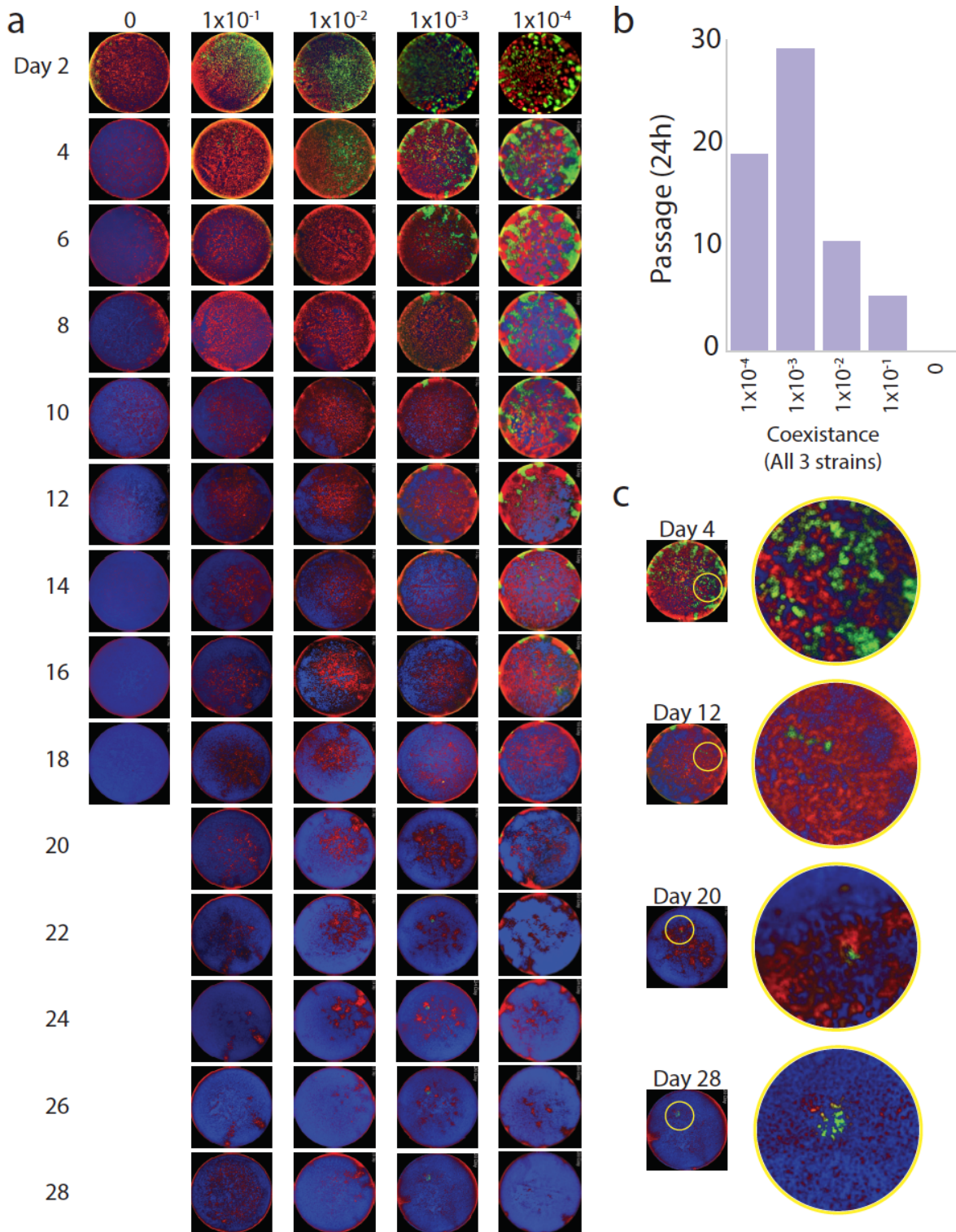

Figure 2.5. The plasmids used in this study

Figure 2.6. Replica plates of all serial dilution densities tested (**a,**) Images of combined three strain plate passages across a range initial starting densities. (**b,**) Duration of three strain coexistence for each starting density. (**c,**) Closeup images of the highlighted regions on day 4, day 12, day 20, and day 28 of the 1 to 1k dilution demonstrating the maintenance of coexistence.
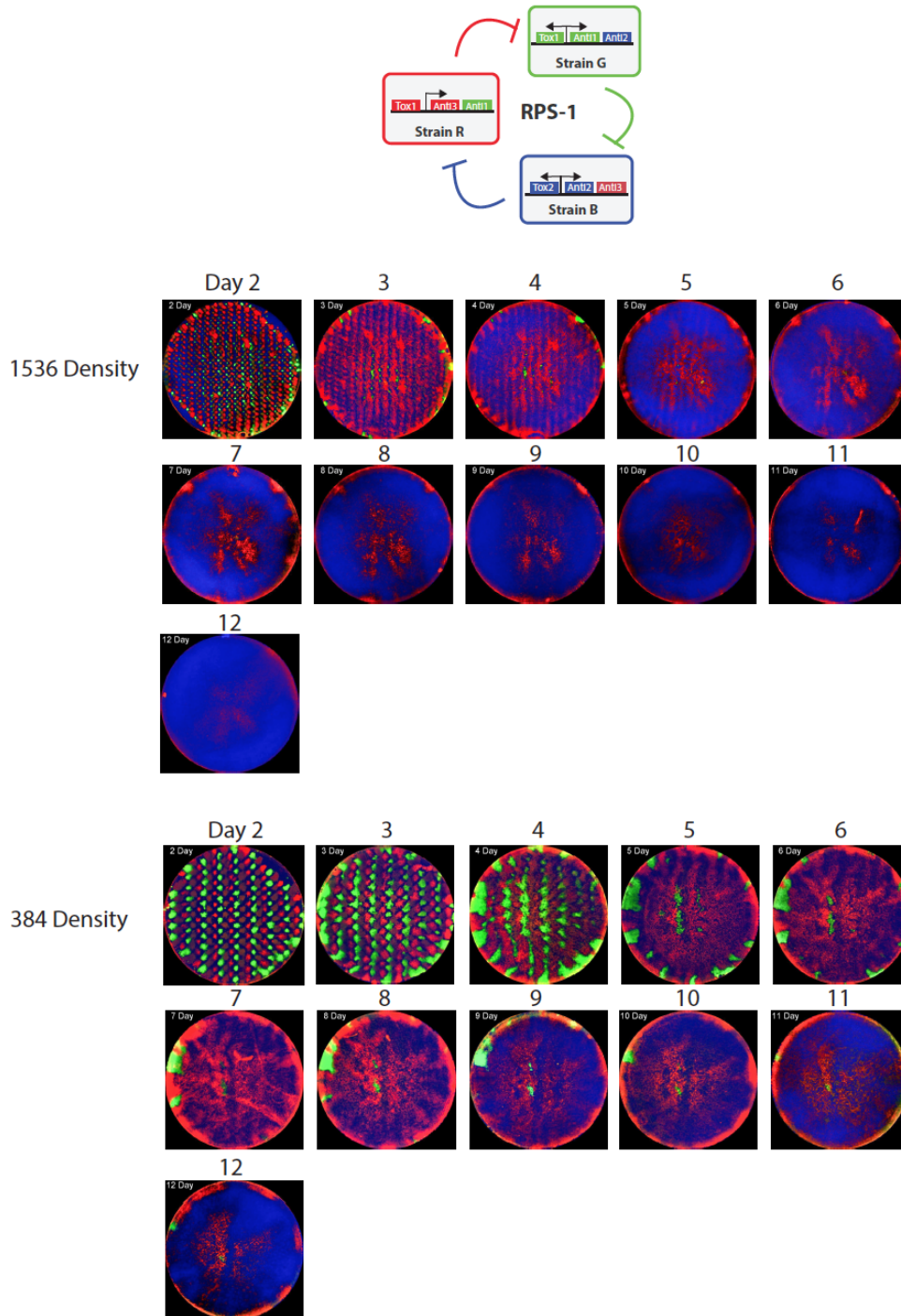
Figure 2.7. Image stills corresponding to the RPS spotted grid passage experiments for 1536 and 384 density starting conditions.
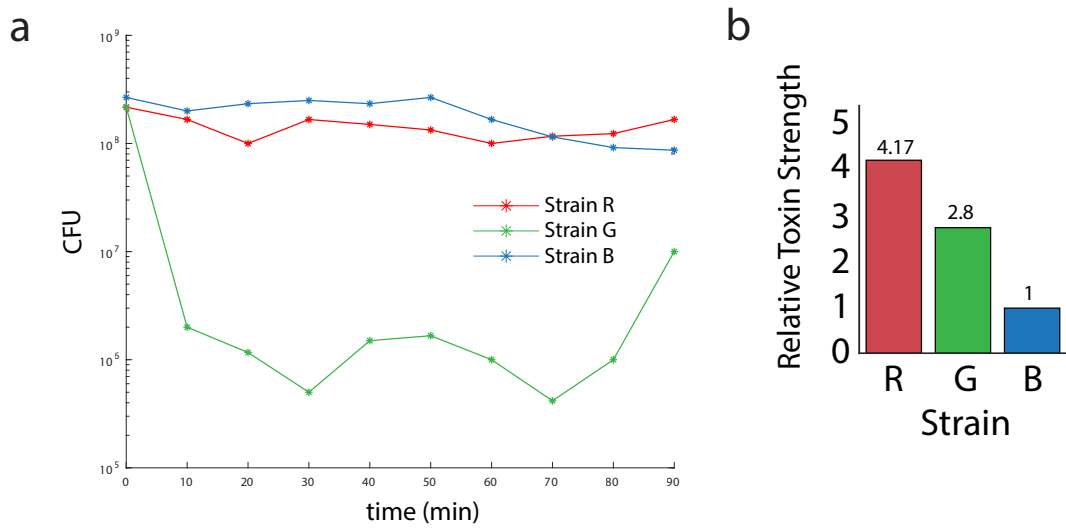
Figure 2.8. (**a,**) Time course results for the liquid kill curve. (**b,**) Toxin strengths of RPS-1 calculated from the kill curves.
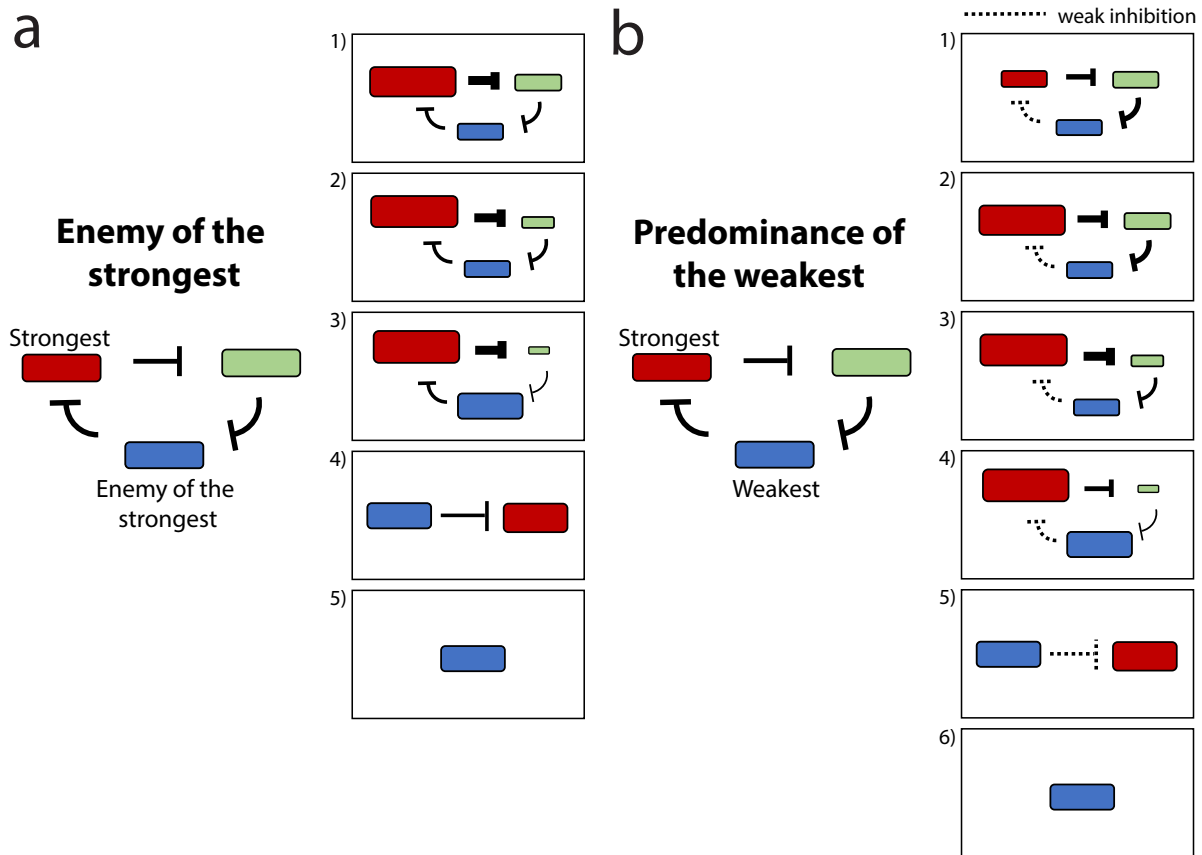
Figure 2.9. Enemy of the strongest or Predominance of the weakest? (**a,**) Schematic representation of "enemy of the strongest". 1) Strain R is the strongest strain, causing strong inhibition of Strain G. 2) As a result of the strong inhibition, Strain G is eliminated faster relative to the other strains. 3) As Strain G is fully eliminated, Strain B can begin to expand. 4) Strain B slowly outcompetes Strain R. 5) Strain B is the final winner. (**b,**) Schematic representation of "predominance of the weakest". 1) Strain B has very weak inhibition of strain R. 2) As a result, Strain R is able to expand faster relative to the other strains. 3) In response to rapid expansion of Strain R, inhibition of Strain G increases. 4) Eventually, Strain G is fully eliminated, enabling Strain B to grow unchecked. 4) Strain B slowly outcompetes Strain R. 5) Strain B is the final winner.
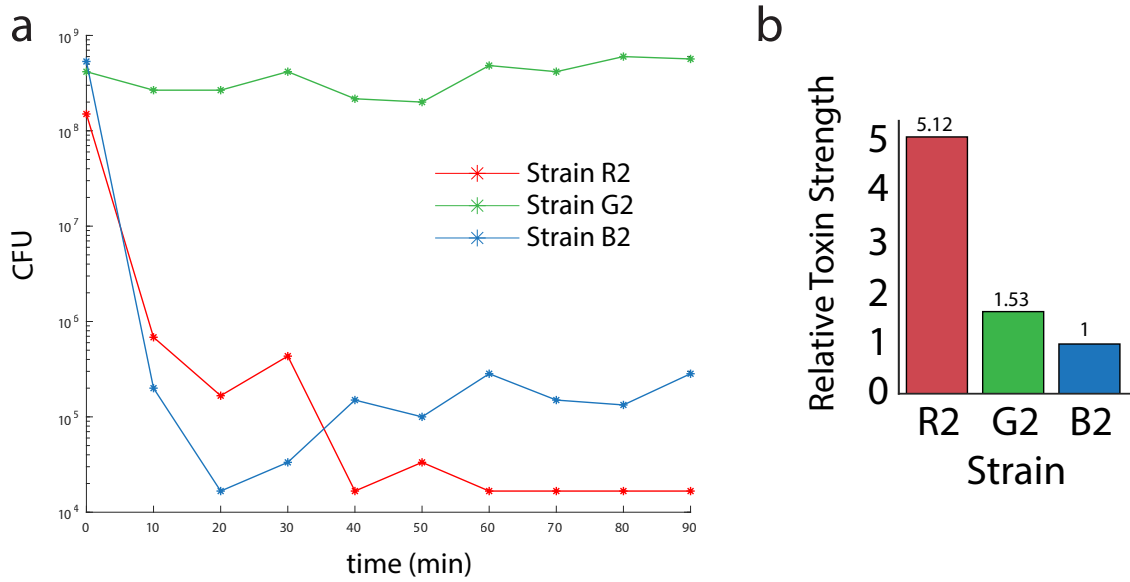
Figure 2.10. Strain hierarchies for RPS-2. (**a,**) Time course results for the liquid kill curve. (**b,**) Toxin strengths of RPS-2 calculated from the kill curves.
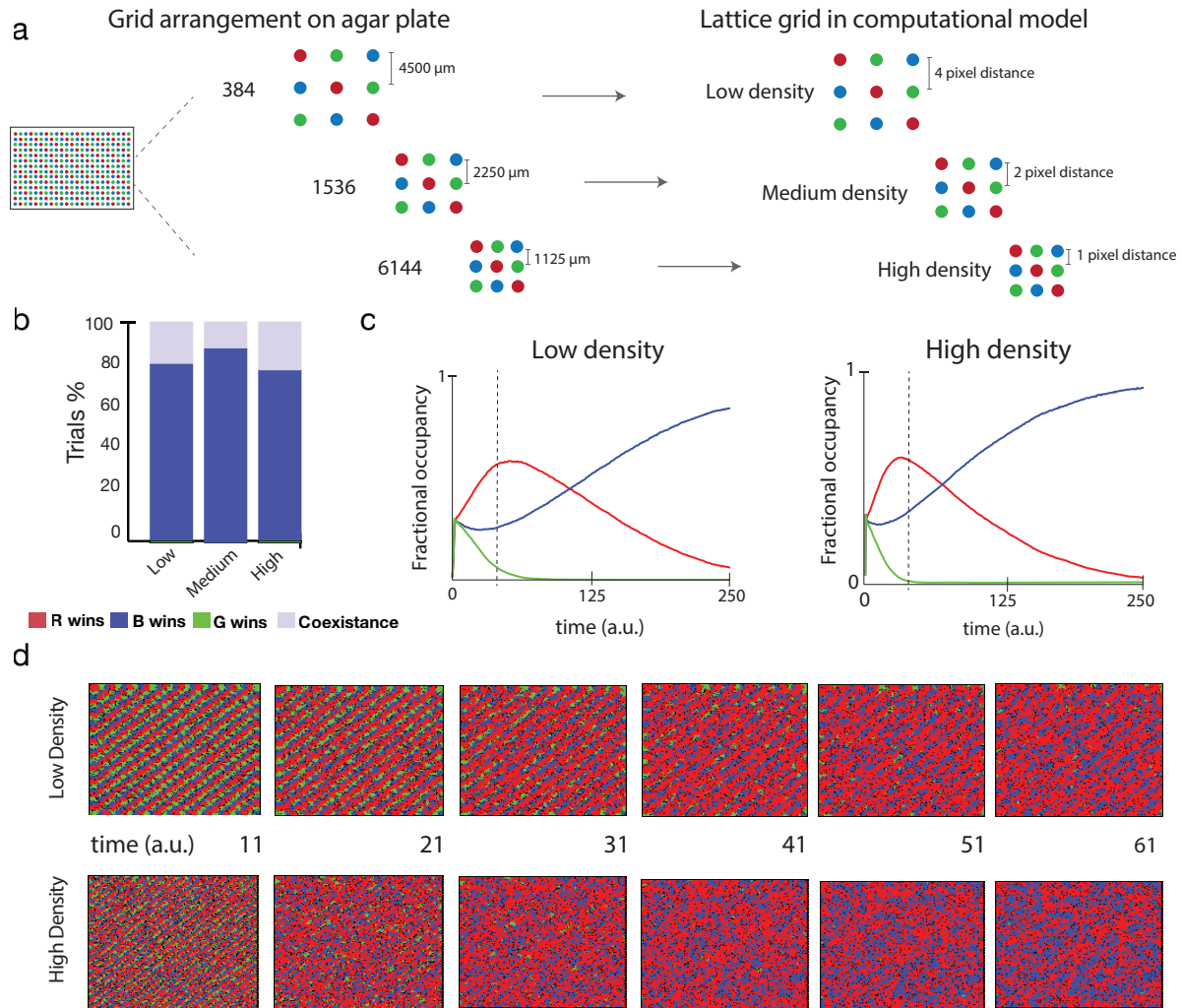
Figure 2.11. Computational model simulates how initial density affects coexistence. (**a,**) Comparison between the spacing between strains on agar plate (on the left) and in the lattice simulation (on the right). (**b,**) The model was simulated with RPS1 parameters (pR= 0.1, pB= 0.23, pG = 0.42) for different initial densities in grid format. Each condition was simulated 100 times and the steady state value for all runs was summarized in this bar graph. (**c,**) Simulated time series for two simulations starting with different grid densities. (**d,**) Representative frames of the lattice grid at different time points for two different initial densities.
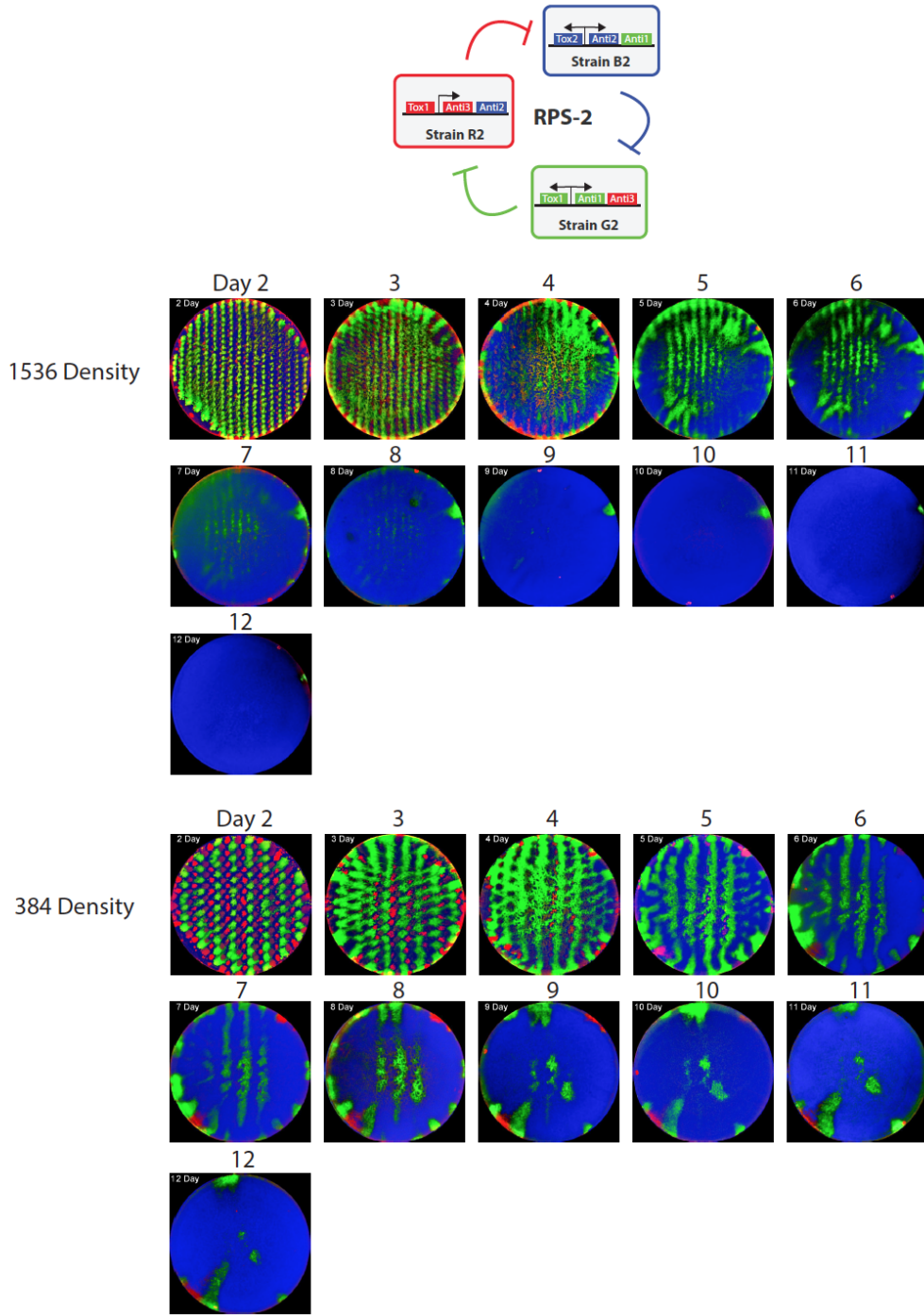
Figure 2.12. Image stills corresponding to the RPS-2 spotted grid passage experiments for 1536 and 384 density starting conditions.
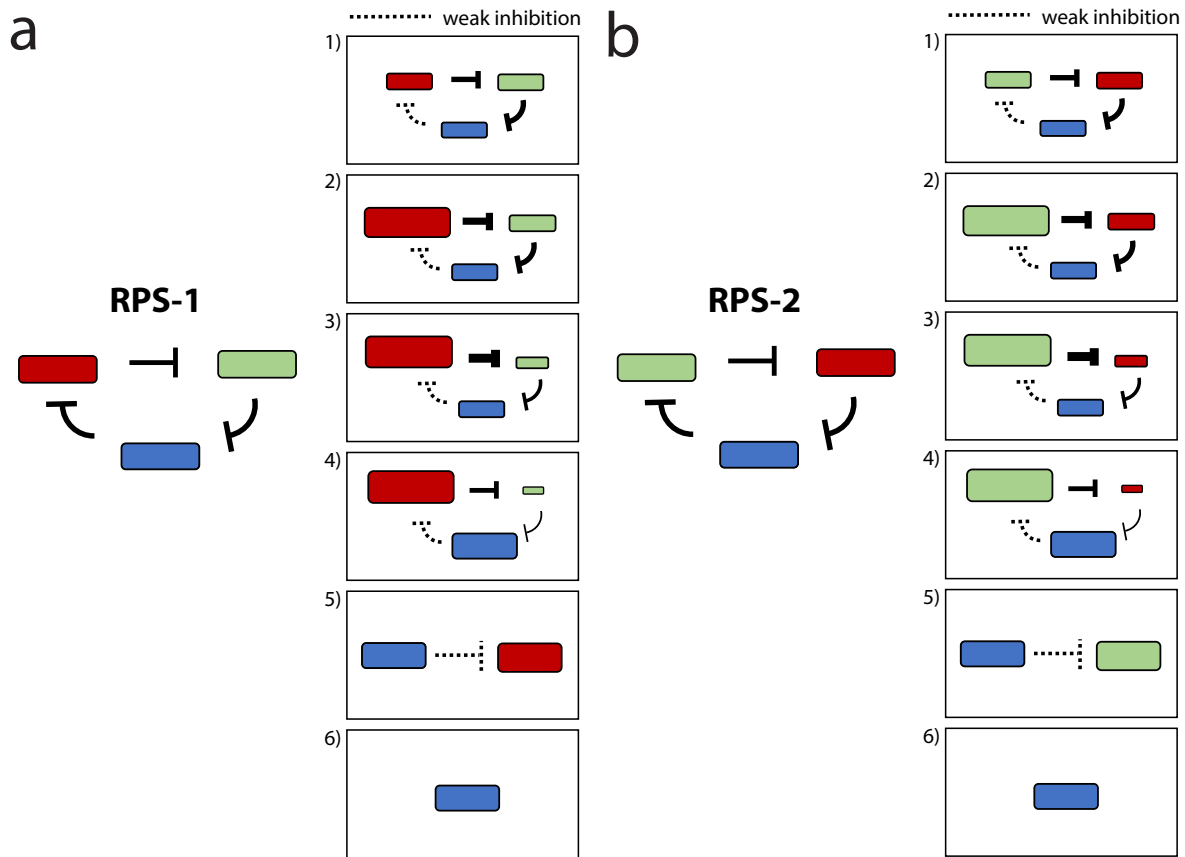
Figure 2.13. Predominance of the weakest for RPS-1 and RPS-2 (**a,**) Schematic representation of "predominance of the weakest" for RPS-1. 1) Strain B has very weak inhibition of strain R. 2) As a result, Strain R is able to expand faster relative to the other strains. 3) In response to rapid expansion of Strain R, inhibition of Strain G increases. 4) Eventually, Strain G is fully eliminated, enabling Strain B to grow unchecked. 4) Strain B slowly outcompetes Strain R. 5) Strain B is the final winner. (**b,**) Schematic representation of "predominance of the weakest" for RPS-2. 1) Strain B2 has very weak inhibition of Strain G2. 2) As a result, Strain G2 is able to expand faster relative to the other strains. 3) In response to rapid expansion of Strain G2, inhibition of Strain R2 increases. 4) Eventually, Strain R2 is fully eliminated, enabling Strain B2 to grow unchecked. 4) Strain B2 slowly outcompetes Strain R2. 5) Strain B2 is the final winner.
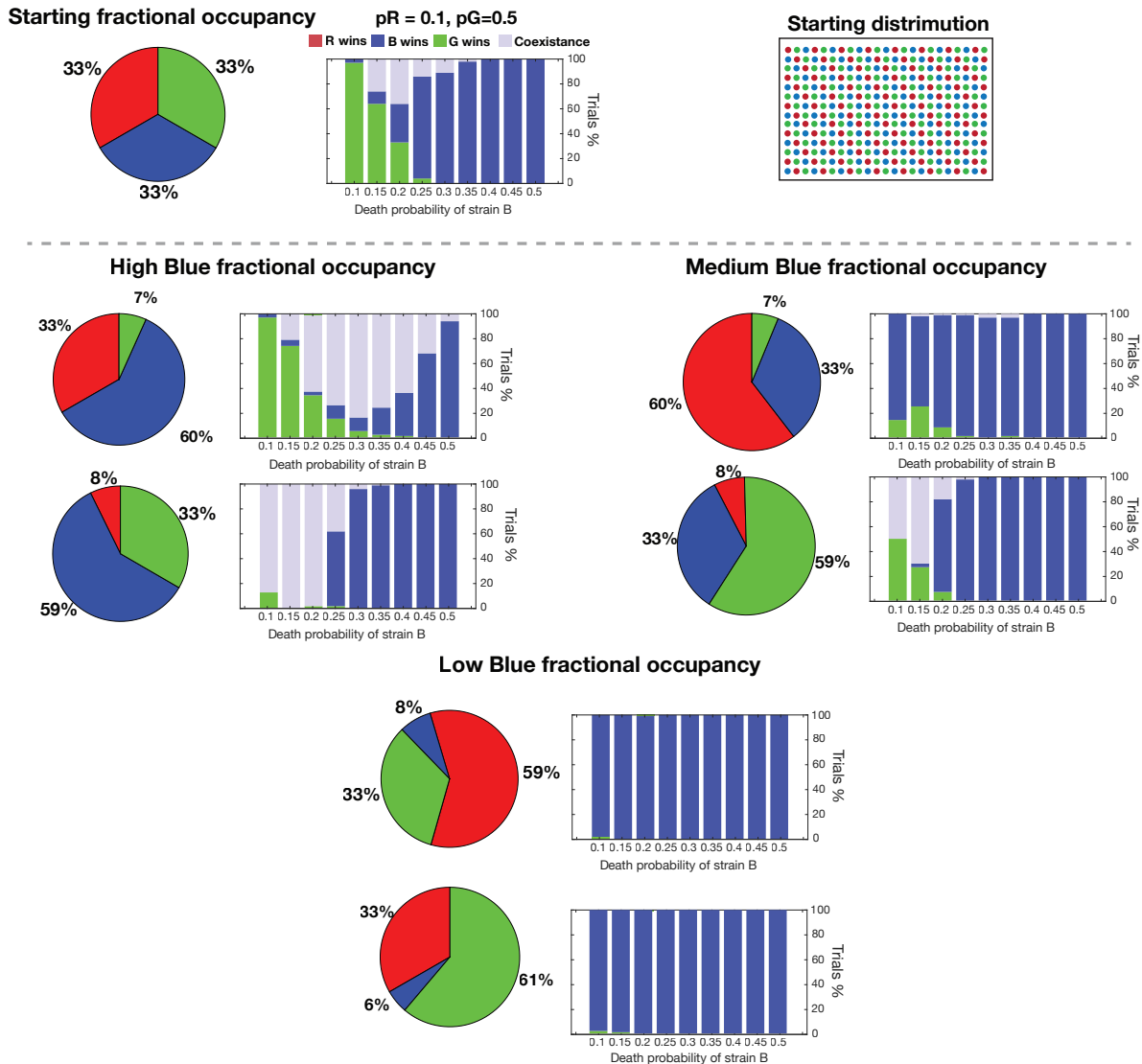
Figure 2.14. Model simulations explore multiple initial conditions in grid format. Pie charts represent the fractional occupancies of the starting allocation of the three strains. The bar charts show the steady state outcomes of multiple simulations for a range of different toxin strength parameters. The bar plots shows the outcome of 100 trials for multiple parameters of Pb (Probability of death of strain B). For all simulations, the probability of death of strain R (0.1) and strain G (0.5) are kept constant.

