2022

# Development of methods to diagnose and predict antibiotic resistance using synthetic biology and computational approaches

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

AND

COLLEGE OF ENGINEERING

Dissertation

# DEVELOPMENT OF METHODS TO DIAGNOSE AND PREDICT ANTIBIOTIC RESISTANCE USING SYNTHETIC BIOLOGY AND COMPUTATIONAL APPROACHES

by

**EMMA ANN BRIARS**

BA, Boston University, 2013

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2022

Approved by

First Reader _____

Ahmad Khalil, PhD
Associate Professor of Biomedical Engineering

Second Reader _____

Daniel Segrè, PhD
Professor of Biology
Professor of Bioinformatics
Professor of Biomedical Engineering
Professor of Physics

## ACKNOWLEDGMENTS

This work would not have been possible without amazing support systems both in and out of graduate school. Finishing up a PhD during a global pandemic with lock downs and lab shifts was not what I would have hoped for, but it was made possible by their support.

First, I would like to thank my advisor, Mo Khalil, for his mentorship and support. Thank you for fostering such a uniquely positive work environment and for always being a source of encouragement and optimism. I would also like to thank Daniel Segrè for giving me my start in bioinformatics back in 2008 as a UROP student, and for supporting me since then. I am also thankful for my additional commitee members: Cathie Klapperich and Mary Dunlop for their expertise and insightful questions.

Next, I would like to thank all of my lab mates, past and present for their individual expertise, thoughtful discussions, and camaraderie over the past few years: James Angstman, Zehua Bao, Abdul Bhuiya, Nate Borders, Dana Braff, Meghan Bragdon, Guilio Chiesa, Allison Drain, Keith Gagnon, Marco Galardini, Sam Gan, Daniel Hart, Zack Heins, Divya Israni, Sungho Jang, Charles Jo, Szilvia Kiriakov, Heidi Klumpe, Colin Kunze, Hui-Shan Li, Chris Mancuso, Hagar Moussa, Nikit Patel, Minhee Park, Arjun Ravikumar, Adam Sanford, Will Shaw, Brandon Wong, Ezira Yimer Wolle, Belle Ye. A special thank you to Marco Galardini for his work on implementing the morbidostat in the eVOLVER, as well as for joining my antibiotic resistance evolution journal club during the pandemic. Another special thanks to Dana Braff for letting me rotate with her as a green synthetic biologist, and for teaching me the basics of toehold sensors and paper-based reactions. I would also like to thank Chris Mancuso, Colin Kunze, Szilvia Kiriakov, and Brandon Wong for

their assistance with various projects over the years. And a final acknowledgement to the Bay of Blankets, for always putting comfort first in a temperature-controlled shared open office space.

I would like to thank my colleagues and friends from the Bioinformatics program and the BDC. A special thank you to Sebastian Carrasco Pro for being my first-year Challenge Project partner, as well as for teaching me how to play squash. Thank you to Dakota Hawkins, Will Hackett, David Jenkins, Xingyi Shi, Sean Corbett, Dave Bray, Alan Pacheco, Lucas Schiffer, Shruthi Bandyadka, Dileep Kishore, Emily Speranza, and Jamie Strampe for combinations of: being great friends, working on the Student organized symposium with me, terrace house watch parties, and being leaders I look up to. A huge thank you to Breanna O'Reilly for being an amazing teammate, for conceptualizing BDC SPIN, and for bringing me along to co-found it. Thank you David McIntyre, Jeremy Tran, and Diana Arguijo Mendoza for jumping onto the BDC SPIN team and being such a joy to work with. I am so proud of the work we have done, and am so excited for this resource that we have created for future members of the BDC. A final shout out to the members of the BDC running club for bringing the fun to fun runs.

Thank you the amazing administrative and support staff that I have worked with in the Khalil Lab, the Biological Design Center, and in the Bioinformatics Program both past and present: David Michaels, Valeria Claret, Kevin Gonzalez, Johanna Vasquez, Caroline Lyman, David King, Mary-Ellen Gipson-FitzPatrick. Thank you to Jess Tytell, Akshata Sonni, and Josh Finkelstein for their past and present support of the BDC SPIN.

To my dearest friends, Georgia, Danny, Jaclyn, Kate, Ashley: I am in debt for your constant encouragement, motivation, and enthusiasm. To my family, Mom,

Dad, Megen, Papa Rock, Mama Rock, Kelsey, Craig, Tommy, Ashley, T, E, Brooks: thank you for your never-ending support. To my husband Pete, thank you for being my best friend and constant throughout this PhD. And to Lilo, thank you for being you.

# DEVELOPMENT OF METHODS TO DIAGNOSE AND PREDICT ANTIBIOTIC RESISTANCE USING SYNTHETIC BIOLOGY AND COMPUTATIONAL APPROACHES

**EMMA ANN BRIARS**

Boston University, Graduate School of Arts and Sciences and College of Engineering, 2022

Major Professor: Ahmad Khalil, Associate Professor of Biomedical Engineering

## ABSTRACT

Antibiotic resistance is a quickly emerging public health crisis, accounting for more than 700,000 annual global deaths. Global human antibiotic overuse and misuse has significantly expedited the rate at which bacteria become resistant to antibiotics. A renewed focus on discovering new antibiotics is one approach to addressing this crisis. However, it alone cannot solve the problem: historically, the introduction of a new antibiotic has consistently, and at times rapidly, been followed by the appearance and dissemination of resistant bacteria. It is thus crucial to develop strategies to improve how we select and deploy antibiotics so that we can control and prevent the emergence and transmission of antibiotic resistance.

Current gold-standard antibiotic susceptibility tests measure bacterial growth, which can take up to 72 hours. However, bacteria exhibit more immediate measurable phenotypes of antibiotic susceptibility, including changes in transcription, after brief antibiotic exposure. In this dissertation I develop a framework for build-

ing a paper-based cell-free toehold sensor antibiotic susceptibility test that can detect differential mRNA expression. I also explore how long-term lab evolution experiments can be used to prospectively uncover transcriptional signatures of antibiotic susceptibility.

Paper-based cell-free systems provide an opportunity for developing clinically tractable nucleic-acid based diagnostics that are low-cost, rapid, and sensitive. I develop a computational workflow to rapidly and easily design toehold switch sensors, amplification primers, and synthetic RNAs. I develop an experimental workflow, based on existing paper-based cell-free technology, for screening toehold sensors, amplifying bacterial mRNA, and deploying sensors for differential mRNA detection. I combine this work to introduce a paper-based cell-free toehold sensor antibiotic susceptibility test that can detect fluoroquinolone-susceptible *E. coli*. Next, I describe a methodology for long-term lab evolution and how it can be used to explore the relationship between a phenotype, such as gene expression, and antibiotic resistance acquisition. Using a set of *E. coli* strains evolved to acquire tetracycline resistance, I explore how each strain's transcriptome changes as resistance increases. Together, this work provides a set of computational and experimental methods that can be used to study the emergence of antibiotic resistance, and improve upon available methods for properly selecting and deploying antibiotics.

## CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| AR | antimicrobial resistance |
| AST | antibiotic susceptibility testing |
| bp | base pair |
| CARD | Comprehensive Antibiotic Resistance Database |
| Cipro | ciprofloxacin |
| CLSI | Clinical & Laboratory Standards Institute |
| CRISPR | clustered regularly interspaced short palindromic repeats |
| Cas | CRISPR-associated proteins |
| *E. coli* | Escherichia coli |
| FC | fold change |
| GFP | green fluorescent protein |
| GO | gene ontology |
| HGT | horizontal gene transfer |
| IDT | Integrated DNA Technologies |
| LAMP | Loop-mediated isothermal amplification |
| MDR | multi-drug resistance |
| MHB | Mueller Hinton Broth |
| MIC | minimum inhibitory concentration |
| NASBA | Nucleic acid sequence-based amplification |
| nt | nucleotide |
| NT | No Treatment |
| OD | optical density |
| PCR | polymerase chain reaction |
| POC | point-of-care |
| RBS | ribosomal binding site |

| | |
|---|---|
| RNA-Seq | RNA-Sequencing |
| RPA | Recombinase polymerase activation |
| RT-qPCR | Reverse transcriptase quantitative PCR |
| PCA | principal component analysis |
| PCR | polymerase chain reaction |
| synRNA | synthetic RNA |
| TE | 1X TE buffer (pH 8.0) |
| TTR | time to response |
| TPM | transcripts per million |
| UTI | urinary tract infection |

# CHAPTER 1
# Introduction

## 1.1 THE ANTIBIOTIC RESISTANCE CRISIS

Antibiotics are a critically important tool in modern healthcare. However, increasing rates of antibiotic resistance pose a real threat on our ability to rely on antibiotics for many aspects of modern medicine, including: surgery, immunosupressive therapy, and the treatment of sepsis and other bacterial infections. Antibiotic use is directly linked to the acquisition of antibiotic resistance, so as we continue to use antibiotics, we will continue to see naturally susceptible bacteria achieve resistance, either through mutation or the transfer of genetic material (MacGowan & Macnaughton (2013)).

Although 1% of bacteria are naturally resistant to new antibiotics (Levine et al. (2006)), the staggering rates at which we are encountering antibiotic resistant infections – up to 95% in some bug-drug combinations – is a global cause for concern. Today, it is estimated that there are more than 700,000 annual global deaths due to antibiotic resistant infections (O'Neill (2020)), and in some countries, up to 90% of *E. coli* infections are resistant to ciprofloxacin (WHO (2020)).

Antibiotic overuse and misuse are major contributors to increased rates of antibiotic resistance (Bronzwaer et al. (2002), Goossens et al. (2005)). In human medicine alone, global antibiotic use increased 40% between 2000-2010 (Van Boeckel et al. (2014)). This, combined with antibiotic use in agriculture, unrestricted antibiotic access in some countries, increased global travel, and decreasing rates of new antibiotic discovery, have contributed to the global rise of antibiotic resistant infections (O'Neill (2020)). More simply, putting antibiotics into the hands of humans

has significantly expedited the rate at which bacteria become resistant to antibiotics.

The WHO has highlighted various scientific, medical, and public health interventions that are needed in order to reduce the increasing rate of antimicrobial infections. These include: promoting the research and development of new antibiotics, implementing best practices for antibiotic use in agriculture, encouraging best practices of antibiotic use among both the general public and healthcare workers, and increasing public awareness of the antibiotic resistance crisis (WHO (2021)). In order for antibiotic best practices to be advocated for by the general public, and implemented by healthcare workers, these groups need access to both diverse sets of antibiotics as well as diagnostic tools. Diagnostics are a crucial implement as they can expedite treatment, enable the use of narrower spectrum antibiotics, and in some cases, reduce the need for antibiotics. As such, there is a critical need for new diagnostics that can identify the bacterial species present in an infection, as well as tools that can assess the antibiotic susceptibility of the infection.

## 1.2 ANTIBIOTIC SUSCEPTIBILITY TESTING

Correctly selecting an antibiotic based on the antibiotic susceptibility profile of a bacterial infection allows clinicians to use targeted antibiotics when appropriate, and preserve stronger, broad spectrum antibiotics for when they are absolutely necessary. Moreover, improved antibiotic stewardship efforts have been linked to improved clinical outcomes, including a reduction in incidence of multi-drug resistant infections (Baur et al. (2017), Karanika et al. (2016)). However, without widespread access to rapid diagnostics that can quickly assess the susceptibility of

an unknown pathogen to a potential antibiotic, physicians must use empirical evidence to determine an initial antibiotic course (Figure 1.1). Unfortunately, relying on empirical evidence leads to frequent antibiotic overuse and misuse, and in some cases, increased mortality and hospital stays (Kollef et al. (2021)). For example, a recent study found that only 20.6% of isolated bacterial cultures from the National Cancer Institute in Sri Lanka were susceptible to the empirical antibiotic selected (Chathuranga et al. (2021)). As rates of antibiotic resistance increase, it is essential that clinicians possess rapid tools for assessing the antibiotic susceptibility of potential bacterial infections.

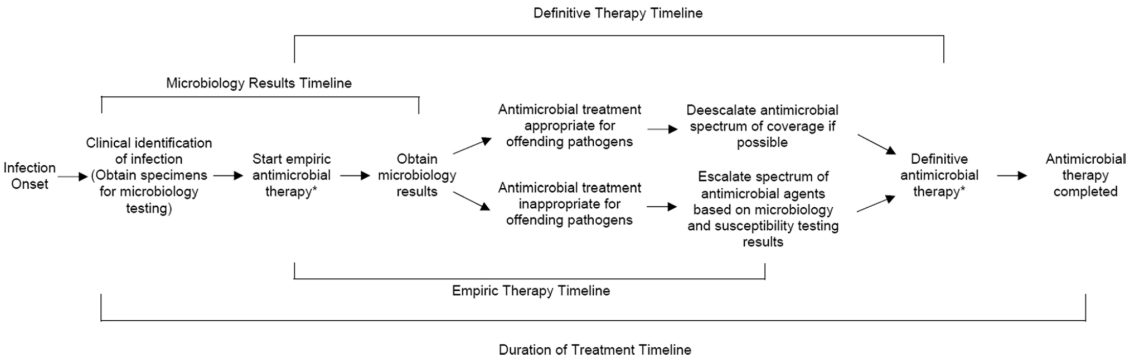



**Figure 1.1:** Overview of the timeline of clinical decision points and antibiotic usage for critically-ill patients

Figure adapted from Kollef et al. (2021). Without rapid diagnostics, the empiric therapy is not always the same as the definitive therapy, leading to unnecessary antibiotic exposures and increased treatment time.

### 1.2.1 Gold-Standard Methods for Antibiotic Susceptibility Testing

The gold-standard methods for clinical antibiotic susceptibility testing (AST) are growth-based phenotypic methods, which assess the ability of an antibiotic to inhibit bacterial growth *in vitro*. These include the disk diffusion assay, which mea-

sures the zone of bacterial growth inhibition around an antibiotic disk; and the broth (or agar) dilution assays, which measures the minimum inhibitory concentration (MIC) of an antibiotic through bacterial culture in the presence of serial dilutions of an antibiotic in media (Arena et al. (2015)). The results of these methods are interpreted as clinical calls of susceptibility or resistance via comparison to the clinical breakpoints set by the Clinical and Laboratory Standards Institute (CLSI) or the European Committee on Antimicrobial Susceptibility Testing (EUCAST). Recently, work has been done to modernize these assays, including the development of microfluidic platforms to streamline assay set-up (Lee et al. (2017), Osaid et al. (2021)); automate reading the results to reduce time to response (TTR) (Hombach et al. (2017)); and democratize the process of interpreting AST results (Fowler et al. (2021)).

Although the gold-standard AST methods are relatively cost-effective and accurate, and despite the highlighted innovations made to streamline these methods, they are still limited by their TTR. Since these methods rely on bacterial growth from a single bacteria colony as the readout, they can take at least 48-72 before a clinical assessment can be made. However, this timescale is days slower than what is needed in a point-of-care setting (Figure 1.1). Thus, AST methods are not always implemented, and even when they are, clinicians still need to make an empirical decision on which antibiotic to use while waiting for the results, creating a large opportunity for antibiotic overuse and misuse.

In the following two subsections I describe alternative molecular approaches to the gold-standard antibiotic susceptibility tests.

### 1.2.2 Genotypic Methods for Antibiotic Susceptibility Testing

Genotypic approaches to AST detect specific genetic elements that are known to confer resistance, include preexisting genomic or proteomic signatures of antibiotic resistance (van Belkum et al. (2020)). This group of molecular approaches uses a one-to-one relationship between genetic element and resistance phenotype for an antibiotic. For some bug-drug combinations, this information has been previously identified and is stored databases such as Comprehensive Antibiotic Resistance Database (CARD) and MEGARes (Alcock et al. (2020), Lakin et al. (2017)). CARD contains both genomic and proteomic genetic elements that are known to confer resistance in certain bug-drug combinations. However, for most bacteria there is not just one genetic element that can be used to infer antibiotic susceptibility, but instead there are numerous genetic elements - including point mutations, plasmids, etc. - that are associated with antibiotic susceptibility for different antibiotics. For example, recent work identified - in *Mycobacterium tuberculosis* alone - 449 unique genetic determinants that increase the MICs across thirteen antibiotics (Carter et al. (2021)). Genotypic diagnostic tests require both a compendium of reference data and the ability to collect sequence information from each sample. This strategy is being developed commercially by companies such as BioFire and Day Zero Diagnostics to provide rapid, genotypic-based methods for AST (Furukawa et al. (2021), Asundi et al. (2020)).

One challenge to genotpyic AST is the dynamic nature of antibiotic resistance. As we continue to use antibiotics and place new selective pressures on bacteria, we will in turn encounter new genomic and proteomic features of antibiotic resistance. Moreover, the absence of certain antibiotic resistance genes does not predict antibiotic susceptibility (Bard & Lee (2018)). In fact, a recent study found that

for 85 isolates of *Bacillus cereus*, the presence of antibiotic resistance genes was a poor predictor of phenotypic resistance (Mills et al. (2021)). Thus, the genetic and proteomic determinants of antibiotic resistance can be both a moving target and non-deterministic, making it a challenging diagnostic approach.

### 1.2.3 Phenotypic Methods for Antibiotic Susceptibility Testing

An alternative approach to detecting antibiotic resistance is phenotypic: introducing the antibiotic to the bacterial sample and observing its response. One benefit of phenotypic AST is that there are far less phenotypes of antibiotic resistance than genotypes, *i.e.* many distinct genetic elements can all lead to the single phenotype of resistance to a specific antibiotic. Additionally, phenotypic methods generally predict antibiotic susceptibility more accurately than genotypic methods since they are directly testing for the susceptibility phenotype (van Belkum et al. (2020)). As described previously, the gold-standard method for AST is a phenotypic test that measures bacterial growth in response to antibiotic exposure. Although this method is extremely accurate, its reliance on observing cell growth is slow and relatively low-throughput.

Recently, work has been done to probe other phenotypes of antibiotic susceptibility for rapid AST. For these methods, bacteria are still exposed to an antibiotic or panel of antibiotics, but a different phenotype of the response is measured. For example, the rate of cell death can be directly linked to antibiotic susceptibility, and can be measured faster than a growth-based output (Kalashnikov et al. (2012)). Antibiotic susceptibility has also been linked to changes in cellular architecture - including cell size and shape - which can be measured faster than a growth-based output (Quach et al. (2016)). Finally, changes in transcription habe also be linked

to antibiotic susceptibility, and can be measured for rapid AST (Sangurdekar et al. (2006), Barczak et al. (2012)).

Together, these new methods demonstrate the potential for advancements in phenotpyic AST. However there is still a need to make these approaches faster, cheaper, and better-suited for point-of-care (POC) settings.

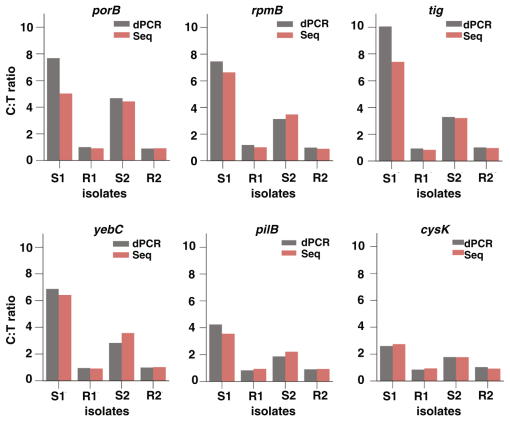
## 1.3 TRANSCRIPTOMES OF ANTIBIOTIC EXPOSURE

Whole-genome expression profiling of bacteria in response to antibiotic exposure has revealed transcriptional signatures of antibiotic exposure (Brazas & Hancock (2005)). The genes involved in these signatures can be grouped into four categories: direct effects (*e.g.*, related to the antibiotic target), indirect effects (*e.g.*, stress response, metabolic changes), secondary effects (downstream to genes targeted by antibiotic), and bystander effects (organism or antibiotic specific). Originally, these gene signatures were used for screening the effects of new antibiotic compounds. More recently, these signatures have been teased apart to provide a relationship between the antibiotic susceptibility of a bacterial strain and its transcriptional signature in response to brief antibiotic exposure.

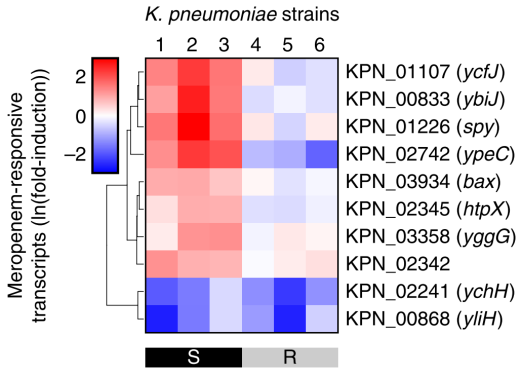
### 1.3.1 Gene Signatures of Antibiotic Susceptibility

The use of microarrays for whole-cell transcriptomic studies allowed for the discovery that susceptible bacteria differentially express stereotypical transcriptional signatures in response to brief antibiotic exposure (15 minutes – 45 minutes). In wild-type *E. coli* cells, brief exposure to norfloxacin - a quinolone with activity against DNA gyrase and topoisomerase IV - elicits changes in gene expression, including genes associated with DNA damage and SOS repair, DNA relaxation,

ATPases, transposons, and targets of FIS and FNR. However in norfloxacin resistant *E. coli* cells, brief antibiotic exposure does not lead to transcriptional changes associated with the antibiotic's effects, but instead leads to changes in cells' energy metabolism and cell division (Sangurdekar et al. (2006)). Signatures of antibiotic susceptibility have since been identified for additional "drug-bug" combinations, including *E. coli*, *P. aeruginosa*, *M. tuberculosis*, *N. gonorrhoea*, and *K. pneumoniae* (Barczak et al. (2012), Khazaei et al. (2018), Bhattacharyya et al. (2019)) (Figure 1.2).




**(a)** Figure adapted from Khazaei et al. (2018). Differential expression of 6 signatures genes in *Neisseria gonorrhoeae* in response to ciprofloxacin as measured by droplet PCR (dPCR) and RNA-sequencing (SEQ) in susceptible (S) and resistant (R) clinical isolates.




**(b)** Figure adapted from Bhattacharyya et al. (2019). Differential expression of signature genes in *Klebsiella pneumoniae* in response to meropenem as measured by a NanoString-based workflow in susceptible (S) and resistant (R) clinical isolates.

**Figure 1.2:** Examples of transcriptional signatures of antibiotic susceptibility

Unlike genotypic signatures of antibiotic susceptibility, these transcriptional signatures are conserved across many strains of one species, and recently have been shown to be conserved across multiple antibiotics within an antibiotic class (Martinsen et al. (2021)). Therefore, transcriptional signatures serve as an opportunity for RNA-based antibiotic susceptibility testing as an innovation in phenotypic

antibiotic susceptibility testing.

## 1.4 NUCLEIC ACID-BASED DIAGNOSTICS FOR RNA DETECTION

Transcriptional signatures of antibiotic susceptibility are conserved within species' and drug classes, and are quickly activated in response to brief antibiotic exposure, making them a promising opportunity for AST. The detection of transcriptional signatures for AST relies on the ability to detect differences in nucleic acids (here, RNA) between bacteria that were and were not briefly exposed to antibiotics. In this section I provide a general overview of nucleic acid-based diagnostics, followed by an introduction to both standard methods and new innovations for RNA detection.

### 1.4.1 Nucleic Acid-Based Diagnostics

Nucleic acid-based diagnostics provide sequence-level information about disease-causing pathogens and their pathogenesis (Paul et al. (2020)). Unlike other molecular diagnostic strategies such as antibody detection, nucleic acid-based diagnostics offer superior sensitivity and specificity for diagnostic applications. These diagnostic tests consists of three steps: nucleic acid extraction, amplification, and amplicon detection. These steps can be run sequentially, or collapsed into one or two total steps. Moreover, these steps can be miniaturized and multiplexed in various paper-based and microfluidic contexts (Basha et al. (2017), Connelly et al. (2015), Rodriguez et al. (2015), and many others). However, the majority of current nucleic acid-based diagnostics focus on the detection of genomic elements such as DNA, viral RNA, or 16S rRNA.

### 1.4.2 Nucleic Acid Amplification

Nucleic acid-based diagnostics frequently employ a nucleic acid amplification step. For DNA, PCR is commonly employed. For POC diagnostics, isothermal amplification techniques are preferred because they do not rely on expensive thermocyclers. Loop-mediated isothermal amplification (LAMP) is run at 65°C and can amplify DNA regions within 30 minutes using multiple sets of primer pairs (Notomi et al. (2000)). Recombinase polymerase activation (RPA) is another isothermal technique that works optimally between 37 – 42°C, requires only one primer pair, and can be as accurate and faster than PCR in amplifying DNA (Piepenburg et al. (2006)). Both LAMP and RPA can be combined with a reverse transcription (RT) step for RNA detection. Nucleic acid sequence based amplification (NASBA) is an isothermal technique that is optimized for RNA detection (Malek et al. (1994)). It works optimally at 41°C (but still has good activity at 37°C), only requires one primer pair, and is one of the only isothermal amplification techniques that preferentially amplifies RNA over DNA. The steps of a NASBA amplification reaction are shown in Figure 1.3.

### 1.4.3 Gold-Standard Methods for RNA Detection

Gold-standard laboratory methods for RNA detection include RT-qPCR and RNA-seq. For the detection of a few RNA transcripts, quantitative reverse transcription PCR (RT-qPCR) is employed (Heid et al. (1996)). In RT-qPCR, RNA is reverse transcribed into cDNA, followed by primer-mediated detection of specific cDNA molecules using quantitative PCR. For the analysis of many transcripts or the entire transcriptome, RNA-seq (RNA-sequencing) is employed (Mortazavi et al. (2008)). In RNA-seq, RNA is reverse transcribed into cDNA, and the resulting

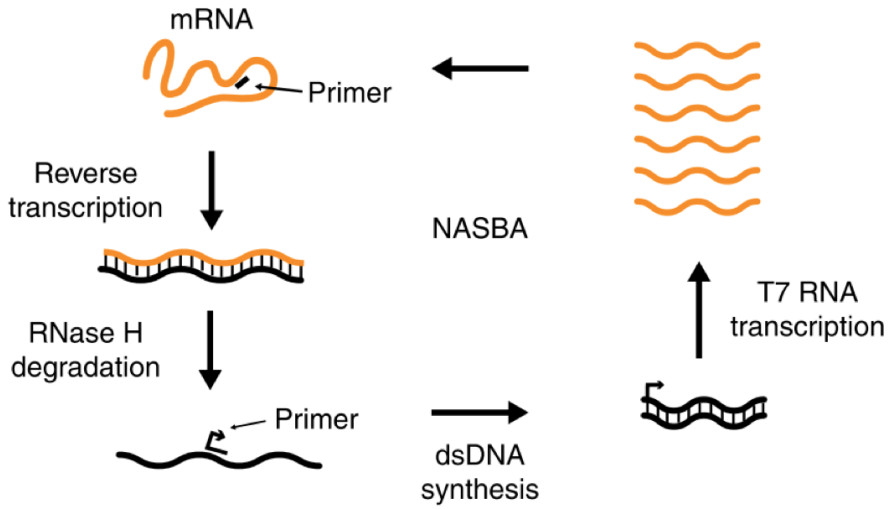



**Figure 1.3:** Overview of nucleic acid sequence-based amplification (NASBA) for isothermal RNA amplification

Figure adapted from Takahashi et al. (2018)

cDNA is chopped up, sequenced using next-generation sequencing methodologies, and the resulting data are bioinformatically translated into information about RNA abundance levels. Both of these methods can be sensitive, fast and high-throughput - making them well-suited for laboratory RNA diagnostic applications. However, they are technologically complicated, requiring expensive, non-portable instrumentation, and requiring skilled personnel. As such, these methods can not employed in many POC and resource-limited settings.

### 1.4.4 Other Emerging Methods for RNA Detection

In an effort to bring RNA-based diagnostics to POC and resource-limited settings, new methodologies for RNA detection have been developed. One hallmark of these strategies is their ability to be multiplexable, portable, and less-reliant on large, expensive equipment.

#### *1.4.4.1 CRISPR-Based RNA Detection*

CRISPR-Cas (clustered regularly interspaced short palindromic repeats; CRISPR-associated proteins) systems can be programmed for single-base specificity and detection of specific nucleic acid sequences (Li et al. (2019)). Recent characterization of different CRISPR systems has identified effector proteins with RNA-guided ribonuclease activity, such as Cas13a (Shmakov et al. (2015), Abudayyeh et al. (2016)). This led to the development of Specific High-Sensitivity Enzymatic Reporter UnLOCKing (SHERLOCK), a platform for RNA sensing based on the programmable RNA cleaving activity of Cas13a (Gootenberg et al. (2017)). RNA detection with SHERLOCK follows the standard nucleic acid detection workflow: input nucleic acid is amplified, T7-mediated reverse transcribed into RNA, followed by CRISPR collateral detection via specific CRISPR RNAs (crRNAs) and Cas13 activity. This approach allows for attamolar sensitivity and has been used for nucleotide resolution detection of Zika virus, Dengue virus, and the SARS-COV-2 (Joung et al. (2020)). Recent work has expanded upon this methodology to design a CRISPR-based broad bacterial species identification panel by designing crRNA guide RNAs for 52 unique bacterial targets of the *topA* gene (Thakku et al. (2021)). Still, these methods have been focused on genotypic detection of bacteria and viruses. Moreover, the single nucleotide specificity of CRISPR-Cas13a methods make them well-suited for species identification, but limits their generalizabilty. For example, the single nucleotide variations in the *topA* gene that make it well-suited for species identification by Cas13a, would therefore prevent that gene from being using in a pan-species gene expression test. Therefore, a molecular sensor with 'wobble' in its specificity might be more adept for generalizable gene expression-based diagnostics.

