Deep Learning for Rapid, Single-Cell Antimicrobial Susceptibility Testing

Aleksander Zagajewski
Keble College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy

Hilary 2023

Acknowledgements

When I started this journey in October of 2018, I had anticipated that a research doctorate would be one of the most challenging yet rewarding undertakings I could sign up for. In part, this is why I did it. The following four and a half years certainly did not disappoint on either measure.

First, I would like to thank my supervisors and collaborators, without whom none of this research would have been possible. Thank you Achilles for not only being an academic supervisor, but also a mentor, leader and friend. Thank you Hafez for patiently teaching a physicist how to deal with biology, and Sonia for always teaching me to inject more of my physics into said biology.

I also want to thank the other members of the Oxford Martin School Consortium on Rapid Antimicrobial Susceptibility Testing: Nicole Stoesser, Christoffer Nellåker, Conor Feehily, Piers Turner, Alison Farrar, Stelios Chatzimichail and others - it was a pleasure to work and learn from you. In similar vein, thank you to other current members of the lab: Christof Hepp, Oliver Pambos, Abhishek Mazumder, Heesoo Uhm, Jagadish Prasad Hazra, Andrew McMahon, Nicolas Shiaelis, Anna Wang, Rasched Haidari, Mirjam Kümmerlin, Qing Zhao; and past members Jun Fan, Rebecca Andrews, Jingwen Xiao and Maabur Sow.

My time in Oxford was made complete and brighter by the range of incredible people I met here and was lucky to call a friend. This list could never be complete, but special thanks to Casey Adam, Marco Grossi, Clara Pavillet, Nikos Athanasiou, Dessislava Veltcheva, Kevin Wang, Nik Gourianov, Micheal Negus, Violeta Stojalnikova, Chris Higgins, Shannon Brunette and James Paterson.

Na końcu tej odysei chciałbym podziękować moim rodzicom, Annie i Przemysławowi. Bez ich poświęceń nic by nie wyszło ze mnie i z mojej edukacji.

Abstract

The rise of antimicrobial resistance (AMR) is one of the most pressing global healthcare challenges, already causing an estimated 1.2 million preventable deaths annually and rising. Crucial to the management of AMR is rapid and specific diagnosis, allowing early and optimized intervention. Unfortunately, current gold-standard antimicrobial susceptibility tests are low-throughput and can take up to 48 hours to produce clinically relevant insights. In this thesis, we propose and evaluate a novel AST approach, based on the deep-learning of single-cell phenotypes directly associated with antimicrobial susceptibility. The phenotypes are revealed by widefield fluorescence microscopy and evaluated automatically by a deep-learning pipeline built on convolutional neural networks (CNNs). We demonstrate our Deep Antimicrobial Susceptibility Phenotyping (DASP) can robustly recognise susceptibility phenotypes associated with 4 representative antibiotics of major antibiotic families, in Escherichia coli, with over 80% single cell accuracy. We then deploy our models trained on susceptible lab strains, to clinical isolates of Escherichia coli treated with one of the antibiotics. Here, we demonstrate the distribution of single-cell phenotypic classification decisions is a reliable indicator of isolate susceptibility around a fixed treatment point, revealing statistically significant ($p<0.001$) differences between untreated and treated cell populations in susceptible isolates, and no difference in resistant isolates. Further, we evaluate the limit of detection, and show this population-level output is indeed sensitive to the resistance status of single cells. Lastly, we investigate the relationship between treatment concentration, the minimum inhibitory concentration (MIC) of the isolate, and the DASP output, and compare this against the gold-standard growth assay. Here, we show that DASP has potential to produce equivalent information to the current gold-standard, but an order on magnitude faster. We conclude this thesis with an outlook on the developmental and mechanistic principles of the phenotypes by studying their time evolution.

Statement of Authorship

Science is always a collaborative endeavour, which the author has chosen to honour by conscious use of the plural, first-person pronoun ‘we’ when describing the contributions of this thesis.

Nevertheless, all work presented herein is the author’s own, except where indicated below, and/or in the text as appropriate:

- Figure 1.3.1 was directly adapted from a presentation by Dr. Conor Feehily.
- MICs of clinical isolates underpinning Figure 3.2.5 B were measured experimentally by Dr. Conor Feehily as part of the wider project family.
- Colony counts supporting Figure 3.2.7 were measured experimentally by Dr. Conor Feehily.
- Growth curve raw data supporting Figure 3.2.11 were measured experimentally by Dr. Conor Feehily.
- The HU-GFP strain supporting results in Chapter 4 was provided by Dr. Hafez el Sayyed and Dr. Olivier Espéli.

Data availability statement

The data and model weights that support the findings of this study, are available from the Oxford University Research Archive: https://ora.ox.ac.uk/objects/uuid:12153432-e8b3-4398-a395-abfb980bd84e. The repository contains all data and models underpinning results from Chapters 2 and 3. Data underpinning results from Chapter 4 are available from the author upon reasonable request.