Strain Name	Host Bacterium	Plasmid	Referenced in Figure
Strain R	**MG1655**	**pML0002** - Colicin E3 + Col E3 Immunity + mKate2 + Col E7 Immunity + Col E1 Lysis Protein	*1a, 1b, 2a-f, 3a-c, 4a-e, Extended Data Figs 1-4*
Strain G	**MG1655**	**pML0003** - Colicin E7 + Col E7 Immunity + sfGFP + Col V Immunity + Col E1 Lysis Protein	*1a, 1b, 2a-f, 3a-c, 4a-e, Extended Data Figs 1-4*
Strain B	**MG1655**	**pML0137** - Colicin V + Colicin V Immunity + Colicin V secretion protein (*CvaA*) + Colicin V secretion/processing ATP-bindingn protein (*CvaB*) + Col E3 Immunity	*1a, 1b, 2a-f, 3a-c, 4a-e, Extended Data Figs 1-4*
Strain R2	**MG1655**	**pML0004** - Colicin E3 + Col E3 Immunity + mKate2 + Col V Immunity + Col E1 Lysis Protein	*3a-c, Extended Data Figs 6,8*
Strain G2	**MG1655**	**pML0001** - Colicin E7 + Col E7 Immunity + sfGFP + Col E3 Immunity + Col E1 Lysis Protein	*3a-c, Extended Data Figs 6,8*
Strain B2	**MG1655**	**pML0146** - Colicin V + Colicin V Immunity + Colicin V secretion protein (*CvaA*) + Colicin V secretion/processing ATP-bindingn protein (*CvaB*) + Col E7 Immunity	*3a-c, Extended Data Figs 6,8*

Figure 2.15. The strains used in this study.

Protein/Gene	AA Sequence
E1 Lysis	MRKRFFVGIFAINLLVGCQANYIPDVQGGTIAPSSSSKLTGIAVQ*
Colicin E3	MSGGDGRGHNTGAHSTSGNINGGPTGLGVGGGASDGSGWSSENNPWGGGSGSGIHWGGGSGHGNGGGNGNSGGGSGTGGNL SAVAAPVAFGFPALSTPGAGGLAVSISAGALSAAIADIMAALKGPFKFGLWGVALYGVLPSQIAKDDPNMMSKIVTSLPADDITESPVSS LPLDKATVNVNVRVVDDVKDERQNISVVSGVPMSVPVVDAKPTERPGVFTASIPGAPVLNISVNNSTPAVQTLSPGVTNNTDKDVRPA GFTQGGNTRDAVIRFPKDSGHNAVYVSVSDVLSPDQVKQRQDEENRRQQEWDATHPVEAAERNYERARAELNQANEDVARNQER QAKAVQVYNSRKSELDAANKTLADAIAEIKQFNRFAHDPMAGGHRMWQMAGLKAQRAQTDVNNKQAAFDAAAKEKSDADAAL SSAMESRKKKEDKKRSAENNLNDEKNKPRKGFKDYGHDYHPAPKTENIKGLGDLKPGIPKTPKQNGGGKRKRWTGDKGRKIYEWDS QHGELEGYRASDGQHLGSFDPKTGNQLKGPDPKRNIKKYL*
Col E3 Immunity	MGLKLDLTWFDKSTEDFKGEEYSKDFGDDGSVMESLGVPFKDNVNNGCFDVIAEWVPLLQPYFNHQIDISDNEYFV SFDYRDGDW*
Colicin E7	MSGGDGRGHNSGAHNTGGNINGGPTGLGGNGGASDGSGWSSENNPWGGGSGSGVHWGGGSGHGNGGGNSN SGGGSNSSVAAPMAFGFPALAAPGAGTLGISVSGEALSAAIADIFAALKGPFKFSAWGIALYGILPSEIAKDDPNMMSK IVTSLPAETVTNVQVSTLPLDQATVSVTKRVTDVVKDTRQHIAVVAGVPMSVPVVNAKPTRTPGVFHASFPGVPSLTVS TVKGLPVSTTLPRGITEDKGRTAVPAGFTFGGGSHEAVIRFPKESGQKPVYVSVTDVLTPAQVKQRQDEEKRLQQEWN DAHPVEVAERNYEQARAELNQANKDVARNQERQAKAVQVYNSRKSELDAANKTLADAKAEIKQFERFAREPMAAGH RMWQMAGLKAQRAQTDVNNKKAAFDAAAKEKSDADVALSSALERRKQKENKEKDAKAKLDKESKRNKPGKATGKG KPVNNKWLNNAGKDLGSPVPDRIANKLRDKEFKSFDDFRKKFWEEVSKDPELSKQFSRNNNDRMKVGKAPKTRTQD VSGKRTSFELHHEKPISQNGGVYDMDNISVVTPKRHIDIHRGK*
Col E7 Immunity	MELKNSISDYTEAEFVQLLKEIEKENVAATDDVLDVLLEHFVKITEHPDGTDLIYYPSDNRDDSPEGIVKEIKEWRAANG KPGFKQG*
CvaC	MRTLTLNELDSVSGGASGRDIAMAIGTLSGQFVAGGIGAAAGGVAGGAIYDYASTHKPNPAMSPSGLGGTIKQKPEGI PSEAWNYAAGRLCNWSPNNLSDVCL*
Cvi	MDRKRTKLELLFAFIINATAIYIALAIYDCVFRGKDFLSMHTFCFSALMSAICYFVGDNYYSISDKIKRRSYENSDSK*
CvaA	MKWQGRAILLPGIPLWLIMLGSIVFITAFLMFIIVGTYSRRVNVSGEVTTWPRAVNIYSGVQGFVVRQFVHEGQLIKKGD PVYLIDISKSTRNGIVTDNHRRDIENQLVRVDNIISRLEESKKITLDTLEKQRLQYTDAFRRSSDIIQRAEEGIKIMKNNMEN YRYYQSKGLINKDQLTNQVALYYQQQNNLLSLSGQNEQNALQITTLESQIQTQAADFDNRIYQMELQRLELQKELVNT DVEGEIIIRALSDGKVDSLSVTVGQMVNTGDSLLQVIPENIENYYLILWVPNDAVPYISAGDKVNIRYEAFPSEKFGQFSA TVKTISRTPASTQEMLTYKGAPQNTPGASVPWYKVIATPEKQIIRYDEKYLPLENGMKAESTLFLEKRRIYQWMLSPFYD MKHSATGPIND*
CvaB	MTNRNFRQIINLLDLRWQRRVPVIHQTETAECGLACLAMICGHFGKNIDLIYLRRKFNLSARGATLAGINGIAEQLGMA TRALSLELDELRVLKTPCILHWDFSHFVVLVSVKRNRYVLHDPARGIRYISREEMSRYFTGVALEVWPGSEFQSETLQTRI SLRSLINSIYGIKRTLAKIFCLSVVIEAINLLMPVGTQLVMDHAIPAGDRGLLTLISAALMFFILLKAATSTLRAWSSLVMSTL INVQWQSGLFDHLLRLPLAFFERRKLGDIQSRFDSLDTLRATFTTSVIGFIMDSIMVVGVCVMMLLYGGYLTWIVLCFTT IYIFIRLVTYGNYRQISEECLVREARAASYFMETLYGIATVKIQGMVGIRGAHWLNMKIDAINSGIKLTRMDLLFGGINTFV TACDQIVILWLGAGLVIDNQMTIGMFVAFSSFRGQFSERVASLTSFLLQLRIMSLHNERIADIALHEKEEKKPEIEIVADMG PISLETNGLSYRYDSQSAPIFSALSLSVAPGESVAITGASGAGKTTLMKVLCGLFEPDSGRVLINGIDIRQIGINNYHRMIA CVMQDDRLFSGSIRENICGFAEEMDEEWMVECARASHIHDVIMNMPMGYETLIGELGEGLSGGQKQRIFIARALYRKP GILFMDEATSALDSESEHFVNVAIKNMNITRVIIAHRETTLRTVDRVISI*

Figure 2.16. The amino acid sequences for the proteins and genes used in this study.

Chapter 3

Machine learning analysis of temporal genome-wide bacterial response to heavy metals

3.1 Introduction

Over the past decades several technologies have been developed to decode the transcriptional regulatory network that guides cells' response to environmental perturbations[55, 87]. These technologies generate large sets of data which usually require sophisticated data analysis techniques to be interpreted[126]. Independent component analysis (ICA) is an unsupervised machine learning technique which was originally developed to deconvolute mixed signals into their individual sources with their relative strengths[23]. ICA has been shown to be particularly successful at extracting biologically relevant transcriptional modules from a wide variety of transcriptomics datasets, particularly obtained via RNA-seq[101, 104]. The main limitations of these datasets is that they can usually only offer a snapshot of the cell's state by providing binary data of the pre and post induction period. In order to address this limitation, we previously developed a microfluidic platform capable of culturing and monitoring over 2000 bacterial strains simultaneously[41]. This technology allows to record the cells response to any environmental perturbation with a time resolution of 10 minutes and for a period of time which can extend up to 14 days.
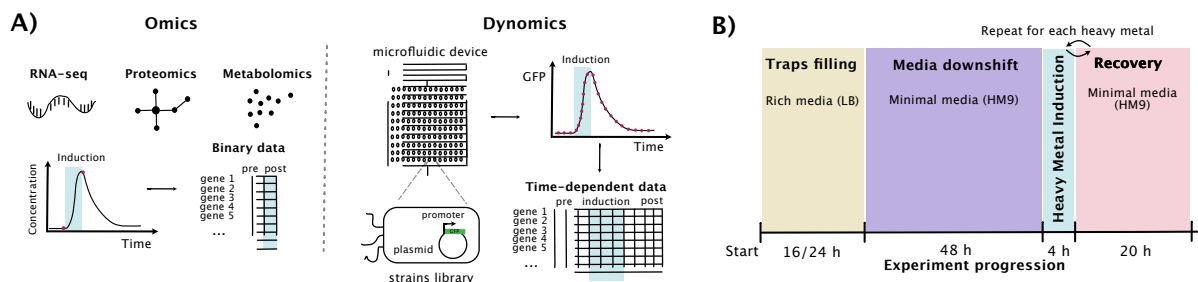
Figure 3.1. Dynomics experimental set up **a** Illustration of the differences between omics technologies (left) and the Dynomics experimental set up consisting of a 2,176-strain microfluidic device. **b** Diagram to illustrate the experiment steps. In chronological order the steps are: growth on rich media to let the cells grow and completely fill the traps, downshift to minimal media for 48 hours, induction with the heavy metal of choice for 4 hours, recovery with minimal media. After 20 hours on minimal media another cycle of induction and recovery can be performed.

In this study, we show how applying ICA to our temporal genome-wide data can lead to unique insights into the temporal patterns of activation of bacterial promoters in response to external perturbations. In order to demonstrate the power of this platform combined with the ICA analysis, we apply this approach to study bacterial response to heavy metals exposure. Heavy metals contamination due to industrial wastewater discharge has become one of the biggest threats to the environment [85]. Many microorganisms have adapted to coexist and resist heavy metals in highly contaminated areas. In addition, some microorganisms can reduce heavy metal ions, lowering their harmfulness[10]. Therefore, bacteria have naturally become great organisms for engineering heavy metal detection systems[41] as well as bioremediation strategies[58]. This is why we focused on expanding our current knowledge of bacterial response mechanism to the exposure of multiple types of heavy metals by investigating a dataset generated by recording the variation over time of the gene expression levels of thousands of different promoters in *E. coli*. Overall, our results confirm patterns of promoter activation previously known from the literature while also identifying new surprising correlations which show the potential of this approach as a hypothesis-generating tool.

3.2 Results

3.2.1 Transcriptional response to heavy metals exposure

Standard omics technologies such as RNA-seq, proteomics and metabolomics can provide accurate quantitative measurements of the cell's behaviour but they usually provide limited or absent temporal information. In fact, most omics datasets are characterized by binary information of the cell status before and after exposure to the inducer of interest. On the other hand, several studies suggest that dynamics is key in how the cells respond to external stimuli and encode information[44, 93], suggesting that there is a need for technologies that are able to capture the real-time response over an extended period of time. In order to address this need, we previously developed a high-throughput platform called Dynomics which can collect genome-wide information of the cells real-time response to a variety of external stimuli. The Dynomics microfluidic device is designed to simultaneously culture 2,176 individual strains in 4 μm tall traps which are continously supplied with fresh media.

For the experiments presented in this study, the strains used belonged to a previously developed *E. coli* library where each strain was engineered with a plasmid containing one of 1,807 native promoters followed by a GFP reporter (Fig. 3.1a). Experiments lasted an average of 7 to 14 days during which the cells were exposed to a different heavy metal every 24 hours (Fig. reffig1MLb). The duration of the induction was 4 hours and it was followed by a period of recovery in minimal media that lasted 20 hours. The average fluorescence expression from each trap was recorded with a time resolution of 10 minutes. The raw fluorescence data collected during the experiment was further processed with background subtraction, smoothing algorithm and normalised by subtraction of the control (promoterless) strains. Furthermore, the data was processed by calculating the log2 of the fold change defined as the ratio of the fluorescence values during the induction window divided by the fluorescence value at the beginning of the induction. For each induction window, we selected six time points spaced every forty minutes in order to minimise noise while retaining temporal information. The data was then organized into a single matrix where each row represented a different promoter and the columns represented the

log2 of the fold change for the six time points considered for each heavy metal induction. We therefore performed ICA (Independent Component Analysis) on this matrix, obtaining the M matrix (promoter coefficients) and the A matrix (activity coefficients) respectively (Fig. 3.2a). We obtained a total of 14 iModulons whose activity profiles and corresponding gene weights are provided in the Supplementary information (Fig 3.5 to 3.19).

Overall, we selected eight biologically significant independently modulated sets of promoters. These sets are referred to as iModulons. Each iModulon was named according to the shape and the predominant condition which characterised the activity profile associated to it (Fig. 3.1b). In particular, we found that the ICA analysis was able to differentiate three main activation patterns associated to the zinc inductions. These patterns were classified as steady activation for promoters whose activation was characterised by a steady increase over the induction window, fast activation for promoters that were activated quickly after the beginning of induction and intermediate activation for the promoters that were the most active in the middle of the activation window. The remaining heavy metals were characterised by the steady state type of activation profile. Each iModulon contains a weighting for each of the promoters in the dataset. Most promoters in each iModulon have a value that is close to zero, therefore in order to extract only the most significant promoters a thresholding method is applied (See Methods). After applying the calculated cut offs, we were able to extract and analyse the most significant promoters for each iModulon (Fig. 3.3). In order to better understand the overall bacterial response to each heavy metal, the promoters were classified according to the phenotypic response they are associated to in the literature as well as their activation profile.

3.2.2 Zinc iModulons

We found out that the significant promoters associated to the zinc iModulons were be associated to four main phenotypic responses: envelope stress response, oxidative stress response, zinc resistance and the dissimilatory nitrate reduction to ammonium pathway (Fig. 3.3a). Our
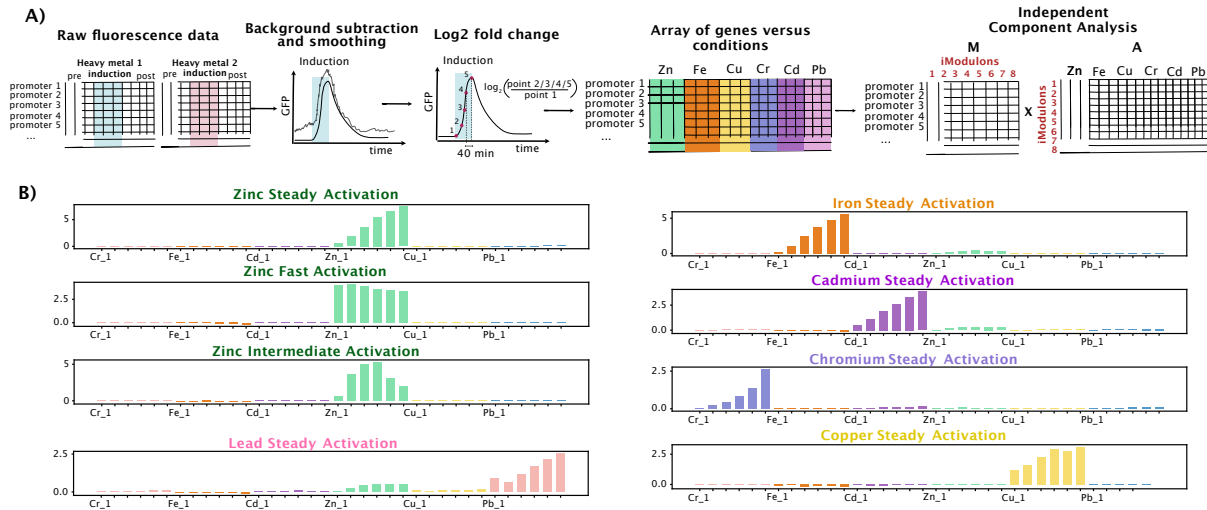
Figure 3.2. Independent Component Analysis (ICA) to analyse heavy metal inductions data from Dynomics experiments. **a** Illustration of the different steps of our data analysis pipeline. The analysis starts with raw fluorescence data which are processed with background signal removal algorithm, normalization by the promoterless strains and smoothing through median filtering. The data is therefore converted to the log2 of the fold change and formatted as a matrix of genes versus conditions (heavy metal inductions). ICA is eventually applied to this matrix obtaining the M (ipromoter coefficients) and A (activity coefficients) matrixes respectively. **b** iModulons activity plots representing different response dynamics to the heavy metals investigated.

results confirm previous findings from the literature while uncovering new associations that were previously unknown. In particular, the ICA analysis applied to our unique Dynomics dataset is able to provide information on the timing of promoter activation, giving us insights on how bacteria dynamically respond to heavy metal exposure over time.

In particular, we find that a set of promoters directly related to the envelope stress response (yfeY, bacA, ropE) and to peptoglycan stress are present in the fast and steady activation profiles as predicted from the literature[86, 125, 82]. Interestingly, we find that the genes that belong to the steady activation profile are associated to peptoglycan synthesis (mipA, cysQ). In addition, we report the activation of glnW which transcribes for glutamine tRNA. Glutamine is a key amino acid in the synthesis of fatty acids, therefore we hypothesise that the expression of this gene is involved in the cells attempt to restore the membrane damage caused by zinc exposure.

Furthermore, the ICA analysis detected several genes involved in the oxidative stress

response, as previously reported in the literature[3], and it was able to distinguish between fast activation (yedY, yfcG, selC) and steady activation (katE, lipA, aldH, nupG). Our results also confirm the activation of the zntA gene which is associated to zinc export as a detoxification mechanism when excess levels of zinc are detected[7]. Interestingly, our analysis is also able to detect the activation of two genes (narZ and nrfE) belonging to the dissimilatory nitrate reduction to ammonium pathway (DNRA)[7]. In particular, we were able to observe the temporal dynamics of the pathway activation since narZ (nitrate reductase) was associated to the fast activation iModulon and nrfE (nitrite reductase) was associated to the intermediate activation profile. To our knowledge, this is the first time that the activation of DNRA pathway in *E. Coli* is linked to the presence of excess zinc concentrations. We hypothesize that in this context DNRA activation can be beneficial to the cells due to the generation of an electron sink that can be used for NADH re-oxidation into NAD+ which is a key metabolites in counteracting DNA damage and oxidative stress[117, 124].

3.2.3 Cadmium iModulons

The cadmium iModulon was characterized by a steady state activation profile and included promoters which are mainly associated to cadmium resistance and oxidative stress. As expected from the literature[7], we find the expression of the zntA which confers cadmium tolerance as well as zinc tolerance as previously described. Furthermore, we detect the activation of the sodB promoter encoding for a superoxide dismutase which is a key in defending the cell against oxidative stress through decomposition of superoxide radical. This finding confirms previous results reported in the literature which showed that superoxide dismutases (SODs) can protect *E. coli* from heavy metal toxicity, particularly from cadmium exposure[36].