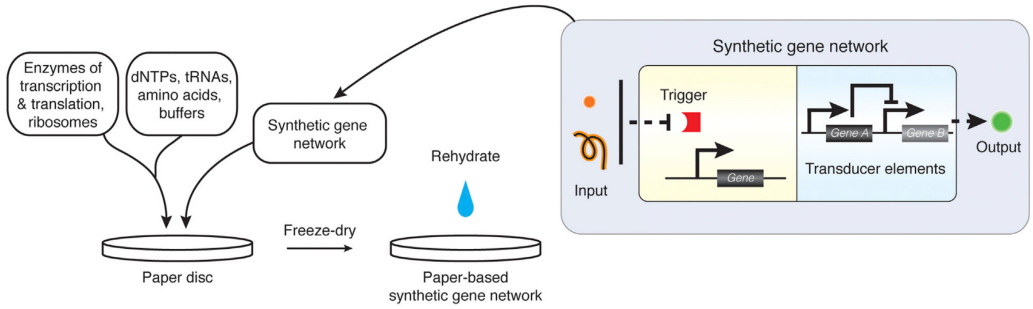
#### *1.4.4.2 Paper-Based Toehold Switch Sensor RNA Detection*

Toehold switch sensors are a programmable class of synthetic riboregulators that control the translation of a reporter gene via the binding of a trans-acting trigger RNA (Green et al. (2014)). The transcriptional and transational machinery needed to power a toehold switch sensor reaction can be accomplished using cell-free extract (Hodgman & Jewett (2012)). Briefly, cell-free systems contain the transcriptional and translational machinery required for *in vitro* protein translation. Together, cell-free extract and toehold switch sensors can be applied to paper discs and freeze-dried for a portable, ready-to-use, RNA detection platform (Pardee et al. (2014)) (Figure 1.4). Toehold switch sensors can be designed to detect virtually any nucleic acid sequence and can tolerate a few mismatches in the toehold region. Panels of toehold sensors have been developed for the rapid and portable detection of the Zika virus (Pardee et al. (2016)), norovirus (Ma et al. (2018), and different bacterial species in the microbiome (Takahashi et al. (2018)). These platforms frequently employ an isothermal RNA amplification step - such as NASBA or RPA - to increase their sensitivity. To date, these methods have focused on the detection of RNA-based genomic elements such as viral RNA or 16S rRNA.

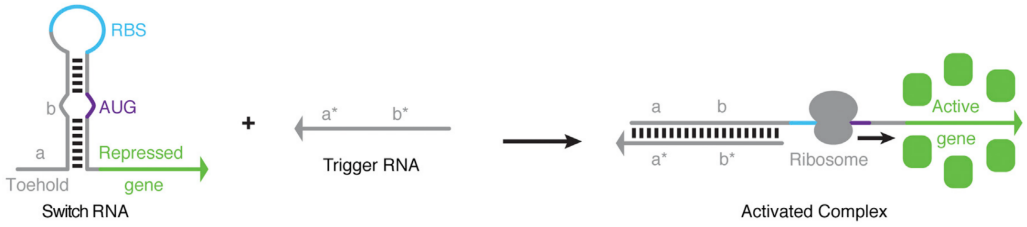
## 1.5 DISSERTATION AIMS

This dissertation aims to improve antibiotic stewardship through the development of methods to diagnose antibiotic resistance in clinically-relevant infections, as well as through the development of methods to predict antibiotic resistance in experimental settings.

In Chapters 2 and 3, I focus on developing fast, sensitive, and portable methods for mRNA detection for transcriptional phenotype-based AST. In Chapter 2,




**(a)** Figure adapted from Pardee et al. (2014). Overview of paper-based synthetic gene networks for sensing applications. Cell-free extract contains enzymes for transcription and translation, ribosomes, and dNTPs, tRNAs, and amino acids. The toehold switch sensor is a synthetic gene network that is activated by a complementary trans-acting RNA (trigger RNA). Reactions can be loaded onto paper and used immediately, or freeze-dried for later use.




**(b)** Figure adapted from Pardee et al. (2014). Overview of toehold switch sensors in the inactive (left) and active (right) states. The binding of a trans-acting RNA (trigger RNA) to the complementary region of the toehold switch allows for the switch to unfold, subsequently revealing a ribosomal binding site (RBS) and translational start site (AUG), leading to translation of the reporter gene (*e.g.*, GFP

**Figure 1.4:** Overview of paper-based synthetic gene networks and toehold switch sensors

I summarize the current approaches for designing toehold switch sensors. I then introduce a streamlined workflow for designing and testing toehold switch sensors for transcriptional phenotype-based AST. In Chapter 3, I introduce a first-of-its-kind toehold switch sensor assay for detecting differential RNA signatures of antibiotic susceptibility, and deploy this assay on a set of clinically relevant *E. coli* strains. This work includes numerous optimizations to the deployment of toehold switch sensors, as well as the characterization of a set of clinically relevant *E. coli*

strains. Finally, in Chapter 4, I describe a method for laboratory evolution of antibiotic resistance and introduce how it can be used for time-resolved analysis of gene expression in increasingly antibiotic resistant strains. Using a small set of *E. coli* strains evolved towards tetracycline resistance, I demonstrate how this method can be used to uncover new signatures of antibiotic susceptibility, as well as probe the evolution of antibiotic resistance.

Together, this work provides a set of computational and experimental methods that can be used to study the emergence of antibiotic resistance, and improve upon available methods for properly selecting and deploying antibiotics.

# CHAPTER 2

# A Comprehensive Framework for Designing, Testing, and Evaluating Toehold Switch Sensors for Detection of Differentially Expressed mRNAs

## 2.1 INTRODUCTION

As described in Chapter 1.4.4.2, toehold switch sensors are a programmable class of synthetic riboregulators that control the translation of a reporter gene via the binding of a trans-acting trigger RNA (Green et al. (2014)). Binding of the trans-acting RNA to the toehold region yields unfolding of the hairpin region of the toehold switch, revealing a ribosomal binding site (RBS) and a translation start site. Unlike CRISPR-based systems, toehold switch sensors can be designed to detect virtually any RNA sequence. This, combined with the portability of toehold switch sensors into paper-based contexts with cell-free extract, makes them a powerful tool for cell-free biosensing.

Toehold switch sensors are programmed to detect specific RNA sequences via encoding a complementary sequence into the toehold region of the switch (Figure 1.4b). The sensor read-out can be customized via the selection of the reporter gene that is translated. When designing toehold switch sensors, the hairpin and linker regions are largely conserved, save for sequence optimizations that have been reported, such as the introduction of a 'Series B' structure that exhibits increased reporter activity and decreased leakiness (Pardee et al. (2016)). Further modifications to the switch structure have yielded interesting new dynamics, including the ability to detect single nucleotide variants (Hong et al. (2020)).

Until recently, less than 1000 toehold switches had been designed and tested in cell-free *in vitro* contexts, making it challenging to systematically identify key de-

sign rules. Various work has been done to try and computationally predict the efficacy of toehold switch sensors using machine learning and thermodynamic modeling. In the beginning of this chapter, I compile all of the available tools and metrics for toehold switch sensor design. I then combine these design considerations into an open-source, easy-to-use tool for toehold switch sensor design (as well as the design of NASBA amplification components). The goal of this work is to provide a first-of-its-kind compendium of information on existing design strategies, as well as to provide a flexible tool for future toehold switch sensor design. The flexibility allows for design changes to be made as new design considerations become available.

## 2.2 CURRENT METHODS FOR DESIGNING TOEHOLD SWITCH SENSORS FOR RNA DETECTION

### 2.2.1 Brute Force Approach to Toehold Switch Sensor Design

The most common approach for toehold switch sensor design is to simply design as many toehold sequences as possible and test them *in vitro* (Green et al. (2014), Pardee et al. (2014), Pardee et al. (2016)). Recent work to design toehold switch sensors for the detection of norovirus provides an update to this brute-force design algorithm (Ma et al. (2018)). First, putative toehold switch sensor sequences are designed for every 36nt continuous region along the target RNA. Second, a set of metrics regarding the thermodynamics of RNA folding are computed to score each design. Third, sequences containing in-frame stop codons are filtered out. The resulting top candidate sequences are then ordered and constructed in plasmids. It was found that scoring these designs by these parameters provides a modest correlation to toehold switch sensor activity, but the authors note that larger libraries

of toehold switch sensors are needed to optimize these design strategies.

### 2.2.2 A Web-Based Tool for Toehold Switch Sensor Design

Most toehold switch sensor design workflows employ some combination of in-house scripts and manual entry. Recently, an iGEM team developed an online web tool to streamline the design process, however it is not open-source (To et al. (2018)). Still, it introduces both the need for and the utility of an automated tool to facilitate the design process. The steps employed to design the toehold switch sensors are as follows. First, the target RNA sequence and design parameters are used as input. Second, a sliding window is used to design a candidate toehold switch for each potential target RNA. For each region, fixed sequences corresponding to the RBS, start codon, and other fixed sequences are attached. Third, each candidate toehold switch sequence is evaluated for its specificity within the target organism, and confirmed to not contain an in-frame stop codon.

### 2.2.3 Deep Learning for Toehold Switch Sensor Design

A new pipeline for large-scale toehold sensor design, synthesis, and evaluation was developed to facilitate large-scale deep-learning on the relationship between structure and function of toehold switch sensors (Angenent-Mari et al. (2020)). In order to generate the large data set required for training deep neural nets, an *in vivo* pooled library approach was employed. Interestingly, it was discovered that the performance of toehold switch sensors *in vivo* differs from their performance in *in vitro* cell-free contexts, likely due to the vast differences between the natural cellular environment and the *in vitro* cell-free environment. Commonly used k-mer searches and thermodynamic modeling provided weak relationships between

toehold sensor sequence and function. However, training feed-forward neural networks on one-hot encoded representations of the full toehold switch sensor sequences provided improvement in the classification scores. However, one challenge with this approach is that it is not simple to translate the generated models into learned sequences or motifs that can be used for future toehold switch sensor design. Together, this work suggests that large-scale toehold sensor screens can be used to predict and identify high functioning toehold switch sensors, but the information gained is not directly portable into developing platforms for cell-free *in vitro* expression of toehold switch sensors.

A complementary study using the data set produced by Angenent-Mari et al. (2020) furthered the modeling performed on the relationship between toehold sequence and performance (Valeri et al. (2020)). This work identified a few new specific toehold switch sequence features that are over-represented in the best performing sensors (Table 2.1). Additionally, this group provides an online tool for predicting ON/OFF values for toehold switch sensors based on the sequence of the toehold region. Currently, this tool is not flexible for user-specified design constraints, nor can it be used for high-throughput toehold switch sensor design. However it is a beneficial, complementary resource to the approach I develop in this Chapter.

## 2.3 DESIGNING TOEHOLD SWITCH SENSORS FOR TRANSCRIPTIONAL PHENOTYPE-BASED AST

As highlighted by Angenent-Mari et al. (2020), the environmental context of the reaction is a critical variable in terms of toehold switch sensor efficacy. This might explain why in many of the above studies, *in silico* predictions and testing of toehold

| Position | Sequence | Note |
|---|---|---|
| 22–24 | AUA, CUA, GUA, UUA | NUA bulge opposite AUG start codon to prevent hybridization to start codon |
| 51, 54, 57 | No U | causes in-frame stop codon |
| overall | N/A | 20-60% GC |

**Table 2.1:** Toehold switch sensor sequence design considerations from Valeri et al. (2020)

switch sensors in different contexts did not always correlated to toehold switch sensor activity in the target context. Thus, for the design, testing and evaluation of toehold switch sensors for phenotpyic AST, it is important to consider the end-point workflow: the use of toehold sensors on NASBA amplified *E. coli* RNA. When designing candidate toehold switch sensor sequences and target RNA regions, the corresponding NASBA primers and amplicon region must also be considered. Moreover, it is useful to be able to rapidly design all of these components so that *in vitro* screens can be quickly performed. Currently, there is no available tool to consider all of these sequences in the design process in an automated way.

## 2.4 MATERIALS & METHODS

This section describes the specific steps taken by the algorithm for automating toehold switch sensor design and NASBA primer design. For brevity, it is referred to as "the workflow". A graphical overview of the workflow is shown in Figure 2.1.

### 2.4.1 Snakemake

This workflow was built using the snakemake workflow management system (Mölder et al. (2021)). The workflow is written in a human-readable, python-based lan-

guage, and allows for scalable and reproducible toehold switch sensor design. Rather than having to keep track of and individually run a batch of scripts to design each toehold switch sensor (and NASBA primers), this snakemake workflow can keep track of both which steps have been completed, and which steps need to be completed, for each target workflow. The workflow is broken down into a set of "rules", each which specify the target inputs and outputs. The operations of each rule are carried out in custom python scripts that interface with the snakemake infrastructure. Moreover, snakemake can parallelize the steps, and the entire design workflow can be run locally or on a shared computing cluster.

### 2.4.2 Installation & Initialization of the Design Workflow Environment

This design workflow is available as an open-source github repository: https://github.com/ebriars/toebuild. A README file provides the instructions for initializing the workflow, and a conda environment is used for reproducibility. A configuration file is used to take user inputs, including the species of interest (*e.g., E. coli*), and a gene, or set of genes, to build toehold switch sensors for. The sequence of each gene is identified by the workflow by querying the NCBI gene database (NCBI (2021)).

### 2.4.3 Using Nupack for Toehold Switch Sensor Design

Nupack is a software suite that allows for the analysis and design of nucleic acids and their structures (Wolfe et al. (2017)). This workflow automates the process of creating nupack design scripts and running nupack. For each gene, a sliding window is used to identify gene subregions. For each subregion, nupack is used to design a toehold switch that can bind to 36nts of this subregion within the spec-

ified design constraints. The design constraints include the desired structures of the switch when it is unbound (inactive), the switch when it is bound by the trans-acting RNA (active), and the trigger RNA. The nupack design file (input to nupack) and the resulting designs (direct output from nupack and a human-friendly simplified file) are automatically generated by the workflow and can be accessed by the user.

### 2.4.4 Filtering Candidate Toehold Switch Sensor Designs

This workflow can produce up to $\text{length}_{\text{gene}} - (\text{length}_{\text{toehold region}} - 1)$ designs. These candidate designs are filtered following a set of both standard and user-defined parameters. In some cases, nupack will design the same toehold switch sensor for slighlty different subregion windows. These duplicate designs are filtered from further analysis. Toehold switch sequences that have an in-frame stop codon, or do not have 20%–60% GC content are removed (as suggested by Valeri et al. (2020)). Toehold switch sequences that have repetitive regions (NNNNN, where N is the same nucleotide) are removed. The user also has the ability to add other custom filters to the desired sensor or trigger sequences by adding them to the configuration file.

### 2.4.5 Designing NASBA Primers and Synthetic RNA

To automate the primer design step, the command-line version of Primer3 is used (Untergasser et al. (2012)). A 350nt region that spans the toehold region is input as the sequence template. Forward and reverse primers are designed with the parameters in Table 2.2. The user has the ability to modify these parameters by modifying the primer design script. The resulting candidate primers are used to design syn-

thetic RNA (synRNA) corresponding to the amplicon region. This synRNA is the region of the gene that spans the forward NASBA primer binding region to the reverse NASBA primer binding region.

| Parameter | Description | Value |
|---|---|---|
| PRIMER_OPT_SIZE | Optimal primer size | 20nts |
| PRIMER_MIN_SIZE | Min. primer size | 19nts |
| PRIMER_MAX_SIZE | Max. primer size | 24nts |
| PRIMER_MIN_GC | Min. GC Content | 40% |
| PRIMER_MAX_GC | Max. GC Content | 60% |
| PRIMER_PRODUCT_SIZE_RANGE | Amplicon size | 180 – 215nts |

**Table 2.2:** Parameters used for NASBA primer design with Primer3

### 2.4.6 Formatting Nucleotide Designs for Ordering and *in vitro* Testing

The methods used for building toehold switch sensors and testing them with trigger RNA and NASBA amplified synRNA are described in the Material & Methods of Chapter 3. In order to facilitate the rapid testing and evaluation of these components, the workflow outputs the designed oligonucleotides in the following formats so that oligonucleotide sequences can be directly ordered (*e.g.*, from IDT).

#### *2.4.6.1 Toehold Switch Sensor*

A T7 promoter is appended to the 5′ end of the toehold switch sensor, and is linked to the switch such that the start of the switch is gGG. The toehold switch sensor design, which is RNA, is converted to DNA via a "U to T" substitution. This resulting oligonucleotide sequence (124 – 126 nts) is ordered as dsDNA (*e.g*, an IDT gBlock).

#### *2.4.6.2 Trigger RNA*

A T7 promoter is appended to the 5′ end of the trigger RNA. The trigger RNA, which is RNA, is converted to DNA via a "U to T" substitution. This resulting oligonucleotide sequence (60nts) is ordered as ssDNA.

#### *2.4.6.3 NASBA Primers*

A T7 promoter is appended to the 5′ end of the forward NASBA primer. The reverse NASBA primer is left as is. These oligonucleotide sequences are ordered as ssDNA.

#### *2.4.6.4 synRNA*

A T7 promoter is appended to the 5′ end of the synRNA sequence. The synRNA, which is RNA, is converted to DNA via a "U to T" substitution. This resulting oligonucleotide sequence (~200nts) is ordered as dsDNA (*e.g*, an IDT gBlock).

## 2.5 RESULTS

### 2.5.1 A Customizable and Automated Workflow for Designing Toehold Switch Sensors for RNA-based AST

The individual steps of the workflow are described in the Material & Methods, and can be visualized by the directed acyclic graph produced by snakemake in Figure 2.1. This is a representation of the workflow for designing three toehold switch sensors for two genes.

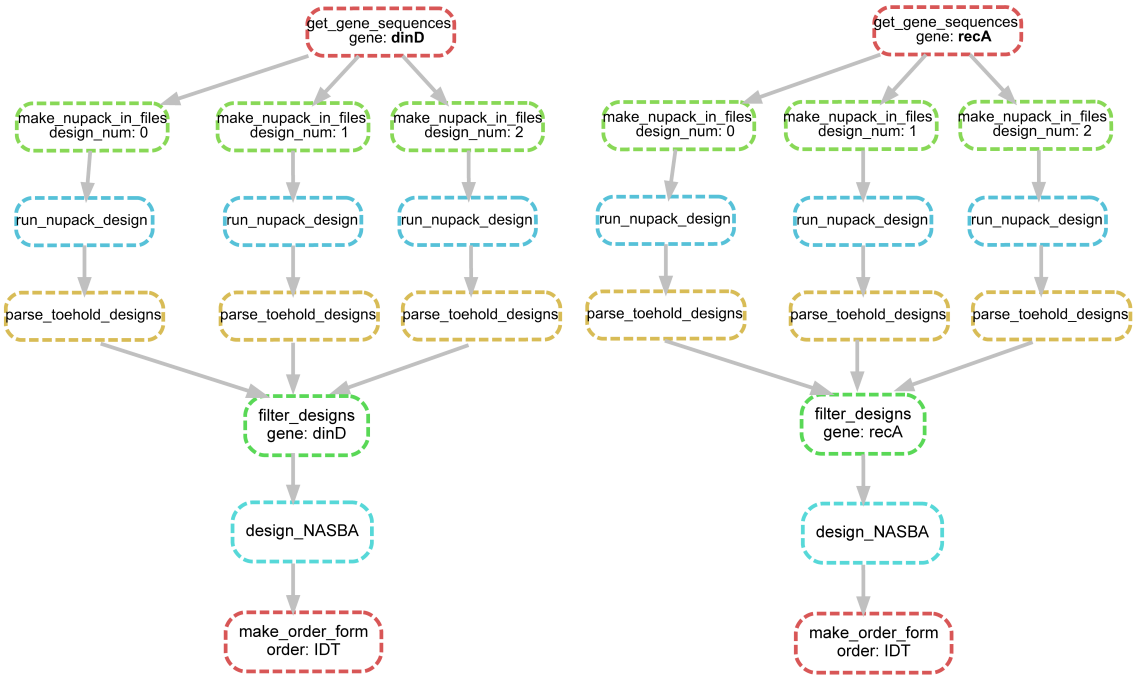



**Figure 2.1:** Workflow for automated toehold sensor design for RNA-based AST

Representation of the steps completed by the snakemake workflow. Here an example for designing three toehold sensors for two genes (dinD and recA).

### 2.5.2 Designing a Set of Toehold Switch Sensors and Components for a Gene Signature of Fluoroquinoline Susceptibility in *E. coli*

Genes associated with fluoroquinolone susceptibility in *E. coli*, and genes with no differential expression (control genes) were identified as described in Chapter 3.3.1 (Tables 3.15, 3.16). The strain *Escherichia coli* str. K-12 substr. MG1655 was used as the reference strain. The workflow described above was used to design candidate toehold switch sensors, trigger RNA, NASBA primers, and synRNA. The resulting oligonucleotide designs listed in Tables 2.6, 2.4, 2.5 were ordered from IDT.

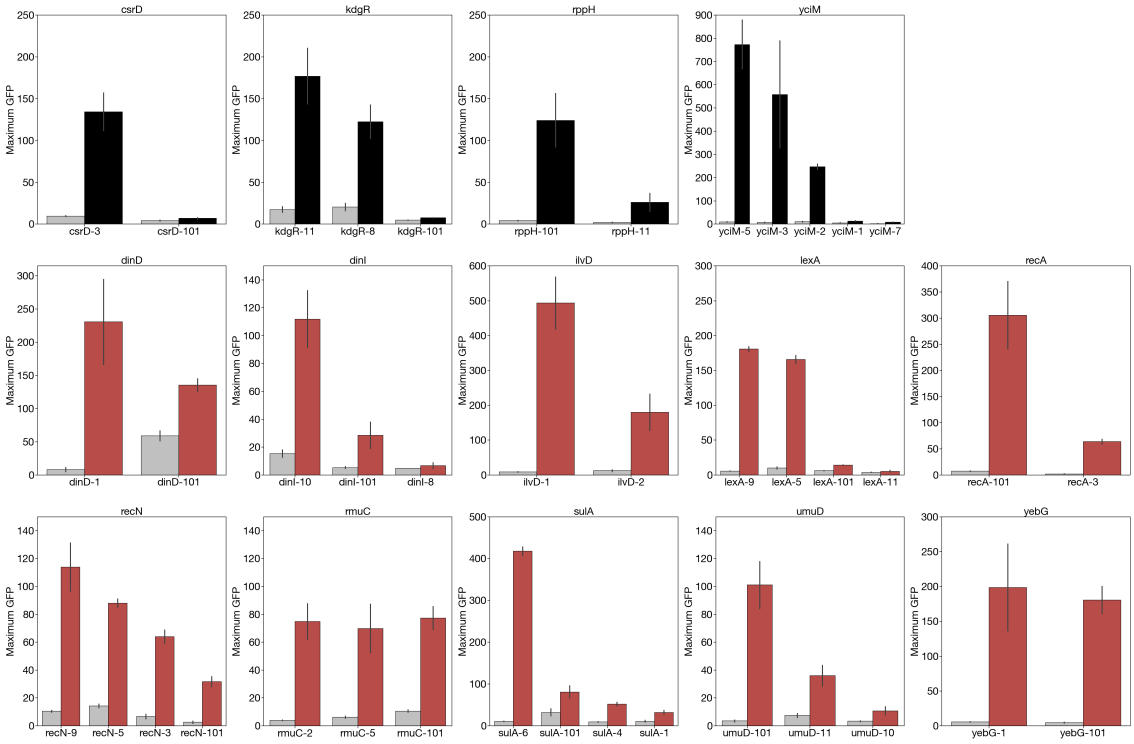



**Figure 2.2:** Screening Candidate Toehold Switch Sensors for the *E. coli* fluoroquinolone susceptibility signature

Sensitivity of each sensor in response to 2$\mu$M trigger RNA (36nt). Replicates are individual paper-based reactions. Top row (black bars) correspond to sensors detecting control genes; Second and third rows (red bars) correspond to sensors detecting signature genes (signature defined in Chapter 3).

### 2.5.3 Screening Candidate Toehold Switch Sensors with Short Trigger RNA

For each gene in the gene signature, a panel of candidate toehold switch sensors and trigger RNAs were generated as described in Chapter 3.2.2 and Chapter 3.2.4, respectively. The sensitivity of each toehold switch sensor was probed by measuring its GFP expression in response to 2$\mu$M of trigger RNA (see 3.2.5 for methods) (Figure 2.2). Each toehold switch sensor was evaluated by two metrics. First, 'GFP activation' was calculated by subtracting the GFP signal of the no RNA control

from the GFP signal of trigger RNA sample. Second, the 'background GFP' value was noted. For each gene, the toehold switch sensor with the highest activation greater than 100, and the lowest background less than 20 was selected as the best per gene sensor. For genes that did not have a sensor that met those criterion, additional sensors were designed, tested, and evaluated. The best sensor for each gene, and its corresponding metrics, are in Table 2.3.

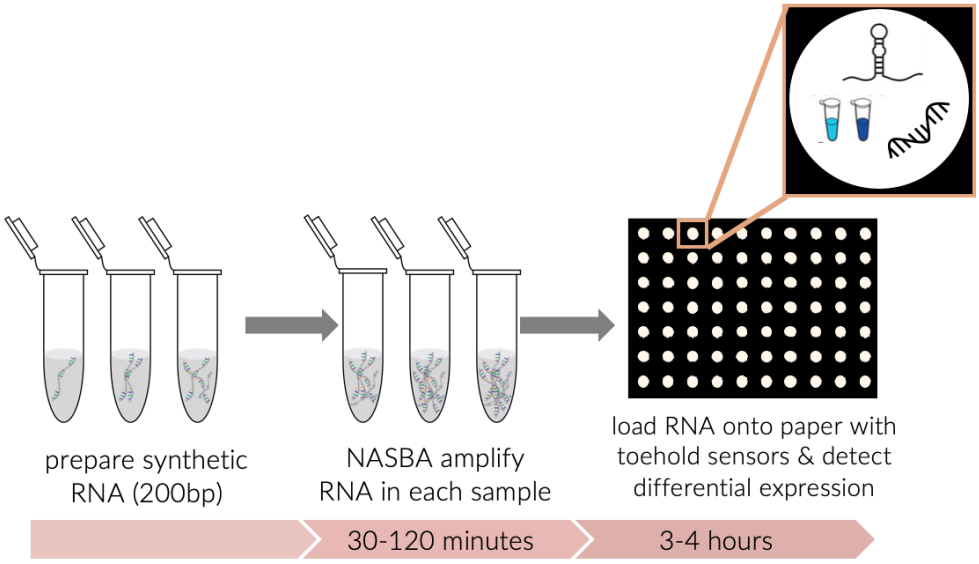



**Figure 2.3:** Workflow for screening toehold switch sensors with NASBA amplified synRNA

Serial dilutions of synRNA are prepared and NASBA amplified, screening different NASBA primer pairs and amplification times. The resulting NASBA amplified RNA is loaded onto paper discs with toehold sensors and other reaction components. Toehold sensor activation is measured by GFP production over time.

### 2.5.4 Screening NASBA Primers and Parameters with synRNA

Although screening toehold switch sensors for GFP activation signal and background GFP levels is an important step first step, toehold sensor activation is

| Sensor | GFP Activation | Background GFP |
|---|---|---|
| csrD-3 | 124.48 | 9.66 |
| kdgR-11 | 159.47 | 17.43 |
| rppH-101 | 119.78 | 4.27 |
| yciM-3 | 551.03 | 7.24 |
| dinD-1 | 222.59 | 8.09 |
| dinI-10 | 96.53 | 15.34 |
| ilvD-1 | 484.72 | 8.72 |
| lexA-9 | 175.10 | 5.56 |
| recA-101 | 298.35 | 7.41 |
| recN-9 | 103.46 | 10.42 |
| rmuC-2 | 71.02 | 3.91 |
| sulA-6 | 406.80 | 10.73 |
| umuD-101 | 97.58 | 3.49 |
| yebG-1 | 192.85 | 5.62 |

**Table 2.3:** Best sensors for each gene in the *E. coli* fluoroquinolone susceptibility signature

GFP activation: difference between maximum GFP signal from trigger RNA and no RNA control. Background GFP: GFP from no RNA control. Control genes are above the horizontal line; signature genes are below the horizontal line (signature defined in Chapter 3).

context dependent. In order to further screen candidate toehold sensors for the purpose of detecting differential RNA from NASBA amplified RNA, I employ an additional screening step with the synRNA designed by the workflow (Figure 2.3). With this approach, each toehold sensor can be screened against synRNA amplified with different primer pairs and for different amplification times. This screening step can be used to identify the primer pairs and NASBA parameters that yield

toehold sensor detection of differential amounts of RNA (*e.g.*, Figure 3.12).

## 2.6 DISCUSSION

Toehold switch sensors can be rapidly designed, tested, and evaluated for emerging disease diagnostics and new applications. Here, I have introduced a framework for facilitating both the *in silico* toehold switch sensor design, and the *in vitro* evaluation of toehold switch sensors. Moreover, this design workflow takes into consideration the context-dependence of toehold switch sensors. This workflow couples toehold sensor design with NASBA amplification primers and synRNA design so that toehold sensors can be rapidly screened not just for their efficacy in detecting short fragments of trigger RNA, but also 200nt NASBA amplified RNA. The set of toehold switch sensors, synRNA, and NASBA primers developed in this framework are used in Chapter 3.

