

Reverse Engineering and Synthetic Perturbation of the Transcriptional Regulatory Network of *Streptococcus* *pneumoniae* to Identify Routes to Antibiotic Potentiation

Matthew Robert Hall

A dissertation for PhD
submitted to the Faculty of
the department of Biology
in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Boston College
Morrissey College of Arts and Sciences
Graduate School

March 2025

© Copyright 2025 Matthew Robert Hall

Reverse Engineering and Synthetic Perturbation of the Transcriptional Regulatory Network of *Streptococcus pneumoniae* to Identify Routes to Antibiotic Potentiation

Matthew Robert Hall

Advisor: Juan C. Ortiz-Marquez, Ph.D.

Abstract

Streptococcus pneumoniae is a bacterial pathogen responsible for over 1 million deaths annually, with 30% of pneumococcal infections in the United States resistant to one or more antibiotics. Here, we characterize the Transcriptional Regulatory Network (TRN) of *S. pneumoniae* strain TIGR4 to understand the regulatory logic of this pathogen and identify strategies for enhancing antibiotic efficacy. We constructed 89 Transcription Factor Induction (TFI) strains, each of which overexpresses a barcoded and HA-tagged transcription factor in response to IPTG. Through a combination of RNAseq, an ensemble of network inference methods, and both *in vivo* and computational motif discovery, we reverse engineered the most comprehensive TRN for *S. pneumoniae* to date, comprised of 3,946 interactions between 88 transcription factors and 1,607 target genes. Transcriptional Regulator Induced Phenotype (TRIP) screens quantify relative fitness values between TFI strains and identify transcriptional pathways critical to metabolism, oxidative stress, zinc intoxication and antibiotic stress. Overexpression of three metabolic regulators, CpsR, SusR and TreR, increase resistance to the cell wall synthesis inhibitors Ceftriaxone and Vancomycin while potentiating the effects of Kanamycin and Levofloxacin. Metal and redox-regulated transcription factors were frequent TRIP hits and are best exemplified by RitR, the redox-sensitive master regulator of metal ion homeostasis, protease activity and

the oxidative stress response. The apparent link between metal ions and oxidative stress led to the identification of zinc as a potentiator of both Kanamycin and Ceftriaxone. Time-kill assay-adapted TRIP screens identified proteostasis as a primary determinant of antibiotic tolerance. Overexpression of the Clp protease repressor CtsR enhanced tolerance, while overexpression of CodY reduced it, an effect reversed by branched-chain amino acid supplementation. This work provides the first comprehensive TRN of *S. pneumoniae* to understand its physiology and identify drug targets, while insights from TRIP screens reveal strategies to potentiate antibiotic treatment of this critical pathogen.

TABLE OF CONTENTS

Table of Contents	v
List of Tables	viii
List of Figures	ix
Abbreviations	xi
Acknowledgements	xiii
Introduction	1
1.1 Antibiotics and Antibiotic Resistance	2
1.2 Overcoming Antibiotic Resistance	5
1.3 Transcription Factors and Antibiotic Resistance	6
1.4 Overview of this Dissertation	8
1.5 References	10
Construction and Validation of our Synthetic Transcription Factor Overexpression System	13
2.0 Introduction	14
2.1 Modern Approaches to Study Gene Regulation	14
2.2 Our Transcription Factor Induction (TFI) System	18
2.2.1 Identification and Classification of Transcription Factors in TIGR4	18
2.2.2 Construction of Our Transcription Factor Induction System	21
2.3 Validation of the TFI Inducible System	26
2.3.1 IPTG Induces Transcription Factor Overexpression.....	26
2.3.2 Transcription Factor Overexpression Alters the Growth Profiles of TFI Strains	27
2.4 Discussion	31
2.5 Materials and Methods	34
2.6 References	37
Reverse Engineering of the Transcriptional Regulatory Network of TIGR4 Aided by Transcription Factor Overexpression	40
3.0 Introduction	41
3.1 Transcription Factor Overexpression Identifies Broad Transcriptomic Changes	43
3.1.1 Constructs of our Synthetic System	44
3.1.2 IPTG Alters the Transcriptome of TIGR4	45
3.1.3 Differential Expression Analysis Fails to Identify All Known Interactions	47
3.2 An Ensemble of Network Inference Methods Predicts Regulatory Interactions based on Covarying Expression	49

3.2.1	Compiling Gene Expression Data for Network Inference	50
3.2.2	Identifying Optimal Combinations of Network Inference Methods	54
3.3	Identification of Transcription Factor Motifs	
	Differentiates Direct and Indirect Regulation	57
3.3.1	ChIPseq Identifies the Binding Motifs of Select Transcription Factors	58
3.3.2	Computational Discovery of Binding Motifs	62
3.4	The Transcriptional Regulatory Network of TIGR4	66
3.4.1	Transcription Factor Activities under Antibiotic Stress	74
3.5	Discussion	76
3.6	Materials and Methods	81
3.7	References	85
 Synthetic Perturbation of the Transcriptional Regulatory Network of TIGR4 Identifies Routes to Antibiotic Potentiation 94		
4.0	Introduction	95
4.1	Metabolic TRIP Screens	98
4.1.1	Metabolic and Oxidative Stress Response Regulators Influence Fitness in Plain Media	98
4.1.2	Fitness Effects of Transcription Factor Overexpression are Obscured in Galactose due to Constitutive Expression of the Leloir Pathway	103
4.1.3	Carbon Catabolite Repression Limits Fitness Effects in GlcNAc	105
4.1.4	Overexpression of <i>manA</i> Increases Fitness when Catabolizing Mannose	106
4.2	Antibiotic Resistance TRIP Screens	108
4.2.1	Overexpression of Certain Metabolic Regulators Increases Resistance to Cell Wall Synthesis Inhibitors	108
4.2.2	Oxidative Stress Response Regulators Influence Kanamycin and Levofloxacin Susceptibilities	113
4.3	Specific Stresses can be Characterized through Abiotic TRIP Screens	117
4.4	Time-kill Assay Adapted TRIP Screens Reveal Transcriptional Regulators' Influence on Antibiotic Tolerance	123
4.5	Comparative Analysis of TRIP Screens Reveals Common Pathways of the Stress Response	135
4.6	Discussion	138
4.7	Materials and Methods	146
4.8	References	149
 Conclusion 154		
5.0	Motivation	155
5.1	Identification and Classification of	

	Transcription Factors in TIGR4	156
5.2	Constructing and Validating	
	Transcription Factor Induction Strains	156
5.3	Reverse Engineering the Transcriptional	
	Regulatory Network of TIGR4	157
	5.3.1 TFI RNAseq Fails to Identify All Regulatory Interactions	158
	5.3.2 An Ensemble of Network Inference Methods	
	Complements TFI RNAseq	158
5.4	Network Precision Improves with	
	Requiring DNA Binding Motifs	159
5.5	The Transcriptional Regulatory Network of TIGR4	161
5.6	TRIP Screens Quantify Fitness Effects of TRN Perturbations	162
	5.6.1 TRIP screens in Alternate Carbon Sources	
	Connect TRN to Metabolic Network	162
	5.6.2 Overexpression of CpsR, TreR and SusR Increase Resistance	
	to CWSI but Increase Susceptibility to	
	Kanamycin and Levofloxacin	163
	5.6.3 The Paramount Importance of Metal Homeostasis and	
	Redox Regulators in the Stress Response	164
	5.6.4 TRIP screens Identify the Determinants of	
	Antibiotic Tolerance	166
5.7	Concluding Remarks	168
5.8	References	169
	Supporting Data	171

LIST OF TABLES

Semi-Defined Minimal Media (SDMM) Recipe	35
Table 1: Network Inference Method Overview	55
Table 2: Transcriptional Regulatory Network Module Enrichments	70
Table 3: Summary of Transcription Factors in TIGR4	70
Minimum Inhibitory Concentrations (MIC)	146

LIST OF FIGURES

Ch1: Introduction

Figure 1: Overview of Antibiotics and Antibiotic Resistance	3
Figure 2: Overview of Gene Regulation by Transcription Factors	7
Figure 3: Organization of this Dissertation	9

Ch2: Construction and Validation of our Synthetic Transcription Factor Overexpression System

Figure 1: Comparison of Methods to Study Gene Regulation	17
Figure 2: Transcription Factor Overview in TIGR4	21
Figure 3: Overview and Cloning of TFI Strains	25
Figure 4: Expression Validation of Select TFI Strains	27
Figure 5: Growth Curve Analysis of TFI Strains	30

Ch3: Reverse Engineering of the Transcriptional Regulatory Network of TIGR4 Aided by Transcription Factor Overexpression

Figure 1: Transcriptional Regulation helps Define Cell State	42
Figure 2: Summary of Transcription Factor Overexpression	45
Figure 3: IPTG Stimulated Genes	46
Figure 4: TFI RNAseq Fails to Identify All Known Target Genes as DEGs	48
Figure 5: Removal of Lowly-Expressed Genes from the Dataset	51
Figure 6: UMAP Representation of our Three Datasets	53
Figure 7: Performance of Individual and Combinations of Inference Methods	56
Figure 8: Direct vs. Indirect Regulation	58
Figure 9: TFI ChIPseq Results	61
Figure 10: Comparison of Binding Motifs Identified by XSTREME and ChIPseq	64
Figure 11: Statistics of Five Networks	65
Figure 12: Scale-Free Network Topology	67
Figure 13: The Transcriptional Regulatory Network of TIGR4	69
Figure 14: Transcription Factor Activity During Antibiotic Treatment	75

Ch4: Synthetic Perturbation of the Transcriptional Regulatory Network of TIGR4 Identifies Routes to Antibiotic Potentiation

Figure 1: Overview of TRIP Screens	98
Figure 2: Metabolic TRIP Screen Results	102
Figure 3: Cell Wall Synthesis Inhibitor TRIP Screen Results	112
Figure 4: Kanamycin and Levofloxacin TRIP Screen Results	115
Figure 5: ROS TRIP Screen Results	120
Figure 6: Zinc Intoxication TRIP Screen Results	122

Figure 7: Tolerance TRIP Screen Overview	126
Figure 8: Tolerance TRIP Screen Results	134
Figure 9: CodY Regulates Glucose Metabolism and the Oxidative Stress Response	135
Figure 10: Meta Analysis of TRIP Screen Results	137
Figure 11: Zinc Potentiates Ceftriaxone and Kanamycin Treatments	142

List of Abbreviations

ADS	Arginine Deiminase System
ATc	Anhydrotetracycline
ATP	Adenosine Triphosphate
AUPR	Area Under Precision Recall
BCAA	Branched Chain Amino Acid
CCR	Carbon Catabolite Repression
CEF	Ceftriaxone
ChIPseq	Chromatin Immunoprecipitation and Sequencing
CWSI	Cell Wall Synthesis Inhibitor
DEG	Differentially Expressed Gene
DREAM	Dialogue on Reverse Engineering Assessment and Methods
DSI	DNA Synthesis Inhibitor
EMSA	Electrophoretic Mobility Shift Assay
FBP	Fructose 1,6-Biphosphate
GlcNAc	N-Acetylglucosamine
HA	Hemagglutinin
HPr	Histidine Protein
HRP	Horseradish Peroxidase
IgG	Immunoglobulin
ILV	Isoleucine, Leucine, Valine
IPTG	Isopropyl β -D-1-Thiogalactopyranoside
KAN	Kanamycin
LEVO	Levofloxacin
LHA	Left Homology Arm
MarR	Multiple Antibiotic Response Regulator
MDR	Multidrug Resistant

MIC	Minimum Inhibitory Concentration
NAM	N-Acetylmuramic Acid
Neu5Ac	N-Acetylneuraminic acid
OEPCR	Overlap Extension Polymerase Chain Reaction
P2TF	Predicted Prokaryotic Transcription Factors
Padj	P-adjusted Value
PBP	Penicillin Binding Protein
PCR	Polymerase Chain Reaction
PQ	Paraquat
PSI	Protein Synthesis Inhibitor
qPCR	Quantitative Polymerase Chain Reaction
RBS	Ribosomal Binding Site
RHA	Right Homology Arm
RNAseq	RNA Sequencing
ROS	Reactive Oxygen Species
RSI	RNA Synthesis Inhibitor
SDMM	Semi-Defined Minimal Media
TA	Toxin-Antitoxin
TCS	Two Component System
TF	Transcription Factor
TFI	Transcription Factor Induction
TPM	Transcripts per Million
TRIP	Transcriptional Regulator Induced Phenotype
TSS	Transcriptional Start Site
UTR	Untranslated Region
VNC	Vancomycin
XRE	Xenobiotic Response Element

ACKNOWLEDGEMENTS

I would like to acknowledge and express my gratitude to everyone who helped make this all possible, starting with and especially my advisor Dr. Juance Ortiz-Marquez. I'd also like to thank all members of the Microbial Systems Biology lab, especially Dr. Bimal Jana, Dr. Federico Rosconi, Dr. Noemí Buján Gómez, Dr. Luisa Nieto Ramirez, Dr. Suyen Espinosa, Dr. Stephen Hummel, Sophie Bodrog and Jon Anthony. Special thanks to Nicoletta Collison, Warren Lin and Sydney Sanchez Tejada for their help with the work in this dissertation.

I sincerely appreciate the guidance and feedback of my Thesis Committee, including Dr. Marc-Jan Gubbels, Dr. Michelle Meyer, Dr. Eddie Geisinger, Dr. Bimal Jana and Dr. Babak Momeni. Your expertise and guidance have been invaluable throughout this process.

I'd also like to thank all members of the Biology Department here at Boston College for all their help and support over the past six years, with special thanks to Dr. Welkin Johnson, Dr. Charlie Hoffman, Dina Goodfriend, Diane Butera, Patrick Chatfield, Collette McLaughlin, Kiel Smith, Jaclyn Girouard and Danny O'Leary. I'd also like to thank Andrew Lonero from the Center for Isotope Geochemistry at Boston College for helping us perform Mass Spectrometry.

Finally, I'd like to thank my family, Mom, Dad, David, Brian and especially my fiancé Emily, for all of the love and support all of these years.

Chapter 1:

Introduction

1.0 Introduction

Life today is the result of 3.8 billion years of continuous natural selection of cells adherent to the fundamental laws of nature. The composition and conditions of primitive Earth facilitated the spontaneous formation of organic molecules^{1–3}, the inherent chemical properties of which created the first cells: self-replicating ribozymes within fatty acid micelles^{4,5}. Selection for the most catalytically efficient and robust ribozymes initiated optimization by natural selection, with competition for resources soon following. This process of evolution by natural selection, uninterrupted for billions of years, is responsible for designing all domains of life and is the defining principle of biology.

As the defining principle of biology, it is unsurprising that natural selection is critically important to medicine. Cancer cells, for example, have a long history of evolving resistance to therapy^{6,7}, and one of today's biggest challenges in oncology is treating drug-resistant malignancies^{8,9}. The zoonotic event that initiated the COVID-19 pandemic allowed us to witness the evolutionary arms race between the virus and our adaptive immune systems in real time. Initially, immunologically naïve hosts were severely infected and treated with monoclonal antibodies, while adaptive immunity through vaccination and prior infection fought back against the virus. This increasing immunity led to the emergence of vaccine- and treatment-resistant variants^{10,11}, necessitating the development of updated vaccines and treatments, and around we go.

1.1 Antibiotics and Antibiotic Resistance

A particularly intricate example of natural selection's influence on medicine can be seen through the antibiotic treatment of bacterial pathogens. Antibiotics themselves are natural chemicals produced by microbes to inhibit the growth of competitors that we have industrialized for use in the clinic. Sir Alexander Flemming's serendipitous discovery of penicillin on a fungi-contaminated Petri dish is often regarded as the origin of this discovery, but several groups also reported the production of antibiotics by microorganisms^{12–14}. The realization that microbes evolved to produce clinically useful antibiotics led to a systematic search for producers of antibiotics and ushered in the Golden Age of Antibiotic Discovery. From 1928 to 1960, 17 families of antibiotics were discovered of bacterial or fungal origin and were adapted for the clinic¹⁴. Solving the chemical structures of these antibiotics and advancements in synthetic chemistry led to the development of semi-synthetic derivatives and, eventually, fully synthetic antibiotics^{14,15}. What were previously deadly bacterial infections could now be treated, and a wide range of medical procedures was now safeguarded, including organ transplantation, cancer treatment and open-heart surgery¹⁴. Overall, human life expectancy increased 30 years across the 20th century¹⁶, with the discovery and widespread clinical use of antibiotics in large part to thank.

All antibiotics target essential cellular processes and either inhibit bacterial growth or kill the target bacteria (Fig. 1A). Many, such as penicillin, act as cell wall synthesis inhibitors (CWSI) by preventing peptidoglycan crosslinking, though the specific molecular targets vary across antibiotic family. Similarly, antibiotics like daptomycin disrupt bacterial

membranes, causing membrane depolarization and ion leakage. Other antibiotics interfere with the central dogma of biology, including DNA synthesis inhibitors (DSI), RNA synthesis inhibitors (RSI) and protein synthesis inhibitors (PSI). The combined diversity of antibiotic targets is crucial for effectively treating all bacterial infections, as some antibiotics have disparate activities against Gram-positive and Gram-negative bacteria, along with other species specificities. Regardless of their mechanism, all antibiotics are classified as either bacteriostatic, which inhibit bacterial growth, or bactericidal, which directly kill bacteria.

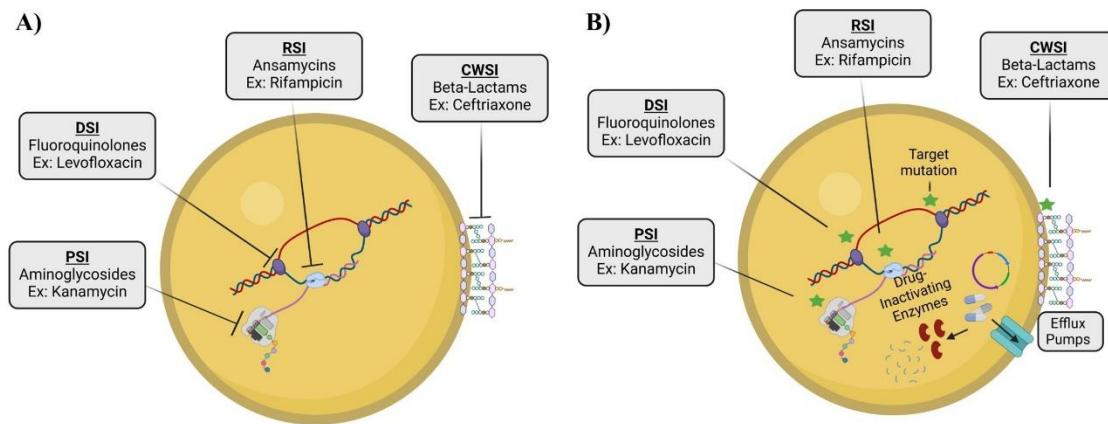


Figure 1. Overview of Antibiotics and Antibiotic Resistance. **A)** Diagram of common antibiotic mechanisms of action for DNA synthesis inhibitors (DSI), RNA synthesis inhibitors (RSI), Protein synthesis inhibitors (PSI) and Cell Wall synthesis inhibitors (CWSI). **B)** Common mechanisms of antibiotic resistance .

Despite their differences, all antibiotics apply an intense selective pressure and resistant mutants rapidly emerge (Fig. 1B). A primary mechanism for antibiotic resistance is through drug target mutations. Mutations are frequently found in DNA topoisomerase in DSI-resistant strains^{17–19}, RNA polymerase for RSI-resistant strains^{20,21}, ribosomes for PSI-resistant strains^{22,23} and penicillin-binding proteins for CWSI-resistant strains^{24,25}. As antibiotics are natural products, bacterial pathogens had evolved resistance mechanisms

even prior to clinical use of antibiotics. The most infamous example of this is penicillinase, a β -lactamase that directly degrades penicillin (a β -lactam antibiotic), which was discovered in penicillin-resistant *Staphylococcus aureus* strains even before penicillin was adapted to the clinic²⁶. Along with actively degrading antibiotics, other antibiotic resistance genes chemically modify antibiotics to inhibit their activity^{27,28}. Alarmingly, transposable elements and plasmids are known to shuttle several antibiotic resistance genes through horizontal gene transfer, simultaneously granting resistance to several antibiotics^{28,29}. This, along with other mechanisms such as increased efflux pump expression^{30,31}, has led to the emergence of multi-drug resistant (MDR) strains.

In addition to developing or acquiring antibiotic resistance, bacteria can survive antibiotic treatment through tolerance and persistence. Antibiotic tolerance and persistence are related phenomena where bacteria endure antibiotic stress by severely downregulating the cellular processes targeted by antibiotics, both of which are linked with slowed growth³². This phenomenon of slow growth hindering antibiotic treatment is best exemplified in the difficulty in eradicating *Mycobacterium tuberculosis*, whose extremely slow growth rates require months of combinatorial antibiotic treatment. Antibiotic tolerance is a population level trait acquired by genetic mutation, while persistence occurs only in a small subpopulation of clonal cells and is therefore regulatory. Alarmingly, antibiotic tolerance and persistence have been shown to precede the development of resistance through genetic mutation and/or horizontal gene transfer³³.

Bacterial infections remain a leading cause of death worldwide, responsible for an estimated 17 million deaths per year¹⁶. The emergence of multidrug resistance forced the

World Health Organization to recognize antibiotic resistance as the fifth greatest threat to human health in 2019³⁴. Unfortunately, the nature of antibiotic treatment—short treatment durations, rapid resistance development, and the need to reserve new antibiotics as last resorts—has prevented significant investment from the pharmaceutical industry. Research into enhancing the efficacy of existing antibiotics and targeting tolerant and persistent cells is crucial to overcoming antibiotic resistance.

1.2 Overcoming Antibiotic Resistance

A promising method to enhance the efficacy of antibiotics is to directly target the mechanism of resistance, either through directly inhibiting the resistance gene or preventing its dissemination. Several studies have identified inhibitors of β -lactamases and efflux pumps, and combinatorial treatment with β -lactam antibiotics and β -lactamase inhibitors have been approved for clinical use^{35–37}. Similarly, inhibitors of conjugation, plasmid replication or segregation have been utilized for their ability to remove plasmid-based resistance genes from bacterial populations³⁸.

An alternative approach is to manipulate bacterial physiology to enhance antibiotic efficacy. As seen in persisters, physiological states can influence susceptibility to antibiotics, and a systematic understanding of which states are most and least susceptible to antibiotics can be used to rationally design antibiotic potentiators. Aminoglycoside treatment of persister cells was shown to be potentiated by specific metabolic stimulation, which in turn activated the proton-motive force required for aminoglycoside uptake³⁹. Recently, pyrimidine supplementation was shown to potentiate antibiotic killing by stimulating carbon metabolism to produce ATP for purine synthesis⁴⁰. Both findings were

dependent on a mechanistic understanding of the metabolic network and demonstrate that rational manipulation of the physiological state of bacterial cells undergoing antibiotic treatment can improve efficacy.

1.3 Transcription Factors and Antibiotic Resistance

All living cells modulate their physiology in response to stimuli to enhance survival and reproductive success. A primary mechanism for this is transcriptional regulation, where the expression of genes is altered to maximize fitness in the current environment. Changes in gene expression are orchestrated through the action of transcription factors, proteins that bind specific DNA-sequences, called operators or motifs, to influence gene expression (Fig. 2). Transcription factors can be activators, repressors or bifunctional depending on context, and are frequently regulated by binding chemical signals or through post-translational modification. Transcription factors regulate all aspects of cellular physiology, including carbon metabolism, nucleotide synthesis, amino acid synthesis, cell wall synthesis, cell division, competence, virulence and various stress responses. Characterization of all transcription factors in a cell creates a transcriptional regulatory network (TRN), which describes the regulatory logic behind cellular physiology.

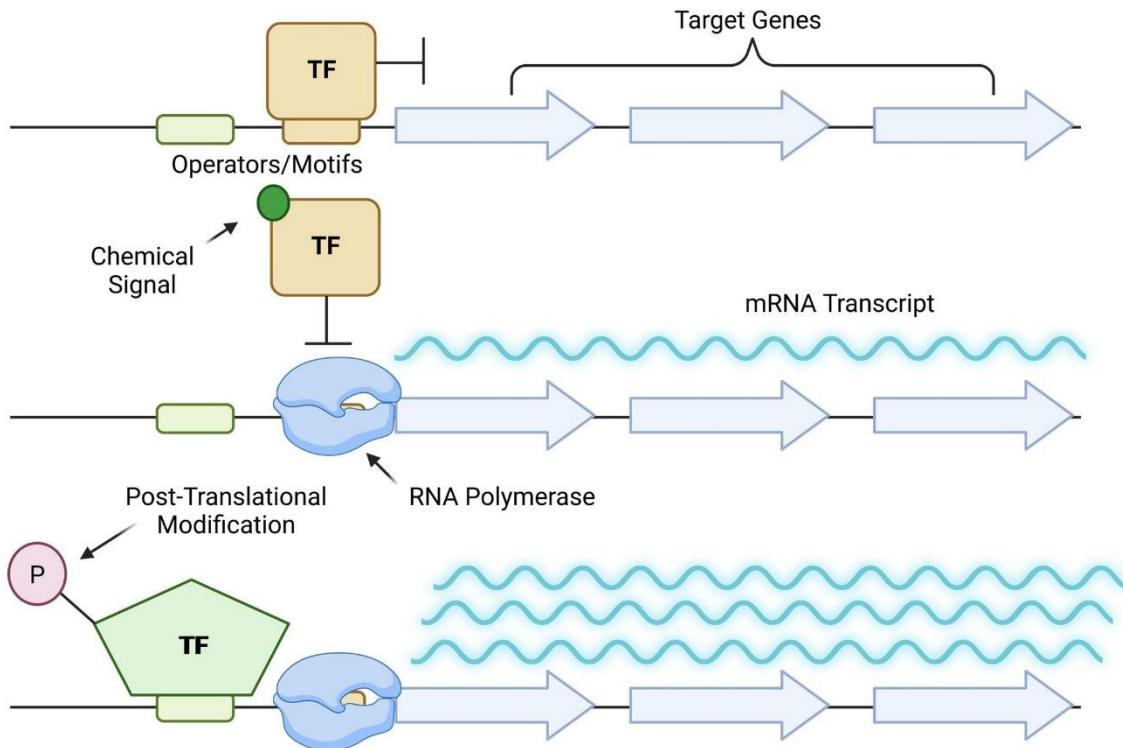


Figure 2. Overview of Gene Regulation by Transcription Factors (TF).

Characterizing the regulatory logic of pathogenic bacteria is crucial for understanding how they regulate virulence, adapt to different niches in the host, survive both immune and antibiotic stress and even develop antibiotic resistance. The transcription factors that regulate these processes can be pharmaceutically targeted and can provide context-specific antibiotic potentiation. A recent study identified a chemical inhibitor of the master regulator of virulence in *S. aureus*, MgrA, that was shown to attenuate virulence *in vivo* and act synergistically with antibiotic treatment⁴¹. Similarly, competence, the state in which bacteria are capable of DNA uptake and assimilation, can be prevented by treatment with competitive inhibitors of the signaling molecule that triggers the transcriptional process of becoming competent⁴². This has the potential to minimize horizontal gene transfer and the development of antibiotic resistance. A characterized TRN not only can identify drug

targets, but can also explain resistance determinants that occur through mutation of transcription factors or their operator sites^{43,44}.

1.4 Overview of this Dissertation

For these reasons, we aimed to characterize the transcriptional regulatory network of *Streptococcus pneumoniae* and identify how this network influences antibiotic susceptibilities. *S. pneumoniae* is a leading cause of community-acquired pneumonia, the seventh leading cause of death in the United States^{45–47}. *S. pneumoniae* is regularly found as a commensal in the nasopharynx, but can travel to secondary sites within the body and cause invasive-pneumococcal diseases such as pneumonia, meningitis, otitis media, osteomyelitis and bacteremia/sepsis. Alarmingly, 30% of pneumococcal infections in the United States are resistant to one or more clinically relevant antibiotics⁴⁸, emphasizing the need to gain a better understanding of antibiotic resistance and identify potentiators. Unlike more common model systems like *E. coli* and *S. aureus*, the transcriptional regulatory network of *S. pneumoniae* has not been characterized. Individual transcription factors have been studied but through varying techniques in different strains, which creates difficulties in reconstructing a comprehensive network. We hypothesize that a mechanistic understanding of global gene regulation and how the transcriptome affects antibiotic efficacy would allow for the rational design of antibiotic potentiators.

We begin by comparing different methods for uncovering transcriptional regulation in bacteria with an emphasis on methods capable of constructing a comprehensive regulatory network and quantifying how transcription factor activity influences antibiotic susceptibility. We identify transcription factor overexpression as the optimal approach and

we construct and validate a synthetic transcription factor overexpression strain for each transcription factor identified in our model strain, TIGR4. Then, we reverse engineer the most comprehensive transcriptional regulatory network for *S. pneumoniae* to date through an ensemble of network inference methods trained on a large, diverse gene expression dataset with binding motif discovery and manual curation. Finally, we perform competition assays between all transcription factor overexpression strains to quantify the fitness effects that alternative transcriptomic states have on various conditions and stresses including both antibiotic resistance and tolerance. This work provides a comprehensive overview of transcriptional regulation in *S. pneumoniae* strain TIGR4 and identifies which transcription factors influence fitness under various stressors providing insight towards specific routes to antibiotic potentiation.

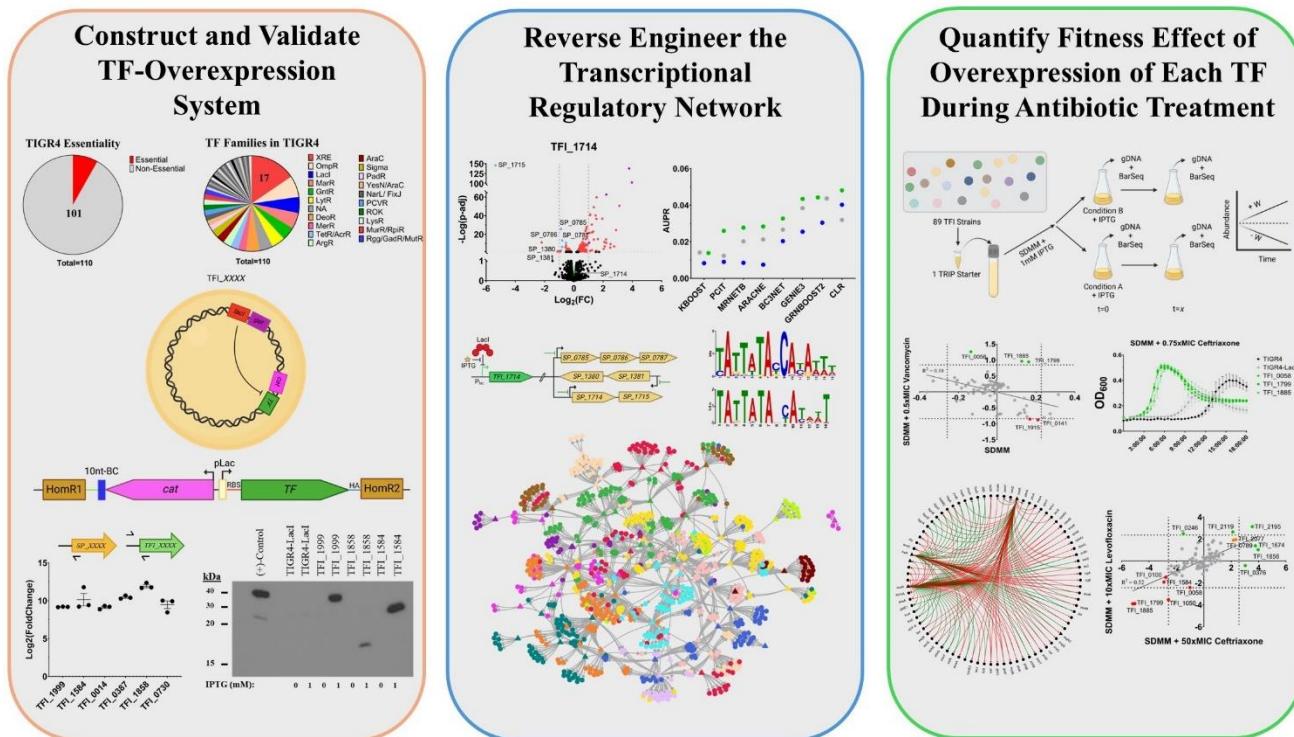


Figure 3. Organization of this Dissertation

1.5 References

1. Miller, S. L. A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* (80-). **117**, 528–529 (1953).
2. ORÓ, J. Mechanism of Synthesis of Adenine from Hydrogen Cyanide under Possible Primitive Earth Conditions. *Nature* **191**, 1193–1194 (1961).
3. Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
4. Babajanyan, S. G. *et al.* Coevolution of reproducers and replicators at the origin of life and the conditions for the origin of genomes. *Proc. Natl. Acad. Sci.* **120**, (2023).
5. Cooper, G. The Origin and Evolution of Cells. in *The Cell: A Molecular Approach* (Sinauer Associates, 2000).
6. Farber, S., Diamond, L. K., Mercer, R. D., Sylvester, R. F. & Wolff, J. A. Temporary Remissions in Acute Leukemia in Children Produced by Folic Acid Antagonist, 4-Aminopteroxy-Glutamic Acid (Aminopterin). *N. Engl. J. Med.* **238**, 787–793 (1948).
7. REICHARD, P., SKÖLD, O. & KLEIN, G. Possible Enzymic Mechanism for the Development of Resistance against Fluorouracil in Ascites Tumours. *Nature* **183**, 939–941 (1959).
8. Cui, Q. *et al.* Modulating ROS to overcome multidrug resistance in cancer. *Drug Resist. Updat.* **41**, 1–25 (2018).
9. Han, H. *et al.* Mesenchymal and stem-like prostate cancer linked to therapy-induced lineage plasticity and metastasis. *Cell Rep.* **39**, 110595 (2022).
10. Collier, D. A. *et al.* Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).
11. Hoffmann, M. *et al.* The Omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic. *Cell* **185**, 447–456.e11 (2022).
12. FLEMING, A. The Discovery of Penicillin. *Br. Med. Bull.* **2**, 4–5 (1944).
13. Frost, W. D. The Antagonism Exhibited By Certain Saprophytic Bacteria Against The Bacillus Typhosus Gaffky. *J. Infect. Dis.* **1**, 599–640 (1904).
14. Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80 (2019).
15. Mitcheltree, M. J. *et al.* A synthetic antibiotic class overcoming bacterial multidrug resistance. *Nature* **599**, 507–512 (2021).
16. Martens, E. & Demain, A. L. The antibiotic resistance crisis, with a focus on the United States. *J. Antibiot. (Tokyo)* **70**, 520–526 (2017).
17. Muñoz, R. & De La Campa, A. G. ParC subunit of DNA topoisomerase IV of *Streptococcus pneumoniae* is a primary target of fluoroquinolones and cooperates with DNA gyrase A subunit in forming resistance phenotype. *Antimicrob. Agents Chemother.* **40**, 2252–2257 (1996).
18. Fukuda, H. & Hiramatsu, K. Primary Targets of Fluoroquinolones in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **43**, 410–412 (1999).
19. Drlica, K. & Zhao, X. DNA gyrase, topoisomerase IV, and the 4-quinolones. *Microbiol. Mol. Biol. Rev.* **61**, 377–392 (1997).

20. Telenti, A. *et al.* Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet* **341**, 647–651 (1993).
21. Jin, D. J. & Gross, C. A. Mapping and sequencing of mutations in the *Escherichia coli* *rpoB* gene that lead to rifampicin resistance. *J. Mol. Biol.* **202**, 45–58 (1988).
22. Apirion, D. & Schlessinger, D. Coresistance to Neomycin and Kanamycin by Mutations in an *Escherichia coli* Locus that Affects Ribosomes. *J. Bacteriol.* **96**, 768–776 (1968).
23. Pihlajamäki, M. *et al.* Ribosomal mutations in *Streptococcus pneumoniae* clinical isolates. *Antimicrob. Agents Chemother.* **46**, 654–8 (2002).
24. Ubukata, K., Nonoguchi, R., Matsuhashi, M. & Konno, M. Expression and inducibility in *Staphylococcus aureus* of the *mecA* gene, which encodes a methicillin-resistant *S. aureus*-specific penicillin-binding protein. *J. Bacteriol.* **171**, 2882–5 (1989).
25. Fishovitz, J., Hermoso, J. A., Chang, M. & Mabashery, S. Penicillin-binding protein 2a of methicillin-resistant *Staphylococcus aureus*. *IUBMB Life* **66**, 572–577 (2014).
26. ABRAHAM, E. P. & CHAIN, E. An Enzyme from Bacteria able to Destroy Penicillin. *Nature* **146**, 837–837 (1940).
27. Shaw, W. V. The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant *Escherichia coli*. *J. Biol. Chem.* **242**, 687–93 (1967).
28. Courvalin, P. & Carlier, C. Transposable multiple antibiotic resistance in *Streptococcus pneumoniae*. *Mol. Gen. Genet. MGG* **205**, 291–297 (1986).
29. Carattoli, A. Plasmids and the spread of resistance. *Int. J. Med. Microbiol.* **303**, 298–304 (2013).
30. McMurry, L., Petrucci, R. E. & Levy, S. B. Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **77**, 3974–3977 (1980).
31. Nikaido, H. Multidrug Resistance in Bacteria. *Annu. Rev. Biochem.* **78**, 119–146 (2009).
32. Balaban, N. Q. *et al.* Definitions and guidelines for research on antibiotic persistence. *Nat. Rev. Microbiol.* **17**, 441–448 (2019).
33. Levin-Reisman, I. *et al.* *Antibiotic tolerance facilitates the evolution of resistance*. <http://science.sciencemag.org/>.
34. World Health Organization. *Ten Threats to Global Health in 2019*. (2019).
35. Sharma, A., Gupta, V. K. & Pathania, R. Efflux pump inhibitors for bacterial pathogens: From bench to bedside. *Indian J. Med. Res.* **149**, 129–145 (2019).
36. Chawla, M., Verma, J., Gupta, R. & Das, B. Antibiotic Potentiators Against Multidrug-Resistant Bacteria: Discovery, Development, and Clinical Relevance. *Front. Microbiol.* **13**, (2022).
37. Darby, E. M. *et al.* Molecular mechanisms of antibiotic resistance revisited. *Nat. Rev. Microbiol.* **21**, 280–295 (2023).
38. Vrancianu, C. O., Popa, L. I., Bleotu, C. & Chifiriuc, M. C. Targeting Plasmids to Limit Acquisition and Transmission of Antimicrobial Resistance. *Front. Microbiol.* **11**, (2020).
39. Allison, K. R., Brynildsen, M. P. & Collins, J. J. Metabolite-enabled eradication of bacterial persisters by aminoglycosides. *Nature* **473**, 216–220 (2011).
40. Yang, J. H. *et al.* A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* **177**, 1649–1661.e9 (2019).

41. Guo, X. *et al.* Thwarting resistance: MgrA inhibition with methylophipogonanone unveils a new battlefield against *S. aureus*. *npj Biofilms Microbiomes* **10**, 15 (2024).
42. Koirala, B. & Tal-Gan, Y. Development of *Streptococcus pneumoniae* Pan-Group Quorum-Sensing Modulators. *ChemBioChem* **21**, 340–345 (2020).
43. Kong, H.-K. *et al.* Fine-tuning carbapenem resistance by reducing porin permeability of bacteria activated in the selection process of conjugation. *Sci. Rep.* **8**, 15248 (2018).
44. Beggs, G. A., Brennan, R. G. & Arshad, M. MarR family proteins are important regulators of clinically relevant antibiotic resistance. *Protein Sci.* **29**, 647–653 (2020).
45. Cillóniz, C., Garcia-Vidal, C., Ceccato, A. & Torres, A. Antimicrobial Resistance Among *Streptococcus pneumoniae*. in *Antimicrobial Resistance in the 21st Century* 13–38 (Springer International Publishing, 2018). doi:10.1007/978-3-319-78538-7_2.
46. File, T. M. & Ramirez, J. A. Community-Acquired Pneumonia. *N. Engl. J. Med.* **389**, 632–641 (2023).
47. Dion, C. & Ashurst, J. *Streptococcus pneumoniae*. *StatPearls* <https://www.ncbi.nlm.nih.gov/books/NBK470537/> (2023).
48. *Antibiotic resistance threats in the United States, 2019*. <https://stacks.cdc.gov/view/cdc/82532> (2019) doi:10.15620/cdc:82532.

Chapter 2:
Construction and Validation of our Synthetic
Transcription Factor Overexpression System

2.0 Introduction

In the 1950s, François Jacob and Jacques Monod pioneered the study of gene regulation by describing the first mechanism of transcriptional regulation. Investigating lactose metabolism in *E. coli*, they identified the enzyme responsible for hydrolyzing lactose to galactose and glucose, β -galactosidase. Hypothesizing that β -galactosidase would be induced by its reactant, they showed that several galactosides, including lactose and isopropyl- β -D-thiogalactoside (IPTG), increased expression of β -galactosidase in *E. coli*. Painstaking purification and biochemical analysis of β -galactosidase and its induction in other bacterial species revealed that some strains constitutively express β -galactosidase. This differential regulation of β -galactosidase expression across bacterial species led Jacob and Monod to postulate that a regulator gene represses expression of the β -galactosidase gene through binding an operator sequence, and this repression is inhibited through interaction with the inducer molecule^{1,2}. Though their proposed model of gene regulation was met with significant criticism, Jacob and Monod's description of transcriptional regulation of *E. coli*'s *lac* operon withstood the test of time and continues to be taught as a textbook example of gene regulation.

2.1 Modern Approaches to Study Gene Regulation

Today, with the massive increase in computational power, it is critical to characterize transcriptional regulation on a genome-wide scale to understand cellular behavior and, eventually, enable computational modeling of cells. Bacteria have evolved not as a collection of distinct modules capable of performing a specific function, but rather as an interconnected array of networks designed to adapt to and survive changing environments.

As these networks define how a bacterium behaves, a comprehensive understanding of how this regulatory array is designed and how it affects the cell's physiology would allow us to manipulate these bacteria for our own gain.

Thanks to technological advancements, uncovering transcriptional regulation and understanding their role in bacterial systems has become widespread. Gene expression profiling, first through DNA microarrays and now next generation sequencing, has enabled genome-wide identification of the mechanisms of transcriptional regulation. RNA sequencing of bacterial cultures catabolizing different sugar sources, for example, can identify which transcription factor(s) regulate expression of specific metabolic pathways. Such studies have identified the regulators of alternate metabolic sources⁴, competence development⁵, infection-relevant stresses⁶, and antibiotic stresses⁷. Further investigation into the regulatory mechanism of these transcription factors, through assays such as DNase footprinting, ChIPseq and DAPseq, can identify their consensus sequence to confirm regulation of the suspected process as well as extrapolate that transcription factor's role across the rest of the genome. Recently, several statistical methods⁸⁻¹² have been developed to reverse engineer regulator networks from gene expression data.

Alternatively, many studies have opted to knock out individual transcription factors to identify the resulting change to the transcriptome and the mutant's phenotype. In *S. pneumoniae* alone, these studies have identified the regulons of the global nutritional regulator CodY¹³, the lactose metabolic regulator LacR¹⁴, the zinc efflux regulator and several two-component systems, including PnpRS¹⁵, CiaRH¹⁶, and the orphan response regulator, RitR¹⁷. Transcription factor knockouts have the added benefit of being able to

phenotype each mutant and identify its role in various physiological processes including virulence and the stress response. For example, *ciaRH* mutants were shown to have attenuated carriage and virulence *in vivo*, a phenotype attributed in part to its regulation of the serine protease, *htrA*¹⁶. Similarly, *nmlR* mutants were implicated in the oxidative stress response through regulation of the alcohol dehydrogenase *adhC* through transcriptomic analysis and then demonstrated to have impaired survival inside human macrophages¹⁸. A significant problem with relying on transcription factor knockouts is that the most important transcription factors, those that are essential, are precluded. In fact, the study that knocked out CodY above was only able to do so because of multiple suppressor mutations¹⁹, calling into question the validity of the study.

More recently, advances in synthetic biology have allowed for both transcription factor overexpression and transcription factor knockdowns. Transcription factor overexpression can amplify its regulatory signature and uncover its biological significance. IPTG-induced expression of a proteolytic-resistant *spx* mutant in *B. subtilis* revealed its role in activating the oxidative stress response while repressing primary metabolism under stress conditions²⁰. Similarly, transcriptional overexpression of the master regulators of flagellar genes in *E. coli*, *flhDC*, extended its role to regulating the oxidative phosphorylation and pentose phosphate pathways²¹. Rustad *et al.* extrapolated this approach to a genome-wide scale by cloning transcription factor overexpression strains for 206 transcription factors in *M. tuberculosis*²². The transcriptomic regulatory signature was identified for 183 of these transcription factors by microarray analysis and the consensus binding signature of 57 of these transcription factors was determined by ChIPseq, creating a comprehensive

transcriptional regulatory network for *M. tuberculosis*²³. Importantly, each transcription factor overexpression strain included a unique DNA barcode, which allowed for phenotypic assays to reveal that overexpression of *mce3R* significantly potentiates the activity of the frontline anti-tuberculosis drug, isoniazid²⁴.

Alternatively, gene knockdown by CRIPSRi and small RNAs (sRNA) have become popular methods for high-throughput screening of diverse cellular processes. These methods rely on repressing transcription and translation, respectively, of a single gene in a large library of mutants to identify the fitness effect each gene has on the tested process. A recent CRIPSRi screen in *S. pneumoniae* demonstrated that knockdown of the response regulator *vicR* lead to irregular cell shape and size and attributed this to its known regulation of *pcsB*²⁵. While transcription factors are targets of these methods and can be phenotypically evaluated, to our knowledge no knockdown has been used to identify the regulon of a transcription factor.



Figure 1. Comparison of Methods to Study Gene Regulation.

2.2 Our Transcription Factor Induction (TFI) System

To accomplish our aims of both mapping the transcriptional regulatory network of *S. pneumoniae* and identifying how the network influences antibiotic efficacy, we decided to create an inducible transcription factor overexpression system. For each transcription factor in *S. pneumoniae* TIGR4's genome, we attempted to clone a Transcription Factor Induction (TFI) strain that is inducible with IPTG, contains a C-terminal HA-tag for ChIPseq and has a unique DNA barcode. A combination of RNAseq and ChIPseq on all TFI strains will allow us to reverse engineer the transcriptional regulatory network, and barcode-assisted competition assays will allow us to identify the regulons that influence antibiotic outcomes. While exact network composition may differ strain to strain, it's important to note that the transcriptional regulatory network of different strains has evolved under similar selective pressures. Therefore, while the exact mechanisms may be different, the overarching principles are likely the same. Differences between the regulatory networks of different strains of *S. pneumoniae* may help explain strain characteristics, while conserved networks indicate essential regulatory processes that may be amenable to therapeutic intervention, especially those involving regulons shown to influence antibiotic outcomes.

2.2.1 Identification and Classification of Transcription Factors in TIGR4

To study transcriptional regulation in TIGR4 by its repertoire of transcription factors, we must first identify which genes are in fact transcription factors. This task had been made the focus of several research groups recently with the aim of identifying transcription factors in microbial communities for synthetic biology^{26–28}. In 2012, Ortet *et al.* created Predicted Prokaryotic Transcription Factors (P2TF), a comprehensive database for prokaryotic transcription factors identified through Reverse Position-Specific BLAST

analysis aimed at identifying conserved DNA-binding domains²⁶. According to P2TF, there are 108 transcription factors in TIGR4. Newer methods, namely PredicTF²⁷ and TransFacPred²⁸, have curated their own databases of transcription factors and trained machine learning models to predict transcription factors in novel genomes. Unfortunately, outdated dependencies prevent users from using PredicTF, and the accuracy of TransFacPred for bacterial transcription factors is questionable as several genes that are clearly annotated as not being transcription factors, including transposases, tRNA deaminases and PTS transporters , were classified as transcription factors. Therefore, we took the list of 108 predicted transcription factors in P2TF, manually curated it by removing genes known not to be transcription factors and identified false negatives by scraping gene annotations that included “helix-turn-helix”, “transcriptional regulator” and “repressor”. Our final list includes 110 putative transcription factors in TIGR4.