We also find that the iModulon reports the cell's response to two of the main effects of oxidative stress: lipid peroxidation and DNA damage response. In particular, we report the activation of promoter mipA which encodes for a scaffolding protein for murein synthesizing

machinery[120] and promoter exfK (also known as bamA) which is involved in of beta-barrel proteins assembly and insertion in the outer membrane[43]. Additionally, we identify three promoters which are known to be involved in nucleotide catabolism repression (deoR[84]), DNA replication (yejK[65]) and DNA recombination and repair respectively (yhcG[61]). Finally, we detect the activation of the phoA promoter which is responsible for the breakdown of organic phosphate esters[1]. We hypothesize that the activation of phoA could be related to the detoxyfication of heavy metals through polyphosphates[63].

3.2.4   Copper iModulons

Similarly to the cadmium iModulon, the copper iModulon was also characterized by a steady state activation profile. As confirmed in the literature, we report the activation of the cueO promoter which encodes for a multicopper oxidase involved in copper tolerance under aerobic conditions[30]. Furthermore, we detect the activation of promoters known to be linked to the oxidative stress response (nrdH[83], ytfE[56]) and to the envelope stress response (yoaE[94]). Interestingly, we also detect the activation of promoter mgsA which is known to be activated during phosphate starvation and glycolysis restriction[102]. Similarly, we report the activation of promoter prpR which is part of the Pho regulon[127]. Therefore, we hypothesize that the activation of these two promoters is part of the cells' attempt to maximize the availability of phosphate to produce polyphosphates as a detoxyfication strategy as previoulsy reported for the cadmium response.

3.2.5   Iron iModulons

The iron iModulon is characterized by a steady state activation profile. Interestingly, we found that three enriched promoters (ugpB, phnC and phoB) which all belong to the Pho regulon[116]. In particular, phoB is the response regulator in two component regulatory system

with PhoR (or CreC) and regulates Pi uptake. The gene phnC is part of the ABC transporter complex PhnCDE involved in phosphonates, phosphate esters, phosphite and phosphate import. Finally, gene ugpB codes for a binding protein-dependent sn-glycerol-3-phosphate transport system which is under the control of the Pho regulon. In particular, ugp-dependent G3P transport activity is present only after growth at limiting concentrations of Pi[107]. From the literature, we know that excess of iron can cause the formation of toxic reactive oxygen species (ROS) through Fenton chemistry. Both iron stress and oxidative DNA damage are successfully prevented by polyposphates which are polymers formed by covalently linked inorganic phosphates[8]. Therefore, we hypothesize that the genes enriched in the Iron iModulons reflect the need for the cell to import additional phosphate in order to form polyphosphates which can be used as a defense mechanism against iron.

3.2.6 Transcriptional response of the recovery post-induction

One of the main advantages of the Dynomics technology is the possibility to track the cells response over an extended period of time. This enabled us to analyse not only the effect of the heavy metal exposure during the induction window but also to explore how the bacteria responded during the recovery period post-induction. In this case, we calculated the logarithm base 2 of the fold change with respect to the end of the induction window(Fig. **??**a). We took into consideration 20 time points spaced 40 minutes apart. Applying the ICA algorithm to this dataset we found 35 iModulons (Fig. 3.20 to 3.54). We focus our analysis on a subset of iModulons whose activity seemed particularly interesting and biologically significant (Fig 4b, Methods).

We report four iModulons whose activity profiles show activation post zinc exposure. Similarly to the induction response, we identify several promoters associated to stress responses such as the SOS response (recX), peptoglycan stress (yjfQ, helD, skp) and oxidative stress (katE). On the other hand, we report the activation of promoters which are related to transcription

and translation processes such us maturation of rRNAs and tRNAs (rne), purine metabolism (yneF), proline tRNAs (proL), methionine synthesis (metF), DNA ligase (ligB). We believe that the activation of this type of promoters suggest a diversion of the cells resources from defense against environmental stressors to growth and repair processes. Interestingly, we find that one of the zinc-related iModulons transcribes for monothiol glutaredoxin which is involved in the biogenesis of iron-sulfur clusters. This confirms previous studies that have shown how excess zinc is associated with the disruption of the iron-sulfur clusters in E. coli[70].

The ICA analysis also identifies two iModulons whose activity profiles are mostly active post cadmium exposure (Fig **??**b). Similarly to the trend we observed for the zinc iModulons, we find that there are several promoters associated to biosynthesis and protein translation processes. This further confirms our prediction that the cells mainly activate stress response processes during the induction itself and switch to recovery and growth processes when the inducer is removed. In particular, for the post-induction cadmium iModulon we report the activation of promoter asd which is involved in L-lysine, L-methionine and L-threonine biosynthesis, promoter tufB which codes for an elongation factor for protein biosynthesis and ykgM which transcribes for a ribosomal protein. Interestingly, ykgM was previously found to be upregulated upon zinc starvation[88, 40]. This physiological response is further supported by the activation of promoter znuC which transcribes for a zinc import ATP-binding protein. Our results therefore align with previous studies based on the analysis of genome-wide temporal gene expression data which suggested that the molecular mechanisms of cadmium toxicity could be partially explained by the disruption in the transcription of genes encoding ribosomal proteins and zinc-binding proteins[121].

Furthermore, we report two iModulons whose activation profiles are active post iron induction. We find that one of the promoters enriched in these iModulons is ompW which transcribes for an outer membrane protein whose expression has been previously observed to be downregulated in iron limitation conditions[130]. Our results show that the expression of ompW is greatly increased post-induction when the toxin is removed from the media. Furthermore, we
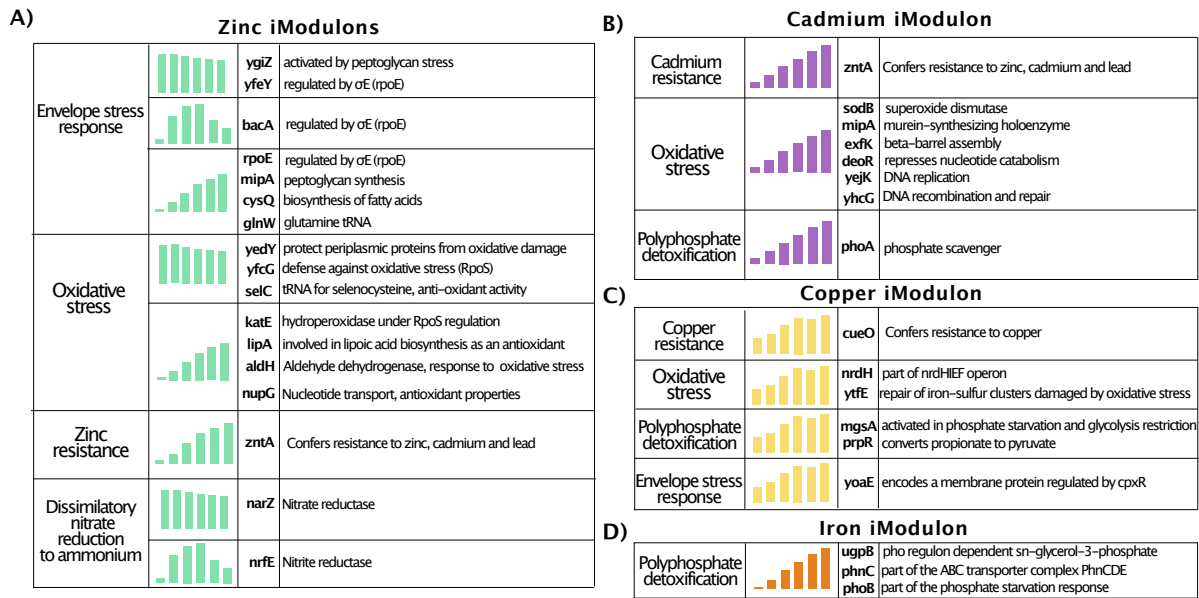
Figure 3.3. Illustration of significant promoters enriched by the ICA analysis. **a** Significant promoters associated to the induction of zinc. Promoters are classified according to the physiological response they are associated to. **b** Significant promoters associated to the induction of cadmium. **c** Significant promoters associated to the induction of copper. **d** Significant promoters associated to the induction of iron.

find that the genes LsrR (associated to quorum sensing regulation) and pppA (part of the type II secretion system). Furthermore, we find a second iron iModulon with a steeper activation profile which is associated with the activation of two genes: rfbB (involved in biogenesis of the bacterial outer membrane) and ydcL (uncharacterized lipoprotein). Finally, we report a copper iModulon linked to the activation of five genes, two of which are involved into amino acid biosynthesis (metF) and DNA replication/transcription processes (helD).

3.3 Discussion

In this study, we combine the power of a genome-scale platform to monitor temporal gene expression[41] with the analytical capabilities of ICA[104]. ICA had been previously shown to be able to extract transcriptional modules with biological significance from a variety of datasets. These datasets generally didn't contain any information on the time-dependent activation of genes under different conditions. Instead, the majority of previous studies were based on binary

data obtained from RNA-seq, proteomics and metabolomics experiments[105, 77, 115, 101].

Here we propose the application of ICA on a dataset obtained from monitoring promoter activation from a library of 1807 unique strains of E. Coli that grow continously in a large-scale microfluidic device. These experiments generate unique datasets where promoter activation is tracked for several days with a resolution of 10 minutes, capturing the genome-wide time-dependent response instead of just the expression level pre and post induction. Therefore, in this study we show how ICA can be applied to our unique dataset to extract information on the timing of promoter activation in response to heavy metals inductions. By first applying our analysis pipeline to the induction window, we observe the richest response for the zinc inductions. In this context, we observe three different iModulons which differentiate fast responders (genes that are activated at the start of the induction window) from intermediate responders (genes that are maximally active in the middle of the induction window) and late responders (genes whose expression steadily increases throughout the induction window).

In particular, we report that our platform is able to resolve the sequence of activation between promoters beloning to the same metabolic pathway. We detect the activation of promoter narZ as an early responder and promoter nrfE as a late responder which are involved in the first and second step of dissimilatory nitrate reduction to ammonium metabolic pathway. This result is a clear example of the power of this platform when used for pathways reconstruction which is a topic of great interest in the scientific community[103]. Therefore, we suggest future studies which are specifically targeted to the investigation of metabolic responses whose pathways are currently unknown or incomplete.

Chapter 3, in part is currently being prepared for submission for publication of the material. Arianna Miano, Anand Sastry, Kevin Rychel, Bernard Palsson Jeff Hasty. The dissertation author was the primary investigator and author of this material.

3.4 Methods

3.4.1 Data collection

Detailed description of the experiments set up can be found in the literature from the original study[40]. In brief, data was obtained from fluorescence values extraction from flat-field corrected images gathered using a custom optical set up. The first step of the experimental set up consisted in arraying the cells using a Singer ROTOR robot so that they could be spotted onto the microfluidic device before glass bonding. Once the device was ready, it was set up inside a custom box kept constant temperature at 37°C for imaging. For media flow, the inlet and outlet were connected to 140mL syringes.

3.4.2 Data processing

The data was first processed by subtracting the local background signal and then dividing the result by the background signal again in order to create a measure of the amplification of the signal over the background. Then, the data were passed through a median filter (scipy.signal.medfilt, kernel_size = 11) and normalised by subtracing and dividing the average expression values of the promoterless strains. The data was further processed by calculating the log_2 of the fold change of the ratio of six time points (spaced 40 minutes apart) with respect to the start of the induction window for the data plotted in Figure 3.2 and 3.2. On the other hand, for the dataset used to produce the results shown in Figure 4 the data was converted in the log_2 of the fold change of the ratio between 20 points (spaced 40 minutes apart) with respect to the end of the induction window. The final dataset which was fed to the ICA algorithm consisted of a matrix where the rows represented all the different promoters (1807) and the columns all the different conditions. Therefore, for the analysis in Figure 1 and 2 the final dataset had a total of 36 columns representing 6 heavy metal inductions, each with 6 time points. Similarly, the dataset behind the results of Figure 4 had dimensions of 1807 rows and 100 columns representing 6 heavy metal inductions, each with 20 points.

3.4.3 Independent component analysis

We used a pipeline for ICA implementation which was previously proposed in other studies[104]. In brief, we run the Scikit-learn70 (v0.19.2) implementation of the FastICA algorithm 100 times with random seeds. The number of components in each iteration was set to the number of components that reconstruct 99% of the variance as calculated by principal component analysis. The resulting source components (M) from all runs were clustered using the Scikit-learn implementation of the DBSCAN algorithm. Therefore, the number of clusters calculated by DBSCAN estimated the dimension of the dataset.

3.4.4 Determination of the gene coefficient threshold

The dataset M contains all the genes coefficients associated to each iModulon. Most of these coefficients have values close to zero which indicates they are not significantly enriched for that iModulon. In order to extract the genes that belong to each iModulon we computed the D'Agostino K^2 test statistic which is a measure of the skew and kurtosis of a sample distribution. As described in previous studies[104], genes with the largest absolute value were iteratively removed and the D'Agostino K2 test statistic was computed for the resulting distribution. The statistic cutoff was kept fixed at 800 for the entire analysis. We only considered positively correlated genes in this analysis.
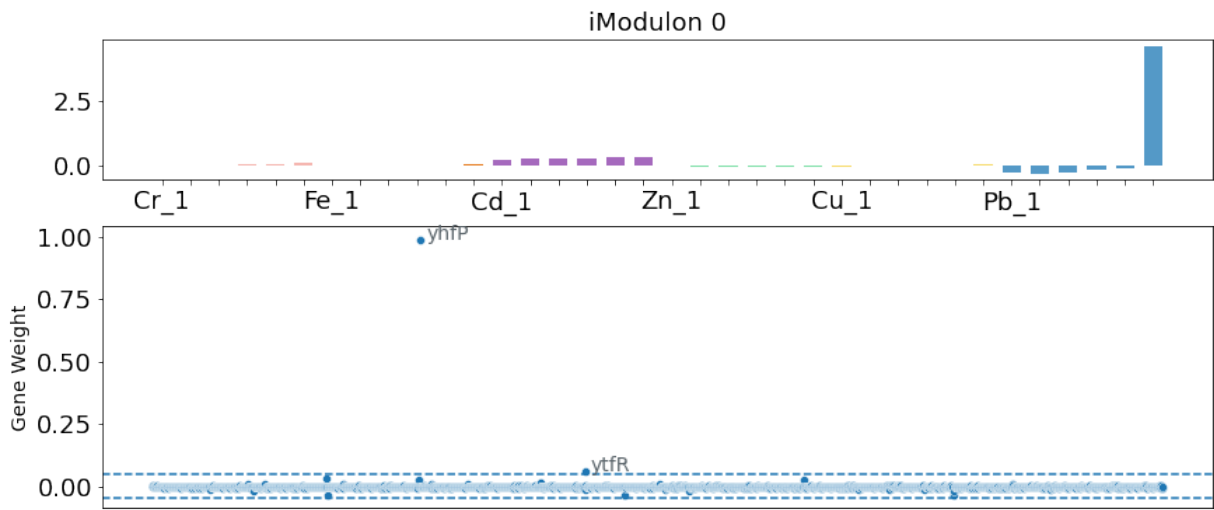
Figure 3.5. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 0. Bottom: Genes weights above threshold for IModulon 0.
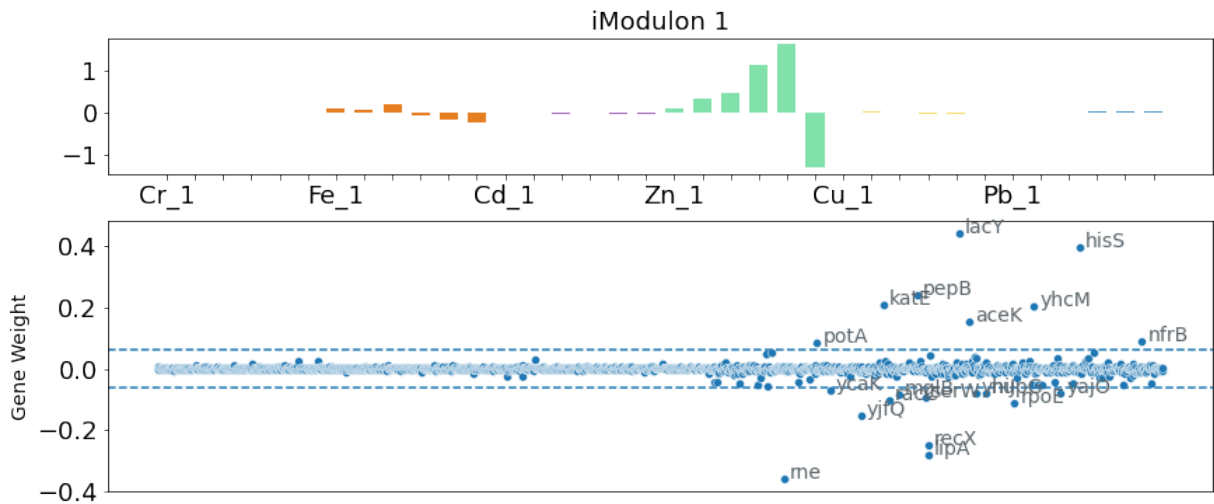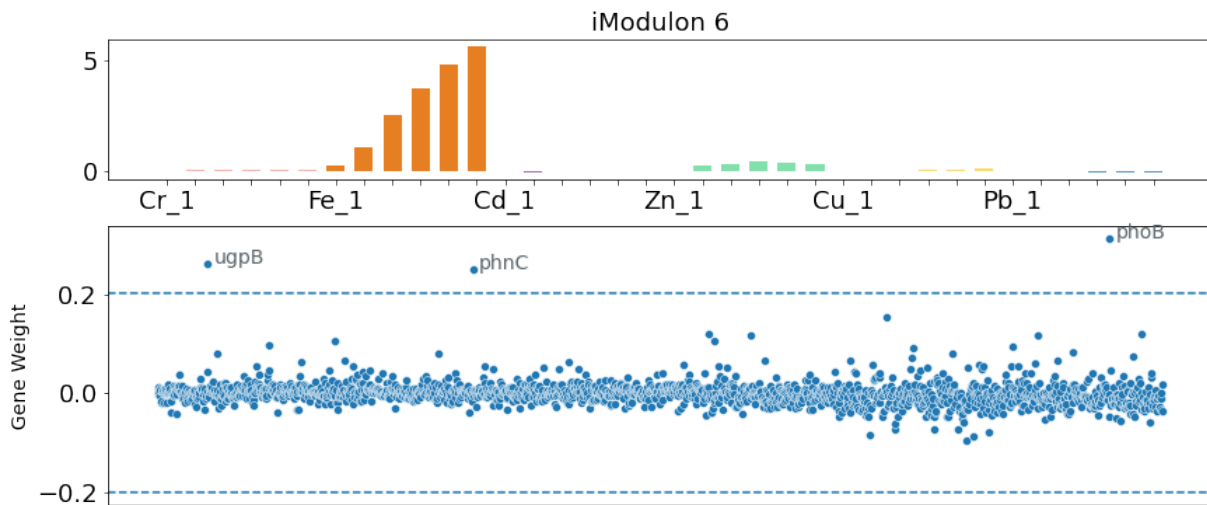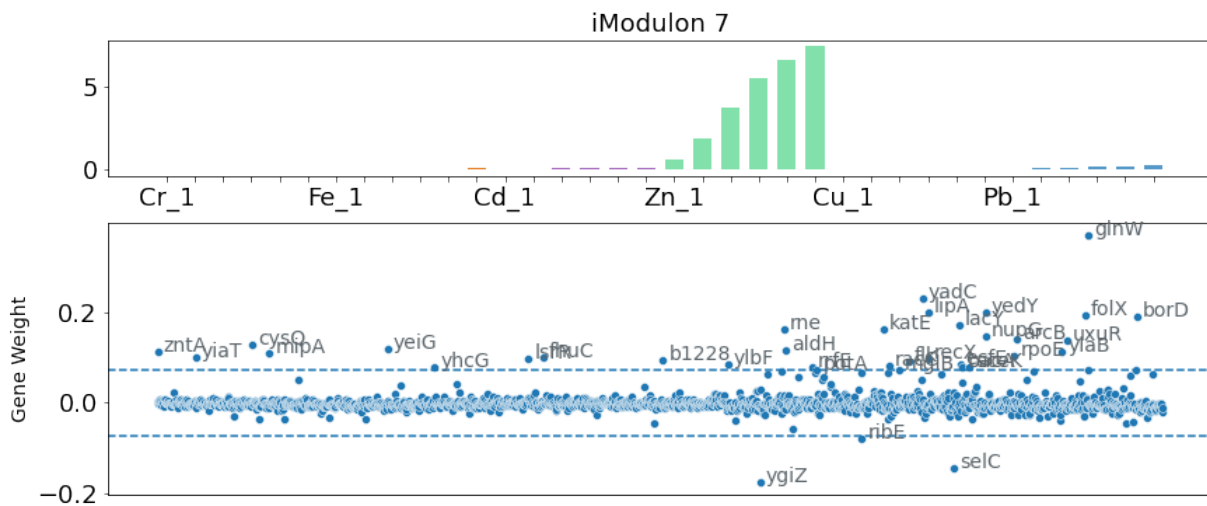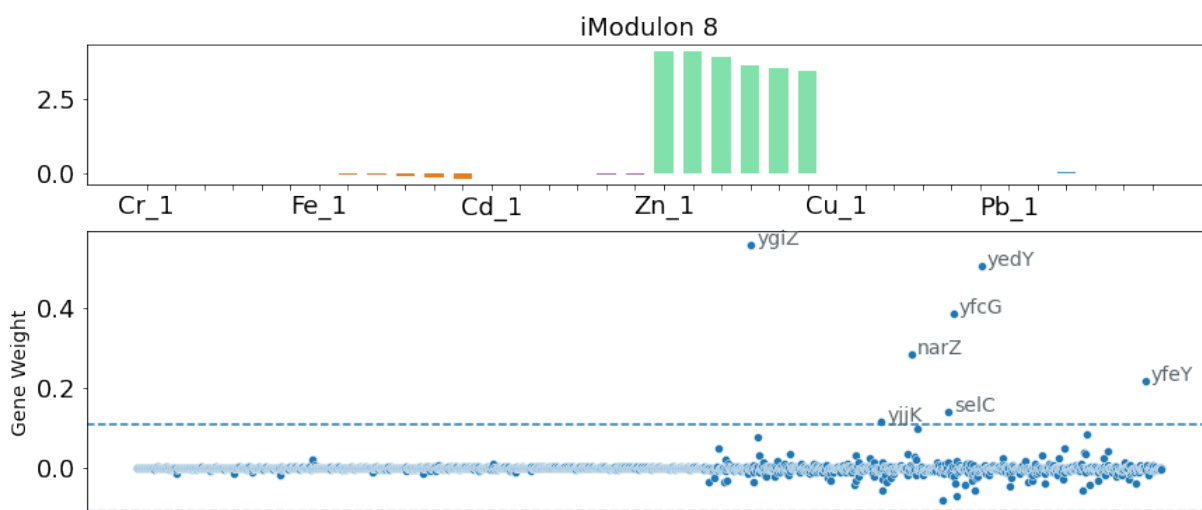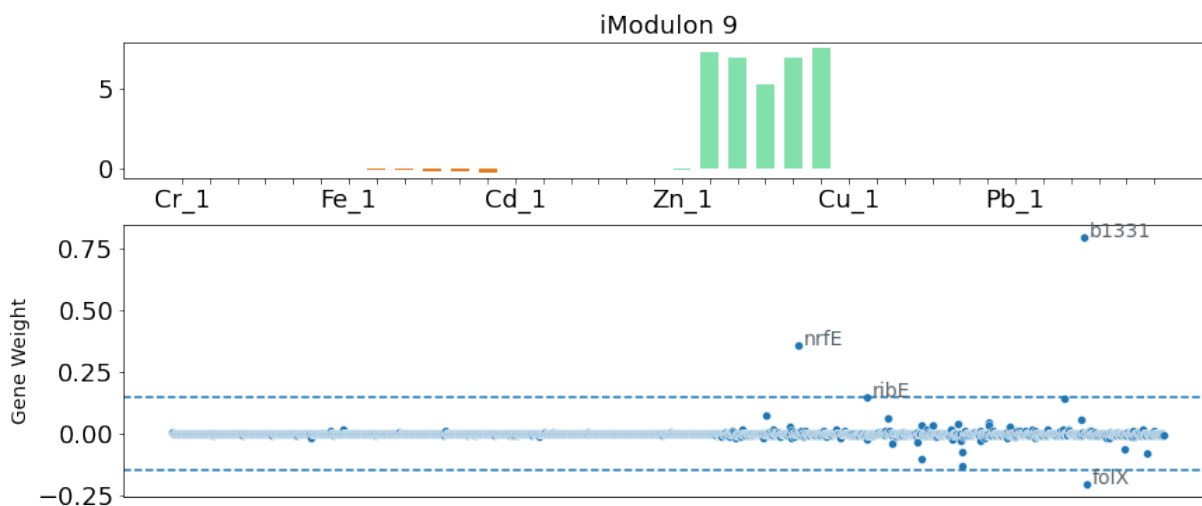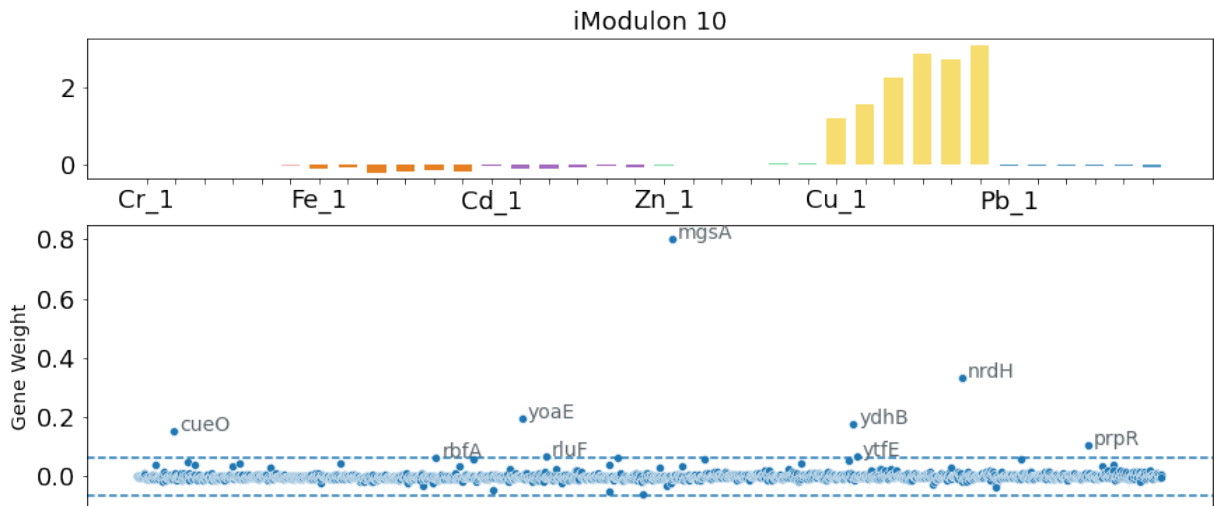