## 2.7 SUPPLEMENTARY TABLES

**Table 2.4:** Oligonucleotides output by design workflow for toehold trigger RNA generation (best sensor per gene)

| **Trigger Name** | **Sequence** | **Type** |
|---|---|---|
| dinD-1-T | AAAGTGTGTGGTGTTTGCTTGCTGTTTT GAATTCACccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| dinI-10-T | TATCAGGAAACGCATACTGAATACGGC GGGAAAGTTccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |

| | | |
|---|---|---|
| ilvD-1-T | CCATTGATTTCAGGAAGCTGGTTGGGT AGAGCATTTccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| lexA-T | ACGCGCAGCAGGAAATCAGCATTCGGC TTGAATAAGccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| recA-101-T | GCGCATGATGGAGCCTTTACCAAATTG TTTCTCAATccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| recN-9-T | ACGTGCGGTCATTTCCTGCAGTAGAGA GGTTTCATTccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| rmuC-2-T | ATTTCGTGGGTCAGGGTATGGCGTTCTT GTGCTTCTccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| sulA-6-T | AGAGTTGCCAGCGCGATTGCTGACCGA GTTGCTGTAccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| umuD-101-T | AATTCCACCATCAATCATAGAATCACC ACTTGCTTTccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| yebG-1-T | CAAGCAGATCCGCCGTATCAAGCATTT TGTCATAAGccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| csrD-3-T | AGTAACGCAATAGTGGAATAAGACGG CTGATTTGCAccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |

| | | |
|---|---|---|
| kdgR-11-T | CCAGTTCAAACAATTTCAGGGTCAGGG AATATTTCTccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| rppH-101-T | CTTCGTGTCCCAACGCACCAAACGTTTC GGTAATTTccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |
| yciM-3-T | CAAGACGGACAATGCCAGTAGAGGGT GTATGCGGTAccctatagtgagtcgtattagcgc | 25nmole DNA Oligo |

**Table 2.5:** synRNA oligonucleotides and corresponding primer pairs output by design workflow for toehold sensor screening (best sensor per gene)

| **synRNA Name** | **synRNA Sequence** | **F Primer** | **R Primer** |
|---|---|---|---|
| dinD-1-synRNA | TAATACGACTCACTATAGGGGCTAT TCATCAGCGGAAGGGGCTGAAAAA GAATCAGAAGATCCTGGATCATAT GGGTTCAACAGAACTGGCGGCTAA TCTCTTTCGAGCTACCCAAACAGAA GAAAAACTCAAGCGGGATGGCGTG AATTCAAAACAGCAAGCAAACACC ACACACTTTGACGTGGGTCGCAAG GTGAGGCAAACCATTCAGGAACTT GGCGGAACCAT | AATTCTA ATACGAC TCACTAT AGGGAGA AGGGCTA TTCATCA GCGGAAG GG | ATGGTTC CGCCAAG TTCCT |

| | | | |
|---|---|---|---|
| dinI-10-synRNA | TAATACGACTCACTATAGGGGCGAAAACTTCTCCATTGCCAGCTGGGGCTATTGACGCCCTGGCTGGCGAACTTTCCCGCCGTATTCAGTATGCGTTTCCTGATAATGAAGGCCACGTATCGGTACGTTATGCCGCAGCGAATAATTTATCGGTTATTGGCGCGACAAAAGAAGATAAACAGCGCATTAGTGAAATTCTCCAGGAAACGTGGGAAAGCGCCGATGACTGGTTT | AATTCTAATACGACTCACTATAGGGAGAAGGGCGAAAACTTCTCCATTGCCA | AAACCAGTCATCGGCGCTTT |
| lexA-synRNA | TAATACGACTCACTATAGGGAATTGTTTCCGGCGCATCACGCGGGATTCGTCTGTTGCAGGAAGAGGAAGAAGGGTTGCCGCTGGTAGGTCGTGTGGCTGCCGGTGAACCACTTCTGGCGCAACAGCATATTGAAGGTCATTATCAGGTCGATCCTTCCTTATTCAAGCCGAATGCTGATTTCCTGCTGCGCGTCAGCGGGATGTCGATGAAAGA | AATTCTAATACGACTCACTATAGGGAGAAGGAATTGTTTCCGGCGCATCAC | TCTTTCATCGACATCCCGCT |

| | | | |
|---|---|---|---|
| recA-101-synRNA | TAATACGACTCACTATAGGGACAAACAGAAAGCGTTGGCGGCAGCACTGGGCCAGATTGAGAAACAATTTGGTAAAGGCTCCATCATGCGCCTGGGTGAAGACCGTTCCATGGATGTGGAAACCATCTCTACCGGTTCGCTTTCACTGGATATCGCGCTTGGGGCAGGTGGTCTGCCGATGGGCCGTATCGTCGAAATCTACGGACCGGAATCTTCCG | AATTCTAATACGACTCACTATAGGGAGAAGGACAAACAGAAAGCGTTGGCG | CGGAAGATTCCGGTCCGTAG |
| recN-9-synRNA | TAATACGACTCACTATAGGGTGGTCAGCACGCTCATCAATTACTCACCAAACCTGAGCACCAAAAATTCCTGCTTGATGGCTATGCCAATGAAACCTCTCTACTGCAGGAAATGACCGCACGTTATCAGTTGTGGCATCAAAGCTGCCGTGACCTCGCGCATCATCAACAGTTAAGTCAGGAACGCGCCGCCCGTGCGGAACTGCTGCAATAC | AATTCTAATACGACTCACTATAGGGTGGTCAGCACGCTCATCAATTACTCACCAAACCTGAGCACCAAAAATTCCTGCTTGATGGCTATGCCAATGAAACCTCTCTACTGCAGGAAATGACCGCACGTTATCAGTTGTGGCATCAAAGCTGCCGTGACCTCGCGCATCATCAACAGTTAAGTCAGGAACGCGCCGCCCGTGCGGAACTGCTGCAATAC | GTATTGCAGCAGTTCCGCAC |

| | | | |
|---|---|---|---|
| rmuC-2-synRNA | TAATACGACTCACTATAGGGGCACAAGAACGCCATACCCTGACCCACGAAATTCGCAATCTCCAGCAACTCAACGCGCAAATGGCCCAGGAAGCGATCAACCTGACGCGCGCTGAAAGGCGACAATAAAACCCAGGG | AATTCTAATACGACTCACTATAGGGAGAAGGGCACAAGAACGCCATACCCT | CCCTGGGTTTTATTGTCGCC |
| sulA-6-synRNA | TAATACGACTCACTATAGGGTCTTCGTCGTTCTCATCCGCAGCAAGTAAAATTGCGCGTGTCTCTACGGAAAACACTACAGCCGGGCTTATCAGTGAAGTTGTCTATCGCGAAGATCAGCCCATGATGACGCAACTTCTACTGTTGCCATTGTTACAGCAACTCGGTCAGCAATCGCGCTGGCAACTCTGGTTAACACCGCAACAAAAACTGAGTCGGGAATGGGTTCAGG | AATTCTAATACGACTCACTATAGGGAGAAGGTCTTCGTCGTTCTCATCCGC | CCTGAACCCATTCCCGACTC |

| | | | |
|---|---|---|---|
| umuD-101-synRNA | TAATACGACTCACTATAGGGTCAGTGTGGCTTTCCTTCACCGGCAGCAGATTACGTTGAACAGCGCATCGATCTGAATCAACTGTTGATCCAGCATCCCAGCGCGACTTACTTCGTCAAAGCAAGTGGTGATTCTATGATTGATGGTGGAATTAGTGACGGTGATTTACTGATTGTCGATAGCGCTATTACCGCCAGCCATGGTG | AATTCTAATACGACTCACTATAGGGAGAAGGTCAGTGTGGCTTTCCTTCACC | CACCATGGCTGGCGGTAATA |
| yebG-1-synRNA | TAATACGACTCACTATAGGGGTCATTCGTGAGGGCGAAGAGAAAATGTCGTTTACCAGCAAAAAGGAAGCCGATGCTTATGACAAAATGCTTGATACGGCGGATCTGCTTGACACCTGGCTGACAAATTCTCCAGTGCAAATGGAAGACGAGCAACGTGAAGCCCTTTCGCTATGGCTGGCAGAACAAAAAGATGTGCTGAGCACCATTCTGAAAACCGGCAAATTACCGTCC | AATTCTAATACGACTCACTATAGGGAGAAGGGTCATTCGTGAGGGCGAAGA | GGACGGTAATTTGCCGGTTT |

| | | | |
|---|---|---|---|
| csrD-3-synRNA | TAATACGACTCACTATAGGGCGGTTACTCGCGAAGGTCAGGTTCATCATCGCGAACTCATGTGCCGCATCTTCGATGGTAATGAAGAGGTTAGCTCGGCGGAGTATATGCCGATGGTCTTGCAGTTTGGCTTATCGGAAGAGTATGACCGTCTGCAAATCAGCCGTCTTATTCCACTATTGCGTTACTGGCCAGAGGAAAATCTGGCGATTCAGGTTACCGTTGAGTCGCT | AATTCTAATACGACTCACTATAGGGAGAAGGCGGTTACTCGC GAAGGTCAG | AGCGACTCAACGGTAACCTG |
| kdgR-11-synRNA | TAATACGACTCACTATAGGGTAGGGATAACCGAGCTGTCGCAGCGCGTCATGATGTCAAAAAGCACCGTTTATCGCTTTTTACAGACCATGAAAACCTTAGGTTATGTGGCGCAGGAAGGGGAGTCGGAGAAATATTCCCTGACCCTGAAATTGTTTGAACTGGGCGCTCGCGCGTTACAAAACGTCGATTTAATTCGTAGCGCAGATATCCAGATGCGTG | AATTCTAATACGACTCACTATAGGGAGAAGGTAGGGATAACCGAGCTGTCG | CACGCATCTGGATATCTGCG |

| | | | |
|---|---|---|---|
| rppH-101-synRNA | TAATACGACTCACTATAGGGCGGAATCAACCCCGGAGAATCCGCAGAGCAGGCGATGTACCGTGAATTGTTTGAAGAAGTAGGATTAAGCCGCAAAGACGTTCGAATCCTTGCTTCAACGCGTAACTGGTTGCGCTACAAATTACCGAAACGTTTGGTGCGTTGGGACACGAAGCCGGTTTGTATCGGCCAAAAACAAAAATGGTTTCTCTTGCAGCTGGT | AATTCTAATACGACTCATAGAGGAGAGGCGGAATCAACCGATC | ACCAGCTGCAAGAGAAACCAT |
| yciM-5-synRNA | TAATACGACTCACTATAGGGGCAGAAGTGCGCAACAAAACAAGCAAGATGAAGCCAACCGCTTGTCGCGTGATTACGTAGCGGGGGTTAACTTCCTGCTTAGTAATCAACAGGATAAAGCGGTAGACCTGTTTCTCGATATGCTTAAAGAGGATACAGGCACCGTTGAAGCCCACCTTACGCTCGGAAACCT | AATTCTAATACGACTCACTATAGAGGTATGGCTGGTATATGGGCCG | CAGGTTTCCGAGCGTAAGGT |

**Table 2.6:** Oligonucleotides output by design workflow for toehold sensor generation (best sensor per gene)

| **Oligo Name** | **Sequence** | **Type** |
|---|---|---|

| | | |
|---|---|---|
| dinD-1-S-GFP | taatacgactcactatagGGAAAGTGTGTGGTGTTTGCTTGCTGTTTTGAATTCACGGACTTTAGAACAGAGGAGATAAAGATGGTGAATTCAAAAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| dinI-10-S-GFP | taatacgactcactatagGGTATCAGGAAACGCATACTGAATACGGCGGGAAAGTTGGACTTTAGAACAGAGGAGATAAAGATGAACTTTCCCGCAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| ilvD-1-S-GFP | taatacgactcactatagGGCCATTGATTTCAGGAAGCTGGTTGGGTAGAGCATTTGGACTTTAGAACAGAGGAGATAAAGATGAAATGCTCTACAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| lexA-9-S-GFP | taatacgactcactatagGGACGCGCAGCAGGAAATCAGCATTCGGCTTGAATAAGGGACTTTAGAACAGAGGAGATAAAGATGCTTATTCAAGCAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| recA-101-S-GFP | taatacgactcactatagGTACTCTTCGTCGATTTGTTCAAACTCTCCGGGCTGCGGACTTTAGAACAGAGGAGATAAAGATGGCAGCCCGGAGTAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |

| | | |
|---|---|---|
| recN-9-S-GFP | taatacgactcactatagGGACGTGCGGTCATTTCCTGCAGTAGAGAGGTTTCATTGGACTTTAGAACAGAGGAGATAAAGATGAATGAAACCTCGAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| rmuC-2-S-GFP | taatacgactcactatagGGATTTCGTGGGTCAGGGTATGGCGTTCTTGTGCTTCTGGACTTTAGAACAGAGGAGATAAAGATGAGAAGCACAAGTAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| sulA-6-S-GFP | taatacgactcactatagGGAGAGTTGCCAGCGCGATTGCTGACCGAGTTGCTGTAGGACTTTAGAACAGAGGAGATAAAGATGTACAGCAACTCAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| umuD-101-S-GFP | taatacgactcactatagGGGAATTCCACCATCAATCATAGAATCACCACTTGCTTTGGACTTTAGAACAGAGGAGATAAAGATGAAAGCAAGTGGAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| yebG-1-S-GFP | taatacgactcactatagGGCAAGCAGATCCGCCGTATCAAGCATTTTGTCATAAGGGACTTTAGAACAGAGGAGATAAAGATGCTTATGACAAACAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |

| | | |
|---|---|---|
| csrD-3-S-GFP | taatacgactcactatagGGAGTAACGCAATAGTGGAATAAGACGGCTGATTTGCAGGACTTTAGAACAGAGGAGATAAAGATGTGCAAATCAGCAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| kdgR-11-S-GFP | taatacgactcactatagGGCCAGTTCAAACAATTTCAGGGTCAGGGAATATTTCTGGACTTTAGAACAGAGGAGATAAAGATGAGAAATATTCCAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| rppH-101-S-GFP | taatacgactcactatagGGCTTCGTGTCCCAACGCACCAAACGTTTCGGTAATTTGGACTTTAGAACAGAGGAGATAAAGATGAAATTACCGAAAAACCTGGCGGCAGCGCAAAAGATGCGTAAA | gBlock |
| yciM-3-S-GFP | taatacgactcactatagGCAAGACGGACAATGCCAGTAGAGGGTGTATGCGGTAGGACTTTAGAACAGAGGAGATAAAGATGTACCGCATACAAAACCTGGCGGCAGCGCAAAAGATGCGTAA | gBlock |

# CHAPTER 3

# A Toehold Sensor Platform for Detecting Ciprofloxacin Susceptible *Escherichia coli*

## 3.1 INTRODUCTION

UTIs are incredibly common community and health-care acquired infections, and are the second most common infection to be prescribed antibiotics (Foxman (2010)). Although UTIs can be caused by bacterial species such as *Klebsiella* or *Enterobacter*, uropathogenic *E. coli* (UPEC) strains are the most prevalent cause of UTIs (WHO (2020), Shaifali et al. (2012), Addis et al. (2021)). Most UPEC strains are susceptible to at least one common oral antibiotic, however the specific antibiotic susceptibility profile of the strain should be used to determine the appropriate antibiotic course (Hooton & Stamm (1997)). More than twenty years ago, 25% to 75% of UPEC isolates were resistant to ampicillin and sulfonamides, and less than 5% were resistant to fluoroquinolones, such as ciprofloxacin (Hooton & Stamm (1997), Grüneberg (1994)). However today, ciprofloxacin resistant UTIs are a major global health concern with up to 92.9% of cases being caused by ciprofloxacin resistant *E. coli* (WHO (2020)). Thus, in only two decades we have seen a dramatic shift in the antibiotic susceptibility profiles of UPEC isolates, and an increased concern for the prevalence of antibiotic resistant strains.

Due to the widespread nature of ciprofloxacin resistant *E. coli* UTIs, there is a great need for rapid antibiotic susceptibility testing (AST) for UTIs. In this chapter, I demonstrate the development, testing and deployment of a proof-of-concept toehold switch sensor platform for detecting fluoroquinolone susceptible *E. coli*.

### 3.1.1 Proposed Workflow for AST Using Paper-Based Toehold Switch Sensor Detection of Signature Transcripts

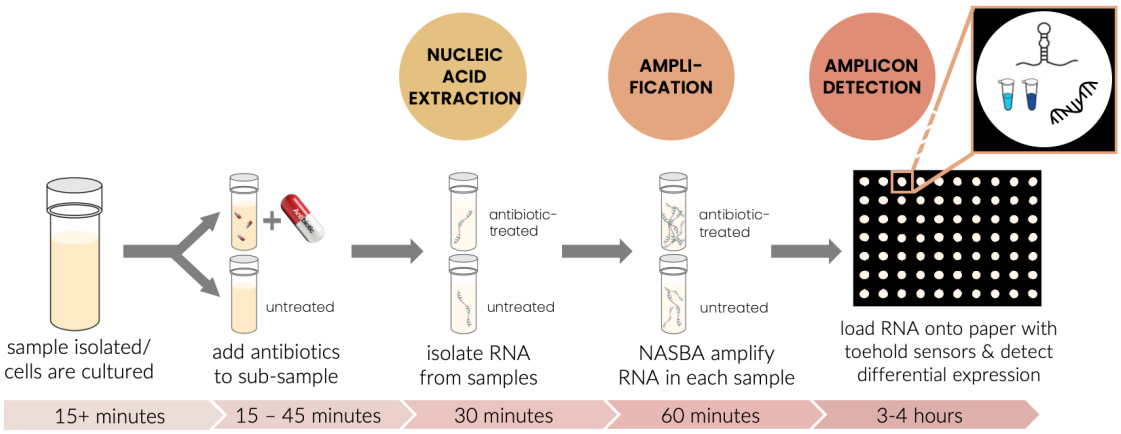
As described in Chapter 1, the gold-standard methods for AST begin with collecting a sample culture, and then waiting for bacterial growth (48-72+ hours) in the presence of antibiotics in order to make a susceptibility or resistance determination. Here, I introduce a faster phenotypic paradigm (Figure 3.1). First, a bacterial sample is isolated and split into two sub-samples. One sub-sample is treated with an antibiotic at the CLSI MIC break-point concentration (*i.e.*, the concentration at which if cells did grow, they would be designated resistant), and the other sub-sample is left untreated. After this brief antibiotic exposure (*e.g.*, 30 minutes), the steps of a nucleic acid-based diagnostic are employed. First, RNA is isolated (nucleic acid extraction). Second, the RNA is amplified using an isothermal amplification technique, nucleic acid sequence-based amplification (NASBA). Third, the amplified RNA is loaded onto paper discs containing toehold switch sensors and cell-free extract for amplicon detection. After I developed and optimized the steps of this workflow, I deploy it on a set of clinically relevant, UPEC strains. The details of each steps in this workflow for assay development and deployment are described in the following Materials & Methods section, and the results are described in the following Results section.

## 3.2 MATERIALS & METHODS

### 3.2.1 Toehold Switch Sensor Design

Candidate toehold switch sensors were designed as described in Chapter 2, and DNA oligonucleotides containing a T7 promoter, toehold switch, and linker region

**Figure 3.1:** Proposed workflow for a toehold sensor based antibiotic susceptibility test




were ordered (Integrated DNA Technologies (IDT)) (Table 2.6).

### 3.2.2 Toehold Switch Sensor Generation

Toehold switch oligonucleotides (gBlocks, IDT) were reconstituted per IDT recommendations in TE buffer to 10 ng/$\mu$L. Extension PCR was employed to attach the gBlock containing the toehold switch to linear DNA containing GFP (GFPmut3b-ASV with T7 terminator) ('Toehold Switch Linker-GFP'). For each toehold switch, extension PCR 1 was set-up on ice following the reaction components listed in Table 3.1. Reactions were run on a thermocycler following the protocol in Table 3.2. The resulting PCR products ('Toehold Switch Product A') were purified using the Qiagen MinElute PCR Purification Kit (Qiagen: 28004). Products were quantified using a nanodrop and validated using a 1% agarose gel. Next, extension PCR 2 was set-up on ice following the reaction components in Table 3.3 to stitch 'Toehold Switch Product A' to 'Toehold Switch Linker-GFP'. Reactions were run on a thermocycler following the protocol in Table 3.4. Again, the resulting PCR products (linearized DNA 'Toehold switch sensors') were purified, quantified, and verified.

Toehold switch sensors were diluted to working stocks of 24nM in water.

| Component | Concentration | Volume $\mu$L |
|---|---|---|
| Sensor-Trigger-Generation-F | 10 $\mu$M | 2.5 |
| Sensor-Linker-Generation-R | 10 $\mu$M | 2.5 |
| Toehold switch DNA oligo | 1 ng/$\mu$L | 1 |
| Phusion Master Mix | N/A | 25 |
| RNase Free Water | N/A | 19 |

**Table 3.1:** Reaction components for toehold switch sensor generation for PCR extension 1

| Cycle(s) | Temperature (°C) | Time (seconds) |
|---|---|---|
| 1 | 98 | 30 |
| 2 - 31 | 98 | 10 |
| | 64 | 20 |
| | 72 | 15 |
| 32 | 72 | 7 |
| 33 | 4 | $\infty$ |

**Table 3.2:** Toehold switch sensor generation PCR protocol for PCR extension 1

### 3.2.3 Trigger RNA and Synthetic RNA Design

RNA trigger (35nt) and synthetic RNA (synRNA) (~200nt) sequences were designed as described in Chapter 2, and DNA oligonucleotides were ordered (IDT) (Tables 2.4, 2.6, 2.5).

| Component | Concentration | Volume $\mu$L |
|---|---|---|
| Sensor-Trigger-Generation-F | 10 $\mu$M | 2.5 |
| Sensor-GFP-Generation-R | 10 $\mu$M | 2.5 |
| Toehold Switch Product A | 1 ng/$\mu$L | 1 |
| Toehold Switch Linker-GFP | 3 ng/$\mu$L | 1 |
| Phusion Master Mix | N/A | 25 |
| RNase Free Water | N/A | 18 |

**Table 3.3:** Reaction components for toehold switch sensor generation for PCR extension 2

| Cycle(s) | Temperature (°C) | Time (seconds) |
|---|---|---|
| 1 | 98 | 30 |
| 2 - 36 | 98 | 10 |
| | 67 | 20 |
| | 72 | 30 |
| 37 | 72 | 7 |
| 38 | 4 | $\infty$ |

**Table 3.4:** Toehold switch sensor generation PCR protocol for PCR extension 2

### 3.2.4 Trigger RNA & synRNA Generation

Trigger DNA oligonucleotides were reconstituted in TE buffer to 100 $\mu$M. The T7 promoter region was made double stranded by annealing a complementary DNA oligonucleotide. The reaction components in Table 3.5 were combined at room temperature, followed by annealing at 95°C for 5 minutes, followed by an hour of cooling to room temperature. synRNA dsDNA oligonucleotides (gBlocks, IDT) were reconstituted following the IDT recommendations in TE buffer to 10 ng/$\mu$L. Trigger DNA and synRNA DNA were *in vitro* transcribed to RNA using T7 me-

diated transcription (HiScribe T7 Quick High Yield RNA Synthesis Kit (NEB): E2050S) following the reaction components in Tables 3.6 and 3.7, and incubated at 37°C for three hours. Following *in vitro* transcription, reactions were treated with DNase I for 15 minutes. The resulting RNA products were purified using the RNA Clean and Concentrator Kit (Zymo: R1017) and quantified using the Qubit RNA Broad-Range (BR) Assay (Invitrogen: Q10210).

| **Component** | **Volume $\mu$L** |
|---|---|
| TE (1X) | 24.2 |
| 5M NaCl | 0.25 |
| ds-T7 (100$\mu$M) | 0.275 |
| Trigger RNA (100 $\mu$M) | 2 |

**Table 3.5:** Reaction components for making Trigger RNA with a dsT7

| **Component** | **Volume $\mu$L** |
|---|---|
| RNase-free Water | 3 |
| NTP Buffer Mix | 10 |
| Trigger RNA with a dsT7 | 5 |
| T7 RNA Polymerase Mix | 2 |

**Table 3.6:** Reaction components for trigger RNA T7-mediated synthesis

### 3.2.5 Paper-based Toehold Sensor Detection of Trigger RNA

A 2mM punch biopsy was used to cut paper discs (Whatman: 1442-042) into each well of a 384 black-walled plate (Corning: 3544) such that the unused wells and border wells were filled with water. Each cell-free toehold sensor reaction was assembled on ice following the reaction components in Table 3.8 using components

| Component | Volume $\mu$L |
|---|---|
| RNase-free Water | 8 |
| NTP Buffer Mix | 10 |
| synRNA (10 ng/$\mu$L) | 10 |
| T7 RNA Polymerase Mix | 2 |

**Table 3.7:** Reaction components for synRNA T7-mediated synthesis

from the NEB PURExpress in vitro Protein Synthesis Kit (NEB: E6800). 1.4$\mu$L of each reaction was spotted onto an individual paper disc, in triplicate. The plate was sealed and immediately transferred to a plate reader set at 37°C for at least two hours. A fluorescent measurement was taken every five minutes.

| Component | Volume ($\mu$L) | Percentage |
|---|---|---|
| NEB Reaction Mix A | 1.6 | 40% |
| NEB Reaction Mix B | 1.2 | 30% |
| RNase Inhibitor (Roche: 3335399001) | 0.02 | 1% |
| Sensor (1.8 nM) | 0.3 | 8% |
| Trigger RNA (2 $\mu$M) | 0.7 | 17% |
| RNase-free water | 0.180 | 5% |

**Table 3.8:** Reaction components for cell-free toehold sensor detection of trigger RNA

### 3.2.6 Nucleic Acid Sequence-Based Amplification (NASBA) of synRNA

NASBA amplification reactions were assembled on ice following the reaction components in Table 3.9, save the enzyme cocktail, using the NASBA Liquid Kit (Life Sciences Advanced Technologies: NWK-1). Reactions were briefly and gently vortexed to mix, and centrifuged to collect at the bottom of the tubes. Reactions were

added to a pre-heated thermocycler (65°C) and incubated at 65°C for five minutes, followed by ten minutes at 41°C (lid temperature: 41°C). NASBA enzyme cocktail was added to each reaction, and reactions were briefly vortexed and centrifuged, then returned to the thermocycler for amplification at 41°C (lid temperature: 41°C) for 30 minutes - 120 minutes. Following amplification, reactions were heat inactivated at 70°C for ten minutes. Upon completion, reactions were immediately removed from the thermocycler, and put on ice.

| Component | Volume ($\mu$L) | Percentage |
|---|---|---|
| 3X NASBA Reaction Buffer | 1.675 | 33.5% |
| 6X NASBA Nucleotide Mix | 0.825 | 16.5% |
| RNase Inhibitor (Roche: 3335399001) | 0.025 | 0.5% |
| NASBA F Primer (12.5 $\mu$M) | 0.05 | 1% |
| NASBA R Primer (12.5 $\mu$M) | 0.05 | 1% |
| synRNA/NASBA amplified synRNA | 1 | 20% |
| RNase-free water | 0.180 | 2.5% |
| NASBA Enzyme Cocktail | 1.25 | 25% |

**Table 3.9:** Reaction components for NASBA amplification of synRNA

### 3.2.7 Paper-based Toehold Sensor Detection of Amplified synRNA

A 2mM punch biopsy was used to cut paper discs (Whatman: 1442-042) into each well of a 384 black-walled plate (Corning: 3544) such that the unused wells and border wells were filled with water. Each toehold sensor reaction was assembled on ice following the reaction components in Table 3.10 using components from the NEB PURExpress in vitro Protein Synthesis Kit (NEB: E6800). 2$\mu$L of each reaction was spotted onto an individual paper disc, in duplicate or triplicate. The plate was

sealed and immediately transferred to a plate reader set at 37°C for at least four hours. A fluorescent measurement was taken every five minutes.

| Component | Volume ($\mu$L) | Percentage |
|---|---|---|
| NEB Reaction Mix A | 0.8 | 40% |
| NEB Reaction Mix B | 0.6 | 30% |
| RNase Inhibitor (Roche: 3335399001) | 0.01 | 1% |
| Sensor (1.8 nM) | 0.15 | 8% |
| synRNA | 0.334 | 17% |
| RNase-free water | 0.106 | 5% |

**Table 3.10:** Reaction components for cell-free toehold sensor detection of synRNA

### 3.2.8 Identifying a Panel of Clinically Relevant *E. coli* Strains Using staramr

Susceptibility to ciprofloxacin was computationally predicted for each *E. coli* strain in a library of strains (library from Galardini et al. (2017)). A fasta file containing the genome assembly for each of the 696 *E. coli* strains was analyzed by staramr (version: 0.4.0) against the ResFinder database (version: Mon, 16 Jul 2018 16:58) (Zankari et al. (2012)). Briefly, staramr scans bacterial genome contigs against the ResFinder database, and provides a summary report of detected antimicrobial resistance genes. The summary report was parsed for strains with predicted ciprofloxacin susceptibility and resistance.

### 3.2.9 Characterizing a Panel of Clinically Relevant *E. coli* Strains with Minimum Inhibitory Concentration (MIC) Determination

Overnight cultures of *E. coli* strains grown in Mueller Hinton Broth (MHB) were diluted 1:10,000 in MHB containing a 3.5X dilution series of ciprofloxacin (604 $\mu$g/mL - 0.0004 $\mu$g/mL) and grown at 37°C with shaking for twenty-four hours. Growth/no growth results were interpreted by eye and confirmed on a plate reader by taking absorbance measurements at $OD_{600}$. MIC was estimated by fitting the data to a modified Gompertz function (Lambert & Pearson (2000)) (3.1) and using the parameters to calculate MIC with Equation 3.2.

$$
\begin{aligned}
&y = A + Ce^{-e^{B(x-M)}}\text{, where} \\
&A\text{: lower asymptote of y} \\
&B\text{: slope parameter} \\
&C\text{: distance between upper and lower asymptote} \\
&M\text{: log concentration of the inflection point}
\end{aligned}
\tag{3.1}
$$

$$
\text{MIC} = 10^{M+1/B} \tag{3.2}
$$

### 3.2.10 Gene Ontology Analysis and Visualization

The gene IDs for each gene in the *E. coli* fluoroquinolone susceptibilty signature were inputs for gene ontology enrichment analysis (Ashburner et al. (2000), gen (2021)). GO terms (biological processes) that were significantly over-represented compared to a reference list of all *E. coli* genes were identified using PANTHER with a false discovery rate p-value < 0.05 (Mi et al. (2021)). To visualize the re-

sulting GO terms, their semantic similarity scores (RSS) were computed and represented in a two-dimensional similarity space using NaviGO (Wei et al. (2017)). The GO terms were colored by their highest-level membership.

### 3.2.11 *E. coli* Strain Growth & Brief Antibiotic Exposure to Elicit Transcriptional Responses

Single colonies of *E. coli* strains were cultured overnight in MHB in a shaking incubator at 37°C. Each sample was diluted 1:1000 and grown to mid-log phase. Each sample was split into two 1mL sub-samples and incubated with either 5 $\mu$L ciprofloxacin ('Cipro'; final concentration: 1 $\mu$g/mL) or 5 $\mu$L water ('NT') for 30 minutes in a shaking incubator at 37°C. Samples were immediately processed for RNA isolation.