Of these 110 transcription factors, only nine are essential in TIGR4 according to Tn-Seq²⁹ (Fig. 2A). These essential transcription factors include the carbon catabolite repressor regulator, CcpA, the global nutritional regulator CodY, both sigma factors (RpoD and ComX), the response regulators VicR and RitR and the cysteine metabolic regulator, CmbR. RpoD is the primary (or housekeeping, sigma70) sigma factor required for transcription initiations as it associates with RNA polymerase and directing the holoenzyme to promoter sequences through recognition of the conserved -10 element. ComX is the only alternate sigma factor in *S. pneumoniae* and it directs the RNA polymerase holoenzyme to the promoters of late genes during competence. There are two identical copies of *comX* in TIGR4 (*SP_0014* and *SP_2006*) and both are classified as essential, although it's unclear if the cell needs both copies. Many of the other essential

transcription factors regulate global metabolism (CcpA, CodY, CmbR) or coordinate global stress responses (VicR, RitR). The final two essential transcription factors in TIGR4 are sangYiaG and SP_0333 which may be cryptic prophage repressors.

Transcription factors can be classified into families based on homology to the prototypical transcription factor first discovered, usually in *E. coli* or *B. subtilis*. Understanding the occurrences of each family within a species can give insight into the regulatory complexity of specific processes. For example, the Xenobiotic Response Element (XRE)-family of transcription factors are known to be involved in stress response with an apparent link to the oxidative stress response^{30,31}. XRE is the most common family found in TIGR4 (Fig. 2B) which is indicative of a transcriptional regulatory network that has evolved under oxidative stress. The second most common family is OmpR, which stems from the 13 two component systems (TCS) found in TIGR4, followed by LacI and MarR families.

Not every gene is present in all strains of a bacterial species. The total genome comprised of all unique genes across all strains of a species is referred to as a species' pangenome. Genes that are a part of the "core" pangenome are found in all strains, while "accessory" genes are found in some strains but not others. Core genes are considered more important than accessory genes because of their conservation across strains. Of the 110 transcription factors in TIGR4, 64 are a part of the core genome and 46 are accessory genes (Fig. 2C). Gene essentiality can also be considered across a pangenome, where universally essential genes are essential across all strains, strain-dependent essential genes are only essential in certain strains and non-essential genes are never essential. Of the 110 transcription factors in TIGR4, only three are universally essential: CodY, RpoD and VicR (Fig. 2D).

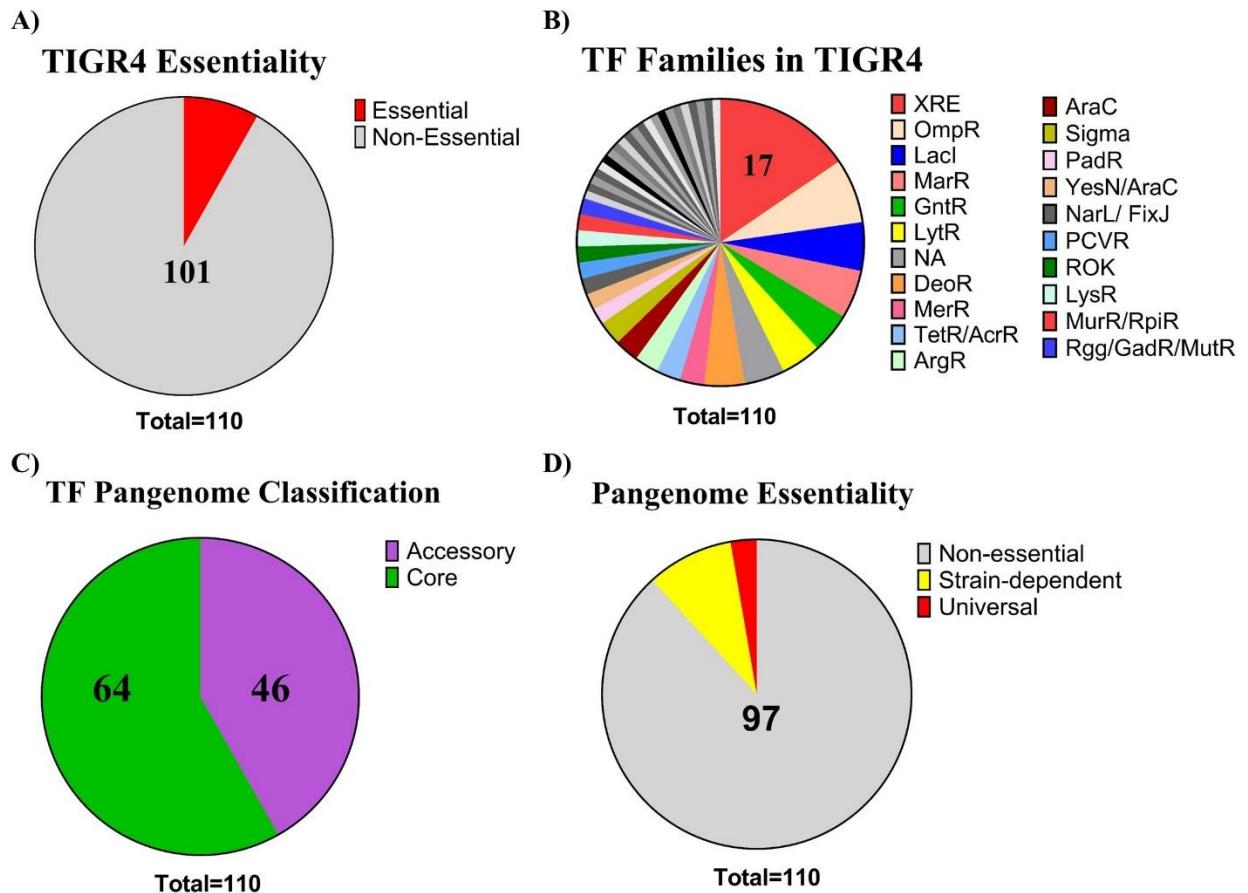


Figure 2. Transcription Factor overview in *S. pneumoniae* strain TIGR4. **A)** Essentiality of all transcription factors (TFs) in TIGR4 according to Tn-Seq. **B)** Distribution of TF families in TIGR4. Only families with 2 or more members shown in legend. **C)** TF essentiality across *S. pneumoniae*'s pangenome according to Tn-Seq. **D)** Classification of each TF in TIGR4 within *S. pneumoniae*'s pangenome.

2.2.2 Construction of our Transcription Factor Induction System

The ideal synthetic system to accomplish our goals is one that is hypersensitive to its inducer and has a large dynamic range. Luckily for us, a recent paper by Sorg *et al.* engineered several inducible gene-networks in *S. pneumoniae*³². The most common small molecule inducible systems in bacteria are based on the TetR and LacI repressors from *E. coli* and are inducible with IPTG and anhydrotetracycline (ATc), respectively. To adapt

these systems to *S. pneumoniae*, Sorg *et al.* first codon optimized *lacI* and *tetR* and cloned them into *S. pneumoniae* under control of a strong constitutive promoter, PF6, along with an optimal ribosomal binding site. Then, several plasmids were engineered with varying *lacO* and *tetO* sites regulating expression of *luciferase* and *green fluorescent protein*. These plasmids were transformed into *lacI*- and *tetR*-expressing *S. pneumoniae* strains, and the dynamic range of each synthetic system was quantified through plate reader assays at varying concentrations of inducer. We decided to use an IPTG-inducible system because of the risk that ATc could induce stress to the cells that would obfuscate future antibiotic assays on these strains. The IPTG-inducible promoter with the largest dynamic range was PL8-2, which contains two *lacO* sites, one in the proximal region (overlapping the transcriptional start site) and one 72.5 bases upstream (measuring center to center of *lacO* sites) in the distal region of its promoter. We therefore decided to design our TFI strains using the PL8-2 inducible promoter.

To clone our TFI strains, we first needed to create the parental strain, TIGR4-LacI, which constitutively expresses the codon-optimized *lacI* repressor. To do so, we simply transformed the plasmid pPEPY-PF6-LacI³² into *S. pneumoniae* strain TIGR4 replacing a transposase (*SP_0028*). TIGR4-LacI is gentamycin-resistant, constitutively-expresses *lacI* and will serve as the parental strain for all TFI strains as well as a control strain for many experiments (Fig. 3A).

Identifying a high-throughput and robust cloning method for TFI strains was more difficult. Initial attempts to clone the PL8-2 regulated transcription factor cassettes with plasmids through *E. coli* failed as several regulators, namely CcpA (SP_1999) and CodY (SP_1584), appear to be toxic to *E. coli* at high concentrations. Every *E. coli* clone sequenced carrying

plasmids of TFI_1999 and TFI_1584 had a frameshift mutation that aligned within the first ~8 amino acids in the coding sequence of the transcription factor. We suspect that these mutations were the result of truncated primer sequences, and the selective pressures of carrying high-copy number plasmids expressing repressor-less PL8-2 controlled global metabolic regulators prevented accurate clones.

We eventually identified an ideal cloning method based on sequential overlap-extension polymerase chain reaction (OEPCR) (Fig. 3B). OEPCR relies on homologous sequences at the ends of DNA molecules, allowing them to anneal on rare occasions during thermocycling. PCR reactions with compatible DNA molecules and DNA polymerases can extend the annealed fragments, and a secondary PCR with primers designed to amplify the combined amplicon selects for the full-length product.

First, each transcription factor was amplified by PCR out of TIGR4's genome using TFI_XXXX_F1/R1 primers that amplify the entire coding sequence of the transcription factor (except the stop codon) and encode a ribosomal binding site (RBS) and HA-tag. The XXXX refers to the locus tag of the transcription factor (i.e., TFI_1999_F1/R1 amplify *ccpA* (*SP_1999*) out of the genome). The left homology arm (LHA) was itself cloned through a two-part OEPCR that produces a full-length product including homology arm 1, a barcoded chloramphenicol resistance gene, the PL8-2 promoter and an RBS. The final product of this reaction contains a library of barcoded amplicons, the sequence of which we don't know until sequencing the final TFI strain. The right homology arm (RHA) was amplified out of the genome and contains homology arm 2 and the HA-tag. The three amplicons: LHA, TF and RHA were then input into a final OEPCR reaction that combined the three based on the RBS shared between the LHA and TF and the HA-tag shared between the TF

and RHA. The final product was selected by amplification with TFI_U_Transform_F/R and then transformed into TIGR4-LacI at the *yabE* (*SP_0648*) locus. Each transformant was Sanger sequenced and was selected as a final TFI strain if it had a completely accurate promoter, coding sequence and HA-tag, as well as a DNA barcode that had a hamming distance of greater than two from all other barcodes of TFI strains. Through this method, we were able to successfully clone TFI strains for 89 out of the 110 transcription factors identified in TIGR4.

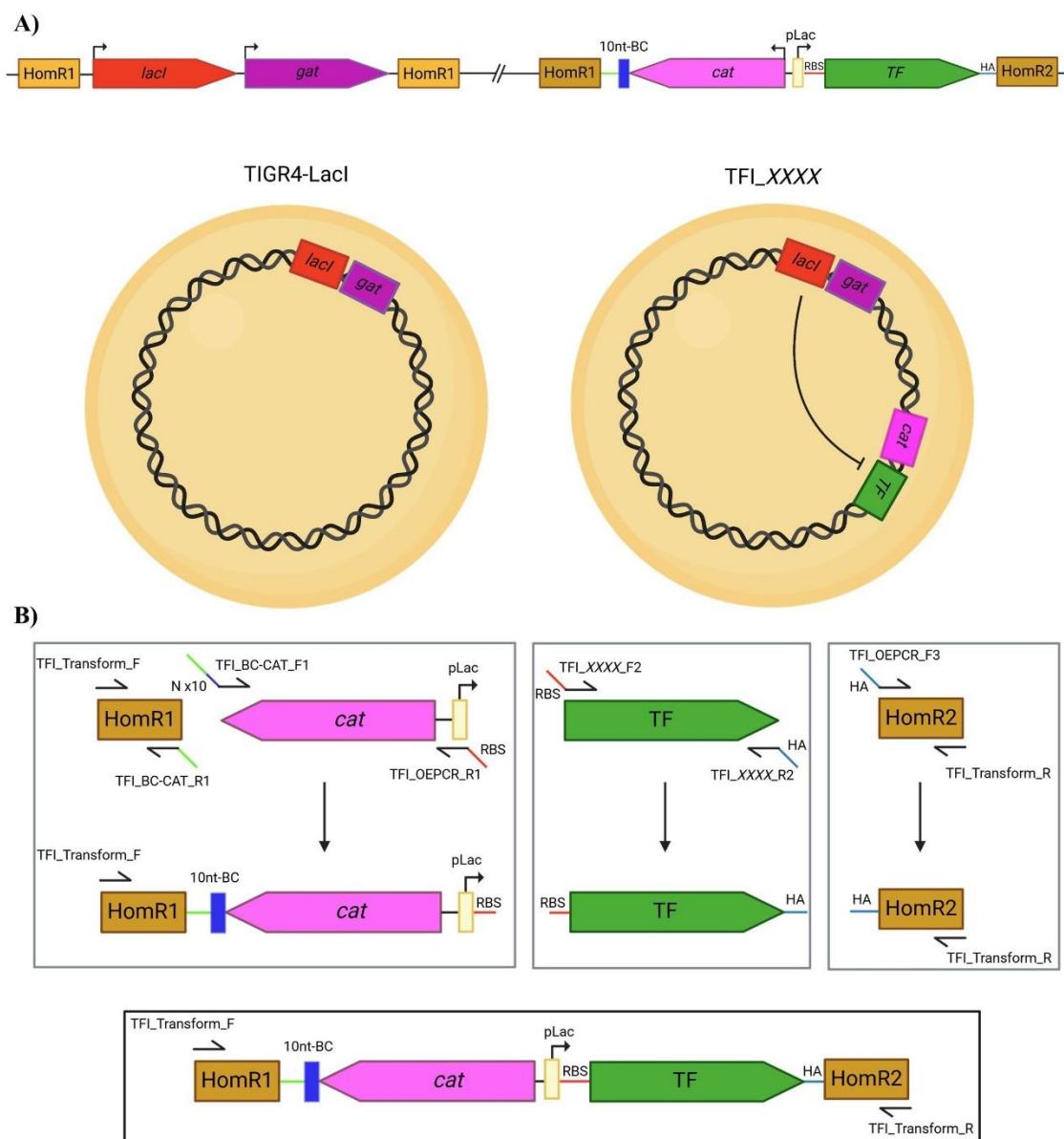


Figure 3. Overview and Cloning Method of TFI Strains. **A)** Schematic of the genomic organization of TFI strains. **B)** Overview of the three-part overlap extension PCR method for TFI strain cloning. Half-arrows represent primers. HomR = Homology Arm. *gat* = gentamycin acetyltransferase. *cat* = chloramphenicol acetyltransferase. BC = barcode. RBS = Ribosomal binding site. HA = Hemagglutinin-tag.

2.3 Validation of the TFI Inducible System

2.3.1 IPTG Induces Transcription Factor Overexpression

To ensure that the inducible system works as designed, we assayed both the transcriptional and translational inducibility of TFI strains by quantitative PCR (qPCR) and western blot, respectively (Fig. 4). Every TFI strain maintains the native copy of its respective transcription factor, so we designed qPCR primers that specifically amplify the synthetic copy of the gene by aligning the forward primer with the TFI 5' untranslated region (UTR). Because every TFI strain has the same 5'-UTR, a single forward primer could be matched with a reverse primer aligning to the beginning of the coding sequence of the transcription factor to amplify only the synthetic copy of the gene. After an hour of induction with 1mM IPTG, all six TFI strains tested showed approximately a ten log₂-fold change in expression (Fig. 4A), demonstrating the large dynamic range of the PL8-2 promoter.

The inducible system was shown to work via qPCR, but to ensure there were no issues in translation of our inducible genes, we performed western blot analysis on several TFI strains (Fig. 4B). Western blots quantify protein expression through a three-step process that includes separation of proteins by size via SDS-PAGE, transferring protein to a membrane, and protein detection/visualization/quantification with antibodies. Every overexpressed transcription factor was designed to encode a C-terminal HA-tag, so to verify protein expression we performed a western blot using an Anti-HA-Horseradish Peroxidase (HRP) antibody. The Anti-HA antibody will recognize and bind our HA-tags and the HRP can be used for visualization through chemiluminescence. Four strains, TIGR4-LacI, TFI_1999, TFI_1858 and TFI_1584, were cultured for one hour in SDMM +

1mM IPTG . All TFI strains demonstrated protein expression only when induced with IPTG, reaffirming a successful inducible system and verifying faithful protein expression.

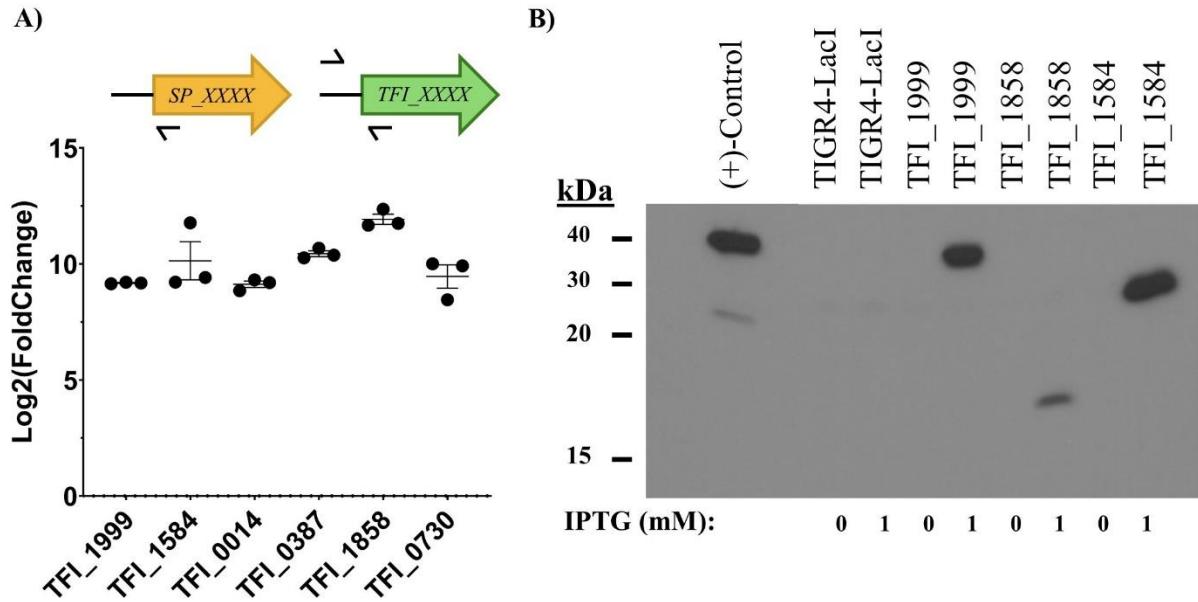


Figure 4. Expression Validation of select TFI Strains. **A)** Top) Schematic showing specificity of TFI qPCR primers. Bottom) qPCR results showing the log₂(fold-change) in expression of the synthetic copy of select transcription factors after 1h of 1mM IPTG induction. **B)** Western blots of TIGR4-LacI and select TFI strains with or without induction by IPTG for 1h.

2.3.2 Transcription Factor Overexpression Alters the Growth Profiles of TFI Strains

With the inducible system validated, we wanted to determine if transcription factor overexpression altered each TFI strains' phenotype. TIGR4, TIGR4-LacI and all TFI strains were cultured in SDMM, SDMM + 10µM IPTG and SDMM + 1mM IPTG with optimal density readings every 30 minutes to create a growth curve profile of all strains across a range of inducer concentrations (Fig. 5A). The synthetic transcription factors should be repressed, slightly expressed and fully expressed without IPTG, with 10µM IPTG and with 1mM IPTG, respectively³². Comparisons of the doubling times, carrying

capacity and stationary phase lengths of all strains at maximum induction reveal that transcription factor overexpression alters the growth profile of the strains (Fig. 5B). The growth profiles of TFI strains induced with a range of inducer concentrations revealed patterns in phenotype changes due to transcription factor overexpression (Fig. 5C). Uninduced strains did show strain-to-strain differences, suggesting leaky expression from the inducible promoters.

All TFI strains can be classified into one or more growth curve profiles: Little-to-No Change, Induced Lyse Earlier, Induced Lyse Later, Diauxic Shift and Rapid Biphasic Lysis (Fig. 5C,D). The most dramatic changes seen in growth profiles were the length of the stationary phase and the induction of autolysis. *S. pneumoniae* undergoes autolysis during stationary phase through the action of LytA which promotes virulence, biofilm formation and may provide resources for surviving cells^{33–35}. LytA achieves autolysis by hydrolyzing the bond between the stem peptides and glycan strands of the cell wall. This process has been shown to be activated when the rate of cell wall degradation outpaces the rate of cell wall synthesis, either through metabolic limitation or antibiotic stress, as well as during extended acidic stress^{33,34}.

The most common growth profile was Induced Lyse Earlier, where the cells induced with IPTG undergo autolysis prior to uninduced conditions and have a shorter stationary phase. This may be indicative of either a reduction in the rate of cell-wall synthesis or an increase in acidity due to transcription factor overexpression. Interestingly, 11 strains both lysed earlier when induced with IPTG and displayed a “diauxic shift” which can be seen as a growth rate blip within exponential phase (Fig. 5D). Diauxic shifts were first described when identifying that bacteria sequentially catabolize available carbon sources in order of

preference rather than sequentially, and the diauxic shift is seen as a short lag during which cells alter expression of metabolic pathways^{1,2}. The overlap between strains that lyse earlier when induced and have a diauxic shift lead us to hypothesize that the earlier lysis is due to altered metabolism in the cell, which may produce a more acidic environment and increase the activity of LytA. Conversely, 12 induced strains underwent autolysis later than when uninduced.

Three strains, TFI_0058, TFI_1799 and TFI_1885, displayed a Rapid Biphasic Lysis growth profile, with an extremely short stationary phase followed by a rapid but incomplete autolysis. These strains overexpress the metabolic regulators CpsR, SusR and TreR, which have been implicated in regulating capsule expression, sucrose metabolism and trehalose metabolism, respectively^{36–38}. This phenotype could be due to either an increased acidic environment or an altered cell wall composition. Interestingly, in Chapter 4, all three strains will be shown to have significantly increased resistance to cell wall synthesis inhibitors and increased susceptibility to Kanamycin and Levofloxacin.

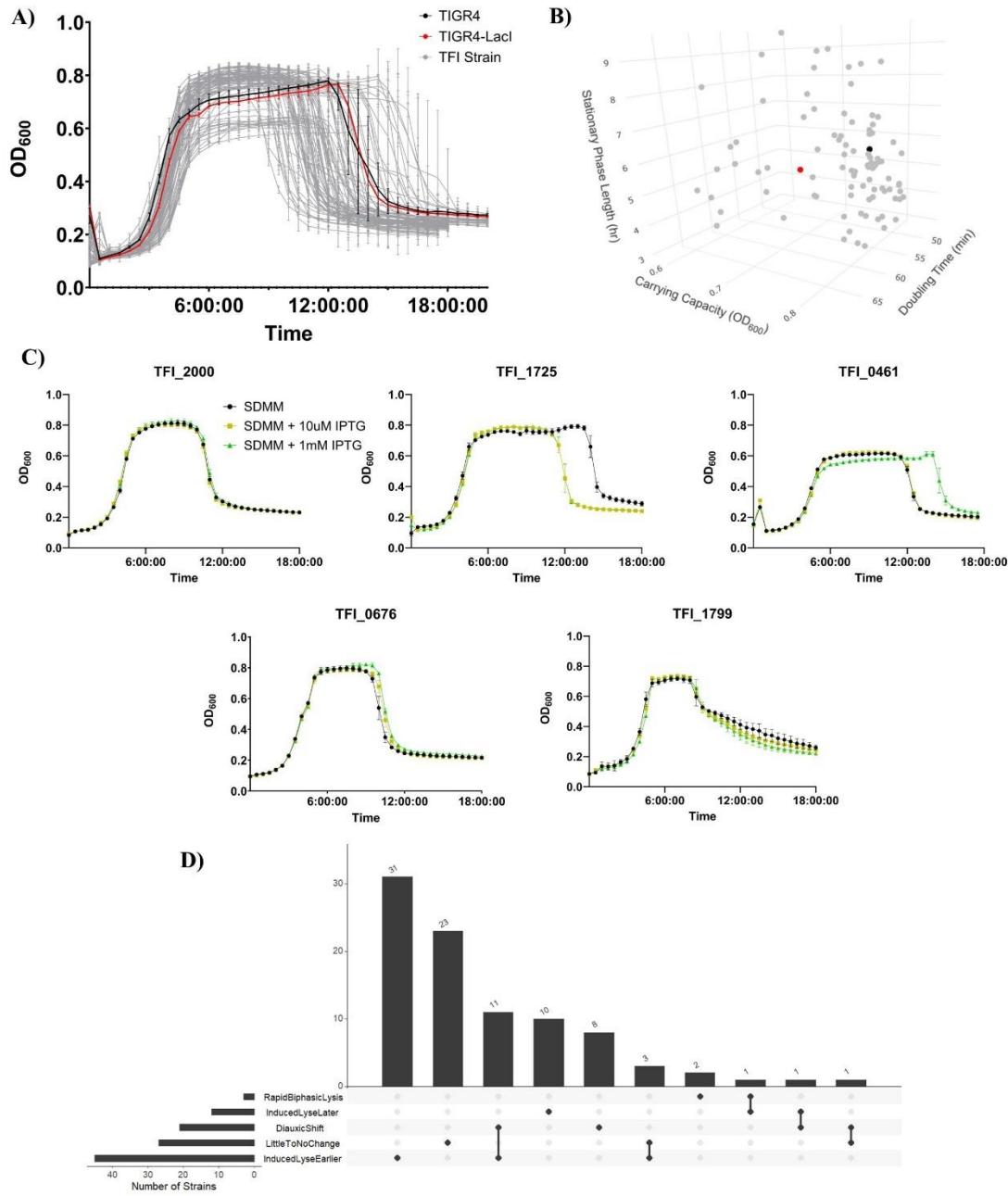


Figure 5. Growth Curve Analysis of TFI Strains. **A)** Optical density plots showing all TIGR4 (black), TIGR4-LacI (red) and all TFI strains (grey) cultured in SDMM + 1mM IPTG. Mean of three replicates shown, error bars represent standard error. **B)** 3D-Scatter plot comparing Doubling Times, Carrying Capacity and Stationary Phase Length for results shown in A. **C)** Representative optical density plots for the five classes of growth curve profiles in SDMM (black), SDMM + 10 μ M IPTG (yellow) and SDMM + 1mM IPTG (green): Top) Little-to-No Change, Induced Lyse Earlier, Induced Lyse Later; Bottom) Diauxic Shift, Rapid Biphasic Lysis. **D)** UpSet plot summarizing growth curve profiles of TFI strains. Vertical bars represent how many strains were classified by the one or two profiles selected below the x-axis. Horizontal bars represent total number of strains to classify as that profile.

2.4 Discussion

Thanks to advancements in research technology, identifying mechanisms of transcriptional regulation in bacteria and their influence on cellular phenotypes has become widespread and achievable through various approaches. Commons approaches include targeted gene expression profiling, transcription factor knockouts, transcription factor knockdowns and transcription factor overexpression systems. Due to its ability to both uncover transcriptional regulation on a genome-wide scale and determine fitness effects of transcription factor overexpression, we chose to utilize a transcription factor overexpression system to characterize the transcriptional regulatory network of TIGR4.

Fortunately, a recent study by Sorg *et al.* designed several synthetic gene networks in *S. pneumoniae* which could be adapted for our cause. We decided to design our Transcription Factor Induction (TFI) system using the PL8-2 promoter from this study because it had the largest dynamic range of all IPTG-inducible promoters tested. We designed a high-throughput cloning method based on overlap-extension PCR that we used to clone TFI strains for 89 transcription factors in TIGR4. Every TFI strain overexpresses a barcoded and HA-tagged transcription factor in response to IPTG. We were unable to attain TFI mutants for several transcription factors (e.g. *SP_0593*, *SP_0661*) whose natural copy is near the locus (*SP_0648*) in which we cloned the synthetic cassettes. We hypothesize this is due to the inhibition of homologous recombination with now three homologous regions (LHA, RHA and the TF coding region).

With our library of TFI strains, we confirmed successful design of the induction system through TFI-specific qPCR and Anti-HA Western blot analysis. TFI-specific qPCR was achieved by designing a TFI-specific forward primer that selectively amplifies the

synthetic copy of the transcription factor. Gene expression increases approximately 1,000-fold after an hour of 1mM IPTG induction according to qPCR, and protein bands were only detectable in TFI strains induced with IPTG.

After verification of the synthetic system, we wanted to see if transcription factor overexpression altered the phenotypes of TFI strains. To achieve this, we profiled the growth of all TFI strains, along with wild-type TIGR4 and TIGR4-LacI, in plain SDMM and SDMM with two concentrations of inducer: 10 μ M and 1mM. There were differences in the doubling times, carrying capacity and stationary phase lengths across TFI strains, demonstrating that we are altering the phenotypes of these strains. Interestingly, all TFI strains could be categorized as one or more of five recurring growth profiles: Little-to-No Change, Induced Lyse Earlier, Induced Lyse Later, Diauxic Shift and Rapid Biphasic Lysis.

The most noticeable differences between TFI strains were their stationary phase lengths and induction of autolysis. Autolysis is achieved through cell wall hydrolysis of LytA, which has been previously shown to be induced by inhibition of cell wall synthesis (either through starvation or antibiotic activity) and by acidic stress. Strains with altered LytA activity may have altered metabolism resulting in different pH or cell wall composition. An extreme example of altered autolysis can be seen in the three TFI strains that display a Rapid Biphasic Lysis phenotype: TFI_0058, TFI_1799 and TFI_1885. These three strains overexpress the metabolic regulators CpsR, SusR and TreR, which regulate capsule expression, sucrose metabolism and trehalose metabolism, respectively. While CpsR is named for its repression of the capsule promoter, it is highly expressed and will be shown in Chapter 3 as being one of the dominant global regulators in *S. pneumoniae*. Interestingly, these same three TFI strains will be shown in Chapter 4 as being resistant to the cell wall

synthesis inhibitors ceftriaxone and vancomycin. Metabolomics on these TFI strains may uncover novel insights linking pneumococcal metabolism to acid stress and/or cell wall composition with implications for antibiotic treatments.

While most TFI strains showed growth curve differences with and without IPTG, it's important to note that the growth profiles of TFI strains differ even without IPTG. This is likely due to leaky expression of the synthetic copy of the transcription factor, which may produce exacerbated effects due to its regulatory activity. For example, many transcription factors are known to regulate expression of their own promoters. Therefore, leaky expression of the synthetic copy may produce amplified effects by regulating its native copy's promoter, leading to larger transcriptomic and phenotypic changes. In hindsight, an ATc-inducible system may have provided better repression when uninduced, but as our focus is on regulation of the antibiotic stress response, we were concerned about ATc's antimicrobial properties and its potential to obscure the stress response being tested. The IPTG-inducible system altered metabolism in all TFI strains as IPTG activates the Leloir pathway in *S. pneumoniae*, which will be shown in Chapter 3.

Overall, we were able to clone TFI strains for 89 transcription factors in TIGR4, validate their construction through qPCR and western blot analysis, and show that overexpressed transcription factors alter phenotypes in recurring patterns. We can now apply transcriptomic and binding analyses to reverse engineer a transcriptional regulatory network for TIGR4 and assay these TFI strains under antibiotic stress to uncover how the transcriptomic state of the cell influences antibiotic outcomes. Together, this will allow us to understand gene regulation of this pathogen on a genome-wide basis, creating the possibility of manipulating its phenotype in such a way to increase antibiotic efficacies.

2.5 Materials and Methods

Cloning: The cassette to clone TIGR4-LacI was amplified from pPEPY-PF6-LacI (Addgene #85589). A three-part overlap-extension PCR created the cassettes to clone each TFI strain. First, pJWV102-PL-dCas9 was modified by restriction cloning to replace the tetracycline resistance gene with a chloramphenicol resistance gene. The left homology arm was created through a two-part overlap extension amplified from the modified pJWV102-PL-dCas9 and the right homology arm was amplified directly from pJWV102-PL-dCas9. The transcription factor was amplified from the TIGR4 genome (NC_003028.3). The three parts were combined with a overlap-extension PCR using Q5 DNA Polymerase (New England Biolabs, NEB). pPEPY-PF6-lacI and pJWV102-PL-dCas9 were gifted from Jan-Willem Veening (Addgene plasmids # 85588 & # 85589).

Transformation: Starter cultures were cultured in Todd-Hewitt both + Yeast extract (THY) plus 0.013N HCl and 0.05% glycine for 2hr. Cells were diluted to OD₆₀₀ of 0.03 before NaOH, Bovine Serum Albumin (BSA) and CaCl were added. Competence Stimulating Peptide (CSP) was added and exactly 14 minutes later, 200ng of amplicon was added. Cells were cultured for 45 minutes before being pelleted and spread onto a TSA blood agar plate with selective antibiotic (Chloramphenicol 4ug/mL, Gentamicin 3ug/mL). Transformants were isolated and cultured the following day, with gDNA extraction and amplicon sequencing of its components. For each TFI strain, the entire synthetic gene, from promoter to HA-tag, and the DNA barcode were sent for Sanger sequencing through Azenta Life Sciences. The R package “DNABarcodes” was used to determine the Hamming distance between every barcode and strains were only selected if the Hamming distance was at least two bases away from all other barcodes.

qPCR: Starter cultures were cultured in SDMM at 37°C in 5% CO₂ until exponential phase before being pelleted by centrifugation, washed in PBS and re-suspended in SDMM at an OD₆₀₀ of 0.1 in biological triplicates. Experimental cultures had 1mM IPTG added and all cultures were incubated at 37°C in 5% CO₂ for one hour. Cells were pelleted by centrifugation and snap frozen. Total RNA was isolated using the RNeasy Mini Kit (Qiagen). 4μg of input RNA was treated with the TURBO DNA-free kit (Ambion) to remove genomic DNA. iScript Supermix was used to reverse transcribe cDNA and qPCR

was performed on the Bio-Rad MYIQ Real-Time PCR System using iTaq Supermix (Bio-Rad) per kit instructions. Log₂(Fold-Change) was calculated per the double delta Ct method.

Western Blot: Starter cultures were cultured in SDMM at 37°C in 5% CO₂ until exponential phase before being pelleted by centrifugation, washed in PBS and re-suspended in SDMM at an OD₆₀₀ of 0.1 in biological triplicates. Experimental cultures had 1mM IPTG added and all cultures were incubated at 37°C in 5% CO₂ for one hour. Cells were pelleted by centrifugation and snap frozen. Pellets were resuspended in reducing Laemmli Buffer and proteins were separated by SDS-PAGE (percentage gel we had) and transferred to a PVDM membrane using a semi-dry transfer for 2 hours at 20 volts. After transfer, the membrane was blocked in Blocking Buffer (3% milk in Tris-buffered Saline (TBS)) with rocking at 4°C. Membranes were then incubated with Anti-HA-HRP antibody (Invitrogen #26183-HRP, diluted 1:500 in 1% milk in TBS) for 1 hour with rocking at room temperature. Membranes were washed 6x for 5 minutes in TBS. Protein bands were detected by X-ray and the Immobilon Western HRP Substrate kit per instructions.

Growth Curves: Individual strain starters were cultured in SDMM + 1mM IPTG at 37°C in 5% CO₂ until exponential phase before being pelleted by centrifugation, washed in PBS (Phosphate-Buffered Saline) and re-suspended in PBS to an OD₆₀₀ of 0.2. Biological triplicates of cells were diluted 1:40 in 96-well plates and cultured in the BioSpa 8 plate readers (Biotek) with OD₆₀₀ readings every 30 minutes. Growth profiles were plotted with Graphpad Prism.

Semi-Defined Minimal Media (SDMM) Recipe:

Reagent	Stock (%)	1L
Acid Hydrolyzed Casein	-	5g
Enzyme Hydrolyzed Casein	-	5g
L-Cys HCl	1	4mL
L-Trp	0.57	1.05mL
L-Asn	0.5	10mL
L-Gln	0.1	10mL
Adenine	0.016	31.25mL
Ca-Pantothenate	0.1	1.2mL
Nicotinic Acid	0.03	1mL

Pyridoxine HCl	0.07	0.428mL
Thiamine HCl	0.09	0.33mL
Riboflavin	0.002	7mL
Biotin	0.0035	0.017mL
Potassium Phosphate Dibasic	-	8.5g
Sodium Acetate	-	2g
Sodium Bicarbonate	-	0.4g
Magnesium Chloride Hexahydrate	-	0.5g
Calcium Chloride	0.6	1mL
Copper Sulfate	0.1	0.5mL
Zinc Sulfate	0.3	0.16mL
Manganese Sulfate	0.07	0.286mL
Yeast Extract	-	0.5g
Uracil	0.1	20mL

pH adjusted to 7.3. Before inoculating cells, add fresh Choline Chloride (0.3%), Iron Sulfate Heptahydrate (0.04%), Catalase (30 U/mL) and Glucose (20mM).

2.6 References

1. Lewis, M. A Tale of Two Repressors. *J. Mol. Biol.* **409**, 14–27 (2011).
2. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).
3. Yang, J. H. *et al.* A White-Box Machine Learning Approach for Revealing Antibiotic Mechanisms of Action. *Cell* **177**, 1649–1661.e9 (2019).
4. Kloosterman, T. G. & Kuipers, O. P. Regulation of Arginine Acquisition and Virulence Gene Expression in the Human Pathogen *Streptococcus pneumoniae* by Transcription Regulators ArgR1 and AhrC. *J. Biol. Chem.* **286**, 44594–44605 (2011).
5. Peterson, S. N. *et al.* Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol. Microbiol.* **51**, 1051–1070 (2004).
6. Aprianto, R., Slager, J., Holsappel, S. & Veening, J.-W. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. *Nucleic Acids Res.* (2018) doi:10.1093/nar/gky750.
7. Lee, J.-H. *et al.* The WblC/WhiB7 Transcription Factor Controls Intrinsic Resistance to Translation-Targeting Antibiotics by Altering Ribosome Composition. *MBio* **11**, (2020).
8. Margolin, A. A. *et al.* ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* **7**, S7 (2006).
9. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS One* **5**, e12776 (2010).
10. Iglesias-Martinez, L. F., De Kegel, B. & Kolch, W. KBoost: a new method to infer gene regulatory networks from gene expression data. *Sci. Rep.* **11**, 15461 (2021).
11. Haury, A.-C., Mordelet, F., Vera-Licona, P. & Vert, J.-P. TIGRESS: Trustful Inference of Gene REgulation using Stability Selection. *BMC Syst. Biol.* **6**, 145 (2012).
12. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 (2012).
13. Hendriksen, W. T. *et al.* CodY of *Streptococcus pneumoniae* : Link between Nutritional Gene Regulation and Colonization. *J. Bacteriol.* **190**, 590–601 (2008).
14. Afzal, M., Shafeeq, S. & Kuipers, O. P. LacR Is a Repressor of lacABCD and LacT Is an Activator of lacTFEG , Constituting the lac Gene Cluster in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* **80**, 5349–5358 (2014).
15. McCluskey, J., Hinds, J., Husain, S., Witney, A. & Mitchell, T. J. A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in *Streptococcus pneumoniae*. *Mol. Microbiol.* **51**, 1661–1675 (2004).
16. Sebert, M. E., Palmer, L. M., Rosenberg, M. & Weiser, J. N. Microarray-Based Identification of htrA , a *Streptococcus pneumoniae* Gene That Is Regulated by the CiaRH Two-Component System and Contributes to Nasopharyngeal Colonization. *Infect. Immun.* **70**, 4059–4067 (2002).
17. Ulijasz, A. T., Andes, D. R., Glasner, J. D. & Weisblum, B. Regulation of Iron Transport in *Streptococcus pneumoniae* by RitR, an Orphan Response Regulator. *J. Bacteriol.* **186**, 8123–8136 (2004).
18. Potter, A. J., Kidd, S. P., McEwan, A. G. & Paton, J. C. The MerR/NmlR Family Transcription Factor of *Streptococcus pneumoniae* Responds to Carbonyl Stress and Modulates Hydrogen Peroxide Production. *J. Bacteriol.* **192**, 4063–4066 (2010).

19. Caymaris, S. *et al.* The global nutritional regulator CodY is an essential protein in the human pathogen *Streptococcus pneumoniae*. *Mol. Microbiol.* **78**, 344–360 (2010).
20. Nakano, S., Küster-Schöck, E., Grossman, A. D. & Zuber, P. Spx-dependent global transcriptional control is induced by thiol-specific oxidative stress in *Bacillus subtilis*. *Proc. Natl. Acad. Sci.* **100**, 13603–13608 (2003).
21. Sun, G. *et al.* Mechanism of *Escherichia coli* Lethality Caused by Overexpression of flhDC, the Flagellar Master Regulator Genes, as Revealed by Transcriptome Analysis. *Int. J. Mol. Sci.* **24**, 14058 (2023).
22. Turkarslan, S. *et al.* A comprehensive map of genome-wide gene regulation in *Mycobacterium tuberculosis* OPEN SUBJECT CATEGORIES » Bacterial genes » High-throughput screening » Regulatory networks » Transcriptomics. (2015) doi:10.1038/sdata.2015.10.
23. Minch, K. J. *et al.* The DNA-binding network of *Mycobacterium tuberculosis*. *Nat. Commun.* **6**, (2015).
24. Rustad, T. R. *et al.* *Mapping and manipulating the Mycobacterium tuberculosis transcriptome using a transcription factor overexpression-derived regulatory network*. <http://genomebiology.com/2014/15/11/502> (2014) doi:10.1186/s13059-014-0502-3.
25. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13**, (2017).
26. Ortet, P., De Luca, G., Whitworth, D. E. & Barakat, M. P2TF: a comprehensive resource for analysis of prokaryotic transcription factors. *BMC Genomics* **13**, 628 (2012).
27. Oliveira Monteiro, L. M. *et al.* PredicTF: prediction of bacterial transcription factors in complex microbial communities using deep learning. *Environ. Microbiome* **17**, 7 (2022).
28. Patiyal, S. *et al.* A hybrid approach for predicting transcription factors. *Front. Bioinforma.* **4**, (2024).
29. Rosconi, F. *et al.* A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat. Microbiol.* **7**, 1580–1592 (2022).
30. Hu, Y. *et al.* The XRE Family Transcriptional Regulator SrtR in *Streptococcus suis* Is Involved in Oxidant Tolerance and Virulence. *Front. Cell. Infect. Microbiol.* **8**, (2019).
31. Si, M. *et al.* MsrR is a thiol-based oxidation-sensing regulator of the XRE family that modulates *C. glutamicum* oxidative stress resistance. *Microb. Cell Fact.* **19**, 189 (2020).
32. Sorg, R. A., Gallay, C., van Maele, L., Sirard, J. C. & Veening, J. W. Synthetic gene-regulatory networks in the opportunistic human pathogen *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 27608–27619 (2020).
33. Mellroth, P. *et al.* LytA, Major Autolysin of *Streptococcus pneumoniae*, Requires Access to Nascent Peptidoglycan. *J. Biol. Chem.* **287**, 11018–11029 (2012).
34. Cortes, P. R., Piñas, G. E., Cian, M. B., Yandar, N. & Echenique, J. Stress-triggered signaling affecting survival or suicide of *Streptococcus pneumoniae*. *Int. J. Med. Microbiol.* **305**, 157–169 (2015).
35. Martner, A., Dahlgren, C., Paton, J. C. & Wold, A. E. Pneumolysin Released during *Streptococcus pneumoniae* Autolysis Is a Potent Activator of Intracellular Oxygen Radical Production in Neutrophils. *Infect. Immun.* **76**, 4079–4087 (2008).
36. Wu, K. *et al.* CpsR, a GntR family regulator, transcriptionally regulates capsular polysaccharide biosynthesis and governs bacterial virulence in *Streptococcus pneumoniae*. *Sci. Rep.* **6**, 29255 (2016).

37. Iyer, R. & Camilli, A. Sucrose metabolism contributes to in vivo fitness of *Streptococcus pneumoniae*. *Mol. Microbiol.* **66**, 1–13 (2007).
38. Baker, J. L. *et al.* Characterization of the Trehalose Utilization Operon in *Streptococcus mutans* Reveals that the TreR Transcriptional Regulator Is Involved in Stress Response Pathways and Toxin Production. *J. Bacteriol.* **200**, (2018).

Chapter 3:

Reverse Engineering of the Transcriptional Regulatory Network of TIGR4 Aided by Transcription Factor Overexpression

3.0 Introduction

All living cells modulate cellular processes in response to stimuli to maximize fitness. A classic example of this is carbon catabolite repression (CCR), in which microorganisms repress genes required for catabolizing secondary carbon sources when the preferred carbon source is present (Fig. 1A). In *S. pneumoniae* and other Firmicutes, CCR is regulated by the transcription factor Control Catabolite Protein A (CcpA) and its co-repressor, Histidine Protein (HPr). HPr serves as a sensor for the metabolic state of the cell. Equilibrium between HPr and phosphorylated-HPr (HPr-Ser~P) is dependent on the concentration of ATP and the glycolytic-metabolite fructose 1,6-biphosphate (FBP) in the cell, where high concentrations of ATP favor the phosphorylation of HPr. HPr-Ser~P and FBP bind CcpA, causing a conformational change in CcpA favoring binding to DNA its recognition sequence, known as *cre* sites. At high concentrations of ATP/FBP, CcpA-HPr-Ser~P binds *cre* sites in the promoters of secondary catabolism genes, repressing their expression. At low concentrations of ATP/FBP, HPr is not phosphorylated and thus *cre* sites remain unoccupied, allowing expression of genes required for catabolizing secondary carbon sources. However, expression of secondary catabolic pathways is also dependent on the availability of that specific carbon source, ensuring that expression only occurs when it's beneficial to the cell¹⁻⁴. CCR exemplifies how multiple signals are processed by transcription factors in the cell to produce an efficient response, allowing bacteria to spend their energy catabolizing the most efficient carbon source currently available.

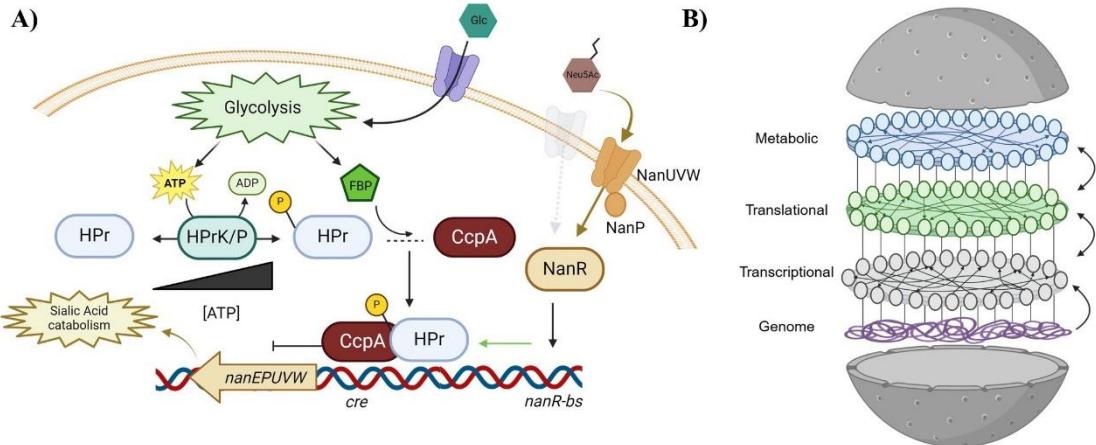


Figure 1. Transcriptional Regulation helps Define Cell State. **A)** When glucose is present, active glycolysis leads to the phosphorylation of HPr, which, aided by FBP, interact with CcpA to promote DNA binding at *cre* sites. This represses expression of sialic acid catabolic genes, *nanEPUVW*. When glucose is absent, HPr is dephosphorylated leading to low levels of expression of *nanEPUVW*. If sialic acid is present, it is transported into the cell by the transporter NanUVW and interacts with the transcription factor NanR to activate a positive feed forward loop to begin sialic acid catabolism. Glucose = Glc. Sialic Acid (N-Acetylneurameric acid) = Neu5Ac. Fructose-1,6-biphosphate = FBP. **B)** Cells can be defined as an interconnected array of networks derived from its genome to regulate and enact cell behavior.