Figure 3.6. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 1. Bottom: Genes weights above threshold for IModulon 1.
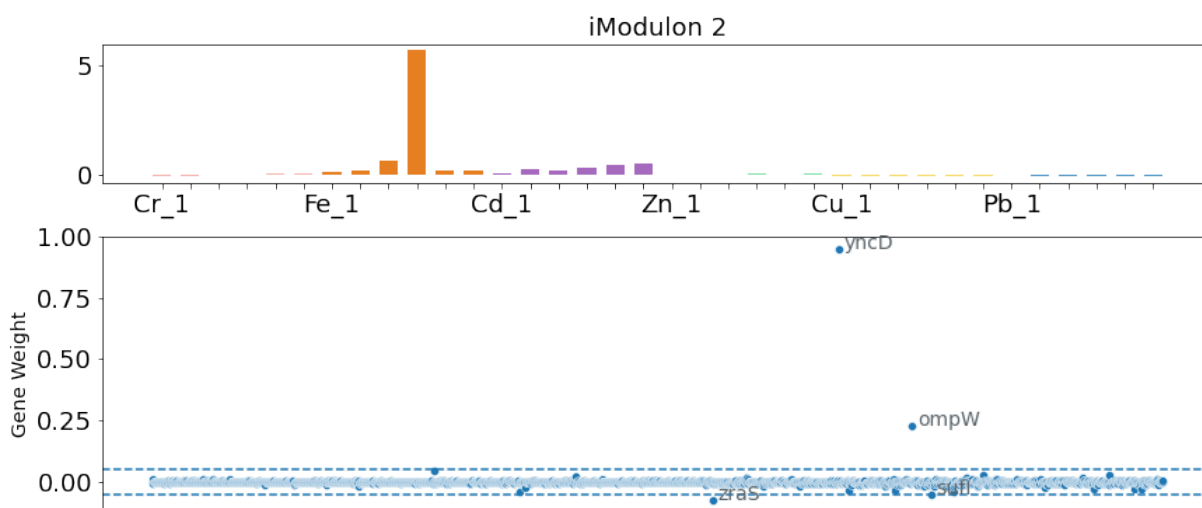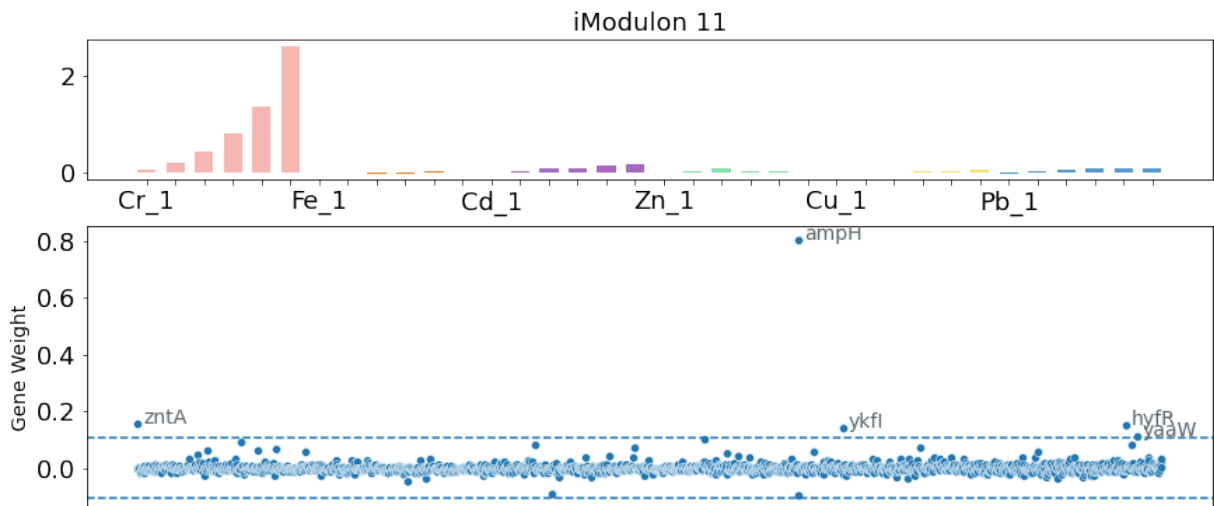
Figure 3.7. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 2. Bottom: Genes weights above threshold for IModulon 2.
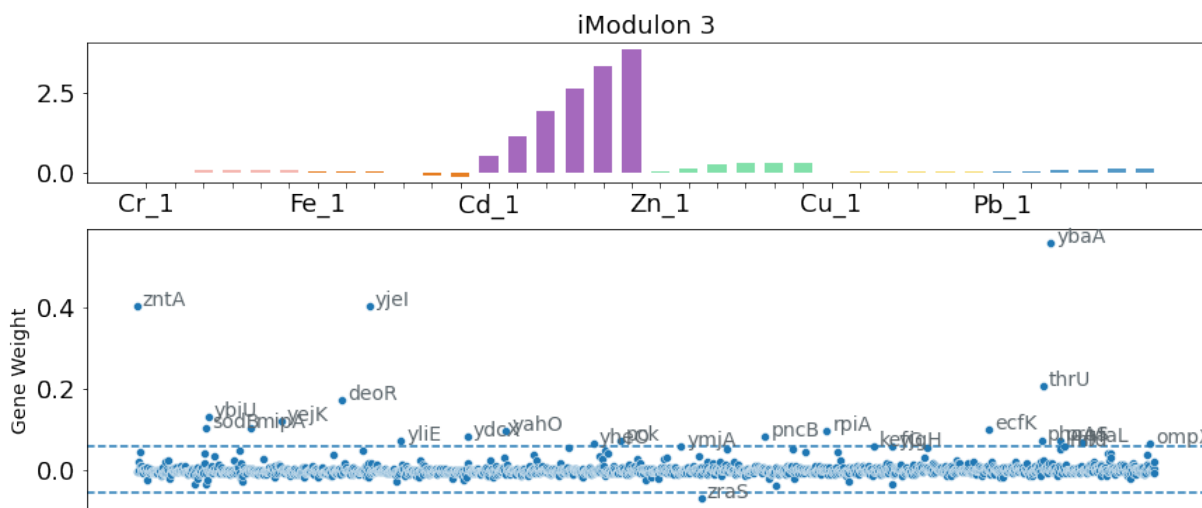
Figure 3.8. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 3. Bottom: Genes weights above threshold for IModulon 3.
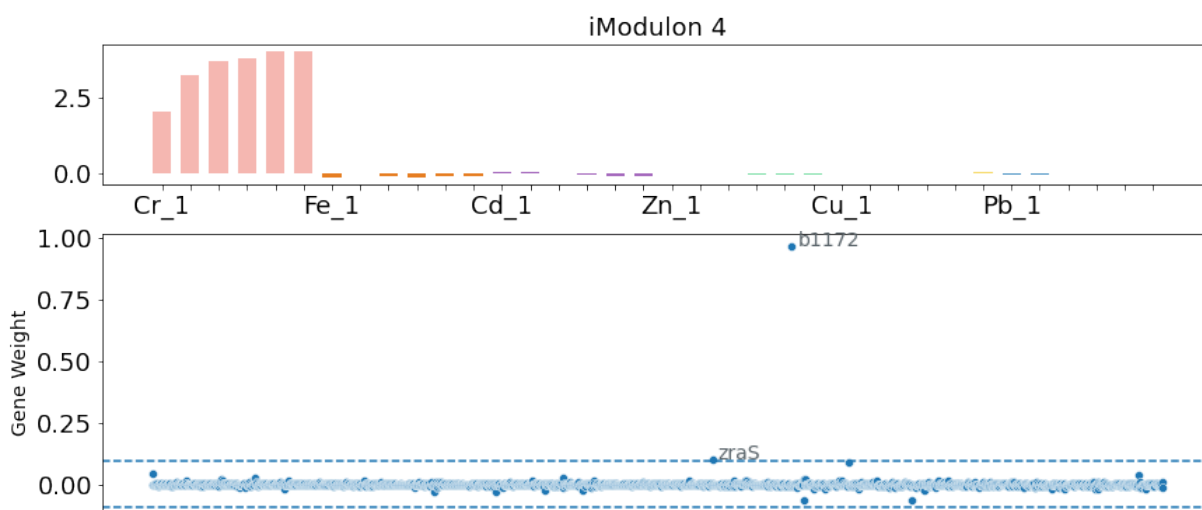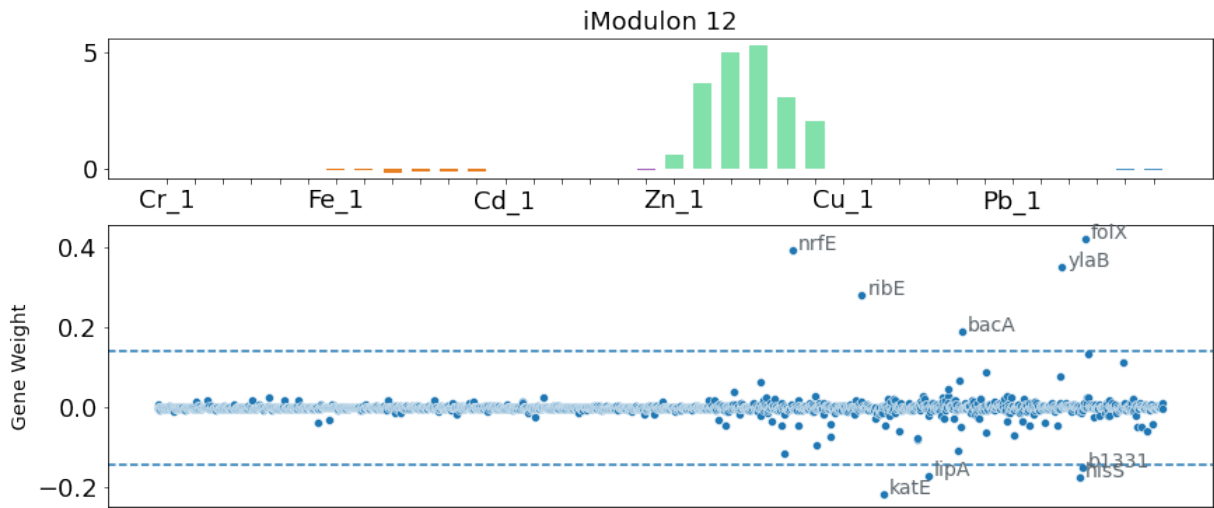
Figure 3.9. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 4. Bottom: Genes weights above threshold for IModulon 4.
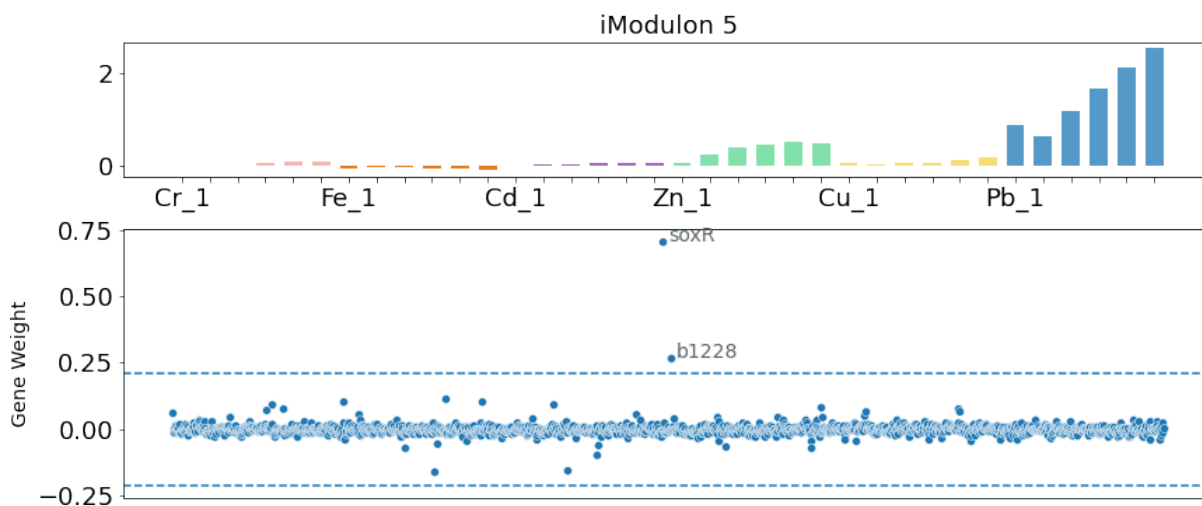
Figure 3.10. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 5. Bottom: Genes weights above threshold for IModulon 5.
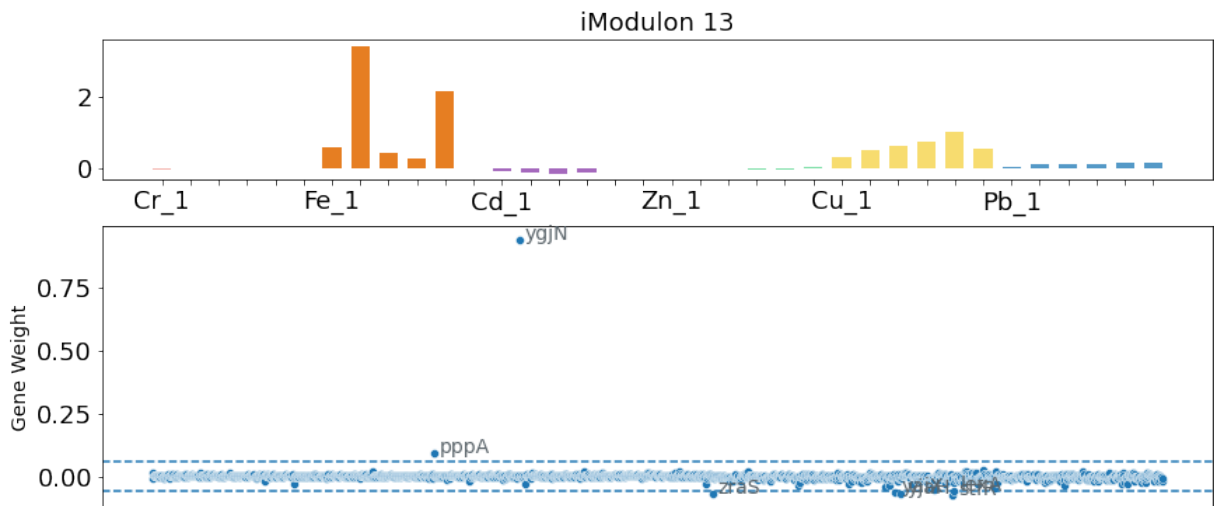
Figure 3.11. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 6. Bottom: Genes weights above threshold for IModulon 6.

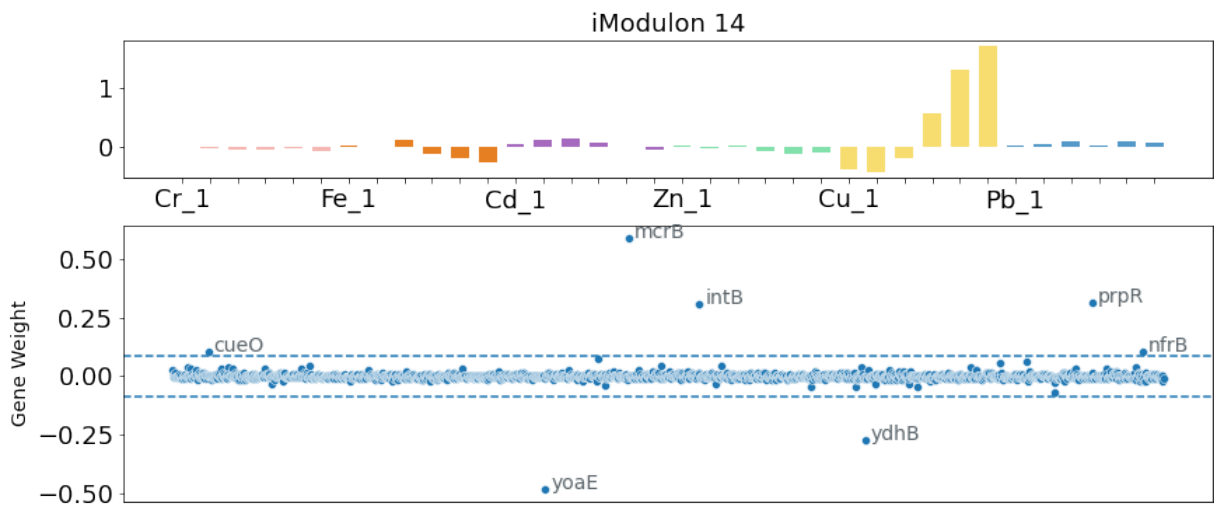

Figure 3.12. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 7. Bottom: Genes weights above threshold for IModulon 7.

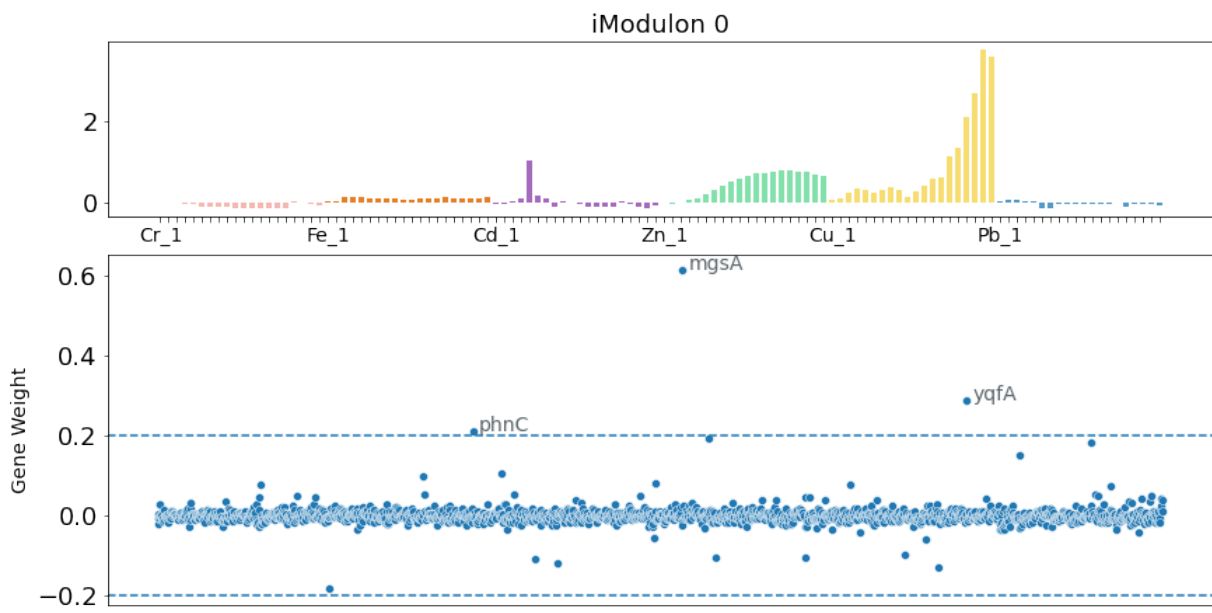**Figure 3.13.** ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 8. Bottom: Genes weights above threshold for IModulon 8.

Figure 3.14. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 9. Bottom: Genes weights above threshold for IModulon 9.
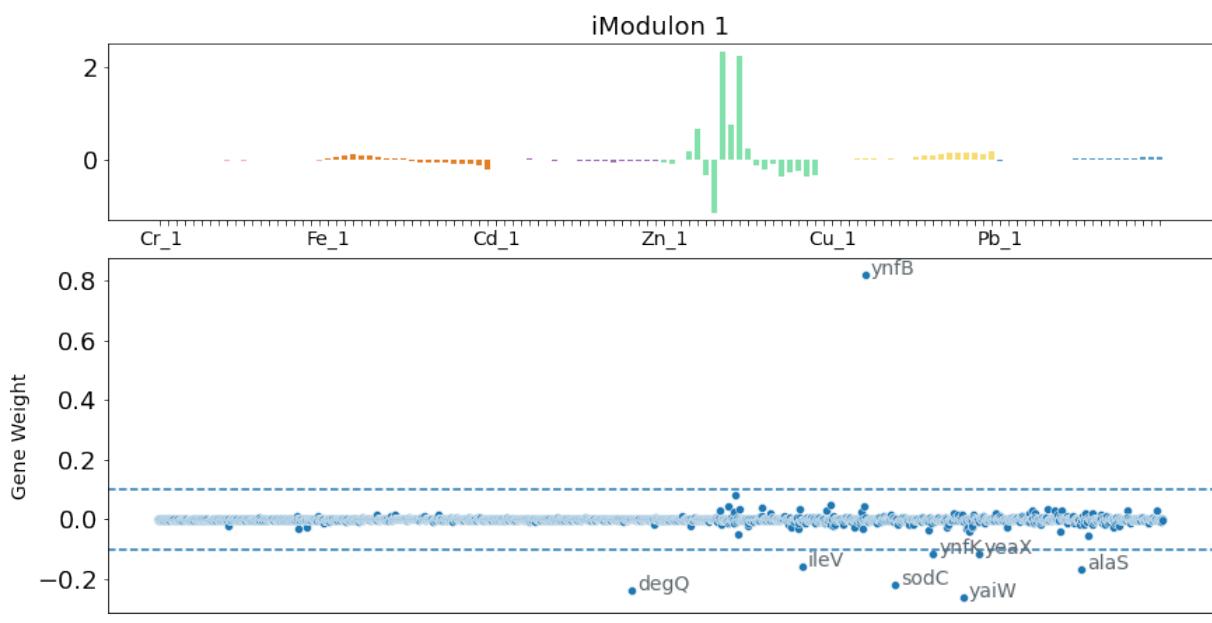
Figure 3.15. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 10. Bottom: Genes weights above threshold for IModulon 10.

Figure 3.16. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 11. Bottom: Genes weights above threshold for IModulon 11.
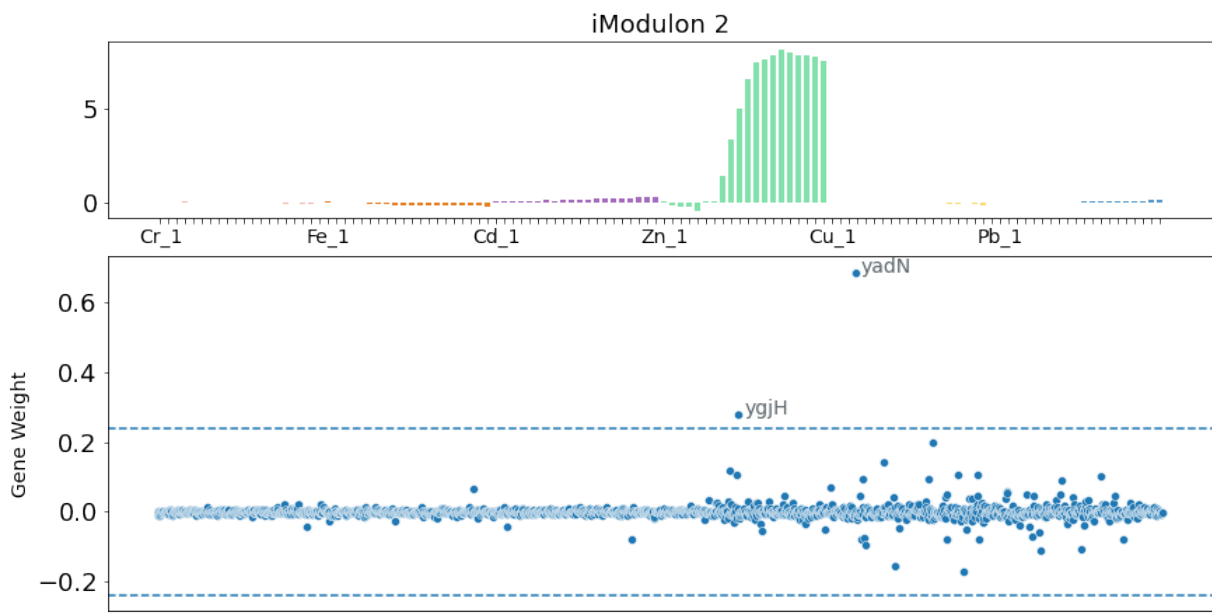
Figure 3.17. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 12. Bottom: Genes weights above threshold for IModulon 12.

Figure 3.18. ICA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 13. Bottom: Genes weights above threshold for IModulon 13.

Figure 3.19. CA results from running the algorithm on the induction window time points. Top: Activation profile plot of iModulon 14. Bottom: Genes weights above threshold for IModulon 14.