### 3.2.12 *E. coli* RNA Isolation

RNA isolation was performed using the RNeasy mini kit (Qiagen: 74104) with the Qiagen RNAprotect Bacteria Reagent (Qiagen: 76506). Briefly, 250 $\mu$L of each sample culture was added to a tube containing 500 $\mu$L RNAprotect Bacteria Reagent, and then processed following the RNAprotect Bacteria Reagent protocl. Each pellet was resuspended in Proteinase K (20 mg/mL; Qiagen: 19131) in TE containing lysozyme (15 mg/mL; Sigma Aldrich: L6876) and incubated at room temperature for ten minutes, and vortexed every two minutes. RNA isolation was done following Protocol #4 of the RNeasy mini kit with the optional DNAse treatment step (Qiagen: 79254), and ultimately eluted in RNase-free water. Isolated RNA was quantified using the Qubit RNA Broad-Range (BR) Assay (Invitrogen: Q10210). Isolated RNA was used in the following experiments: RT-qPCR, NASBA amplifi-

cation of *E. coli* RNA.

### 3.2.13 RT-qPCR

Transcriptional response to brief antibiotic exposure in the set of clinically relevant *E. coli* strains was probed by RT-qPCR using the iTaq Universal SYBR Green One-Step Kit (Bio-Rad: 1725151). *E. coli* strains were treated with brief antibiotic exposure and RNA was isolated as described above. RT-qPCR reactions were assembled on ice following the reaction components in Table 3.11. 10 $\mu$L of each reaction were input in duplicate into the wells of a 384-well white-walled plate. No reverse transcriptase (NRT) and no template (NTC) controls were always performed. RT-qPCR reactions were run following the protocol in Table 3.12. The primers for each signature gene and control gene are listed in Table 3.20. Only data from wells with one observed melting curve was observed was used for analysis. $dC_t$ values were calculated by subtracting either the NT signature gene $C_t$ values or the Cipro signature gene $C_t$ value from their respective NT or Cipro control gene $C_t$ values. Each gene's $ddC_t$ values were calculated by subtracting the Cipro $dC_t$ values from the NT $dC_t$ values. Fold changes were calculated as $2^{ddC_t}$, and replicates were used to compute the mean and standard deviation.

### 3.2.14 NASBA Amplification of *E. coli* RNA

NASBA amplification was carried out similar to the amplification of synRNA. Isolated RNA was diluted to 0.1ng - 1.0 ng in water. Each reaction always had an amplification control with no RNA. The current optimal gene specific NASBA parameters are in Table 3.13.

| Component | Volume ($\mu$L) | Percentage |
|---|---|---|
| SYBR Green Master Mix | 5 | 50% |
| Reverse transcriptase | 0.125 | 1.25% |
| Forward Primer (10 $\mu$M) | 0.3 | 3% |
| Reverse Primer (10 $\mu$M) | 0.3 | 3% |
| RNA (1 ng/$\mu$L) | 1 | 10% |
| RNase-free water | 3.275 | 32.75% |

**Table 3.11:** Reaction components for RT-qPCR of *E. coli* strains after brief ciprofloxacin exposure

| Function | Time | Temperature (°C) | # of Cycles |
|---|---|---|---|
| Reverse Transcription | 10 minutes | 50 | 1 |
| Polymerase Activation & DNA Denaturation | 5 minutes | 95 | 1 |
| Denaturation | 15 seconds | 95 | 40 |
| Annealing & Extension | 60 seconds | 60 | |
| Acquisition | x | x | |
| Melt-Curve Analysis | x | x | x |

**Table 3.12:** RT-qPCR parameters

| Gene | Input RNA (ng/uL) | Amplification Time |
|---|---|---|
| csrD | 1 | 120 minutes |
| dinD | 1 | 120 minutes |
| sulA | 1 | 120 minutes |
| umuD | 1 | 120 minutes |
| yebG | 0.1 | 60 minutes |

**Table 3.13:** NASBA parameters for *E. coli* RNA amplification

### 3.2.15 Paper-based Toehold Sensor Detection of NASBA Amplified *E. coli* RNA

A 2mM punch biopsy was used to cut paper discs (Whatman: 1442-042) into each well of a 384 black-walled plate (Corning: 3544) such that the unused wells and border wells were filled with water. Each toehold sensor reaction was assembled on ice following the reaction components in Table 3.14 using components from the NEB PURExpress in vitro Protein Synthesis Kit (NEB: E6800). 2$\mu$L of each reaction was spotted onto an individual paper disc, in duplicate or triplicate. The plate was sealed and immediately transferred to a plate reader set at 37°C for at least four hours. A fluorescent measurement was taken every five minutes.

| Component | Volume ($\mu$L) | Percentage |
|---|---|---|
| NEB Reaction Mix A | 0.8 | 40% |
| NEB Reaction Mix B | 0.6 | 30% |
| RNase Inhibitor (Roche: 3335399001) | 0.01 | 1% |
| Sensor (1.8 nM) | 0.15 | 8% |
| *E. coli* RNA/NASBA Amplified *E. coli* RNA | 0.334 | 17% |
| RNase-free water | 0.106 | 5% |

**Table 3.14:** Reaction components for cell-free toehold sensor detection of *E. coli* RNA

### 3.2.16 Best Practices for Working with RNA Samples

The following best practices were used for all of the above methods employing RNA. Prior to each experiment, all surfaces and pipettes were cleaned with RNase AWAY Surface Decontaminant (ThermoFisher: 7000TS1). Filter tips were used as much as possible to reduce contamination by aerosols. RNase-free PCR tubes with caps were used for the set-up of all RT-qPCR, NASBA, and toehold sensor reac-

tions, and care was taken to leave extra space between tubes and keep caps closed to prevent cross-contamination. RNA samples were frequently re-quantified with the Qubit BR assay, as well as probed for their integrity using the RNA IQ assay (Invitrogen: Q33221). All RNA samples were stored at -80°C in RNase-free water, and were either sub-aliquoted or limited in their free-thaw cycles.

### 3.2.17 Data Analysis for Paper-based Toehold Sensor Reactions

Custom python notebooks were used for data analysis and visualization. Raw fluorescent measurements were first smoothed using a three-point moving average. Each measurement was then normalized to its corresponding 'Negative Control' (no sensor, no trigger). These background adjusted measurements were then adjusted to zero such that there were no remaining negative values. To calculate the relative maximum GFP for each sensor-trigger pair, an average was computed on the maximum value of the sensor-trigger pair divided by the maximum value of the corresponding sensor-no trigger pair (labelled as 'GFP' on plots for parity). For the toehold sensor detection of *E. coli* RNA, a ciprofloxacin transcript response (CTR) value was calculated by normalizing either the NT signature gene relative maximum GFP values or the Cipro signature gene relative maximum GFP values from their respective NT or Cipro control gene relative maximum GFP values, and then calculating the fold change by dividing the normalized Cipro relative maximum GFP values from the normalized NT relative maximum GFP values. The CTR is then computed as the $log_2(\text{Fold Change})$ (Bhattacharyya et al. (2019)).

## 3.3 RESULTS

### 3.3.1 Transcriptional Signatures Associated with Fluoroquinolone Susceptibility in *E. coli*

As described in Chapter 1, brief antibiotic exposure can elicit stereotypical transcriptional responses in antibiotic susceptible cells. A panel of clinical isolates of *E. coli* were used to identify ten genes that are differentially expressed between fluoroquinolone susceptible and resistant *E. coli* in response to brief (30 minutes) exposure to three fluoroquinolones: ciprofloxacin, levofloxacin, moxifloxacin (Martinsen et al. (2021), Bhattacharyya et al. (2019)). Fluoroquinolones are chemotherapeutic bactericidal antibiotics and act by interfering with DNA replication. The differentially expressed genes are shown in (Figure 3.2 and Table 3.15). A set of control genes were also identified (Table 3.16).

| Gene Name | Gene ID | Encodes |
|---|---|---|
| dinD | b3645 | DNA damage-inducible protein D |
| dinI | b1061 | DNA damage-inducible protein I |
| ilvD | b3771 | dihydroxy-acid dehydratase |
| lexA | b4043 | DNA-binding transcriptional repressor LexA |
| recA | b2699 | DNA recombination/repair protein RecA |
| recN | b2616 | DNA repair protein RecN |
| rmuC | b3832 | putative recombination limiting protein RmuC |
| sulA | b0958 | cell division inhibitor SulA |
| umuD | b1183 | DNA polymerase V protein UmuD |
| yebG | b1848 | DNA damage-inducible protein YebG |

**Table 3.15:** List of ten genes in the *E. coli* fluoroquinolone susceptibility transcript signature

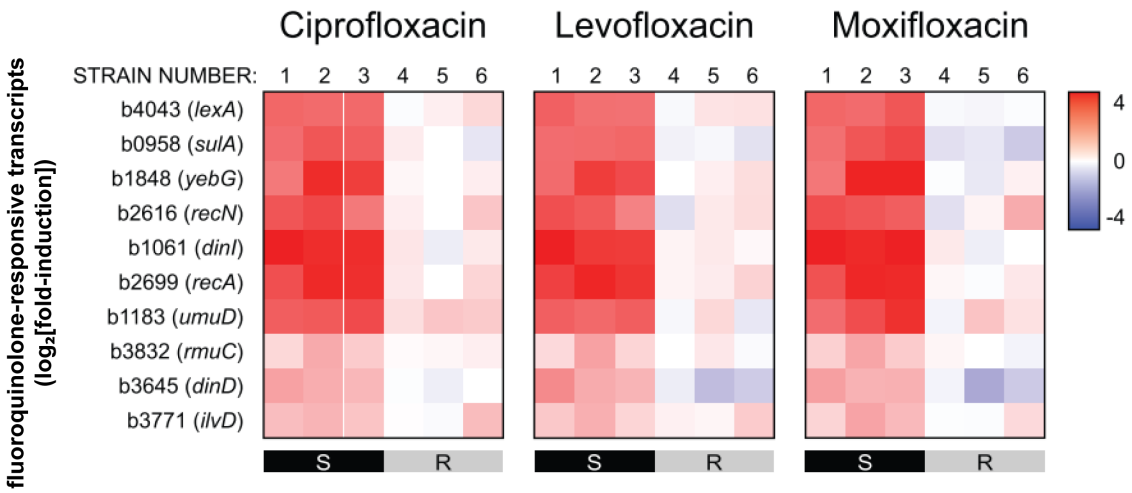



**Figure 3.2:** Differential expression of a fluoroquinolone-induced 10-transcript gene signature in *E. coli*

Figure adapted from Martinsen et al. (2021). RNA-Sequencing was performed and heatmaps of the normalized, log-transformed fold induction (treated/untreated) data show the differential expression of a fluoroquinolone-induced 10-transcript signatures for three fluoroquinolone drugs across six strains of *E. coli* that are either susceptible to the antibiotic ('S') or resistant ('R').

| Gene Name | Gene ID | Encodes |
|---|---|---|
| csrD | b3252 | regulator of CsrB and CsrC decay |
| kdgR | b1827 | DNA-binding transcriptional repressor KdgR |
| rppH | b2830 | RNA pyrophosphohydrolase |

**Table 3.16:** List of control genes in the *E. coli* fluoroquinolone susceptibility transcript signature

### 3.3.2 Gene Ontology Enrichment Analysis of the Fluoroquinolone Susceptibility Signature

To uncover the significance of the biological functions of this gene set, gene ontology (GO) enrichment analysis was performed on the ten genes listed in (Table 3.15. The set of significantly shared GO terms that are over-represented in the list were

visulized using NaviGO (Figure 3.3). As expected, the biological functions of the genes in the signature are associated with cell stress and response to stimuli. Interestingly, this signature seems to be composed of three hubs of GO terms that span different niches of cellular response to an antibiotic. The first are GO terms broadly associated with the extracellular response to a stimulus. The second are GO terms associated with the intracellular response to a stimulus. And the third are GO terms specifically associated with the DNA response to a stimulus. This suggests that the gene signature is composed of both broad stress response genes, as well as fluoroquinolone-specific response genes.

### 3.3.3 Identifying a Panel of Clinically Relevant *E. coli* Strains for AST Platform Development and Deployment

From a panel of *E. coli* strains (Galardini et al. (2017)), six *E. coli* strains were predicted by staramr to have an "I/R" (Intermediate/Resistant) phenotype to ciprofloxacin with 100% identity (Table 3.17). Five of these strains are pathogenic *E. coli* strains, four of which are UPEC strains. In order to build a proof-of-concept toehold switch sensor platform for detecting fluoroquinolone susceptible *E. coli*, a set of three susceptible and three intermediate/resistant *E. coli* strains were selected to use as a test bed. For the remaining of this chapter, this set of six *E. coli* strains were used for all experiments.

### 3.3.4 Set of *E. coli* Strains have Varying Ciprofloxcain Susceptibilities

The minimum inhibitory concentration (MIC) of each *E. coli* strain was experimentally determined using the broth dilution method. The MIC was estimated by fitting the growth curve data to a modified Gompertz Equation 3.1 and calculat-

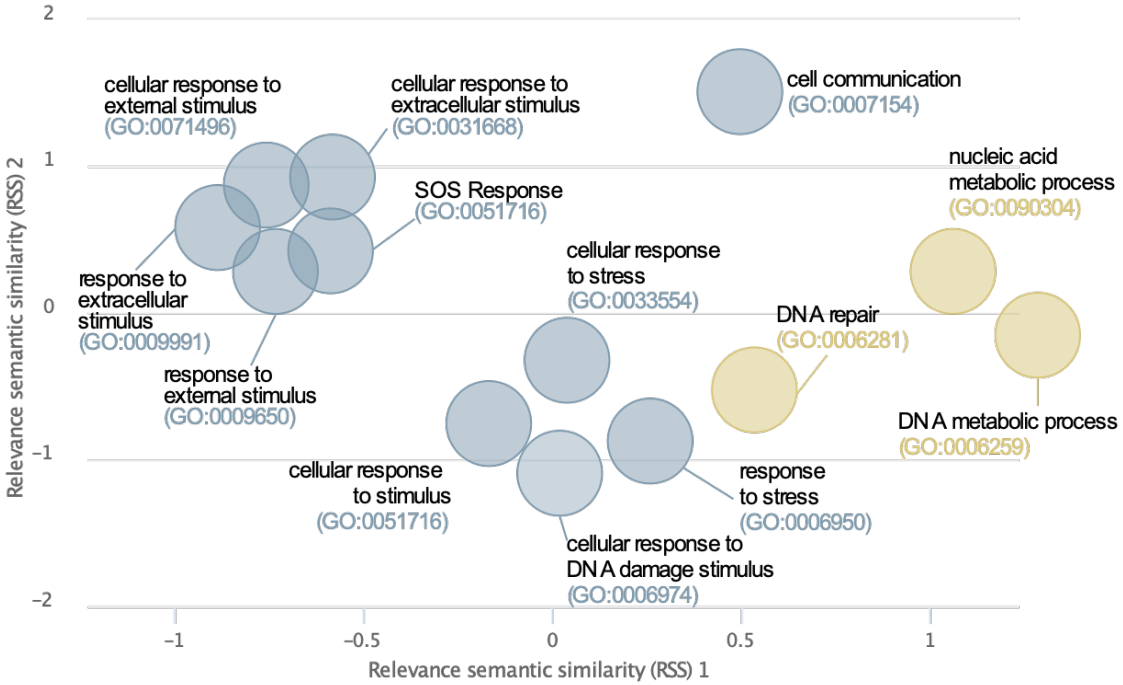



**Figure 3.3:** Visualization of Gene Ontology Terms for 10-transcript gene signature in *E. coli*

Visualization of the significant gene ontology (GO) terms associated with the 10-transcript signature associated with fluoroquinolone susceptibility in *E. coli*. GO terms are plotted by their computed relevance semantic similarity scores (RSS). GO terms in blue are associated with SOS response; GO terms in yellow are associated with DNA repair.

ing the MIC with Equation 3.2, and the resulting value was compared to the CLSI break-points to determine susceptibility or resistance. The MIC results confirmed the resistance phenotypes predicted by staramr (Figure 3.4). Thus, resulting in a set of *E. coli* strains with known genotypic and phenotypic ciprofloxacin susceptibility.

| Strain Name | Predicted Phenotpye (staramr) | Type | Source |
|---|---|---|---|
| MG1655 | ciprofloxacin susceptible | N/A | Lab strain |
| NILS 30 | ciprofloxacin susceptible | Pathogenic | Urine |
| NILS 33 | ciprofloxacin susceptible | Pathogenic | Urine |
| NILS 55 | ciprofloxacin I/R | Pathogenic | Unlisted |
| NILS 64 | ciprofloxacin I/R | Pathogenic | Urine |
| NILS 78 | ciprofloxacin I/R | Pathogenic | Urine |

**Table 3.17:** *E. coli* strains used for development of toehold sensor platform

### 3.3.5 Set of *E. coli* Strains Have Expected Transcriptional Signatures in Response to Brief Ciprofloxacin Exposure

Although this set of *E. coli* strains showed a growth-based phenotype of ciprofloxacin susceptibility, I next sought to determine if these strains also exhibit the predicted transcriptional signature of ciprofloxacin susceptibility. To probe this signature in this set of *E. coli* strains, I used RT-qPCR to measure the differential response to ciprofloxacin of the ten signature genes across the set *E. coli* strains. A comparison of MIC results and gene expression results are shown in Figure 3.5 for representative susceptible and resistant strains. Gene expression results for all strains are shown in Figure 3.6. A set of three control genes were used, and the normalized fold change values were comparable across each of them (Figure 3.7). The susceptible strains show differential expression of the ten signature genes, relative to each control gene. Conversely, the resistant strains do not show differential expression of the ten signature genes relative to each control gene. Interestingly, the intermediate strain (NILS 78) has differential expression of some, but not all, of the ten signature genes. Hierarchical clustering the fold change values separates the

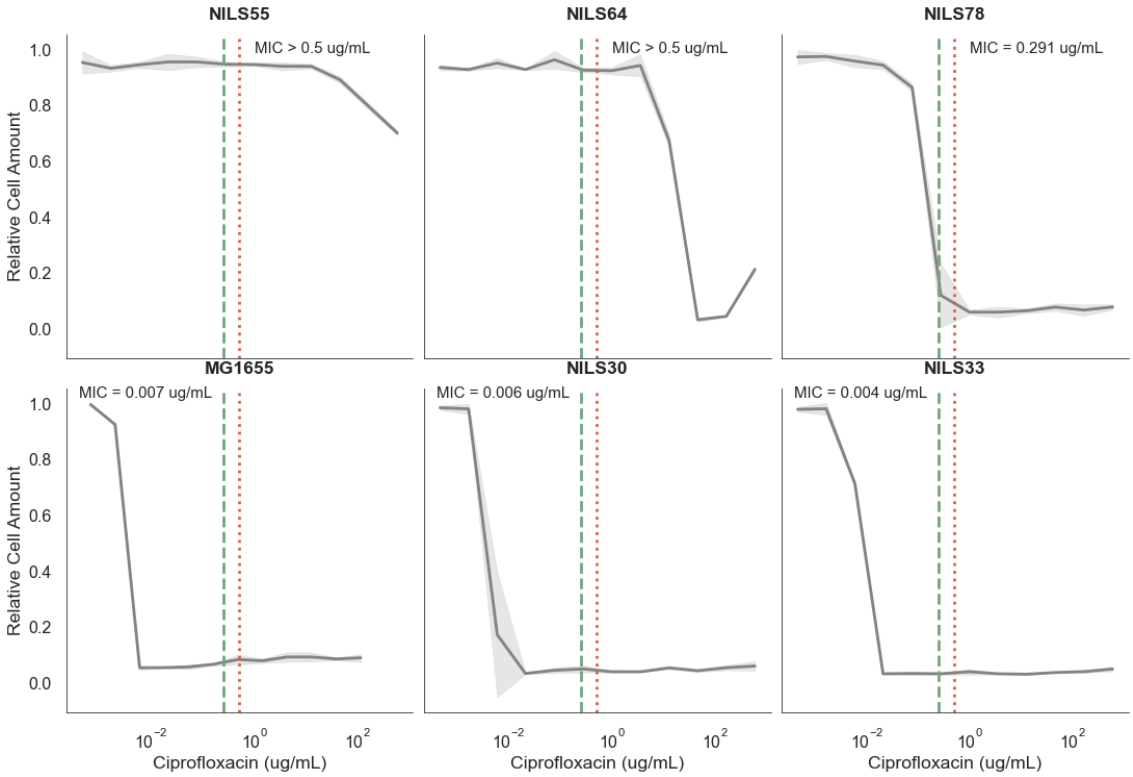



**Figure 3.4:** Minimum inhibitory concentration (MIC) measurements for a panel of six *E. coli* strains

Relative cell amount ($OD_600$ normalized to maximum value) v. concentration of ciprofloxacin in a serial dilution of ciprofloxacin exposure. Green vertical line (left) corresponds to CLSI threshold for susceptibility; red vertical line (right) corresponds to CLSI threshold for resistance. MIC is estimated using a modified Gomptertz function (3.1)

resistant strains from the susceptible strains.

### 3.3.6 A Panel of Toehold Switch Sensors for Detecting Ciprofloxacin Susceptible *E. coli* are Specific

A library of toehold sensors were designed, built, and screened as described in Chapter 2, resulting in one toehold sensor per gene in the fluoroquinolone response signature. To demonstrate the specificity of the toehold sensor panel, each

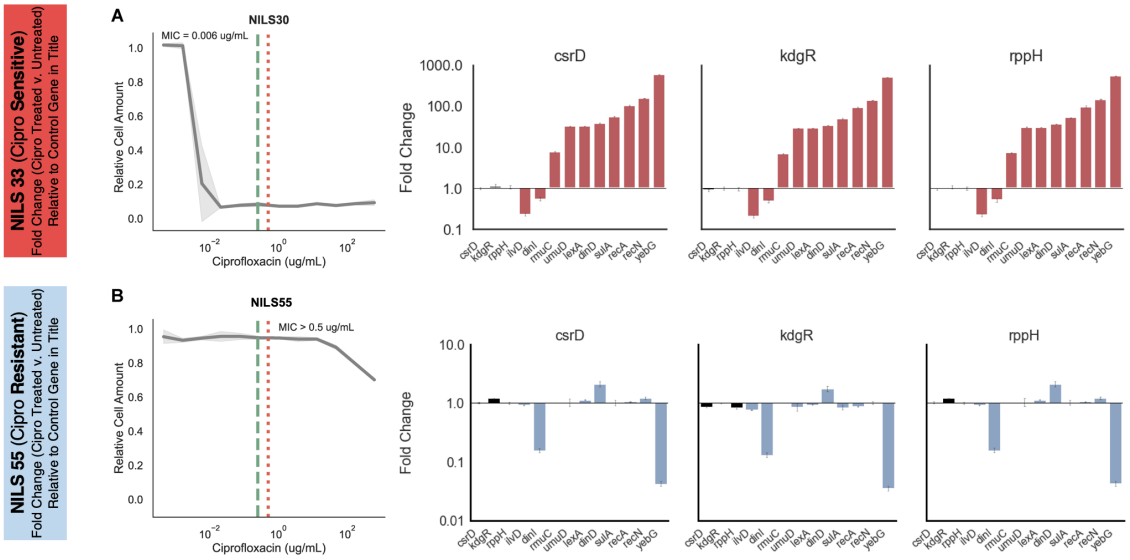



**Figure 3.5:** Comparing the MIC to gene expression (RT-qPCR) of signature genes in two *E. coli* strains.

(A): Susceptibility of NILS30 to ciprofloxacin via broth dilution assay (left) and gene expression across panel of signature genes (right). (B) Resistance of NILS55 to ciprofloxacin via broth dilution assay (left) and gene expression panel of signature genes (right). Gene in title is control gene.

toehold sensor was deployed against each 35nt trigger RNA in paper-based, cell-free reactions. Each toehold sensor is specific to its intended RNA target and showed little-to-no off-target activation (Figure 3.8).

### 3.3.7 Optimization of the NASBA Workflow for Context-Dependent Differential mRNA Detection

NASBA has typically been employed to amplify RNA for downstream applications where the readout is binary: 'no' signal v. 'yes' signal. For those applications, the input RNA is amplified such that the greatest amount of amplicon can be produced for subsequent RNA detection. This is achieved through running the amplification reaction to completion (*i.e.*, for more than two hours). Other work

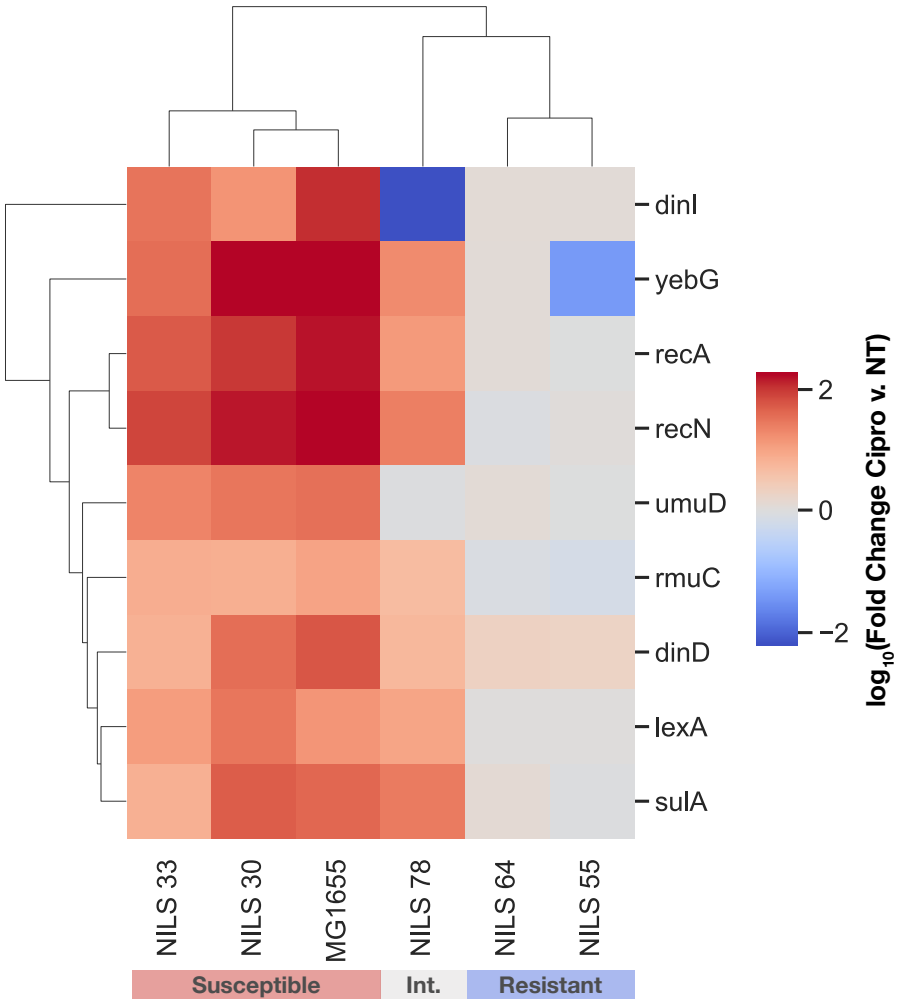
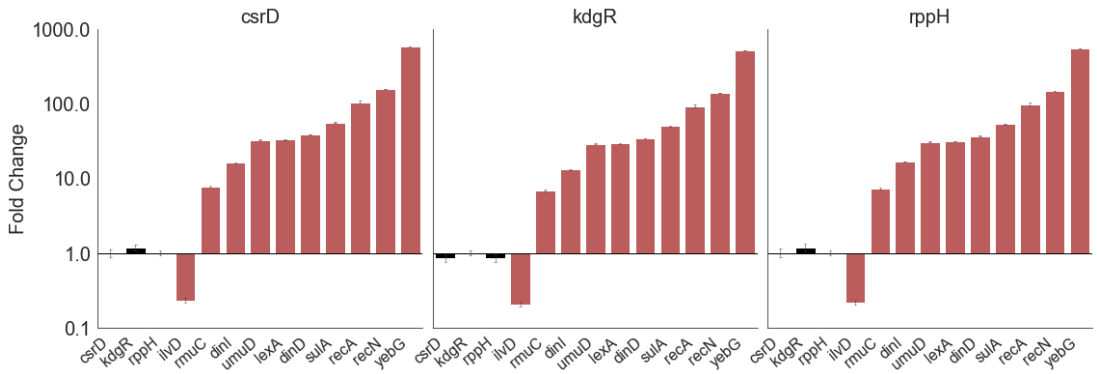


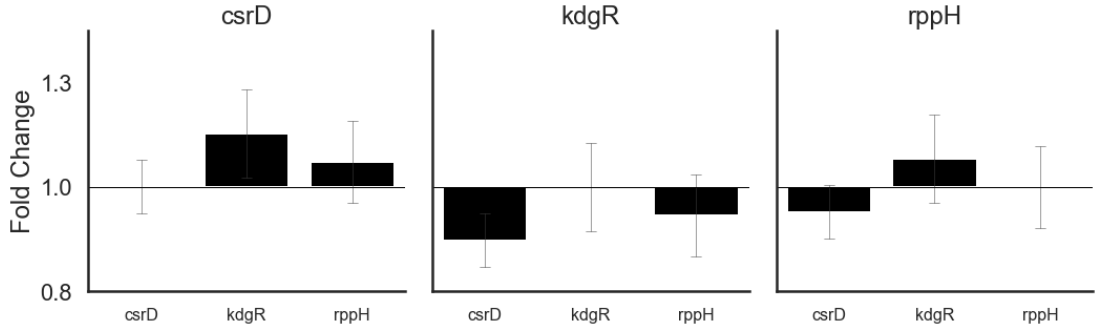

**Figure 3.6:** Gene expression (RT-qPCR) of signature genes in set of *E. coli* strains

*E. coli* strains labelled as resistant, intermediate ("Int."), or susceptible as confirmed by broth dilution MIC.

has used NASBA for quantitative RNA detection, however this been done with *in vitro* transcribed RNA (Takahashi et al. (2018)). In order to use NASBA amplification followed by toehold sensor detection of mRNA from antibiotic treated cells, I sought to modify the traditional NASBA paradigm in the context of this AST workflow. A summary of the NASBA and toehold sensor detection workflow and




**(a)** Fold Change as determined by RT-qPCR of RNA isolated from NILS30 following brief exposure to ciprofloxacin. Plot headers designate the control gene used as reference to calculate Fold Change.




**(b)** Fold Change as determined by RT-qPCR of NILS30 brief exposure to ciprofloxacin for the control genes only.

**Figure 3.7:** Consistency of control gene expression by RT-qPCR of NILS30

points I identified as potential parameters to be tuned are described in Figure 3.9. These optimizations used both synRNA and *E. coli* RNA for context-dependent improvement of the workflow.