Understanding the regulatory logic of a cell allows for control of its behavior and the potential for industrial and medical applications. Perhaps the greatest feat of genetic engineering came from Genentech's synthetic production of human insulin using, by today's standard, a simple overexpression system in *Escherichia coli*⁵. Today, sustainable production of biofuels has been significantly enhanced through a combination of repressor knockout and synthetic overexpression of a transcriptional activator in *Saccharomyces cerevisiae*⁶. More complex synthetic gene circuits are being engineered from an ever-increasing list of parts including transcription factors, their known binding sites and promoters of varying strengths^{7–10}. Fine tuning and complete control of these genetic circuits is critical to future success of novel medical technologies such as CRISPR/Cas9

editing and gene therapy. An understanding of the regulatory logic behind virulence can also identify drug targets, with specific transcriptional regulators in *Staphylococcus aureus* and cancer cells having been targeted in antibiotic¹¹ and cancer treatment¹².

As a single-celled organism, all of *S. pneumoniae*'s behavioral decisions must be encoded in its genome (Fig. 1B). If we can understand this regulatory logic on a genome-wide scale, we should also be able to develop strategies to potentiate antibiotics by synthetically predisposing bacterial cells to a state that is most susceptible to antibiotic killing. It's known that the metabolic state of bacteria influences the efficacy of certain antibiotics, identifying metabolic stimulation with glucose as an antibiotic potentiator^{13,14}. It's also been shown that combination treatment with antibiotics that are Strongly Dependent on Metabolism (SDM) and Weakly Dependent on Metabolism (WDM) improves lethality¹⁵. Further, many antibiotic-resistant strains become resistant through regulatory mutations, including through the overexpression of multidrug efflux pumps^{16,17}. Identifying mutations in the regulators or non-coding regulatory motifs that influence antibiotic susceptibility is impossible without a well-characterized transcriptional regulatory network. For these reasons, we aim to characterize the transcriptional regulatory network of *S. pneumoniae* strain TIGR4.

3.1 Transcription Factor Overexpression Identifies Broad Transcriptomic Changes

To define the transcriptional regulatory network of *S. pneumoniae*, we first aimed to identify the regulatory signature of every transcription factor. By overexpressing each transcription factor through our Transcription Factor Induction (TFI) strains, we can

identify its target genes through their resulting differential expression. To this end, RNA sequencing (RNAseq) was performed on each TFI strain induced with isopropyl β -D-1-thiogalactopyranoside (IPTG). Every TFI strain was sequenced in biological triplicates and each sample has an average of 3.9 million aligned coding sequence reads per sample. To identify differential expression caused only by transcription factor overexpression, all strains were cultured in Semi-Defined Minimal Media (SDMM) to exponential phase. Exponential growth of bacterial cultures is approximately a steady-state system, so profiling the transcriptome of each induced-TFI strain in this phase minimized noise. Importantly, cells were precultured with IPTG before being reinoculated into fresh media to ensure transcription factor overexpression before exponential growth. This helped maintain a clean, steady-state system because the expression of the induced transcription factor will no longer be increasing. Genes with an adjusted p-value of less than 0.05 were considered significant (Padj) and genes with both an adjusted p-value of less than 0.05 and a log₂Fold-Change of $\geq |1|$ were considered differentially expressed genes (DEG). RNA sequencing performed on all 89 TFI strains identified 23 DEGs and 145 Padjs on average. Across all experiments, 491 (23% of total) and 1,671 (78% of total) genes were found to be DEGs and Padjs at least once, respectively.

3.1.1 Constructs of our Synthetic System.

Upon analysis of our TFI RNA sequencing experiments, it became clear that our synthetic TFI system creates two inherent constructs: varying levels of transcription factor overexpression and IPTG-stimulated genes. Despite using the same synthetic system, different transcription factors were overexpressed to different levels (Fig. 2A). As our TFI system maintains the native copy of each transcription factor, the magnitude of

overexpression of each transcription factor varied inversely with its average natural expression (Fig. 2B). Transcription factors with high natural expression, such as *fabT* (*SP_0416*), are only slightly overexpressed after induction with IPTG. Conversely, lowly-expressed transcription factors such as *nanR2* (*SP_1331*) have much greater fold-changes. There was no correlation between the level of transcription factor overexpression and the number of resulting differentially expressed genes, possibly because transcriptional regulation relies upon many cellular signals and not solely on the regulator's expression level (Fig. 2C).

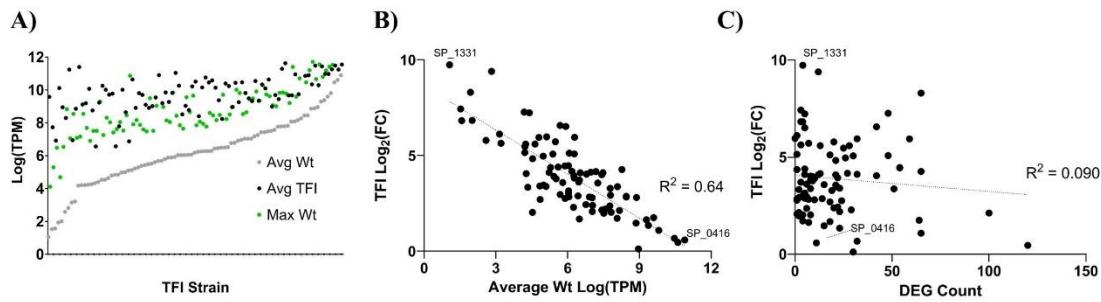


Figure 2. Summary of Transcription Factor Overexpression. **A)** Comparison of the expression levels of the average wild-type (Wt), maximum wild-type and average TFI expression of each transcription factor, ordered by increasing average Wt expression. **B)** The differential expression of each transcription factor in its TFI strain varied inversely to its average Wt expression due to the expression of the native copy of the transcription factor. **C)** There was no relationship between the degree of overexpression of a transcription factor to the number of differentially expressed genes (DEG) identified.

3.1.2 IPTG Alters the Transcriptome of TIGR4

As IPTG is a molecular mimic of a signaling molecule known to influence gene expression in *E. coli*¹⁸, one can expect that IPTG would influence gene expression in other bacterial species as well. To characterize the transcriptomic response of TIGR4 to IPTG, we identified recurring differential expression across all TFI RNAseq experiments, as well as wild-type TIGR4 treated with IPTG. We reasoned that IPTG-stimulated genes would be

repeatedly identified as a differentially expressed gene and differ in the same direction (up- or down-regulated). We identified 23 and 5 genes that were determined to Padj and DEG, respectively, in at least 50% of all TFI strains. Of these genes, only four were consistently regulated in the same direction, the galactose metabolic operon (*SP_1851-1854*) (Fig. 3). GalR (*SP_1854*) and the transgenic LacI used in our synthetic TFI system share 44% positive identities in their primary sequences, so we suspect that GalR is directly responding to IPTG akin to LacI in *E. coli*. Indeed, all four genes in the operon were repeatedly activated in the presence of IPTG, suggesting a conserved allosteric derepression of the operon signaled by IPTG. Interestingly, as SDMM contains glucose, if the galactose metabolic operon in TIGR4 was controlled by CCR, one would expect that the presence of IPTG would not lead to derepression because CcpA-HPr-Ser~P would be repressing expression. There is no identifiable *cre* site within the promoter of the galactose metabolic operon, indicating that it is not regulated by CCR in TIGR4 but only by repression by GalR.

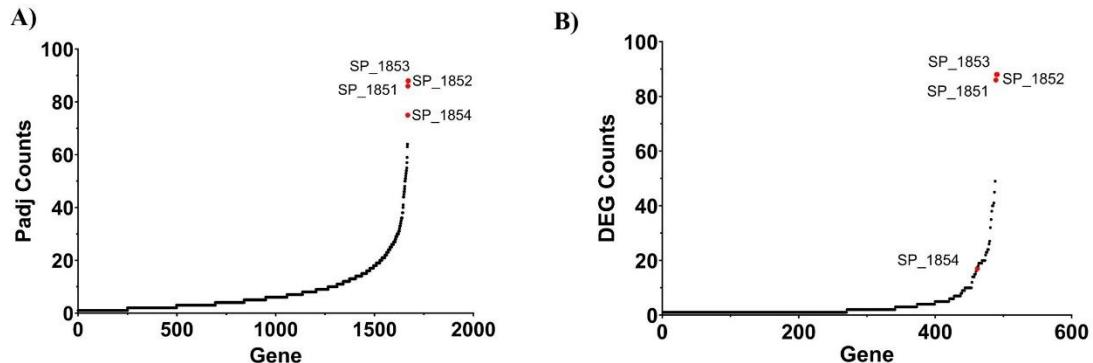


Figure 3. IPTG Stimulated Genes. Visualization of the galactose metabolic operon (*SP_1851-1854*) as being respondent to IPTG through frequent identification as a **A)** Padj and **B)** DEG in TFI RNAseq Experiments.

3.1.3 Differential Expression Analysis Fails to Identify All Known Interactions

RNAseq analysis aims to identify differentially expressed genes (DEG), which are defined as genes with a $\log_2\text{FC}$ of $\geq |1|$ with an adjusted p-value (Padj) of < 0.05 . Therefore, to be considered a DEG, a target gene's expression must either be doubled or halved by transcription factor overexpression. Unfortunately, many previously known regulatory interactions were not classified as DEGs because they did not reach the $\log_2\text{FC}$ threshold (Fig. 4). This is best exemplified by the TFI RNAseq results of YhcF (*SP_1714*), a transcription factor that has been previously predicted by comparative genomics to regulate three operons, *SP_1714-1715*, *SP_1381-1380* and *SP_0785-786*, all of which encode ABC transporters¹⁹. Of the three regulatory interactions, TFI RNAseq was only able to identify self-repression of *SP_1715* as a DEG. The other two operons were identified as being significant according to their Padj values, yet all failed to reach the $\log_2\text{FC}$ cutoff (Fig. 4A). Interestingly, overexpression of *SP_1714* was not identified as either a DEG or a Padj because overexpression of the synthetic copy of the gene strongly repressed expression of the native copy, leading to no change in overall expression. Overexpression of other transcription factors was able to identify some of their known target genes as DEGs, as in the case of PtvR²⁰ (*SP_0100*, Fig. 4B), while overexpression of other transcription factors such as TreR (*SP_1885*) fully captured its known target genes as DEGs¹⁹ (Fig. 4C).

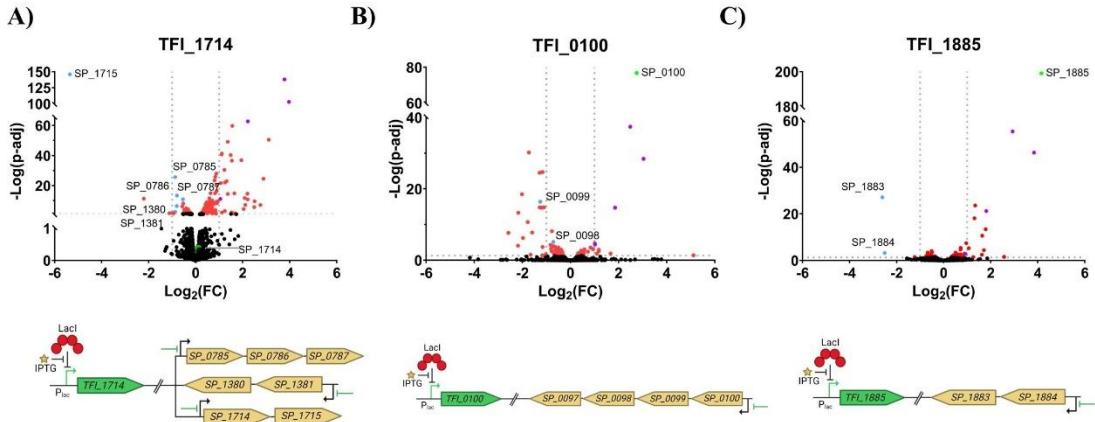


Figure 4. TFI RNAseq Fails to Identify All Known Target Genes as DEGs. **Top)** Volcano plots of the results of TFI RNAseq from **A)** TFI_1714, **B)** TFI_0100 and **C)** SP_1885. Overexpressed transcription factors are shown in green, known target genes are shown in blue, Padj are shown in red and IPTG-stimulated genes are shown in purple. Vertical dotted lines identify the log₂(FC) cutoff and the horizontal dotted line identifies the significance threshold. **Bottom)** TFI regulatory diagrams for known target genes of **A)** SP_1714, **B)** SP_0100 and **C)** SP_1885.

There are several possible explanations as to why known regulatory interactions were not identified as DEGs in our TFI RNAseq experiments. Many transcription factors rely on cellular signals (i.e., availability of a specific effector molecule) that without, transcription factor overexpression can only slightly influence transcription of a target gene. Also, many genes are regulated by two or more transcription factors that work together to fine-tune their expression^{21,22}. Overexpression of just one of the transcription factors that regulates expression of a target gene may only slightly alter expression of the target gene. Lastly, all RNAseq experiments suffer from mean-dependent variance, where lowly expressed genes have a larger variance than highly expressed genes^{23,24}. Therefore, transcription factors that regulate lowly expressed genes are more likely to produce DEGs through overexpression than transcription factors that regulate highly expressed genes. While identifying DEGs to call regulatory interactions failed for several known interactions, there

were still many known interactions that were identified as significant DEGs. For this reason, when looking at differential expression analysis of TFI RNAseq, we focused more on genes with significant Padj values than classical DEGs. Still, it's clear that TFI RNAseq alone cannot identify all regulatory interactions to create a comprehensive transcriptional regulatory network.

3.2 An Ensemble of Network Inference Methods Predicts Regulatory Interactions based on Covarying Expression

Reverse engineering of a transcriptional regulatory network from transcriptomic data is a longstanding computational problem that has been attempted to be answered by many research groups with many different approaches. A recent DREAM (Dialogue on Reverse Engineering Assessment and Methods) challenge was put together to try and identify the best performing network inference method. 35 network inference methods, classified into six categories based on their statistical approach, were tested for accuracy in mapping the transcriptional regulatory networks of *E. coli*, *S. aureus*, *S. cerevisiae* and an in-silico model²⁵. Researchers found that no one method repeatedly outperformed the others and that the best approach is to use the “wisdom of the crowds,” that is, to use many inference methods and combine the results into a “community” network. This community network almost always outperformed each individual method, and the diversification of methods blends and minimizes biases. Therefore, we also have taken to the “wisdom of the crowds” to complement TFI RNAseq in our effort to map the transcriptional regulatory network of TIGR4.

3.2.1 Compiling Gene Expression Data for Network Inference

High quality, diverse transcriptomic data is required to accurately reverse engineer a transcriptional regulatory network. Our TFI gene expression dataset attempts to accomplish this by overexpressing each individual transcription factor, the aggregation of which shows each regulon stimulated above a consistent background. However, as previously mentioned, many transcription factors respond to varying levels of effector molecules, which may or may not be present during exponential growth in SDMM. AdcR (*SP_2172*) has been shown in *S. pneumoniae* strain D39 to repress the expression of the zinc importer, *adcABC* (*SP_2169-2171*), in response to intracellular zinc concentrations²⁶. SDMM contains both zinc sulfate and yeast extract, thus during exponential growth in SDMM one would expect that AdcR would be repressing expression of *adcABC*. In that case, synthetically overexpressing *adcR* would not result in differential expression of its known target genes because they are already repressed. As expected, TFI_2172 RNAseq failed to identify *adcABC* as a Padj.

To compare transcriptomes of different TFI strains, they must be identified in the same conditions, but that inevitably leads to missed transcriptomic states due to a single culture condition. Therefore, to create a more diverse dataset we included previous gene expression data collected from our lab on antibiotic-treated TIGR4. Ultimately, we want to leverage our knowledge of the transcriptional regulatory network of *S. pneumoniae* to improve antibiotic treatments, so this had the added benefit of analyzing the transcriptome under antibiotic stress. Our starting dataset includes expression data of 1,892 genes across 604 total RNAseq samples, including 271 TFI samples, 76 protein synthesis inhibitor (PSI) samples, 18 RNA synthesis inhibitor (RSI) samples, 120 cell wall synthesis inhibitor

(CWSI), 36 DNA synthesis inhibitor (DSI) samples, 71 control samples and 12 *in vivo* samples²⁷. Gene expression data were normalized to transcripts per million (TPM) to allow inter-experiment comparison. Initial clustering analysis of all TFI transcriptomes showed clear batch effects between TFI RNAseq experiments performed in two separate lab spaces, requiring the use of ComBat-Seq²⁴ to correct. To optimize network reconstruction, we then cleaned the data by removing genes with either an average or median log(TPM) of less than one and a Tn-Seq fitness value of greater than 0.8 in SDMM²⁸ (Fig. 5A). The inherent mean-dependent variance in gene expression results in more variance supplied to the dataset from these lowly expressed genes, which also tend to be phenotypically less important as essential genes are shown to be expressed to higher levels than nonessential genes (Fig. 5B). In total, 44 genes (2.33% of all genes) were removed for this reason.

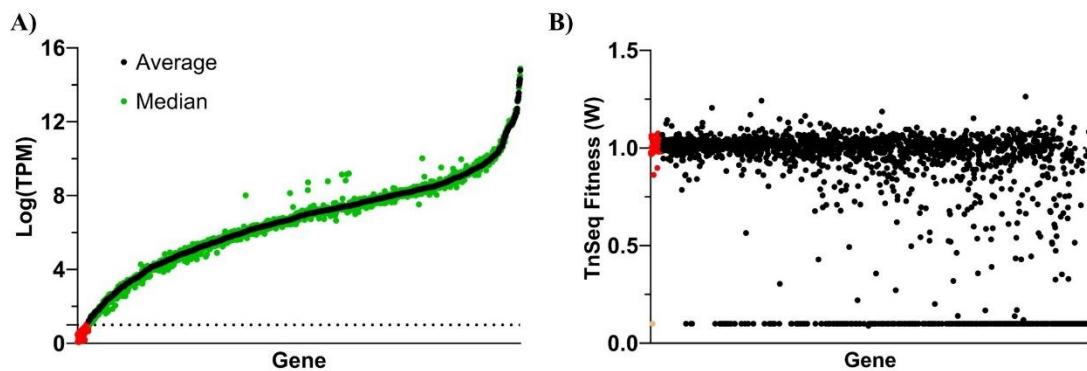


Figure 5. Removal of Lowly-Expressed Genes from the Dataset. **A)** Average and Median expression of each gene in TIGR4 according to its log(TPM) values. **B)** Fitness values of each gene in TIGR4 according to Tn-Seq experiments in SDMM. Genes removed for having either an average or median expression value of less than one and a fitness value of greater than 0.8 are shown in red. Genes spared from removal for phenotypic importance shown in tan. Genes are ordered across the x-axis by increasing average expression in both plots.

There are intuitive strengths and weaknesses to using larger, more diverse datasets compared to using smaller, cleaner datasets when reverse engineering transcriptional

regulatory networks from gene expression data. With larger and more diverse datasets, you capture the transcriptome across a larger space, but also input more noise. Inversely, smaller, cleaner datasets result in less noise but have a less comprehensive view of the possible transcriptomic states. Without knowing which datasets would perform best, we created three different gene expression datasets to identify which combination of gene expression data result in the best reconstructed network (Fig. 6). All datasets contain the same 1,848 genes, and the first dataset, referred to as Dataset A, contains all 604 samples mentioned above. From there, we subsetted Dataset A into Dataset B by removing outliers identified through Uniform Manifold Approximation and Projection (UMAP). UMAP is a non-linear dimension reduction technique used to visualize large datasets on a two-dimensional plane, which we prefer to the similar technique t-SNE as it emphasizes separation of dissimilar clusters. By visualizing all gene expression datasets on a two-dimensional plane, we can identify and remove outlier samples, thus creating a cleaner dataset. Through this, 167 samples were removed from Dataset A to create Dataset B. Finally, Dataset C was created by removing non-centroid biological triplicates from Dataset B to remove within-sample noise. To identify the centroid of a set of replicates, we calculated the pairwise Pearson's correlation coefficient between each sample. Within sets of replicates, we kept the sample that had the greatest sum of correlation coefficients. Within replicate correlation coefficients were often above 0.98, so which replicate was chosen likely has little influence on the resulting Dataset C.

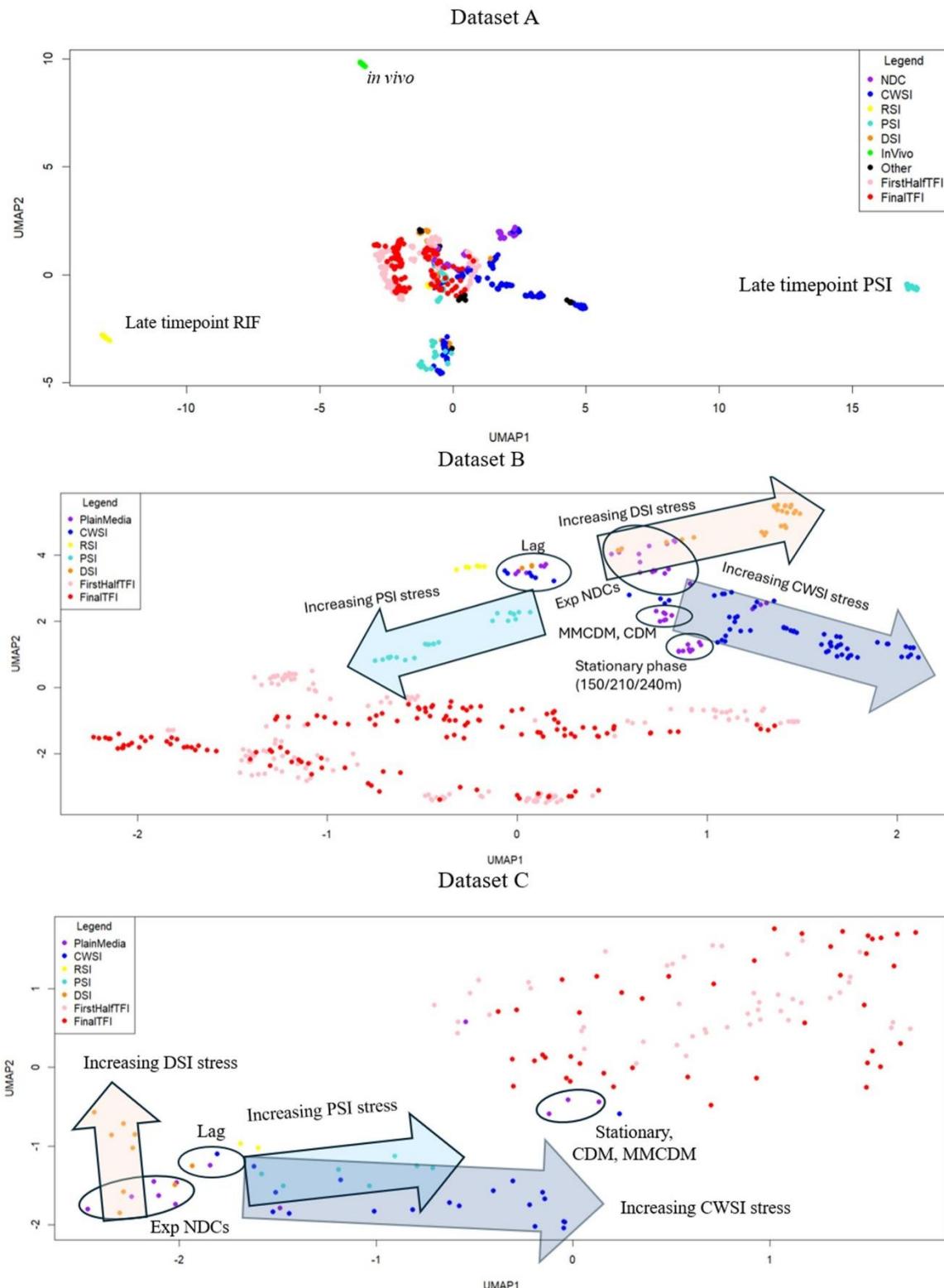


Figure 6. UMAP Representation of our three Datasets. The two sets of ComBat-Seq-adjusted TFI RNAseq batches are shown in pink and red. Cell Wall Synthesis Inhibitors = CWSI. RNA Synthesis Inhibitors = RSI. Protein Synthesis Inhibitors = PSI. DNA Synthesis Inhibitors = DSI.

3.2.2 Identifying Optimal Combinations of Network Inference Methods

We then input each dataset independently into eight published network inference methods (Table 1)^{29–35}. The methods we chose have diverse mathematical approaches including mutual information theory, regression, machine learning (including both random forests and gradient boosting machines) and partial correlations. Predicted interactions identified in the DREAM challenge were graded against the known regulatory networks of *E. coli*, *S. cerevisiae* and the *in silico* model network. However, like *S. aureus*, *S. pneumoniae* does not have a previously mapped transcriptional regulatory network to compare to, hence this study. Therefore, performance of each network inference method was graded by calculating the Area Under Precision Recall (AUPR) relative to predicted interactions in TIGR4 according to the RegPrecise database¹⁹. RegPrecise is a manually curated database of predicted regulatory interactions for bacteria, but instead of predicting interactions based on gene expression data, it relies on comparative genomics. Many transcription factors have previously been experimentally characterized, typically in model organisms such as *E. coli* or *B. subtilis*. RegPrecise leverages this information with multiple genomic sequences within a taxonomic group to predict regulatory interactions in other organisms. While it is not as accurate as the gold standard networks available for *E. coli* or *S. cerevisiae*, it allows for relative comparisons of our network inference methods. RegPrecise is not intended to be a comprehensive regulatory network, so we expect many true interactions to be missing from this dataset and thus low precision values.

Method	Description	Predicted Interactions	Precision	Recall	AUPR
ARACNE	Mutual Information theory with Relevance Network Filtering, false positives minimized by Data Processing Inequality	212	0.293	0.053	2.8%
BC3NET	Bootstrapping Mutual Information theory ensemble with significance filtering	342	0.155	0.049	3.3%
CLR	Mutual Information theory with normalization and significance filtering	5,000	0.039	0.170	4.8%
GENIE3	Random Forest regressor for each gene in the network	5,099	0.038	0.182	4.3%
GRNBOOST2	Gradient Boosting Machine regressor for each gene in the network	8,174	0.030	0.235	4.4%
KBOOST	Kernal PCA predictions for each gene with Bayesian Model Averaging	9,555	0.018	0.167	1.4%
MRNETB	Backwards elimination of features followed by Mutual Information theory with filtering to minimize redundancy	5,000	0.033	0.141	2.8%
PCIT	Partial Correlations with Mutual Information, false positives minimized by Data Processing Inequality	10,000	0.021	0.199	2.6%

Table 1. Network Inference Method Overview. Description and performance of each individual network inference method used. Performance shown from each methods results on Dataset A. AUPR = Area Under Precision Recall.

For seven of the eight inference methods, Dataset A outperformed the rest according to its AUPR (Fig. 7A), suggesting that larger, more diverse datasets improve network reconstruction. The only inference method where Dataset A did not perform the best was KBOOST, which was the lowest scoring inference method tested. This trend continues down to Dataset B as, again, for seven of eight methods, Dataset B outperformed Dataset C (Fig. 7A).

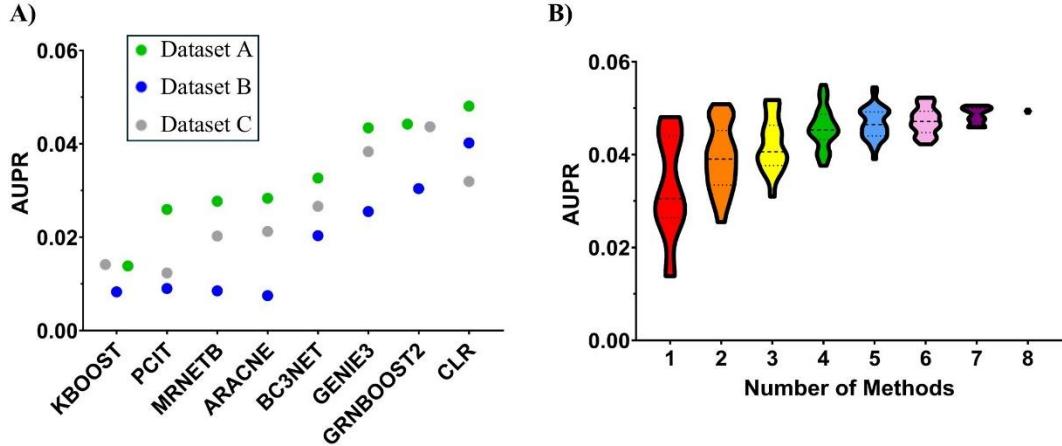


Figure 7. Performance of Individual and Combinations of Inference Methods. **A)** Area Under Precision Recall (AUPR) of all eight network inference methods across the three datasets: Dataset A (green), Dataset B (blue) and Dataset C (grey). **B)** Violin plot showing the spread of AUPRs for combinations of different numbers of network inference methods.

The power of using an ensemble of inference methods comes from combining the inferences made by each method into a community network. As different methods rank predicted interactions on different statistical measures (i.e., p-values and correlation coefficients), we used the average rank method to combine each method's inferences. The average rank method simply ranks each predicted interaction made by a method, then takes the average rank of an interaction across all methods. Interactions that are not called by an inference method are punished by giving them a ranking equal to the maximum ranking (equal to the maximum number of interactions predicted by a method) plus one. This ensures that interactions that are repeatedly predicted have lower average ranks than interactions that are only called in a single inference method.

To identify the optimal combination of inference methods, we randomly removed one through seven of the network inference methods and recalculated the average rank of each interaction and the ensemble's AUPR, and repeated this 1,000 times (Fig. 7B). As expected,

using more than one inference method increased overall AUPR, but this improvement plateaued at four inference methods. In fact, the optimal network came from just four of the eight network inference methods: GENIE3, ARACNE, CLR and GRNBOOST2, which were four of the five highest scoring individual methods. This network, called our “Ensemble” network, contains 11,228 predicted regulatory interactions with 28.4% recall, 2.64% precision and an AUPR of 5.5%. While that is low in terms of classifiers, it is within range of inference methods used on the *E. coli* TRN²⁵ and is not only evaluating our network construction but also the feasibility of mapping a TRN from gene expression data alone.

3.3 Identification of Transcription Factor Motifs Differentiates Direct and Indirect Regulation

Perhaps the most difficult aspect of predicting regulatory interactions from gene expression data is differentiating direct and indirect regulation. Direct regulation occurs when a transcription factor binds the promoter of a gene or operon and directly affects its expression (Fig. 8A), while indirect regulation occurs as a downstream effect of direct regulation. Indirect regulation can occur via a regulatory cascade, for example where one transcription factor regulates expression of another transcription factor, which then represses expression of its target genes (Fig. 8A). Alternatively, indirect regulation can occur due to the cellular changes that occur due to direct regulation. For example, overexpression of an activator of a zinc efflux pump would reduce the intracellular concentration of zinc in the cell, and several other regulons may respond to this change (Fig. 8B).

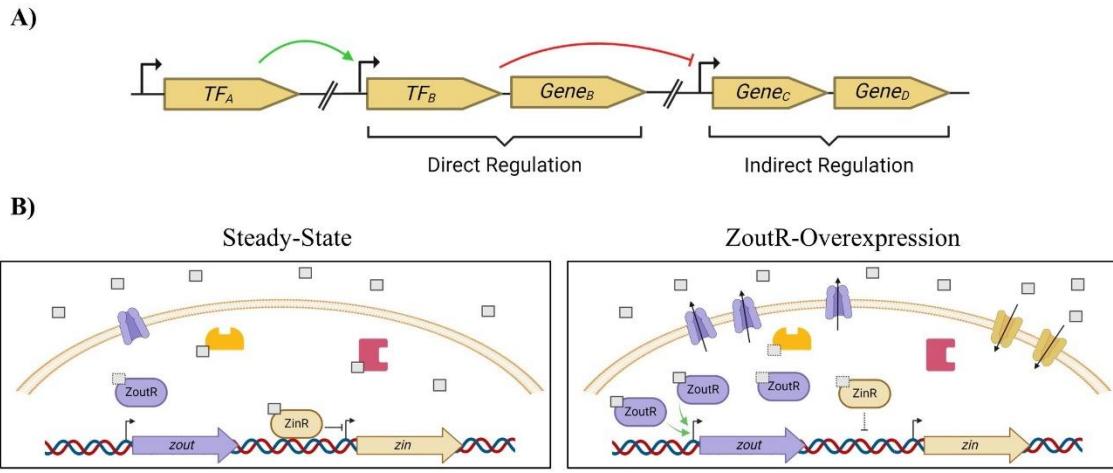


Figure 8. Direct vs Indirect Regulation. **A)** Regulatory diagram demonstrating direct regulation vs. indirect regulation via a regulatory cascade. **B)** Hypothetical scenario where overexpression of a zinc-efflux activator, ZoutR, indirectly results in activation of a zinc importer, Zin, by changing the intracellular concentration of zinc (grey boxes). Zinc exporter, Zout, shown in purple; zinc importer, Zin, shown in tan.

Both TFI RNAseq and network inference approaches struggle differentiating direct and indirect regulation. For TFI RNAseq, how can we identify which genes with significant Padj values are directly regulated by the overexpressed transcription factor, and which are a downstream result? Similarly, when using the network inference methods, how can we identify direct regulation when several regulons have overlapping functions in the cell and strong expression correlation? For direct regulation to occur, transcription factors must bind DNA in the promoter region of its target genes. On that basis, we decided to require a binding motif within a promoter region to call a high-confidence interaction. By doing so we will select only direct regulatory interactions and minimize false positives in our network.

3.3.1 ChIPseq Identifies the Binding Motifs of Select Transcription Factors.

A classical method to identify the binding motif of a transcription factor is chromatin immunoprecipitation and sequencing (ChIPseq). Briefly, cells are chemically crosslinked

through formaldehyde which covalently binds transcription factors to their DNA-binding site. Cells are then sonicated for the dual purpose of both lysing the cells and shearing the chromosome into ~500 base pair fragments. DNA fragments that are bound to HA-tagged (TFI) transcription factors are purified out using anti-HA antibodies and extensively washed. Then, the transcription factor is degraded with heat and proteinases and the remaining DNA is purified and sequenced. In our analytical pipeline, sequencing reads are first trimmed with Cutadapt³⁶ and then aligned to the TIGR4 genome (NC_003028.3) with BWA Aln³⁷. Significant peaks were called using HOMER³⁸ with a required 2.5-fold enrichment relative to control and no local filtering. Two types of controls were tested: Input and IgG controls. Input controls include no immunoprecipitation and are simply the sequencing of sonicated DNA. IgG controls mimic the entire ChIPseq protocol but, instead of an anti-HA antibody, IgG controls use an IgG control antibody that won't recognize the HA-tagged transcription factors and represent the DNA sequenced for non-specific binding that occurs throughout the immunoprecipitation process. Visual analysis of both sets of controls compared to test samples show that only IgG controls identify recurrent peaks likely due to non-specific binding of DNA. For this reason, IgG controls were used for our analysis (Fig. 9A). Transcription factor binding motifs were predicted with MEME-ChIP, a motif discovery algorithm designed specifically for ChIPseq analysis with expected motifs to fall centrally on a ~500 base pair sequence^{39,40}.

ChIPseq was performed on six TFI strains: TFI_1999, TFI_0058, TFI_1131, TFI_2062, TFI_0376 and TFI_1674. These strains were selected because they have varying expression levels, and several have known binding motifs to compare our results to. All experiments were performed in biological duplicate, and each sample had over two million aligned

paired end reads. Unfortunately, only two of the six strains provided enough significant peaks to identify binding motifs. These two strains were TFI_1999 and TFI_0058, which overexpress CcpA and CpsR, respectively, and both are global metabolic regulators that are known to bind DNA when glucose is present^{1,41}. Biological duplicates of these TFI strains had significant peaks called mostly within 30 bases of each other (Fig. 9B) and the normalized tag count between matching peaks was extremely correlated (Fig. 9C). Histograms of the distance between called peaks and transcriptional start sites (TSS) and start codons indicate that binding events occur primarily near promoter regions (Fig. 9D). Importantly, the motifs identified by ChIPseq align to the known motifs of both CcpA and CpsR (Fig. 9E). Peaks called far away from both TSSs and start codons may represent DNA sequences with high similarity to each transcription factor's binding motif.

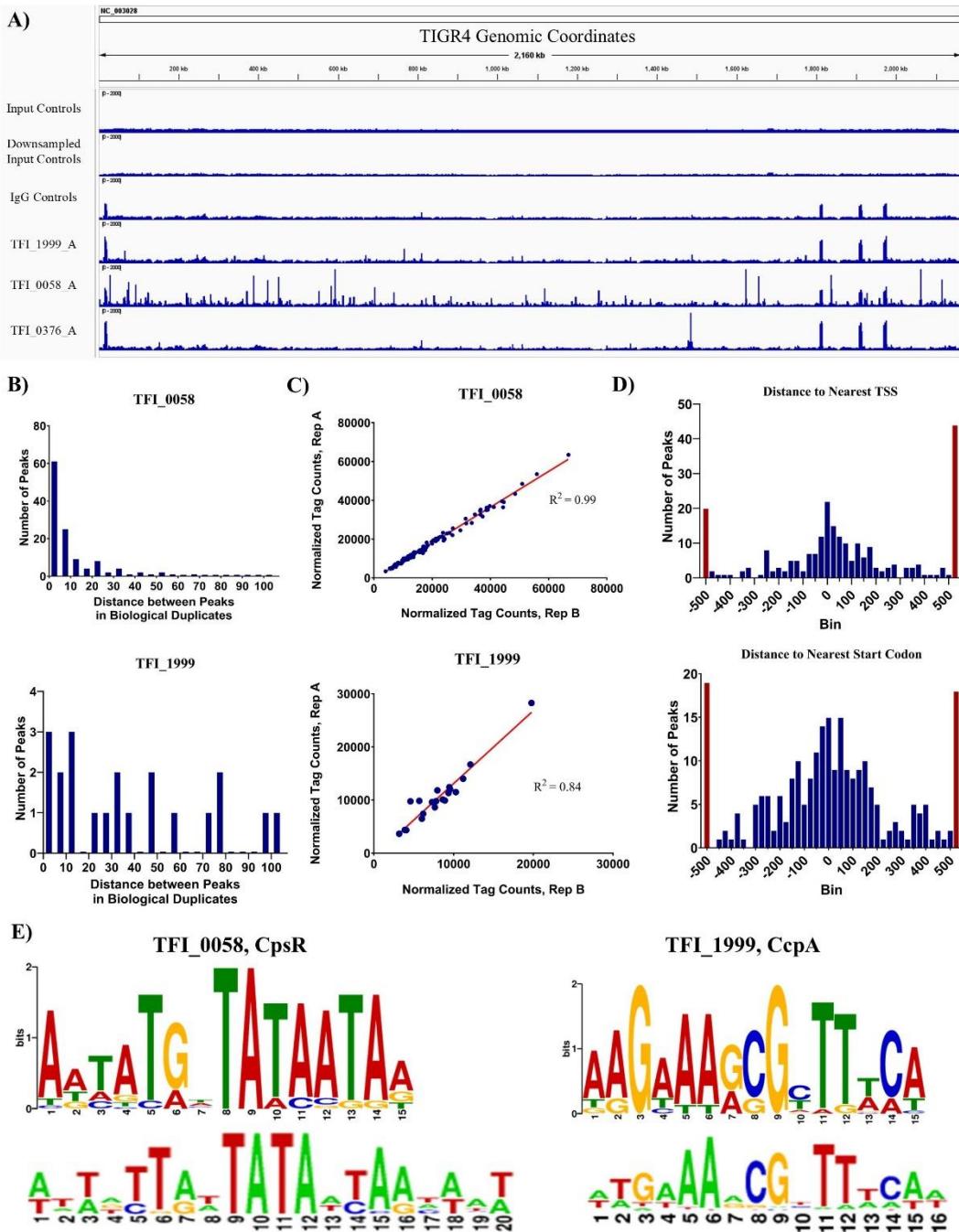


Figure 9. TFI ChIPseq Results. **A)** Visualization of ChIPseq reads aligned to TIGR4's genome. **B)** Histogram showing distances from center of closest significant peaks between the two biological replicates of **Top)** TFI_0058 and **Bottom)** TFI_1999. **C)** Correlations between the normalized tag counts of each aligned peak in both biological duplicates in **Top)** TFI_0058 and **Bottom)** TFI_1999. **D)** Histograms showing distances of peaks from both TFI_0058 and TFI_1999 ChIPseq replicates relative to **Top)** Transcriptional Start Sites (TSS) and **Bottom)** Start Codons. Red bars represent all peaks outside of the boundaries of the x-axis. **E)** Comparisons of binding motifs of CpsR (Left) and CcpA (Right) identified by TFI ChIPseq (Top) and reported motifs in RegPrecise (Bottom).

There are several possible reasons we were only able to obtain binding motifs for CcpA and CpsR. As previously mentioned, these transcription factors are global regulators known to bind DNA when glucose is present. This suggests that both transcription factors had genome-wide binding during the tested condition which increases likelihood of motif discovery. The other transcription factors may not have had the required conditions for DNA binding, or binding was limited to few locations in the genome minimizing chances of successful immunoprecipitation. It's also possible that the C-terminal HA-tag is obscured when bound to DNA or through dimerization, or the HA-tag inhibited DNA binding altogether. TFI strains were designed with a small, flexible linker sequence between the C-terminus of the transcription factor and the HA-tag to minimize this risk, but protein structures or dimerization requirements between different transcription factors may have impacted this. It's impossible to rule out user error, but the positive results for CcpA and CpsR suggest this is not the case. ChIPseq is a time-consuming, laborious and costly technique that failed to identify binding motifs for 67% of TFI strains tested, so we decided to look for alternative methods to differentiate direct and indirect regulation to compile the transcriptional regulatory network.

3.2.2 Computational Discovery of Binding Motifs

With ChIPseq failing to identify binding motifs for several transcription factors, we decided to computationally predict the binding motifs of each transcription factor by interrogating suspected target promoter regions for enriched motifs. First, we identified the promoter sequences of all predicted interactions in our top Ensemble (GENIE3-ARACNE-CLR-GRNBOOST2) network and all genes determined to be Padj in TFI RNAseq. To do this, we created a list of transcript IDs for each possible transcript in TIGR4

according to known operon structure and transcriptional start sites⁴². Then, the combined list of predicted interactions and Padj were converted from target gene to transcript ID with preference given to high-confidence transcriptional start sites. The promoter of each transcript ID was scraped from the TIGR4 genome (NC_003028.3), taking 200 bases upstream and downstream of the transcriptional start site. Finally, a fasta file of all the promoters of all predicted target genes for every transcription factor was created. The fasta file of predicted promoters of all transcription factors was input into XSTREME, a comprehensive motif discovery algorithm in the meme-suite to predict the binding sequences of each transcription factor⁴³. XSTREME may identify several motifs and so a thorough literature search was undertaken to identify known motifs of transcription factors or their homologs and family members. This improves motif identification by having a reference motif length, architecture (i.e. palindromic, direct repeats, etc.) and motif spacing, to compare XTREME-identified motifs to. Motif identification for global regulators is expected to be more accurate than for specific regulators due to the greater number of true motifs across the genome. Importantly, the binding motifs predicted by XSTREME for CcpA (*SP_1999*) and CpsR (*SP_0058*) matched the binding motifs determined by ChIPseq (Fig. 10). The final predicted regulons for all transcription factors were constructed by identifying all occurrences of their motifs within the TIGR4 genome that are within 600 bases of the transcriptional start site of a transcript ID that was identified either by the ensemble of inference methods or RNAseq. This “Combined” network contains 8,786 regulatory interactions and significantly outperformed our Ensemble inferred network, with a precision of 4.75% (up from 2.64%), a recall of 43.4% (up from 28.4%) and an F1 score of 0.086 (up from 0.048), all relative to predicted interactions in

TIGR4 according to the RegPrecise database¹⁹ (Fig. 11). Note that it is not possible to accurately calculate AUPR on this network of combined RNAseq data, inference method prediction and motif presence because there is no way to accurately rank interactions by confidence level.

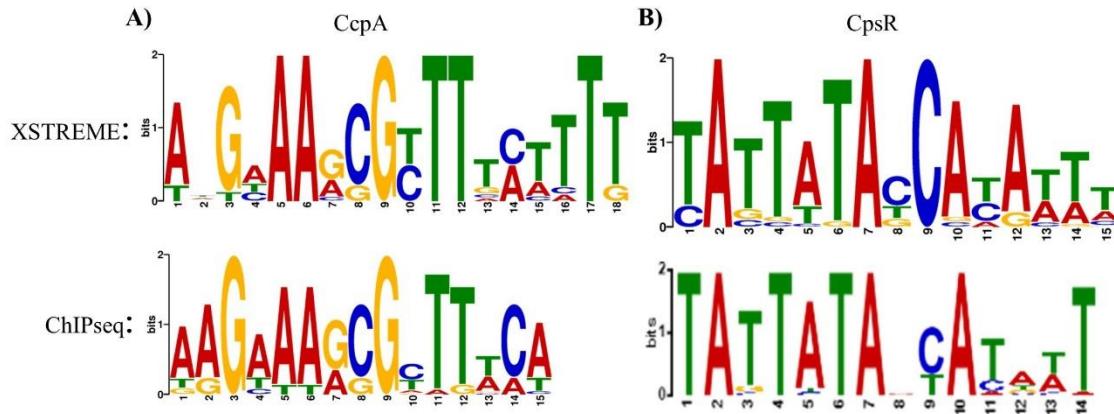


Figure 10. Comparison of Binding Motifs Identified by XSTREME and ChIPseq. Predicted binding motifs for **A)** CcpA and **B)** CpsR via XSTREME (top) and ChIPseq (bottom).

While this combined network greatly improved recall, it still contains a lot of false positives, as evidenced by the total number of interactions, its low precision (4.75%) and low F1 score (0.086). In the *E. coli* TRN, on average each gene is regulated by two transcription factors^{21,22,44}. Assuming a similar distribution in *S. pneumoniae*, we can expect approximately 4,000 interactions. We therefore decided to prune the network further using motif data, RNAseq data, inference method rankings, natural expression of each transcription factor, literature sources and genomic context. This resulted in our “Combined Trimmed” network of 3,663 interactions with a precision of 12.0%, a recall of 43.4% and an F1 score of 0.191 (Fig. 11). Lastly, we manually added interactions from the eight transcription factors for which we could not identify a binding motif, resulting in our Final Network that has 3,867 interactions with a precision of 14.4%, a recall of 54.1% and

an F1 score of 0.227 (Fig. 11). We then added known interactions that were missed by our approach to create the most accurate network possible. Most missed interactions involved regulation by one of two global metabolic regulators, CodY and CcpA. These transcription factors have naturally high expression and are regulated by interaction with different nutritional signals. Therefore, differential regulation of their target genes is not dependent on their expression but on cell state, limiting inference methods that only use expression levels. Also, these transcription factors are expected to act as repressors during exponential growth in SDMM, so overexpression would not result in significant downstream effects.