Figure 3.20. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 0. Bottom: Genes weights above threshold for IModulon 0.
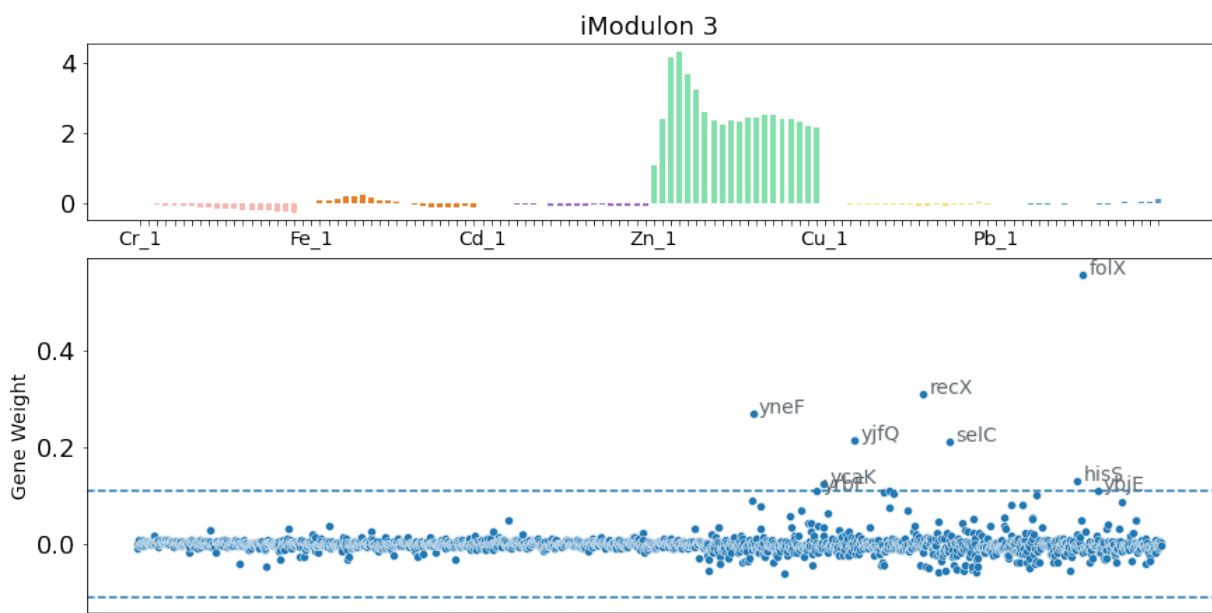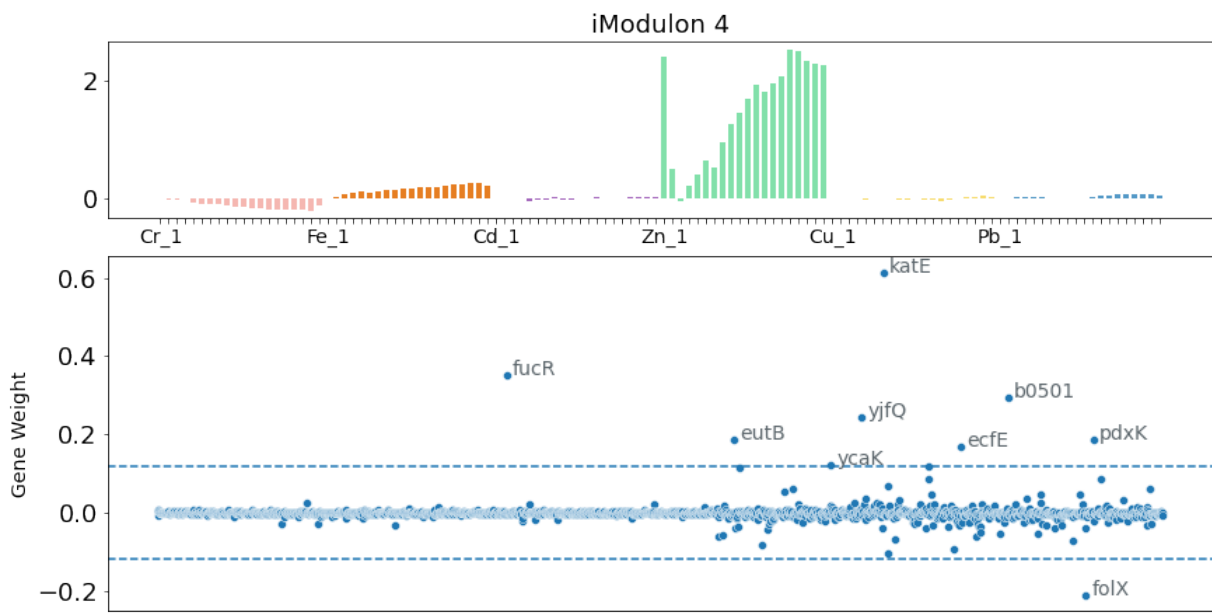

Figure 3.21. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 1. Bottom: Genes weights above threshold for IModulon 1.
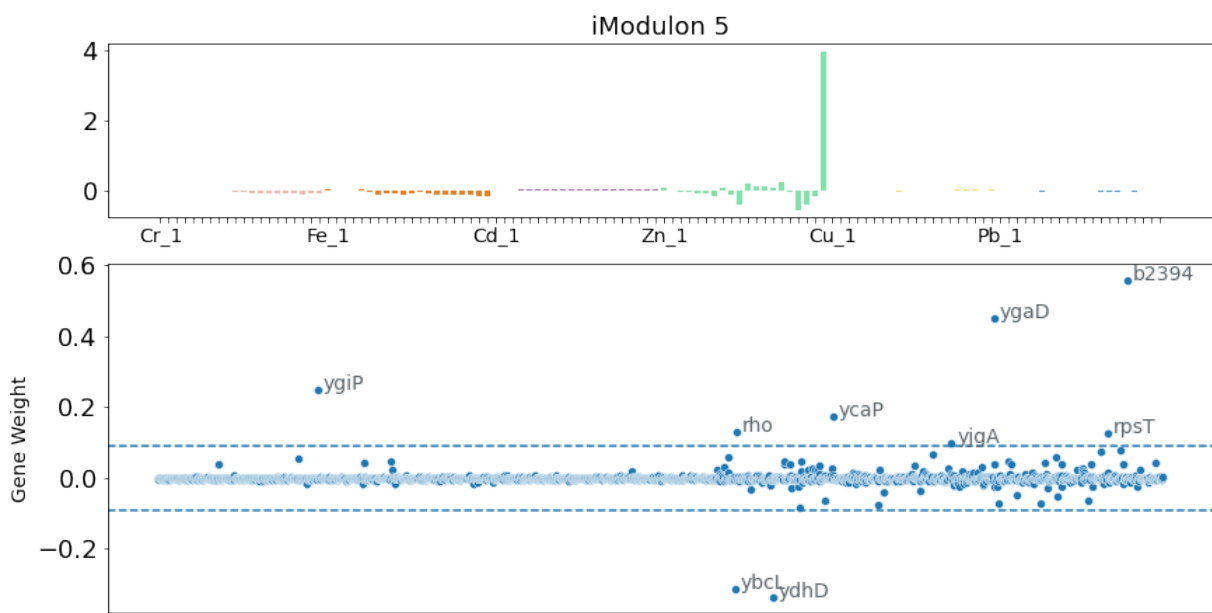
Figure 3.22. CA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 2. Bottom: Genes weights above threshold for IModulon 2.

Figure 3.23. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 3. Bottom: Genes weights above threshold for IModulon 3.
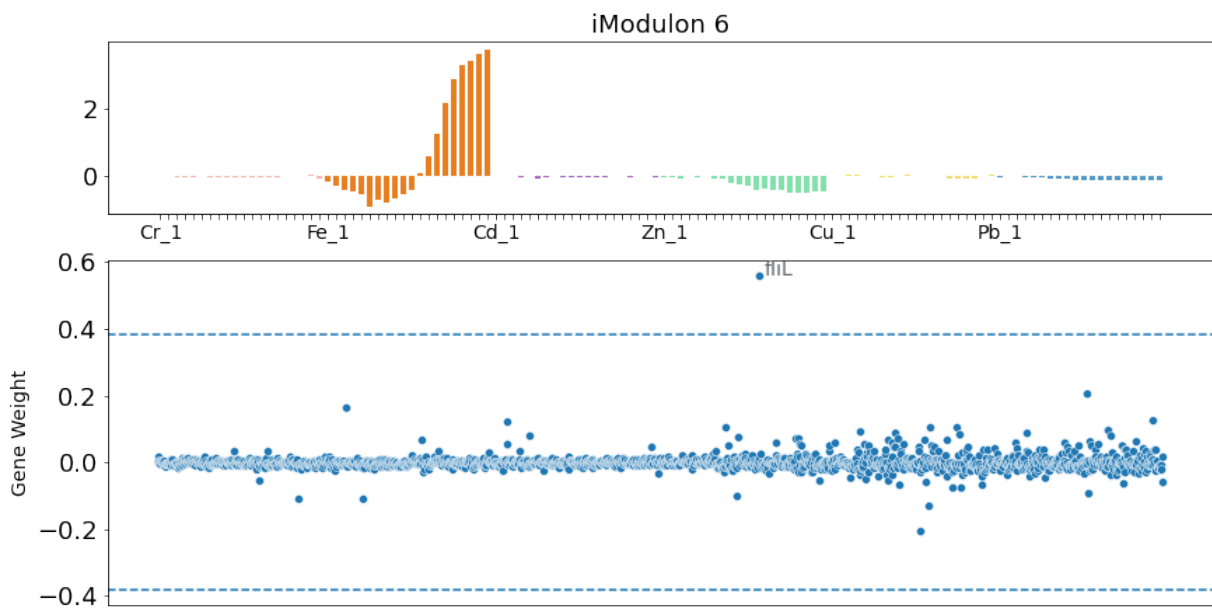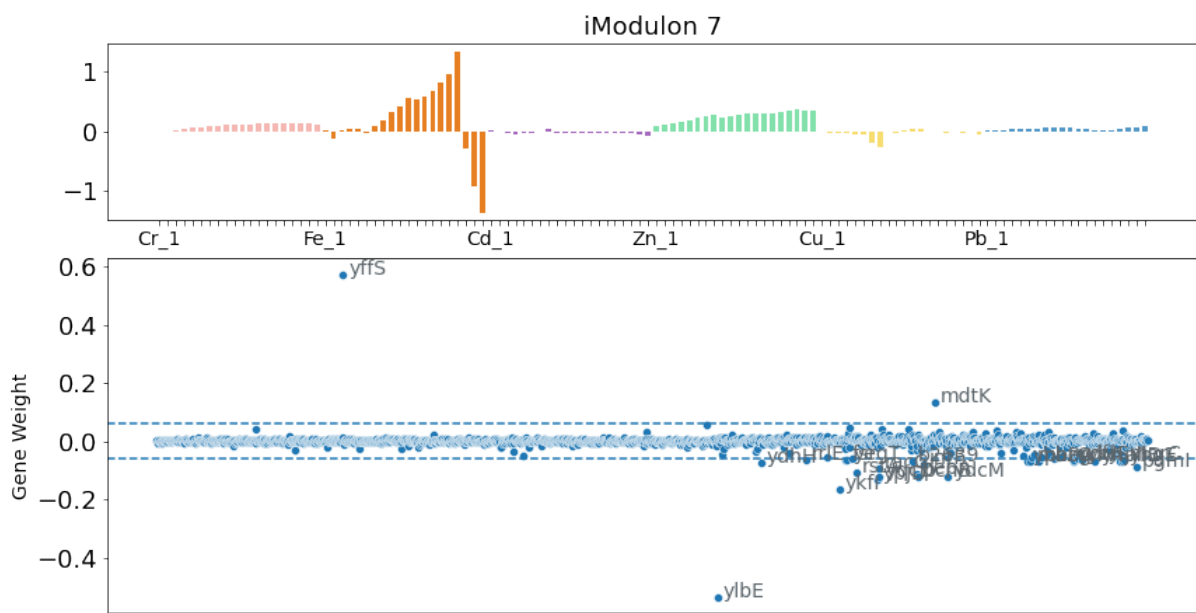
Figure 3.24. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 4. Bottom: Genes weights above threshold for IModulon 4.

Figure 3.25. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 5. Bottom: Genes weights above threshold for IModulon 5.

Figure 3.26. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 6. Bottom: Genes weights above threshold for IModulon 6.

Figure 3.27. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 7. Bottom: Genes weights above threshold for IModulon 7.
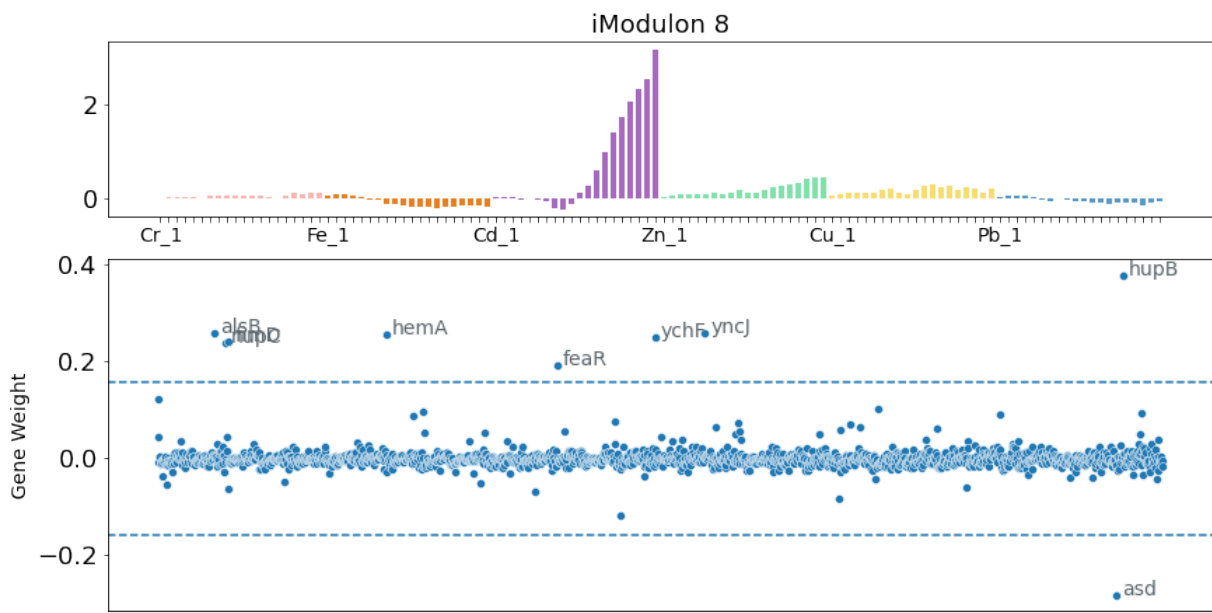
Figure 3.28. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 8. Bottom: Genes weights above threshold for IModulon 8.
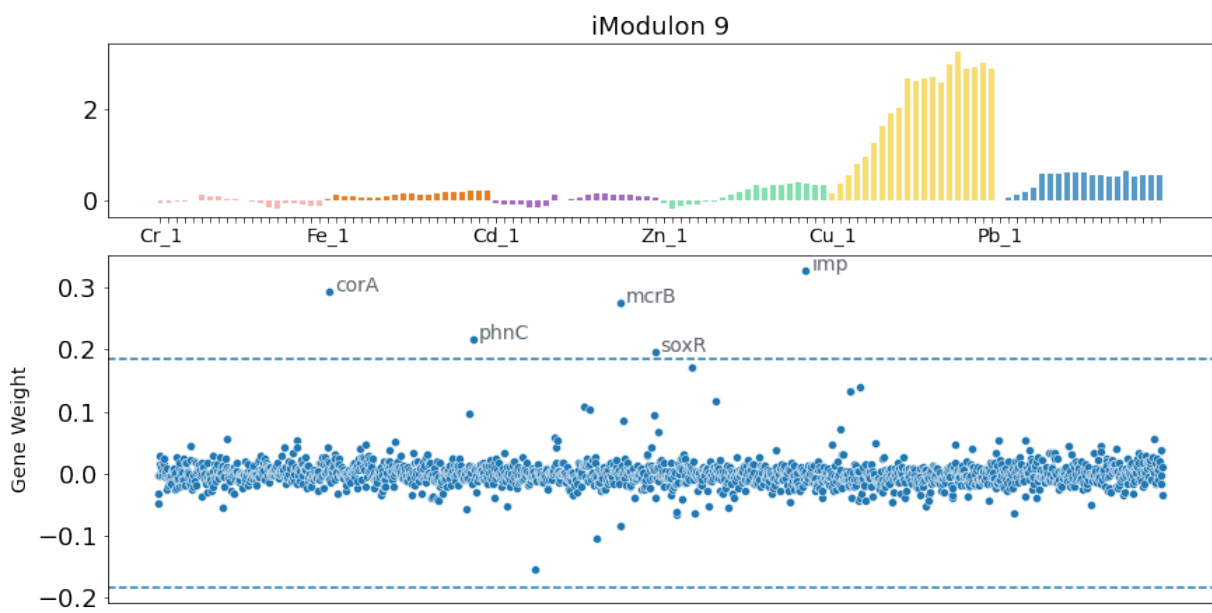

Figure 3.29. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 9. Bottom: Genes weights above threshold for IModulon 9.
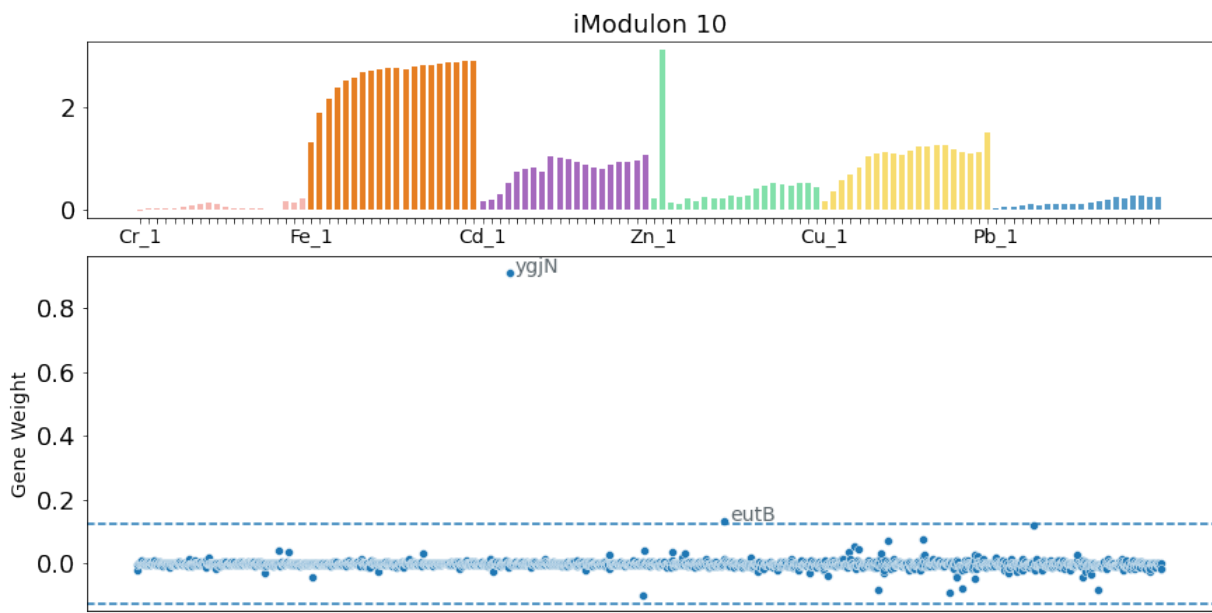
Figure 3.30. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 10. Bottom: Genes weights above threshold for IModulon 10.
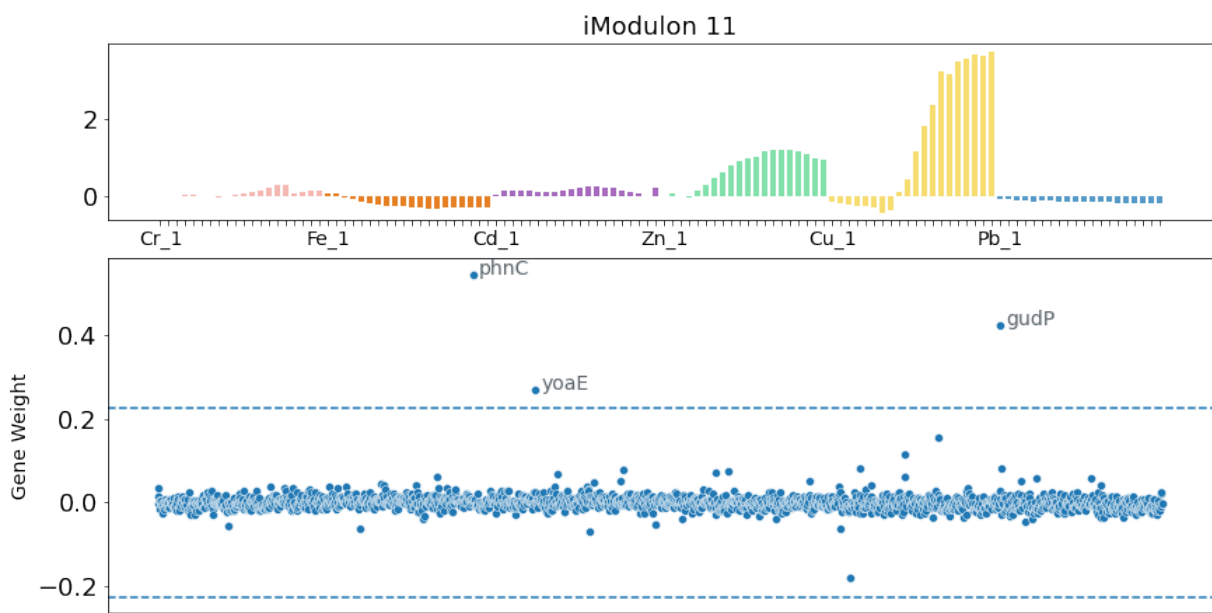

Figure 3.31. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 11. Bottom: Genes weights above threshold for IModulon 11.

Figure 3.32. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 12. Bottom: Genes weights above threshold for IModulon 12.
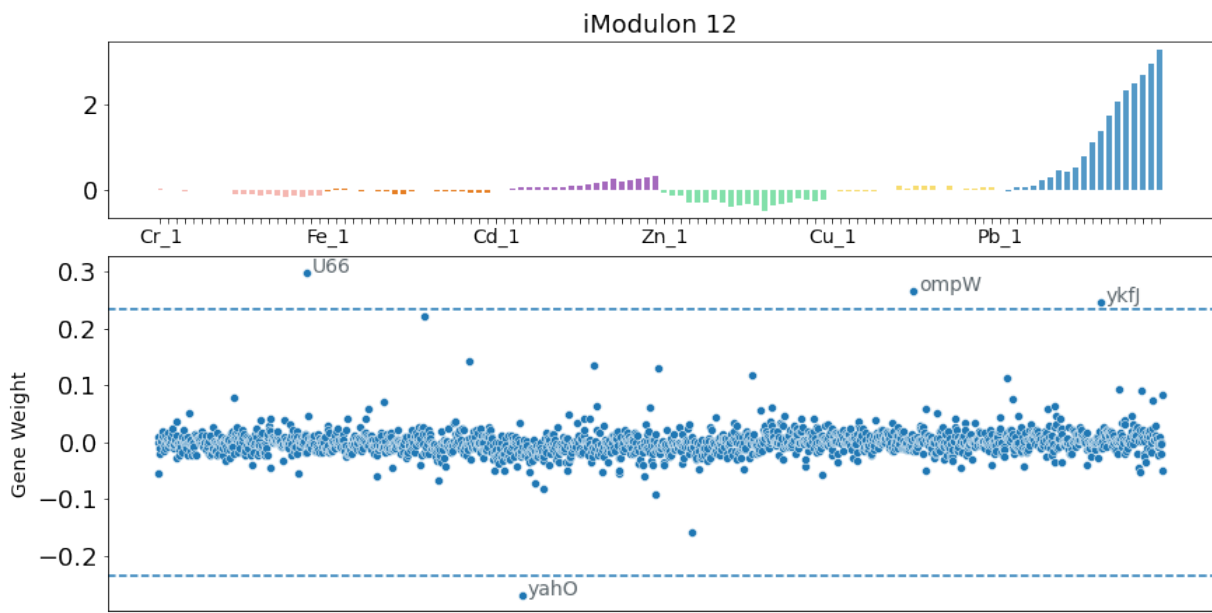

Figure 3.33. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 13. Bottom: Genes weights above threshold for IModulon 13.

Figure 3.34. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 14. Bottom: Genes weights above threshold for IModulon 14.

Figure 3.35. CA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 15. Bottom: Genes weights above threshold for IModulon 15.
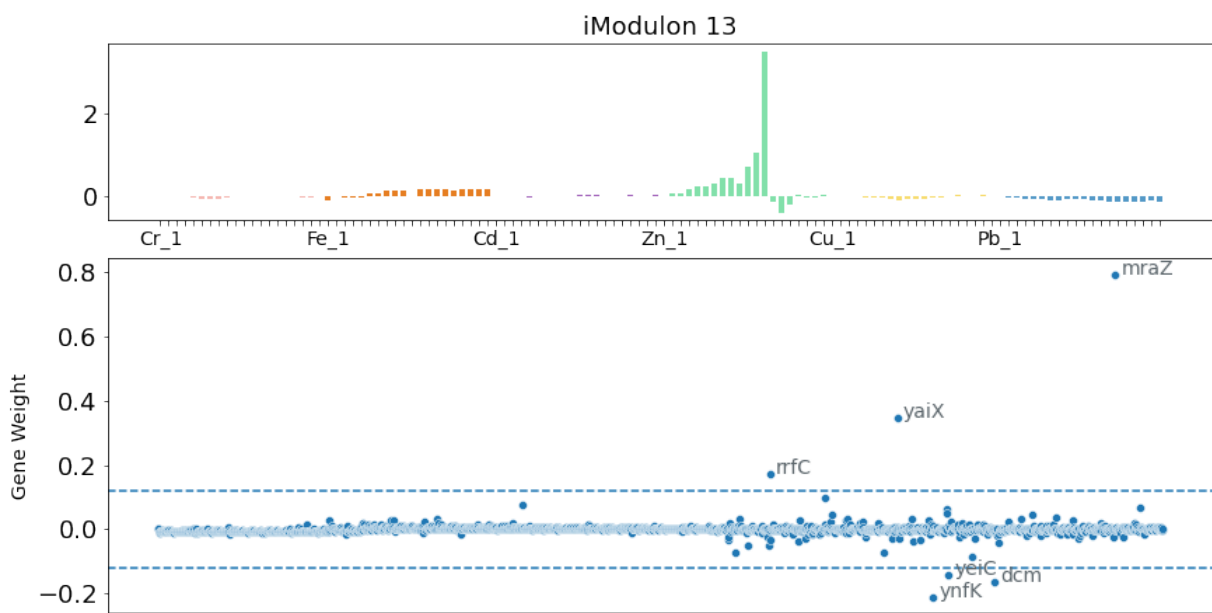
Figure 3.36. CA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 16. Bottom: Genes weights above threshold for IModulon 16.