#### *3.3.7.1 Toehold Switch Sensors can Detect Equimolar Concentrations of Un-amplified RNA*

It has previously been shown that the GFP production from toehold sensor detection is proportional to different amounts of input trigger RNA, without NASBA

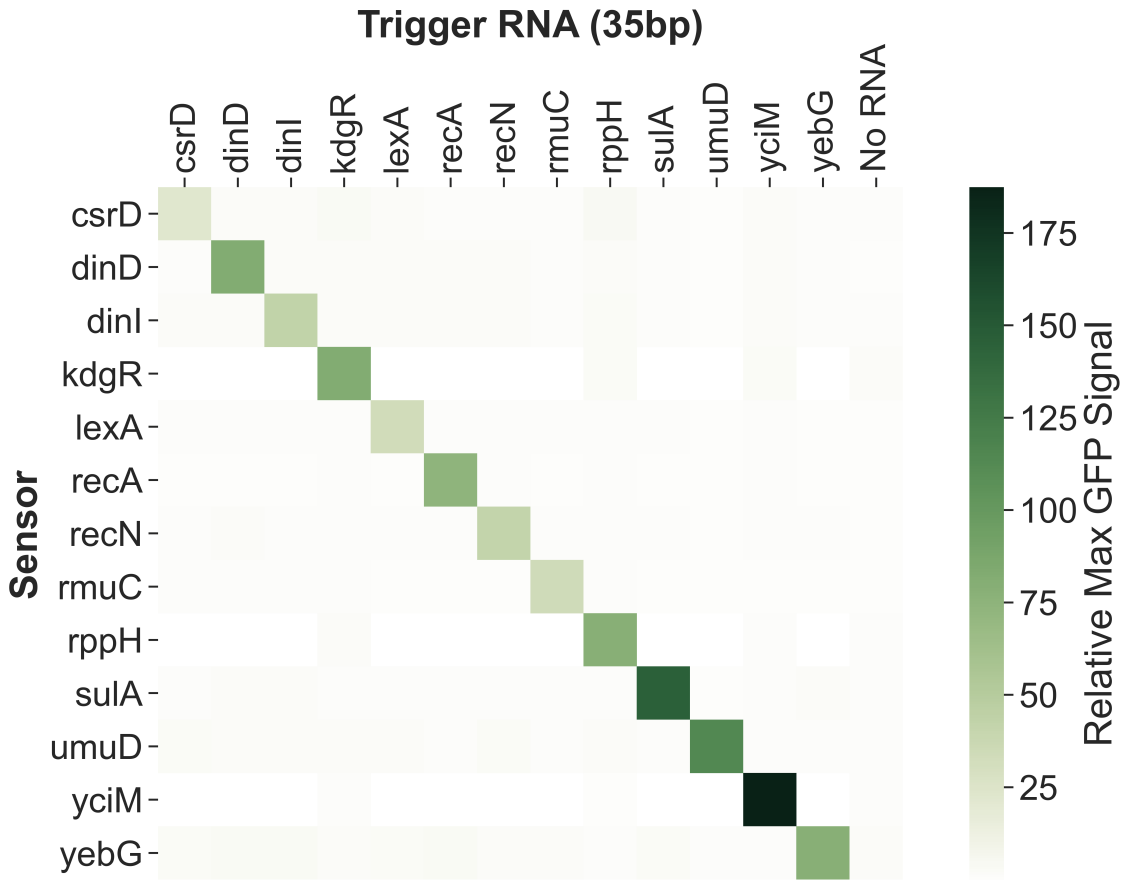
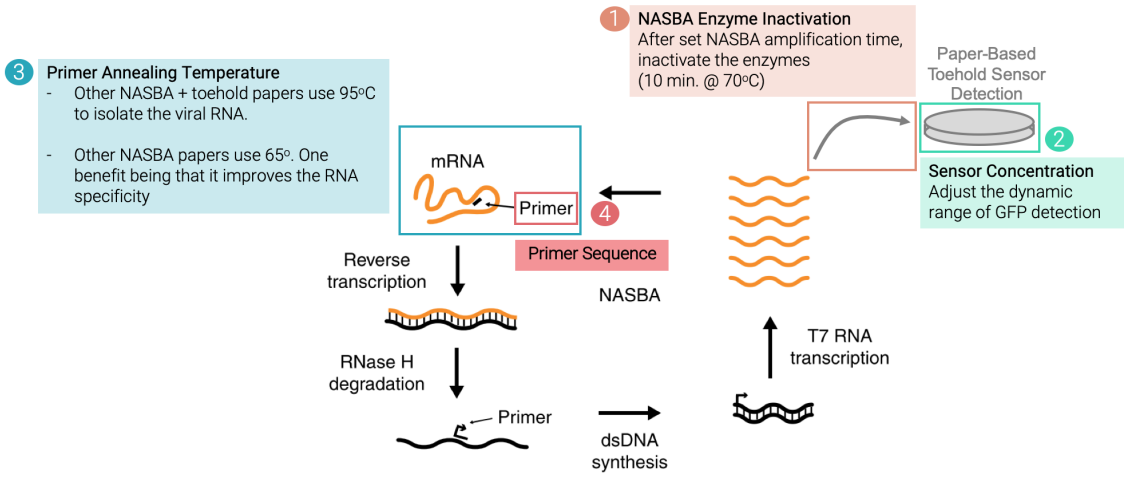



**Figure 3.8:** Specificity of the best toehold sensor for each gene in the fluoroquinolone response signature, and controls

Rows: Toehold sensors; Columns: 35nt trigger RNA, and a water control (NA); Color Scale: Mean relative maximum GFP signal acquired across technical replicates

amplification (Pardee et al. (2016), Takahashi et al. (2018)). To see if this was true for the 200nt synRNA, serial dilutions of equimolar amounts of trigger RNA and synRNA were detected with the toehold switch sensors. It has been demonstrated previously that the limit of detection for toehold sensor detection of RNA fragments (without NASBA) is in the nM range (Pardee et al. (2016)). Data for the detection of yebG trigger RNA and synRNA is shown in Figure 3.10. With the exception of the highest RNA concentration (9.62 $\mu$M), GFP production is proportional to the

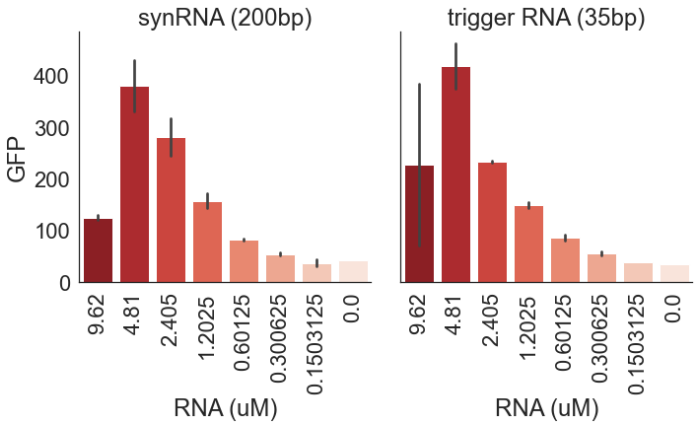
**Figure 3.9:** NASBA process and corresponding points of optimization



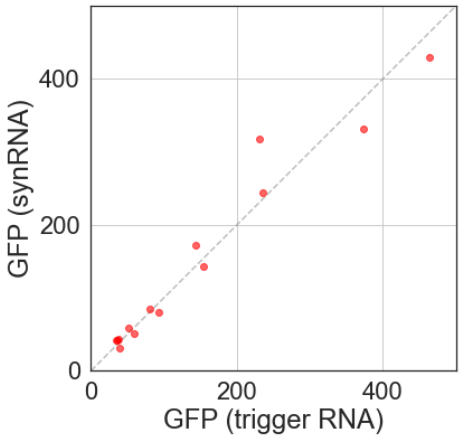

amount of input RNA for both trigger RNA and synRNA (Figure 3.10a). Moreover, there is a linear relationship between the GFP production of equimolar amounts of the two types of RNAs (Figure 3.10b). One hypothesis for the reduction in GFP production for the highest RNA concentration (9.62 $\mu$M) is that by there being too much RNA in the sample, there is too much toehold sensor activation, and therefore overwhelmed the translational machinery in the cell-free mixture.

### *3.3.7.2 Heat Inactivation of NASBA Enzymes Preserves Relative RNA Concentrations*

NASBA amplification is optimal at 41°C, but has been shown to have almost comparable activity at 37°C (Pardee et al. (2016)). However, the toehold sensor detection, which is run subsequent to the NASBA amplification, is optimized to run at 37°C. This overlap in activity has the potential for NASBA reactions to carry over from amplification to toehold sensor detection, and unintentionally run to completion. One strategy to reduce this unwanted phenomenon is to inactivate the enzymes present in the NASBA mixture. Theses enzymes and their corresponding




**(a)** Maximum GFP signal of toehold sensor detection of equimolar amounts of yebG trigger RNA (left) or yebG synRNA (right)




**(b)** Comparison of maximum GFP values from toehold sensor detection for equimolar amounts of yebG trigger RNA and synRNA

**Figure 3.10:** Toehold sensor detection of serial dilutions of equimolar amounts of trigger RNA and synRNA.

heat inactivation protocols are listed in Table 3.18. The effect of inlcuding a heat inactivation step following *E. coli* RNA amplification was compared to no heat inactivation by measuring the subsequent toehold sensor production of GFP. The heat inactivation protocol employed was a 10 minute incubation at 70°C, followed by immediately putting the reactions on ice. By including the heat inactivation

step, there is a more linear relationship between input RNA amount (pre-NASBA) and GFP production than when this step was omitted (Figure 3.11).

| Enzyme | Heat Inactivation Protocol | Citation |
|---|---|---|
| Avian Myeloblastosis Virus Reverse Transcriptase | 70-100°C, followed by a 5-minute incubation on ice | Ahokas & Erkkilä (1993) |
| Rnase H | 65°C for 20 minutes | https://www.neb.com/products/m0297-rnase-h#Product%20Information |
| T7 RNA Polymerase | 70°C for 10 minutes | https://www.nzytech.com/files/brochures/MB080_T7%20RNA%20polymerase.pdf |

**Table 3.18:** Enzymes present in the NASBA enzyme mixture and their recommended heat inactivation protocols

#### *3.3.7.3 The Panel of Toehold Switch Sensors for Detecting Ciprofloxacin Susceptible E. coli are Sensitive and Log-linear*

After the introduction of heat inactivation into the NASBA protocol, I next sought to determine the sensitivity each toehold switch sensor in detecting NASBA amplified synRNA. A range of synRNA concentrations were NASBA amplified for 60 minutes (before completion), followed by toehold sensor detection (Figure 3.12). All of the toehold switch sensors are able to detect NASBA amplified synRNA within log-linear regimes ($R^2 > 0.75$). These data suggest that toehold switch sensors can detect differential amounts of NASBA amplified RNA from the nM to fM range.

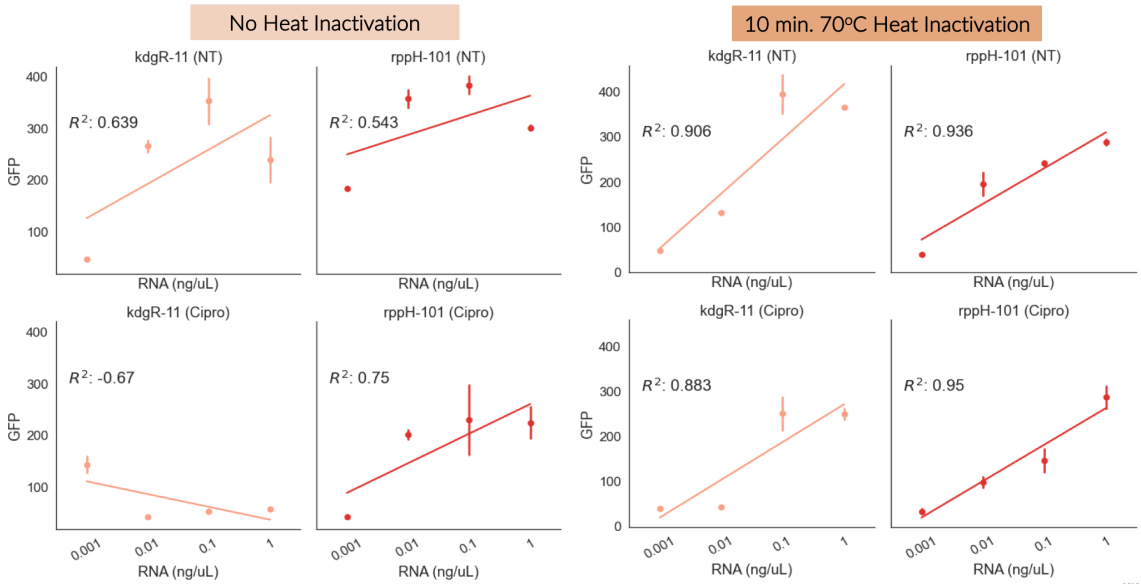



**Figure 3.11:** Heat inactivation of NASBA enzyme preserves relative *E. coli* RNA concentrations

Left: No heat inactivation; Right: 10 min. 70°C heat inactivation. RNA samples from both untreated (NT) and ciprofloxacin (Cipro) treated *E. coli*.

#### 3.3.7.4 *Toehold Switch Sensor GFP Production is Dependent on Input RNA Concentration to NASBA Amplification and Toehold Sensor Concentration*

Previous studies have used a toehold sensor concentration of 1.8nM for toehold sensor detection of NASBA amplified RNA. To see if this concentration was appropriate for the toehold sensor detection of NASBA amplified *E. coli* RNA for differential detection of signature genes, a screen of sensor concentrations and RNA concentrations (input to NASBA) was performed for dinD expression in the susceptible strain ECOR-28 after brief antibiotic exposure (Figure 3.13, Figure 3.14). The goal of these experiments was to determine which combination(s) of RNA concentrations and sensor concentrations yield a good dynamic range in the proposed AST context: differential detection of signature genes in Ciprofloxacin

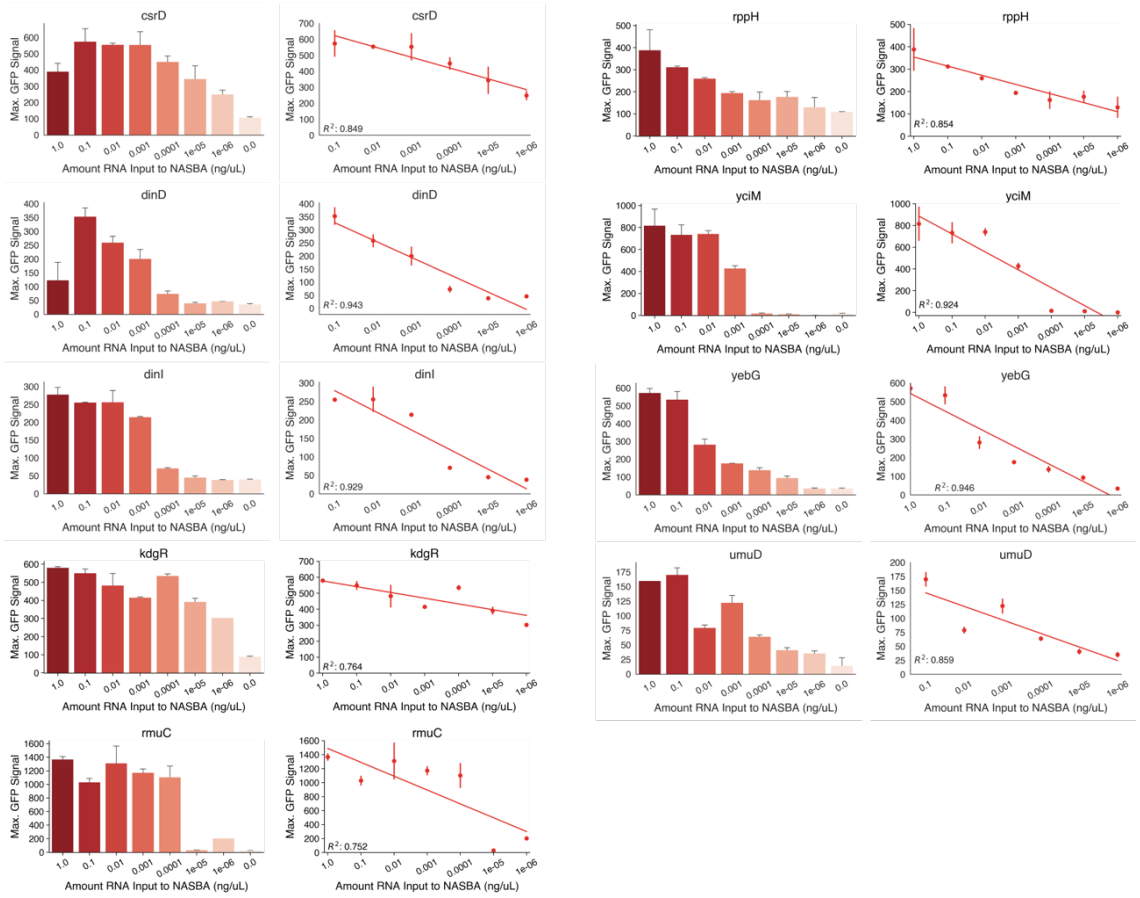



**Figure 3.12:** Sensitivity and concentration dependence of panel of toehold switch sensors in detecting NASBA amplified synRNA

(Left plots) GFP production of toehold sensor detection of NASBA amplified synRNA across a range of synRNA concentrations. (Right plots) Log-linear fit of GFP v. RNA input amount with $R^2$ value. (0.1ng/$\mu$L ~1.3nM synRNA)

treated cells versus untreated cells. Figure 3.13 shows the middle ranges of input *E. col* RNA concentrations tested (0.1ng – 10ng). Across the sensor concentrations, when 0.1ng/$\mu$L or 1ng/$\mu$L RNA is input into NASBA, there is differential detection of dinD in the Cipro treated versus untreated cells. The dynamic range is greatest when 1nM or 2nM sensor is used, and for 1ng/$\mu$L of input RNA. As RNA concentration increases to 100ng$\mu$L, there is no differential detection of RNA

and in fact, the GFP production is similar to when there is no input RNA (Figure 3.14). These results are further illustrated in Figure 3.15. 1nM – 2nM sensor concentration is appropriate for the toehold sensor detection of NASBA amplified *E. coli* RNA when the amount of RNA input to NASBA is within a certain range (for dinD, 0.1ng – 1ng). For both the untreated and Cipro treated samples, as RNA input concentration increases, so does the GFP production - until there is too much RNA in the sample (10ng/$\mu$L for Cipro treated, less than 100ng/$\mu$L for untreated). Since the brief antibiotic exposure increases the amount of the signature transcript, it is expected that these two samples would have different limits. Together, this data demonstrates that it is possible to identify a regime of NASBA parameters and toehold sensor parameters where differential transcript detection from NASBA amplified *E. coli* RNA can occur.

#### *3.3.7.5 Tuning RNA Levels through RNA Input Amounts and NASBA Amplification Time*

For end-point reactions, NASBA is typically run for at least two hours. However data for the detection of *E. coli* 16S mRNA suggested that tuning the amplification time can modulate the amount of resulting amplicon (Takahashi et al. (2018)). To identify the NASBA amplification times that preserve relative amounts of input RNA, a combination of experiments with synRNA and *E. coli* RNA were used. Since this variable is likely linked to the amount of input RNA, these two parameters were tested together. Thus, these experiments scanned the space of amplification times and input RNA concentrations for both synRNA and *E. coli* RNA. Ultimately, the parameters that lead to differential detection of synRNA are not always the same for *E. coli* RNA, so their respective parameters are used for the

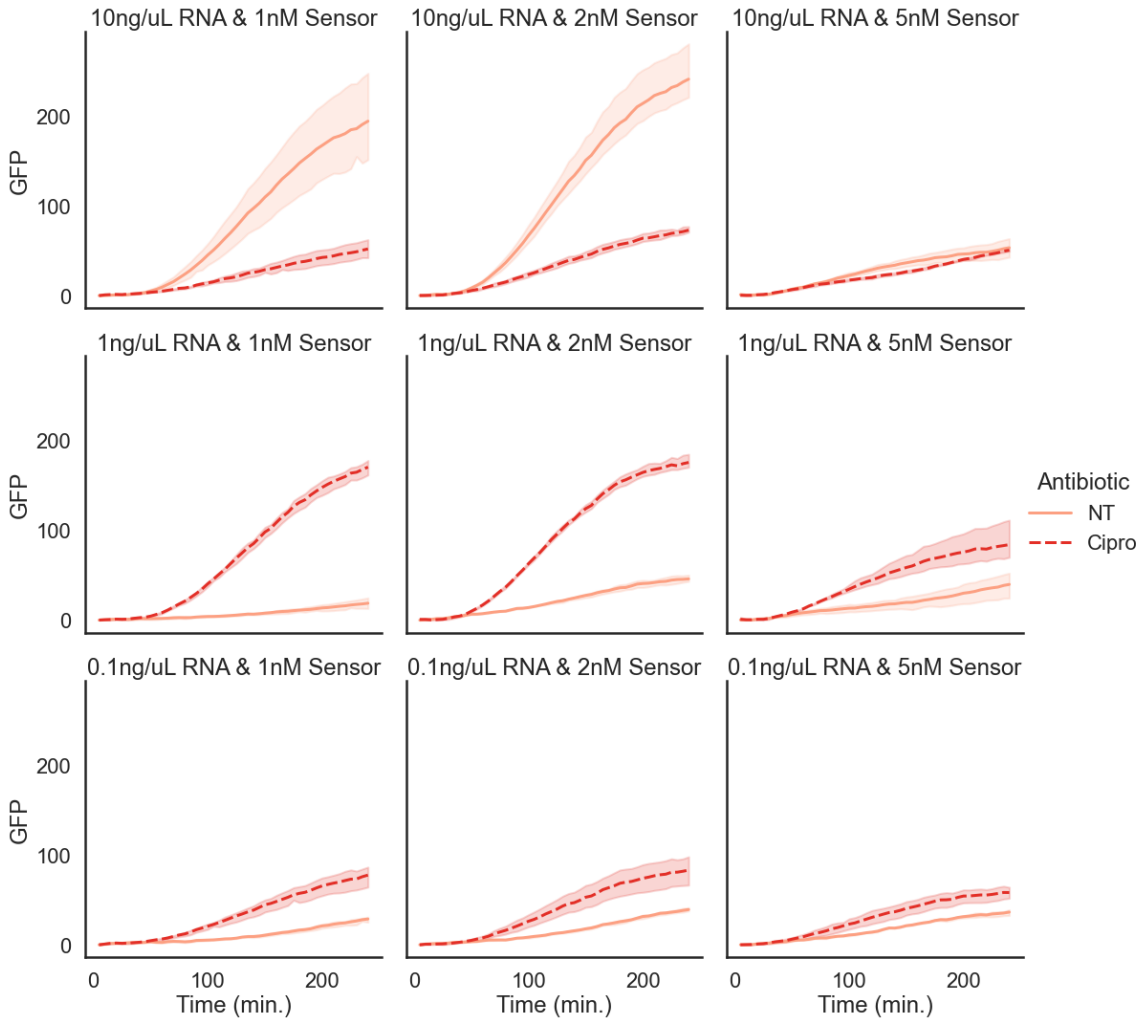



**Figure 3.13:** Optimization of *E. coli* RNA concentration and sensor concentration for NASBA amplification and toehold sensor detection

Time-course of GFP production from toehold sensor detection of NASBA amplified (60 minutes) RNA from *E. coli* strain ECOR-28 (susceptible) after brief ciprofloxacin exposure.

respective experiments. For synRNA, amplification times from 30 minutes to 60 minutes preserve relative input RNA amounts. For total *E. coli* RNA, amplification times from 60 minutes to 120 minutes preserve relative input RNA amounts.

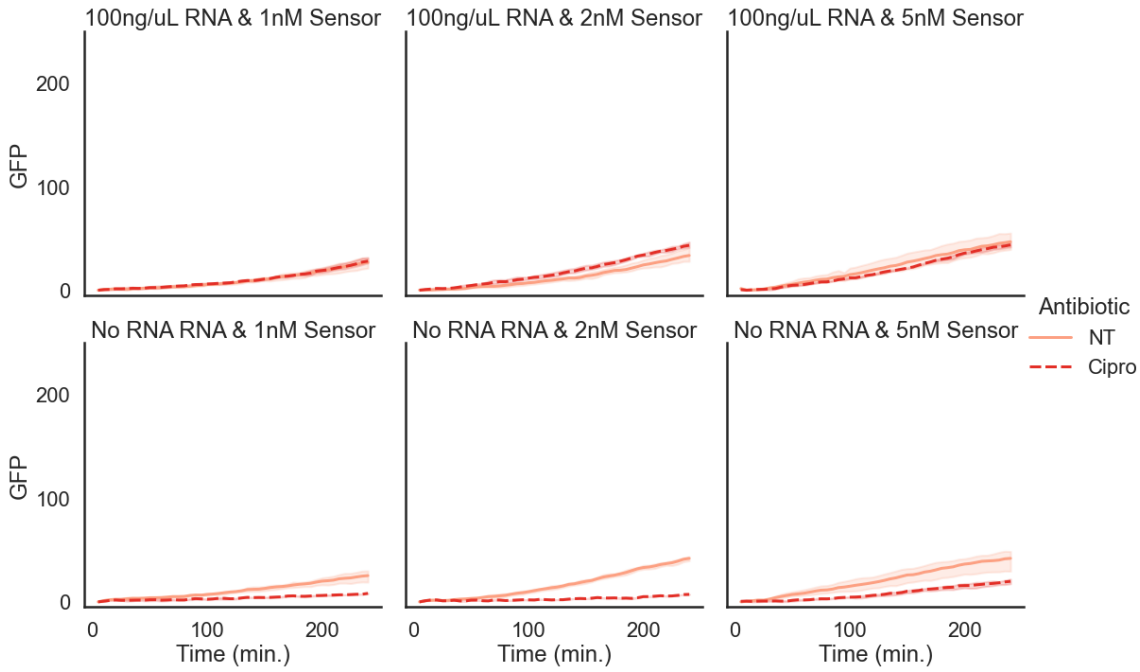
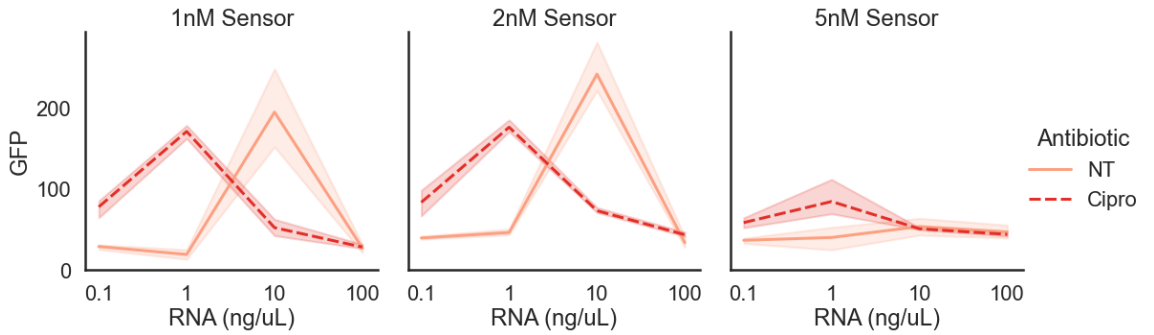



**Figure 3.14:** Optimization of *E. coli* RNA concentration and sensor concentration for NASBA amplification and toehold sensor detection

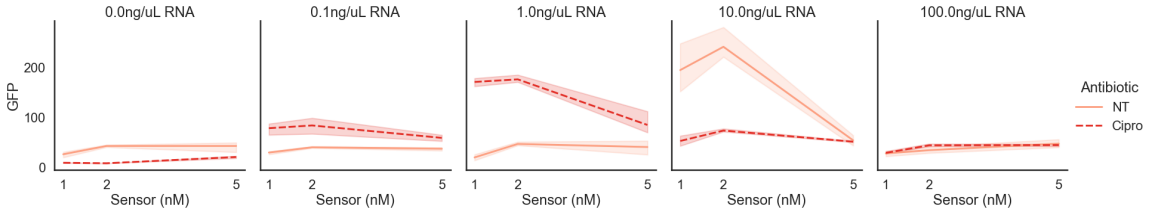
Time-course of GFP production from toehold sensor detection of NASBA amplified (60 minutes) RNA from *E. coli* strain ECOR-28 (susceptible) after brief ciprofloxacin exposure.

### 3.3.8 Paper-Based Toehold Sensor Detection of Ciprofloxacin Susceptibility in a Panel of *E. coli* Strains

Finally, I demonstrate the complete workflow (Figure 3.1) for detecting ciprofloxacin susceptible *E. coli* via toehold switch sensor detection of differentially expressed signature transcripts. The set of *E. coli* strains (Table 3.17) were briefly exposed to Ciprofloxacin to elicit the transcriptional response, and RNA was isolated and NASBA amplified using the previously described optimized workflow. NASBA amplified RNA was detected on paper-based reactions with toehold switch sensors for the *E. coli* fluoroquinolone signature and a control gene (csrD), and GFP measurements were used to calculate a ciprofloxacin transcript response (Figure 3.16).




**(a)** GFP production per toehold sensor concentration. 1nM and 2nM sensor show similar response to input RNA concentrations; 5nM sensor has reduction in response.




**(b)**

**Figure 3.15:** Optimization of *E. coli* RNA concentration and sensor concentration for NASBA amplification and toehold sensor detection

This workflow for paper-based toehold sensor detection of differentially expressed signature genes was able to measure increased expression of the signature genes in the susceptible strains (NILS33, NILS30, MG1655), and no change in expression or decreased expression of the signature genes in the resistant strains (NILS55, NILS64). As was the case in the RT-qPCR results, toehold sensor detection of signature genes in the intermediate strain (NILS78) shows up-regulation of some, but not all signature genes. Hierarchical clustering of the ciprofloxacin transcript response values groups the resistant strains seprate from the susceptible strains. The intermediate strain (NILS78) is grouped with NILS33, which is the least susceptible, susceptible strain. Future work will provide optimized parameters for the remaining genes in the signature, which will likely improve the interpretation of results from intermediate strains. Nevertheless, susceptible strains are distinctly

identified by this method.