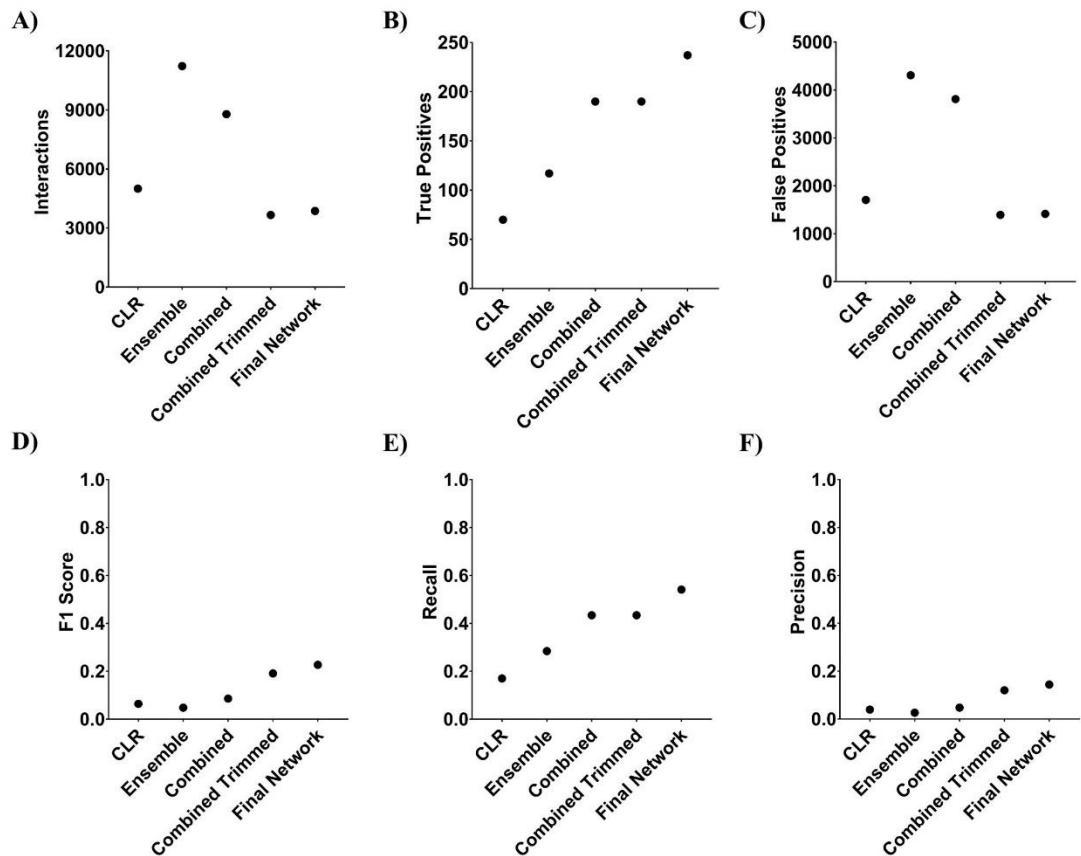


Figure 11. Statistics of Five Networks. Comparison of **A**) the number of total interactions, **B**) the number of true positives, **C**) the number of false positives, **D**) the F1 score, **E**) Recall and **F**) Precision of five select networks. B through F are all calculated in relation to the known interactions from RegPrecise.

3.4 The Transcriptional Regulatory Network of TIGR4

The transcriptional regulatory network of TIGR4 that we assembled through the combination of TFI RNAseq, an ensemble of network inference methods and the identification of predicted binding sites is the most comprehensive TRN for *S. pneumoniae* to date. It includes 3,946 transcription factor – target gene interactions involving 88 transcription factors and 1,607 target genes. Relative to the comparative genomics-based predicted interactions from RegPrecise, our Final Network achieved a recall of 54.1% and a precision of 14.4%, both of which are 10% greater than previous iterations of the network. The average and median out-degree of each transcription factor is 45 and 20, respectively. The degree-distribution in this network follows a power law and represents a scale free network, as expected, with a few global regulators dominating the regulatory network while most transcription factors regulate a small number of genes (Fig. 12). The average and median degree of each target gene is 2.46 and 2, respectively, which is consistent with *E. coli*'s regulatory network^{21,22,44}.

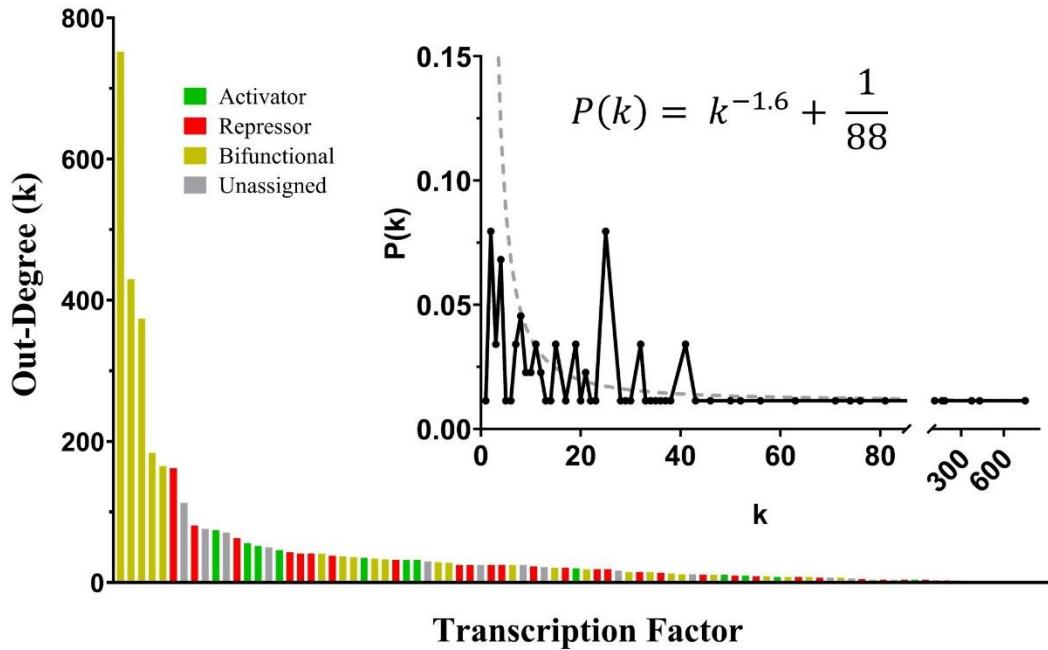


Figure 12. Scale-free Network Topology. Each transcription factor is classified as an activator (green), repressor (red), bifunctional (yellow) or unassigned (grey) depending on mode of regulation determined in TFI RNAseq. Transcription factors in decreasing order of target genes. Inlet: Degree-distribution plot comparing degree, k , with the probability of a transcription factor having that degree, $P(k)$, fit to a power function (grey dotted curve).

The network can be broken down into fifteen modules of five or more genes that are enriched in one or more cellular processes that represent the regulatory logic of these processes (Fig. 13, Table 2). We used the Louvain method⁴⁵ for identifying modularity and tested for gene set enrichment both manually and with FUNAGE-Pro, a gene set enrichment platform designed specifically for prokaryotes⁴⁶. To manually calculate gene set enrichment, we first compiled the gene ontology (GO) biological processes annotations of each gene from FUNAGE-Pro and converted each annotation to its root node. Because genes can be multifunctional and annotations can vary in specificity, GO annotations are hierarchical and represented in an “ancestral tree”, where the children nodes are more specific than their parent nodes. Without converting each annotation to its root node, sibling

annotations would not be a part of the same gene set, and thus not accounted for in gene set enrichment analysis. After converting each annotation to its root, a one-tailed hypergeometric test was performed to identify significantly enriched biological processes within each module. FUNAGE-Pro was only able to identify enrichment in eight modules, possibly due to not accounting for the hierarchical structure of GO annotations. For these eight modules, manual enrichment analysis found similar enriched annotations (Table 2). Removing the dominance of the top 5 global regulators reveals the interconnectedness of the regulatory network between related cellular processes (Fig. 13B). A summary of each transcription factor in TIGR4 can be found in Table 3.

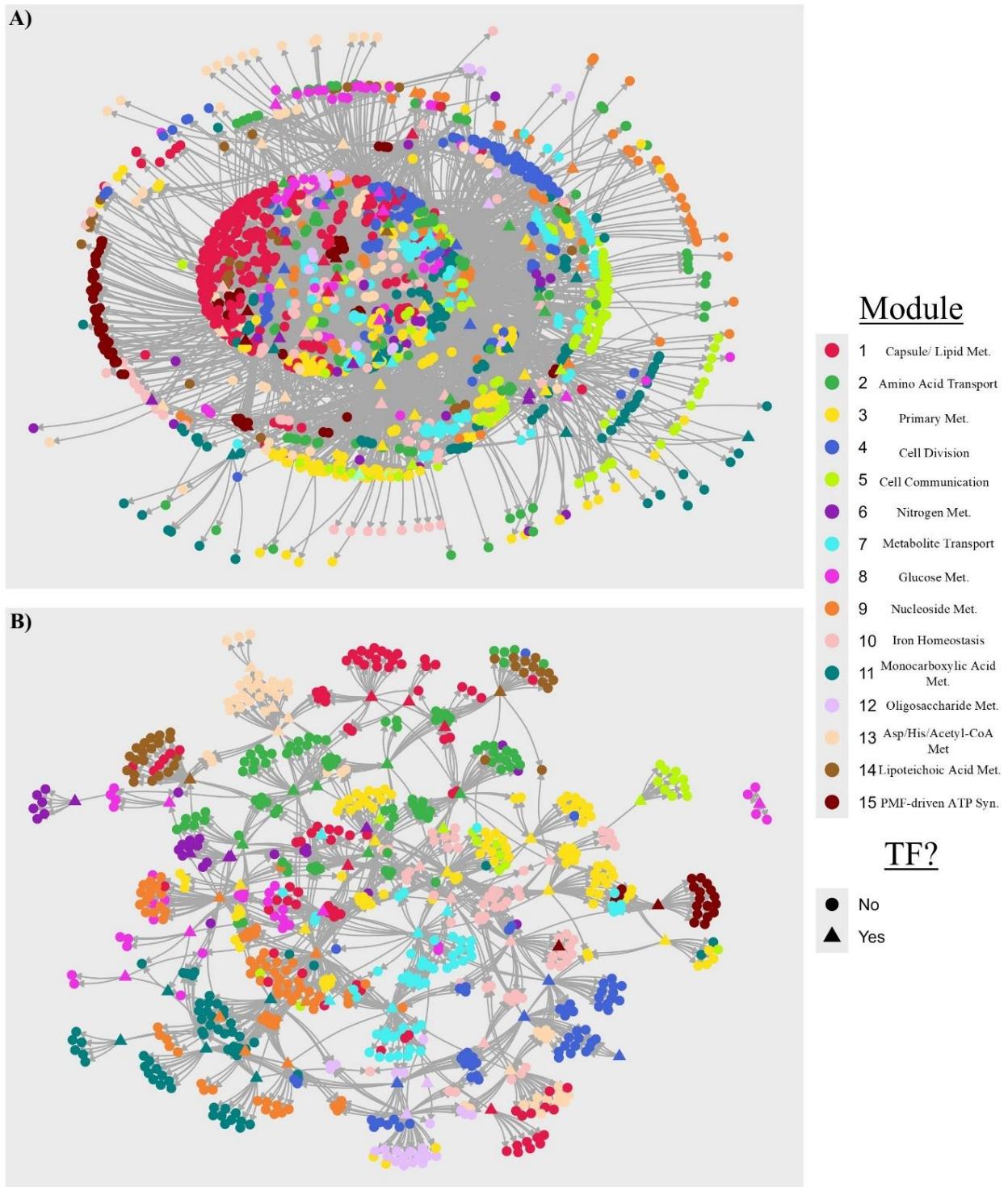


Figure 13. The Transcriptional Regulatory Network of TIGR4. A) Complete and B) Simplified schematics of the Transcriptional Regulatory Network of TIGR4. Genes colored by network modularity. Met. = Metabolism. Syn. = Synthesis.

Module	# Genes	FUNAGE-Pro Enrichments	Manual Enrichments
1	277	Aromatic amino acid biosynthesis, Chorismite biosynthetic processes , Lipopolysaccharide biosynthesis, Capsule polysaccharide biosynthesis , Membrane lipid biosynthesis , Acetoin Catabolism	Cellular metabolic compound salvage, Macromolecule modification, Chorismite metabolism , Vitamin metabolism, Capsule organization , Membrane lipid metabolism , Phosphorous metabolism
2	115	None	Amino acid transport, Lactose metabolism, Galactose metabolism, Energy reserve metabolism
3	197	Zinc ion transport , DNA catabolism, Carbohydrate metabolism , 'de novo' UMP biosynthesis, N-acetyleneuraminate catabolism, Mannose transport	Transition metal ion transport , Response to nutrient levels, Primary metabolism , Poly(ribitol-phosphate) teichoic acid metabolism
4	141	Amino Acid metabolism, Division Septum assembly , Protein transport, Phosphate ion transport , 'de novo' Pyrimidine biosynthesis	Cellular component assembly, Reproductive process , Organic substance transport, Inorganic anion transport
5	116	None	Cell communication, Regulation of Amide metabolism, DNA metabolism, Monocarboxylic acid metabolism, Amino Acid metabolism
6	38	Valine biosynthesis , Isoleucine biosynthesis , Amino acid transport , Arginine Metabolism	Amino acid metabolism , Nitrogen compound transport , Arginine metabolism , Vitamin metabolism
7	96	None	Organic substance transport, Generation of precursor metabolites and energy, D-Allose metabolism, Cell wall macromolecule biosynthesis, O-antigen metabolism
8	64	None	Diaminopimelate metabolism, Glucose metabolism, Nucleotide metabolism, Sulfur compound metabolism
9	83	Pyridoxal phosphate biosynthesis, Double-strand break repair via Homologous Recombination , 'de novo' IMP biosynthesis	Nucleotide metabolism , Double strand break repair , Aromatic compound metabolism
10	84	Siderophore-dependent iron import, Iron-sulfur cluster assembly , Iron ion homeostasis, Arginine catabolism to Ornithine , Glycerol metabolism	Ornithine Metabolism , Sulfur compound metabolism , Polyol metabolism
11	134	None	Monocarboxylic acid metabolism, Malonyl-CoA metabolism
12	26	None	Glucose metabolism, Nitrogen compound transport, Trehalose metabolism, Glycine metabolism, Oligosaccharide metabolism, Disaccharide transport
13	84	None	Aspartate metabolism, Histidine metabolism, Acetyl-CoA metabolism, Nicotinamide nucleotide metabolism, diaminopimelate metabolism, Nitrogen cycle metabolism, siderophore metabolism, Clearance of foreign DNA
14	36	Lipoteichoic acid biosynthesis , Ribosome	Macromolecule metabolism, Lipoteichoic acid metabolism , Dicarboxylic acid metabolism
15	84	None	PMF-driven ATP synthesis, Metal ion transport, Tetrahydrafolate metabolism, Amide metabolism, Selenocysteine metabolism, Response to antibiotic

Table 2. Transcriptional Regulatory Network Module Enrichments.

Table 3. Summary of Transcription Factors in TIGR4.

Locus	TF	TF Family	Role	Core?	Essential?	Ref
<i>SP_0014</i>	ComX	Sigma	Competence	Accessory	Non-essential	47–50
<i>SP_0018</i>	ComW	ComW	Competence	Core	Non-essential	51,52
<i>SP_0058</i>	CpsR	GntR	Metabolism, Capsule	Core	Non-essential	19,53
<i>SP_0083</i>	SaeR (TCS08)	OmpR	Virulence, Metabolism	Core	Non-essential	54,55
<i>SP_0100</i>	PtvR	PadR	Stress	Core	Non-essential	19,20
<i>SP_0141</i>	MutR1	XRE	Stress	Accessory	Non-essential	56
<i>SP_0156</i>	YesN (TCS07)	YesN/AraC	Glycan Metabolism	Accessory	Non-essential	54,55,57
<i>SP_0161</i>		LytR	Stress	Core	Non-essential	58
<i>SP_0163</i>	TprC	XRE	Stress	Accessory	Non-essential	59,60
<i>SP_0189</i>	SpxA2	NA	OSR	Core	Strain-dependent	61

<i>SP_0246</i>	GlpR	DeoR	Sugar Metabolism	Core	Non-essential	19
<i>SP_0247</i>	LsrR	DeoR	Metabolism	Core	Non-essential	62,63
<i>SP_0330</i>		LacI	Metabolism	Core	Non-essential	19
<i>SP_0333</i>		XRE	Stress	Core	Non-essential	
<i>SP_0376</i>	RitR	OmpR	Iron, OSR, Proteostasis	Core	Strain-dependent	64,65
<i>SP_0387</i>	LiaR (TCS03)	NarL/ FixJ	Stress	Core	Non-essential	54,55
<i>SP_0416</i>	FabT	MarR	Stress, Capsule	Core	Strain-dependent	19,53,66,67
<i>SP_0461</i>	RlrA	PCVR	Virulence	Accessory	Non-essential	68
<i>SP_0473</i>	XylR	ROK	Xylose Metabolism	Accessory	Non-essential	19,69
<i>SP_0501</i>	GlnR	MerR	Glutamine Metabolism	Core	Strain-dependent	19,70–72
<i>SP_0515</i>	HrcA	HrcA	Heat Shock	Core	Non-essential	19,73–75
<i>SP_0525</i>	BlpS	LytR	Bacteriocins	Core	Non-essential	76,77
<i>SP_0526</i>	BlpR (TCS13)	LytR	Bacteriocins	Accessory	Non-essential	54,55,78
<i>SP_0584</i>		NA	?	Accessory	Non-essential	
<i>SP_0593</i>	CdaR	PucR	Sugar Metabolism	Core	Non-essential	79
<i>SP_0603</i>	VncR (TCS10)	OmpR	Stress	Core	Non-essential	54,55
<i>SP_0661</i>	ZmpR (TCS09)	YesN/AraC	Virulence	Core	Non-essential	54,55
<i>SP_0676</i>	MtaR	LysR	Methionine	Core	Non-essential	19,80
<i>SP_0727</i>	CopY	CopY	Copper	Core	Non-essential	19,81
<i>SP_0739</i>	SoxR	MerR	OSR	Accessory	Non-essential	82
<i>SP_0743</i>	AcrR	TetR/AcrR	Efflux Pump	Core	Non-essential	83
<i>SP_0789</i>	PadR	PadR	Phenolic Acid Stress	Core	Non-essential	20
<i>SP_0798</i>	CiaR (TCS05)	OmpR	Competence, Heat Shock, OSR, Virulence	Core	Non-essential	54,55,72,84
<i>SP_0875</i>	FruR	DeoR	Fructose Metabolism	Core	Non-essential	19
<i>SP_0888</i>	AbrB	AbrB	Toxin/Antitoxin	Accessory	Non-essential	85,86
<i>SP_0893</i>	ArgR2	ArgR	Arginine Deiminase System	Core	Non-essential	87,88
<i>SP_0926</i>		AraC	-	Accessory	Non-essential	
<i>SP_0927</i>	CmbR	LysR	Cysteine/Methionine Metabolism	Core	Strain-dependent	19,89
<i>SP_0951</i>	TfoX	TfoX/Sxy	-	Accessory	Non-essential	
<i>SP_1030</i>	YozG	XRE	Stress	Accessory	Non-essential	90
<i>SP_1050</i>	PezA	XRE	Toxin/Antitoxin	Accessory	Non-essential	85,91
<i>SP_1057</i>	TprB	XRE	Stress	Accessory	Non-essential	59,92

<i>SP_1073</i>	RpoD	Sigma	Housekeeping Sigma	Core	Universal	93,94
<i>SP_1090</i>	Rex	Rex	Redox Homeostasis	Core	Non-essential	19,95
<i>SP_1115</i>	RggD	XRE	Ribosomal protection, Metabolism	Accessory	Non-essential	56,96
<i>SP_1130</i>	YiaG	XRE	Stress	Accessory	Non-essential	
<i>SP_1131</i>		XRE	Phage?	Accessory	Non-essential	
<i>SP_1134</i>	Rha	Rha	Phage?	Accessory	Non-essential	
<i>SP_1144</i>	HigA	XRE	Toxin/Antitoxin	Accessory	Non-essential	85
<i>SP_1182</i>	LacR	DeoR	Lactose Metabolism	Core	Non-essential	19,97
<i>SP_1203</i>	AhrC	ArgR	Arginine Uptake	Core	Non-essential	87,88,98
<i>SP_1224</i>	RelB	NA	Toxin/Antitoxin	Accessory	Non-essential	85
<i>SP_1227</i>	VicR (TCS02)	OmpR	Cell Wall Turnover, Fatty Acid Synthesis	Core	Universal	54,55,99
<i>SP_1234</i>	NiaR	NadR	NAD Synthesis	Core	Non-essential	19
<i>SP_1324</i>	NagC	ROK	-	Accessory	Non-essential	
<i>SP_1331</i>	NanR2	MurR/RpiR	Sialic Acid Metabolism	Accessory	Non-essential	19,100,101
<i>SP_1393</i>	PdxR	MocR	Pyridoxine Metabolism	Accessory	Non-essential	102
<i>SP_1423</i>		XRE	-	Core	Non-essential	
<i>SP_1433</i>	AraC	AraC	Sugar Metabolism	Accessory	Non-essential	103
<i>SP_1446</i>	NagR	GntR	GlcNAc Utilization	Accessory	Non-essential	19,104,105
<i>SP_1584</i>	CodY	CodY	Global Nutrient Regulaor	Core	Universal	19,106–110
<i>SP_1633</i>	SirR (TSC01)	OmpR	Stress	Core	Non-essential	52,53,109, 110
<i>SP_1636</i>	SifR	Rrf2	Iron Uptake	Core	Non-essential	113
<i>SP_1638</i>	PsaR	DtxR	Managese Uptake	Core	Non-essential	19,114–117
<i>SP_1674</i>	NanR	MurR/RpiR	Sialic Acid Metabolism	Core	Non-essential	19,100,101
<i>SP_1713</i>	NrdR	NrdR	Ribonucleotide Reductase Regulator	Core	Strain-dependent	19,118
<i>SP_1714</i>	YheF	GntR	Transporters	Core	Strain-dependent	19,119
<i>SP_1725</i>	ScrR	LacI	Sucrose Metabolism	Accessory	Non-essential	19,120,121
<i>SP_1741</i>	YefM	Phd/YefM	Toxin/Antitoxin	Accessory	Non-essential	85
<i>SP_1774</i>		MarR	OSR	Core	Non-essential	122
<i>SP_1786</i>	HicB	XRE	Toxin/Antitoxin	Core	Non-essential	85
<i>SP_1799</i>	SusR	LacI	Sucrose Metabolism	Core	Non-essential	19,120,121

<i>SP_1800</i>	MgrA	PCRV	Virulence	Accessory	Non-essential	123,124
<i>SP_1809</i>		XRE	-	Accessory	Non-essential	
<i>SP_1821</i>		LacI	Metal Uptake	Core	Non-essential	
<i>SP_1854</i>	GalR	LacI	Galactose Metabolism	Core	Non-essential	19,125,126
<i>SP_1856</i>	NmlR	MerR	OSR	Core	Non-essential	19,69,81
<i>SP_1858</i>	SzcA	TetR/AcrR	Zinc Efflux	Core	Non-essential	19,81,127
<i>SP_1863</i>		MarR	Osmoprotectant Import	Core	Non-essential	128
<i>SP_1885</i>	TreR	GntR	Trehalose Metabolism	Accessory	Non-essential	19,129
<i>SP_1899</i>	MsmR	AraC	Sugar Metabolism	Core	Non-essential	130,131
<i>SP_1900</i>	BirA	BirA	Biotin Processing	Core	Strain-dependent	19,132
<i>SP_1904</i>		Rgg/GadR/MutR	-	Accessory	Non-essential	133,134
<i>SP_1915</i>		LytR	OSR	Core	Non-essential	
<i>SP_1920</i>		MarR	Transporter	Accessory	Non-essential	19
<i>SP_1922</i>	YebC	YebC/PmpR	Acid Stress, Capsule	Core	Non-essential	135
<i>SP_1936</i>		XRE	Toxin/Antitoxin	Accessory	Non-essential	85
<i>SP_1946</i>	PlcR	XRE	Quorum Sensing, Bacteriocin	Accessory	Non-essential	136
<i>SP_1979</i>	PurR	PurR	Puring Synthesis	Core	Non-essential	19,137
<i>SP_1989</i>	HipB	XRE	Toxin/Antitoxin	Accessory	Non-essential	85,138
<i>SP_1999</i>	CcpA	CcpA	Carbon Catabolite Repression	Core	Strain-dependent	1,17,69, 137–139
<i>SP_2000</i>	DesR (TCS11)	NarL/ FixJ	Biofilm, Sugar Metabolism	Core	Non-essential	54,55
<i>SP_2006</i>	ComX	Sigma	Competence	Accessory	Non-essential	47–50
<i>SP_2020</i>	BguR	GntR	Biofilm, Sugar Transport	Accessory	Non-essential	142
<i>SP_2062</i>		MarR	Efflux Pumps	Accessory	Non-essential	66
<i>SP_2077</i>	ArgR	ArgR	Arginine Uptake	Core	Non-essential	87,88
<i>SP_2082</i>	Pnpr (TCS04)	OmpR	Manganese Uptake, Sugar Metabolism	Accessory	Non-essential	54,55,143
<i>SP_2090</i>		NA	-	Accessory	Non-essential	
<i>SP_2112</i>	MalR	LacI	Maltose Metabolism	Core	Non-essential	19,126
<i>SP_2119</i>		XRE	Stress, Membrane Composition	Core	Non-essential	
<i>SP_2123</i>	RggM	Rgg/GadR/MutR	Quorum Sensing	Accessory	Non-essential	56,92,134
<i>SP_2168</i>	FucR	DeoR	Fucose Metabolism	Accessory	Non-essential	19
<i>SP_2172</i>	AdcR	MarR	Zinc Import	Core	Non-essential	19,26,144
<i>SP_2187</i>		PCVR	Glycan Metabolism	Accessory	Non-essential	145

<i>SP_2193</i>	CbpR (TCS06)	OmpR	Virulence, Adherence, Complement Evasion	Accessory	Non-essential	54,55
<i>SP_2195</i>	CtsR	CtsR	Proteostasis	Core	Non-essential	19,75,81,146
<i>SP_2234</i>	PipR	TetR/AcrR	Phage	Core	Non-essential	19
<i>SP_2235</i>	ComE (TSC12)	LytR	Competence	Accessory	Non-essential	52,53,73, 145,146

3.4.1 Transcription Factor Activities under Antibiotic Stress

With an established architecture of the transcriptional regulatory network of TIGR4, we can identify transcription factor activities during antibiotic treatment to understand how the cells are responding to the different stresses (Fig. 14). DecouplR¹⁴⁹ uses an ensemble of computational methods to infer transcription factor activity across antibiotic treatments when fed DESeq data and our transcriptional regulatory network. It returns a Score, a metric that represents transcription factor activity where positive and negative values represent active and inactive transcription factors, respectively, as well as a p-value for significance. Identifying transcription factor activities across various time-course antibiotic experiments reveal several regulators with significant responses.

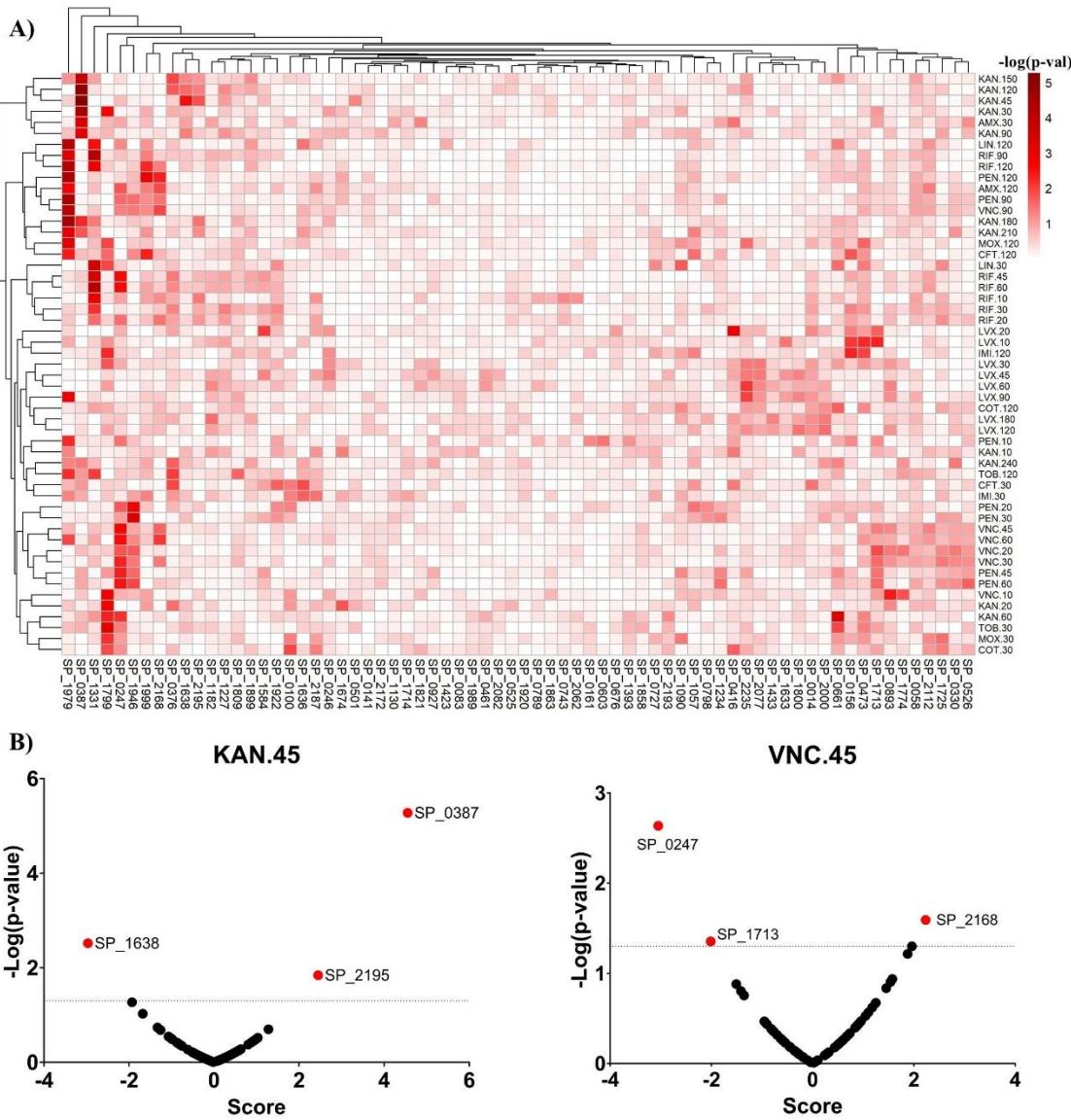


Figure 14. Transcription Factor Activity during Antibiotic Treatments. **A)** Heatmap showing the significance of each TF's activity during various antibiotic treatment timepoints as determined by decouplR. **B)** Volcano plots of Score (activity, + = active, - = inactive) and log-transformed p-values of TF activity in kanamycin (Left) and vancomycin (Right) treatment.

DecouplR identified CtsR (*SP_2195*), LiaR (*SP_0387*) and PsaR (*SP_1638*) as significant regulators in response to Kanamycin (Fig. 14B). CtsR regulates the expression of the Clp protease and is expected to be impactful during translational stress of kanamycin. Similarly, LiaR is the response regulator of TCS03 which senses and regulates cell wall stress through regulation of a putative multidrug efflux pump and the heat shock stress response regulator,

HrcA¹⁵⁰. Lastly, PsaR regulates manganese homeostasis through *psaABC* as well as part of the oxidative stress response through regulation of *tpxD*¹¹⁴. These three transcription factors regulate proteostasis and the oxidative stress response which align with the known translational and oxidative stress conferred by kanamycin treatment, and demonstrate the generalizable insight our network-based approach has in evaluating gene expression data.

Conversely, the three transcription factors found to have a significant role in adapting the transcriptome to vancomycin are three metabolic regulators, LsrR (*SP_0247*), NrdR (*SP_1713*) and FucR (*SP_2168*) (Fig. 14B). LsrR regulates expression of ribosomal proteins and pyruvate formate lyase, a key enzyme central to pyruvate metabolism. NrdR is an ATP-sensing regulator of deoxynucleotide synthesis through ribonucleotide reductase and FucR regulates fucose metabolism. How exactly these regulons impact the response to vancomycin is unclear, but metabolic adaptation appears critical to vancomycin stress. This differs from the importance of proteostasis regulators during translational stress induced by kanamycin, and indicates that distinct pathways within *S. pneumoniae*'s transcriptional regulatory network have evolved to overcome various stresses.

3.5 Discussion

S. pneumoniae has evolved tremendous adaptability as an opportunistic pathogen, navigating vastly different environments—whether residing as a commensal in the nasopharynx or causing invasive infections the lungs, bloodstream, bones or brain. In these diverse niches, the bacterium must contend with competing microbial species, evade the host immune system and survive antibiotic attacks. Its ability to catabolize 32 distinct sugars encountered in the host requires precise metabolic regulation¹⁵¹ and employs

complex quorum sensing systems capable of cell-to-cell communication^{50,152,153}. Critical to *S. pneumoniae*'s ability to survive across such variable and hostile environments is a tightly regulated and efficient transcriptional response, yet little is known about its transcriptional regulatory network.

To date, no systematic effort has been made to map transcriptional regulation in *S. pneumoniae* on a genome-wide scale. Previous studies have largely focused on individual transcription factors, typically using knockout strains out to identify DEGs as the transcription factor's regulon. However, these regulons are prone to include false positive interactions, likely resulting from indirect regulation, and may miss true regulatory interactions when target genes are not classified as DEGs, either by insufficient magnitude of differential expression or the limited experimental conditions assayed. Additionally, these studies span several strains of *S. pneumoniae*, posing the challenge of integrating data from different strains into a unified network.

Here we present the first most comprehensive transcriptional regulatory network of *S. pneumoniae* by identifying regulatory interactions for 88 of the 108 transcription factors (81%) identified in TIGR4. Regulatory interactions were predicted from three complementary approaches: 1) gene expression analysis of transcription factor overexpressing strains, 2) an ensemble of network inference methods trained on 604 RNAseq samples, and 3) transcription factor binding motif discovery. The combined use of transcription factor overexpression gene expression profiling and an ensemble of network inference methods trained on a large, diverse gene expression dataset ensures we probed the transcriptome in all its states. Perhaps the greatest challenge to reverse engineering transcriptional regulatory networks is to differentiate direct and indirect

regulation, and to do this we predicted a DNA-binding motif for 78 of the 108 transcription factors (72%). Fortunately, transcription factors that regulate a single operon are more difficult to predict binding motifs for, reducing the risk of many false positive interactions arising from these transcription factors. Together with manual curation, our approach yielded a network of 3,946 total regulatory interactions organized into 15 modules enriched in specific biological processes. This includes novel predicted regulons for 51 transcription factors encompassing 3,507 novel interactions.

Reconstructing comprehensive transcriptional regulatory networks is inherently challenging due to highly correlated expression of regulons involved in similar cellular processes and the independency between transcription factor expression and activity. Our network achieved a precision and recall of 54.1% and 14.4%, respectively, relative to the RegPrecise database. As the RegPrecise database is predicted by comparative genomics and not intended to be comprehensive, low precision is expected and reflects novel interactions identified in our study. Interactions predicted in RegPrecise but absent in our network were neither significantly differentially regulated by transcription factor overexpression nor ranked in our ensemble of network inference methods. We retroactively added these missed interactions to the final network to ensure we created the most comprehensive network, but it is possible that true interactions were still missed, likely due to not testing the correct environment. The activity of many transcription factors relies on the presence or absence of specific signals, such as an effector molecule, that if absent, would lead to no differential activity of the transcription factor. This motivated the complementary approach of using an ensemble of network inference methods to predict interactions from our lab's dataset of 604 RNAseq samples. However, our lab primarily

studies antibiotic resistance, so this dataset is enriched in antibiotic-treated samples and control samples in SDMM. Further, stressful conditions such as antibiotic treatment are known to cause transcriptional entropy, which may reduce the effectiveness of network inference methods due to the added noise. Future RNAseq experiments in minimal media with varied supplements—such as alternate carbon or nitrogen sources and variation of metal ion concentrations—could improve our network’s coverage. Additionally, further experimental verification of the predicted binding motifs would improve the network, though the signal-to-noise ratio in ChIPseq within our system remains a limitation. Of the six TFI strains we assayed with ChIPseq, only two produced binding motifs, both of which are global regulators that act as repressors in the condition they were assayed. Fortunately, our network would help identify experimental conditions for future DNA-binding assays, and developments in assay technology, such as adaptation of CUT&Tag to prokaryotes, will improve outcomes. Alternatively, *in vitro* methods such as DNase-footprinting or DAPseq could identify the binding motifs of each transcription factor, which could be then used to prune interactions predicted computationally or through TFI RNAseq.

By feeding the network architecture and antibiotic-treated gene expression data to an ensemble of activity inference methods compiled through decouplR, we were able to identify the which transcription factors coordinate antibiotic stress responses. For the protein synthesis inhibitor, kanamycin, we showed that proteostasis regulators, including the regulator of the Clp protease, CtsR, were paramount in responding to the translational stress. Conversely, several metabolic regulators were shown to regulate response to the cell wall synthesis inhibitor, vancomycin. DecouplR depends on a defined “mode of regulation” which ranges from -1 (strong repressor) to 1 (strong activator), which is

difficult to determine for our network. We used both the correlation coefficient between transcription factor and target gene and a binary classification of activator or repressor as determined by TFI RNAseq as modes of regulation, and there was little difference in the results. Neither of these are optimal though as transcription factor expression doesn't necessarily equal activity and not all interactions were identified by TFI RNAseq and thus can't be classified as activator or repressor. Future methods for extracting transcription factor activity from gene expression profiles would be improved with context-specific modes of regulation, although identifying these would be extremely difficult. While there's room for improvement, we still were able to identify significant regulons in response to various antibiotics which represent expected and novel insights.

In addition to inferring transcription factor activity through antibiotic treated gene expression data, we designed our TFI strains with unique DNA barcodes to allow for phenotyping of each strain. Overexpressing each individual transcription factor in pooled competition assays, we can synthetically stimulate each regulon to assess each transcription factor's effect on antibiotic stress response. Repeating this assay in several infection-relevant conditions can identify which transcription factors influence fitness the most, and network analysis of these results will identify routes towards antibiotic potentiation

3.6 Materials and Methods

RNA Sequencing: Every TFI strain was treated with 1mM IPTG to induce TF-overexpression and subjected to RNA sequencing. Parallel cultures of biological triplicates of each TFI strain were grown to exponential phase in SDMM and SDMM + IPTG before being re-inoculated into fresh media at an OD of 0.1 and cultured for 2h. Pretreatment of the TF-overexpressing samples was critical to ensure there was high overexpression at t=0 and samples were collected after 2h to ensure all strains were in exponential phase to approximate a steady-state system. After 2h of growth, cell pellets were collected via centrifugation and snap frozen. RNA was isolated using the RNeasy Mini Kit (Qiagen) and 650ng of RNA was used as input for library prep as described previously with minor modifications. Briefly, total RNA was fragmented with heat and treated with FastAP (Thermo Scientific) and TURBO DNase (Invitrogen). RNA barcodes were ligated onto DNase-treated RNA and samples were pooled using the RNA Clean & Concentrator kit (Zymo Research). rRNA was removed with a TIGR4-specific RNase-H based rRNA depletion step as described in with TIGR4-specific DNA probes. cDNA was generated with AffinityScript Reverse Transcriptase (Agilent) and a second barcode was ligated before Q5 PCR (New England Biolabs) enrichment. Samples were pooled and sequenced with a P2 100-cycle kit on the NextSeq2000 (Illumina) with a 10% PhiX spike, resulting in a depth of ~3.9 million assigned reads per sample.

RNA sequencing reads were processed with the Aerobio analysis pipeline. Briefly, reads were demultiplexed based off their unique combination of barcodes and quality filtered. Bowtie2 was used to align reads to the TIGR4 genome (NC_003028) and reads were quantified with featureCounts. Differential expression was calculated between SDMM and SDMM + IPTG samples using DESeq2¹⁵⁴. Genes with a log2-fold change of $\geq |1|$ with an adjusted p-value (padj) of < 0.05 were determined to be differentially expressed genes (referred to as DEGs, see TFI_DEGs.csv). Many known target genes did not reach the log2-fold change cutoff from TF-overexpression alone, so genes with significant padj values were considered differentially expressed as well (referred to as Padjs, see TFI_Padjs.csv).

Network Inference Methods: To directly compare sequencing results between runs, the transcripts per million (TPM) of each gene was calculated for each run because TPM

normalizes for both gene length and sequencing depth. To do this, read counts were first divided by the length of the gene in kilobases to determine reads per kilobase (RPK). The RPKs were summed within a single sequencing run and this sum was divided by 1 million to determine the scaling factor. Individual genes' RPKs were then divided by that run's scaling factor to produce its TPM value. TPM + 1 values were then log-normalized, resulting in $\log_2(\text{TPM}+1)$ values which will from now on be referred to as TPMs. All three datasets (discussed in section __) were independently analyzed with 8 network inference methods: KBOOST, PCIT, MRNETB, ARACNE, BC3NET, GENIE3, GRNBOOST2 and CLR. Analysis was carried out via BIO-INSIGHT through Docker. For each method, every predicted interaction that did not include a transcription factor was removed. Predicted interactions from any single method was capped at 10,000 and all interactions were ordered by decreasing significance. Interactions from all methods were then concatenated into final datasets ("Merged_DatasetA.csv", "Merged_DatasetB.csv" and "Merged_DatasetC.csv"). To combine the results of each inference method, the average rank of each interaction was calculated across the dataset. Interactions that fail to be called by an inference method are punished by giving them a ranking of the maximum ranking (equal to the maximum number of interactions predicted by a method) plus one. This ensures that interactions that are repeatedly predicted have lower average ranks than interactions that are only called in a single inference method. The recall, precision, area under precision recall (AUPR) and F1 score was calculated for the total community network in R. To identify the optimal combination of inference methods, the total community network was randomly subset to one to seven methods and retested for precision, recall, AUPR and F1 score. This process was repeated 1,000 times to identify the optimal network.

ChIPseq: TFI starter cultures were pretreated with IPTG to induce transcription factor overexpression before being reinoculated into 40mL of SDMM + IPTG at an OD of 0.1 and cultured for 2h. 4mL of Complete Fixing solution (70% volume Fixing base: 50mM Tris, 100mM NaCl, 0.5mM EGTA, 1mM EDTA, pH 8.0 + 30% volume 36.5% formaldehyde) was added and the cells were rotated at room temperature for 30 minutes. Crosslinking was quenched with the addition of 4mL Quenching solution (50mM Tris, 100mM NaCl, 0.5mM EGTA, 1 mM EDTA, 1.25M glycine, pH 8.0) and rotation at 4°C for 30 minutes. Cells were washed three times with ice cold PBS and snap frozen. Pellets

were thawed and resuspended in 4mL of IP Lysis Buffer (25mM Tris, 150mM NaCl, 1mM EDTA, 1% NP40, 5% glycerol, pH 7.4) supplemented with 1mM PMSF and 1X Roche Complete Protease Inhibitor Tablet. Cells were lysed and DNA was sheared with sonication on ice at 35% amplitude for 30 seconds on and off for a total of 10 minutes on. Immunoprecipitation was carried out overnight at 4°C with Pierce Protein G Magnetic Beads (Thermo Scientific) and 5 ug HA Tag Monoclonal Antibody (2-2.2.14, Invitrogen), according to instructions. Transcription factor-DNA complexes were eluted off the beads with Elution Buffer (50 mM Tris-HCl 10mM EDTA, 1% SDS, pH 8.0) freshly supplemented with 1uL of Proteinase K (20mg/mL) and incubated overnight at 65°C to both reverse crosslinking and degrade proteins. The elutant was treated with RNase A and DNA was purified using the QIAquick PCR Purification kit (Qiagen). Libraries were prepped for sequencing using the NEB Ultra II DNA Library prep kit (New England Biolabs) and sequenced on an Illumina NextSeq 2000. Adapters and low-quality bases were removed from the sequencing reads by cutadapt and reads were aligned to the TIGR4 genome with the Burrows-Wheeler Aligner. Significant peaks were called relative to no-antibody controls using MACS2 and binding motifs were identified within peaks by MEME-ChIP.

Motif Discovery and Analysis: A list of transcript IDs (TIDs) was created for each possible transcript in TIGR4 according to known operon structure and transcriptional start sites. Then, for every transcription factor, a list of potential target genes was created by combining all interactions identified in our optimal inferred network with all target genes identified as significantly differentially expressed (according to padj values only) in TFI RNAseq. Potential target genes were converted to potential target TIDs with preference given to TIDs with high-confidence transcriptional start sites. The promoter of each transcript ID was scraped from the TIGR4 genome (NC_003028.3) taking 200 bases up- and downstream of the transcriptional start site. Finally, a fasta file of all the promoters of all predicted target genes for every transcription factor was created. The fasta file of predicted promoters of all transcription factors was input into XSTREME, a comprehensive motif discovery algorithm in the MEME-suite to predict the binding sequences of each transcription factor. A thorough literature review was conducted to identify which motif identified by XSTREME was the most likely motif for each

transcription factor. For specific regulators with only a single or few target promoters, these promoters were searched directly as opposed to all predicted interaction by the inference methods and RNAseq. Once a motif was identified, FIMO was used to identify all occurrences of it in all TIGR4 promoters. This was included to avoid missing motifs of specific regulators that occurred outside of the promoters in which they were originally identified and to find the exact genomic sequences.

Module Detection and Enrichment: Modularity was identified within the network using the Louvain method within the R package igraph. The statistical enrichment of each module was calculated by both FUANGE-Pro and manually. To manually calculate gene set enrichment, we first compiled the gene ontology (GO) biological processes annotations of each gene in our network from FUNAGE-Pro and converted each annotation to its root node. “Isa” relationships within GO annotations, provided by the AnnotationDbi package in Bioconductor, were used to create the network structure of GO terms in R. The “biological processes” node was deleted and then all children nodes were converted to their root node. This ensures that each gene is annotated to the same specificity and results in each parent node inheriting their children nodes. After converting each annotation to its root, a one-tailed hypergeometric test was performed to identify significantly enriched biological processes within each module.