Figure 3.37. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 17. Bottom: Genes weights above threshold for IModulon 17.
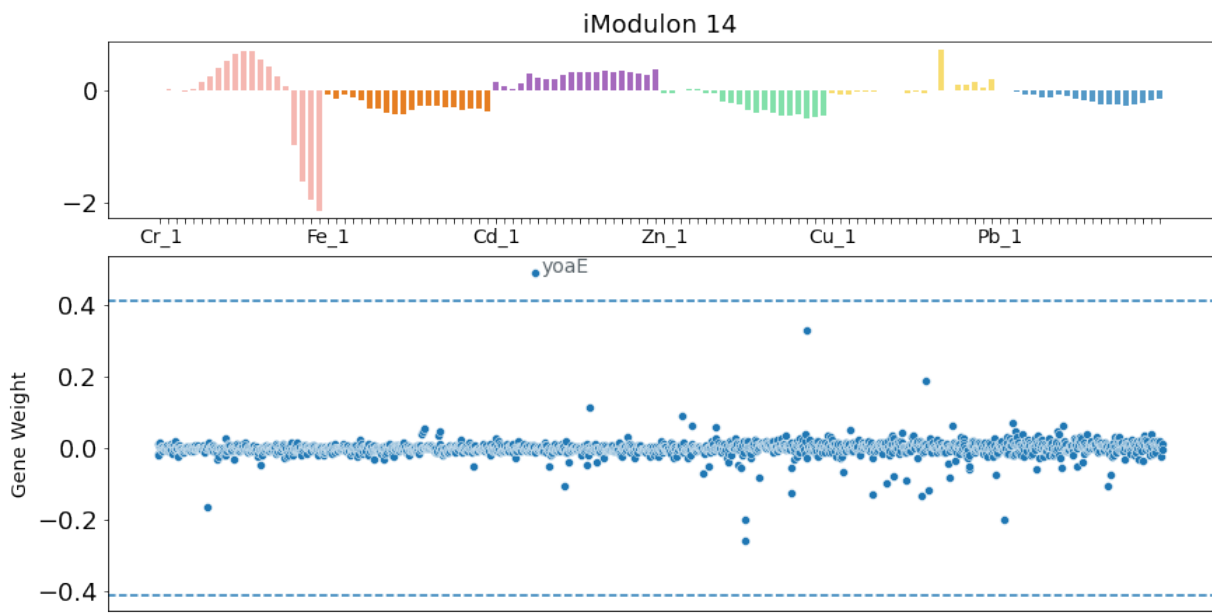
Figure 3.38. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 18. Bottom: Genes weights above threshold for IModulon 18.
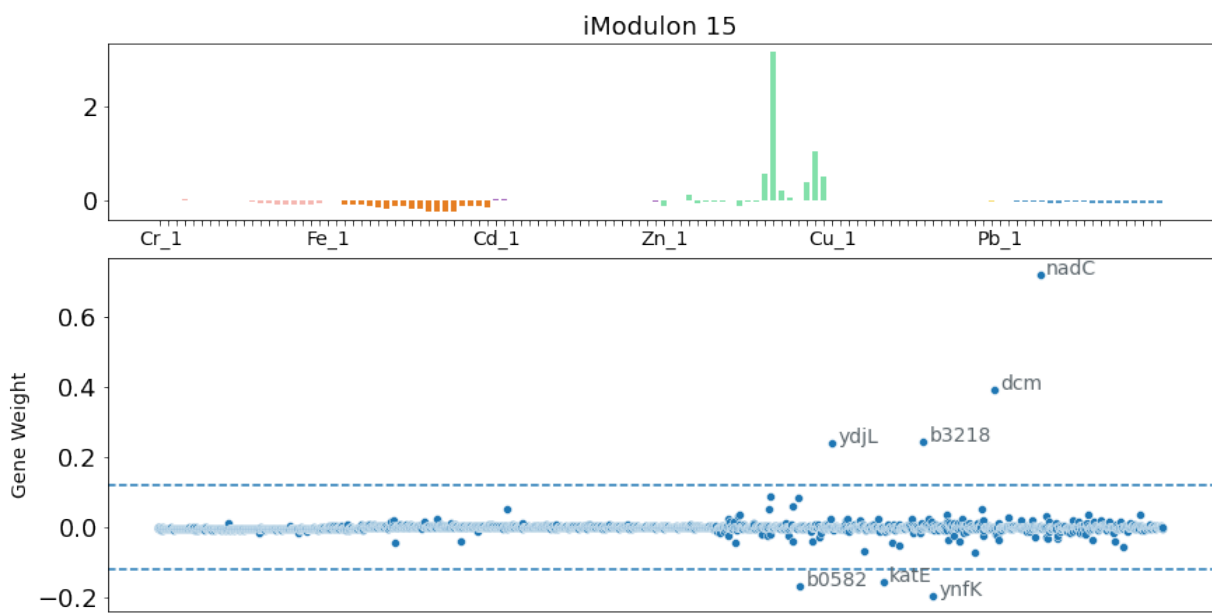

Figure 3.39. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 19. Bottom: Genes weights above threshold for IModulon 19.
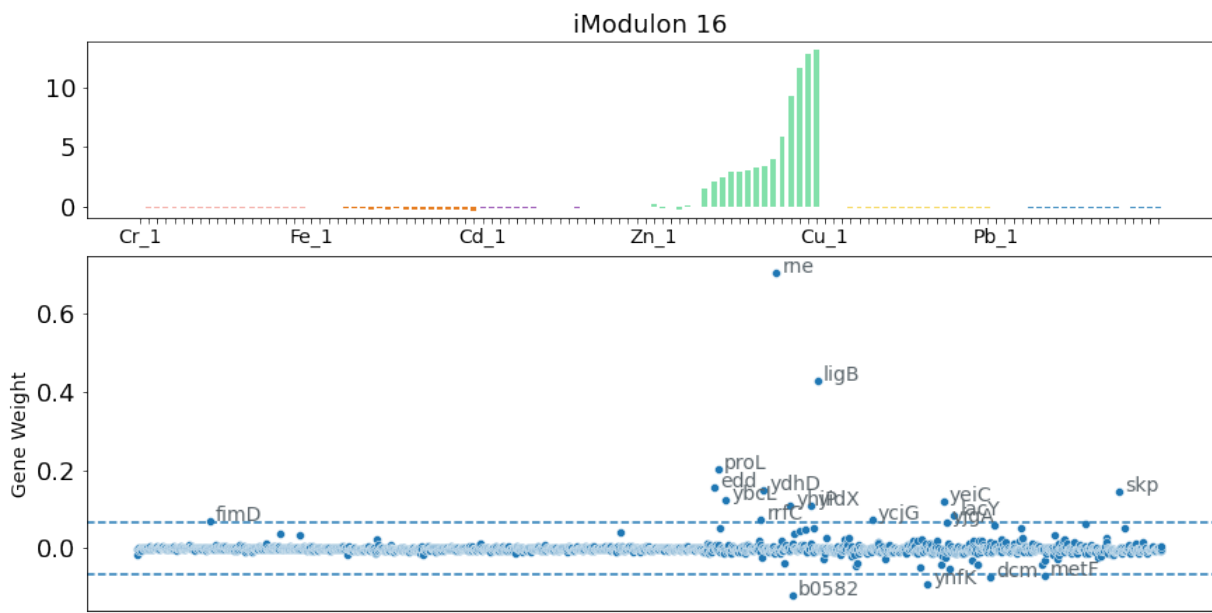
Figure 3.40. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 20. Bottom: Genes weights above threshold for IModulon 20.

Figure 3.41. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 21. Bottom: Genes weights above threshold for IModulon 21.
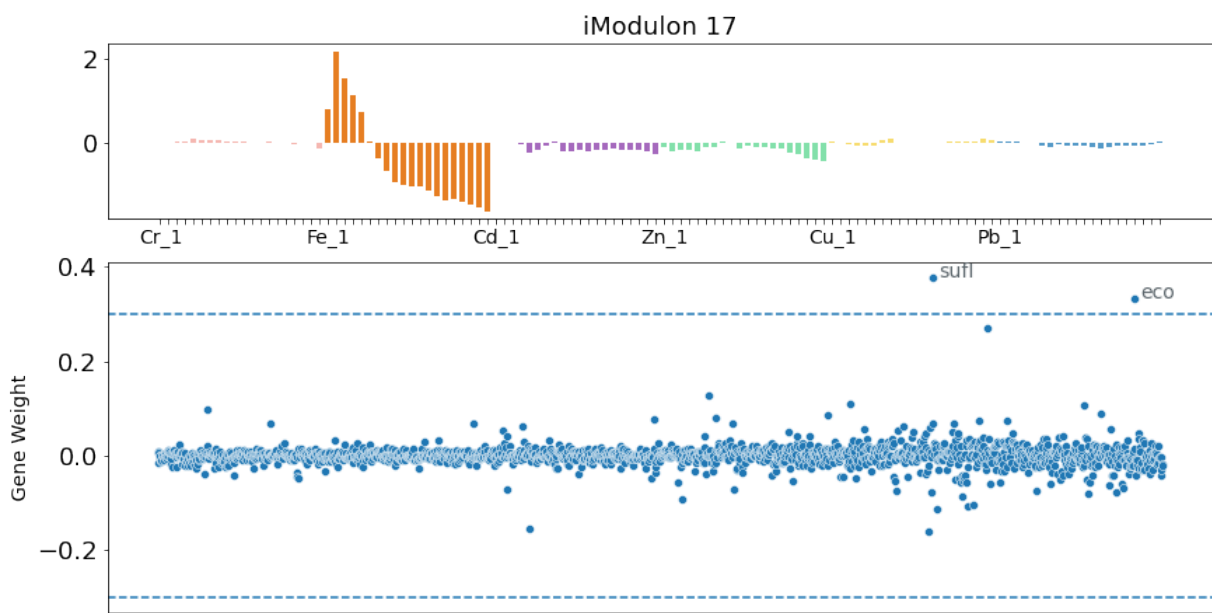
Figure 3.42. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 22. Bottom: Genes weights above threshold for IModulon 22.

Figure 3.43. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 23. Bottom: Genes weights above threshold for IModulon 23.

Figure 3.44. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 24. Bottom: Genes weights above threshold for IModulon 24.
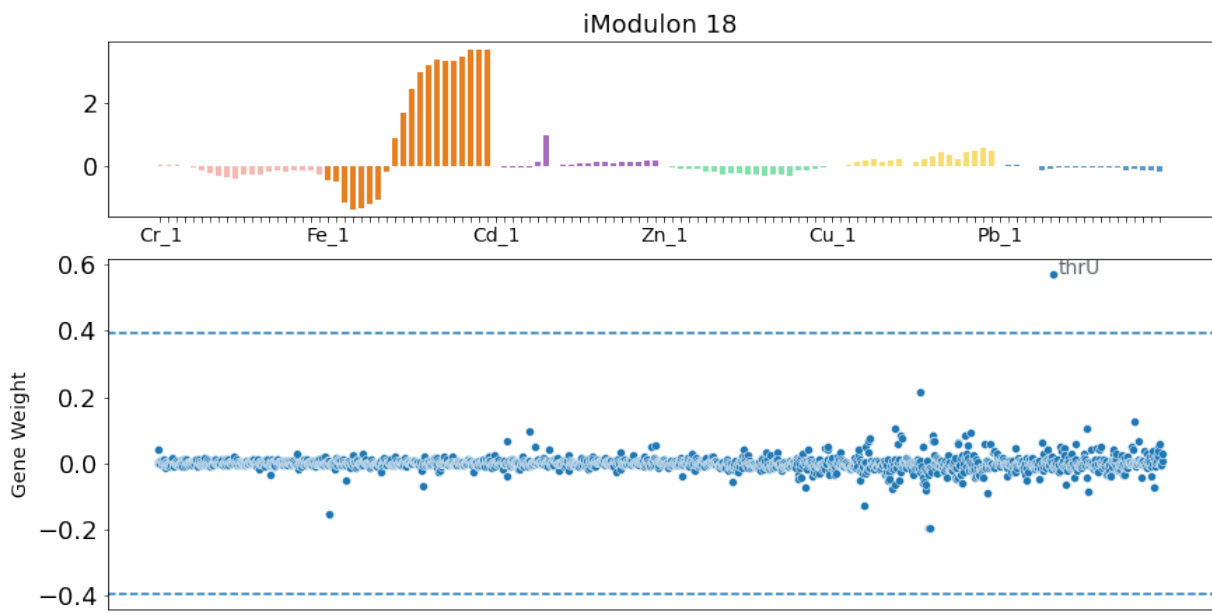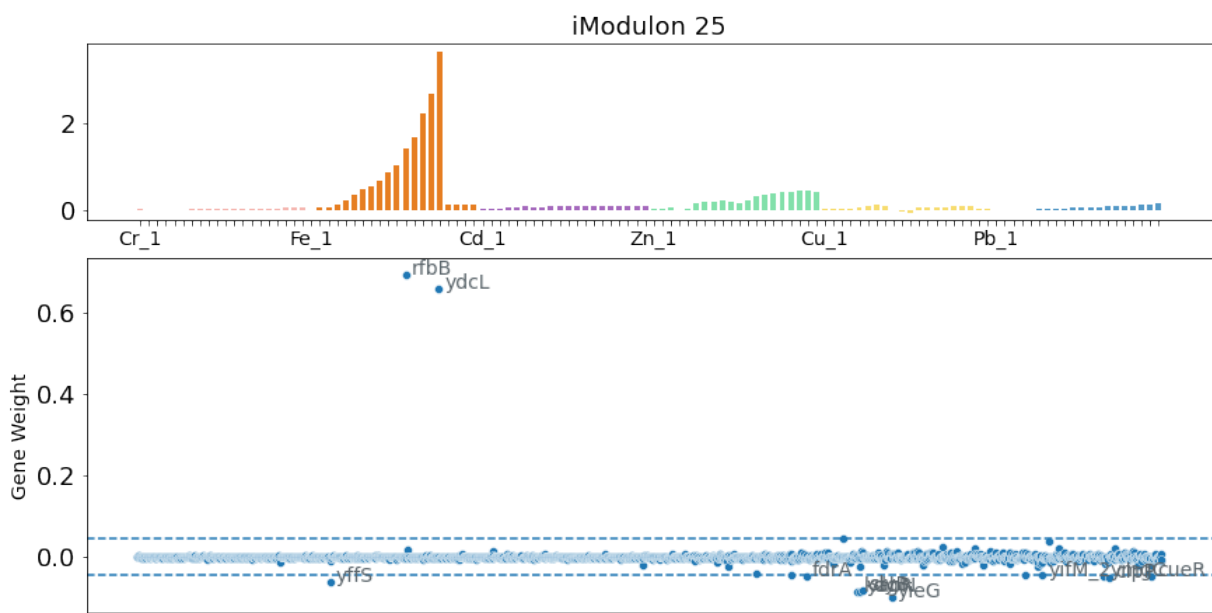

Figure 3.45. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 25. Bottom: Genes weights above threshold for IModulon 25.
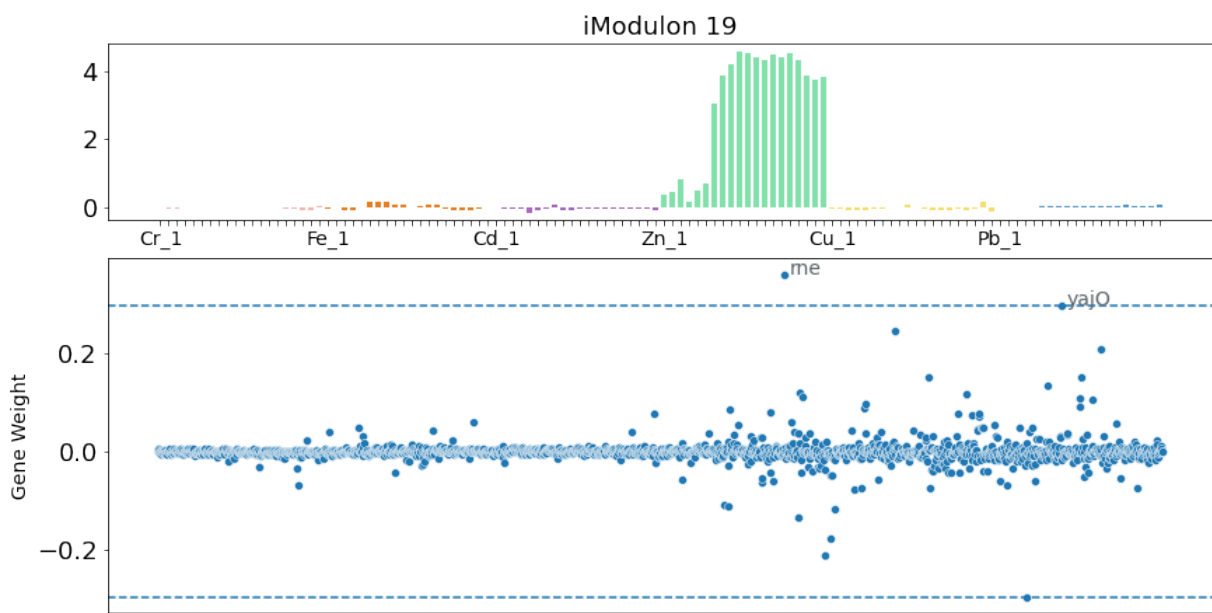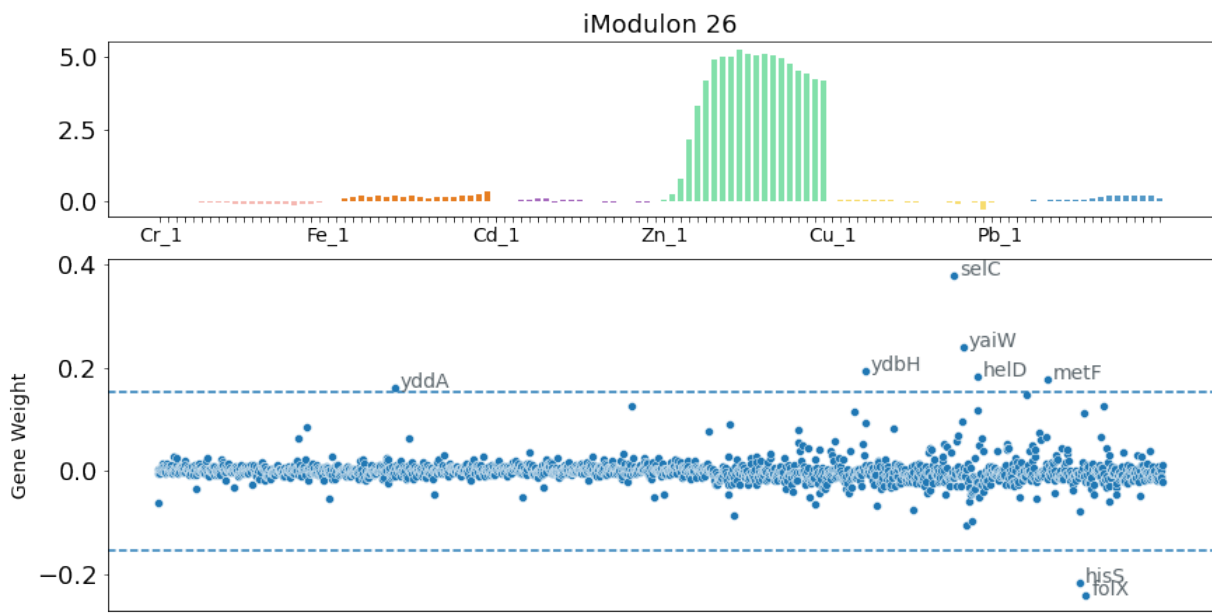
Figure 3.46. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 26. Bottom: Genes weights above threshold for IModulon 26.
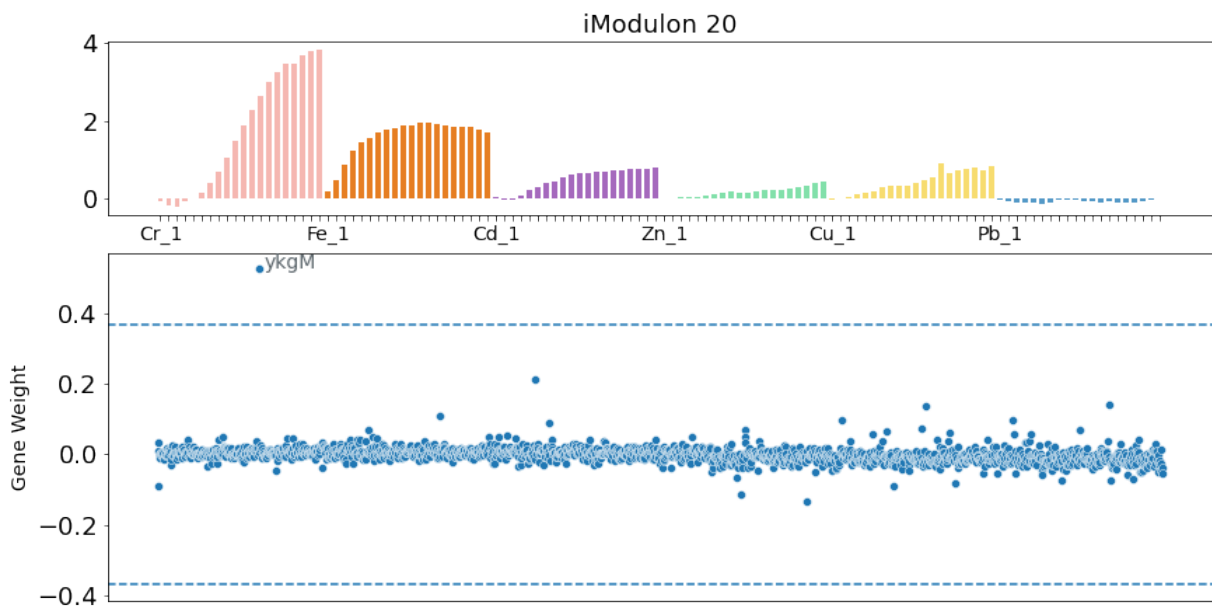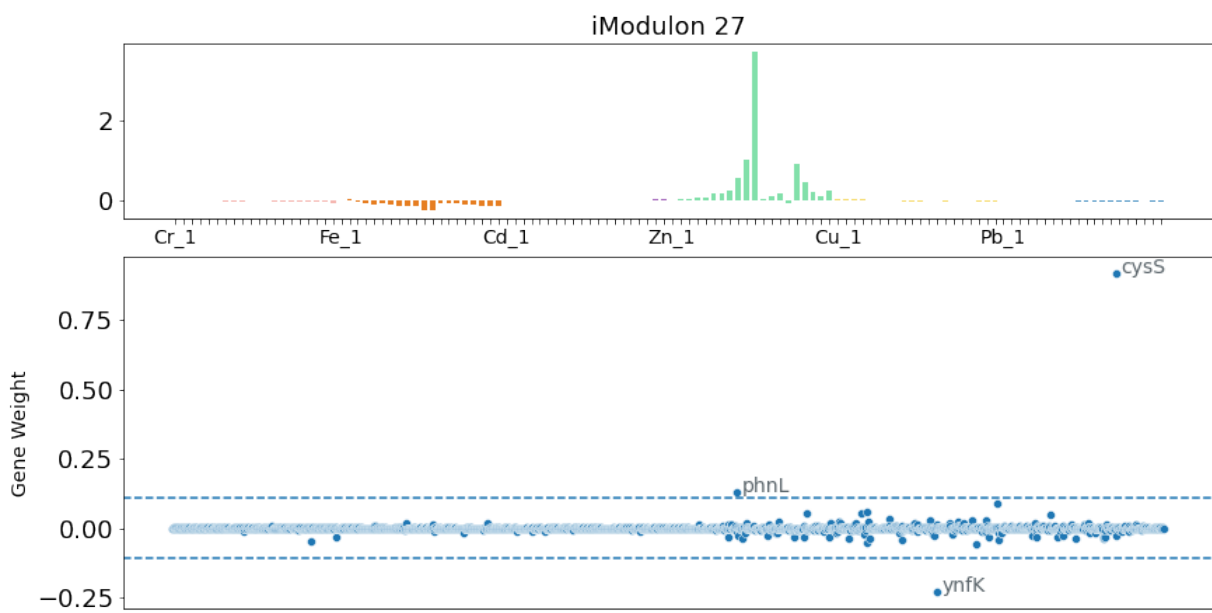