Code availability statement

All code used to generate the results is available publicly on the Kapanidis Laboratory GitHub account, accessible at: https://github.com/KapanidisLab/Deep-Learning-and-Single-Cell-Phenotyping-for-Rapid-Antimicrobial-Susceptibility-Testing. Detailed guidance on using the code is available upon request from the author.

Contents

List of Figures **viii**

List of Abbreviations **xi**

1 Introduction **1**

1.1 Bacterial Infections and Modern Antibiotics 1

1.1.1 Host-Pathogen Interactions in Bacterial Disease 1

1.1.2 History and Modes of Action of Modern Antibiotics 5

1.2 The Rise of Antimicrobial Resistance 11

1.2.1 Biochemical Mechanisms of Resistance 11

1.2.2 Current Timeline 15

1.2.3 Drivers of Resistance 17

1.3 Strategies to tackle Antimicrobial Resistance 17

1.3.1 Novel Antibiotics 17

1.3.2 Public Health Initiatives 18

1.3.3 Antimicrobial Susceptibility Testing - Current Approaches . 19

1.3.4 Antimicrobial Susceptibility Testing - Novel Developments . 22

1.4 Introduction to Deep Learning 26

1.4.1 A Brief Overview of Machine Learning 26

1.4.2 Artificial Neural Networks for Supervised Learning 29

1.4.3 Neural Networks in Computer Vision 35

2 Deep Antimicrobial Susceptibility Phenotyping (DASP) assay design, development and training. **40**

2.1 DASP Concept 41

2.2 DASP Design and Training 43

2.2.1 Phenotypes 43

2.2.2 Training and Computational Experiments 48

2.3 DASP Validation Results 53

2.3.1 Segmentation 53

2.3.2 Phenotype Classification 56

2.4 Classifier Explainability 59

2.5 Experimental Methods 63
2.6 Discussion 70
2.6.1 Segmentation and classification of phenotypes 70
2.6.2 Experimental design 72
2.6.3 Future directions in optimising phenotypes 74
2.6.4 Generalisability of findings 75
2.7 Conclusions 76

3 DASP Deployment on Clinical Isolates 77
3.1 Phenotype Heterogeneity in Clinical Isolates 77
3.1.1 Ciprofloxacin-susceptible phenotypes 79
3.1.2 Ciprofloxacin-resistant phenotypes 81
3.2 Segmentation and Classification of Clinical Isolates 82
3.2.1 Clinical isolates at a fixed treatment point 83
3.2.2 Mixed samples and the Limit of Detection 92
3.2.3 Antibiotic titration allows MIC quantification 97
3.3 Experimental Methods 101
3.4 Discussion 105
3.4.1 Deployment of models on clinical isolates 105
3.4.2 Sources of errors in the classifier 107
3.4.3 Generalisability of findings 109
3.4.4 Future directions 110
3.5 Conclusion 112

4 DASP Time Evolution 113
4.1 Phenotypes as a function of time 113
4.2 DASP as a function of time 123
4.2.1 DASP as a function of single-cell phenotype development . . 123
4.2.2 Aggregate DASP response as a function of time 129
4.3 Experimental Methods 130
4.4 Discussion 133
4.4.1 Deployment of models on timelapses 133
4.4.2 Future directions 134
4.5 Conclusions 136

5 Conclusions 137
5.1 The state of the field of Antimicrobial Susceptibility Testing 137
5.2 The literature gap - diagnostic utility of CNN-derived features and subcellular phenotypes. 139
5.3 Addressing the gap - scientific contributions of this thesis 140
5.4 Extensions and Future work 144
5.5 Closing remarks 150

References **151**

List of Figures

1.1.1 Antibiotic targets in a model bacterial cell 11
1.2.1 Timeline of resistance development and AMR response severity, including selected key events. 16
1.3.1 An representative gold-standard phenotypic AST pipeline 22
1.3.2 Emerging ASTs by mode and reporting outcome. 23
1.4.1 Functional representation of an artificial neuron 32
1.4.2 CNN architectures visualised 37