3.7 References

1. Carvalho, S. M., Kloosterman, T. G., Kuipers, O. P. & Neves, A. R. CcpA Ensures Optimal Metabolic Fitness of *Streptococcus pneumoniae*. *PLoS One* **6**, e26707 (2011).
2. Schumacher, M. A. *et al.* Structural Basis for Allosteric Control of the Transcription Regulator CcpA by the Phosphoprotein HPr-Ser46-P. *Cell* **118**, 731–741 (2004).
3. Görke, B. & Stölke, J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613–624 (2008).
4. Fleming, E., Lazinski, D. W. & Camilli, A. Carbon catabolite repression by seryl phosphorylated HPr is essential to *S. pneumoniae* in carbohydrate-rich environments. *Mol. Microbiol.* **97**, 360–380 (2015).
5. Goeddel, D. V *et al.* Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proc. Natl. Acad. Sci.* **76**, 106–110 (1979).
6. Songdech, P. *et al.* Increased production of isobutanol from xylose through metabolic engineering of *Saccharomyces cerevisiae* overexpressing transcription factor Znf1 and exogenous genes. *FEMS Yeast Res.* **24**, (2024).
7. Sorg, R. A., Gallay, C., van Maele, L., Sirard, J. C. & Veening, J. W. Synthetic gene-regulatory networks in the opportunistic human pathogen *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 27608–27619 (2020).
8. Olson, E. J., Hartsough, L. A., Landry, B. P., Shroff, R. & Tabor, J. J. Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat. Methods* **11**, 449–455 (2014).
9. Sprinzak, D. & Elowitz, M. B. Reconstruction of genetic circuits. *Nature* **438**, 443–448 (2005).
10. Nadezhdin, E., Murphy, N., Dalchau, N., Phillips, A. & Locke, J. C. W. Stochastic pulsing of gene expression enables the generation of spatial patterns in *Bacillus subtilis* biofilms. *Nat. Commun.* **11**, 950 (2020).
11. Guo, X. *et al.* Thwarting resistance: MgrA inhibition with methylophipogonanone unveils a new battlefield against *S. aureus*. *npj Biofilms Microbiomes* **10**, 15 (2024).
12. Vervoort, S. J. *et al.* Targeting transcription cycles in cancer. *Nat. Rev. Cancer* **22**, 5–24 (2022).
13. Allison, K. R., Brynildsen, M. P. & Collins, J. J. Metabolite-enabled eradication of bacterial persisters by aminoglycosides. *Nature* **473**, 216–220 (2011).
14. Stokes, J. M., Lopatkin, A. J., Lobritz, M. A. & Collins, J. J. Bacterial Metabolism and Antibiotic Efficacy. *Cell Metab.* **30**, 251–259 (2019).
15. Zheng, E. J., Stokes, J. M. & Collins, J. J. Eradicating Bacterial Persisters with Combinations of Strongly and Weakly Metabolism-Dependent Antibiotics. *Cell Chem. Biol.* **27**, 1544–1552.e3 (2020).
16. Lin, J., Akiba, M., Sahin, O. & Zhang, Q. CmeR Functions as a Transcriptional Repressor for the Multidrug Efflux Pump CmeABC in *Campylobacter jejuni*. *Antimicrob. Agents Chemother.* **49**, 1067–1075 (2005).
17. Veleba, M., Higgins, P. G., Gonzalez, G., Seifert, H. & Schneiders, T. Characterization of RarA, a Novel AraC Family Multidrug Resistance Regulator in *Klebsiella pneumoniae*. *Antimicrob. Agents Chemother.* **56**, 4450–4458 (2012).
18. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* **3**, 318–356 (1961).

19. Novichkov, P. S. *et al.* RegPrecise 3.0 – A resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* **14**, 745 (2013).
20. Liu, X. *et al.* Transcriptional Repressor PtvR Regulates Phenotypic Tolerance to Vancomycin in *Streptococcus pneumoniae*. *J. Bacteriol.* **199**, (2017).
21. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957 (2005).
22. Balleza, E. *et al.* Regulation by transcription factors in bacteria: beyond description. *FEMS Microbiol. Rev.* **33**, 133–151 (2009).
23. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
24. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genomics Bioinforma.* **2**, (2020).
25. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 (2012).
26. Reyes-Caballero, H. *et al.* The Metalloregulatory Zinc Site in *Streptococcus pneumoniae* AdcR, a Zinc-activated MarR Family Repressor. *J. Mol. Biol.* **403**, 197–216 (2010).
27. Zhu, Z. *et al.* Entropy of a bacterial stress response is a generalizable predictor for fitness and antibiotic sensitivity. *Nat. Commun.* **11**, (2020).
28. Wood, S. *et al.* A Pangenomic Perspective on the Emergence, Maintenance, and Predictability of Antibiotic Resistance. doi:10.1007/978-3-030-38281-0_8.
29. Margolin, A. A. *et al.* ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. *BMC Bioinformatics* **7**, S7 (2006).
30. de Matos Simoes, R. & Emmert-Streib, F. Bagging Statistical Network Inference from Large-Scale Gene Expression Data. *PLoS One* **7**, e33624 (2012).
31. Faith, J. J. *et al.* Large-Scale Mapping and Validation of *Escherichia coli* Transcriptional Regulation from a Compendium of Expression Profiles. *PLoS Biol.* **5**, e8 (2007).
32. Huynh-Thu, V. A., Irrthum, A., Wehenkel, L. & Geurts, P. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS One* **5**, e12776 (2010).
33. Moerman, T. *et al.* GRNBoost2 and Arboreto: efficient and scalable inference of gene regulatory networks. *Bioinformatics* **35**, 2159–2161 (2019).
34. Iglesias-Martinez, L. F., De Kegel, B. & Kolch, W. KBoost: a new method to infer gene regulatory networks from gene expression data. *Sci. Rep.* **11**, 15461 (2021).
35. Reverter, A. & Chan, E. K. F. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* **24**, 2491–2497 (2008).
36. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
37. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
38. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–89 (2010).
39. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).

40. Machanick, P. & Bailey, T. L. MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
41. Wu, K. *et al.* CpsR, a GntR family regulator, transcriptionally regulates capsular polysaccharide biosynthesis and governs bacterial virulence in *Streptococcus pneumoniae*. *Sci. Rep.* **6**, 29255 (2016).
42. Warrier, I. *et al.* The Transcriptional landscape of *Streptococcus pneumoniae* TIGR4 reveals a complex operon architecture and abundant riboregulation critical for growth and virulence. *PLOS Pathog.* **14**, e1007461 (2018).
43. Grant, C. E. & Bailey, T. L. XSTREME: Comprehensive motif analysis of biological sequence datasets. *bioRxiv* (2021) doi:10.1101/2021.09.02.458722.
44. Madan Babu, M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. doi:10.1093/nar/gkg210.
45. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, P10008 (2008).
46. de Jong, A., Kuipers, O. P. & Kok, J. FUNAGE-Pro: comprehensive web server for gene set enrichment analysis of prokaryotes. *Nucleic Acids Res.* **50**, (2022).
47. Peterson, S. N. *et al.* Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. *Mol. Microbiol.* **51**, 1051–1070 (2004).
48. Peterson, S., Cline, R. T., Tettelin, H., Sharov, V. & Morrison, D. A. Gene Expression Analysis of the *Streptococcus pneumoniae* Competence Regulons by Use of DNA Microarrays. *J. Bacteriol.* **182**, 6192–6202 (2000).
49. Piotrowski, A., Luo, P. & Morrison, D. A. Competence for Genetic Transformation in *Streptococcus pneumoniae*: Termination of Activity of the Alternative Sigma Factor ComX Is Independent of Proteolysis of ComX and ComW. *J. Bacteriol.* **191**, 3359–3366 (2009).
50. Luo, P., Li, H. & Morrison, D. A. ComX is a unique link between multiple quorum sensing outputs and competence in *Streptococcus pneumoniae*. *Mol. Microbiol.* **50**, 623–633 (2003).
51. Tovpeko, Y., Bai, J. & Morrison, D. A. Competence for Genetic Transformation in *Streptococcus pneumoniae*: Mutations in σ A Bypass the ComW Requirement for Late Gene Expression. *J. Bacteriol.* **198**, 2370–2378 (2016).
52. Inmisi, N. L., Prehna, G. & Morrison, D. A. The pneumococcal σX activator, ComW, is a DNA-binding protein critical for natural transformation. *J. Biol. Chem.* **294**, 11101–11118 (2019).
53. Wu, K. *et al.* CpsR, a GntR family regulator, transcriptionally regulates capsular polysaccharide biosynthesis and governs bacterial virulence in *Streptococcus pneumoniae*. (2016) doi:10.1038/srep29255.
54. Gómez-Mejía, A., Gámez, G. & Hammerschmidt, S. *Streptococcus pneumoniae* two-component regulatory systems: The interplay of the pneumococcus with its environment. *Int. J. Med. Microbiol.* **308**, (2018).
55. Pettersen, J. S., Nielsen, F. D., Andreassen, P. R., Møller-Jensen, J. & Jørgensen, M. G. A comprehensive analysis of pneumococcal two-component system regulatory networks. *NAR Genomics Bioinforma.* **6**, (2024).
56. Bortoni, M. E., Terra, V. S., Hinds, J., Andrew, P. W. & Yesilkaya, H. The pneumococcal response to oxidative stress includes a role for Rgg. *Microbiology* **155**, 4123–4134 (2009).
57. Andreassen, P. R. *et al.* Host-glycan metabolism is regulated by a species-conserved two-component system in *Streptococcus pneumoniae*. *PLOS Pathog.* **16**, e1008332 (2020).

58. Sidote, D. J., Barbieri, C. M., Wu, T. & Stock, A. M. Structure of the *Staphylococcus aureus* AgrA LytTR Domain Bound to DNA Reveals a Beta Fold with an Unusual Mode of Binding. *Structure* **16**, 727–735 (2008).
59. Aggarwal, S. D., Yesilkaya, H., Dawid, S. & Hiller, N. L. The pneumococcal social network. *PLOS Pathog.* **16**, e1008931 (2020).
60. Long, X., Wang, X., Mao, D., Wu, W. & Luo, Y. A Novel XRE-Type Regulator Mediates Phage Lytic Development and Multiple Host Metabolic Processes in *Pseudomonas aeruginosa*. *Microbiol. Spectr.* **10**, (2022).
61. Rojas-Tapias, D. F. & Helmann, J. D. Roles and regulation of Spx family transcription factors in *Bacillus subtilis* and related species. in 279–323 (2019). doi:10.1016/bs.ampbs.2019.05.003.
62. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **130**, 797–810 (2007).
63. Lempp, M. *et al.* Systematic identification of metabolites controlling gene expression in *E. coli*. doi:10.1038/s41467-019-12474-1.
64. Ulijasz, A. T., Andes, D. R., Glasner, J. D. & Weisblum, B. Regulation of Iron Transport in *Streptococcus pneumoniae* by RitR, an Orphan Response Regulator. *J. Bacteriol.* **186**, 8123–8136 (2004).
65. Glanville, D. G. *et al.* RitR is an archetype for a novel family of redox sensors in the streptococci that has evolved from two-component response regulators and is required for pneumococcal colonization. *PLOS Pathog.* **14**, e1007052 (2018).
66. Beggs, G. A., Brennan, R. G. & Arshad, M. MarR family proteins are important regulators of clinically relevant antibiotic resistance. *Protein Sci.* **29**, 647–653 (2020).
67. Duval. MarA, SoxS and Rob of *Escherichia coli* – Global Regulators of Multidrug Resistance, Virulence and Stress Response. *Int. J. Biotechnol. Wellness Ind.* (2013) doi:10.6000/1927-3037.2013.02.03.2.
68. Hava, D. L., Hemsley, C. J. & Camilli, A. Transcriptional Regulation in the *Streptococcus pneumoniae* rlrA Pathogenicity Islet by RlrA. *J. Bacteriol.* **185**, 413–421 (2003).
69. Potter, A. J., Kidd, S. P., McEwan, A. G. & Paton, J. C. The MerR/NmlR Family Transcription Factor of *Streptococcus pneumoniae* Responds to Carbonyl Stress and Modulates Hydrogen Peroxide Production. *J. Bacteriol.* **192**, 4063–4066 (2010).
70. Hendriksen, W. T. *et al.* Site-Specific Contributions of Glutamine-Dependent Regulator GlnR and GlnR-Regulated Genes to Virulence of *Streptococcus pneumoniae*. *Infect. Immun.* **76**, 1230–1238 (2008).
71. Al-Bayati, F. A. Y. *et al.* Pneumococcal galactose catabolism is controlled by multiple regulators acting on pyruvate formate lyase. *Sci. Rep.* **7**, 43587 (2017).
72. El Khoury, J. Y., Boucher, N., Bergeron, M. G., Leprohon, P. & Ouellette, M. Penicillin induces alterations in glutamine metabolism in *Streptococcus pneumoniae*. *Sci. Rep.* **7**, 14587 (2017).
73. Kim, S.-W., Bae, Y.-G., Pyo, S.-N. & Rhee, D.-K. Differential Regulation of the Genes of the *Streptococcus pneumoniae* dnaK Operon by Ca⁺⁺. *Mol. Cells* **23**, 239–245 (2007).
74. Kwon, H.-Y., Kim, E.-H., Tran, T. D. H., Pyo, S.-N. & Rhee, D.-K. Reduction-Sensitive and Cysteine Residue-Mediated *Streptococcus pneumoniae* HrcA Oligomerization In Vitro. *Mol. Cells* **27**, 149–157 (2009).
75. Dagkessamanskaia, A. *et al.* Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of ciaR mutant cells.

- Mol. Microbiol.* **51**, 1071–1086 (2004).
76. Kjos, M. *et al.* Expression of *Streptococcus pneumoniae* Bacteriocins Is Induced by Antibiotics via Regulatory Interplay with the Competence System. *PLOS Pathog.* **12**, e1005422 (2016).
 77. Proutière, A. *et al.* Characterization of a Four-Component Regulatory System Controlling Bacteriocin Production in *Streptococcus gallolyticus*. *MBio* **12**, (2021).
 78. Kjos, M. *et al.* Expression of *Streptococcus pneumoniae* Bacteriocins Is Induced by Antibiotics via Regulatory Interplay with the Competence System. *PLoS Pathog.* **12**, (2016).
 79. Sastry, A. V. *et al.* The *Escherichia coli* transcriptome mostly consists of independently regulated modules. *Nat. Commun.* **10**, 5536 (2019).
 80. Liu, M., Prakash, C., Nauta, A., Siezen, R. J. & Francke, C. Computational Analysis of Cysteine and Methionine Metabolism and Its Regulation in Dairy Starter and Related Bacteria. *J. Bacteriol.* **194**, 3522–3533 (2012).
 81. Fritsch, V. N. *et al.* The MerR-family regulator NmlR is involved in the defense against oxidative stress in *Streptococcus pneumoniae*. *Mol. Microbiol.* **119**, 191–207 (2023).
 82. Greenberg, J. T., Monach, P., Chou, J. H., Josephy, P. D. & Demple, B. Positive control of a global antioxidant defense regulon activated by superoxide-generating agents in *Escherichia coli*. *Proc. Natl. Acad. Sci.* **87**, 6181–6185 (1990).
 83. Cuthbertson, L. & Nodwell, J. R. The TetR Family of Regulators. (2013) doi:10.1128/MMBR.00018-13.
 84. Ibrahim, Y. M., Kerr, A. R., McCluskey, J. & Mitchell, T. J. Control of Virulence by the Two-Component System CiaR/H Is Mediated via HtrA, a Major Virulence Factor of *Streptococcus pneumoniae*. *J. Bacteriol.* **186**, 5258–5266 (2004).
 85. Fraikin, N., Goormaghtigh, F. & Van Melderen, L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *J. Bacteriol.* **202**, (2020).
 86. Lehnher, H., Maguin, E., Jafri, S. & Yarmolinsky, M. B. Plasmid Addiction Genes of Bacteriophage P1: doc, which Causes Cell Death on Curing of Prophage, and phd, which Prevents Host Death when Prophage is Retained. *J. Mol. Biol.* **233**, 414–428 (1993).
 87. Schulz, C. *et al.* Regulation of the Arginine Deiminase System by ArgR2 Interferes with Arginine Metabolism and Fitness of *Streptococcus pneumoniae*. *MBio* **5**, (2014).
 88. Kloosterman, T. G. & Kuipers, O. P. Regulation of Arginine Acquisition and Virulence Gene Expression in the Human Pathogen *Streptococcus pneumoniae* by Transcription Regulators ArgR1 and AhrC. *J. Biol. Chem.* **286**, 44594–44605 (2011).
 89. Kovaleva, G. Y. & Gelfand, M. S. Transcriptional regulation of the methionine and cysteine transport and metabolism in streptococci. *FEMS Microbiol. Lett.* **276**, 207–215 (2007).
 90. Hu, Y. *et al.* The XRE Family Transcriptional Regulator SrtR in *Streptococcus suis* Is Involved in Oxidant Tolerance and Virulence. *Front. Cell. Infect. Microbiol.* **8**, (2019).
 91. Chan, W. T. & Espinosa, M. The *Streptococcus pneumoniae* pezAT Toxin–Antitoxin System Reduces β-Lactam Resistance and Genetic Competence. *Front. Microbiol.* **7**, (2016).
 92. Cuevas, R. A. *et al.* A novel streptococcal cell–cell communication peptide promotes pneumococcal virulence and biofilm formation. *Mol. Microbiol.* **105**, 554–571 (2017).
 93. Solano-Collado, V. *et al.* Recognition of Streptococcal Promoters by the Pneumococcal SigA Protein. *Front. Mol. Biosci.* **8**, (2021).
 94. Shimada, T., Yamazaki, Y., Tanaka, K. & Ishihama, A. The Whole Set of Constitutive Promoters

Recognized by RNA Polymerase RpoD Holoenzyme of Escherichia coli. *PLoS One* **9**, e90447 (2014).

95. Franza, T. *et al.* NAD⁺ pool depletion as a signal for the Rex regulon involved in Streptococcus agalactiae virulence. *PLOS Pathog.* **17**, e1009791 (2021).
96. Sharkey, L. K. R., Edwards, T. A. & O'Neill, A. J. ABC-F Proteins Mediate Antibiotic Resistance through Ribosomal Protection. *MBio* **7**, (2016).
97. Afzal, M., Shafeeq, S. & Kuipers, O. P. LacR Is a Repressor of lacABCD and LacT Is an Activator of lacTFEG , Constituting the lac Gene Cluster in Streptococcus pneumoniae. *Appl. Environ. Microbiol.* **80**, 5349–5358 (2014).
98. Schaumburg, C. S. & Tan, M. Arginine-Dependent Gene Regulation via the ArgR Repressor Is Species Specific in Chlamydia. *J. Bacteriol.* **188**, 919–927 (2006).
99. Ng, W.-L., Tsui, H.-C. T. & Winkler, M. E. Regulation of the pspA Virulence Factor and Essential pcsB Murein Biosynthetic Genes by the Phosphorylated VicR (YycF) Response Regulator in Streptococcus pneumoniae. *J. Bacteriol.* **187**, 7444–7459 (2005).
100. Afzal, M., Shafeeq, S., Ahmed, H. & Kuipers, O. P. Sialic Acid-Mediated Gene Expression in Streptococcus pneumoniae and Role of NanR as a Transcriptional Activator of the nan Gene Cluster. *Appl. Environ. Microbiol.* **81**, 3121–3131 (2015).
101. Horne, C. R. *et al.* Mechanism of NanR gene repression and allosteric induction of bacterial sialic acid metabolism. *Nat. Commun.* **12**, 1988 (2021).
102. El Qaidi, S., Yang, J., Zhang, J.-R., Metzger, D. W. & Bai, G. The Vitamin B 6 Biosynthesis Pathway in Streptococcus pneumoniae Is Controlled by Pyridoxal 5'-Phosphate and the Transcription Factor PdxR and Has an Impact on Ear Infection. *J. Bacteriol.* **195**, 2187–2196 (2013).
103. Yang, J., Tauschek, M. & Robins-Browne, R. M. Control of bacterial virulence by AraC-like regulators that respond to chemical signals. *Trends Microbiol.* **19**, 128–135 (2011).
104. Afzal, M., Shafeeq, S., Manzoor, I., Henriques-Normark, B. & Kuipers, O. P. N-acetylglucosamine-Mediated Expression of nagA and nagB in Streptococcus pneumoniae. *Front. Cell. Infect. Microbiol.* **6**, (2016).
105. Gaugué, I., Oberto, J. & Plumbridge, J. Regulation of amino sugar utilization in B acillus subtilis by the GntR family regulators, NagR and GamR. *Mol. Microbiol.* **92**, 100–115 (2014).
106. Brinsmade, S. R., Kleijn, R. J., Sauer, U. & Sonenshein, A. L. Regulation of CodY Activity through Modulation of Intracellular Branched-Chain Amino Acid Pools. *J. Bacteriol.* **192**, 6357–6368 (2010).
107. Hajaj, B. *et al.* CodY Regulates Thiol Peroxidase Expression as Part of the Pneumococcal Defense Mechanism against H₂O₂ Stress. *Front. Cell. Infect. Microbiol.* **7**, (2017).
108. Caymaris, S. *et al.* The global nutritional regulator CodY is an essential protein in the human pathogen Streptococcus pneumoniae. *Mol. Microbiol.* **78**, 344–360 (2010).
109. Zhu, J. *et al.* Co-regulation of CodY and (p)ppGpp synthetases on morphology and pathogenesis of Streptococcus suis. *Microbiol. Res.* **223–225**, 88–98 (2019).
110. Geiger, T. & Wolz, C. Intersection of the stringent response and the CodY regulon in low GC Gram-positive bacteria. *Int. J. Med. Microbiol.* **304**, 150–155 (2014).
111. Yu, W.-L. *et al.* TCS01 Two-Component System Influenced the Virulence of Streptococcus pneumoniae by Regulating PcpA. *Infect. Immun.* **91**, (2023).

112. Diagne, A. M. *et al.* Identification of a two-component regulatory system involved in antimicrobial peptide resistance in *Streptococcus pneumoniae*. *PLOS Pathog.* **18**, e1010458 (2022).
113. Zhang, Y., Martin, J. E., Edmonds, K. A., Winkler, M. E. & Giedroc, D. P. SifR is an Rrf2-family quinone sensor associated with catechol iron uptake in *Streptococcus pneumoniae* D39. *J. Biol. Chem.* **298**, 102046 (2022).
114. Kloosterman, T. G., Witwicki, R. M., Van Der Kooi-Pol, M. M., Bijlsma, J. J. E. & Kuipers, O. P. Opposite Effects of Mn²⁺ and Zn²⁺ on PsaR-Mediated Expression of the Virulence Genes pcpA, prtA, and psaBCA of *Streptococcus pneumoniae*. *J. Bacteriol.* **190**, 66 (2008).
115. Martin, J. E. *et al.* A Mn-sensing riboswitch activates expression of a Mn²⁺/Ca²⁺ ATPase transporter in *Streptococcus*. *Nucleic Acids Res.* **47**, 6885–6899 (2019).
116. Hendriksen, W. T. *et al.* Strain-specific impact of PsaR of *Streptococcus pneumoniae* on global gene expression and virulence. *Microbiology* **155**, 1569–1579 (2009).
117. Manzoor, I., Shafeeq, S. & Kuipers, O. P. Ni²⁺-Dependent and PsaR-Mediated Regulation of the Virulence Genes pcpA, psaBCA, and prtA in *Streptococcus pneumoniae*. *PLoS One* **10**, e0142839 (2015).
118. Rozman Grinberg, I. *et al.* A nucleotide-sensing oligomerization mechanism that controls NrdR-dependent transcription of ribonucleotide reductases. *Nat. Commun.* **13**, 2700 (2022).
119. Rigali, S., Derouaux, A., Giannotta, F. & Dusart, J. Subdivision of the Helix-Turn-Helix GntR Family of Bacterial Regulators in the FadR, HutC, MocR, and YtrA Subfamilies. *J. Biol. Chem.* **277**, 12507–12515 (2002).
120. Iyer, R. & Camilli, A. Sucrose metabolism contributes to in vivo fitness of *Streptococcus pneumoniae*. *Mol. Microbiol.* **66**, 1–13 (2007).
121. Ravcheev, D. A. *et al.* Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front. Microbiol.* **5**, (2014).
122. Kaatz, G. W., DeMarco, C. E. & Seo, S. M. MepR, a Repressor of the *Staphylococcus aureus* MATE Family Multidrug Efflux Pump MepA, Is a Substrate-Responsive Regulatory Protein. *Antimicrob. Agents Chemother.* **50**, 1276–1281 (2006).
123. Kreikemeyer, B. *et al.* Genomic organization, structure, regulation and pathogenic role of pilus constituents in major pathogenic Streptococci and Enterococci. *Int. J. Med. Microbiol.* **301**, 240–251 (2011).
124. Hemsley, C., Joyce, E., Hava, D. L., Kawale, A. & Camilli, A. MgrA, an Orthologue of Mga, Acts as a Transcriptional Repressor of the Genes within the rlrA Pathogenicity Islet in *Streptococcus pneumoniae*. *J. Bacteriol.* **185**, 6640–6647 (2003).
125. Mclean, K. T. *et al.* Site-Specific Mutations of GalR Affect Galactose Metabolism in *Streptococcus pneumoniae*. (2020) doi:10.1128/JB.00180.
126. Ravcheev, D. A. *et al.* Comparative genomics and evolution of regulons of the LacI-family transcription factors. (2014) doi:10.3389/fmicb.2014.00294.
127. Kloosterman, T. G., van der Kooi-Pol, M. M., Bijlsma, J. J. E. & Kuipers, O. P. The novel transcriptional regulator SczA mediates protection against Zn²⁺ stress by activation of the Zn²⁺-resistance gene czcD in *Streptococcus pneumoniae*. *Mol. Microbiol.* **65**, (2007).
128. Checroun, C. & Gutierrez, C. Dependent regulation of yehZYXW, which encodes a putative osmoprotectant ABC transporter of *Escherichia coli*. *FEMS Microbiol. Lett.* **236**, 221–226 (2004).
129. Baker, J. L. *et al.* Characterization of the Trehalose Utilization Operon in *Streptococcus mutans* Reveals that the TreR Transcriptional Regulator Is Involved in Stress Response Pathways and

Toxin Production. *J. Bacteriol.* **200**, (2018).

130. Nakata, M., Podbielski, A. & Kreikemeyer, B. MsmR, a specific positive regulator of the *Streptococcus pyogenes* FCT pathogenicity region and cytolsin-mediated translocation system genes. *Mol. Microbiol.* **57**, 786–803 (2005).
131. Zhao, D. *et al.* MsmR1, a global transcription factor, regulates polymyxin synthesis and carbohydrate metabolism in *Paenibacillus polymyxa* SC2. *Front. Microbiol.* **13**, (2022).
132. Henke, S. K. & Cronan, J. E. The *Staphylococcus aureus* group II biotin protein ligase BirA is an effective regulator of biotin operon transcription and requires the DNA binding domain for full enzymatic activity. *Mol. Microbiol.* **102**, (2016).
133. Hu, D., Laczkovich, I., Federle, M. J. & Morrison, D. A. Identification and Characterization of Negative Regulators of Rgg1518 Quorum Sensing in *Streptococcus pneumoniae*. *J. Bacteriol.* **205**, e0008723 (2023).
134. Fleuchot, B. *et al.* Rgg proteins associated with internalized small hydrophobic peptides: a new quorum-sensing mechanism in streptococci. *Mol. Microbiol.* **80**, 1102–1119 (2011).
135. Choi, E., Jeon, H., Oh, C. & Hwang, J. Elucidation of a Novel Role of YebC in Surface Polysaccharides Regulation of *Escherichia coli* bipA-Deletion. *Front. Microbiol.* **11**, (2020).
136. Grenha, R. *et al.* Structural basis for the activation mechanism of the PlcR virulence regulator by the quorum-sensing signal peptide PapR. *Proc. Natl. Acad. Sci.* **110**, 1047–1052 (2013).
137. Weng, M., Nagy, P. L. & Zalkin, H. Identification of the *Bacillus subtilis* pur operon repressor. *Proc. Natl. Acad. Sci.* **92**, 7455–7459 (1995).
138. Huang, C. Y., Gonzalez-Lopez, C., Henry, C., Mijakovic, I. & Ryan, K. R. hipBA toxin-antitoxin systems mediate persistence in *Caulobacter crescentus*. *Sci. Rep.* **10**, 2865 (2020).
139. Singh, K. D., Schmalisch, M. H., Stölke, J. & Görke, B. Carbon Catabolite Repression in *Bacillus subtilis* : Quantitative Analysis of Repression Exerted by Different Carbon Sources. *J. Bacteriol.* **190**, 7275–7284 (2008).
140. Debroy, S. *et al.* A Multi-Serotype Approach Clarifies the Catabolite Control Protein A Regulon in the Major Human Pathogen Group A *Streptococcus* OPEN. (2016) doi:10.1038/srep32442.
141. Zhang, Y. *et al.* comCDE (Competence) Operon Is Regulated by CcpA in *Streptococcus pneumoniae* D39. *Microbiol. Spectr.* (2023) doi:10.1128/spectrum.00012-23.
142. Shafeeq, S., Kuipers, O. P. & Kloosterman, T. G. Cellobiose-Mediated Gene Expression in *Streptococcus pneumoniae*: A Repressor Function of the Novel GntR-Type Regulator BguR. *PLoS One* **8**, e57586 (2013).
143. McCluskey, J., Hinds, J., Husain, S., Witney, A. & Mitchell, T. J. A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in *Streptococcus pneumoniae*. *Mol. Microbiol.* **51**, 1661–1675 (2004).
144. Loisel, E. *et al.* AdcAII, A New Pneumococcal Zn-Binding Protein Homologous with ABC Transporters: Biochemical and Structural Analysis. *J. Mol. Biol.* **381**, 594–606 (2008).
145. Tobisch, S., Stölke, J. & Hecker, M. Regulation of the lic Operon of *Bacillus subtilis* and Characterization of Potential Phosphorylation Sites of the LicR Regulator Protein by Site-Directed Mutagenesis. *J. Bacteriol.* **181**, 4995–5003 (1999).
146. Chastanet, A., Prudhomme, M., Claverys, J.-P. & Msadek, T. Regulation of *Streptococcus pneumoniae* clp Genes and Their Role in Competence Development and Stress Survival. *J. Bacteriol.* **183**, 7295–7307 (2001).

147. Johnston, C. H. *et al.* The alternative sigma factor σX mediates competence shut-off at the cell pole in *Streptococcus pneumoniae*. *Elife* **9**, (2020).
148. Pestova, E. V., Håvarstein, L. S. & Morrison, D. A. Regulation of competence for genetic transformation in *Streptococcus pneumoniae* by an auto-induced peptide pheromone and a two-component regulatory system. *Mol. Microbiol.* **21**, 853–862 (1996).
149. Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinforma. Adv.* **2**, (2022).
150. Eldholm, V. *et al.* The Pneumococcal Cell Envelope Stress-Sensing System LiaFSR Is Activated by Murein Hydrolases and Lipid II-Interacting Antibiotics. *J. Bacteriol.* **192**, 1761–1773 (2010).
151. Bidossi, A. *et al.* A Functional Genomics Approach to Establish the Complement of Carbohydrate Transporters in *Streptococcus pneumoniae*. *PLoS One* **7**, e33320 (2012).
152. Hoover, S. E. *et al.* A new quorum-sensing system (TprA/PhrA) for *S. pneumoniae* D39 that regulates a lantibiotic biosynthesis gene cluster. *Mol. Microbiol.* **97**, 229–243 (2015).
153. Laczkovich, I. *et al.* Discovery of Unannotated Small Open Reading Frames in *Streptococcus pneumoniae* D39 Involved in Quorum Sensing and Virulence Using Ribosome Profiling. *MBio* **13**, (2022).
154. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

Chapter 4:

Synthetic Perturbation of the Transcriptional
Regulatory Network of TIGR4 Identifies
Routes to Antibiotic Potentiation

4.0 Introduction

Antibiotic resistance is often thought to be a binary classification; resistant or susceptible. This comes from our evolving understanding of this evolutionary outcome, where the first examples of antibiotic-resistant bacteria could be classified by the presence or absence of specific resistance genes^{1–3}. Today, we understand bacteria as complex, fluid systems where stresses such as antibiotics are sensed, absorbed and responded to by an array of interconnected networks. Therefore, the state of this physiological array at the time of antibiotic treatment will influence its outcome. Antibiotic persisters, a subpopulation of cells that drastically reduce their metabolic output to survive stressful conditions, are an extreme example of this. This state has been shown to be induced by various stressors and, although not fully understood, can be achieved through a combination of transcriptional⁴, translational^{5,6} and proteolytic^{6,7} mechanisms. Persister cells overexpress genes critical to surviving the current stress and have an increased likelihood of developing antibiotic resistance through genetic mutation or horizontal gene transfer^{8,9}. Alarmingly, persister cells initially are genetically identical to antibiotic susceptible cells, demonstrating that within every bacterial cell there exists the potential to survive antibiotic treatment and develop resistance. Understanding which physiological states are most influential on antibiotic efficacies, and how these states come to be, is critical to improving antibiotic treatments and combatting antibiotic resistance.

Our library of transcription factor overexpression strains gives us the unique ability to assess the influence that the transcriptional regulatory network has over antibiotic outcomes. First, identifying phenotypically important transcriptional responses to

antibiotics is fundamental to understanding how bacteria respond to antibiotic stress. Previously it has been shown that antibiotic stress leads to transcriptional entropy (disorder) within bacteria, which obscures transcriptional regulation that is beneficial to the cell from the antibiotic-induced chaos¹⁰. Eventually, understanding which regulatory events impact antibiotic outcomes opens the possibility of targeting the stress response to potentiate antibiotics. Second, by synthetically stimulating each regulon through transcription factor overexpression prior to antibiotic treatment, we can understand which cell states are most and least susceptible to antibiotics. This would allow for future treatment plans aimed at predisposing bacterial cells to antibiotic treatment, such as metabolic stimulation of persister cells to induce growth^{11,12}. An understanding of how to predispose bacterial cells to antibiotic killing and how to inhibit their stress response is critical to rationally design antibiotic potentiators.

To determine the fitness effect of overexpressing each transcription factor in *S. pneumoniae* strain TIGR4, we completed several Transcriptional Regulator Induced Phenotype (TRIP) screens¹³. Briefly, all transcription factor induction (TFI) strains are combined into a TRIP pool, induced with IPTG and cultured across varying conditions including antibiotic treatment. The relative fitness effect of overexpressing each transcription factor is calculated by tracking each strain's relative abundance across time through barcode sequencing. Strains that increase in relative abundance throughout the experiment have relatively greater fitness than the rest of the pool and indicate that overexpression of that transcription factor is beneficial to the cell under the tested condition (Fig. 1A, 1B). Strains with significant fitness effects have relative fitness values that are greater than two standard deviations from the average of the relative fitness values of all TFI strains (Fig. 1C), as

well as an adjusted p-value of less than 0.05. To visualize specific fitness effects to the condition tested (i.e. Kanamycin treatment), TRIP results are compared between the tested condition and the baseline media (Fig. 1D). TRIP screens are a high throughput, transcriptional regulatory network-based approach that allow for evaluation of transcription factor activity in diverse cell states and can capture emergent phenotypes from differential regulation of entire regulons. They also complement mutagenesis-based screens well in that we assay the differential regulation of all genes, including essential genes, rather than gene knockouts. The totality of TRIP screen results will help mechanistically describe transcriptional regulation involved in *S. pneumoniae*'s stress response and identify routes towards antibiotic potentiation.

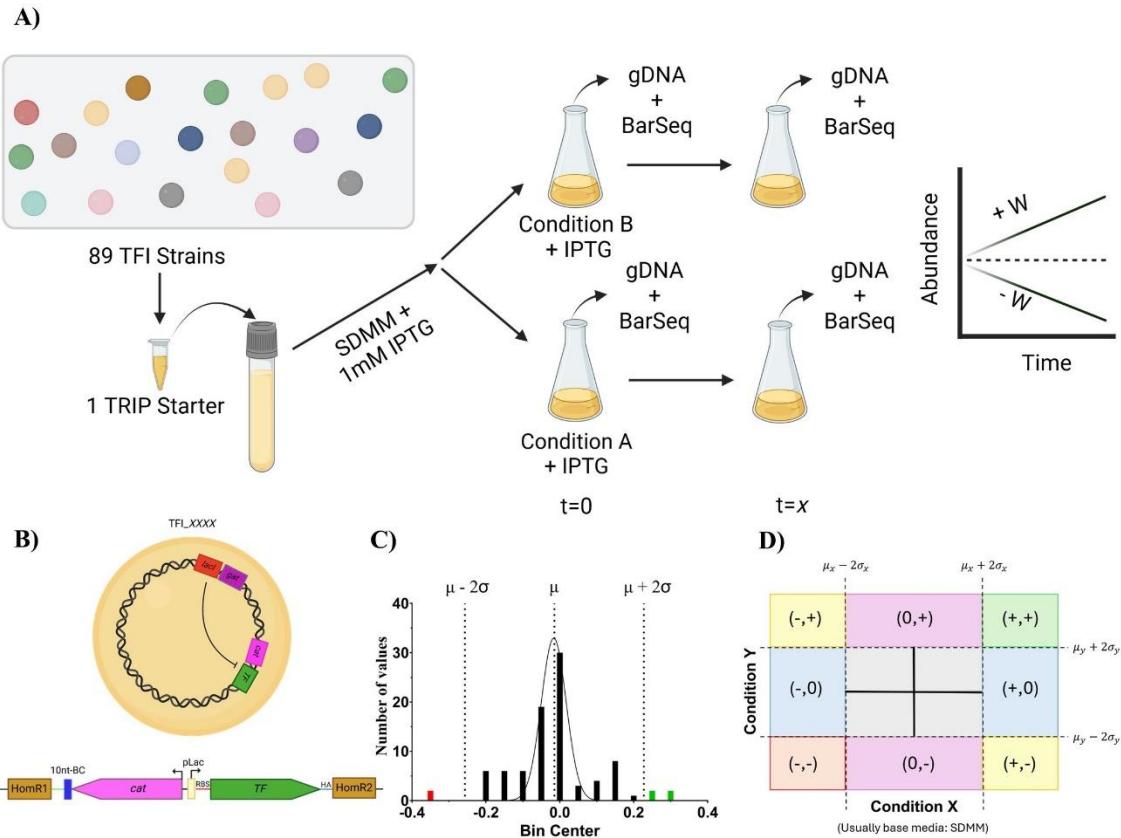


Figure 1. Overview of TRIP Screens. **A)** Schematic representation of TRIP screens. **B)** Diagram of TFI strain organization. **C)** Histogram of relative fitness results for a TRIP screen. Colored bars represent significant results. **D)** Organization of TRIP Specific Fitness plots which plot the average fitness ($\text{Log}_2(\text{Fold-Change in Abundance})$) of all TFI strains across three biological replicates. Parenthesis represent significant fitness results in (x,y) conditions. $W = \text{Relative Fitness}$. $\mu = \text{Mean}$. $\sigma = \text{Standard Deviation}$.

4.1 Metabolic TRIP Screens

4.1.1 Metabolic and Oxidative Stress Response Regulators Influence Fitness in Plain Media

We first completed TRIP screens in plain SDMM with glucose as the carbon source to identify the baseline fitness effect of overexpressing each transcription factor in rich media. Several TFI strains have significant fitness defects in plain media (Fig. 2A, 2B). TFI_2112, TFI_0893 and TFI_0246 all are significantly less fit in plain media and overexpress metabolic regulators. TFI_2112 overexpresses the maltose operon repressor MalR, which

activates expression of several genes involved in maltose catabolism (*dexB*, *malPQX*), suggesting these cells may be wasting energy while glucose is present. *malR* itself is under carbon catabolite repression, which indicates that *S. pneumoniae* has evolved to repress its regulon when glucose is present. TFI_0893 overexpresses the arginine deiminase system (ADS) activator ArgR2, which activates import and metabolism of arginine to produce ATP, ornithine and ammonia¹⁴. This system is under catabolite repression though, so overexpression of ArgR2 did not affect expression of the ADS system in SDMM, but did result in the second most Padj genes of all TFI strains with. The effector molecule that regulates expression of ArgR2 activity is arginine itself, so in this scenario, slow growth may be due to “quenching” of arginine within the cell and system-wide metabolic changes¹⁴. TFI_0246 overexpresses a metabolic regulator DeoR. DeoR is a strong repressor of a glycyl-radical enzyme activating protein (SP_0245) as well as the transcription factor, *lsrR* (SP_0247). Glycyl-radical enzyme activating proteins are required to activate glycyl-radical enzymes within the cell that catalyze difficult metabolic reactions. While the target of SP_0245 remains uncharacterized¹⁵, genomic organization suggests it may be pyruvate formate lyase A (*pflD*, SP_0251). Interestingly, LsrR (SP_0247) represses expression of a key metabolic operon, encoding *gldA*, *fsa*, and the possible target of SP_0245, *pflD*. The fitness defect of TFI_0246 in base media therefore may also be due to indirect regulation of this metabolic operon via LsrR and the required glycyl-radical enzyme activating protein (SP_0245). Along with those metabolic regulators, TFI_1920 also has a significant fitness defect in plain media and overexpresses an unnamed MarR family transcription factor. This MarR transcription factor regulates

expression of a multidrug efflux pump (*mdlB*, *SP_1918* and *SP_1919*), suggesting its slow growth may result from activation of an export-based stress response.

Conversely, four TFI strains had significant growth advantages in plain media: TFI_0376, TFI_0501, TFI_1115 and TFI_1131 (Fig. 2A, 2B). TFI_0376 overexpresses the orphan response regulator, RitR, which regulates the most genes of any transcription factor in TIGR4 and is primarily involved in regulating metal transport, primarily iron, protease expression and the oxidative stress response^{16,17}. Similarly, TFI_1115 overexpresses RggD which only directly regulates an ABC-F type ribosomal protection factor, but overexpression of *rggD* in SDMM resulted in the most Padj genes of any TFI strain with 735. Differential regulation of this ribosomal protection factor could influence translational efficiency within the cell, resulting in whole-system changes. In *S. pneumoniae* strain D39, an *rggD* mutant was shown to alter both carbohydrate and amino acid metabolism, while also inhibiting the oxidative stress response and virulence¹⁸. Overexpression of both RitR and RggD may result in faster growth in plain SDMM due to a preemptive induction of the oxidative stress response, particularly proteostasis, which is critical in *S. pneumoniae* due to the large quantity of H₂O₂ it produces without encoding a catalase to detoxify. Similarly, TFI_0501 overexpresses GlnR, which regulates glutamine import and synthesis^{19,20}. Glutamine availability is critical to cells for nitrogen-dependent synthesis of both amino acids and nucleotides, both of which are required for fast growth. Also, glutamine is a source of ammonia in the cell, which serves as a buffer for acidic conditions that would arise in cells with increased metabolism. Taken together, overexpression of GlnR could ensure the cell has an abundance of glutamine for nitrogen-dependent metabolism and ammonia for buffering against the acidic conditions generated through increased

metabolism. Lastly, TFI_1131 overexpresses an uncharacterized XRE-family transcription factor, but *SP_1131* is naturally very lowly expressed and we were not able to identify any of its target genes. Monoculture growth curve validations of significant hits in plain media TRIP screens show very small differences in growth in plain media, indicating the fitness effect of overexpressing these transcription factors is minor in plain media, likely because these strains are well adapted to these conditions (Fig. 2C). The ability of our TRIP screens to quantify these apparently minor fitness emphasizes the sensitivity of barcode sequencing and our assay as a whole.

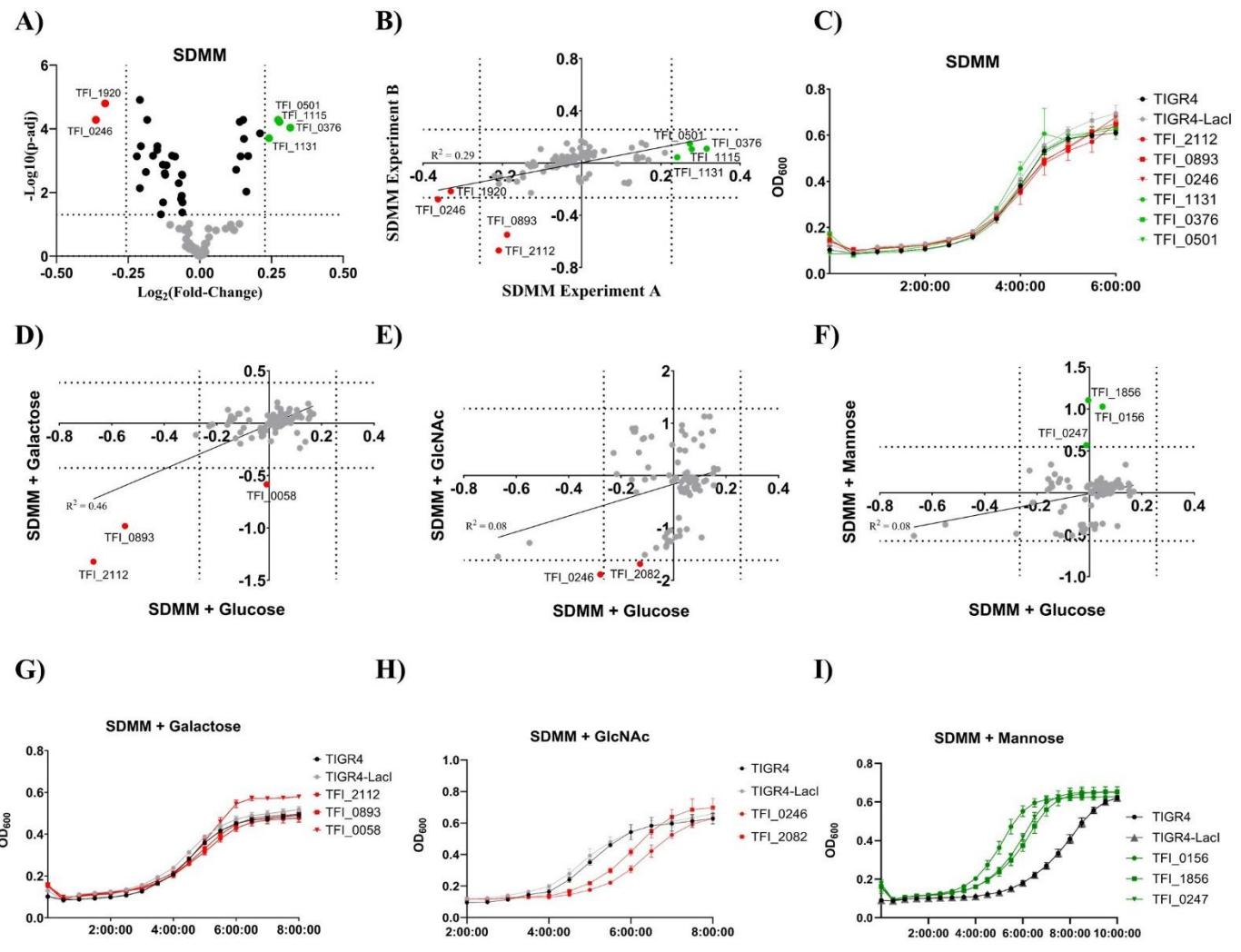


Figure 2. Metabolic TRIP Screen results and validations. **A)** Volcano plot of TRIP results in baseline SDMM. **B)** Specific fitness plots comparing two TRIP screens in baseline SDMM. **C)** Monoculture growth curve validation of TRIP results in SDMM. **D, E, F)** Specific fitness plots of TRIP screens in **D)** Galactose, **E)** N-acetylglucosamine (GlcNAc) and **F)** Mannose. **G, H, I)** Monoculture growth curve validations in SDMM + **G)** Galactose, **H)** GlcNAc, **I)** Mannose. For all specific fitness plots, the average fitness ($\text{Log}_2(\text{Fold-Change in Abundance})$) of each TFI strain across three biological replicates is shown. For all growth curves, the average of four replicates is shown, error bars represent standard error.

We next performed TRIP screens in alternate carbon sources to identify which regulons contribute to the metabolism of specific carbon sources. We hypothesized that these TRIP screens would identify TFI strains that overexpress regulators of key metabolic enzymes as being significantly more fit than other TFI strains. If this were the case, it would lend credence to the reliability of our TRIP screens as we'd be able to recapture true biology. We completed TRIP screens in three alternate carbon sources: galactose, mannose and N-acetylglucosamine (GlcNAc). These three carbon sources were chosen because they are all available within the host and *S. pneumoniae* is known to be able to catabolize²¹. Differential availability of glucose and galactose across host tissues, such as the nasopharynx, lung and blood, are known to influence *S. pneumoniae* metabolism and virulence²². Therefore, an understanding of the transcriptional drivers to metabolic adaptation may also help understand disease progression within the host.