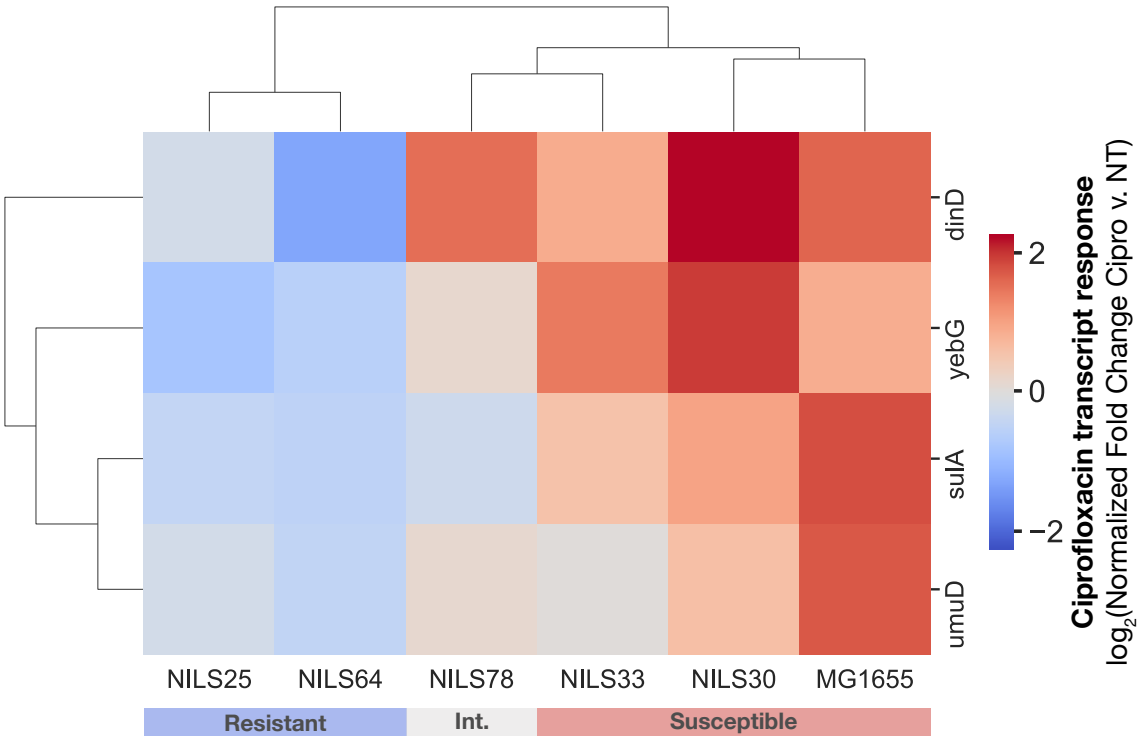



**Figure 3.16:** Paper-based toehold sensor detection of ciprofloxacin susceptibility in a panel of *E. coli* strains.

*E. coli* strains were briefly exposed to ciprofloxacin, and RNA was isolated and amplified with NASBA following the gene-specific parameters in Table 3.13. Color bar is the ciprofloxacin transcript response: normalized ciprofloxacin versus untreated signal.

Although this method is not meant to be as quantitative as RT-qPCR, it is useful to see how the ciprofloxacin signature response as determined by RT-qPCR compares to the signature response as determined by this paper-based toehold sensor platform. The log transformed Fold Change values were compared for the genes present in both assays (Figure 3.17). There is relatively good correlation between the Fold Change as determined by the two, distinct, methodologies.

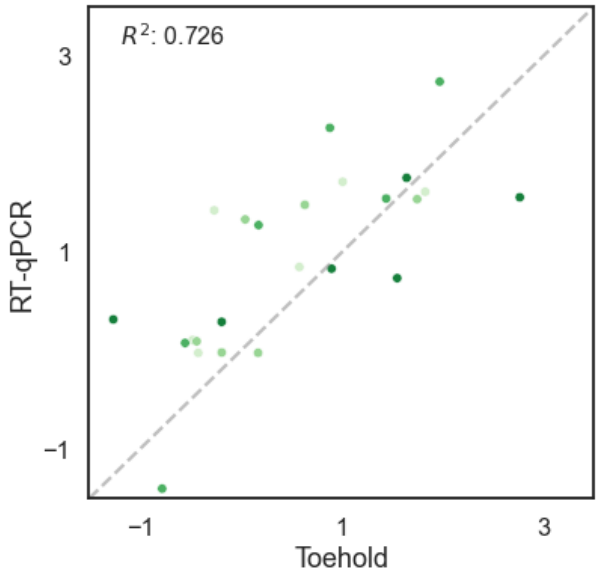



**Figure 3.17:** Correlation of Ciprofloxacin transcript response as determined by RT-qPCR and the paper-based toehold sensor platform

Toehold: $log_2$Fold Change; RT-qPCR: $log_{10}$Fold Change. For both toehold sensor detection and RT-qPCR, fold change is calculated as Cipro v. NT for signature genes normalized by a control gene.

## 3.4 DISCUSSION

Here I have introduced a first-of-its-kind toehold switch sensor workflow for detecting differential RNA signatures of antibiotic susceptibility, and deployed this assay on a set of clinically relevant *E. coli* strains. I characterized a set of clinically relevant *E. coli* strains with different genotypic and phenotypic ciprofloxacin susceptibility profiles. Using the work done in Chapter 2, I was able to identify a set of toehold switch sensors with the greatest efficacy for each gene. These toehold switch sensors can differentially detect equimolar amounts of short 35nt trigger RNA and 200nt synRNA. I explored numerous optimizations to identify the parameters that optimize the context-specific toehold sensor detection of NASBA amplified *E. coli* RNA. Toehold switch sensors can differentially detect NASBA am-

plified synRNA from the nM to fM range when NASBA is not run to completion. NASBA can be kept from running to completion by adding a heat inactivation step and reducing the amplification time. Toehold switch sensors can detect differential amounts of NASBA amplified *E. coli* RNA within gene specific regimes of sensor concentration, NASBA amplification time, and input RNA amount. Together, these parameters were used to detect ciprofloxacin susceptibility in a set of *E. coli* strains, with results comparable to RT-qPCR. Future work could deploy this workflow on additional *E. coli* strains and in response to additional fluoroquinolone drugs such that statistically significant models can be made to translate toehold switch sensor GFP production to a classification of susceptibility or resistance.

## 3.5 SUPPLEMENTARY TABLES

**Table 3.19:** Primers used for trigger RNA and toehold switch sensor generation

| Name | Primer Sequence |
|---|---|
| Sensor-Trigger-Generation-F | GCGCTAATACGACTCACTATAGGG |
| Sensor-Linker-Generation-R | CATCTTTTGCGCTGCCG |
| Sensor-GFP-Generation-R | TCTCAAATGCCTGAGGTTTCAG |
| ds-T7 | GCGCTAATACGACTCACTATAGGG |

**Table 3.20:** List of primers for RT-qPCR of *E. coli* fluoroquinolone susceptibility signature

| Gene | Forward Primer | Reverse Primer | Type of Response |
|---|---|---|---|
| csrD | CTGGGCGGCCATTCATAGAT | CATTCGTTGATGACCACCGC | Control |
| dinD | GCTATTCATCAGCGGAAGGG | CATGGTTCCGCCAAGTTCCT | Control |
| dinI | AAACCAGTCATCGGCGCTTT | GCGAAAACTTCTCCATTGCCA | Control |
| ilvD | CTATGAAGGCCCGAAAGGCG | TCGATAGCGATCAGGTCACCA | Control |
| kdgR | TTTCTTTGGTCAGGCGGGAG | CAGCGCGTCATGATGTCAAA | Signature |
| lexA | AATTGTTTCCGGCGCATCAC | TCTTTCATCGACATCCCGCT | Signature |
| recA | GTCAGCGTGGTTTTACCGGA | CAGCACTGGGCCAGATTGA | Signature |
| recN | TCACCAAACCTGAGCACCAA | TCTCCGGGCTGCGGATTAAA | Signature |
| rmuC | CCGCTACGTGAACAACTGGA | CGTCAATACTACCTCGCCCC | Signature |
| rppH | GCAACCAGTTACGCGTTGAA | ACCGCCCAAACGTAGGTATC | Signature |
| sulA | ACTGTAATTGCCCGTGCGTA | GCCCATGATGACGCAACTTC | Signature |
| umuD | ATCCCAGCGCGACTTACTTC | TACGCGCTGTTCATGGGAAT | Signature |
| yciM | TGGTATATGGGCCGCAGAAG | AGGTTTCCGAGCGTAAGGTG | Signature |
| yebG | GACCTGTGGGGACGGTAATTT | GCAAAAAGGAAGCCGATGCT | Signature |

# CHAPTER 4

# Transcriptomics of the Evolution of Antibiotic Resistance

## 4.1 INTRODUCTION

As described in Chapter 1, widespread global antibiotic use has led to increased rates of antibiotic resistance, making antibiotic resistant infections a major global health concern. In an effort to slow the rate of antibiotic resistance acquisition, it is critical that we improve our antibiotic stewardship efforts and reduce unnecessary antibiotic use. Still, antibiotic resistance will always be a feature of antibiotic usage, and as such, it is critical that we have methodologies to study it.

When susceptible bacteria are under external antibiotic pressures, they experience biological stress due to cell membrane damage, changes in protein synthesis, or alterations in DNA supercoiling. These stress conditions increase the amount of mutational events, leading to an increase in the rate of resistant variants emerging in the population (Baquero & Blázquez (1997)). Over the past century, consistent and widespread environmental antibiotic pressures have continuously allowed for the emergence of bacterial variants with numerous phenotypes and genotypes of resistance.

As resistant variants continue to emerge, there is a need to be able to identify both the genetic and phenotypic markers of resistance. The majority of existing efforts have been retroactive: identifying specific mutations, genes, or genotypes of bacteria that have observable phenotypes of resistance. However, a more challenging phenomenon to study is the coupling between genetic or phenotypic markers and the evolution of antibiotic resistance. As was described in Chapters 2 and 3, transcriptional signatures are an observable phenotype of antibiotic suscepti-

bility. In this Chapter, I investigate if we can use gene expression as a marker of antibiotic resistance emerging in a population. First, I provide a brief summary of antibiotic resistance evolution and how it has been studied in *in vitro*. Second, I introduce a framework for utilizing long-term evolution experiments to control the evolution of antibiotic resistance and study how a strain's transcriptome changes as resistance increases. Third, I provide preliminary results for the lab evolution of tetracycline resistance in *E. coli*, and the transcriptional changes observed. Together, this work provides a framework for future studies aimed at gleaning the connection between antibiotic resistance evolution and gene expression.

### 4.1.1 Evolution of Antibiotic Resistance

The evolution of bacteria to evade antibiotics is not a unique phenomenon; bacteria are extremely adept to evade numerous environmental challenges, including extreme pH, extreme temperature, and nutrient depletion (Rodríguez-Rojas et al. (2013)). Widespread antibiotic use has allowed for bacteria to develop mechanisms to evade antibiotics through many mechanisms, including changes in cell permeability, altering the antibiotic target, and horizontal gene transfer (andersson 2003, davies 1994, livermore 2003).

The classical view of antibiotic resistance evolution is that prolonged antibiotic treatment selects for pre-existing antibiotic resistant variants. This can occur when genetic mutations that confer antibiotic resistance spontaneously arise during replication, and antibiotic treatment will fortuitously select for them. This type of spontaneous resistance is of particular concern in infections that are treated with monotherapies or get insufficient amount of antibiotics, such as tuberculosis and gonorrhoea (Llewelyn et al. (2017)).

Beyond spontaneous mutations, antibiotics themselves can be mutagenic. The general stress response that they induce can increase mutation rates, but also mutations can be induced from oxidative damage (*e.g.*, by bactericidal antibiotics), the SOS response (*e.g.*, from sub-inhibitory concentrations), and nucleotide-pool unbalancing (*e.g.*, by folic acid antagonists) (Kohanski et al. (2007), Baharoglu & Mazel (2011), Genther et al. (1977)). Horizontal gene transfer (HGT), or the lateral transfer of resistance genes between organisms, is another major mechanism for the acquisition of resistance. Although HGT can happen spontaneously, some antibiotics can increase the rates of HGT and antibiotic damage can promote the mobilization of genetic elements (Rodríguez-Rojas et al. (2013)).

### 4.1.2 Methods for *in vitro* Antibiotic Resistance Evolution

Lab evolution experiments allow for time-resolved measurements, making them well-suited for studying the evolution of antibiotic resistance. Some of the most well-known long-term evolution experiments are those carried out by Lenski, *et. al.* (Lenski et al. (1991), Lenski (2004)). In this batch culture set-up, bacterial populations are manually propagated in specific conditions (*e.g.*, nutrients, temperature, antibiotics), and samples are stored throughout the course of the experiment. These batch culture evolution experiments are relatively easy to implement, and have been used to generate experiments with samples that have evolved over 50,000 generation times (Grant et al. (2020)). Evolved strains can be compared to their ancestral strains to uncover the genetic links behind evolvability and antibiotic resistance phenotypes (Card et al. (2019), Card et al. (2021)).

The MEGA-plate (microbial evolution growth arena) is an experimental device that allows for evolution of antibiotic resistance in both time and space (Baym

et al. (2016)). In this set-up, bacteria are cultured on a large agar plate with different regimes of increasing antibiotic concentration. Distinct bacterial lineages can then migrate and evolve across the plate. Individual isolates can be sequenced to identify distinct antibiotic-resistance conferring mutations, and their phenotypes of antibiotic susceptibility can be measured. This approach offers a unique method to visually observe the evolution of antibiotic resistance.

Continuous culture systems offer more fine-tine control over the conditions in a long-term evolution experiment. These systems can be implemented to maintain predefined constant growth rates (chemostat), or use feedback control algorithms to maintain population features such as population density (turbidostat) or antibiotic resistance (morbidostat) (Koga & Humphrey (1967), Sorgeloos et al. (1976), Toprak et al. (2013)). Of relevance for this work, the morbidostat is a continuous culture device that can be used to study the evolution of antibiotic resistance through dynamic control of antibiotic concentration in a fixed volume with fixed intervals of refreshed media containing antibiotics (Figure 4.1).

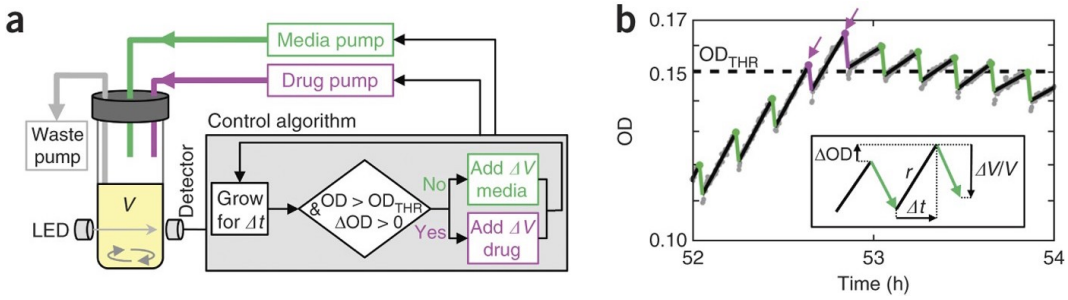



**Figure 4.1:** Morbidostat overview figure adapted from (Toprak et al. (2013)).

(A) Morbidostat set-up and control algorithm. (B) Representative bacterial growth curve (black line). Green lines denote culture dilution with media. Purple lines denote culture dilution with antibiotics.

One challenge with many of these methods is that they are cumbersome and difficult to use to study complex environments or dynamics. To address this challenge, in work led by Brandon Wong, PhD and Chris Mancuso, PhD, our lab developed the eVOLVER, a novel platform for scalable programmable continuous culture (Wong et al. (2018)). Additionally, recent work in our lab done by Marco Galardini, PhD built out a morbidostat control algorithm for the eVOLVER (unpublished).

### 4.1.3 A Proof-of-Concept Framework for Time Resolved Transcriptomics in Lab Evolution Experiments

As described in Chapter 1, and leveraged in Chapters 2 and 3, antibiotic susceptible bacteria exhibit stereotypical transcriptional signatures. However, these have largely been studied as end-point measurements: measuring gene expression from bacteria that are already susceptible or resistant, but not as they adapt from one phenotype to the other (as seen in Figure 4.2). In this Chapter, I probe gene expression in response to antibiotic exposure not just at these end-points, but also over time as bacteria evolve antibiotic resistance. This framework can be used to (1) identify drug-bug specific signatures of antibiotic susceptibility and resistance, and (2) identify transcriptional markers of the emergence of antibiotic resistance. Together, this can provide time-resolved gene expression measurements as bacteria traverse the evolutionary landscape in response to antibiotic selective pressures.

By conducting long-term evolution experiments and probing their transcriptome over time, we can observe the transcriptional dynamics in concert with mutational analysis (via genome sequencing), antibiotic resistance (via growth condi-

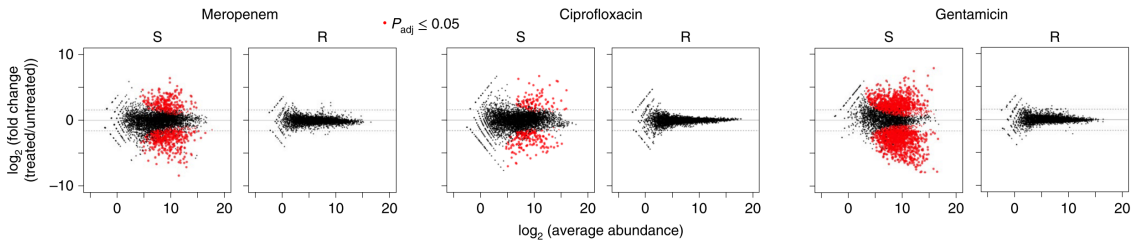



**Figure 4.2:** Gene expression of susceptible and resistant isolates of *K. pneumoniae*

Figure adapted from Bhattacharyya et al. (2019). RNA-Sequencing data from two susceptible (S) and two resistant (R) *K. pneumoniae* strains briefly exposed to the corresponding antibiotics.

tions), and other phenotypes. We can begin to ask questions such as: do susceptibility gene signatures progressively dampen as cells become more resistant (continuous expression), or are they discrete; what is the effect of specific resistance-conferring mutations on gene expression? Together, this approach can provide an expanded view of the evolution of antibiotic resistance.

In an experiment led by Marco Galardini, PhD, a panel of *E. coli* strains with known genetic backgrounds and known susceptibility to tetracycline were continuously cultured in the eVOLVER platform under a morbidostat feedback control algorithm. This algorithm increased tetracycline concentrations in the media while maintaining growth rate of the population, thus evolving tetracycline resistance. At regular intervals throughout the course of the experiment, cultures were sampled and frozen as glycerol stocks. This resulted in an experimental data set of 338 samples across 16 ancestral strains, 37 total time-points, and vial replicates. This data set provides the framework for measuring transcriptional phenotypes of antibiotic resistance across time and genetic background.

## 4.2 MATERIALS & METHODS

### 4.2.1 RNA-Sequencing of Lab Evolved *E. coli* with Time-Resolved Tetracycline Resistance

#### *4.2.1.1 Strain Growth & Brief Antibiotic Exposure*

From the set of *E. coli* strains that were evolved in the framework described in 4.1.3, three strains each from three time-points (9 samples), and the wild-type lab strain MG1655 in replicate (11 samples total) were cultured overnight in a shaking incubator at 37°C (Table 4.1). Each sample was diluted 1:000 and grown to mid-log phase. Each sample was split into two sub-samples and treated with either tetracycline (final concentration: 8 $\mu$g/mL) or water for 30 minutes in a shaking incubator at 37°C, to induce a transcriptioanl response. The resulting 22 samples were immediately processed for RNA isolation.

#### *4.2.1.2 RNA Isolation & Library Preparation*

RNA isolation was performed using the RNeasy Mini Kit (Qiagen: 74104) with the Qiagen RNAprotect Bacteria Reagent (Qiagen: 76506). Briefly, 250$\mu$L of each sample culture was added to a tube containing 500$\mu$L RNAprotect Bacteria Reagent, and then processed following the RNAprotect Bacteria Reagent steps. Each pellet was resuspended in Proteinase K (20mg/mL; Qiagen: 19131) in TE containing lysozyme (15mg/mL; Sigma Aldrich: L6876) and incubated at room temperature for ten minutes, and vortexed every two minutes. The samples were then processed following Protocol #4 of the RNeasy mini kit. The isolated RNA was then transferred to the DAMP Lab (https://www.damplab.org/) for library preparation. RNA quality was assessed using an Agilent Bioanalyzer with the RNA 6000

Pico Kit (Agilent: 5067-1513). Enrichment of mRNA was accomplished by using the MICROBExpress Kit (ThermoFisher: AM1905). cDNA libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina (NEB: E7775).

#### *4.2.1.3 RNA-Sequencing*

The cDNA libraries were sequenced by the Boston University Medical Campus Microarray and Sequencing Core. Library quality was first assessed using an Agilent Bioanalyzer, and all samples were satisfactory for sequencing. All libraries were sequenced on an Illumina NextSeq 500 across four lanes, generating single-end, un-stranded, 150bp reads. FASTQ generation was performed in Illumina BaseSpace.

### 4.2.2 Analysis Pipeline for Pre-Processing RNA-Sequencing Data

The following analysis steps were built into a pipeline using the snakemake workflow management system (available at: https://github.com/ebriars/TREE-Analysis). Each sample was coded following the scheme in Table 4.1. The complete list of software packages and versions can be found in Table 4.2. All of the software requirements can be reproduced using the provided conda environment in the repository. An overview of the workflow is visualized in Figure 4.3. By building this workflow using the snakemake workflow management system (Mölder et al. (2021)). The workflow is written in a human-readable, python-based language, and allows for scalable and reproducible analysis. The workflow is broken down into a set of "rules", each which specify the target inputs and outputs. The operations of each rule are carried out in custom python and bash scripts that interface with the snakemake infrastructure. Moreover, snakemake can parallelize the steps, and the

entire design workflow can be run locally or on a shared computing cluster.

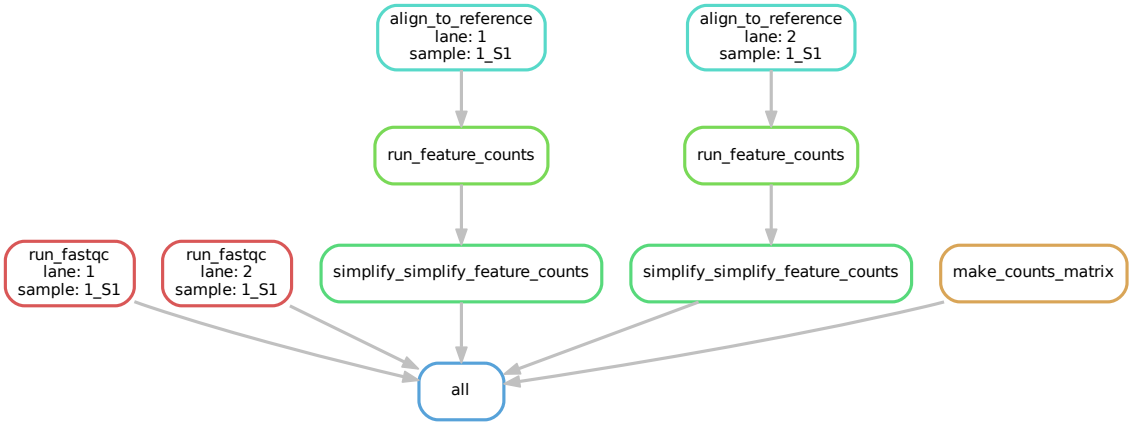



**Figure 4.3:** Snakemake workflow for RNA-Seq analysis of evolved *E. coli* strains

Example workflow for two FASTQ files corresponding to two lanes of the same sample. Final output is a feature counts matrix of per sample gene counts that can be used for downstream analysis.

#### *4.2.2.1 FASTQ Quality Control*

Each raw FASTQ file was analyzed using FASTQC (Andrews (2010)), and an aggregate report was generated using multiQC (Ewels et al. (2016)) (Figure 4.1). Reports were manually screened for quality assessment.

#### *4.2.2.2 Read Alignment & Calculating Transcript Counts*

Sequencing reads were aligned to the *E. coli* str. K-12 substr. MG1655 reference genome (GenBank Assembly Accession: GCA_000005845.2) using the Burrows-Wheeler Aligner (`bwa mem`) (Li & Durbin (2009)). Alignment metrics were analyzed using `samtools flagstat`. The `featureCounts` program from Subread was used to calculate gene counts for each sample (Liao et al. (2014)). Alignment coordinates were mapped to gene names using the *E. coli* str. K-12 substr

MG1655 GTF file (GCF_000005845.2_ASM584v2.gtf). Feature counts were computed for genes using the `-t gene` parameter. The resulting per sample feature counts were reduced to a single column of data for each sample, where each row is a gene and the value is the count. A feature counts matrix was created by combining each column of per sample data into a matrix with 4419 rows (genes) and 88 columns (samples). A reduced feature counts matrix was created by collapsing each sample's four columns (from four sequencing lanes) into one column. This reduced feature counts matrix is the final output of this Snakemake workflow and was then used for downstream analysis.

### 4.2.3 Custom RNA-Seq Analysis of Evolved *E. coli* strains

The following analyses were completed using custom `R` scripts which are also available in the github repository.

#### *4.2.3.1 Required Data & Metadata for Analysis*

The experiment's feature counts matrix was prepared by the snakemake workflow, resulting in a genes by samples matrix with 4419 rows and 22 columns. A metadata matrix was manually prepared, resulting in a samples by features matrix with 22 rows and 4 columns. The features included were: Sample Name, Strain, Treatment, and Time-Point. Additionally, a file containing the gene IDs for rRNA and tRNA genes was prepared from the *E. coli* str. K-12 substr MG1655 GTF file. Finally, a file containing the gene lengths for each gene in the *E. coli* str. K-12 substr MG1655 GTF file was prepared.

#### 4.2.3.2 *Transcript Abundance and Principal Component Analysis (PCA)*

Rows in the feature counts matrix corresponding to rRNA or tRNA genes were removed. Rows where all samples had a count of zero were removed. Transcripts per million (TPM) were calculated as reads per kilobase per million. TPM values were log-transformed and centered. Log-transformed TPM values were used for principal component analysis (PCA)

#### 4.2.3.3 *Differential Expression Analysis*

Differential expression analysis was performed using DESeq2 on raw transcript counts (Love et al. (2014)). DESeq2 was selected since this experiment has a small number of biological replicates (Schurch et al. (2016)). Additionally, since this is a small data set with a small number of biological replicates, statistical significance is not always achieved with the standard significance thresholds of $|log_2(\text{Fold Change})| > 2$ and adjusted p-value $< 0.05$. As such, when it is relevant, data is shown for a range of significance thresholds to provide an illustration of the trends in the data, without purporting statistical significance.

#### 4.2.3.4 *Gene ontology analysis and visualization*

Gene IDs from gene sets produced by differential expression analysis were inputs for gene ontology enrichment analysis (Ashburner et al. (2000), gen (2021)). GO terms (biological processes) that were significantly over-represented compared to a reference list of all *E. coli* genes were identified using PANTHER with a false discovery rate p-value $< 0.05$ (Mi et al. (2021)). To visualize the resulting GO terms, their semantic similarity scores (RSS) were computed and represented in a two-dimensional similarity space using NaviGO (Wei et al. (2017)).

| Code | Strain | Time-Point | Treatment | MIC ($\mu$g/mL) |
|---|---|---|---|---|
| 7 | ECOR-09 | 0 hours | No Treatment | 3 |
| 17 | IAI14 | 0 hours | No Treatment | 1.5 |
| 1 | IAI40 | 0 hours | No Treatment | 3 |
| 5 | ECOR-09 | 0 hours | Tetracycline (8 $\mu$g/mL) | 3 |
| 21 | IAI14 | 0 hours | Tetracycline (8 $\mu$g/mL) | 1.5 |
| 19 | IAI40 | 0 hours | Tetracycline (8 $\mu$g/mL) | 3 |
| 12 | ECOR-09 | 43.02 hours | No Treatment | 3 < MIC < 24 |
| 22 | IAI14 | 43.02 hours | No Treatment | 3 < MIC < 24 |
| 4 | IAI40 | 43.02 hours | No Treatment | 3 < MIC < 24 |
| 16 | ECOR-09 | 43.02 hours | Tetracycline (8 $\mu$g/mL) | 3 < MIC < 24 |
| 18 | IAI14 | 43.02 hours | Tetracycline (8 $\mu$g/mL) | 3 < MIC < 24 |
| 3 | IAI40 | 43.02 hours | Tetracycline (8 $\mu$g/mL) | 3 < MIC < 24 |
| 14 | ECOR-09 | 762.54 hours | No Treatment | 24 |
| 2 | IAI14 | 762.54 hours | No Treatment | 24 |
| 8 | IAI40 | 236.54 hours | No Treatment | 24-32 |
| 15 | ECOR-09 | 762.54 hours | Tetracycline (8 $\mu$g/mL) | 24 |
| 13 | IAI14 | 762.54 hours | Tetracycline (8 $\mu$g/mL) | 24 |
| 20 | IAI40 | 236.54 hours | Tetracycline (8 $\mu$g/mL) | 24-32 |
| 11 | MG1655 | N/A | No Treatment | <1 |
| 10 | MG1655 | N/A | No Treatment | <1 |
| 6 | MG1655 | N/A | Tetracycline (8 $\mu$g/mL) | <1 |
| 9 | MG1655 | N/A | Tetracycline (8 $\mu$g/mL) | <1 |

**Table 4.1:** *E. coli* strains, time-points, and tetracycline susceptibilities of the evolved samples processed for RNA-Seq

## 4.3 RESULTS

### 4.3.1 Tetracycline Susceptibility Over Time of the Evolved Strains in the RNA-Seq Data Set

For each sample, their tetracycline susceptibility (MIC) was estimated by the eVOLVER using the amount of tetracycline in the media (Figure 4.4). Their initial time-point and final time-point MICs were also measured using tetracycline MIC strips (Table 4.1). The clinical breakpoints for tetracycline (4 $\mu$g/mL,

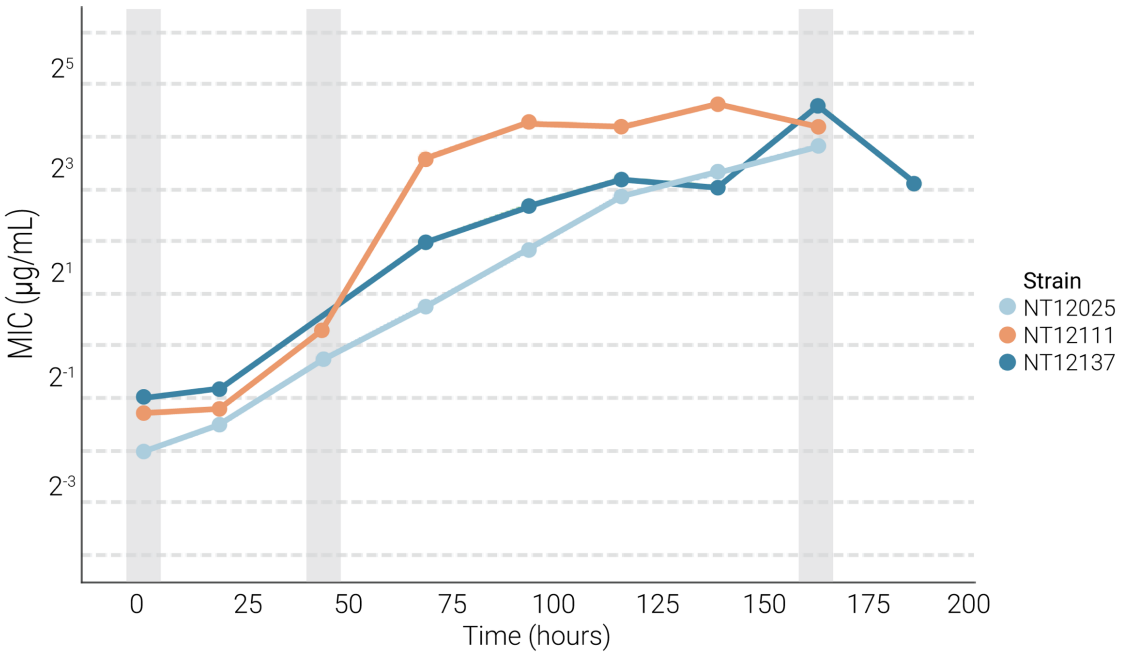



**Figure 4.4:** Estimated tetracycline MIC over time for three evolved strains evolved processed for RNA-Seq

### 4.3.2 Characterization of the RNA-Seq Data Set

Overall, this experiment produced a high quality initial data set. There were no observed lane effects, and each sample had between 1.1-2.4 million reads (with

the exception of Sample 2). All samples had a high percentage of duplicate reads, likely due to the high number of rRNA and tRNA remaining in the sample (Figure 4.5, Figure 4.6).