Figure 3.47. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 27. Bottom: Genes weights above threshold for IModulon 27.
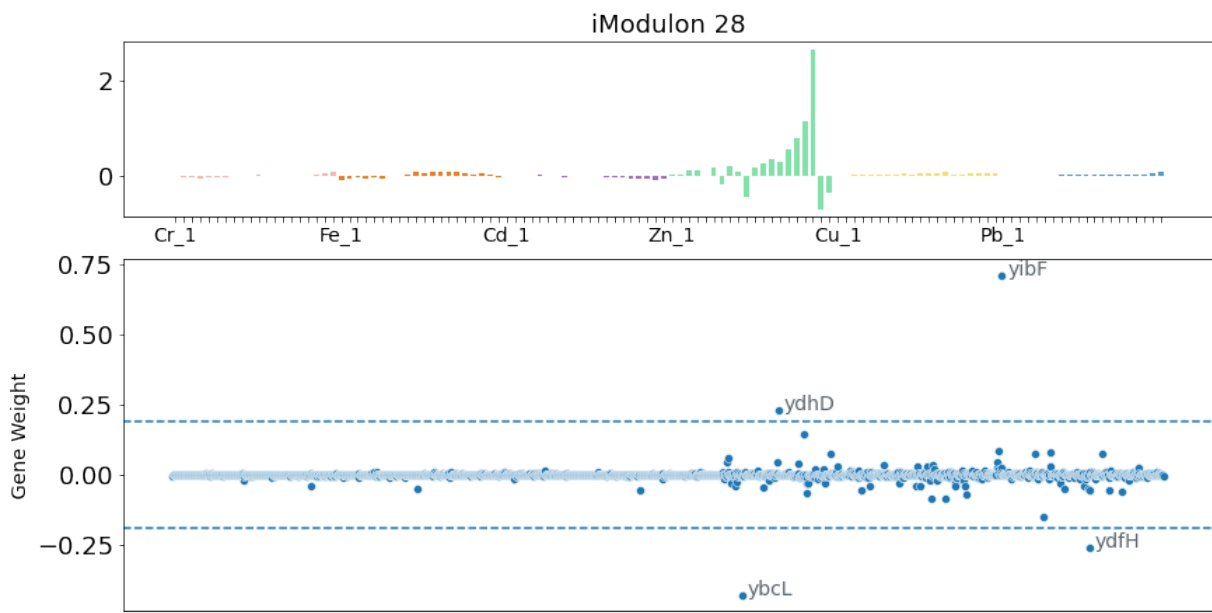
Figure 3.48. CA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 28. Bottom: Genes weights above threshold for IModulon 28.
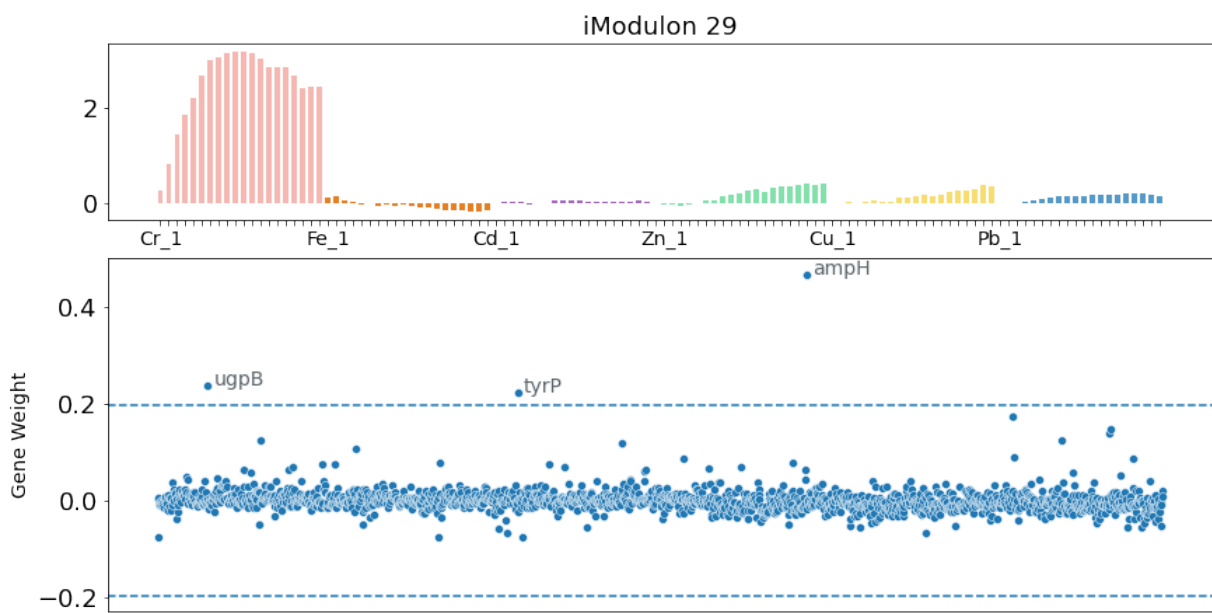
Figure 3.49. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 29. Bottom: Genes weights above threshold for IModulon 29.
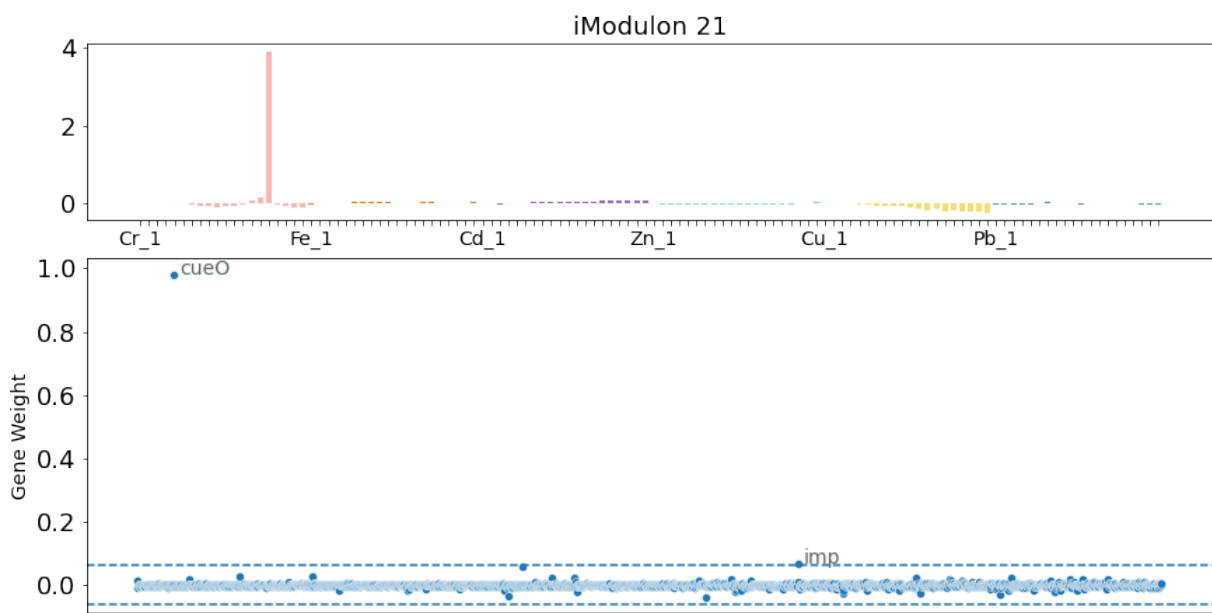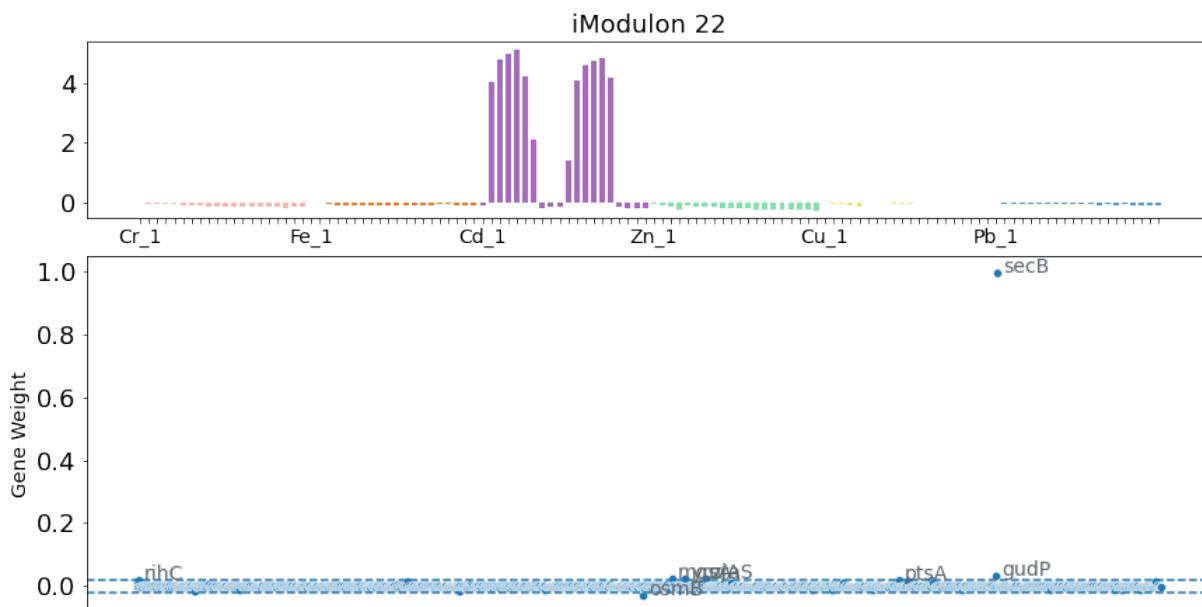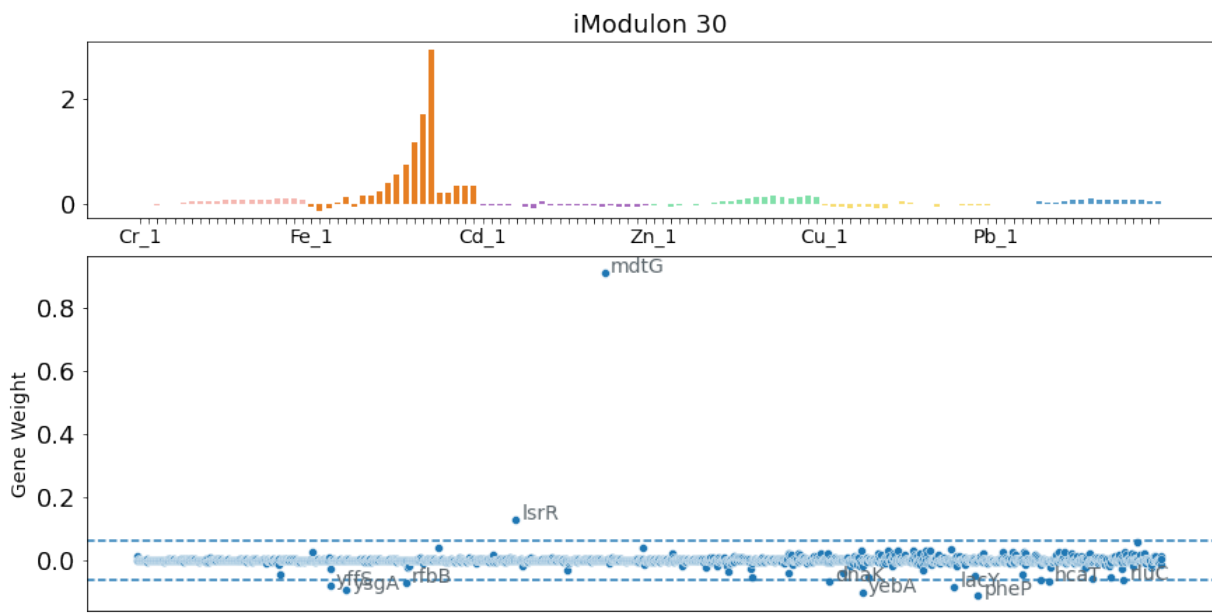
Figure 3.50. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 30. Bottom: Genes weights above threshold for IModulon 30.
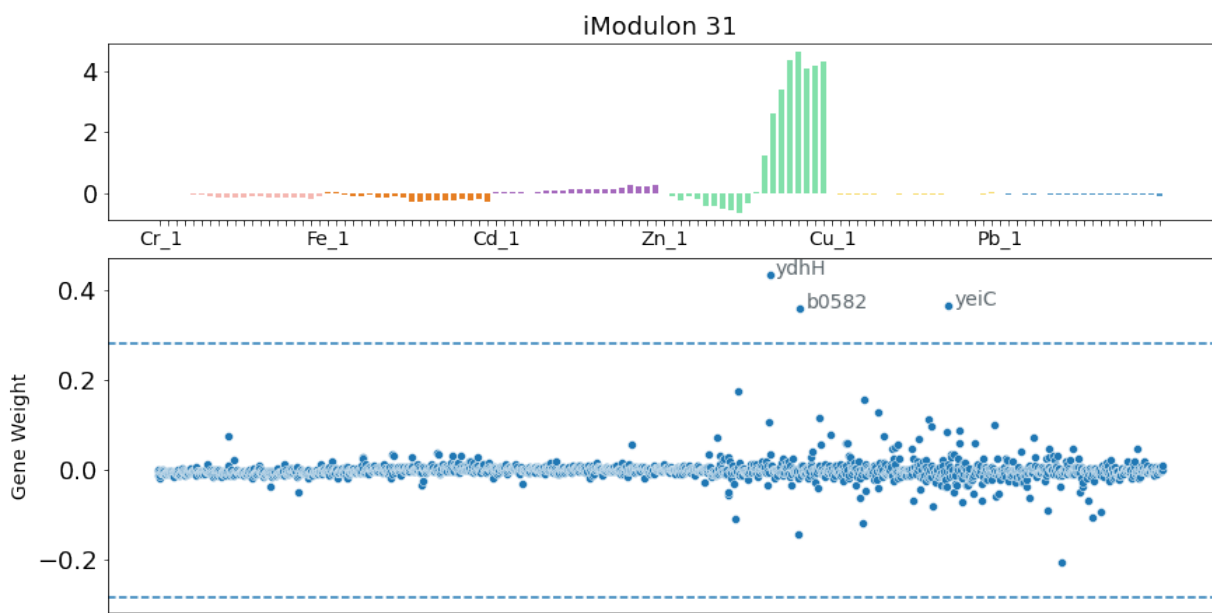
Figure 3.51. CA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 31. Bottom: Genes weights above threshold for IModulon 31.
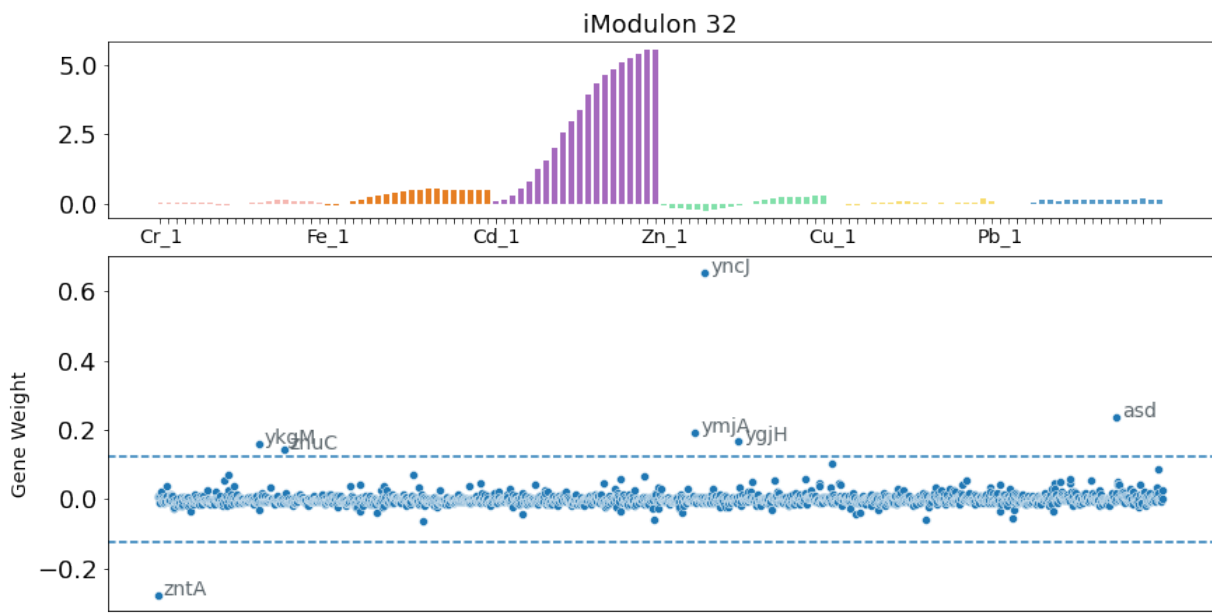
Figure 3.52. CA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 32. Bottom: Genes weights above threshold for IModulon 32.
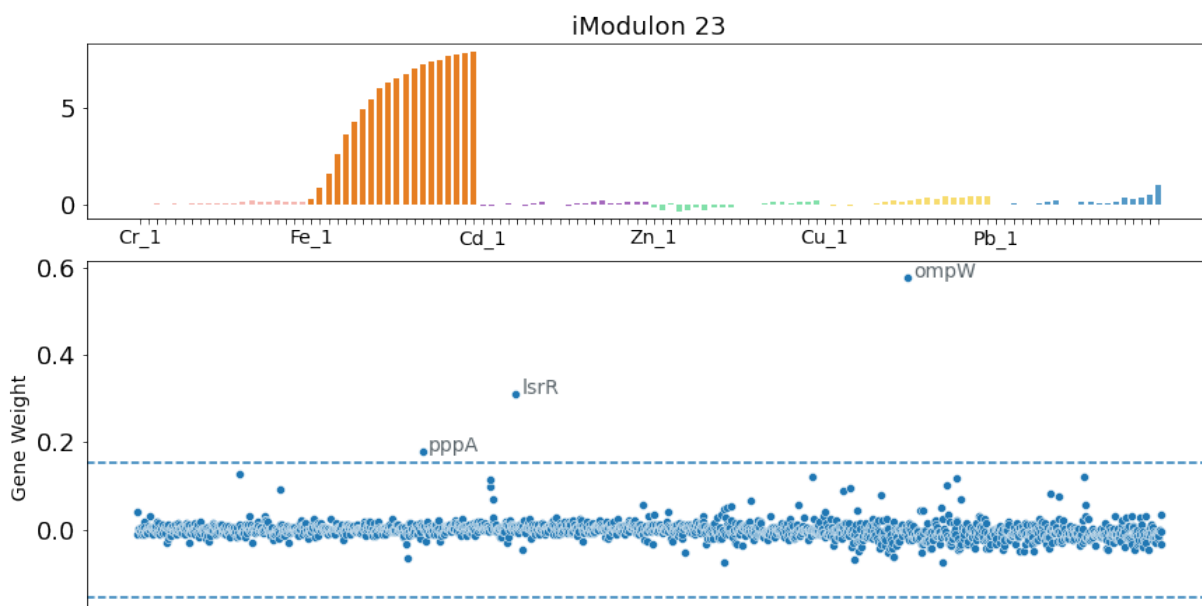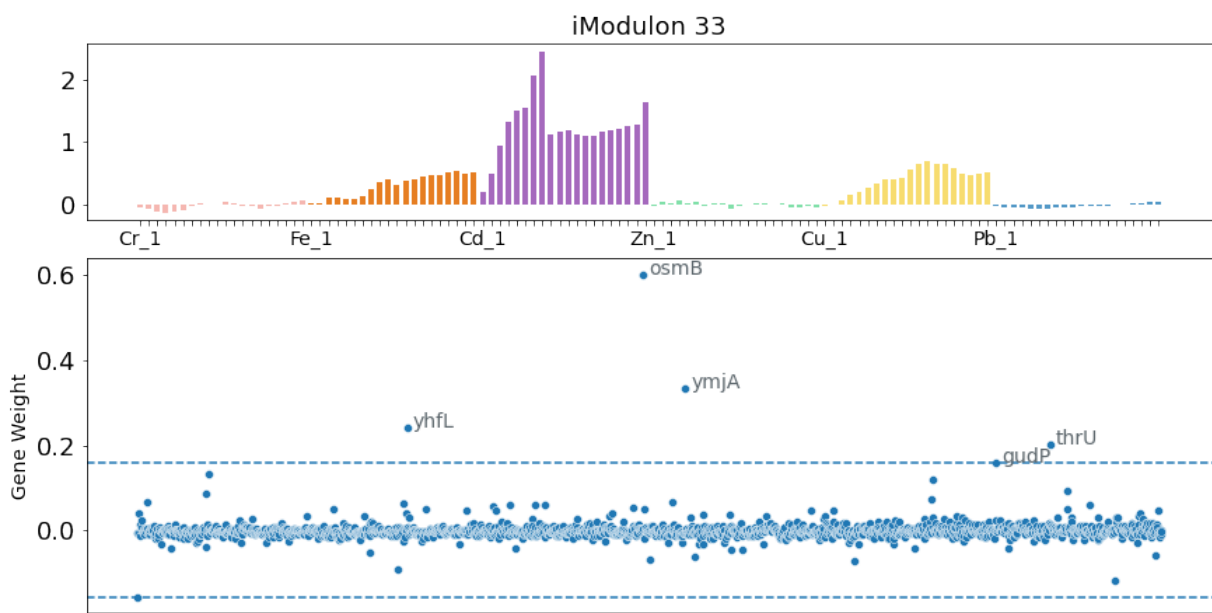

Figure 3.53. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 33. Bottom: Genes weights above threshold for IModulon 33.
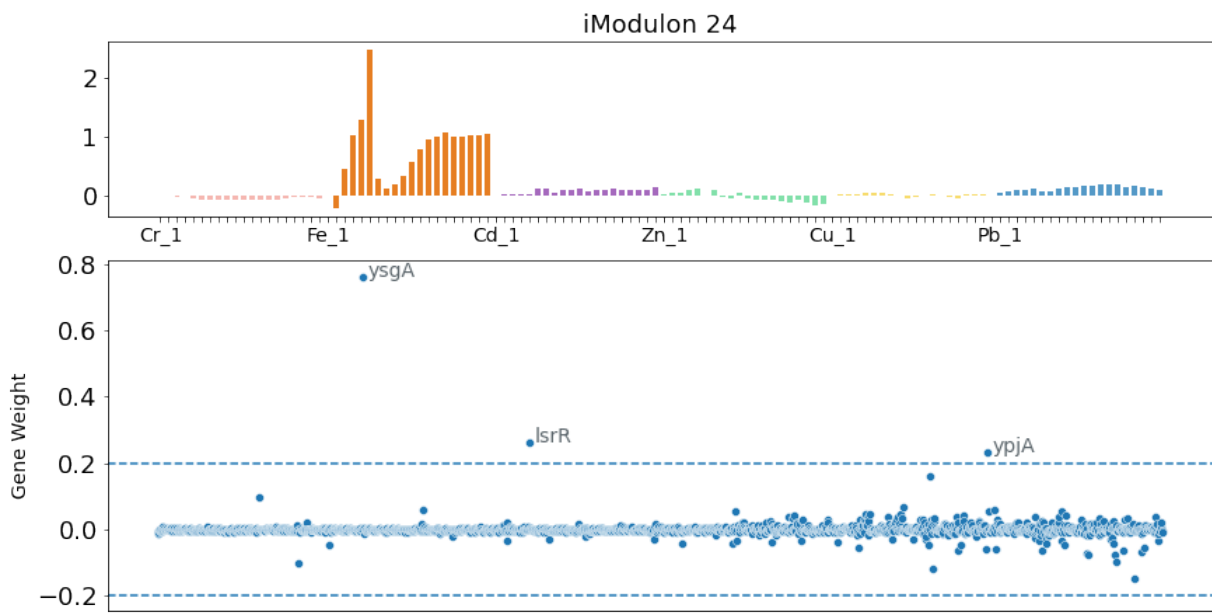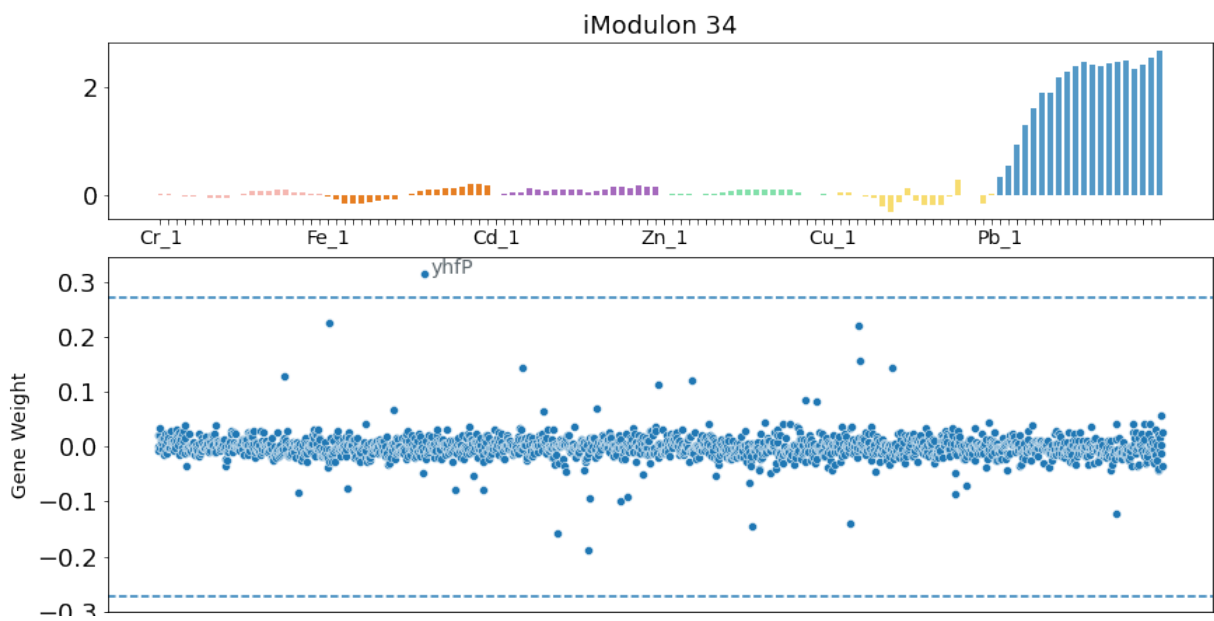
Figure 3.54. ICA results from running the algorithm on the data post induction. Top: Activation profile plot of iModulon 34. Bottom: Genes weights above threshold for IModulon 34.

Bibliography

[1] DEEPAK K Agrawal and BARRY L Wanner. A phoa structural gene mutation that conditionally affects formation of the enzyme bacterial alkaline phosphatase. *Journal of bacteriology*, 172(6):3180–3190, 1990.

[2] PP Avelino, BF de Oliveira, and RS Trintin. Predominance of the weakest species in lotka-volterra and may-leonard formulations of the rock-paper-scissors model. *Physical Review E*, 100(4):042209, 2019.

[3] Reuven Babai and Eliora Z Ron. An escherichia coli gene responsive to heavy metals. *FEMS microbiology letters*, 167(2):107–111, 1998.

[4] Frederick K Balagaddé, Hao Song, Jun Ozaki, Cynthia H Collins, Matthew Barnet, Frances H Arnold, Stephen R Quake, and Lingchong You. A synthetic escherichia coli predator–prey ecosystem. *Molecular systems biology*, 4(1), 2008.

[5] Subhayu Basu, Yoram Gerchman, Cynthia H Collins, Frances H Arnold, and Ron Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037):1130, 2005.

[6] Maria A Bauer, Katharina Kainz, Didac Carmona-Gutierrez, and Frank Madeo. Microbial wars: competition in ecological niches and within the microbiome. *Microbial Cell*, 5(5):215, 2018.

[7] Steven J Beard, Rohani Hashim, Jorge Membrillo-Hernández, Martin N Hughes, and Robert K Poole. Zinc (ii) tolerance in escherichia coli k-12: evidence that the znta gene (o732) encodes a cation transport atpase. *Molecular microbiology*, 25(5):883–891, 1997.

[8] François Beaufay, Ellen Quarles, Allison Franz, Olivia Katamanin, Wei-Yun Wholey, and Ursula Jakob. Polyphosphate functions in vivo as an iron chelator and fenton reaction inhibitor. *Mbio*, 11(4), 2020.

[9] K Bernhardt, EJ Carter, NS Chand, J Lee, Y Xu, X Zhu, JW Ajioka, JM Goncalves, J Haseloff, G Micklem, et al. New tools for self-organised pattern formation. *IET Synthetic Biology*, 1(1):29–31, 2007.

[10] Jyoti Bisht and NSK Harsh. Utilizing aspergillus niger for bioremediation of tannery effluent. *Octa Journal of Environmental Research*, 2(1), 2014.

[11] Philip Bittihn, M Omar Din, Lev S Tsimring, and Jeff Hasty. Rational engineering of synthetic microbial systems: from single cells to consortia. *Current opinion in microbiology*, 45:92–99, 2018.