4.1.2 Fitness effects of Transcription Factor Overexpression are obscured in TRIP screens due to Constitutive Expression of the Leloir Pathway

For TRIP screens in galactose, there were relatively little fitness effects beyond the baseline (Fig. 2D). Two TFI strains, TFI_2112 and TFI_0893, which overexpress MalR and ArgR2, respectively, showed significantly reduced fitness in SDMM + galactose, but also displayed significantly reduced fitness in SDMM + glucose. TFI_0058 overexpresses the global metabolic regulator CpsR and had a specific significant fitness defect in SDMM + galactose. CpsR, named after its role in repressing capsule expression, is known to be allosterically regulated by glucose²³. In terms of capsule regulation, the lack of glucose would lead to repression of capsule expression at its promoter, however, it is unclear how glucose availability influences regulation of its many other target genes. Like SDMM +

glucose, significant hits identified in SDMM + galactose TRIP screens show little effect in monoculture growth curves (Fig. 2G).

The lack of large fitness effects of transcription factor overexpression in TIGR4 in SDMM + galactose may be due to the two parallel metabolic pathways for galactose: the Tagatose-6-phosphate pathway and the Leloir pathway. The key metabolic operons within these pathways are the lactose gene cluster, which is split into two operons *lac operon I* (*lacABCD*, *SP_1190-1193*) and the *lac operon II* (*lacTFEG*, *SP_1187-1184*), and the galactose (*SP_1851-1854*) metabolic operon²⁴. *Lac operon I* is repressed by LacR (*SP_1182*) while *lac operon II* is activated by derepression of *lac operon I* and the subsequent activity of the antiterminator, LacT (*SP_1187*). *Lac operon I* contains the genes for the enzymes within the Tagatose-6-Phosphate pathway while the genes in *lac operon II* are primarily involved in lactose transport. The galactose metabolic operon encodes enzymes within the Leloir pathway and is regulated by GalR (*SP_1854*). In *S. pneumoniae*, none of these operons are regulated through carbon catabolite repression²⁴. Interestingly, neither TFI_1854 nor TFI_1182 have significant fitness effects in galactose. In fact, neither strain has any significant fitness effects in any TRIP screen completed. For GalR, we hypothesize that this is because IPTG appears to allosterically inhibit GalR::DNA binding, resulting in expression of the galactose metabolic operon (*SP_1851-1854*) and activation of the Leloir pathway in all TFI strains. Overexpression of LacR, however, did result in strong repression of *lac operon I*, which indicates LacR is not allosterically inhibited by IPTG like GalR and LacI in *E. coli*. The lack of fitness defect of TFI_1182 in galactose is likely due to an activated and parallel Leloir pathway.

4.1.3 Carbon Catabolite Repression limits Fitness Effects in GlcNAc

N-Acetylglucosamine (GlcNAc) is the most abundant monosaccharide in human mucus and is a critical reagent for synthesis of peptidoglycan in the cell wall²⁵. Catabolism of GlcNAc is dependent on the enzymatic activities of NagA and NagB, which catalyze the conversions of GlcNAc-6-P to GlcN-6-P and subsequently GlcN-6-P to F-6-P, respectively. Both *nagA* and *nagB* are transcriptionally regulated by NagR (*SP_1446*) and catabolite repression²⁶. Our initial hypothesis was that overexpression of *nagR* would result in significant fitness changes when cultured on GlcNAc, however overlapping regulation by CcpA via catabolite repression may minimize the actual effect *nagR*-overexpression has on the cell.

In TRIP screens cultured in GlcNAc, TFI_1446, which overexpresses NagR, had no significant fitness effect (Fig. 2E). This is likely due to catabolite repression, as regardless of NagR expression, whenever glucose is present *nagAB* will be repressed. As expected, overexpression of NagR in SDMM + IPTG with glucose resulted in no differential expression of *nagA* or *nagB*, presumably because these genes were under catabolite repression. Along with CcpA (*SP_1999*), NagB is regulated by several other global metabolic regulators, including CodY (*SP_1584*), RitR (*SP_0376*) and CpsR (*SP_0058*), indicating its importance in cellular metabolism, particularly peptidoglycan turnover. This multi-factor regulation could explain why overexpression of just one of its regulators failed to significantly alter the cells' phenotypes.

Two TFI strains had significantly negative fitness in GlcNAc: TFI_0246 and TFI_2082 (Fig. 2E). TFI_0246 overexpresses the uncharacterized glycol-radical activating enzyme repressor, DeoR, which had significantly negative fitness in SDMM + glucose as well (Fig.

2A, 2B), indicating this result may be a general fitness defect. TFI_2082 overexpresses the response regulator of TCS04, PnpR, and had significantly reduced fitness when catabolizing GlcNAc. PnpR is known to regulate manganese homeostasis through regulation of *psaABC*, but its link to GlcNAc metabolism is unclear²⁷. TFI_2082 has a noticeable but insignificant fitness defect in glucose as well, so it also may not be a specific fitness defect to GlcNAc catabolism. Monoculture growth curve validations of TFI_0246 and TFI_2082 show greater fitness defects relative to TIGR4-LacI than growth curve validations in galactose, likely due to a lack of an alternative and IPTG-induced GlcNAc metabolic pathway (Fig. 2H).

4.1.4 Overexpression of *manA* Increases Fitness when Catabolizing Mannose

There were three TFI strains with significant fitness advantages when catabolizing mannose: TFI_0156, TFI_1856 and TFI_0247 (Fig. 2F). The key reaction in the catabolism of mannose is the conversion of mannose-6-phosphate to fructose-6-phosphate (F-6-P) by ManA (*SP_0736*). According to the transcriptional regulatory network, *manA* is only regulated by the response regulator YesM (*SP_0156*) of the two-component system YesMN, which has been shown to be involved in host glycan metabolism^{28,29}. As expected, overexpression of YesM had significant, positive fitness that was specific to the presence of mannose. Interestingly, two other TFI strains had significantly positive fitness in mannose, TFI_0247, which overexpresses LsrR, and TFI_1856, which overexpresses NmlR. LsrR contains a putative sugar binding domain and is of the DeoR family of transcription factors which are known to regulate sugar metabolism^{30–32}. In TIGR4, LsrR regulates expression of a key metabolic operon encoding *gldA*, *fsa*, and *pflD*. Interestingly, this transcription factor lies directly downstream of DeoR (*SP_0246*) which has been

shown to be a strong repressor of LsrR and have significant fitness defects in SDMM + glucose (Fig. 2A, 2B) and SDMM + galactose (Fig. 2D). The only direct target of NmlR is the alcohol dehydrogenase, *adhC*, which is known to be involved in protection against redox stress through glutathione recycling but also can be important to growth on mannose through detoxification of toxic byproducts including reactive aldehydes and dicarbonyl compounds^{33,34}. Despite not being identified as direct regulators of *manA*, overexpression of both NmlR and LsrR resulted in upregulation of *manA* ($\text{Log}_2\text{FC} = 0.44$ and 0.65 , $P_{\text{adj}} = 0.0096$ and $1.31\text{E-}6$, respectively).

Overall, it appears that differential fitness of TFI strains catabolizing alternate carbon sources is dependent on the redundancy within the metabolic regulatory network and differential expression of specific metabolic enzymes caused by transcription factor overexpression in glucose. TRIP screens in galactose showed few significant hits because of the redundancy of the Tagatose-6-Phosphate and Leloir pathways and the constitutive expression of the Leloir pathway due to IPTG's effect on GalR. Catabolism of mannose and GlcNAc, on the other hand, are dependent on the expression of specific enzymes, ManA and NagAB, respectively. Three TFI strains, TFI_0156, TFI_0247 and TFI_1856, all have significant fitness advantages while catabolizing mannose because they all overexpress ManA in glucose, which the cells were cultured in prior to re-inoculation in mannose. This means these strains already have ManA present and able to immediately catabolize mannose, while the other strains must rewire their metabolic networks, likely through YesM (*SP_0156*), to express ManA and restart growth (akin to the diauxic growth pattern of bacteria cultured in multiple carbon sources). Ultimately, this supports our hypothesis that synthetic perturbations of the transcriptional regulatory network will

identify regulators, pathways and transcriptomes that influence antibiotic treatment outcomes.

4.2 Antibiotic Resistance TRIP Screens

After showing that synthetic perturbation of the transcriptional regulatory network of *S. pneumoniae* can alter fitness to alternate carbon sources in a rational manner, we turned to antibiotics to attempt to identify which phenotypic states the bacterial cells are most, and least, susceptible to antibiotics. Faster-growing cells are generally more susceptible to antibiotics, as their accelerated growth requires more output from the core processes targeted by the antibiotics. It has also been shown that all bactericidal antibiotics, regardless of target, result in accumulation of reactive oxygen species (ROS), leading to cell death³⁵. Therefore, we hypothesized that TRIP screens in a variety of antibiotic classes would identify trends in which regulons influence antibiotic susceptibility, and that for bactericidal antibiotics, these regulons would be enriched in regulons involved in the oxidative stress response.

4.2.1 Overexpression of Certain Metabolic Regulators Increases Resistance to Cell Wall Synthesis Inhibitors

Ceftriaxone and vancomycin are both cell wall synthesis inhibitors (CWSI), but they have distinct molecular targets and are in different antibiotic classes. Bacterial cell walls are matrices of peptidoglycan, which is a polymer consisting of a polysaccharide backbone of alternating N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (NAM) sugars, with pentapeptides attached to each NAM residue³⁶. Penicillin binding proteins (PBPs) recognize and bind the terminal amino acids on these pentapeptides, D-alanyl-D-alanine, and catalyze crosslinking between adjacent pentapeptides³⁷. Ceftriaxone is a

cephalosporin, a subfamily of beta-lactam antibiotics, which irreversibly bind and inhibit these PBPs to inhibit peptidoglycan crosslinking. Vancomycin, on the other hand, is a glycopeptide which binds the terminal D-alanyl-D-alanine residues on peptidoglycan and prevents crosslinking by PBPs via substrate sequestration³⁸.

TRIP screens in ceftriaxone and vancomycin identified several metabolic regulators that influence susceptibility to cell wall synthesis inhibitors (Fig. 3). As expected, there was a general trend where the TFI strains with positive fitness in plain SDMM had lower fitness in both CWSIs (Fig. 3A, 3B). Overexpression of three metabolic regulators, CpsR (*SP_0058*), SusR (*SP_1799*) and TreR (*SP_1885*) have significantly increased resistance to both ceftriaxone and vancomycin. These three strains were the only three strains in monoculture growth curve analysis to display the “Rapid Biphasic Lysis” phenotype in Chapter 2. CpsR, named for its role in repressing capsule expression, was previously identified as having significantly negative fitness when cultured with galactose as its sole carbon source (Fig. 2D). CpsR is known to repress capsule expression in the presence of glucose, so overexpression in SDMM resulted in no significant differential expression of the capsule operon²³. Interestingly, TFI RNAseq on both TFI_1799 and TFI_1885 revealed repression of the capsule operon, and cell pellets of these three TFI strains are noticeably smaller/more compact than other TFI strains. Along with repression of the capsule, overexpression of SusR and TreR resulted in the differential expression of several cell wall synthesis related genes (represents *wcaKA*, activates *dltB*, *pbp3* and *eno*) and activation of the oxidative stress response through *gor*, *trxA* and *pdrM* (Fig. 3E). Penicillin binding protein 3 (*pbp3*) is the only low molecular weight *pfp* in *S. pneumoniae* and has been shown to be beta-lactam-sensitive and should not influence ceftriaxone or vancomycin

sensitivity³⁹. Overexpression of SusR also resulted in repression of *murMN*, which catalyzes indirect branching of the cell wall and are required for high levels of penicillin resistance⁴⁰.

A possible connection between overexpression of these metabolic regulators and peptidoglycan structure is the acid stress produced by bacterial metabolism. Bacterial metabolism produces acidic by-products, and specific metabolic pathways generate more acidic stress than others³³. For example, in *Streptococcus mutans*, growth on trehalose resulted in less lactic acid production and higher pH than when grown on either sucrose or glucose⁴¹. Bacteria are known to alter both membrane⁴² and peptidoglycan structure⁴³ under acid stress. Interestingly, MurM has recently been implicated in regulation of the stringent response in *S. pneumoniae* by incorporating mischarged seryl-tRNA^{Ala} into the cell wall and decreasing the pool of mischarged tRNAs⁴⁴. A *murMN* mutant was shown to have growth defects and rapid lysis phenotype in acidic conditions, a phenotype that was rescued by a deletion of the autolysin, LytA⁴⁴. LytA activity is dependent on the arrest of cell wall synthesis, either through nutrient depletion, cell wall synthesis inhibitors or acidic stress⁴⁴⁻⁴⁶.

Together, TFI_0058, TFI_1799 and TFI_1885's phenotypes of rapid stationary phase lysis, increased resistance to cell wall synthesis inhibitors and gene expression profiles suggests a concomitant increase in acidic stress and decreased cell wall synthesis. Altered metabolism can increase cell acidity, and activation of several components of the oxidative stress response could be indicative of increased acidic stress. Both the rapid lysis and resistance to CWSIs could be due to repressed cell wall synthesis, as seen via repression

of *murMN* and *wcaKA*. It's unclear, however, if repression of cell wall synthesis is directly due to metabolic regulation or as a response to acidic stress. The only other strains with significant negative fitness in the two cell wall synthesis inhibitors tested were TFI_1131 and TFI_0246, strains that had significantly positive and negative fitness in plain SDMM, respectively and therefore may not be directly due to cell wall synthesis inhibition.

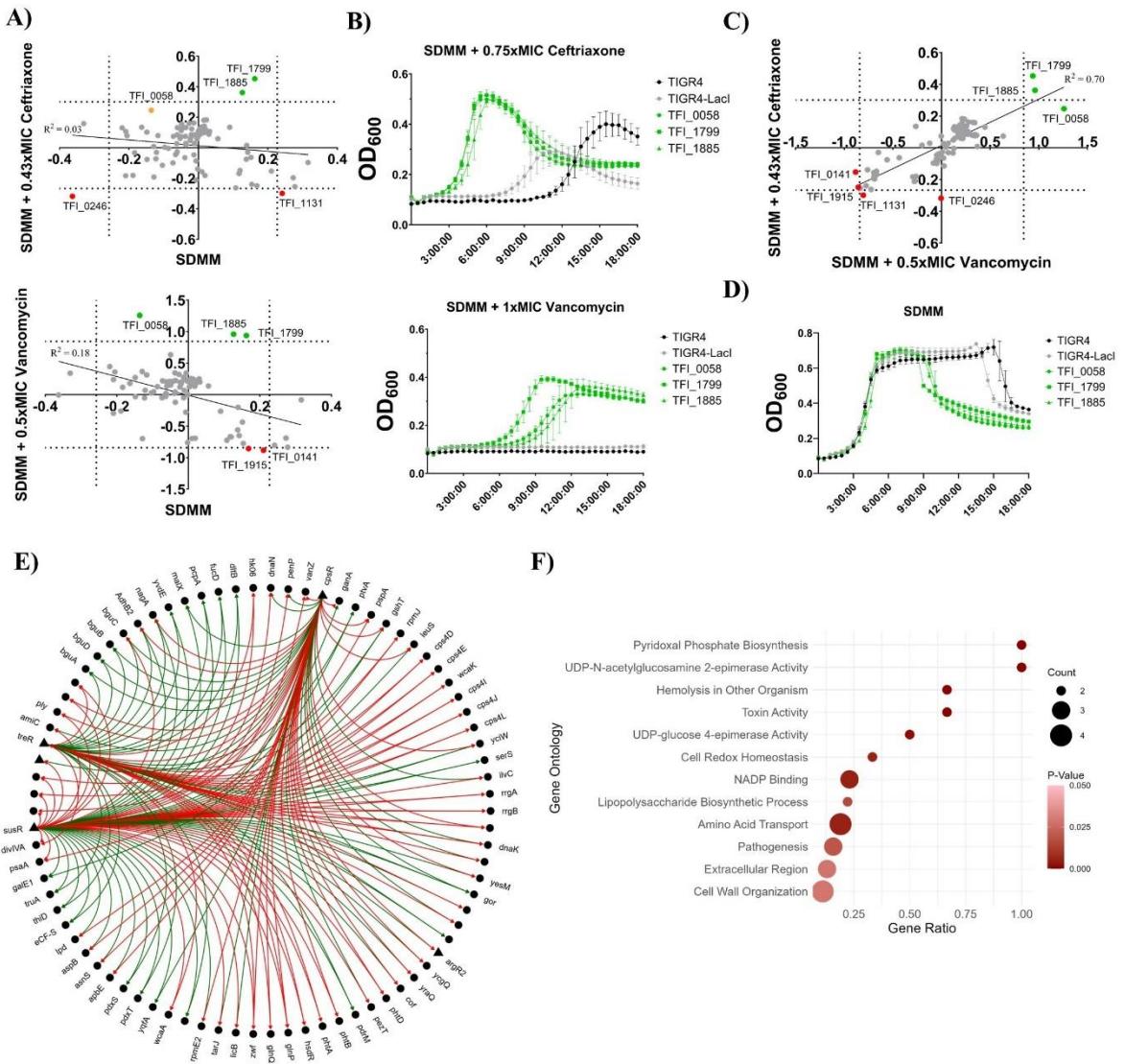


Figure 3. Cell wall synthesis inhibitor TRIP Screen results and validations. **A)** Specific fitness plots comparing two TRIP screens against baseline SDMM in **Top) Ceftriaxone** and **Bottom) Vancomycin**. **B)** Monoculture growth curve validation of TRIP results in **Top) SDMM + Ceftriaxone** and **Bottom) SDMM + Vancomycin**. **C)** Specific fitness plots comparing TRIP results in Ceftriaxone and Vancomycin. **D)** Monoculture growth curves of the three “rapid biphasic lysis” strains. **E)** Network representation of genes found to be differentially regulated (by PadJ) by two or more of CpsR, SusR and TreR overexpression. Green and red edges represent activation and repression, respectively. **F)** Enrichment of Gene Ontology terms of genes found to be differentially regulated (by PadJ) by two or more of CpsR, SusR and TreR overexpression. For all specific fitness plots, the average fitness ($\text{Log}_2(\text{Fold-Change in Abundance})$) of each TFI strain across three biological replicates is shown. For all growth curves, average of four replicates shown, error bars represent standard error.

4.2.2 Oxidative Stress Response Regulators Influence Kanamycin and Levofloxacin Susceptibilities

TRIP screens were completed under two other antibiotic stresses: Kanamycin and Levofloxacin (Fig. 4). Kanamycin is an aminoglycoside antibiotic that targets the 30S subunit of the bacterial ribosome to inhibit protein synthesis. Interaction between Kanamycin and the ribosome results either in translation infidelity, where the incorrect amino acid is added to a growing peptide chain, or complete inhibition of ribosome translocation³⁸. At higher concentrations, Kanamycin can also bind and disrupt bacterial membranes⁴⁷. Levofloxacin, on the other hand, is a fluoroquinolone antibiotic that inhibits DNA synthesis. Depending on the bacterial species, Levofloxacin can inhibit one of two enzymes, topoisomerase IV and DNA gyrase, which are responsible for separating duplicated strands of DNA and relaxing supercoiled DNA, respectively^{48,49}. For *S. pneumoniae* and other gram-positives, adaptive evolution experiments with Levofloxacin identified resistance mutations within the *parC* gene of topoisomerase IV, indicating this is the primary target of Levofloxacin⁴⁸. Both Kanamycin and Levofloxacin are bactericidal antibiotics that are known to produce oxidative stress prior to cell death.

TRIP screens completed at 0.67xMIC Kanamycin reveal several TFI strains with positive fitness effects (Fig. 4A). TFI_0376, which overexpresses the orphan response regulator RitR, had significantly positive fitness in TRIP screens and clear resistance in monoculture growth curves (Fig. 4B). RitR is post-translationally regulated through oxidation of a cysteine residue¹⁷ and overexpression in SDMM results in repression of the pneumococcal iron uptake transporter (*piuABCD*, *SP_1869-1872*) as well as differential regulation of

several genes involved in the oxidative stress response (*lctO*, *spxB*, *dpr*, *trxA*, *trxB*, *tpxD*), several proteases (*ClpEX*, *htrA*, *ftsH*, *zmpB*, *zmpC*, *ptrB*, *htpX*, *ppqL*, *SP_0617*) and several metal transporters (*nlpA*, *mnhH*, *corA*, *ccmA*, *piaABCD*, *psaABC*, *afuA*, *adcABC*). Gene set enrichment analysis of RitR's regulon identify enrichment in several related Gene Ontology terms including “Iron Ion Homeostasis”, “Cell Redox Homeostasis” and “Metalloendopeptidase Activity” (Fig. 4C). RitR's regulation of several proteases is critical to degrade the misfolded proteins caused by Kanamycin, while regulation of the oxidative stress response and metal homeostasis is critical to detoxify the reactive oxygen species caused by the bactericidal antibiotic. While not significant, TFI_2195, which overexpresses the regulator of the Clp protease, CtsR, had positive fitness in Kanamycin treatment and reemphasizes the importance of protease activity during translational stress.

TFI_1774 and TFI_0727 overexpress regulators with overlapping functions with RitR and have positive but insignificant fitness effects in Kanamycin treatment. TFI_1774, which overexpresses an unnamed MarR family transcription factor, regulates the expression of thioredoxin (*trxA*, *SP_1776*), which is an important antioxidant in the oxidative stress response, and TFI_0727 overexpresses the regulator of copper export, CopY, which regulates the metal transporter *zntA* (*SP_0729*). Interestingly, copper levels were shown to be elevated in response to bactericidal antibiotics in *E. coli*, and *E. coli*'s MarR, which TFI_1774 is homologous to, is regulated through oxidation of a conserved cysteine, similar to RitR⁵⁰. Clearly, regulation of protease activity, metal transport and the oxidative stress response is critical to *S. pneumoniae*'s stress response to Kanamycin, which appears to be redox regulated.

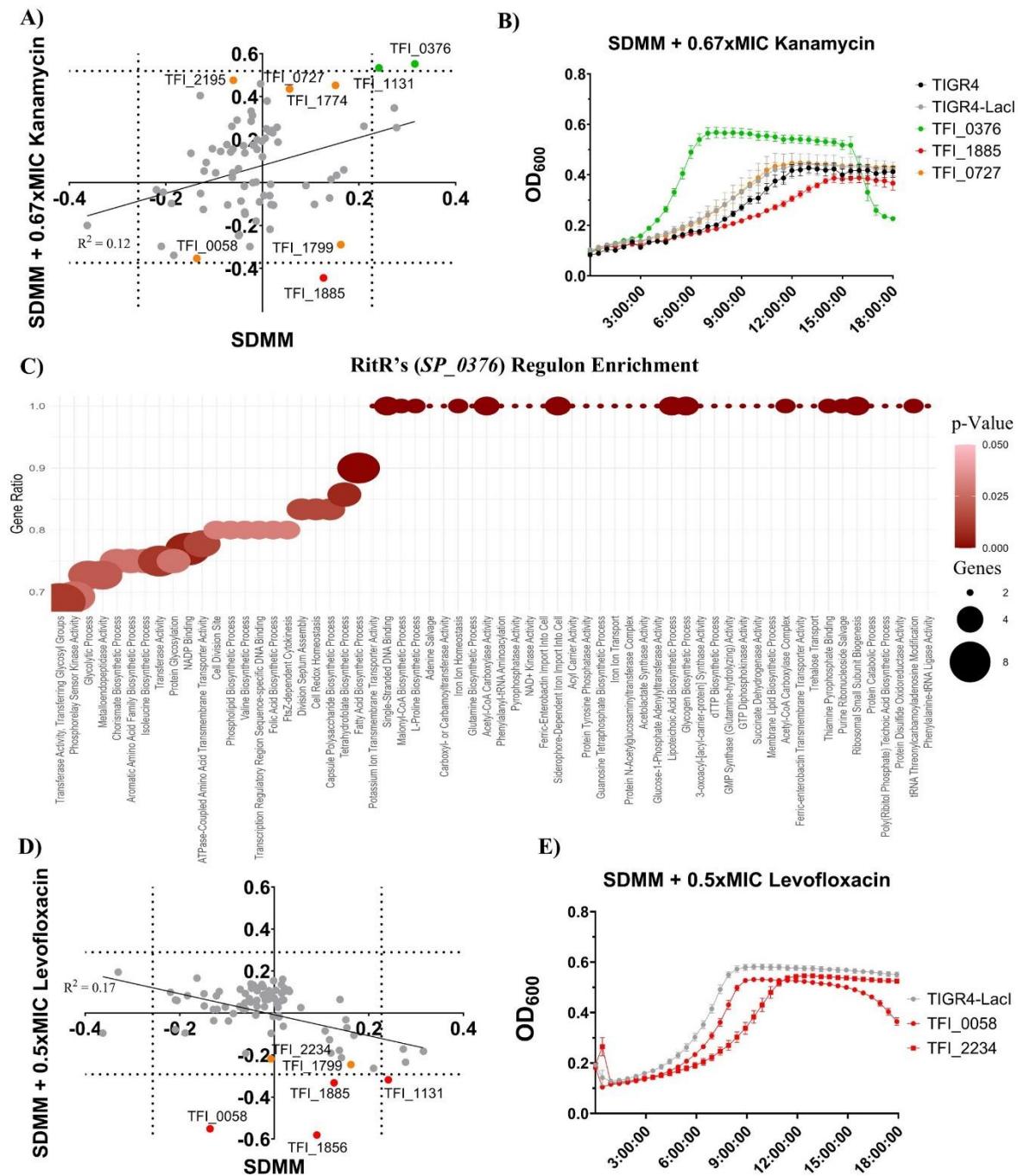


Figure 4. Kanamycin and Levofloxacin TRIP Screen results and validations. **A)** Specific fitness plot comparing TRIP screen results between baseline SDMM and Kanamycin. **B)** Monoculture growth curve validation of TRIP results in Kanamycin. **C)** Dot plot of Gene Ontology Enrichment in RitR's regulon. **D)** Specific fitness plot comparing TRIP screen results between baseline SDMM and Levofloxacin. **E)** Monoculture growth curve validation of TRIP results in Levofloxacin. For all specific fitness plots, the average fitness ($\log_2(\text{Fold-Change in Abundance})$) of each TFI strain across three biological replicates is shown. For all growth curves, average of four replicates shown, error bars represent standard error.

The only TFI strain with significantly decreased fitness during Kanamycin treatment is TFI_1885, which overexpresses the trehalose metabolic regulator, TreR. This strain was one of three TFI strains with a “Rapid Biphasic Lysis” growth profile and significantly increased fitness to cell wall synthesis inhibitors. Both other “Rapid Biphasic Lysis” strains, TFI_0058 and TFI_1799, had negative but insignificant fitness in Kanamycin. Interestingly, both TFI_1885 and TFI_0058 had significant fitness defects in Levofloxacin as well, and TFI_1799 had an insignificant fitness defect (Fig. 4D). The negative fitness of these strains in both Kanamycin and Levofloxacin suggest they are susceptible to oxidative stress, which supports our earlier hypothesis that metabolic rewiring of these strains increases cellular acidity. Somehow, this metabolic rewiring protects against cell-wall synthesis inhibitors, perhaps through peptidoglycan downregulation or restructuring, but potentiates oxidative stressors.

Two other strains with significant fitness defects in Levofloxacin are TFI_1131 and TFI_1856 (Fig 4D). The significant negative fitness of TFI_1131, which overexpresses an uncharacterized XRE-family transcription factor with no identifiable target genes, is likely due to its increased fitness in plain SDMM, reemphasizing the fact that faster growing cells are more susceptible to antibiotic treatment. TFI_1856, overexpressing the redox-sensitive regulator of *adhC*, NmlR, also had decreased fitness to Kanamycin treatment, likely due to its importance in the oxidative stress response. Interestingly, in TRIP screens assaying the tolerance of each TFI strain, TFI_1856 demonstrated increased tolerance to Levofloxacin (Fig. 8C). This suggests the “fitness defect” displayed here, at sub-MIC concentrations of Levofloxacin, may be a tolerant growth arrest, as fitness here is calculated based on population expansion.

Interestingly, TFI_2234, which overexpresses a putative phage regulator, had negative but insignificant fitness in Levofloxacin. This strain also displayed significant fitness defects in elevated concentrations of zinc (Fig. 6). Overexpression of the transcriptional regulator of deoxynucleotide synthesis, *nrdR* (*SP_1713*), had no fitness effect to Levofloxacin.

4.3 Specific Stresses can be Characterized through Abiotic TRIP Screens

Despite their current commercialization and the rational design of synthetic derivatives, antibiotics are a natural mechanism used by microbes to kill niche and resource competitors. As we have evolved alongside bacterial pathogens, it is unsurprising that we have evolved similar mechanisms to inhibit bacterial growth. Several stresses that *S. pneumoniae* encounters in the host can be simplified down to the same fundamental stresses caused by antibiotic treatments. A particularly relevant example of this is the oxidative burst of the immune system, where neutrophils and macrophages release large amounts of reactive oxygen species (ROS) to kill pathogens. ROS damage all aspects of cellular physiology and lead to cell death by oxidizing proteins, damaging DNA and disrupting membrane structure. It has been recently shown that all bactericidal antibiotics, regardless of target, result in the production of ROS through the Fenton reaction leading to cell death³⁵. The Fenton reaction produces hydroxy radicals through the reduction of hydrogen peroxide by ferrous iron (Fe^{2+}), which is particularly relevant in *S. pneumoniae* as it produces large amounts of hydrogen peroxide as a means of both microbial competition and virulence. Our TRIP screens in bactericidal antibiotics further support this narrative as several regulators of the oxidative stress response were shown to significantly influence antibiotic susceptibilities.

Closely linked to the oxidative burst of the immune system and the ROS-induced cell death pathway of bactericidal antibiotics is another chemical stress encountered by *S. pneumoniae* in the host, zinc intoxication. Transition metals, such as zinc, manganese and iron, are essential cofactors for many cellular processes, and up to 40% of all proteins require a metal cofactor⁵¹. At high concentrations, however, these essential metals, especially zinc and copper, become toxic due to mismetallation, where proteins are inhibited through binding of a noncognate metal. Our immune defense uses this as an antimicrobial strategy. Macrophages deploy a lipopolysaccharide-inducible zinc importer that senses intracellular bacteria and overwhelms them with toxic levels of zinc⁵². In *S. pneumoniae*, zinc intoxication is best known for mismetallation of the manganese importer PsaA (*SP_1650*) by zinc, which leads to manganese starvation. Manganese itself is a critical cofactor for the oxidative stress response, suggesting that zinc intoxication acts synergistically with oxidative stress caused by both antibiotics and the immune system. Further support for the synergistic action of zinc intoxication and oxidative stress is that, along with the reduction of hydrogen peroxide to ROS, the Fenton reaction oxidizes ferrous iron (Fe^{2+}) to ferric iron (Fe^{3+}), which inhibits the activity of enzymes that require a ferrous iron cofactor and creates an opportunity for zinc mismetallation. Overexpression of several regulators of metal homeostasis were shown to have significant fitness defects in antibiotic treatments, especially to the protein synthesis inhibitor kanamycin (Fig. 4A).

We therefore hypothesized that dysregulation of the oxidative stress response and metal ion homeostasis in *S. pneumoniae* would potentiate the lethal effects of both antibiotics and the immune system. To gain a network-level understanding of these processes, we completed TRIP screens under the oxidative stress of removing catalase from SDMM, zinc

intoxication, and treatment with paraquat. Paraquat is a highly toxic redox cycling compound that generates large quantities of ROS by reacting with electron carriers within the cell. For this reason, paraquat is commonly used to study the oxidative stress response in both bacterial⁵³ and human⁵⁴ cells. *S. pneumoniae* is typically cultured in the presence of catalase to detoxify the hydrogen peroxide it naturally produces, so paraquat TRIP screens were completed in SDMM without catalase.

In TRIP screens in SDMM without catalase no strains had significantly positive fitness, while two strains, TFI_1638 and TFI_1856, had significantly negative fitness (Fig. 5A). TFI_1856 overexpresses NmlR which directly regulates the alcohol dehydrogenase *adhC* (*SP_1855*) and has been shown to be important in *S. pneumoniae*'s oxidative stress response³⁴. Previous TRIP screens show TFI_1856 to have a significant fitness advantage when catabolizing mannose (Fig. 2F) and a significant fitness defect in Levofloxacin (Fig. 4D). However, in the next section TFI_1856 will be shown to display increased tolerance to several antibiotics, so these fitness defects may be a misrepresentation of tolerance-achieving growth arrest.

TFI_1638 had a significant fitness defect without catalase and overexpresses the manganese regulator, PsaR. Manganese is known to be involved in the oxidative stress response and PsaR regulates manganese homeostasis through *psaABC* (*SP_1648-1650*), an ABC-type manganese transport system, as well as other genes involved in the oxidative stress response and virulence. Co-transcribed with *psaABC* is thiol peroxidase, *tpxD* (*SP_1651*), which protects the cell from ROS by reducing peroxides⁵⁵. Interestingly, TFI_1638 was the only strain with significantly positive fitness when treated with paraquat

(Fig. 5B). TFI_1638's positive fitness in the oxidative stress induced by paraquat is likely due to activation of the oxidative stress response through manganese and *tpxD*. Interestingly, PsaR also regulates several virulence genes, including *pcpA* (*SP_2136*) and *prtA* (*SP_0641*), genes involved in host adhesion and an immune system evading peptidase, respectively. This suggests an evolved regulatory link between the oxidative stress response, sensing of the immune system, and virulence.

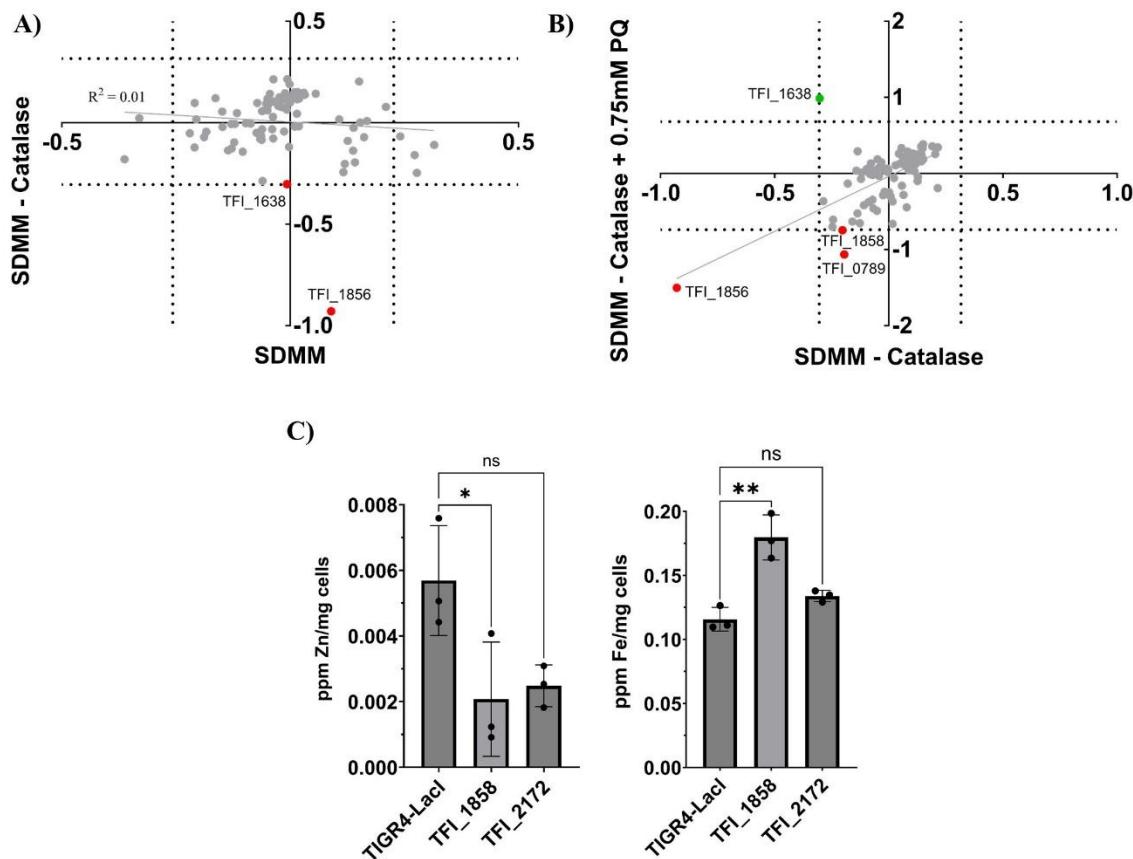


Figure 5. ROS TRIP Screen Results and Validation. Specific fitness plots for **A**) SDMM without catalase and **B**) SDMM without catalase plus Paraquat (PQ) **C**) ICP-MS results of total cell associated Zinc (left) and Iron (Right) in parts per million (ppm) per milligram of dry cell weight. Significance testing by one-way analysis of variance (ANOVA) with Dunnett multiple comparisons (ns = not significant; * = $P < 0.05$; ** = $P < 0.01$). For all specific fitness plots, the average fitness ($\log_2(\text{Fold-Change in Abundance})$) of each TFI strain across three biological replicates is shown.

TFI_1856, TFI_0789 and TFI_1858 all had negative fitness in the oxidative stress induced by paraquat. TFI_1856 had significantly negative fitness in both the control condition (SDMM – catalase) and when treated with paraquat, while TFI_0789 and TFI_1858 had negative but insignificant fitness in the control. TFI_0789 overexpresses the phenolic acid stress regulator, PadR, and is involved in the oxidative stress response through regulation of *ydhF* (*SP_0791*), an aldo/keto reductase, several uncharacterized genes, and *fabG* (*SP_0793*). Phenolic acid stress in bacteria has been shown to induce membrane damage^{56,57}. The activities of YdhF and FabG would be beneficial to the cell under oxidative stress as YdhF can reduce and detoxify the lipid aldehydes produced by ROS while FabG catalyzes the initial reduction step in fatty acid biosynthesis. Lastly, TFI_1858 overexpresses the zinc efflux regulator, SzcA, which is known to be connected to the oxidative stress response through regulating zinc homeostasis⁵⁸. The other zinc regulator, AdcR, had similar negative fitness in paraquat, but failed to reach significance. Interestingly, overexpression of both the zinc efflux regulator, SzcA, lead to a significant increase in cell-associated iron, while an insignificant increase was seen in overexpression of AdcR (Fig. 5C). This aligns a recent study in *Streptococcus agalactiae* that showed *szcA* mutants had significant increases in total cell associated iron⁵⁹ and suggests dysregulated zinc homeostasis may result in increased oxidative stress via the iron-catalyzed Fenton reaction.

Zinc intoxication TRIP screens revealed one drastically less fit TFI strain: TFI_2234 (Fig. 3). TFI_2234 overexpresses the Phage Infection Protein regulator, PipR, which regulates two transmembrane proteins that may serve as bacteriophage receptors. Bacteriophages are viruses that infect bacteria and have garnered significant research interest to develop

“phage therapies” to treat bacterial infections. Different genome versions of TIGR4 show *pip* (*phage infection protein*, *SP_2232*) as having an internal stop codon, which would provide resistance to phage infection. However, our lab’s whole genome sequence of TIGR4 (CP035239.1) shows complete versions of *pip* and the other unannotated transmembrane protein, *SP_2231*. Interestingly, zinc⁶⁰ and other metals⁶¹ have been shown to enhance bacteriophage infectivity of diverse bacterial species, and fluoroquinolones such as Levofloxacin are known to be frequent prophage inducers⁶². Taken together, this supports the idea that the combination of zinc intoxication and PipR overexpression led to activation of a bacteriophage in TIGR4.

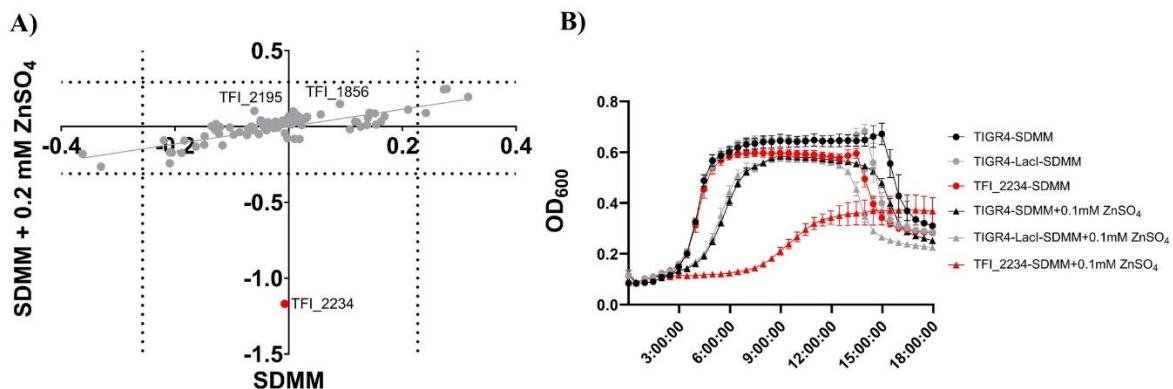


Figure 6. Zinc intoxication TRIP Screen Results and Validation. **A)** Specific fitness plots for SDMM + 0.2mM + Zinc sulfate (ZnSO₄); the average fitness (Log₂(Fold-Change in Abundance)) of each TFI strain across three biological replicates is shown. **B)** Monoculture growth curve validation of SDMM + Zinc sulfate TRIP screen results; the average of four replicates shown, error bars represent standard error.

Overall, there was a very high correlation between TRIP results in plain SDMM and in SDMM + zinc. The coefficient of determination of a linear regression between TRIP results in SDMM and with zinc was 0.21, but if you remove the outlier TFI_2234 it becomes 0.66.

This indicates that the excess zinc has little differential fitness effect across TFI strains, except for TFI_2234.

There were two TFI strains that appear to have slightly, but statistically insignificant, positive specific fitness to zinc: TFI_2195 and TFI_1856. As mentioned previously, TFI_1856 overexpresses NmlR which directly regulates the alcohol dehydrogenase *adhC* (*SP_1855*) and has been shown to be important in *S. pneumoniae*'s oxidative stress response⁶³. Despite apparently only directly regulating *adhC*, overexpression of NmlR in plain SDMM revealed overexpression of two zinc efflux pumps *czcD* (*SP_1857*) and *zntA* (*SP_0729*). It is noteworthy that *nmlR* lies directly upstream *czcD*, yet they are transcribed from different promoters and there is no identifiable NmlR motif in the promoter of *czcD*. Interestingly, activity of an alcohol dehydrogenase in *E. coli* was inhibited by mismetallation with zinc⁶⁴, which may explain TFI_1856's activation of zinc efflux pumps and the genomic pairing of *nmlR* and *czcD*. TFI_2195 overexpresses the Clp protease regulator, CtsR. The Clp protease is critical for proteostasis through ATP-dependent degradation of misfolded proteins, and overexpression of *ctsR* in plain SDMM revealed repression of the Clp protease and differential expression of many oxidative stress response genes, including *gor*, *pdxST*, *dnaK* and *hrcA*. Overexpression of CtsR and NmlR resulting in positive fitness in elevated zinc concentrations further supports the link between zinc intoxication, protein misfolding and the oxidative stress response.

4.4 Time-kill Assay Adapted TRIP Screens Reveal Transcriptional Regulators' Influence on Antibiotic Tolerance

Antibiotic tolerance is the ability of bacterial cells *to survive* antibiotic treatment. This differs from antibiotic resistance, where bacterial cells are able *to grow* in the presence of

antibiotics. Tolerance has been linked to slow growth and lesser metabolic output, minimizing antibiotic stress by severely downregulating the cellular processes targeted by antibiotics. Persistence is a phenomenon where a subset of clonal cells displays extreme tolerance to antibiotics and have been long linked with antibiotic treatment failure^{65,66}.

Toxin-antitoxin (TA) systems were the first identified mechanism to achieve persistence, specifically through the *hipAB* system in *E. coli*⁶⁷. Proteolytic degradation of the antitoxin, HipB, allows HipA to phosphorylate and inactivate the glutamyl-tRNA synthetase GltX, which in turn halts translation and activates the stringent response⁶⁸. There are five TA systems in TIGR4, all of which target the same processes as antibiotics: DNA replication, translation and cell wall synthesis^{69,70}. Of these, *pezAT* (*SP_1050-51*) is unique in that it is the only known TA system to target cell wall synthesis and has been shown to protect against the cell wall synthesis inhibitor ampicillin⁷¹. Interestingly, the RNase activity of the toxin YeoB (*SP_1740*) has been reengineered to target oncogenic micro-RNA's to achieve selective killing of certain breast cancer cells⁷².

Along with TA systems, the stringent response has been linked to antibiotic tolerance through metabolic downregulation of key cellular processes during nutrient limitation and other stresses. Under stressful conditions, the alarmone (p)ppGpp is produced by RSH (RelA-SpoT Homolog) in gram-positives. Accumulation of this alarmone leads to inhibition of translation, DNA replication, and GTP synthesis, with a simultaneous increase in expression of genes involved in amino acid synthesis, the oxidative stress response and virulence^{44,73}. CodY (*SP_1584*) is a global nutritional transcriptional regulator that has long been linked to the stringent response, and tangentially tolerance, because it represses

global metabolism in response to depleted branch chain amino acid (BCAA) and, in some species, GTP pools^{74,75}.

Recent studies have identified translational deficiency and accumulation of protein aggregates as a key driver of antibiotic tolerance. Single-cell RNAseq analysis of tetracycline-, a bacteriostatic protein synthesis inhibitor, -treated cells revealed that these cells clustered with persisters⁶. Conversely, ATP-deficient cells have been shown to produce protein aggregates that correlate with increased antibiotic tolerance. These cells then require ATP-dependent proteolysis to resuscitate, with protein disaggregation preceding growth resumption. To this end, a synthetic system to generate a proteolytic queue at the protease ClpXP through overexpression of SsrA-tagged peptides was shown in *E. coli* to increase antibiotic tolerance to both Ampicillin and Ciprofloxacin⁷⁶. Importantly, an antibiotic that activates the ClpP protease, acyldepsipeptide (ADEP 4), has been shown to effectively kill persisters in an ATP-independent manner⁷⁷.

Time-kill assays are commonly used to assay persistence and tolerance, where bacterial cultures are treated with high concentrations of antibiotics and the fraction of surviving cells are determined at set time points. The vast majority of cells rapidly die at these concentrations of antibiotics, but tolerant populations die at a significantly reduced rate. Persisters, on the other hand, are classified by a biphasic death curve, where the majority of cells die quickly but a small subset is notably tolerant to the antibiotics⁷⁸. Persister cells greatly reduce their metabolic output, express genes critical for surviving the antibiotic stress and have been shown to precede the development of resistance through genetic

mutation and/or horizontal gene transfer^{6,9}. As persister cells are initially genetically identical to susceptible cells, their development must be regulatory.

Therefore, to gain a better understanding of how the transcriptional regulatory network of *S. pneumoniae* influences antibiotic tolerance, and perhaps identify routes to persistence, we adapted time-kill assays to TRIP screens in several antibiotics. We hypothesized that transcriptional regulators of toxin-antitoxin systems, the stringent response, and protease activity would be significant hits in our tolerance TRIP screens. Luckily, knockout of one transcription factor, PtvR (*SP_0100*) has been shown to significantly increase tolerance to vancomycin⁷⁹, so its TFI strain can serve as an internal positive control.