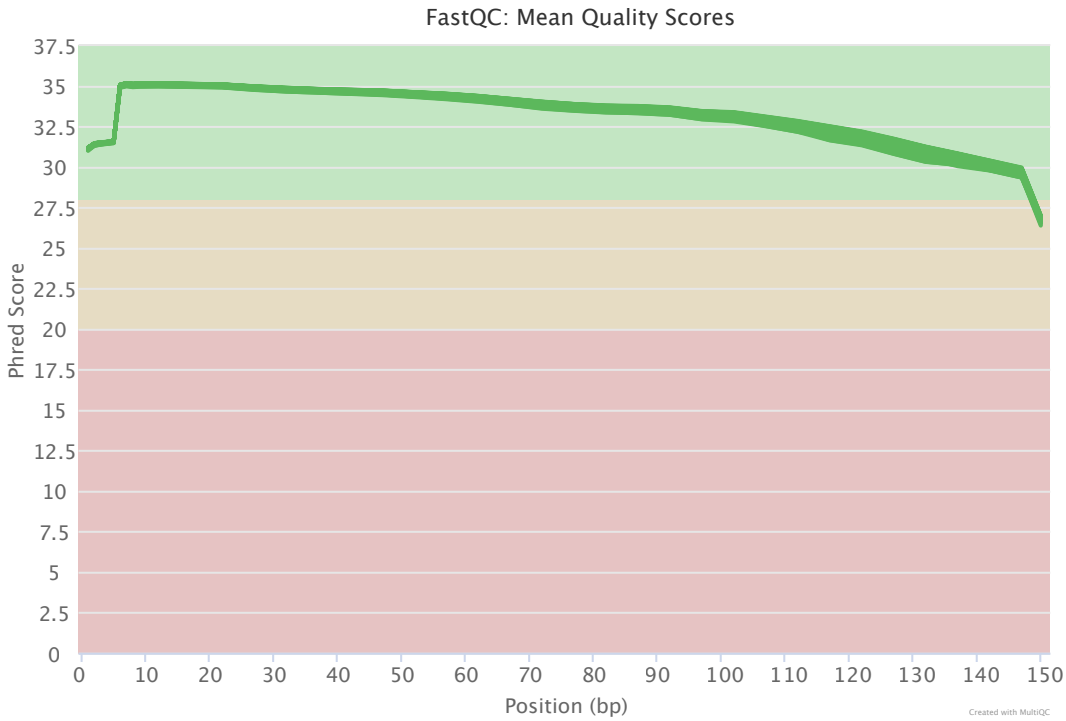



**Figure 4.5:** Summary of FASTQC per base quality results for RNA-Seq of evolved *E. coli* strains

Figures generated by multiqc. Mean quality (Phred) score for all of the reads each sample across each lane at each position.

PCA of all samples suggests that, as expected, the majority of the variance in gene expression is due to the strain and tetracycline exposure as seen in Figure 4.7. Additionally, the MG1655 biological replicates (the only biological replicates in this study) cluster close to each other, suggesting that the gene expression is consistent across different individual samples.

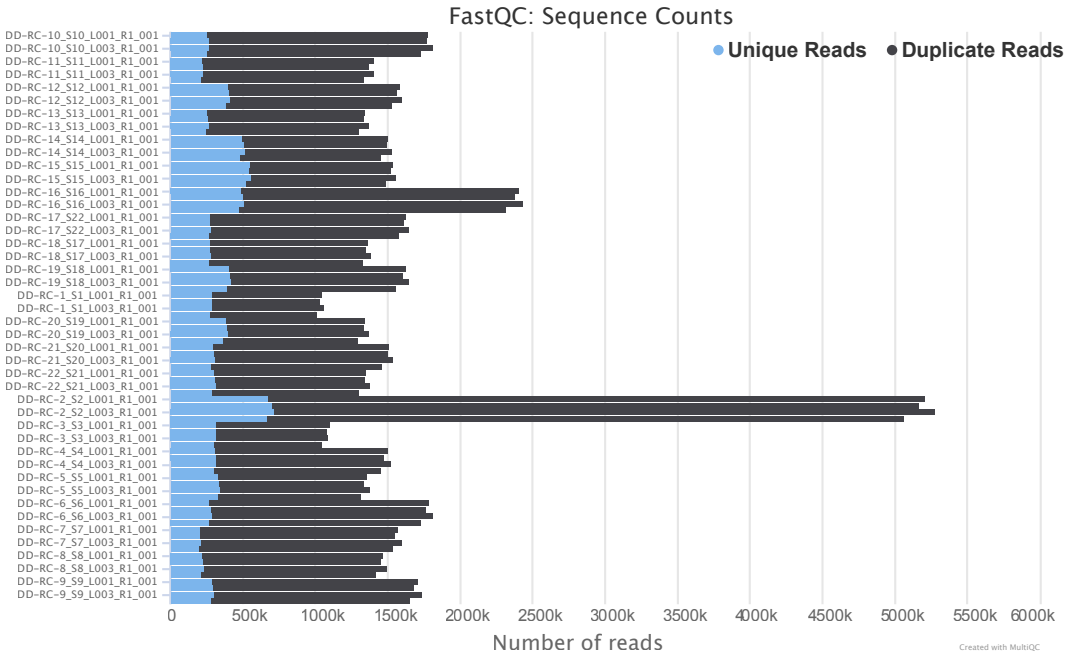



**Figure 4.6:** Summary of number of reads per sample from RNA-Seq of evolved *E. coli* strains

Figures generated by multiqc. Total number of reads for each sample across each lane.

### 4.3.3 Total Number of Differentially Expressed Genes Decreases as *E. coli* Become Resistant to Tetracycline

The overall differential gene expression was compared between the susceptible (time = 0 hours), intermediate (time = 43 hours), and resistant (time = 236–762 hours) samples in response to brief tetracycline exposure (Figure 4.8). As the strains are continuously cultured in the eVOLVER with constant growth rate and increasing tetracycline concentration as controlled by the morbidostat control algorithm, the amount of genes that are differentially expressed in response to brief tetracycline exposure dampens over time (Figure 4.9). Interestingly, the amount of differentially expressed genes between the initial time-point and after 43 hours is only slightly reduced. By sequencing additional samples in this data set from addi-

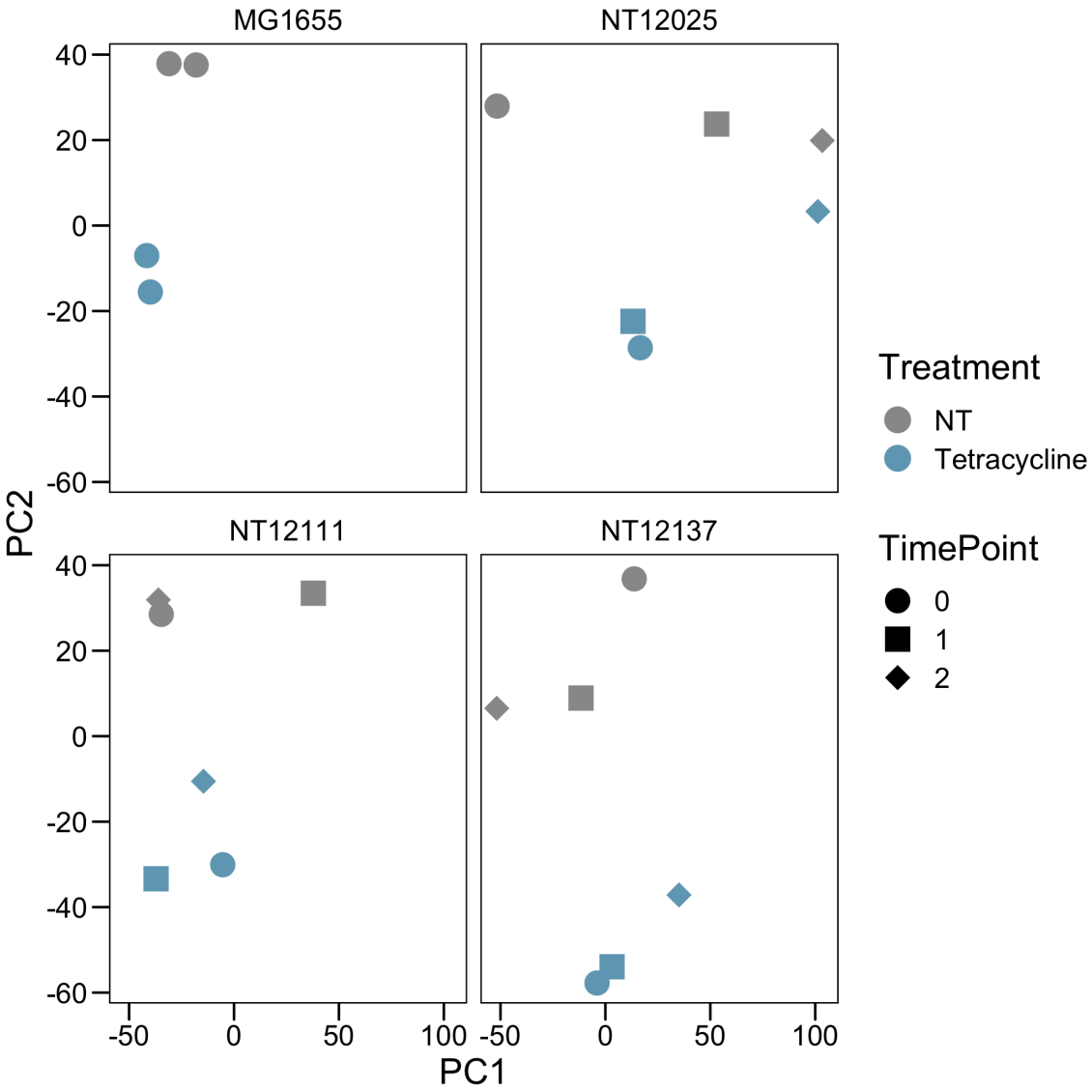



**Figure 4.7:** PCA of all samples from the dataset of RNA-Seq of evolved *E. coli* strains

Time-points correspond to the initial (0), intermediate (1), and final (2) time-points listed in Table 4.1

tional time-points corresponding to increasing levels of tetracycline resistance, this curve could be expanded to concretely identify if (a) there is an inflection point in resistance where differential gene expression shuts off, or (b) gene expression dampens over time.

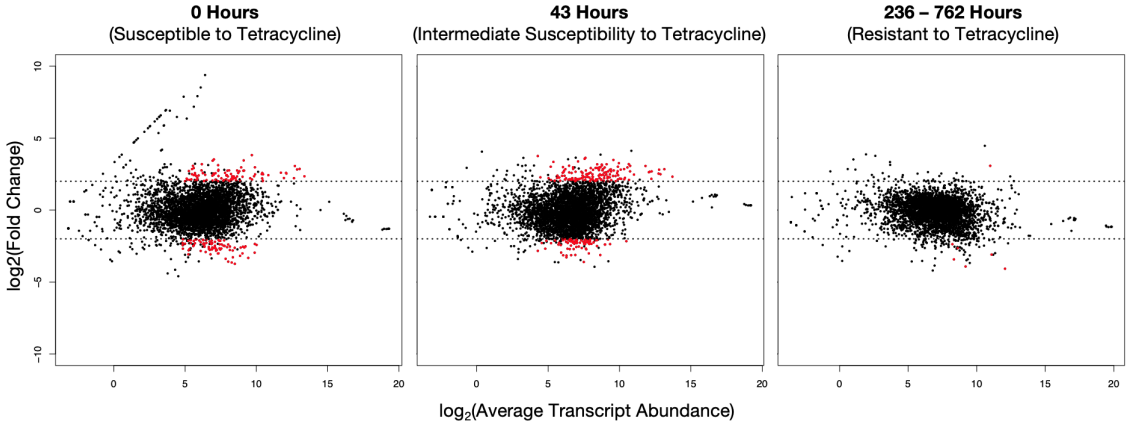



**Figure 4.8:** Differential expression versus transcript abundance across time in *E. coli* strains evolved towards tetracylcine resistance

Differential expression (brief tetracycline exposure v. NT) of all evolved strains at each time point. Each dot represents one gene, and are colored red when the differential expression is significant ($|log_2$(Fold Change)$| > 2$ and adjusted p-value $< 0.05$).

### 4.3.4 Identification of Differentially Expressed Genes Unique to Levels of Tetracycline Resistance

As expected, the susceptible samples exhibit a transcriptional response to the tetracycline exposure. The resistant strains, however, only differentially express a few genes in response to tetracycline exposure (Figure 4.10). The specific genes that are in each subset of genes that are differentially expressed in response to brief tetracycline exposure for each time-point were further analyzed. Interestingly, the genes that are differentially expressed at the initial time-point only marginally overlap with the genes that are differentially expressed at the intermediate time-point (Figure 4.11). Both the susceptible and intermediate samples have unique sets of genes that are differentially expressed ($|log_2$(Fold Change)$| > 2.5$ and adjusted p-value $< 0.05$). These two gene sets were probed using gene onotology analysis to identify over-represented biological features. The enriched GO terms

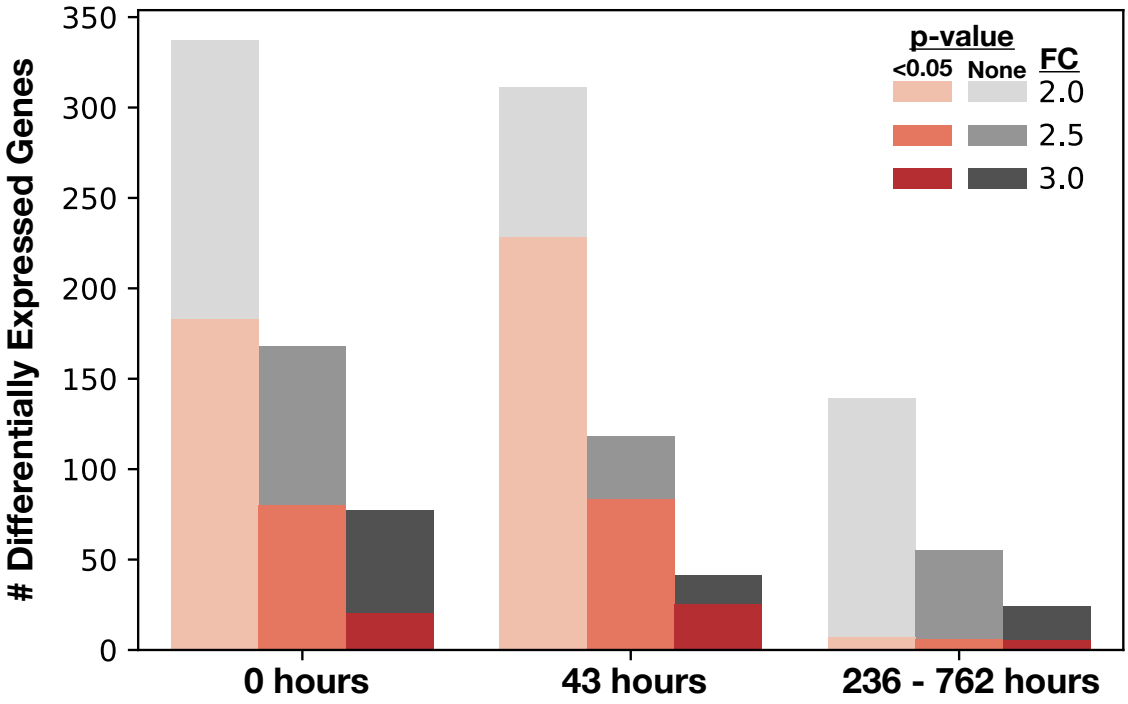



**Figure 4.9:** Number of differentially expressed genes across time in *E. coli* strains evolved towards tetracycline resistance

Comparison of number of differentially expressed (brief tetracycline exposure v. NT) genes with different log2 Fold Change (FC) and adjusted p-value cut-offs.

for the initial time-point and the intermediate time-point are visualized in Figure 4.12 and Figure 4.13, respectively. There were no enriched GO terms from the gene set from the final time-point. In the susceptible samples, there appear to be a few regimes of GO terms. The largest group of GO terms are associated with amino acid transport, and likely correspond to a tetracycline-specific response since tetracycline is a bacteriostatic drug that binds thee 30S ribosomal subunit and prevents incoming aminoacyl tRNAs from binding. The remaining GO terms are more as-

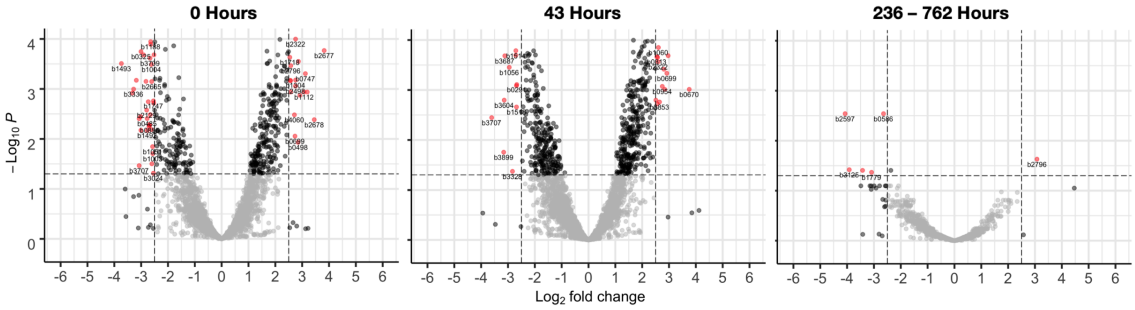



**Figure 4.10:** Volcano plots of differential expression in *E. coli* strains evolved towards tetracylcine resistance

Differential expression (brief tetracycline exposure v. NT) of all evolved strains at each time point. Each dot represents one gene, and are colored red when the differential expression is significant ($|log_2$(Fold Change)$| > 2.5$ and adjusted p-value $< 0.05$).

sociated with general stress responses. In the intermediate samples, the majority of GO terms cluster in one group associated with protein localization.

Interestingly, there are only four genes that are uniquely differentially expressed in the resistant strains (Table 4.2). One of these genes, gapA, has been shown to be affected by evolution under conditions of chronic heat stress (Riehle et al. (2003)), suggesting that this is due to the experimental set-up. However, upon examining the other genes in this set, we see potential associations with mechanisms of tetracycline resistant. For example, the over expression of raiA in stress conditions has been associated with faster growth restart and an increase in 70S ribosome content, allowing for protection of the ribosome (Lang et al. (2021)). Additionally, the protein encoded by entF is involved with enterobactin synthsesis. Enterobactin is synthesized in response to iron deprivation (Reichert et al. (1992)). Interestingly, it has been shown that the accumulation and activity of tetracycline can be inhibited in *E. coli* in the presence of iron (Avery et al. (2004)). One hypothesis is that these *E. coli* strains evolved to tetracycline resistance are up-regulating entF to encour-

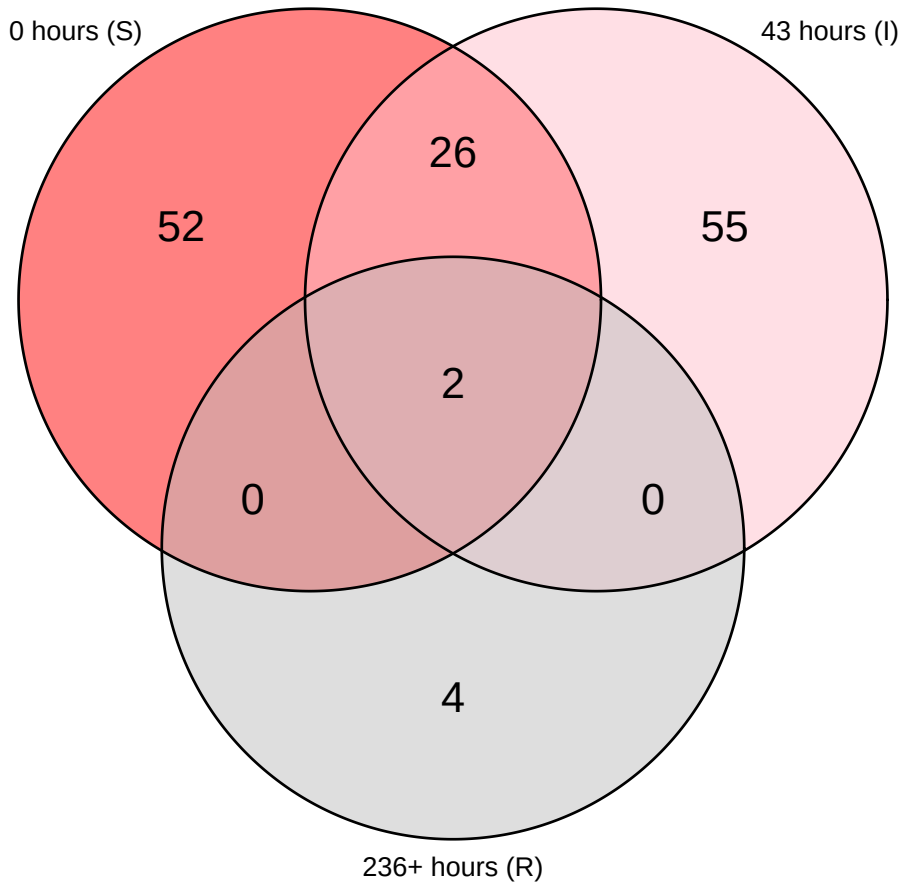



**Figure 4.11:** Intersection of differentially expressed genes between each time-point

age enterobactin production in order to increase iron uptake, and therefore inhibit tetracycline accumulation. Follow-up studies could be used to explore these hypotheses generated by this work.

## 4.4 DISCUSSION

In this Chapter, I have introduced a framework for reproducible RNA-seq analysis of bacterial strains evolved to antibiotic resistance over time. I have used this

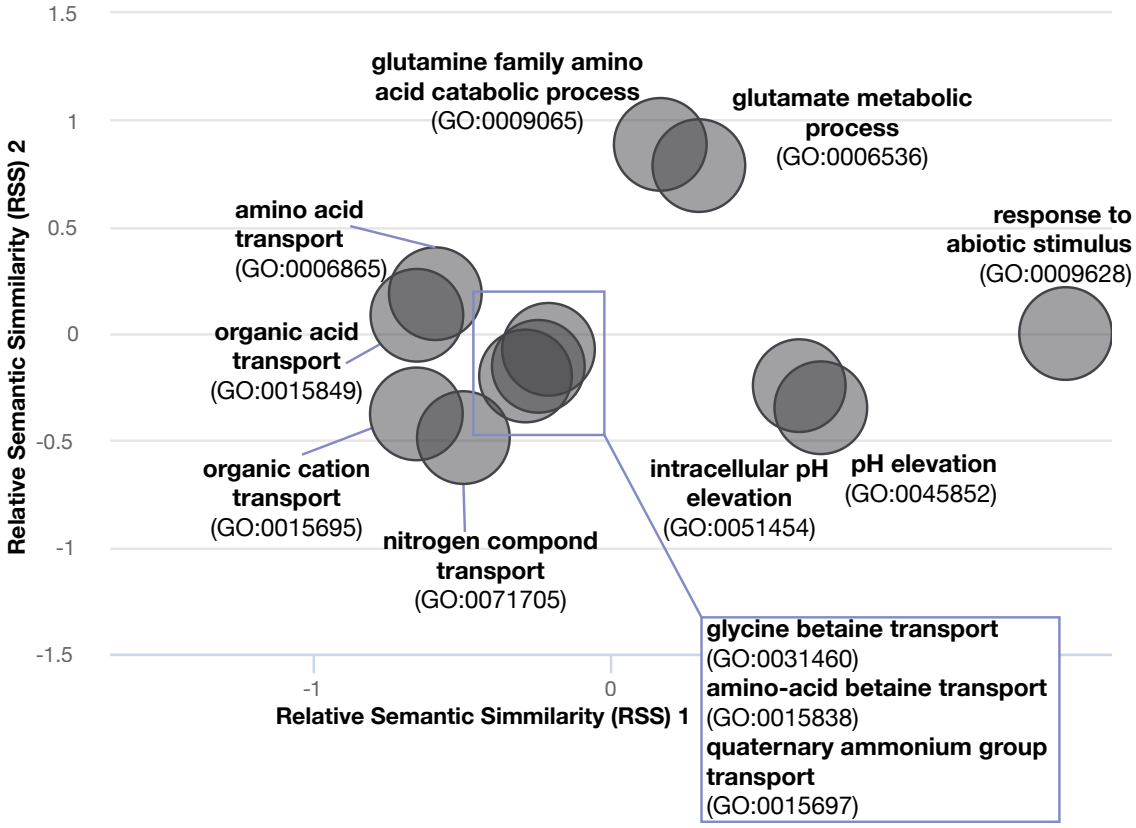



**Figure 4.12:** Visualization of enriched GO terms in gene set of uniquely differentially expressed genes in the samples from the initial time-point

framework to explore the distinct transcriptional phenotypes of *E. coli* strains with different susceptibilities to tetracycline throughout their evolution towards tetracycline resistance. In this data set, I identified distinct gene sets in the susceptible and intermediate strains that are differentially expressed in response to brief tetracycline exposure. I also identified sets of genes whose expression continuously increases or decreases as the strain becomes more resistant to tetracycline. Together, this work provides a framework that could be replicated with additional samples from the tetracycline evolution experiment introduced here, or for the design and

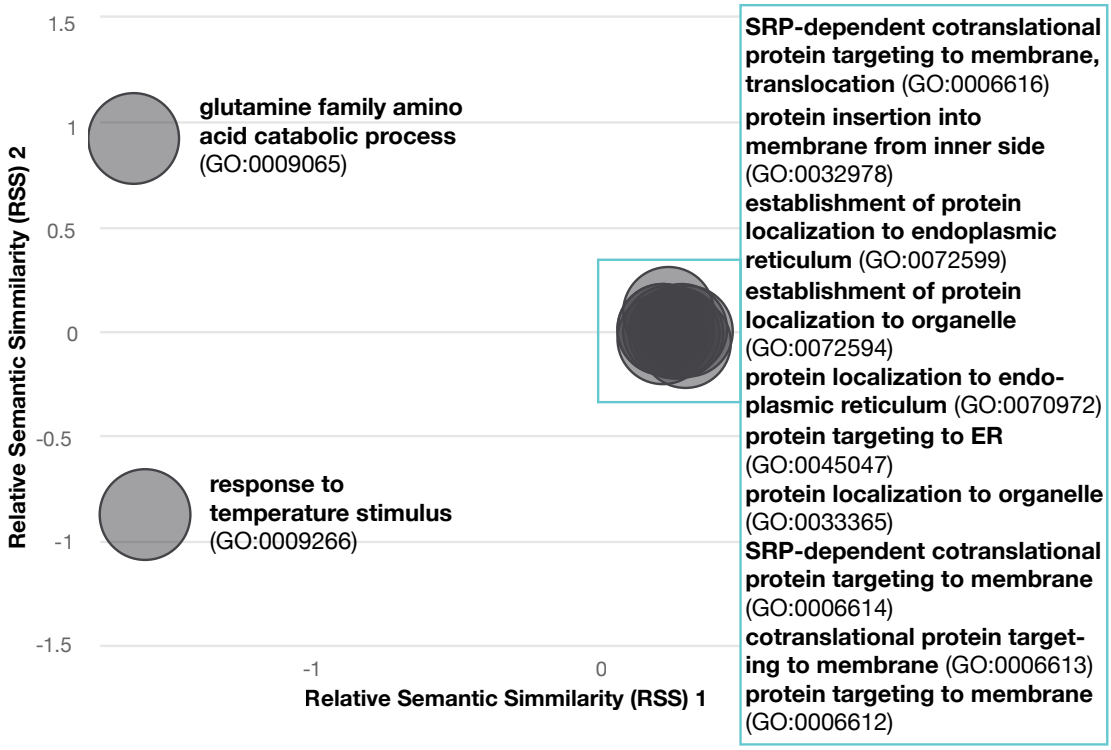



**Figure 4.13:** Visualization of enriched GO terms in gene set of uniquely differentially expressed genes in the samples from the intermediate time-point

analysis of future evolution experiments. The inclusion of additional strains could allow for the discovery of a signature of tetracycline susceptibility for *E. coli*.