[12] Hüseyin Boz. p-coumaric acid in cereals: presence, antioxidant and antimicrobial effects. *International journal of food science & technology*, 50(11):2323–2328, 2015.

[13] Katie Brenner, Lingchong You, and Frances H Arnold. Engineering microbial consortia: a new frontier in synthetic biology. *Trends in biotechnology*, 26(9):483–489, 2008.

[14] LW Buss and JBC Jackson. Competitive networks: nontransitive competitive relationships in cryptic coral reef environments. *The American Naturalist*, 113(2):223–234, 1979.

[15] Eva María Camacho, Beatriz Mesa-Pereira, Carlos Medina, Amando Flores, and Eduardo Santero. Engineering salmonella as intracellular factory for effective killing of tumour cells. *Scientific reports*, 6:30591, 2016.

[16] Eric Cascales, Susan K Buchanan, Denis Duché, Colin Kleanthous, Roland Lloubes, Kathleen Postle, Margaret Riley, Stephen Slatin, and Daniele Cavard. Colicin biology. *Microbiology and molecular biology reviews*, 71(1):158–229, 2007.

[17] Sunitha M Chacko, Kottayath G Nevin, R Dhanyakrishnan, and B Prakash Kumar. Protective effect of p-coumaric acid against doxorubicin induced toxicity in h9c2 cardiomyoblast cell lines. *Toxicology Reports*, 2:1213–1221, 2015.

[18] Clement TY Chan, Jeong Wook Lee, D Ewen Cameron, Caleb J Bashor, and James J Collins. 'deadman' and 'passcode' microbial kill switches for bacterial containment. *Nature chemical biology*, 12(2):82, 2016.

[19] Lin Chao and Bruce R Levin. Structured habitats and the evolution of anticompetitor toxins in bacteria. *Proceedings of the National Academy of Sciences*, 78(10):6324–6328, 1981.

[20] Ye Chen, Jae Kyoung Kim, Andrew J Hirning, Krešimir Josić, and Matthew R Bennett. Emergent genetic oscillations in a synthetic microbial consortium. *Science*, 349(6251):986–989, 2015.

[21] Peter Chesson. Mechanisms of maintenance of species diversity. *Annual review of Ecology and Systematics*, 31(1):343–366, 2000.

[22] 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*, 25(7):1057–1063, 2019.

[23] Pierre Comon. Independent component analysis, a new concept? *Signal processing*, 36(3):287–314, 1994.

[24] Katharine Z Coyte and Seth Rakoff-Nahoum. Understanding competition and cooperation within the mammalian gut microbiome. *Current Biology*, 29(11):R538–R544, 2019.

[25] Katharine Z Coyte, Jonas Schluter, and Kevin R Foster. The ecology of the microbiome: networks, competition, and stability. *Science*, 350(6261):663–666, 2015.

[26] Tamás L Czárán, Rolf F Hoekstra, and Ludo Pagie. Chemical warfare between microbes promotes biodiversity. *Proceedings of the National Academy of Sciences*, 99(2):786–790, 2002.

[27] Tal Danino, Octavio Mondragón-Palomino, Lev Tsimring, and Jeff Hasty. A synchronized quorum of genetic clocks. *Nature*, 463(7279):326, 2010.

[28] Vanessa M D'Costa, Katherine M McGrann, Donald W Hughes, and Gerard D Wright. Sampling the antibiotic resistome. *Science*, 311(5759):374–377, 2006.

[29] M Omar Din, Tal Danino, Arthur Prindle, Matt Skalak, Jangir Selimkhanov, Kaitlin Allen, Ellixis Julio, Eta Atolia, Lev S Tsimring, Sangeeta N Bhatia, et al. Synchronized cycles of bacterial lysis for in vivo delivery. *Nature*, 536(7614):81, 2016.

[30] Karrera Y Djoko, Lee Xin Chong, Anthony G Wedd, and Zhiguang Xiao. Reaction mechanisms of the multicopper oxidase cueo from escherichia coli support its functional role as a cuprous oxidase. *Journal of the American Chemical Society*, 132(6):2005–2015, 2010.

[31] Knut Drescher, Carey D Nadell, Howard A Stone, Ned S Wingreen, and Bonnie L Bassler. Solutions to the public goods dilemma in bacterial biofilms. *Current Biology*, 24(1):50–55, 2014.

[32] PAUL V Dunlap and EP Greenberg. Control of vibrio fischeri luminescence gene expression in escherichia coli by cyclic amp and cyclic amp receptor protein. *Journal of bacteriology*, 164(1):45–50, 1985.

[33] Mike S Ferry, Ivan A Razinkov, and Jeff Hasty. Microfluidics for synthetic biology: from design to execution. In *Methods in enzymology*, volume 497, pages 295–372. Elsevier, 2011.

[34] Marcus Frean and Edward R Abraham. Rock–scissors–paper and the survival of the

weakest. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1474):1323–1327, 2001.

[35] Leonor García-Bayona and Laurie E Comstock. Bacterial antagonism in host-associated microbial communities. *Science*, 361(6408):eaat2456, 2018.

[36] Claire Geslin, Jenny Llanos, Daniel Prieur, and Christian Jeanthon. The manganese and iron superoxide dismutases protect escherichia coli from heavy metal toxicity. *Research in microbiology*, 152(10):901–905, 2001.

[37] Daniel G Gibson, Lei Young, Ray-Yuan Chuang, J Craig Venter, Clyde A Hutchison, and Hamilton O Smith. Enzymatic assembly of dna molecules up to several hundred kilobases. *Nature methods*, 6(5):343–345, 2009.

[38] LYNNE Gilson, Hare Khrisna Mahanty, and R Kolter. Four plasmid genes are required for colicin v synthesis, export, and immunity. *Journal of bacteriology*, 169(6):2466–2470, 1987.

[39] Oscar Godoy, Daniel B Stouffer, Nathan JB Kraft, and Jonathan M Levine. Intransitivity is infrequent and fails to promote annual plant coexistence without pairwise niche differences. *Ecology*, 98(5):1193–1200, 2017.

[40] Alison I Graham, Stuart Hunt, Sarah L Stokes, Neil Bramall, Josephine Bunch, Alan G Cox, Cameron W McLeod, and Robert K Poole. Severe zinc depletion of escherichia coli: roles for high affinity zinc binding by zint, zinc transport and zinc-independent proteins. *Journal of Biological Chemistry*, 284(27):18377–18389, 2009.

[41] 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*, 117(6):3301–3306, 2020.

[42] Antonina Hafner, Martha L Bulyk, Ashwini Jambhekar, and Galit Lahav. The multiple mechanisms that regulate p53 activity and cell fate. *Nature reviews. Molecular cell biology*, 20:199–210, 2019.

[43] Christine L Hagan, Seokhee Kim, and Daniel Kahne. Reconstitution of outer membrane protein assembly from purified components. *Science*, 328(5980):890–892, 2010.

[44] Nan Hao, Bogdan A Budnik, Jeremy Gunawardena, and Erin K O'Shea. Tunable signal processing through modular control of transcription factor translocation. *Science*, 339(6118):460–464, 2013.

[45] Michael P Hassell, Hugh N Comins, and Robert M May. Species coexistence and self-

organizing spatial dynamics. *Nature*, 370(6487):290–292, 1994.

[46] Qian He, Mauro Mobilia, and Uwe C Täuber. Spatial rock-paper-scissors models with inhomogeneous reaction rates. *Physical Review E*, 82(5):051909, 2010.

[47] Stefan Hennig, Gerhard Rödel, and Kai Ostermann. Artificial cell-cell communication as an emerging tool in synthetic biology applications. *Journal of biological engineering*, 9(1):1–12, 2015.

[48] Michael E Hibbing, Clay Fuqua, Matthew R Parsek, and S Brook Peterson. Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, 8(1):15–25, 2010.

[49] John Warren Huntley and Michał Kowalewski. Strong coupling of predation intensity and diversity in the phanerozoic fossil record. *Proceedings of the National Academy of Sciences*, 104(38):15006–15010, 2007.

[50] Curtis Huttenhower, Dirk Gevers, Rob Knight, Sahar Abubucker, Jonathan H Badger, Asif T Chinwalla, Heather H Creasy, Ashlee M Earl, Michael G FitzGerald, Robert S Fulton, et al. Structure, function and diversity of the healthy human microbiome. *nature*, 486(7402):207, 2012.

[51] In Young Hwang, Elvin Koh, Adison Wong, John C March, William E Bentley, Yung Seng Lee, and Matthew Wook Chang. Engineered probiotic escherichia coli can eliminate and prevent pseudomonas aeruginosa gut infection in animal models. *Nature communications*, 8:15028, 2017.

[52] Vincent M Isabella, Binh N Ha, Mary Joan Castillo, David J Lubkowicz, Sarah E Rowe, Yves A Millet, Cami L Anderson, Ning Li, Adam B Fisher, Kip A West, et al. Development of a synthetic live bacterial therapeutic for the human metabolic disease phenylketonuria. *Nature biotechnology*, 2018.

[53] Premkumar Jayaraman, Maciej B Holowko, Jing Wui Yeoh, Sierin Lim, and Chueh Loo Poh. Repurposing a two-component system-based biosensor for the killing of vibrio cholerae. *ACS synthetic biology*, 6(7):1403–1415, 2017.

[54] Luo-Luo Jiang, Tao Zhou, Matjaž Perc, and Bing-Hong Wang. Effects of competition on pattern formation in the rock-paper-scissors game. *Physical Review E*, 84(2):021912, 2011.

[55] Andrew R Joyce and Bernhard Ø Palsson. The model organism as a system: integrating'omics' data sets. *Nature reviews Molecular cell biology*, 7(3):198–210, 2006.

[56] Marta C Justino, Cláudia C Almeida, Miguel Teixeira, and Lígia M Saraiva. Escherichia

coli di-iron ytfe protein is necessary for the repair of stress-damaged iron-sulfur clusters. *Journal of Biological Chemistry*, 282(14):10352–10359, 2007.

[57] Sun-Young Kang, Jae Kyoung Lee, Jae-Hyuk Jang, Bang Yeon Hwang, and Young-Soo Hong. Production of phenylacetyl-homoserine lactone analogs by artificial biosynthetic pathway in escherichia coli. *Microbial cell factories*, 14(1):191, 2015.

[58] Meena Kapahi and Sarita Sachdeva. Bioremediation options for heavy metal pollution. *Journal of Health and Pollution*, 9(24), 2019.

[59] Benjamin Kerr, Margaret A Riley, Marcus W Feldman, and Brendan JM Bohannan. Local dispersal promotes biodiversity in a real-life game of rock–paper–scissors. *Nature*, 418(6894):171–174, 2002.

[60] Benjamin C Kirkup and Margaret A Riley. Antibiotic-mediated antagonism leads to a bacterial game of rock–paper–scissors in vivo. *Nature*, 428(6981):412–414, 2004.

[61] Jan Kosinski, Marcin Feder, and Janusz M Bujnicki. The pd-(d/e) xk superfamily revisited: identification of new members among proteins involved in dna metabolism and functional predictions for domains of (hitherto) unknown function. *BMC bioinformatics*, 6(1):1–13, 2005.

[62] NAV Krishna, MD Nadeem, MP Saradhi, B Mahendran, and S Bharathi. Cumulative activity of the p-coumaric acid and syringaldehyde for antimicrobial activity of different microbial strains. *Euro. J. Exp. Biol*, 4:40–43, 2014.

[63] Tatiana Kulakovskaya. Inorganic polyphosphates and heavy metal resistance in microorganisms. *World Journal of Microbiology and Biotechnology*, 34(9):1–8, 2018.

[64] Nicolas Kylilis, Zoltan A Tuza, Guy-Bart Stan, and Karen M Polizzi. Tools for engineering coordinated system behaviour in synthetic microbial consortia. *Nature communications*, 9(1):2677, 2018.

[65] Chong Lee and Kenneth J Marians. Characterization of the nucleoid-associated protein yejk. *Journal of Biological Chemistry*, 288(44):31503–31516, 2013.

[66] Diana Lee and Carl J Douglas. Two divergent members of a tobacco 4-coumarate: coenzyme a ligase (4cl) gene family (cdna structure, gene inheritance and expression, and properties of recombinant proteins). *Plant Physiology*, 112(1):193–205, 1996.

[67] Jonathan M Levine, Jordi Bascompte, Peter B Adler, and Stefano Allesina. Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, 546(7656):56–64, 2017.

[68] Jonathan M Levine and Janneke HilleRisLambers. The importance of niches for the maintenance of species diversity. *Nature*, 461(7261):254–257, 2009.

[69] Roie Levy and Elhanan Borenstein. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences*, 110(31):12804–12809, 2013.

[70] Jianghui Li, Xiaojun Ren, Bingqian Fan, Zhaoyang Huang, Wu Wang, Huaibin Zhou, Zhefeng Lou, Huangen Ding, Jianxin Lyu, and Guoqiang Tan. Zinc toxicity and iron-sulfur cluster biogenesis in escherichia coli. *Applied and environmental microbiology*, 85(9):e01967–18, 2019.

[71] Michael J Liao, M Omar Din, Lev Tsimring, and Jeff Hasty. Rock-paper-scissors: Engineered population dynamics increase genetic stability. *Science*, 365(6457):1045–1049, 2019.

[72] Ferdinand CO Los, Tara M Randis, Raffi V Aroian, and Adam J Ratner. Role of pore-forming toxins in bacterial infectious diseases. *Microbiology and Molecular Biology Reviews*, 77(2):173–207, 2013.

[73] Catherine A Lozupone, Jesse I Stombaugh, Jeffrey I Gordon, Janet K Jansson, and Rob Knight. Diversity, stability and resilience of the human gut microbiota. *Nature*, 489(7415):220–230, 2012.

[74] Derek S Lundberg, Sarah L Lebeis, Sur Herrera Paredes, Scott Yourstone, Jase Gehring, Stephanie Malfatti, Julien Tremblay, Anna Engelbrektson, Victor Kunin, Tijana Glavina Del Rio, et al. Defining the core arabidopsis thaliana root microbiome. *Nature*, 488(7409):86–90, 2012.

[75] Hadeel Majeed, Adam Lampert, Lusine Ghazaryan, and Osnat Gillor. The weak shall inherit: bacteriocin-mediated interactions in bacterial populations. *PLoS one*, 8(5), 2013.

[76] Ning Mao, Andres Cubillos-Ruiz, D Ewen Cameron, and James J Collins. Probiotic strains detect and suppress cholera in mice. *Science translational medicine*, 10(445):eaao2586, 2018.

[77] John Luke McConn, Cameron R Lamoureux, Saugat Poudel, Bernhard O Palsson, and Anand V Sastry. Optimal dimensionality selection for independent component analysis of transcriptomic data. *BMC bioinformatics*, 22(1):1–13, 2021.

[78] Edward A Meighen. Genetics of bacterial bioluminescence. *Annual review of genetics*, 28(1):117–139, 1994.

[79] J Menezes, B Moura, and TA Pereira. Uneven rock-paper-scissors models: Patterns and

coexistence. *EPL (Europhysics Letters)*, 126(1):18003, 2019.

[80] Hans Metz. *The geometry of ecological interactions: simplifying spatial complexity*. Cambridge University Press, 2000.

[81] Takahiko Mitani, Kana Ota, Nobuya Inaba, Kunihiro Kishida, and Hajime A Koyama. Antimicrobial activity of the phenolic compounds of prunus mume against enterobacteria. *Biological and Pharmaceutical Bulletin*, 41(2):208–212, 2018.

[82] Angela M Mitchell and Thomas J Silhavy. Envelope stress responses: balancing damage repair and toxicity. *Nature Reviews Microbiology*, 17(7):417–428, 2019.

[83] Fernando Monje-Casas, Juan Jurado, Marıa-José Prieto-Álamo, Arne Holmgren, and Carmen Pueyo. Expression analysis of the nrdhief operon fromescherichia coli: Conditions that trigger the transcript level in vivo. *Journal of Biological Chemistry*, 276(21):18031–18037, 2001.

[84] Lisbeth Mortensen, Gert Dandanell, and Karin Hammer. Purification and characterization of the deor repressor of escherichia coli. *The EMBO Journal*, 8(1):325–331, 1989.

[85] Ghazala Mustafa and Setsuko Komatsu. Toxicity of heavy metals and metal-containing nanoparticles on plants. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1864(8):932–944, 2016.

[86] Dietrich H Nies and Simon Silver. *Molecular microbiology of heavy metals*, volume 6. Springer Science & Business Media, 2007.

[87] Shane Thomas O'Donnell, R Paul Ross, and Catherine Stanton. The progress of multi-omics technologies: determining function in lactic acid bacteria using a systems level approach. *Frontiers in microbiology*, 10:3084, 2020.

[88] Ekaterina M Panina, Andrey A Mironov, and Mikhail S Gelfand. Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proceedings of the National Academy of Sciences*, 100(17):9912–9917, 2003.

[89] Danielle B Pedrolli, Nathan V Ribeiro, Patrick N Squizato, Victor N de Jesus, Daniel A Cozetto, Team AQA Unesp at iGEM, et al. Engineering microbial living therapeutics: the synthetic biology toolbox. *Trends in biotechnology*, 2018.

[90] Jiao Peng, Ting-ting Zheng, Yue Liang, Li-fang Duan, Yao-dong Zhang, Li-Jun Wang, Guang-ming He, and Hai-tao Xiao. p-coumaric acid protects human lens epithelial cells against oxidative stress-induced apoptosis by mapk signaling. *Oxidative medicine and cellular longevity*, 2018, 2018.

[91] Dietmar H Pieper and Walter Reineke. Engineering bacteria for bioremediation. *Current opinion in biotechnology*, 11(3):262–270, 2000.

[92] 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, 2012.

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

[94] MANIDIPA Raha, I Kawagishi, V Müller, M Kihara, and RM Macnab. Escherichia coli produces a cytoplasmic alpha-amylase, amya. *Journal of bacteriology*, 174(20):6644–6652, 1992.

[95] Seth Rakoff-Nahoum, Kevin R Foster, and Laurie E Comstock. The evolution of cooperation within the gut microbiota. *Nature*, 533(7602):255–259, 2016.

[96] Tobias Reichenbach, Mauro Mobilia, and Erwin Frey. Mobility promotes and jeopardizes biodiversity in rock–paper–scissors games. *Nature*, 448(7157):1046–1049, 2007.

[97] Steven M Reppert and David R Weaver. Coordination of circadian timing in mammals. *Nature*, 418(6901):935, 2002.

[98] David T Riglar and Pamela A Silver. Engineering bacteria for diagnostic and therapeutic applications. *Nature Reviews Microbiology*, 16(4):214, 2018.

[99] Margaret A Riley and David M Gordon. The ecological role of bacteriocins in bacterial competition. *Trends in microbiology*, 7(3):129–133, 1999.

[100] Timothy J Rudge, Paul J Steiner, Andrew Phillips, and Jim Haseloff. Computational modeling of synthetic microbial biofilms. *ACS Synthetic Biology*, 1(8):345–352, 2012.

[101] Kevin Rychel, Anand V Sastry, and Bernhard O Palsson. Machine learning uncovers independently regulated modules in the bacillus subtilis transcriptome. *Nature communications*, 11(1):1–10, 2020.

[102] Dana Saadat and David HT Harrison. Identification of catalytic bases in the active site of escherichia coli methylglyoxal synthase: cloning, expression, and functional characterization of conserved aspartic acid residues. *Biochemistry*, 37(28):10074–10086, 1998.

[103] Rajib Saha, Anupam Chowdhury, and Costas D Maranas. Recent advances in the reconstruction of metabolic models and integration of omics data. *Current opinion in biotechnology*, 29:39–45, 2014.

[104] Anand V Sastry, Ye Gao, Richard Szubin, Ying Hefner, Sibei Xu, Donghyuk Kim, Kumari Sonal Choudhary, Laurence Yang, Zachary A King, and Bernhard O Palsson. The escherichia coli transcriptome mostly consists of independently regulated modules. *Nature communications*, 10(1):1–14, 2019.

[105] Anand V Sastry, Alyssa Hu, David Heckmann, Saugat Poudel, Erol Kavvas, and Bernhard O Palsson. Independent component analysis recovers consistent regulatory signals from disparate datasets. *PLoS computational biology*, 17(2):e1008647, 2021.

[106] Amy L Schaefer, EP Greenberg, Colin M Oliver, Yasuhiro Oda, Jean J Huang, Gili Bittan-Banin, Caroline M Peres, Silke Schmidt, Katarina Juhaszova, Janice R Sufrin, et al. A new class of homoserine lactone quorum-sensing signals. *Nature*, 454(7204):595, 2008.

[107] HERBERT Schweizer, THOMAS Grussenmeyer, and WINFRIED Boos. Mapping of two ugp genes coding for the pho regulon-dependent sn-glycerol-3-phosphate transport system of escherichia coli. *Journal of bacteriology*, 150(3):1164–1171, 1982.

[108] Spencer R Scott, M Omar Din, Philip Bittihn, Liyang Xiong, Lev S Tsimring, and Jeff Hasty. A stabilized microbial ecosystem of self-limiting bacteria using synthetic quorum-regulated lysis. *Nature microbiology*, 2(8):17083, 2017.

[109] Spencer R Scott and Jeff Hasty. Quorum sensing communication modules for microbial consortia. *ACS synthetic biology*, 5(9):969–977, 2016.

[110] Jasmine Shong, Yao-Ming Huang, Christopher Bystroff, and Cynthia H Collins. Directed evolution of the quorum-sensing regulator esar for increased signal sensitivity. *ACS chemical biology*, 8(4):789–795, 2013.

[111] Barry Sinervo and Curt M Lively. The rock–paper–scissors game and the evolution of alternative male strategies. *Nature*, 380(6571):240–243, 1996.

[112] Santiago Soliveres, Fernando T Maestre, Werner Ulrich, Peter Manning, Steffen Boch, Matthew A Bowker, Daniel Prati, Manuel Delgado-Baquerizo, José L Quero, Ingo Schöning, et al. Intransitive competition is widespread in plant communities and maintains their species richness. *Ecology letters*, 18(8):790–798, 2015.

[113] Hao Song, Stephen Payne, Meagan Gray, and Lingchong You. Spatiotemporal modulation of biodiversity in a synthetic chemical-mediated ecosystem. *Nature chemical biology*, 5(12):929, 2009.

[114] Attila Szolnoki, Mauro Mobilia, Luo-Luo Jiang, Bartosz Szczesny, Alastair M Rucklidge, and Matjaž Perc. Cyclic dominance in evolutionary games: a review. *Journal of the Royal Society Interface*, 11(100):20140735, 2014.

[115] Justin Tan, Anand V Sastry, Karoline S Fremming, Sara P Bjørn, Alexandra Hoffmeyer, Sangwoo Seo, Bjørn G Voldborg, and Bernhard O Palsson. Independent component analysis of e. coli's transcriptome reveals the cellular processes that respond to heterologous gene expression. *Metabolic Engineering*, 61:360–368, 2020.

[116] Natalia Pasternak Taschner, Ezra Yagil, and Beny Spira. A differential effect of σs on the expression of the pho regulon genes of escherichia coli. *Microbiology*, 150(9):2985–2992, 2004.

[117] James M Tiedje. Ecology of denitrification and dissimilatory nitrate reduction to ammonium. *Biology of anaerobic microorganisms*, pages 179–244, 1988.

[118] Thomas R Turner, Euan K James, and Philip S Poole. The plant microbiome. *Genome biology*, 14(6):209, 2013.

[119] Adrian J Verster, Benjamin D Ross, Matthew C Radey, Yiqiao Bao, Andrew L Goodman, Joseph D Mougous, and Elhanan Borenstein. The landscape of type vi secretion across human gut microbiomes reveals its role in community composition. *Cell host & microbe*, 22(3):411–419, 2017.

[120] Waldemar Vollmer, Moritz von Rechenberg, and Joachim-Volker Höltje. Demonstration of molecular interactions between the murein polymerase pbp1b, the lytic transglycosylase mlta, and the scaffolding protein mipa of escherichia coli. *Journal of Biological Chemistry*, 274(10):6726–6734, 1999.

[121] Anyou Wang and David E Crowley. Global gene expression responses to cadmium toxicity in escherichia coli. *Journal of bacteriology*, 187(9):3259–3266, 2005.

[122] Stuart A West, Ashleigh S Griffin, Andy Gardner, and Stephen P Diggle. Social evolution theory for microorganisms. *Nature reviews microbiology*, 4(8):597–607, 2006.

[123] Yilin Wu, Yi Jiang, A Dale Kaiser, and Mark Alber. Self-organization in bacterial swarming: lessons from myxobacteria. *Physical biology*, 8(5):055003, 2011.

[124] Na Xie, Lu Zhang, Wei Gao, Canhua Huang, Peter Ernst Huber, Xiaobo Zhou, Changlong Li, Guobo Shen, and Bingwen Zou. Nad+ metabolism: pathophysiologic mechanisms and therapeutic potential. *Signal transduction and targeted therapy*, 5(1):1–37, 2020.

[125] Yuan Xue, Jossef Osborn, Anand Panchal, and Jay L Mellies. The rpoe stress response pathway mediates reduction of the virulence of enteropathogenic escherichia coli by zinc. *Applied and environmental microbiology*, 81(11):3766–3774, 2015.

[126] Ryo Yamada, Daigo Okada, Juan Wang, Tapati Basak, and Satoshi Koyama. Interpretation of omics data analyses. *Journal of human genetics*, 66(1):93–102, 2021.

[127] Chi Yang, Tzu-Wen Huang, Shiau-Yi Wen, Chun-Yang Chang, Shih-Feng Tsai, Whei-Fen Wu, and Chuan-Hsiung Chang. Genome-wide phob binding and gene expression profiles reveal the hierarchical gene regulatory network of phosphate starvation in escherichia coli. *PloS one*, 7(10):e47314, 2012.

[128] Lingchong You, Robert Sidney Cox III, Ron Weiss, and Frances H Arnold. Programmed population control by cell–cell communication and regulated killing. *Nature*, 428(6985):868, 2004.

[129] Dapeng Zhang, Robson F de Souza, Vivek Anantharaman, Lakshminarayan M Iyer, and L Aravind. Polymorphic toxin systems: comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics. *Biology direct*, 7(1):18, 2012.

[130] Pengfei Zhang, Zhicang Ye, Chen Ye, Haijie Zou, Zhichao Gao, and Jianyi Pan. Ompw is positively regulated by iron via fur, and negatively regulated by soxs contribution to oxidative stress resistance in escherichia coli. *Microbial pathogenesis*, 138:103808, 2020.