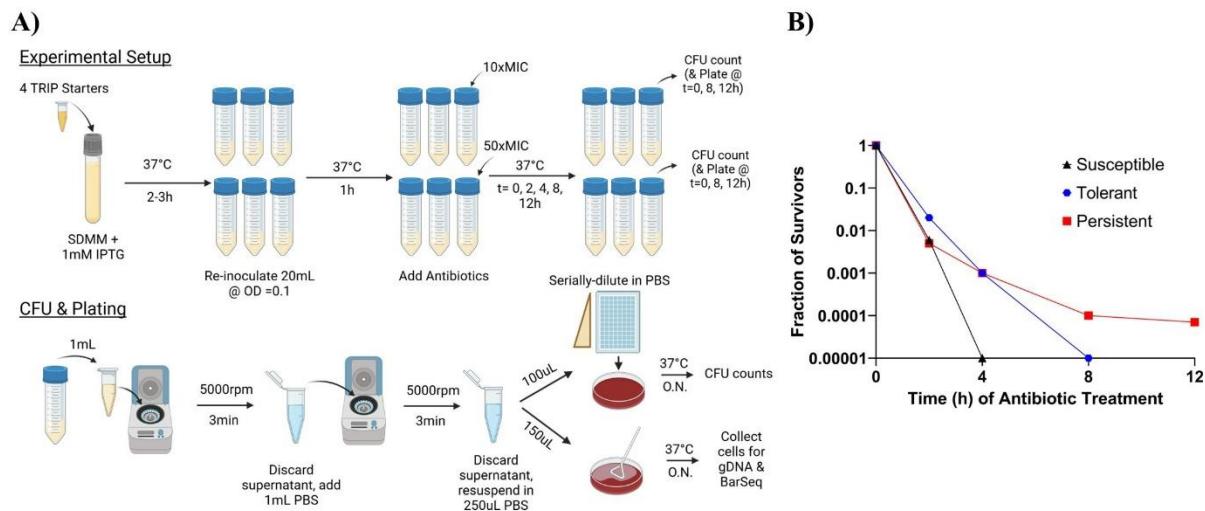


Figure 7. Tolerance TRIP Screen Overview. **A)** Experimental diagram of Time-Kill Assay adapted TRIP Screens. **B)** Hypothetical time-kill assays demonstrating death curves of susceptible, tolerant and persistent (sub)populations of cells.

Tolerance TRIP screens were completed in three different antibiotics: Levofloxacin, Ceftriaxone and Vancomycin in two high concentrations of antibiotics (Fig. 8). Tolerance

and persistence are both the ability to survive extreme concentrations of antibiotics and are thought to be dose-independent, so using two concentrations both well above the MIC should produce the same results. High concentrations of Vancomycin decimated the TRIP pools, leaving very few cells alive after 8h of treatment and large amounts of noise in the data. Interestingly, only tolerance TRIP screens in Levofloxacin produced the biphasic death curve indicative of persistence (Fig. 8A). Paradoxically, there were more surviving cells in higher concentrations of both Ceftriaxone and Vancomycin. This may be attributed to the Eagle effect, where concentrations of antibiotics well above their optimal bactericidal concentrations decrease bactericidal activity^{80,81}.

While the data is noisy, especially for Vancomycin which had the greatest bactericidal activity, the same strains appear to have increased and decreased tolerance to the three drugs tested (Fig. 8B). Tolerant TRIP screens in Levofloxacin and Ceftriaxone have high correlation (Fig. 8C), which indicates that there are common mechanisms to survive stress stemming from different antibiotics. Six TFI strains had significantly increased tolerance to either Levofloxacin or Ceftriaxone: TFI_0246, TFI_1674, TFI_2119, TFI_0376, TFI_1856 and TFI_2195.

Importantly, of the TFI strains with positive fitness in tolerance TRIP screens, TFI_2195, which overexpresses the ClpXP protease regulator, CtsR, was the only strain to have significantly increased tolerance to both Levofloxacin and Ceftriaxone (Fig. 8C). This confirms that repression of protease activity in *S. pneumoniae* leads to increased tolerance/persistence as has recently been demonstrated in *E. coli*⁶ and even targeted with the anti-persistence and ClpP-activating drug, ADEP4 in *S. aureus*⁷⁷. This tolerance

phenotype derived from repressed protease activity has now been shown across three bacterial species and across diverse antibiotics, indicating a conserved mechanism among pathogenic bacteria. Reduction in ATP levels has long been linked to tolerance and persistence^{78,82} and recently has been shown to produce protein aggregates seen in persister and viable but non-culturable (VNBC) cells⁷, which may be due to the ATP requirement of ClpXP. This may also link antibiotic tolerance to the development of antibiotic resistance as ComX, the competence sigma factor responsible for expressing the late competence genes, is known to be a ClpXP target⁸³ and CRISPRi repression of both ClpP and ClpX have been shown to trigger competence⁸⁴. Many stress response regulators, including antitoxins and transcription factors, are regulated by proteolysis, so a genome wide understanding of which regulators are degraded by ClpXP may provide further insight into the development of tolerance and persistence.

TFI_0246 overexpresses the metabolic repressor DeoR, that was previously found to have significant fitness defects in plain media (Fig. 2), presumably through repression of a glycyl-radical enzyme activating protein (*SP_0245*). SP_0245 may indirectly regulate the activity of pyruvate formate lyase A (*pflD*, *SP_0251*), which catalyzes the conversion of pyruvate to acetyl-CoA during anaerobic metabolism. This slow growth phenotype may predispose cells for increased tolerance.

TFI_1674 overexpresses NanR, the sialic acid (N-acetylneuraminic acid, Neu5Ac) utilization regulator. Neu5Ac is one of the most important carbohydrates for *S. pneumoniae* because of its potential as a carbon source and its role as a signaling molecule for biofilm formation and the expression of virulence genes^{85–87}. NanR regulates expression of NanA,

a neuraminidase which promotes colonization of the nasopharynx by cleaving sialic acid residues from host glycans and presenting host cell receptors for adherence⁸⁵. Overexpression of NanR in SDMM + glucose revealed activation of the capsule operon, several cell wall synthesis genes (*rfaB*, *wcaA*, *wcaK*, *murMN*, *murE*) and glycosyltransferases (*glyBCDEFG*). NanR overexpression also resulted in repression of the F0F1 ATP synthase complex (*atpABCDEFG*), activation of *deoR* and the toxin, *pezT*. Overexpression of *deoR* also displays increased tolerance, and interestingly, overexpression of *pezT*'s cognate antitoxin, *pezA* (*SP_1050*), was shown to have decreased tolerance. TFI_1674 also had negative but insignificant fitness TRIP screens in SDMM + glucose. Taken together, overexpression of NanR increases expression of capsule and cell wall synthesis genes while slowing metabolism and ATP synthesis which leads to increased tolerance to antibiotics.

TFI_2119 overexpresses an unnamed transcription factor of the xenobiotic response element (XRE) family which are known to be involved in various stress responses including heat shock and oxidative stress^{88,89}. In plain SDMM, overexpression of SP_2119 lead to strong repression of an upstream operon, *SP_2117-SP_2115*, which encodes an excreted peptide (*SP_2115*), a membrane protease involved in post-translational processing of membrane-associated proteins (*ydiL*, *SP_2116*) and a Brp/Blh family β-carotene 15,15'-monooxygenase (*brp*, *SP_2117*). β-carotene 15,15'-monooxygenases catalyze the cleavage of β-carotenes to all-trans retinal. Under normal conditions in *E. coli*, Brp has been proposed to alter membrane composition by anchoring either β-carotene or processed retinal to the cell membrane. In stressful conditions, it is proposed that *brp* is expressed from an alternative start codon, producing a shorter peptide that, through an unknown

mechanism, increases osmotic stress tolerance⁹⁰. Notably, overexpression of both the full-length and truncated versions of Brp were shown to increase salt tolerance⁹⁰, and carotenes are known to be radical scavengers which protect against oxidative stress and regulate membrane flexibility^{91,92}. We hypothesize that an unknown stressor in tolerance TRIP screens in TFI_2119 lead to an increase in expression of *brp*, which in turn altered the membrane composition to protect against oxidative and/or osmotic stress.

Two strains previously identified as having significant fitness effects in several previous TRIP screens also have significant tolerance to Ceftriaxone, TFI_0376 and TFI_1856. TFI_0376 overexpresses the orphan response regulator, RitR and was shown to have significant increase in fitness in both plain SDMM and during kanamycin treatment. RitR is redox sensitive and regulates several processes related to the oxidative stress response, including iron and other metals uptake, antioxidants and several proteases. The emergent phenotype of overexpression of RitR may have predisposed the cells for survival of antibiotic stress through regulation of metal homeostasis, protease activity and the oxidative stress response. Another redox sensitive transcription factor involved in the oxidative stress response is TFI_1856, which overexpresses NmlR and regulates expression of *adhC*. TFI_1856 has been shown to have significant fitness effects when catabolizing mannose (Fig. 2F, 2I), in sub-MIC levofloxacin treatment (Fig. 4D) and in ROS-conditions (Fig. 5A, 5B). Overexpression of NmlR in sub-MIC levofloxacin and under oxidative stress appeared to result in reduced fitness, but may have instead been a tolerance phenotype through reduced growth.

Conversely, six strains had significantly reduced tolerance to ceftriaxone and/or levofloxacin: TFI_0100, TFI_1584, TFI_1050, TFI_0058, TFI_1799 and TFI_1885. TFI_0100 overexpresses the regulator of phenotypic tolerance to vancomycin, PtvR, and served as an internal control. *ptvR* mutants in *S. pneumoniae* were previously shown to have increased tolerance to vancomycin through derepression of the *ptv* operon⁷⁹, and here overexpression of PtvR lead to a significant decrease in tolerance. Interestingly, here we show that overexpression of PtvR and the subsequent repression of the *ptv* operon leads to decreased tolerance to not only the cell wall synthesis inhibitor, Ceftriaxone, but also to the DNA synthesis inhibitor, Levofloxacin.

Another TFI strain that we suspected may play a role in regulating tolerance is TFI_1584, which overexpresses the nutritional regulator, CodY, and showed significantly reduced tolerance in our TRIP screens. CodY's ability to bind to DNA is enhanced through interaction with BCAAs and oxidation of its two cysteine residues, which in turn induces expression of thiol peroxidase (*tpxD*), an important antioxidant, as well as alters global metabolism by repressing uptake and biosynthesis of BCAA while shifting glucose utilization away from glycolysis and towards the pentose phosphate pathway⁹³ (Fig. 9).

TpxD catalyzes the reduction of toxic peroxides, such as hydrogen peroxide (H₂O₂), but becomes oxidized in the process. NADPH is then required to reduce oxidized *TpxD* back into its active form. The metabolic shift of glucose away from glycolysis towards the pentose phosphate pathway helps protect against oxidative stress by increasing NADPH which helps maintain reduced *TpxD*. Downregulation of glycolysis also reduces oxidative stress by reducing the production of pyruvate, the reactant for the main source of pneumococcal H₂O₂, *SpxB*, as well as a potential reactant for the synthesis of BCAA⁹⁴.

Synthesis of BCAA (including Isoleucine, Leucine and Valine (ILV)) can be inhibited by oxidative stress or copper toxicity through inactivation of the iron-sulfur cluster dependent isopropylmalate isomerase (*SP_1255*)^{95,96}.

Supplementation of TIGR4-LacI with 5mM of ILV lead to a 1,000-fold increase in survivors after 8h of 10xMIC of Ceftriaxone treatment (Fig. 8D). The presence of both BCAA and H₂O₂ both activate expression of *tpxD* and shift glucose metabolism towards the pentose phosphate pathway through CodY, priming the cell to survive oxidative stress. Further, as tolerance has been intimately linked with proteostasis, we suspect supplementation of BCAA may help the cell resume translation and growth after the antibiotic stress is alleviated.

TFI_1050 overexpresses the antitoxin PezA and had decreased tolerance to both Ceftriaxone and Levofloxacin. Under normal conditions, PezA neutralizes the activity of the toxin PezT and represses expression of its operon. However, in stressful conditions, PezA is preferentially degraded, leaving the toxin PezT to inhibit cell wall synthesis⁷¹. By overexpressing the antitoxin PezA, we are synthetically repressing the activity of PezT, which apparently decreases tolerance. This was the only antitoxin of the five TA systems in TIGR4 for which a TFI strain was cloned, so it would be interesting to see if overexpression of the other antitoxins result in reduced tolerance. While TA systems are known to be redundant, it's important to note that overexpression of just one of the five antitoxins leads to a quantifiable and significant decrease in tolerance.

Lastly, the three “Rapid Biphasic Lysers”, TFI_1799, TFI_1885 and TFI_0058, all had significantly reduced tolerance. These strains, overexpressing SusR, TreR and CpsR,

respectively, are metabolic regulators that were previously shown to have significant increases in resistance to two cell wall synthesis inhibitors, Ceftriaxone and Vancomycin (Fig. 3) and significant increases in susceptibility to both Kanamycin and Levofloxacin (Fig. 4). We suspect that the metabolic rewiring by overexpressing these metabolic regulators leads to more acidic environment, leading to greater susceptibility to Kanamycin, Levofloxacin, autolysin activity and high concentrations of antibiotics.

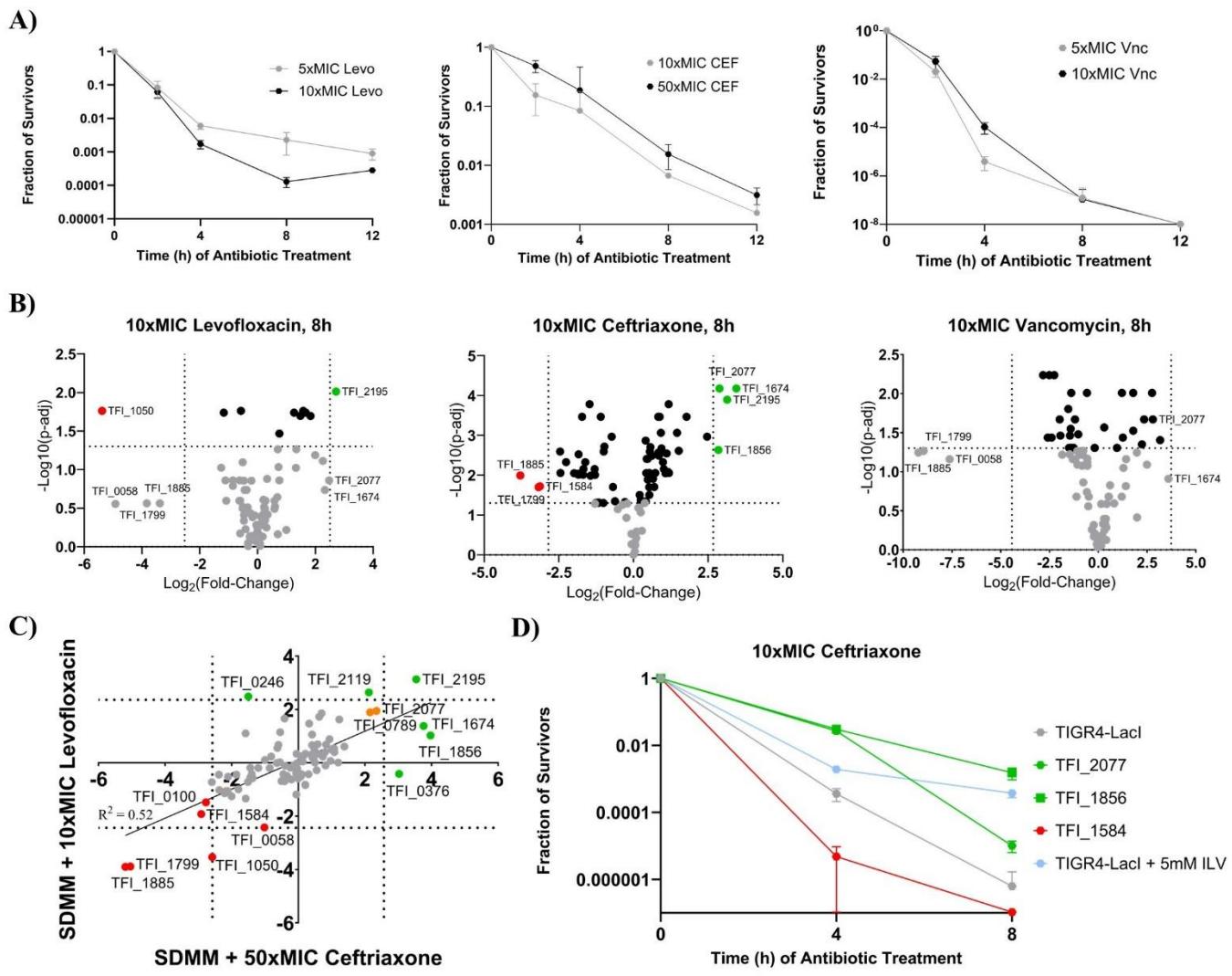


Figure 8. Tolerance TRIP Screen Results and Validation. **A)** Survivorship curves of Tolerance TRIP screen results across three antibiotics with two concentrations each. Average of three replicates show, error bars represent standard error. **B)** Volcano plots of Tolerance TRIP screens for each antibiotic tested. **C)** Correlation of Tolerance TRIP screen results for Levofloxacin (DSI) and Ceftriaxone (CWSI); the average fitness ($\text{Log}_2(\text{Fold-Change in Abundance})$) of each TFI strain across three biological replicates is shown. **D)** Monoculture time-kill assay validation of Tolerance TRIP screen results for select strains or supplementation with Isoleucine, leucine and Valine (ILV). Average of three replicates shown, error bars represent standard error.

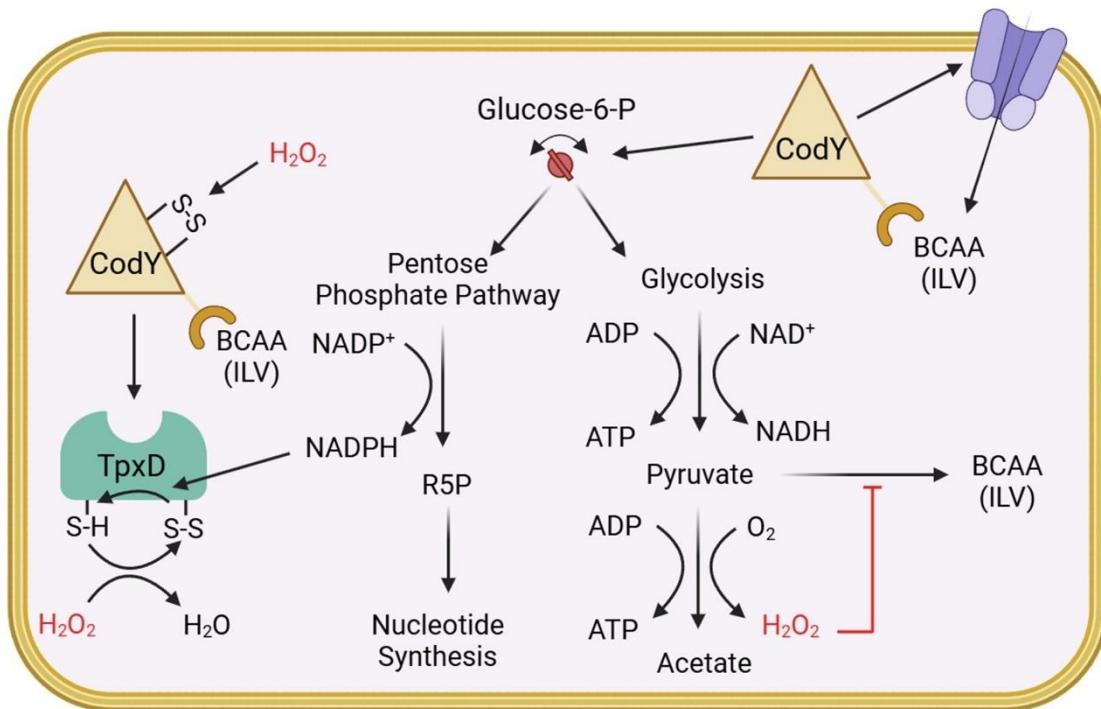


Figure 9. CodY Regulates Glucose Metabolism and the Oxidative Stress Response. Simplified schematic of CodY’s dual role in regulating glucose metabolism and the oxidative stress response through sensing hydrogen peroxide (H_2O_2) and branch chain amino acids (BCAA; Isoleucine, Leucine and Valine (ILV)).

4.5 Comparative Analysis of TRIP Screens Reveals Common Pathways of the Stress Response

Meta analysis of all TRIP screens reveal how antibiotics and other stressors are sensed by TIGR4 and what transcriptional pathways most influence susceptibility to these stresses (Fig. 10). To gain an understanding of how different stresses impact TIGR4, we clustered TRIP screen results using UMAP to visualize similarities between conditions (Fig. 10A). As expected, all metabolic TRIP screens cluster together. TRIP screens cultured on galactose cluster closest to the base media of SDMM + glucose possibly because our inducer IPTG activates the Leloir pathway for galactose metabolism in all TFI strains.

TRIP screens at a low concentration of Kanamycin (0.5x MIC) and with excess zinc also cluster near SDMM, indicative of the low stress levels these conditions applied to our TRIP pools. There are two opposing clusters representing stressful states: the oxidative stress cluster (top right, Fig. 10A) and the tolerance cluster (bottom left, Fig. 10A). As all bactericidal antibiotics tested are known to produce oxidative stress, it is reaffirming to see nearly all antibiotic TRIP screens shifting away from base media towards the oxidative stress cluster. Treatment with the DNA synthesis inhibitor Levofloxacin clusters within the oxidative stress cluster. The only antibiotic TRIP screens to shift towards the tolerance cluster are the higher concentration Kanamycin (0.67xMIC) TRIP samples, which provides further evidence for the causal link between translational stress and antibiotic tolerance.

Aggregating the regulons of transcription factors overexpressed in TFI strains found to be significant hits in TRIP screens helps identify which pathways are most influential to the stress response. Gene set enrichment analysis of this combined regulon identified the gene categories critical to *S. pneumoniae*'s stress response, which include diverse cellular processes. Several enriched categories are involved in metabolism (“Glucose Metabolic Process”, “Carbohydrate Transmembrane Transport”, “Tetrahydrofolate Biosynthetic Processes”), proteostasis (“Amino Acid Transport”, “Metalloendopeptidase Activity”), membrane lipid synthesis (“Fatty Acid Biosynthesis”, “Membrane Lipid Biosynthetic Process”), cell wall synthesis (“Lipoteichoic Acid Biosynthetic Process”), the oxidative stress response (“Cell Redox Homeostasis”, “Protein Disulfide Oxidoreductase”) and metal homeostasis (“Zinc Ion Binding”, “Zinc Ion Transport”, “Iron Ion Homeostasis”). Together, this demonstrates that overcoming stress is a system-wide response with multiple critically important processes.

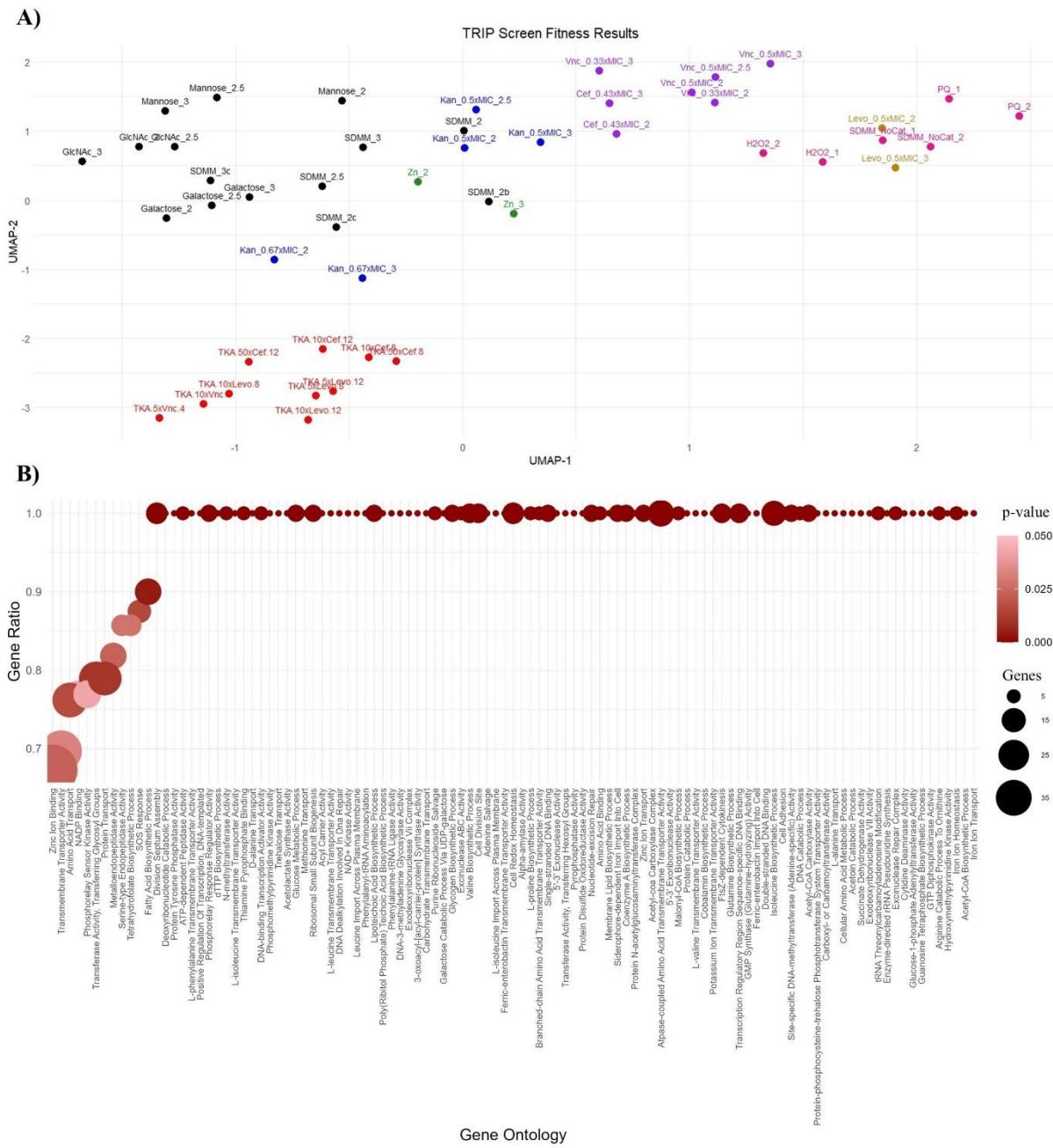


Figure 10. Meta Analysis of TRIP Screen Results. **A)** UMAP representation of TRIP Screen Results. SDMM = Semi-Defined Minimal Media, Cef = Ceftriaxone, Vnc = Vancomycin, Kan = Kanamycin, Levo = Levofloxacin, NoCat = No Catalase, PQ = Paraquat, TKA = Time-Kill Assay, MIC = Minimum Inhibitory Concentration. Numbers in experiment names represent hours of treatment. **B)** Gene Ontology Terms Enriched in TFI strains with Significant TRIP Results. Strains include: TFI_0058 (CpsR), TFI_0100 (PtvR), TFI_0376 (RitR), TFI_0789 (PadR), TFI_1050 (PezA), TFI_1584 (CodY), TFI_1638 (PsaR), TFI_1799 (SusR), TFI_1856 (NmlR), TFI_1858 (SczA), TFI_1885 (TreR), TFI_2195 (CtsR) and TFI_2234 (PipR).

4.6 Discussion

Understanding pathogenic bacteria as complex, fluid systems designed to survive stressful conditions to replicate reshapes antibiotic treatment from a medical to an engineering problem. Advances in synthetic biology for both research and industrial purposes have demonstrated that we can control the behavior of these single-celled organisms, but to do that on a system wide scale, we must understand the effects of its regulatory networks. To understand the role that the transcriptional regulatory network has on cell fitness, we performed Transcriptional Regulator Induced Phenotype (TRIP) screens in a variety of conditions targeted to understand metabolic fitness, abiotic and antibiotic resistance and antibiotic tolerance.

Initial TRIP screens quantified the fitness effects of transcription factor overexpression in rich media (SDMM + glucose, Fig. 2). Validation of significant hits via monoculture growth curves showed little differences, demonstrating the sensitivity of TRIP screens (Fig. 2C). TRIP screens in media with alternate carbon sources revealed the impact of metabolic redundancy and carbon catabolite repression (CCR). There are two pathways capable of catabolizing galactose in *S. pneumoniae*: the Leloir and tagatose-6-phosphate pathways. Unfortunately, the Leloir pathway is regulated by GalR (SP_1854), which is derepressed by IPTG, the same inducer molecule of our synthetic system (See Section 3.1.2). Therefore, in all TFI strains, the Leloir pathway was activated and that obfuscated TRIP results in SDMM + galactose (Fig. 2D). Similarly, catabolism of GlcNAc is dependent on the activities of NagA and NagB, which are dually regulated by NagR (SP_1446) and CCR (via CcpA). Because of this, overexpression of NagR (TFI_1446) showed no fitness effect in SDMM + GlcNAc (Fig. 2E). Catabolism of mannose, however, is dependent on the

activity of ManA, which is not under CCR and does not have an alternative pathway. Therefore, overexpression of three transcription factors that upregulated *manA* expression in SDMM + glucose had positive fitness in TRIP screens in SDMM + mannose (Fig. 2F). A comparison of TRIP results in mannose, GlcNAc and galactose emphasize the impact that alternate pathways and overlapping regulation have on TRIP screens.

TRIP screens completed in four antibiotics (Ceftriaxone, Vancomycin, Kanamycin and Levofloxacin) demonstrate that overexpression of transcriptional regulators can influence susceptibility to antibiotic stress. The three TFI strains with a “Rapid Biphasic Lysis” (Chapter 2) growth profile, TFI_0058, TFI_1799 and TFI_1885, all had increased fitness to the cell wall synthesis inhibitors Ceftriaxone and Vancomycin (Fig. 3B), and negative fitness in both Kanamycin and Levofloxacin (Fig. 4). These strains overexpress the metabolic regulators CpsR, SusR and TreR, and TFI RNAseq revealed repression of the capsule operon and several peptidoglycan synthesis genes, as well as activation of the oxidative stress response (Fig. 3E). We hypothesize that the metabolic rewiring resulting from overexpression of these TFs increases cellular acidity and either directly or indirectly represses cell wall synthesis, which can also explain the rapid induction of autolysis by LytA in stationary phase. Further studies, particularly metabolomic approaches, are needed to fully connect these regulators and their effects on cellular acidity, cell wall expression/structure and their opposing fitness effects in specific antibiotic treatments.

Regulators of protease activity and metal homeostasis greatly influence susceptibility to the protein synthesis inhibitor Kanamycin (Fig. 4A). Overexpression of two protease activity regulators, CtsR and RitR, were shown to have reduced susceptibility to Kanamycin, along with regulators of copper homeostasis (CopY) and the thioredoxin

regulator (SP_1774, a MarR-family TF). Notably, several of these strains may be post-translationally regulated by metal concentrations and/or oxidative stress. CopY directly binds Cu²⁺^{97,98}, and MarR in *E. coli* has been shown to be regulated by Cu²⁺-catalyzed oxidation of a conserved cysteine residue⁵⁰. RitR, being the lone orphan response regulator in *S. pneumoniae*, is regulated via H₂O₂-mediated oxidation of Cys128 resulting in dimerization and DNA-binding¹⁷. Interestingly, CuCl₂ was the only metal salt shown to increase RitR dimerization, but had no effect on DNA-binding without H₂O₂¹⁷. Metals, oxidative stress and Kanamycin are intertwined through production of ROS by bactericidal antibiotics like Kanamycin through the iron-dependent Fenton reaction, as well as the manganese-dependent oxidative stress response and protein mismetallation. RitR regulates metal homeostasis, protease activity and detoxification of oxidative stress, granting TFI_0376 the greatest fitness advantage during Kanamycin treatment (Fig. 4B).

TRIP screens can also be used to identify fitness effects of transcription factor overexpression in specific stresses like oxidative stress (Fig. 5) and zinc intoxication (Fig. 6), both of which mimic mechanisms used by our immune systems to fight pathogenic bacteria. As expected, overexpression of transcriptional regulators of the oxidative stress response (NmlR and PadR) and metal homeostasis (SzczA and PsaR) had significant fitness effects under oxidative stress. PadR's regulon coordinates the detoxification of damaged membranes and initializes synthesis of replacement lipids, while oxidized NmlR activates expression of *adhC*, which has previously been shown to be required for surviving oxidative stress and virulence⁹⁹. The two metal regulators with significant fitness effects in oxidative stress are SzczA and PsaR. PsaR regulates manganese homeostasis, which is an

important antioxidant, while SzcA regulates zinc homeostasis which interferes with the oxidative stress response through mismetallation. Overexpression of the other zinc regulator, AdcR, had negative but insignificant fitness defects in the oxidative stress TRIP screens. We suspect metal dysregulation potentiates oxidative stress by both inhibiting the oxidative stress response and increasing ROS production, which is supported by observed increases in cell associated iron in TFI_1858 and TFI_2172 (Fig. 5C). Recently, the zinc ionophore PBT2 was shown to induce zinc toxicity and break ampicillin resistance in a pneumococcal lung infection¹⁰⁰. This was attributed to zinc's inhibition of GlmU, an essential enzyme involved in peptidoglycan synthesis. Our work suggests that zinc intoxication may also increase production of ROS and inhibits several proteins including GlmU, the manganese importer PsaABC¹⁰¹ and iron-sulfur cluster dependent enzymes. We confirmed that zinc potentiates the activity of cell wall synthesis inhibitors and show it also potentiates the activity of Kanamycin in wild-type TIGR4 (Fig. 11). Potentiation of protein synthesis inhibition appears to be much more dramatic than that of cell wall synthesis inhibition, likely due to zinc's capacity for mismetallation. Interestingly, *adhC* (*SP_1885*) is annotated as being zinc-dependent and is genetically linked to zinc homeostasis through *szcA* (*SP_1858*) and its primary target gene, the zinc efflux pump *czcD* (*SP_1857*).

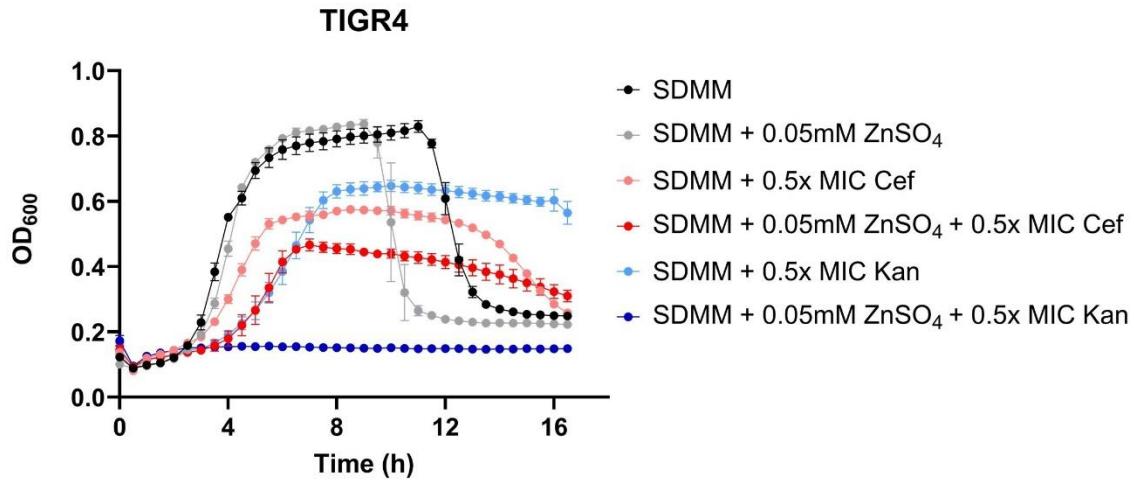


Figure 11. Zinc Potentiates Ceftriaxone and Kanamycin Treatments. Average of four replicates shown, error bars represent standard error. ZnSO₄ = Zinc sulfate, Cef = Ceftriaxone, Kan = Kanamycin.

TRIP screens assaying TIGR4's response to zinc intoxication revealed on TFI strain with drastically reduced fitness: TFI_2234 (Fig. 6). TFI_2234 overexpresses the Phage Infection Protein regulator, PipR, which regulates two transmembrane proteins that appear to serve as bacteriophage receptors. Zinc⁶⁰ and other metals⁶¹ have been shown to enhance bacteriophage infectivity of diverse bacterial species, suggesting that overexpression of PipR at elevated concentrations of zinc induces prophage induction. Interestingly, TFI_2234 had reduced but insignificant fitness during Levofloxacin treatment (Fig. 4), which aligns with previous studies that show fluoroquinolones are frequent prophage inducers⁶².

Lastly, we were able to adapt TRIP screens to time-kill assays to identify which transcription factors influence antibiotic tolerance (Fig. 8). Overall, tolerance TRIP screens in Levofloxacin and Ceftriaxone have high correlation, indicating the same mechanisms are used to survive intense antibiotic stress irrespective of initial target. The only TFI strain

with significantly increased tolerance to both antibiotics is TFI_2195, which overexpresses the ClpXP protease repressor, CtsR. This further supports the role of proteostasis in the induction of tolerance and persistence, as has been shown in both *E. coli*^{6,7} and *S. aureus*⁷⁷. Unsupervised clustering of TRIP screens shows Kanamycin fitness results clustering near tolerance fitness results (Fig. 10A). A loss of protease activity in tolerant cells would directly lead to an increase in competence and antibiotic resistance as the competence sigma factor, ComX, is a known target of ClpXP⁸¹.

Strains overexpressing several other redox-sensitive and metabolic regulators also increased antibiotic tolerance (Fig. 8). The orphan response regulator RitR and the *adhC*-regulator NmlR both had significant increases in tolerance and are both activated through oxidation of conserved cysteines. Overexpression of NmlR in sub-MIC Levofloxacin and under oxidative stress appeared to result in reduced fitness, but may have instead been a tolerance phenotype through reduced growth. Similarly, overexpression of PadR, which has a similar fitness profile to NmlR, displayed increased tolerance, albeit insignificant. Interestingly, lipid peroxidation and their toxic by-products have been proposed to be involved in the formation of therapy-induced senescence in cancer cells¹⁰². Overexpression of two metabolic regulators with slow growth phenotypes, NanR and DeoR, also lead to an increase in tolerance.

Six strains had significantly reduced tolerance to Ceftriaxone and/or Levofloxacin (Fig. 8C) including expected hits including TFI_0100 and TFI_1584, which overexpress the Phenotypic Tolerance to Vancomycin repressor, PtvR, and the global nutritional regulator, CodY. *ptvR* mutants were previously shown to have increased tolerance to Vancomycin

through derepression of the *ptv* operon⁷⁹. Here we have validated those findings and shown that overexpression of PtvR leads to decreased tolerance to both cell wall synthesis and DNA synthesis inhibitors. CodY plays a multifaceted role in regulating cellular metabolism, the oxidative stress response, and BCAA uptake and biosynthesis (Fig. 9). Supplementation of TIGR4-LacI with 5mM ILV increased survival of 8h of Ceftriaxone treatment 1,000-fold, likely due to the pre-induction of the oxidative stress response by CodY in response to both BCAA and H₂O₂. Interestingly, *S. pneumoniae* appears to be auxotrophic for BCAs in chemically defined media, but encodes the genes necessary for *de novo* synthesis⁹⁴. This could be due to *S. pneumoniae*'s production of as a virulence factor and allelopathic strategy, as production of H₂O₂ requires the reactant of BCAs, pyruvate, and results in oxidative stress which would inhibit the enzymes required for BCAA biosynthesis. Together this identifies CodY, the ABC uptake system for BCAs and iron-sulfur clusters as valuable drug targets.

The final four strains with decreased tolerance to antibiotics were the three TFI strains with the “Rapid Biphasic Lysis” growth profiles and the antitoxin, PezT. TFI_0058, TFI_1799, and TFI_1885 overexpress the metabolic regulators CpsR, SusR and TreR, respectively, and have shown significant increases in resistance to two cell wall synthesis inhibitors, Ceftriaxone and Vancomycin (Fig. 3) and significant increases in susceptibility to both Kanamycin and Levofloxacin (Fig. 4). We suspect that the metabolic rewiring by overexpressing these metabolic regulators leads to more acidic environment, leading to greater susceptibility to Kanamycin, Levofloxacin, autolysin activity and high concentrations of antibiotics. The final strain with significant decrease in tolerance was TFI_1050, which overexpresses the antitoxin, PezT. Toxin-antitoxin systems have long

been linked to antibiotic tolerance, and even though they are known to be redundant, overexpression of just one of the five antitoxins in TIGR4 lead to a significant decrease in tolerance.

Our TRIP screens have shown that transcription factor overexpression is a powerful tool to identify the effects transcriptional perturbation has on cell fitness in a variety of situations. UMAP representation of TRIP screen results compare how different stresses are sensed and responded to by *S. pneumoniae*, and further demonstrate the link between bactericidal antibiotics and oxidative stress as well as between dysfunctional proteostasis and antibiotic tolerance. Across antibiotic TRIP screens, metabolic regulators, metal ion regulators and oxidative stress response regulators were recurring significant hits, indicating the importance of these processes during antibiotic stress. This supports our hypothesis that a comprehensive knowledge of the regulatory mechanisms of *S. pneumoniae* can identify routes to antibiotic potentiation, such as specific metabolic stimulation or zinc intoxication. Identifying transcription factors or other regulatory mechanisms that are redox regulated, regulated by proteolysis or require metal cofactors are important next steps to mechanistically define the stress response. Targeting these transcription factors and/or their regulons will help overcome the challenges of antibiotic resistance.

4.7 Materials and Methods

TRIP Screens: TRIP starters were cultured in SDMM + 1mM IPTG at 37°C in 5% CO₂ until exponential phase before being pelleted by centrifugation, washed in PBS and re-inoculated into tested conditions at an OD of 0.1. All tested conditions contained 1mM IPTG and each experiment had three or more biological replicates. At each timepoint indicated, 5mL of cells were removed from each replicate, pelleted and frozen. DNA was extracted from each sample using the Qiagen DNeasy kit with a Gram-positive lysis buffer (Tris 10 mM pH 7.9; EDTA 10 mM; Tween 0.1%, Triton 1.2 %) and the addition of RNase and Proteinase K degradation steps. 10ng of chromosomal DNA from each sample was input into the first PCR with primers TRIP_F1 and TRIP_R1 and 10 PCR-cycles. DNA was purified using 1.8x AxyMag PCR clean up with 38uL elutant, which was all input into a second PCR, again with 10 cycles and now using PCR primers TRIP_D5xx and TRIP_D7xx to add Illumina adapter and index sequences. The final PCR product was purified with 1x AxyMag PCR clean up and eluted in 10uL. Each amplicon was quantified and pooled for sequencing on Illumina NextSeq 2000 sequenced 75:8:8 with 40% PhiX spike. Bartender was used to identify and quantify barcode counts within each Fastq file, and each sample's barcode counts were matched to each TFI strains known barcodes to create a barcode count matrix. Barcode counts were converted to Reads per Million (RPM) according to the equation:

$$\frac{BC_{i,j}}{\left(\sum_{i=1}^n BC_{i,j} \right) / 1,000,000} = RPM_{i,j}$$

Fitness (W) of each sample was calculated as the fold-change in abundance relative to time 0:

$$W_{i,j} = \frac{RPM_{i,j}}{Avg(RPM)_{i,j=0}}$$

An unpaired, two-sided t-test was performed on the RPM values of each biological replicate relative to the RPM values of the same condition at time zero and adjusted for multiple hypotheses testing using the Benjamini-Hochberg method. Results were considered significant if the average fitness of the biological replicates was equal to or

greater than two standard deviations away from the mean and the p-adjusted value was less than 0.05.

Tolerance TRIP Screens: TRIP starters were cultured in SDMM + 1mM IPTG at 37°C in 5% CO₂ until exponential phase before being pelleted by centrifugation, washed in PBS and re-inoculated as biological triplicates into fresh media at an OD₆₀₀ of 0.1. After 1h of growth to ensure exponential growth, high concentrations of antibiotics (5x, 10x, or 50x MIC) were added to each culture. At each timepoint, 1mL of culture was removed, cells were pelleted by centrifugation and washed with PBS. Cells were then serially diluted in PBS and drip-plated onto TSA blood agar plates, and the remaining cells were plated on a second TSA blood agar plate for recovery. Both plates were incubated overnight at 37°C in 5% CO₂. The following day, colonies were counted from the drip plate to determine CFU counts and cells were scraped from the recovery plate for gDNA extraction and barcode sequencing. Library prep, sequencing and data analysis is the same as previous TRIP screens.

Minimum Inhibitory Concentrations (MIC):

Antibiotic	MIC ($\mu\text{g/mL}$)
Ceftriaxone	0.02
Vancomycin	0.39
Kanamycin	150
Levofloxacin	2

Growth Curves: Individual strain starters were cultured in SDMM + 1mM IPTG at 37°C in 5% CO₂ until exponential phase before being pelleted by centrifugation, washed in PBS and re-suspended in PBS to an OD₆₀₀ of 0.2. Biological triplicates of cells were diluted 1:40 in 96-well plates and cultured in the BioSpa 8 plate readers (Biotek) with OD₆₀₀ readings every 30 minutes. Growth profiles were plotted with Graphpad Prism.

Inductively Coupled Plasma Mass Spectrometry (ICP-MS): Individual strain starters were cultured in SDMM + 1mM IPTG at 37°C in 5% CO₂ until exponential phase before being pelleted by centrifugation and re-inoculated in SDMM + 1mM IPTG for 2h. Cell were pelleted by centrifugation, transferred to pre-weighed Eppendorf tube, washed twice with PBS + 0.5mM EDTA and once with pre-chelated PBS (Chelex 100, Bio Rad). Cell pellets were dried overnight at 95°C. Dried pellets were solubilized in 30% TraceMetal Nitric Acid (Fisher Chemical), boiled for 10 minutes at 95°C and diluted to 5% TraceMetal Nitric Acid in 10mL. Samples were run on Agilent 8900 ICPMS in solution mode.

Statistical Analysis: Unifold Manifold Approximation and Projection was carried out in R (4.4.2) using the “umap” package and Gene Ontology enrichment by calculated by FUANGE-Pro. Plots were made with the “ggplot2” package.