Additionally, this framework can be used for hypothesis generation and testing for mechanisms of susceptibility and resistance. In this Chapter, I highlighted the genes that are uniquely differentially expressed in the resistant strains. I performed a literature search to generate hypotheses for why these genes are present in this gene set, but follow-up mechanistic studies could be performed to test these hypotheses. This would be useful for identifying the functional features of susceptibility and resistance.

| GeneID | Name | Encodes | Note |
|---|---|---|---|
| b0586 | entF | apo-serine activating enzyme | involved with enterobactin synthesis |
| b1779 | gapA | glyceraldehyde-3-phosphate dehydrogenase | Expression affected by evolution and oxygen limitation |
| b2597 | raiA | ribosome-associated inhibitor A | ribosome protection in stress conditions |
| b3125 | garR | tartronate semialdehyde reductase | |

**Table 4.2:** Genes that are differentially expressed only in *E. coli* strains with lab-evolved resistance to tetracycline

A deliberate design choice for this framework was to study the evolution of antibiotic resistance in these different *E. coli* strains with the lab strain *E. coli* MG1655 as the reference. This allows for any differentially expressed signature genes identified to be generalizable across all strains, and not just features specific to each strain (*i.e.*, if each strain was analyzed individually in comparison to its reference). However, if one wanted to couple this transcriptional analysis with mutational analysis, for example, it would make sense to modify this workflow such that each species' reference genome is used.

## 4.5 SUPPLEMENTARY TABLES

| Package | Version |
|---|---|
| bwa mem | |
| fastqc | 0.11.7 |
| multiqc | 1.10.1 |
| samtools | 1.10 |
| snakemake-minimal | 5.4.5 |
| subread | 1.6.2 |

**Table 4.3:** List of software packages and versions used by snakemake pipeline for RNA-Seq analysis of evolved *E. coli*

# CHAPTER 5

# Conclusions & Future Perspectives

## 5.1 SUMMARY OF WORK

This dissertation aims to improve antibiotic stewardship through the development of methods to diagnose antibiotic resistance in clinically-relevant infections, as well as through the development of methods to predict antibiotic resistance in experimental settings. In Chapter 2, I introduced a framework for facilitating both the *in silico* toehold switch sensor design, and the *in vitro* evaluation of toehold switch sensors. In Chapter 3, I introduced a context-optimized workflow for toehold sensor detection of NASBA amplified differentially expressed mRNA. I introduced a toehold switch sensor platform for detecting differential RNA signatures of antibiotic susceptibility, and deploy this assay on a set of clinically relevant *E. coli* strains. In Chapter 4, I introduced a framework for reproducible RNA-seq analysis of bacterial strains evolved to antibiotic resistance over time, and highlight some interesting transcriptional features observed in a set of *E. coli* strains evolved towards tetracycline resistance.

## 5.2 FUTURE PERSPECTIVES

### 5.2.1 Deploying the AST for *E. coli* Susceptibility to Fluoroquinolones on Additional Clinically Relevant Strains

The simplest future direction of this work is to deploy the platform developed in Chapter 3 on additional, clinically relevant strains of *E. coli*. This data set could then be used to build a set of rules for classifying a test sample as susceptible or resistant, based on the toehold sensor detection. Moreover, this data set could be

expanded to additional fluoroquinolone antibiotics.

### 5.2.2 Exploring Known Transcriptional Signatures of Antibiotic Susceptibility

For currently known signatures of antibiotic susceptibility (*e.g.* the *E. coli* fluoroquinolone susceptibility signature used in Chapter 3), the framework introduced in Chapter 4 could be used to follow the signature gene expression over time as strains evolve towards tetracycline resistance. By teasing apart where in the resistance evolutionary landscape each gene's expression is modulated, one could assign weights to each gene in the signature. For example, genes that quickly (in evolutionary time corresponding to increased resistance) would be weighted such that detected differential expression strongly correlates with susceptibility, and the lack of differential expression suggests emerging resistance.

### 5.2.3 Identifying New Signatures of Antibiotic Susceptibility

The framework introduced in Chapter 4 could be used to identify new transcriptional signatures of antibiotic susceptibility. Moreover, these time-resolved evolution experiments done in the eVOLVER can easily be paired with genotypic measurements and other phenotypic measurements. Not only could this be used to identify end-point signatures of antibiotic susceptibility, but this can also be used to identify early transcriptional markers (or other measurable phenotypes) of susceptibility. This could be extremely useful for emerging infections that are not well-characterized, or for bacteria for which early resistance detection has favorable clinical outcomes, *e.g.* sepsis (Rivers et al. (2001)).

### 5.2.4 Exploring Drug-Bug Pairs and Collateral Susceptibility

Time-resolved transcriptional measurements throughout the evolution of antibiotic resistance could be used to probe the currently mixed literature on collateral sensitivity. If a bacterial species is resistant to antibiotic X, what is the likelihood that it is resistant to antibiotic Y (concurrent resistance), and what is the likelihood that it is susceptible to antibiotic Z (collateral sensitivity). Identifying collateral sensitivity is of clinical relevance because it would allow for antibiotic cycling in treatment plans, thus significantly reducing the occurrence of antibiotic resistance due to single antibiotic use. Some *in vitro* evolution studies have demonstrated the occurrence of collateral sensitivity for *E. coli* (Pál et al. (2015), Apjok et al. (2019)). However, a recent retrospective analysis of almost 500,000 antimicrobial susceptibility tests acquired over four years from 23 hospitals in Pittsburgh, PA found that at the species level, only 0.7% of antibiotic pairs had collateral sensitivity. Instead, more than 53% of antibiotic pairs had concurrent resistance (Beckley & Wright (2021)). Additionally, a similar trend was seen in a smaller *in vivo* study (Cherny et al. (2021)). Although these *in vivo* data suggest that there are not generalizable rules for collateral sensitivity within a species, the high incidence of concurrent resistance does provide empirical evidence for choosing antibiotics that avoid cross-resistance. Additionally, these data could be informative both for the development of future evolution experiments, and for considerations in identifying new transcriptional signatures of antibiotic susceptibility. For example, signatures could be designed such that they cover the smallest set of antibiotics for which there is either direct or concurrent resistance information for.

## BIBLIOGRAPHY

(2021). The gene ontology resource: enriching a gold mine. *Nucleic Acids Research*, *49*(D1), D325–D334.

Abudayyeh, O. O., Gootenberg, J. S., Konermann, S., Joung, J., Slaymaker, I. M., Cox, D. B., Shmakov, S., Makarova, K. S., Semenova, E., Minakhin, L., et al. (2016). C2c2 is a single-component programmable rna-guided rna-targeting crispr effector. *Science*, *353*(6299).

Addis, T., Mekonnen, Y., Ayenew, Z., Fentaw, S., & Biazin, H. (2021). Bacterial uropathogens and burden of antimicrobial resistance pattern in urine specimens referred to ethiopian public health institute. *PloS one*, *16*(11), e0259602.

Ahokas, H., & Erkkilä, M. J. (1993). Interference of pcr amplification by the polyamines, spermine and spermidine. *Genome Research*, *3*(1), 65–68.

Alcock, B. P., Raphenya, A. R., Lau, T. T., Tsang, K. K., Bouchard, M., Edalatmand, A., Huynh, W., Nguyen, A.-L. V., Cheng, A. A., Liu, S., et al. (2020). Card 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research*, *48*(D1), D517–D525.

Andrews, S. (2010). Fastqc: A quality control tool for high throughput sequence data. Http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Angenent-Mari, N. M., Garruss, A. S., Soenksen, L. R., Church, G., & Collins, J. J. (2020). A deep learning approach to programmable rna switches. *Nature communications*, *11*(1), 1–12.

Apjok, G., Boross, G., Nyerges, Á., Fekete, G., Lázár, V., Papp, B., Pál, C., & Csörgő, B. (2019). Limited evolutionary conservation of the phenotypic effects of antibiotic resistance mutations. *Molecular biology and evolution*, *36*(8), 1601–1611.

Arena, F., Viaggi, B., Galli, L., & Rossolini, G. M. (2015). Antibiotic susceptibility testing: present and future. *The Pediatric infectious disease journal*, *34*(10), 1128–1130.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., et al. (2000). Gene ontology: tool for the unification of biology. *Nature genetics*, *25*(1), 25–29.

Asundi, A., Billings, N., Rogers, Z. H., Roche, S., Bombard, J., Kumcu, M. E., Cunden, L. S., Nanayakkara, I. A., Tsui, C., Knysh, P., et al. (2020). 153. pilot study

of a novel whole-genome sequencing based rapid bacterial identification assay in patients with bacteremia. In *Open Forum Infectious Diseases*, vol. 7, (pp. S206–S206). Oxford University Press US.

Avery, A. M., Goddard, H. J., Sumner, E. R., & Avery, S. V. (2004). Iron blocks the accumulation and activity of tetracyclines in bacteria. *Antimicrobial agents and chemotherapy*, *48*(5), 1892–1894.

Baharoglu, Z., & Mazel, D. (2011). Vibrio cholerae triggers sos and mutagenesis in response to a wide range of antibiotics: a route towards multiresistance. *Antimicrobial agents and chemotherapy*, *55*(5), 2438–2441.

Baquero, F., & Blázquez, J. (1997). Evolution of antibiotic resistance. *Trends in ecology & evolution*, *12*(12), 482–487.

Barczak, A. K., Gomez, J. E., Kaufmann, B. B., Hinson, E. R., Cosimi, L., Borowsky, M. L., Onderdonk, A. B., Stanley, S. A., Kaur, D., Bryant, K. F., et al. (2012). Rna signatures allow rapid identification of pathogens and antibiotic susceptibilities. *Proceedings of the national academy of sciences*, *109*(16), 6217–6222.

Bard, J. D., & Lee, F. (2018). Why can't we just use pcr? the role of genotypic versus phenotypic testing for antimicrobial resistance testing. *Clinical microbiology newsletter*, *40*(11), 87–95.

Basha, I. H. K., Ho, E. T. W., Yousuff, C. M., & Hamid, N. H. B. (2017). Towards multiplex molecular diagnosis—a review of microfluidic genomics technologies. *Micromachines*, *8*(9), 266.

Baur, D., Gladstone, B. P., Burkert, F., Carrara, E., Foschi, F., Döbele, S., & Tacconelli, E. (2017). Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and clostridium difficile infection: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, *17*(9), 990–1001.

Baym, M., Lieberman, T. D., Kelsic, E. D., Chait, R., Gross, R., Yelin, I., & Kishony, R. (2016). Spatiotemporal microbial evolution on antibiotic landscapes. *Science*, *353*(6304), 1147–1151.

Beckley, A. M., & Wright, E. S. (2021). Identification of antibiotic pairs that evade concurrent resistance via a retrospective analysis of antimicrobial susceptibility test results. *The Lancet Microbe*.

Bhattacharyya, R. P., Bandyopadhyay, N., Ma, P., Son, S. S., Liu, J., He, L. L., Wu, L., Khafizov, R., Boykin, R., Cerqueira, G. C., et al. (2019). Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. *Nature medicine*, *25*(12), 1858–1864.

Brazas, M. D., & Hancock, R. E. (2005). Using microarray gene signatures to elucidate mechanisms of antibiotic action and resistance. *Drug discovery today*, *10*(18), 1245–1252.

Bronzwaer, S. L., Cars, O., Buchholz, U., Mölstad, S., Goettsch, W., Veldhuijzen, I. K., Kool, J. L., Sprenger, M. J., Degener, J. E., et al. (2002). The relationship between antimicrobial use and antimicrobial resistance in europe. *Emerging infectious diseases*, *8*(3), 278.

Card, K. J., LaBar, T., Gomez, J. B., & Lenski, R. E. (2019). Historical contingency in the evolution of antibiotic resistance after decades of relaxed selection. *PLoS biology*, *17*(10), e3000397.

Card, K. J., Thomas, M. D., Graves, J. L., Barrick, J. E., & Lenski, R. E. (2021). Genomic evolution of antibiotic resistance is contingent on genetic background following a long-term experiment with escherichia coli. *Proceedings of the National Academy of Sciences*, *118*(5).

Carter, J. J., Consortium, C., et al. (2021). Quantitative measurement of antibiotic resistance in mycobacterium tuberculosis reveals genetic determinants of resistance and susceptibility in a target gene approach. *bioRxiv*.

Chathuranga, G., Dissanayake, T., Fernando, N., & Wanigatunge, C. (2021). Appropriateness of the empirical antibiotics prescribed and their concordance with national guidelines for three selected infections among cancer patients in a tertiary care centre in sri lanka. *International Journal of Microbiology*, *2021*.

Cherny, S. S., Nevo, D., Baraz, A., Baruch, S., Lewin-Epstein, O., Stein, G. Y., & Obolski, U. (2021). Revealing antibiotic cross-resistance patterns in hospitalized patients through bayesian network modelling. *Journal of Antimicrobial Chemotherapy*, *76*(1), 239–248.

Connelly, J. T., Rolland, J. P., & Whitesides, G. M. (2015). "paper machine" for molecular diagnostics. *Analytical chemistry*, *87*(15), 7595–7601.

Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). Multiqc: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.

Fowler, P. W., Wright, C., Spiers-Bowers, H., Zhu, T., Baeten, E. M., Hoosdally, S. W., Cruz, A. L. G., Roohi, A., Kouchaki, S., Walker, T. M., et al. (2021). Bashthebug: a crowd of volunteers reproducibly and accurately measure the minimum inhibitory concentrations of 13 antitubercular drugs from photographs of 96-well broth microdilution plates. *bioRxiv*.

Foxman, B. (2010). The epidemiology of urinary tract infection. *Nature Reviews Urology*, *7*(12), 653–660.

Furukawa, D., Kim, B., & Jeng, A. (2021). Real-life utilization of biofire® filmarray® pneumonia panel as an antibiotic stewardship tool. *Infectious Diseases*, *53*(4), 308–313.

Galardini, M., Koumoutsi, A., Herrera-Dominguez, L., Varela, J. A. C., Telzerow, A., Wagih, O., Wartel, M., Clermont, O., Denamur, E., Typas, A., et al. (2017). Phenotype inference in an escherichia coli strain panel. *Elife*, *6*, e31035.

Genther, C., Schoeny, R., Loper, J., & Smith, C. (1977). Mutagenic studies of folic acid antagonists. *Antimicrobial agents and chemotherapy*, *12*(1), 84–92.

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M., Group, E. P., et al. (2005). Outpatient antibiotic use in europe and association with resistance: a cross-national database study. *The Lancet*, *365*(9459), 579–587.

Gootenberg, J. S., Abudayyeh, O. O., Lee, J. W., Essletzbichler, P., Dy, A. J., Joung, J., Verdine, V., Donghia, N., Daringer, N. M., Freije, C. A., et al. (2017). Nucleic acid detection with crispr-cas13a/c2c2. *Science*, *356*(6336), 438–442.

Grant, N. A., Abdel Magid, A., Franklin, J., Dufour, Y., & Lenski, R. E. (2020). Changes in cell size and shape during 50,000 generations of experimental evolution with escherichia coli. *Journal of Bacteriology*, *203*(10), e00469–20.

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: de-novo-designed regulators of gene expression. *Cell*, *159*(4), 925–939.

Grüneberg, R. (1994). Changes in urinary pathogens and their antibiotic sensitivities, 1971–1992. *Journal of Antimicrobial Chemotherapy*, *33*(suppl_A), 1–8.

Heid, C. A., Stevens, J., Livak, K. J., & Williams, P. M. (1996). Real time quantitative pcr. *Genome research*, *6*(10), 986–994.

Hodgman, C. E., & Jewett, M. C. (2012). Cell-free synthetic biology: thinking outside the cell. *Metabolic engineering*, *14*(3), 261–269.

Hombach, M., Jetter, M., Keller, P. M., Blöchliger, N., Kolesnik-Goldmann, N., & Böttger, E. C. (2017). Rapid detection of esbl, carbapenemases, mrsa and other important resistance phenotypes within 6–8 h by automated disc diffusion antibiotic susceptibility testing. *Journal of Antimicrobial Chemotherapy*, *72*(11), 3063–3069.

Hong, F., Ma, D., Wu, K., Mina, L. A., Luiten, R. C., Liu, Y., Yan, H., & Green, A. A. (2020). Precise and programmable detection of mutations using ultraspecific riboregulators. *Cell*, *180*(5), 1018–1032.

Hooton, T. M., & Stamm, W. E. (1997). Diagnosis and treatment of uncomplicated urinary tract infection. *Infectious Disease Clinics*, *11*(3), 551–581.

Joung, J., Ladha, A., Saito, M., Segel, M., Bruneau, R., Huang, M.-l. W., Kim, N.-G., Yu, X., Li, J., Walker, B. D., et al. (2020). Point-of-care testing for covid-19 using sherlock diagnostics. *MedRxiv*.

Kalashnikov, M., Lee, J. C., Campbell, J., Sharon, A., & Sauer-Budge, A. F. (2012). A microfluidic platform for rapid, stress-induced antibiotic susceptibility testing of staphylococcus aureus. *Lab on a chip*, *12*(21), 4523–4532.

Karanika, S., Paudel, S., Grigoras, C., Kalbasi, A., & Mylonakis, E. (2016). Systematic review and meta-analysis of clinical and economic outcomes from the implementation of hospital-based antimicrobial stewardship programs. *Antimicrobial agents and chemotherapy*, *60*(8), 4840–4852.

Khazaei, T., Barlow, J. T., Schoepp, N. G., & Ismagilov, R. F. (2018). Rna markers enable phenotypic test of antibiotic susceptibility in neisseria gonorrhoeae after 10 minutes of ciprofloxacin exposure. *Scientific reports*, *8*(1), 1–10.

Koga, S., & Humphrey, A. (1967). Study of the dynamic behavior of the chemostat system. *Biotechnology and Bioengineering*, *9*(3), 375–386.

Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A., & Collins, J. J. (2007). A common mechanism of cellular death induced by bactericidal antibiotics. *Cell*, *130*(5), 797–810.

Kollef, M. H., Shorr, A. F., Bassetti, M., Timsit, J.-F., Micek, S. T., Michelson, A. P., & Garnacho-Montero, J. (2021). Timing of antibiotic therapy in the icu. *Critical Care*, *25*(1), 1–10.

Lakin, S. M., Dean, C., Noyes, N. R., Dettenwanger, A., Ross, A. S., Doster, E., Rovira, P., Abdo, Z., Jones, K. L., Ruiz, J., et al. (2017). Megares: an antimicrobial resistance database for high throughput sequencing. *Nucleic acids research*, *45*(D1), D574–D580.

Lambert, R., & Pearson, J. (2000). Susceptibility testing: accurate and reproducible minimum inhibitory concentration (mic) and non-inhibitory concentration (nic) values. *Journal of applied microbiology*, *88*(5), 784–790.

Lang, M., Krin, E., Korlowski, C., Sismeiro, O., Varet, H., Coppée, J.-Y., Mazel, D., & Baharoglu, Z. (2021). Sleeping ribosomes: Bacterial signaling triggers raia mediated persistence to aminoglycosides. *Iscience*, *24*(10), 103128.

Lee, W.-B., Fu, C.-Y., Chang, W.-H., You, H.-L., Wang, C.-H., Lee, M. S., & Lee, G.-B. (2017). A microfluidic device for antimicrobial susceptibility testing based on a broth dilution method. *Biosensors and Bioelectronics*, *87*, 669–678.

Lenski, R. E. (2004). Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium escherichia coli. *Plant breeding reviews*, *24*(2), 225–266.

Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-term experimental evolution in escherichia coli. i. adaptation and divergence during 2,000 generations. *The American Naturalist*, *138*(6), 1315–1341.

Levine, A. D., Meyer, M. T., & Kish, G. (2006). Evaluation of the persistence of micropollutants through pure-oxygen activated sludge nitrification and denitrification. *Water environment research*, *78*(11), 2276–2285.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with burrows–wheeler transform. *bioinformatics*, *25*(14), 1754–1760.

Li, Y., Li, S., Wang, J., & Liu, G. (2019). Crispr/cas systems towards next-generation biosensing. *Trends in biotechnology*, *37*(7), 730–743.

Liao, Y., Smyth, G. K., & Shi, W. (2014). featurecounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930.

Llewelyn, M. J., Fitzpatrick, J. M., Darwin, E., Gorton, C., Paul, J., Peto, T. E., Yardley, L., Hopkins, S., Walker, A. S., et al. (2017). The antibiotic course has had its day. *Bmj*, *358*.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, *15*(12), 1–21.

Ma, D., Shen, L., Wu, K., Diehnelt, C. W., & Green, A. A. (2018). Low-cost detection of norovirus using paper-based cell-free systems and synbody-based viral enrichment. *Synthetic Biology*, *3*(1), ysy018.

MacGowan, A., & Macnaughton, E. (2013). Antibiotic resistance. *Medicine*, *41*(11), 642–648.

Malek, L., Sooknanan, R., & Compton, J. (1994). Nucleic acid sequence-based amplification (nasba™). In *Protocols for Nucleic Acid Analysis by Nonradioactive Probes*, (pp. 253–260). Springer.

Martinsen, M. A., Jaramillo Cartagena, A., & Bhattacharyya, R. P. (2021). Core antibiotic-induced transcriptional signatures reflect susceptibility to all members of an antibiotic class. *Antimicrobial Agents and Chemotherapy*, *65*(6), e02296–20.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T., & Thomas, P. D. (2021). Panther version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive api. *Nucleic acids research*, *49*(D1), D394–D403.

Mills, E., Sullivan, E., & Kovac, J. (2021). Comparative analysis of bacillus cereus group isolates resistance using disk diffusion and broth microdilution and the correlation between resistance phenotypes and genotypes. *bioRxiv*.

Mölder, F., Jablonski, K. P., Letcher, B., Hall, M. B., Tomkins-Tinch, C. H., Sochat, V., Forster, J., Lee, S., Twardziok, S. O., Kanitz, A., et al. (2021). Sustainable data analysis with snakemake. *F1000Research*, *10*.

Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., & Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature methods*, *5*(7), 621–628.

NCBI (2021). Gene [internet]. *Bethesda (MD): National Library of Medicine (US)*.

Notomi, T., Okayama, H., Masubuchi, H., Yonekawa, T., Watanabe, K., Amino, N., & Hase, T. (2000). Loop-mediated isothermal amplification of dna. *Nucleic acids research*, *28*(12), e63–e63.

O'Neill, J. (2020). Antimicrobial resistance: tackling a crisis for the health and wealth of nations. december, 2014. *Commissioned by HM Government and the Wellcome Trust*.

Osaid, M., Chen, Y.-S., Wang, C.-H., Sinha, A., Lee, W.-B., Gopinathan, P., Wu, H.-B., & Lee, G.-B. (2021). A multiplexed nanoliter array-based microfluidic platform for quick, automatic antimicrobial susceptibility testing. *Lab on a Chip*.

Pál, C., Papp, B., & Lázár, V. (2015). Collateral sensitivity of antibiotic-resistant microbes. *Trends in microbiology*, *23*(7), 401–407.

Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., & Collins, J. J. (2014). based synthetic gene networks. *Cell*, *159*(4), 940–954.

Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., et al. (2016). Rapid, low-cost detection of zika virus using programmable biomolecular components. *Cell*, *165*(5), 1255–1266.

Paul, R., Ostermann, E., & Wei, Q. (2020). Advances in point-of-care nucleic acid extraction technologies for rapid diagnosis of human and plant diseases. *Biosensors and Bioelectronics*, *169*, 112592.

Piepenburg, O., Williams, C. H., Stemple, D. L., & Armes, N. A. (2006). Dna detection using recombination proteins. *PLoS biology*, *4*(7), e204.

Quach, D., Sakoulas, G., Nizet, V., Pogliano, J., & Pogliano, K. (2016). Bacterial cytological profiling (bcp) as a rapid and accurate antimicrobial susceptibility testing method for staphylococcus aureus. *EBioMedicine*, *4*, 95–103.

Reichert, J., Sakaitani, M., & Walsh, C. T. (1992). Characterization of entf as a serine-activating enzyme. *Protein Science*, *1*(4), 549–556.

Riehle, M. M., Bennett, A. F., Lenski, R. E., & Long, A. D. (2003). Evolutionary changes in heat-inducible gene expression in lines of escherichia coli adapted to high temperature. *Physiological genomics*, *14*(1), 47–58.

Rivers, E., Nguyen, B., Havstad, S., Ressler, J., Muzzin, A., Knoblich, B., Peterson, E., & Tomlanovich, M. (2001). Early goal-directed therapy in the treatment of severe sepsis and septic shock. *New England Journal of Medicine*, *345*(19), 1368–1377.

Rodriguez, N. M., Linnes, J. C., Fan, A., Ellenson, C. K., Pollock, N. R., & Klapperich, C. M. (2015). based rna extraction, in situ isothermal amplification, and lateral flow detection for low-cost, rapid diagnosis of influenza a (h1n1) from clinical specimens. *Analytical chemistry*, *87*(15), 7872–7879.

Rodríguez-Rojas, A., Rodríguez-Beltrán, J., Couce, A., & Blázquez, J. (2013). Antibiotics and antibiotic resistance: a bitter fight against evolution. *International Journal of Medical Microbiology*, *303*(6-7), 293–297.

Sangurdekar, D. P., Srienc, F., & Khodursky, A. B. (2006). A classification based framework for quantitative description of large-scale microarray data. *Genome biology*, *7*(4), 1–17.

Schurch, N. J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., Wrobel, N., Gharbi, K., Simpson, G. G., Owen-Hughes, T., et al. (2016). How many biological replicates are needed in an rna-seq experiment and which differential expression tool should you use? *Rna*, *22*(6), 839–851.

Shaifali, I., Gupta, U., Mahmood, S. E., & Ahmed, J. (2012). Antibiotic susceptibility patterns of urinary pathogens in female outpatients. *North American journal of medical sciences*, *4*(4), 163.

Shmakov, S., Abudayyeh, O. O., Makarova, K. S., Wolf, Y. I., Gootenberg, J. S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K., et al. (2015). Discovery and functional characterization of diverse class 2 crispr-cas systems. *Molecular cell*, *60*(3), 385–397.

Sorgeloos, P., Van Outryve, E., Persoone, G., & Cattoir-Reynaerts, A. (1976). New type of turbidostat with intermittent determination of cell density outside the culture vessel. *Applied and Environmental Microbiology*, *31*(3), 327–331.

Takahashi, M. K., Tan, X., Dy, A. J., Braff, D., Akana, R. T., Furuta, Y., Donghia, N., Ananthakrishnan, A., & Collins, J. J. (2018). A low-cost paper-based synthetic biology platform for analyzing gut microbiota and host biomarkers. *Nature communications*, *9*(1), 1–12.

Thakku, S. G., Ackerman, C. M., Myhrvold, C., Bhattacharyya, R. P., Livny, J., Ma, P., Gomez, G. I., Sabeti, P. C., Blainey, P. C., & Hung, D. T. (2021). Multiplexed detection of bacterial nucleic acids using cas13 in droplet microarrays. *bioRxiv*.

To, A. C.-Y., Chu, D. H.-T., Wang, A. R., Li, F. C.-Y., Chiu, A. W.-O., Gao, D. Y., Choi, C. H. J., Kong, S.-K., Chan, T.-F., Chan, K.-M., et al. (2018). A comprehensive web tool for toehold switch design. *Bioinformatics*, *34*(16), 2862–2864.

Toprak, E., Veres, A., Yildiz, S., Pedraza, J. M., Chait, R., Paulsson, J., & Kishony, R. (2013). Building a morbidostat: an automated continuous-culture device for studying bacterial drug resistance under dynamically sustained drug inhibition. *Nature protocols*, *8*(3), 555–567.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3—new capabilities and interfaces. *Nucleic acids research*, *40*(15), e115–e115.

Valeri, J. A., Collins, K. M., Ramesh, P., Alcantar, M. A., Lepe, B. A., Lu, T. K., & Camacho, D. M. (2020). Sequence-to-function deep learning frameworks for engineered riboregulators. *Nature communications*, *11*(1), 1–14.

van Belkum, A., Burnham, C.-A. D., Rossen, J. W., Mallard, F., Rochas, O., & Dunne, W. M. (2020). Innovative and rapid antimicrobial susceptibility testing systems. *Nature Reviews Microbiology*, *18*(5), 299–311.

Van Boeckel, T. P., Gandra, S., Ashok, A., Caudron, Q., Grenfell, B. T., Levin, S. A., & Laxminarayan, R. (2014). Global antibiotic consumption 2000 to 2010: an analysis of national pharmaceutical sales data. *The Lancet infectious diseases*, *14*(8), 742–750.

Wei, Q., Khan, I. K., Ding, Z., Yerneni, S., & Kihara, D. (2017). Navigo: interactive tool for visualization and functional similarity and coherence analysis with gene ontology. *Bmc Bioinformatics*, *18*(1), 1–13.

WHO (2020). Global antimicrobial resistance surveillance system (glass) report: early implementation 2020. 2020. *Contract*, (978-92), 4–000558.

WHO (2021). Who policy guidance on integrated antimicrobial stewardship activities.

Wolfe, B. R., Porubsky, N. J., Zadeh, J. N., Dirks, R. M., & Pierce, N. A. (2017). Constrained multistate sequence design for nucleic acid reaction pathway engineering. *Journal of the American Chemical Society*, *139*(8), 3134–3144.

Wong, B. G., Mancuso, C. P., Kiriakov, S., Bashor, C. J., & Khalil, A. S. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with evolver. *Nature biotechnology*, *36*(7), 614–623.

Zankari, E., Hasman, H., Cosentino, S., Vestergaard, M., Rasmussen, S., Lund, O., Aarestrup, F. M., & Larsen, M. V. (2012). Identification of acquired antimicrobial resistance genes. *Journal of antimicrobial chemotherapy*, *67*(11), 2640–2644.

## CURRICULUM VITAE