4.8 References

1. ABRAHAM, E. P. & CHAIN, E. An Enzyme from Bacteria able to Destroy Penicillin. *Nature* **146**, 837–837 (1940).
2. Lobanovska, M. & Pilla, G. Penicillin’s Discovery and Antibiotic Resistance: Lessons for the Future? *Yale J. Biol. Med.* **90**, 135–145 (2017).
3. Kayser, F. H., Benner, E. J., Troy, R. & Hoeprich, P. D. MODE OF RESISTANCE AGAINST β -LACTAM ANTIBIOTICS IN STAPHYLOCOCCI. *Ann. N. Y. Acad. Sci.* **182**, 106–117 (1971).
4. Long, Y. *et al.* Identification of novel genes that promote persister formation by repressing transcription and cell division in *Pseudomonas aeruginosa*. *J. Antimicrob. Chemother.* **74**, 2575–2587 (2019).
5. Huang, C. Y., Gonzalez-Lopez, C., Henry, C., Mijakovic, I. & Ryan, K. R. hipBA toxin-antitoxin systems mediate persistence in *Caulobacter crescentus*. *Sci. Rep.* **10**, 2865 (2020).
6. Blattman, S. B. *et al.* Identification and genetic dissection of convergent persister cell states. *Nature* **636**, 438–446 (2024).
7. Pu, Y. *et al.* ATP-Dependent Dynamic Protein Aggregation Regulates Bacterial Dormancy Depth Critical for Antibiotic Tolerance. *Mol. Cell* **73**, 143–156.e4 (2019).
8. Wu, S., Yu, P.-L., Wheeler, D. & Flint, S. Transcriptomic study on persistence and survival of *Listeria monocytogenes* following lethal treatment with nisin. *J. Glob. Antimicrob. Resist.* **15**, 25–31 (2018).
9. Levin-Reisman, I. *et al.* Antibiotic tolerance facilitates the evolution of resistance. <http://science.sciencemag.org/>.
10. Zhu, Z. *et al.* Entropy of a bacterial stress response is a generalizable predictor for fitness and antibiotic sensitivity. *Nat. Commun.* **11**, (2020).
11. Allison, K. R., Brynildsen, M. P. & Collins, J. J. Metabolite-enabled eradication of bacterial persisters by aminoglycosides. *Nature* **473**, 216–220 (2011).
12. Meylan, S., Andrews, I. W. & Collins, J. J. Targeting Antibiotic Tolerance, Pathogen by Pathogen. *Cell* **172**, 1228–1238 (2018).
13. Ma, S. *et al.* Transcriptional regulator-induced phenotype screen reveals drug potentiators in *Mycobacterium tuberculosis*. *Nat. Microbiol.* **6**, (2021).
14. Schulz, C. *et al.* Regulation of the Arginine Deiminase System by ArgR2 Interferes with Arginine Metabolism and Fitness of *Streptococcus pneumoniae*. *MBio* **5**, (2014).
15. Yesilkaya, H. *et al.* Pyruvate Formate Lyase Is Required for Pneumococcal Fermentative Metabolism and Virulence. *Infect. Immun.* **77**, 5418–5427 (2009).
16. Ulijasz, A. T., Andes, D. R., Glasner, J. D. & Weisblum, B. Regulation of Iron Transport in *Streptococcus pneumoniae* by RitR, an Orphan Response Regulator. *J. Bacteriol.* **186**, 8123–8136 (2004).
17. Glanville, D. G. *et al.* RitR is an archetype for a novel family of redox sensors in the streptococci that has evolved from two-component response regulators and is required for pneumococcal colonization. *PLOS Pathog.* **14**, e1007052 (2018).
18. Bortoni, M. E., Terra, V. S., Hinds, J., Andrew, P. W. & Yesilkaya, H. The pneumococcal response to oxidative stress includes a role for Rgg. *Microbiology* **155**, 4123–4134 (2009).
19. Hendriksen, W. T. *et al.* Site-Specific Contributions of Glutamine-Dependent Regulator GlnR and

- GlnR-Regulated Genes to Virulence of *Streptococcus pneumoniae*. *Infect. Immun.* **76**, 1230–1238 (2008).
20. Kloosterman, T. G. *et al.* Regulation of Glutamine and Glutamate Metabolism by GlnR and GlnA in *Streptococcus pneumoniae*. *J. Biol. Chem.* **281**, 25097–25109 (2006).
 21. Bidossi, A. *et al.* A Functional Genomics Approach to Establish the Complement of Carbohydrate Transporters in *Streptococcus pneumoniae*. *PLoS One* **7**, e33320 (2012).
 22. Carvalho, S. M., Kloosterman, T. G., Kuipers, O. P. & Neves, A. R. CcpA Ensures Optimal Metabolic Fitness of *Streptococcus pneumoniae*. *PLoS One* **6**, e26707 (2011).
 23. Wu, K. *et al.* CpsR, a GntR family regulator, transcriptionally regulates capsular polysaccharide biosynthesis and governs bacterial virulence in *Streptococcus pneumoniae*. *Sci. Rep.* **6**, 29255 (2016).
 24. Afzal, M., Shafeeq, S. & Kuipers, O. P. LacR Is a Repressor of lacABCD and LacT Is an Activator of lacTFEG , Constituting the lac Gene Cluster in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* **80**, 5349–5358 (2014).
 25. Aprianto, R., Slager, J., Holsappel, S. & Veening, J.-W. High-resolution analysis of the pneumococcal transcriptome under a wide range of infection-relevant conditions. *Nucleic Acids Res.* (2018) doi:10.1093/nar/gky750.
 26. Afzal, M., Shafeeq, S., Manzoor, I., Henriques-Normark, B. & Kuipers, O. P. N-acetylglucosamine-Mediated Expression of nagA and nagB in *Streptococcus pneumoniae*. *Front. Cell. Infect. Microbiol.* **6**, (2016).
 27. McCluskey, J., Hinds, J., Husain, S., Witney, A. & Mitchell, T. J. A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in *Streptococcus pneumoniae*. *Mol. Microbiol.* **51**, 1661–1675 (2004).
 28. Andreassen, P. R. *et al.* Host-glycan metabolism is regulated by a species-conserved two-component system in *Streptococcus pneumoniae*. *PLOS Pathog.* **16**, e1008332 (2020).
 29. Pettersen, J. S., Nielsen, F. D., Andreassen, P. R., Møller-Jensen, J. & Jørgensen, M. G. A comprehensive analysis of pneumococcal two-component system regulatory networks. *NAR Genomics Bioinforma.* **6**, (2024).
 30. Šoltysová, M. *et al.* Structural insight into DNA recognition by bacterial transcriptional regulators of the SorC/DeoR family. *Acta Crystallogr. Sect. D Struct. Biol.* **77**, 1411–1424 (2021).
 31. Liang, Z. *et al.* Glycerol metabolic repressor GlpR contributes to *Streptococcus suis* oxidative stress resistance and virulence. *Microbes Infect.* **27**, 105307 (2025).
 32. Turner, S. E. *et al.* A DeoR-Type Transcription Regulator Is Required for Sugar-Induced Expression of Type III Secretion-Encoding Genes in *Pseudomonas syringae* pv. *tomato* DC3000. *Mol. Plant-Microbe Interact.* **33**, 509–518 (2020).
 33. Okado-Matsumoto, A. & Fridovich, I. The Role of α,β -Dicarbonyl Compounds in the Toxicity of Short Chain Sugars. *J. Biol. Chem.* **275**, 34853–34857 (2000).
 34. Fritsch, V. N. *et al.* The MerR-family regulator NmlR is involved in the defense against oxidative stress in *Streptococcus pneumoniae*. *Mol. Microbiol.* **119**, 191–207 (2023).
 35. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **130**, 797–810 (2007).
 36. Garde, S., Chodisetti, P. K. & Reddy, M. Peptidoglycan: Structure, Synthesis, and Regulation. *EcoSal Plus* **9**, (2021).

37. El Khoury, J. Y., Boucher, N., Bergeron, M. G., Leprohon, P. & Ouellette, M. Penicillin induces alterations in glutamine metabolism in *Streptococcus pneumoniae*. *Sci. Rep.* **7**, 14587 (2017).
38. Hutchings, M. I., Truman, A. W. & Wilkinson, B. Antibiotics: past, present and future. *Curr. Opin. Microbiol.* **51**, 72–80 (2019).
39. HAKENBECK, R. & KOHIYAMA, M. Purification of Penicillin-Binding Protein 3 from *Streptococcus pneumoniae*. *Eur. J. Biochem.* **127**, 231–236 (1982).
40. Filipe, S. R. & Tomasz, A. Inhibition of the expression of penicillin resistance in *Streptococcus pneumoniae* by inactivation of cell wall muropeptide branching genes. *Proc. Natl. Acad. Sci.* **97**, 4891–4896 (2000).
41. Neta, T., Takada, K. & Hirasawa, M. Low-cariogenicity of trehalose as a substrate. *J. Dent.* **28**, 571–576 (2000).
42. Guan, N. & Liu, L. Microbial response to acid stress: mechanisms and applications. *Appl. Microbiol. Biotechnol.* **104**, 51–65 (2020).
43. Mueller, E. A. & Levin, P. A. Bacterial Cell Wall Quality Control during Environmental Stress. *MBio* **11**, (2020).
44. Aggarwal, S. D. *et al.* A molecular link between cell wall biosynthesis, translation fidelity, and stringent response in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci.* **118**, (2021).
45. Mellroth, P. *et al.* LytA, Major Autolysin of *Streptococcus pneumoniae*, Requires Access to Nascent Peptidoglycan. *J. Biol. Chem.* **287**, 11018–11029 (2012).
46. Cortes, P. R., Piñas, G. E., Cian, M. B., Yandar, N. & Echenique, J. Stress-triggered signaling affecting survival or suicide of *Streptococcus pneumoniae*. *Int. J. Med. Microbiol.* **305**, 157–169 (2015).
47. John, T. *et al.* How kanamycin A interacts with bacterial and mammalian mimetic membranes. *Biochim. Biophys. Acta - Biomembr.* **1859**, 2242–2252 (2017).
48. Muñoz, R. & De La Campa, A. G. ParC subunit of DNA topoisomerase IV of *Streptococcus pneumoniae* is a primary target of fluoroquinolones and cooperates with DNA gyrase A subunit in forming resistance phenotype. *Antimicrob. Agents Chemother.* **40**, 2252–2257 (1996).
49. Ferrández, M.-J. & de la Campa, A. G. The Fluoroquinolone Levofloxacin Triggers the Transcriptional Activation of Iron Transport Genes That Contribute to Cell Death in *Streptococcus pneumoniae*. *Antimicrob. Agents Chemother.* **58**, 247–257 (2014).
50. Hao, Z. *et al.* The multiple antibiotic resistance regulator MarR is a copper sensor in *Escherichia coli*. *Nat. Chem. Biol.* **10**, 21–28 (2014).
51. Kunkle, D. E. & Skaar, E. P. Moving metals: How microbes deliver metal cofactors to metalloproteins. *Mol. Microbiol.* **120**, 547–554 (2023).
52. Stocks, C. J. *et al.* Frontline Science: LPS-inducible SLC30A1 drives human macrophage-mediated zinc toxicity against intracellular *Escherichia coli*. *J. Leukoc. Biol.* **109**, 287–297 (2021).
53. Rychel, K. *et al.* Laboratory evolution, transcriptomics, and modeling reveal mechanisms of paraquat tolerance. *Cell Rep.* **42**, 113105 (2023).
54. Tatjana, V., Domitille, S. & Jean-Charles, S. Paraquat-induced cholesterol biosynthesis proteins dysregulation in human brain microvascular endothelial cells. *Sci. Rep.* **11**, 18137 (2021).
55. Hernandez-Morfa, M. *et al.* The oxidative stress response of *Streptococcus pneumoniae*: its contribution to both extracellular and intracellular survival. *Front. Microbiol.* **14**, (2023).
56. Araya-Cloutier, C., Vincken, J.-P., van Ederen, R., den Besten, H. M. W. & Gruppen, H. Rapid

- membrane permeabilization of *Listeria monocytogenes* and *Escherichia coli* induced by antibacterial prenylated phenolic compounds from legumes. *Food Chem.* **240**, 147–155 (2018).
57. Aribisala, J. O., S'thebe, N. W. & Sabiu, S. In silico exploration of phenolics as modulators of penicillin binding protein (PBP) 2 \times of *Streptococcus pneumoniae*. *Sci. Rep.* **14**, 8788 (2024).
 58. Kloosterman, T. G., van der Kooi-Pol, M. M., Bijlsma, J. J. E. & Kuipers, O. P. The novel transcriptional regulator SczA mediates protection against Zn 2+ stress by activation of the Zn 2+ - resistance gene czcD in *Streptococcus pneumoniae*. *Mol. Microbiol.* **65**, (2007).
 59. Sullivan, M. J., Goh, K. G. K. & Ulett, G. C. Cellular Management of Zinc in Group B *Streptococcus* Supports Bacterial Resistance against Metal Intoxication and Promotes Disseminated Infection. *mSphere* **6**, (2021).
 60. Li, X. *et al.* Exploring the Benefits of Metal Ions in Phage Cocktail for the Treatment of Methicillin-Resistant *Staphylococcus aureus* (MRSA) Infection. *Infect. Drug Resist.* **Volume 15**, 2689–2702 (2022).
 61. Ma, L., Green, S. I., Trautner, B. W., Ramig, R. F. & Maresso, A. W. Metals Enhance the Killing of Bacteria by Bacteriophage in Human Blood. *Sci. Rep.* **8**, 2326 (2018).
 62. Henrot, C. & Petit, M.-A. Signals triggering prophage induction in the gut microbiota. *Mol. Microbiol.* **118**, 494–502 (2022).
 63. Fritsch, V. N. *et al.* The MerR -family regulator NmlR is involved in the defense against oxidative stress in *Streptococcus pneumoniae*. *Mol. Microbiol.* **119**, 191–207 (2023).
 64. Echave, P., Tamarit, J., Cabiscol, E. & Ros, J. Novel Antioxidant Role of Alcohol Dehydrogenase E from *Escherichia coli*. *J. Biol. Chem.* **278**, 30193–30198 (2003).
 65. Bigger, J. TREATMENT OF STAPHYLOCOCCAL INFECTIONS WITH PENICILLIN BY INTERMITTENT STERILISATION. *Lancet* **244**, 497–500 (1944).
 66. Huemer, M., Mairpady Shambat, S., Brugger, S. D. & Zinkernagel, A. S. Antibiotic resistance and persistence—Implications for human health and treatment perspectives. *EMBO Rep.* **21**, (2020).
 67. Moyed, H. S. & Bertrand, K. P. *hipA*, a newly recognized gene of *Escherichia coli* K-12 that affects frequency of persistence after inhibition of murein synthesis. *J. Bacteriol.* **155**, 768–775 (1983).
 68. Kaspy, I. *et al.* *HipA*-mediated antibiotic persistence via phosphorylation of the glutamyl-tRNA-synthetase. *Nat. Commun.* **4**, 3001 (2013).
 69. Xie, Y. *et al.* TADB 2.0: an updated database of bacterial type II toxin–antitoxin loci. *Nucleic Acids Res.* **46**, D749–D753 (2018).
 70. Fraikin, N., Goormaghtigh, F. & Van Melderen, L. Type II Toxin-Antitoxin Systems: Evolution and Revolutions. *J. Bacteriol.* **202**, (2020).
 71. Chan, W. T. & Espinosa, M. The *Streptococcus pneumoniae* *pezAT* Toxin–Antitoxin System Reduces β-Lactam Resistance and Genetic Competence. *Front. Microbiol.* **7**, (2016).
 72. Houri, H., Ghalavand, Z., Faghihloo, E., Fallah, F. & Mohammadi-Yeganeh, S. Exploiting *yoeB* - *yefM* toxin-antitoxin system of *Streptococcus pneumoniae* on the selective killing of miR-21 overexpressing breast cancer cell line (MCF-7). *J. Cell. Physiol.* **235**, 2925–2936 (2020).
 73. Salzer, A. & Wolz, C. Role of (p)ppGpp in antibiotic resistance, tolerance, persistence and survival in Firmicutes. *microLife* **4**, (2023).
 74. Zhu, J. *et al.* Co-regulation of *CodY* and (p)ppGpp synthetases on morphology and pathogenesis of *Streptococcus suis*. *Microbiol. Res.* **223–225**, 88–98 (2019).

75. Geiger, T. & Wolz, C. Intersection of the stringent response and the CodY regulon in low GC Gram-positive bacteria. *Int. J. Med. Microbiol.* **304**, 150–155 (2014).
76. Ahlawat, S. & Morrison, D. A. ClpXP Degrades SsrA-Tagged Proteins in *Streptococcus pneumoniae*. *J. Bacteriol.* **191**, 2894–2898 (2009).
77. Conlon, B. P. *et al.* Activated ClpP kills persisters and eradicates a chronic biofilm infection. *Nature* **503**, 365–370 (2013).
78. Balaban, N. Q. *et al.* Definitions and guidelines for research on antibiotic persistence. *Nat. Rev. Microbiol.* **17**, 441–448 (2019).
79. Liu, X. *et al.* Transcriptional Repressor PtVR Regulates Phenotypic Tolerance to Vancomycin in *Streptococcus pneumoniae*. *J. Bacteriol.* **199**, (2017).
80. Eagle, H. A Paradoxical Zone Phenomenon in the Bactericidal Action of Penicillin in Vitro. *Science (80-.).* **107**, 44–45 (1948).
81. Prasetyoputri, A., Jarrad, A. M., Cooper, M. A. & Blaskovich, M. A. T. The Eagle Effect and Antibiotic-Induced Persistence: Two Sides of the Same Coin? *Trends Microbiol.* **27**, 339–354 (2019).
82. Deter, H. S., Hossain, T. & Butzin, N. C. Antibiotic tolerance is associated with a broad and complex transcriptional response in *E. coli*. *Sci. Rep.* **11**, (2021).
83. Tovpeko, Y., Bai, J. & Morrison, D. A. Competence for Genetic Transformation in *Streptococcus pneumoniae*: Mutations in σ A Bypass the ComW Requirement for Late Gene Expression. *J. Bacteriol.* **198**, 2370–2378 (2016).
84. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in *Streptococcus pneumoniae*. *Mol. Syst. Biol.* **13**, (2017).
85. Afzal, M., Shafeeq, S., Ahmed, H. & Kuipers, O. P. Sialic Acid-Mediated Gene Expression in *Streptococcus pneumoniae* and Role of NanR as a Transcriptional Activator of the nan Gene Cluster. *Appl. Environ. Microbiol.* **81**, 3121–3131 (2015).
86. Gualdi, L. *et al.* Regulation of neuraminidase expression in *Streptococcus pneumoniae*. *BMC Microbiol.* **12**, 200 (2012).
87. Dalia, A. B., Standish, A. J. & Weiser, J. N. Three Surface Exoglycosidases from *Streptococcus pneumoniae*, NanA, BgaA, and StrH, Promote Resistance to Opsonophagocytic Killing by Human Neutrophils. *Infect. Immun.* **78**, 2108–2116 (2010).
88. Si, M. *et al.* MsrR is a thiol-based oxidation-sensing regulator of the XRE family that modulates *C. glutamicum* oxidative stress resistance. *Microb. Cell Fact.* **19**, 189 (2020).
89. Hu, Y. *et al.* The XRE Family Transcriptional Regulator SrtR in *Streptococcus suis* Is Involved in Oxidant Tolerance and Virulence. *Front. Cell. Infect. Microbiol.* **8**, (2019).
90. Culligan, E. P., Sleator, R. D., Marchesi, J. R. & Hill, C. Metagenomic Identification of a Novel Salt Tolerance Gene from the Human Gut Microbiome Which Encodes a Membrane Protein with Homology to a brp/blh-Family β-Carotene 15,15'-Monooxygenase. *PLoS One* **9**, e103318 (2014).
91. Giani, M. & Martínez-Espinosa, R. M. Carotenoids as a Protection Mechanism against Oxidative Stress in *Haloferax mediterranei*. *Antioxidants* **9**, 1060 (2020).
92. Dieser, M., Greenwood, M. & Foreman, C. M. Carotenoid Pigmentation in Antarctic Heterotrophic Bacteria as a Strategy to Withstand Environmental Stresses. *Arctic, Antarct. Alp. Res.* **42**, 396–405 (2010).
93. Hajaj, B. *et al.* CodY Regulates Thiol Peroxidase Expression as Part of the Pneumococcal Defense

- Mechanism against H₂O₂ Stress. *Front. Cell. Infect. Microbiol.* **7**, (2017).
- 94. Härtel, T. *et al.* Characterization of Central Carbon Metabolism of *Streptococcus pneumoniae* by Isotopologue Profiling. *J. Biol. Chem.* **287**, 4260–4274 (2012).
 - 95. Macomber, L. & Imlay, J. A. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc. Natl. Acad. Sci.* **106**, 8344–8349 (2009).
 - 96. Sullivan, M. J., Terán, I., Goh, K. G. K. & Ulett, G. C. Resisting death by metal: metabolism and Cu/Zn homeostasis in bacteria. *Emerg. Top. Life Sci.* **8**, 45–56 (2024).
 - 97. Novichkov, P. S. *et al.* RegPrecise 3.0 – A resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* **14**, 745 (2013).
 - 98. O'Brien, H. *et al.* Rules of Expansion: an Updated Consensus Operator Site for the CopR-CopY Family of Bacterial Copper Exporter System Repressors. *mSphere* **5**, (2020).
 - 99. Stroher, U. H. *et al.* A Pneumococcal MerR-Like Regulator and S-nitrosoglutathione Reductase Are Required for Systemic Virulence. *J. Infect. Dis.* **196**, 1820–1826 (2007).
 - 100. Brazel, E. B. *et al.* Dysregulation of *Streptococcus pneumoniae* zinc homeostasis breaks ampicillin resistance in a pneumonia infection model. *Cell Rep.* **38**, 110202 (2022).
 - 101. McDevitt, C. A. *et al.* A Molecular Mechanism for Bacterial Susceptibility to Zinc. *PLoS Pathog.* **7**, e1002357 (2011).
 - 102. Flor, A. C. & Kron, S. J. Lipid-derived reactive aldehydes link oxidative stress to cell senescence. *Cell Death Dis.* **7**, e2366–e2366 (2016).

Chapter 5:

Conclusion

5.0 Motivation

Antibiotic resistance poses a dire threat to global health, undermining our ability to treat bacterial infections worldwide. The emergence of multidrug-resistant strains has rendered many antibiotics ineffective, necessitating the development of novel antibiotic strategies. While traditional approaches focus on discovering new antibiotics or inhibiting resistance mechanisms^{1–4}, an alternative and complementary strategy lies in targeting bacterial transcriptional regulation. Transcription factors play a crucial role in modulating virulence, stress responses, and antibiotic tolerance, making them promising drug targets^{5,6}. By inhibiting virulence-regulating transcription factors, we can attenuate pathogenicity, reducing the burden of infection. Additionally, a deeper understanding of transcriptional regulatory networks (TRNs) can enable the rational manipulation of bacterial physiology to potentiate the efficacy of existing antibiotics. Both approaches leverage transcriptional regulation as a means to combat antibiotic resistance, either by identifying novel drug targets or by enhancing antibiotic activity against them.

Unfortunately, beyond the most common model systems—*E. coli* and *S. aureus*—comprehensive transcriptional regulatory networks don't exist for bacterial pathogens including *S. pneumoniae*. *S. pneumoniae* is an opportunistic pathogen commonly found as a commensal in the nasopharynx but can invade disparate niches within the body to cause infection, a process reliant on transcriptional regulation of metabolic, virulence and stress response pathways. Alarmingly, *S. pneumoniae* is estimated to be responsible for over 1 million deaths per year globally⁷, and 30% of pneumococcal infections in the United States are resistant to one or more clinically relevant antibiotics⁸. This work aimed to both

characterize the TRN of *S. pneumoniae* and quantify the fitness effects that different states of the TRN have during antibiotic treatment to identify novel drug targets and antibiotic potentiators.

5.1 Identification and Classification of Transcription Factors in TIGR4

We begin by identifying, classifying and constructing an overexpression strain for all transcription factors our model strain of *S. pneumoniae*, TIGR4. In total, we identified 110 transcription factors split across 20 transcription factor families. 21% of all transcription factors belong to either the Xenobiotic Response Element (XRE) or Multiple Antibiotic Response Regulator (MarR) families, indicating a large portion of the TRN is devoted to overcoming stress. 58% of the transcription factors in TIGR4 are in *S. pneumoniae*'s core genome and only three are universally essential: *rpoD*, *codY* and *vicR*⁹. *RpoD* encodes the housekeeping sigma factor required for transcription, *codY* encodes a global nutritional regulator that's heavily linked with the oxidative stress response and *vicR* encodes a response regulator required for peptidoglycan synthesis and cell division¹⁰. These three genes are potential drug targets, with functional overlap between targeting *rpoD* and *vicR* with existing RNA-synthesis inhibitors and cell wall synthesis inhibitors. The link between metabolism, the oxidative stress response and overall fitness exemplified by *codY* is a recurrent theme throughout this work.

5.2 Constructing and Validating Transcription Factor Induction Strains

To achieve our aims of first reverse engineering the TRN of *S. pneumoniae* and manipulating the TRN to identify states susceptible to antibiotics, we cloned Transcription Factor Induction (TFI) strains for 89 TFs. Each TFI strain encodes a transcription factor

under an inducible promoter with a C-terminal HA tag and a unique DNA barcode. For our synthetic system, we decided to use the IPTG-inducible PL8-2 promoter for its large dynamic range and to avoid potential confusion using an ATc-inducible promoter during antibiotic assays¹¹. The system was validated using TFI-specific qPCR and Western blot, and growth curve profiling of all TFI strains reveal transcription factor overexpression as capable of causing quantifiable phenotypic changes. The growth profiles of TFI strains could be classified into one or more categories, indicative of converging phenotypes based on different perturbations of the TRN. The most dramatic changes seen in growth profiles were the length of the stationary phase and the induction of autolysis, with the most extreme being the three Rapid Biphasic Lysis strains which overexpress the metabolic regulators CpsR, TreR and SusR. There are differences between the growth profiles of TFI strains without IPTG, indicative of leaky expression. This may have been reduced had we used ATc-inducible promoters, but then we'd have multiple antibiotic stress signals during our antibiotic assays.

5.3 Reverse Engineering the Transcriptional Regulatory Network of TIGR4

The original plan to reverse engineer a comprehensive TRN for *S. pneumoniae* was to subject all TFI strains to both gene expression profiling by RNAseq and DNA-binding analysis by ChIPseq. RNAseq would identify candidate target genes by their differential expression, and ChIPseq would differentiate direct from indirect regulation by identifying binding motifs in true target promoters. All experiments take place in exponential growth in rich media, ensuring a consistent, steady state environment. However, this presented several challenges to overcome.

5.3.1 TFI RNAseq Fails to Identify All Regulatory Interactions

First, RNAseq on overexpressed transcription factors failed to identify many known target genes as differentially expressed genes (DEGs). In many cases, true target genes were significantly regulated according to p-adjusted values (Padjs), but failed to reach the log₂fold-change cutoff. In other cases, known interactions were missed altogether. This is likely due to a combination of factors, including regulation by the native copy of the transcription factor, regulation by other transcription factors, the inherent mean-dependent variance in gene expression and testing in a single environment. In the case of repressors, if the native copy of the transcription factor is repressing a target gene, then overexpression of the transcription factor won't affect the target gene's expression because it is already repressed. In fact, the majority of missed interactions in our network were from one of two global metabolic regulators, CodY and CcpA, both of which are expected to serve primarily as repressors in rich media. Also, the activity of transcription factors is typically regulated through binding a specific molecule or post-translational modification, often through a phosphorylation cascade. By performing RNAseq in rich media, transcription factors that respond to signals stemming from nutritional depletion or stress conditions were likely inactive, leading to missed interactions. For these reasons, we had to find a complimentary approach to identify target genes in our gene expression data.

5.3.2 An Ensemble of Network Inference Methods Complements TFI RNAseq

A complimentary approach to identify regulatory interactions is to computationally predict interactions from gene expression data¹². High quality, diverse transcriptomic data is critical to the success of this approach, so we combined all TFI RNAseq experiments with

our lab's previous RNAseq experiments. This dataset is comprised of 604 total RNAseq samples, including 271 TFI samples, 250 antibiotic samples, 71 control samples and 12 *in vivo* samples. Concerned with the noise stemming from the large percentage of samples with antibiotic treatment, we pruned this dataset into two smaller subsets, first by removing outlier or misclustering samples, and second by removing non-centroid replicates. Analysis of all three datasets revealed that the initial dataset was best, indicating that the addition of more data outweighs the reduction of noise.

Constrained optimization of which inference methods to use identified the optimal network stemming from a combination of the GENIE3, ARACNE, CLR and GRNBOOST2 methods. This network contains 11,228 predicted regulatory interactions with 28.4% recall, 2.64% precision and an AUPR of 5.5%. While that is low in terms of classifiers, it is within the range of inference methods used on the *E. coli* TRN and is not only evaluating our network construction but also the feasibility of mapping a TRN from gene expression data alone. Still, the low recall is indicative of missing interactions and the low precision is indicative of false positives. To improve recall, we combined predicted interactions in the network with all significantly regulated genes (according to adjusted p-values only) identified by TFI RNAseq. A more diverse dataset including RNAseq samples with alternate carbon or nitrogen sources, metabolite depletions or specific stressors like zinc intoxication could improve network inference by representing the transcriptome in more states.

5.4 Network Precision Improves with Requiring DNA Binding Motifs

To minimize false positives in our network, we required the presence of a binding motif within a promoter region to call a true interaction as transcription factors must bind specific

DNA sequences to influence gene expression. We performed ChIPseq on six TFI strains that overexpress transcription factors with varying expression levels and known binding motifs, using the same conditions as TFI RNAseq. Unfortunately, we were only able to identify binding motifs for two transcription factors: CcpA and CpsR, both of which matched their known motif from literature. Both CcpA and CpsR are global metabolic regulators that are expected to be repressors during exponential growth in rich media, which would increase the likelihood of capturing DNA::TF complexes by immunoprecipitation. ChIPseq is a time-consuming, laborious and costly technique that failed to identify binding motifs for 67% of TFI strains tested, necessitating an alternate approach identify binding motifs. Alternate *in vitro* methods like DNase Footprinting or DAPseq may have been better suited to identify the binding motifs of all transcription factors.

Computationally we were able to predict the binding motifs by interrogating suspected target promoter regions for enriched motifs. The promoters of all candidate target genes identified by TFI RNAseq or our ensemble of inference methods were analyzed XSTREME, a comprehensive motif discovery algorithm^{13,14}. Fortunately, the motifs identified by XSTREME for CcpA and CpsR matched those identified via ChIPseq and the known literature values. We subsetted the list of potential interactions by keeping only interactions with predicted motifs, improving precision to 4.75% (from 2.64%) and recall to 43.4% (from 28.4%). The improved precision demonstrates the value in removing false positives by requiring a binding motif and the improved recall demonstrates the importance of using both TFI RNAseq and network inference methods. Predicting binding motifs for each transcription factor also allows for mutational analysis to discover how non-coding

mutations may influence *S. pneumoniae* physiology specifically in antibiotic resistant strains.

Our network still contained many false positives, as evidenced by both its total number of interactions and low precision. Assuming the TRN of *S. pneumoniae* has a similar distribution of interactions as *E. coli*, we can expect approximately 4,000 interactions (~2 for every gene)^{15,16}. We therefore decided to prune the network further using motif data, RNAseq data, inference method rankings, natural expression of each transcription factor, literature sources and genomic context. After manually adding interactions from the eight transcription factors for which we could not identify a binding motif, our final network achieved a precision of 14.4% and a recall of 54.1%. Still, not all interactions were found, as evidenced by our recall, so known interactions were added to make the most complete network possible.

5.5 The Transcriptional Regulatory Network of TIGR4

The transcriptional regulatory network of TIGR4 that we assembled through the combination of TFI RNAseq, an ensemble of network inference methods and the identification of predicted binding sites contains 3,946 transcription factor – target gene interactions involving 88 transcription factors and 1,607 target genes. This represents 3,506 novel interactions spanning important cellular processes such as metabolism, various stress responses and virulence. This network includes interactions for all 64 transcription factors found in *S. pneumoniae*'s core genome, demonstrating its applicability to other strains. The average and median out-degree of all transcription factors in the network is 45 and 20, respectively, while every gene is regulated by an average of two transcription factors. The

network has a scale-free network topology and can be broken down into fifteen interconnected modules of five or more genes that are enriched in one or more cellular processes. Re-analyzing antibiotic treated RNAseq samples in a network perspective identifies transcription factors involved in specific stress responses, such as the importance that the Clp protease regulator, CtsR, has during kanamycin treatment. Identifying the effectors that regulate transcription factor activity would improve the network and allow for context-specific analysis of transcriptional regulation. A recent approach to achieve this is through time-matched RNAseq and metabolomics and can identify transcription factor effectors in high throughput¹⁷. Regarding *S. pneumoniae*'s stress response, it would be critical to identify the transcription factors affected by metal binding, oxidative stress and those that are regulated through proteolysis.

5.6 TRIP Screens Quantify Fitness Effects of TRN Perturbations

To determine the fitness effect of overexpressing each transcription factor in *S. pneumoniae* strain TIGR4, we completed several Transcriptional Regulator Induced Phenotype (TRIP) screens¹⁸. For these, all TFI strains are pooled and cultured under various conditions, with barcode sequencing used to track and quantify fitness of each individual strain. By overexpressing each transcription factor, we can synthetically create dozens of distinct TRN states and assay how the transcriptome influences fitness to a variety of conditions.

5.6.1 TRIP screens in Alternate Carbon Sources Connect TRN to Metabolic Network

Metabolic TRIP screens were performed in glucose, galactose, N-acetylglucosamine (GlcNAc) and mannose because each of these sugars are both available in the host and known to be catabolized by *S. pneumoniae*. TRIP screens in SDMM + glucose were able

to identify several TFI strains with statistically significant fitness changes, both positive and negative. Monoculture growth curve validation of these hits show little differences, demonstrating both the sensitivity of our method and how well-adapted these strains are to growth in rich media. Redundancy in the transcriptional and metabolic networks obscured TRIP screens in GlcNAc and galactose. No fitness advantages were identified in SDMM + GlcNAc because the primary enzymes required for its catabolism, NagA and NagB, are under carbon catabolite repression. Similarly, TRIP screens completed in galactose were obscured by the accidental induction of the Leloir pathway in all TFI strains by our inducer, IPTG. TRIP screens completed in mannose revealed three TFI strains with significantly positive fitness. These strains overexpress YesM, LsrR and NmlR, all of which upregulated expression of *manA*, the key catabolic enzyme of mannose, in TFI RNAseq. These positive fitness effects were likely due to preventing a diauxic shift by pre-induction of *manA*, as overexpression of YesM, LsrR and NmlR already did so.

5.6.2 Overexpression of CpsR, TreR and SusR Increase Resistance to CWSI but Increase Susceptibility to Kanamycin and Levofloxacin

All three TFI strains with the “Rapid Biphasic Lysis” growth curve profile in SDMM + glucose have significant resistance to the cell wall synthesis inhibitors Ceftriaxone and Vancomycin and significant susceptibility to Kanamycin and Levofloxacin. These strains overexpress the metabolic regulators CpsR, SusR and TreR, and TFI RNAseq revealed extensive overlap between the effects overexpressing each had on the transcriptome. Overexpression of both SusR and TreR resulted in repression of the capsule, while CpsR is named for its role in repressing capsule expression¹⁹, as well as differential expression of several cell wall synthesis related genes and activation of the oxidative stress response.

Gene enrichment analysis of genes differentially regulated by two or more of these three strains reveal significant enrichment in “Cell Wall Organization”, “Lipopolysaccharide Biosynthetic Process” and “Cell Redox Homeostasis”, among others. This suggests that the metabolic perturbations caused by overexpression of these three transcription factors leads to an increase in cellular acidity and either directly or indirectly represses cell wall synthesis, which explains the resistance to cell wall synthesis inhibitors, rapid induction of autolysis by LytA and susceptibility to the bactericidal antibiotics Kanamycin and Levofloxacin. Metabolomic analysis of these strains may reveal the exact mechanism(s) by which these transcription factors increase resistance to CWSIs while increasing susceptibility to Kanamycin and Levofloxacin, ultimately identifying candidates for antibiotic potentiation.

5.6.3 The Paramount Importance of Metal Homeostasis and Redox Regulators in the Stress Response

A recurring theme across TRIP screens is the overlapping importance that metal homeostasis and the redox regulation have on surviving stressful conditions. Transition metals, such as zinc, copper, manganese and iron, are essential cofactors for many cellular processes but can be toxic at high concentrations due to their reactivity and capacity for protein inhibition by mismetallation. Transition metals, the oxidative stress response and antibiotic treatment are all linked through the iron-dependent Fenton reaction that produces ROS during bactericidal antibiotic treatment²⁰. Our immune systems have evolved both metal intoxication and oxidative stressors to kill bacterial pathogens, indicating convergent defense mechanisms between our immune cells and antibiotic-producing microbes.

The importance of metal homeostasis and redox regulation is best exemplified by the orphan response regulator, RitR. RitR is a redox-sensitive²¹, global regulator that regulates expression of several metal transporters, oxidative stress genes and proteases. Overexpression led to significant increase in resistance to the protein synthesis inhibitor, Kanamycin. Other transcription factors with positive fitness in Kanamycin treatment include the regulator of copper export, CopY, an unnamed MarR family TF that regulates the antioxidant, thioredoxin, and the Clp protease regulator, CtsR. Several of these transcription factors, including RitR, CopY and possibly the unnamed MarR TF, directly respond to either metal binding or oxidation of cysteine residues^{21–23}.

To deconstruct the total stress response to specific stresses that occur during antibiotic treatment and from the immune system, TRIP screens were completed during zinc intoxication and oxidative stress. As expected, overexpression of the Clp protease regulator CtsR in elevated zinc concentrations revealed the importance of protease activity due to mismetallation. Overexpression of the phage infection protein regulator, PipR, with elevated zinc concentrations appears to have activated a bacteriophage in TIGR4. Interestingly, overexpression of PipR also had reduced fitness during Levofloxacin treatment, which aligns with previous studies that show fluoroquinolones are frequent prophage inducers²⁴, which may be related to metal dysregulation.

Zinc and manganese regulators had opposing fitness effects when overexpressed in the toxic redox cycling compound paraquat. Overexpression of the manganese regulator PsaR led to an increase in fitness, likely due to its regulation of the manganese importer *psaABC* and the antioxidant, *tpxD*. Alternatively, overexpression of the zinc efflux regulator, SzcA, and to a lesser extent the zinc influx regulator AdcR, had decreases in fitness.

Overexpression of both SzcA and AdcR lead to an increase in total cell-associated iron by ICP-MS and suggests dysregulated zinc homeostasis may result in increased oxidative stress via the iron-catalyzed Fenton reaction. Recently a zinc ionophore designed to impose zinc intoxication on *S. pneumoniae* was identified as a potent potentiator capable of breaking ampicillin resistance *in vivo*²⁵. We validated this finding as well as extended it by demonstrating zinc's potential as a potentiator of kanamycin.

5.6.4 TRIP screens Identify the Determinants of Antibiotic Tolerance

TRIP screens were adapted to time-kill assays to discover the transcriptional regulators involved in antibiotic tolerance. Tolerance TRIP screens completed with antibiotics with different targets produced highly correlated results, indicating tolerance mechanisms are agnostic to the initial stress. The only TFI strain with significantly increased tolerance to both Levofloxacin and Ceftriaxone is TFI_2195, which overexpresses the Clp protease repressor, CtsR. UMAP clustering shows Kanamycin TRIP results clustering with Tolerance TRIP results, further supporting the role of proteostasis in the induction of tolerance and persistence. Importantly, an Clp-activating drug, ADEP 4, has been shown to effectively kill persisters²⁶ and validates our TRIP screens as identifying useful drug targets.

Inversely, overexpression of the global nutritional regulator CodY led to a significant and validated decrease in antibiotic tolerance. Activated by either interaction with BCAA or oxidation of two conserved cysteines, CodY represses BCAA synthesis, alters glucose metabolism and activates expression of *tpxD*, an important antioxidant²⁷. Time-kill assays of TIGR4-LacI supplemented with 5mM of BCAs resulted in increased tolerance. We suspect this result can be attributed to a pre-induction of the oxidative stress response by

CodY through sensing BCAAs and H₂O₂ and may also ease resumption of growth after the antibiotic stress is gone by increasing the cellular pool of these amino acids. This is even more evidence for a translational role in antibiotic tolerance, and suggests that amino acid depletion, specifically leucine, isoleucine and valine, can potentiate antibiotic treatment. For pneumococcal infections, this could be performed through monoclonal antibodies targeting the BCAA importer, LivKHMGF (*SP_0749-SP_0753*).

Several other transcription factors that regulate the oxidative stress response demonstrated increased antibiotic tolerance. NmlR and ArgR are both connected with glutathione, an important antioxidant, were identified by TRIP screens and validated in monoculture time-kill assays as increasing tolerance. NmlR activates expression of an alcohol dehydrogenase, *adhC*, which detoxifies S-nitrosoglutathione to maintain glutathione pools. *AdhC* has previously been shown to be required for surviving oxidative stress and for virulence, but its conservation in humans may make it a difficult drug target^{29,30}. ArgR regulates arginine uptake, which may increase glutathione levels due to increased glutamate, a key precursor of glutathione. Further metabolomic studies may reveal a clearer mechanism. Similarly, PadR regulates genes responsible for clearance of oxidized membrane lipids and synthesis of replacements, which suggests lipid peroxidation is a significant threat to cell survival during antibiotic stress. Interestingly, lipid peroxidation and their toxic by-products have been proposed to be involved in the formation of therapy-induced senescence in cancer cells³¹. This suggests a broadly conserved role for membrane damage in cellular adaptation to stress.

5.7 Concluding Remarks

Here, we present the most comprehensive Transcriptional Regulatory Network of *S. pneumoniae* to date, comprising 3,946 transcription factor – target gene interactions involving 88 transcription factors and 1,607 target genes. Our TRIP screens quantified the fitness effects of the network under various antibiotic stresses, revealing zinc intoxication as an antibiotic potentiator, proteostasis as a main determinant of antibiotic tolerance and the pivotal role of metabolic regulation in the stress response. Encouragingly, many of our insights align with previous observations in other bacterial pathogens, highlighting conserved and targetable mechanisms that provide a promising, unified framework for advancing our fight against antibiotic resistance.

5.8 References

1. Sharma, A., Gupta, V. K. & Pathania, R. Efflux pump inhibitors for bacterial pathogens: From bench to bedside. *Indian J. Med. Res.* **149**, 129–145 (2019).
2. Chawla, M., Verma, J., Gupta, R. & Das, B. Antibiotic Potentiators Against Multidrug-Resistant Bacteria: Discovery, Development, and Clinical Relevance. *Front. Microbiol.* **13**, (2022).
3. Darby, E. M. *et al.* Molecular mechanisms of antibiotic resistance revisited. *Nat. Rev. Microbiol.* **21**, 280–295 (2023).
4. Vrancianu, C. O., Popa, L. I., Bleotu, C. & Chifiriuc, M. C. Targeting Plasmids to Limit Acquisition and Transmission of Antimicrobial Resistance. *Front. Microbiol.* **11**, (2020).
5. Guo, X. *et al.* Thwarting resistance: MgrA inhibition with methylophopogonanone unveils a new battlefield against *S. aureus*. *npj Biofilms Microbiomes* **10**, 15 (2024).
6. Koirala, B. & Tal-Gan, Y. Development of *Streptococcus pneumoniae* Pan-Group Quorum-Sensing Modulators. *ChemBioChem* **21**, 340–345 (2020).
7. Troeger, C. *et al.* Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory infections in 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect. Dis.* **18**, 1191–1210 (2018).
8. *Antibiotic resistance threats in the United States, 2019*. <https://stacks.cdc.gov/view/cdc/82532> (2019) doi:10.15620/cdc:82532.
9. Rosconi, F. *et al.* A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat. Microbiol.* **7**, 1580–1592 (2022).
10. Ng, W.-L., Tsui, H.-C. T. & Winkler, M. E. Regulation of the *sspA* Virulence Factor and Essential *pcnB* Murein Biosynthetic Genes by the Phosphorylated *VicR* (*YycF*) Response Regulator in *Streptococcus pneumoniae*. *J. Bacteriol.* **187**, 7444–7459 (2005).
11. Sorg, R. A., Gallay, C., van Maele, L., Sirard, J. C. & Veening, J. W. Synthetic gene-regulatory networks in the opportunistic human pathogen *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 27608–27619 (2020).
12. Marbach, D. *et al.* Wisdom of crowds for robust gene network inference. *Nat. Methods* **9**, 796–804 (2012).
13. Grant, C. E. & Bailey, T. L. XSTREME: Comprehensive motif analysis of biological sequence datasets. *bioRxiv* (2021) doi:10.1101/2021.09.02.458722.
14. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, (2009).
15. Madan Babu, M. & Teichmann, S. A. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. doi:10.1093/nar/gkg210.
16. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118**, 4947–4957 (2005).
17. Lempp, M. *et al.* Systematic identification of metabolites controlling gene expression in *E. coli*. doi:10.1038/s41467-019-12474-1.
18. Ma, S. *et al.* Transcriptional regulator-induced phenotype screen reveals drug potentiators in *Mycobacterium tuberculosis*. *Nat. Microbiol.* **6**, (2021).
19. Wu, K. *et al.* CpsR, a GntR family regulator, transcriptionally regulates capsular polysaccharide biosynthesis and governs bacterial virulence in *Streptococcus pneumoniae*. (2016)

doi:10.1038/srep29255.

20. Kohanski, M. A., Dwyer, D. J., Hayete, B., Lawrence, C. A. & Collins, J. J. A Common Mechanism of Cellular Death Induced by Bactericidal Antibiotics. *Cell* **130**, 797–810 (2007).
21. Glanville, D. G. *et al.* RitR is an archetype for a novel family of redox sensors in the streptococci that has evolved from two-component response regulators and is required for pneumococcal colonization. *PLOS Pathog.* **14**, e1007052 (2018).
22. Hao, Z. *et al.* The multiple antibiotic resistance regulator MarR is a copper sensor in Escherichia coli. *Nat. Chem. Biol.* **10**, 21–28 (2014).
23. O'Brien, H. *et al.* Rules of Expansion: an Updated Consensus Operator Site for the CopR-CopY Family of Bacterial Copper Exporter System Repressors. *mSphere* **5**, (2020).
24. Henrot, C. & Petit, M.-A. Signals triggering prophage induction in the gut microbiota. *Mol. Microbiol.* **118**, 494–502 (2022).
25. McDevitt, C. A. *et al.* A Molecular Mechanism for Bacterial Susceptibility to Zinc. *PLoS Pathog.* **7**, e1002357 (2011).
26. Conlon, B. P. *et al.* Activated ClpP kills persisters and eradicates a chronic biofilm infection. *Nature* **503**, 365–370 (2013).
27. Hajaj, B. *et al.* CodY Regulates Thiol Peroxidase Expression as Part of the Pneumococcal Defense Mechanism against H₂O₂ Stress. *Front. Cell. Infect. Microbiol.* **7**, (2017).
28. Macomber, L. & Imlay, J. A. The iron-sulfur clusters of dehydratases are primary intracellular targets of copper toxicity. *Proc. Natl. Acad. Sci.* **106**, 8344–8349 (2009).
29. Liu, L. *et al.* A metabolic enzyme for S-nitrosothiol conserved from bacteria to humans. *Nature* **410**, 490–494 (2001).
30. Stroher, U. H. *et al.* A Pneumococcal MerR-Like Regulator and S -nitrosoglutathione Reductase Are Required for Systemic Virulence. *J. Infect. Dis.* **196**, 1820–1826 (2007).
31. Flor, A. C. & Kron, S. J. Lipid-derived reactive aldehydes link oxidative stress to cell senescence. *Cell Death Dis.* **7**, e2366–e2366 (2016).

Supporting Data:

Supporting Data files can be found at:

https://drive.google.com/drive/folders/11VQKBb13b3V1Sgzqi28cfnqY8akLjvy2?usp=drive_link

S1_TIGR4_TRN.csv:

The complete Transcriptional Regulatory Network for TIGR4 includes Transcription factor, Target Gene, TFI RNAseq results, Network Inferences, Binding Motifs and Transcript ID information.

S2_TFI_RNAseq.csv:

All statistically significant TFI RNAseq data (Padjs). DEGs are all entries with Log2FC greater than |1|. Raw sequencing data can be found in BioProject PRJNA1152972

S3_TIGR4_GeneExpressionData.csv:

Gene Expression Data (in TPM) input into our ensemble of network inference methods. Referred to as “Dataset A” in text and outperformed data subsets.

S4_Eensemle_Inferences.csv:

Predicted regulatory interactions from our highest-performing ensemble of network inference methods (GENIE3-ARACNE-CLR-GRNBOOST2) combined by the Average Rank Method.

S5_All_Inferences.csv:

Predicted regulatory interactions from all network inference methods.

S6_TIGR4_TF_BindingMotifs.csv:

Predicted Binding sites for all Transcription factors across all promoters in the TIGR4 genome.

S7_TIGR4_Transcript_IDs.csv:

Transcript IDs (TIDs) for all genes in TIGR4.

S8_TRIP_FitnessValues.csv:

TRIP Fitness values for all completed TRIP screens.

S9_TRIP_BCcounts.csv:

Raw barcode counts for all TFI strains across all TRIP screens.

S10_Primers.csv:

List of primers used throughout this work, including cloning primers and TRIP screen primers.