2.1.1 DASP: an AST assay based on single-cell phenotyping and deep learning 42
2.1.2 DASP experimental wetlab process 43
2.2.1 Phenotypes, and analysis pipeline 44
2.2.2 Representative FoVs of untreated, ciprofloxacin-treated and gentamicin-treated cells. 46
2.2.3 Representative FoVs of rifampicin-treated and co-amoxiclav treated cells. 47
2.2.4 Setup of computational experiments 49
2.2.5 Example training logs of holdout classifiers 52
2.2.6 Example weight histograms of holdout classifiers 54
2.3.1 Segmentation in dense and sparse FoVs 55
2.3.2 Precision-Recall and Confidence-Recall curves of the segmenter, evaluated on the holdout test experiment. 56
2.3.3 Binary classification of resistant and susceptible phenotypes in the holdout experiment. 58
2.3.4 Full binary classification results of resistant and susceptible phenotypes across all computational experiments. 60
2.4.1 Saliency mapping of randomly selected, correctly classified susceptible cells 62
2.4.2 Saliency mapping of randomly selected, misclassified susceptible cells 64
2.5.1 Treatment conditions and abberviations used to generate susceptible phenotypes. 65
2.5.2 Microscopy of samples on agarose pads 66

3.1.1 Untreated and susceptible phenotypes in clinical isolates 80
3.1.2 Untreated and resistant phenotypes in clinical isolates 82
3.2.1 Representative FoVs and phenotype detections of ciprofloxacin-susceptible clinical isolates EC1 and EC2, in both untreated and ciprofloxacin treated conditions. 84
3.2.2 Representative FoVs and phenotype detections of ciprofloxacin-susceptible clinical isolates EC3 and EC4, in both untreated and ciprofloxacin treated conditions. 85
3.2.3 Representative FoVs and phenotype detections of ciprofloxacin-resistant clinical isolates EC5 and EC6, in both untreated and ciprofloxacin treated conditions. 86
3.2.4 Distribution of single-cell classifications is sensitive to the resistance of clinical isolates. 88
3.2.5 Changes in detection distribution upon treatment correlate with the degree of resistance in clinical isolates. 90
3.2.6 Total numbers of cells detected in each of the biological replicate experiments of clinical isolates. 91
3.2.7 Proportion of resistant cells is a linear function of dilution of resistant culture in susceptible cells. 94
3.2.8 Total numbers of cells detected in each of the biological replicate experiments of the mixed samples, in each dilution step. 95
3.2.9 Residual analysis and limit of detection of resistant cells 96
3.2.10 Normal probability plot of the linear fit residuals. 97
3.2.11 DASP provides equivalent information to a 24-hour growth assay, but within 30 minutes. 99
3.2.12 Total numbers of cells detected in each of the biological replicate experiments of ciprofloxacin titrations, in every isolate and ciprofloxacin concentration 100
3.3.1 Genotypes and MICs of clinical isolates 102

4.1.1 Comparison of staining methodology for live and fixed cell imaging 115
4.1.2 Timelapse of untreated HU-GFP reveals expected cell division and nucleoid replication. 117
4.1.3 Kymograph of a single, representative HU-GFP *E.coli* cell with no antibiotic treatment, undergoing cell division. 117
4.1.4 Timelapse of ciprofloxacin-treated HU-GFP reveals reveals a halt of cell division and condensation of both nucleoids towards cell centre. 118
4.1.5 Kymograph of a single, representative HU-GFP *E.coli* cell under ciprofloxacin treatment. 118

4.1.6 Timelapse of gentamicin-treated HU-GFP reveals reveals a halt of cell division and condensation of both nucleoids, but without merging. 119
4.1.7 Kymograph of a single, representative HU-GFP *E.coli* cell under gentamicin treatment. 119
4.1.8 Timelapse of rifampicin-treated HU-GFP reveals a halt of cell division and decompaction of both nucleoids. 120
4.1.9 Kymograph of a single, representative HU-GFP *E.coli* cell under rifampicin treatment. 120
4.1.10 Timelapse of co-amoxiclav treated HU-GFP reveals reveals loss of membrane integrity. 121
4.1.11 Co-amoxiclav treatment leads to a loss of membrane fluorescence intensity over time until catastrophic cell lysis occurs. 122
4.2.1 Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, ciprofloxacin-susceptible binary classifier. 124
4.2.2 Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, gentamicin-susceptible binary classifier. 126
4.2.3 Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, rifampicin-susceptible binary classifier. 127
4.2.4 Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, co-amoxiclav susceptible binary classifier. 128
4.2.5 Population response to ciprofloxacin treatment in live HU-GFP *E.coli*. 130
4.2.6 Population response to gentamicin treatment in live HU-GFP *E.coli*. 131
4.2.7 Population response to rifampicin treatment in live HU-GFP *E.coli*. 131
4.2.8 Population response to co-amoxiclav treatment in live HU-GFP *E.coli*. 132

List of Abbreviations

AMR	Antimicrobial resistance
PBP	Penicillin binding protein
PABA	Para-amino-benzoic acid
AME	Aminoglycoside modifying enzyme
RPP	Ribosome protection protein
AMS	Antimicrobial stewardship
AST	Antimicrobial Susceptibility Testing
CFU	Colony Forming Unit
OD, $\mathbf{OD_{600}}$	Optical Density at 600nm
EUCAST	European Committee on Antimicrobial Susceptibility Testing
MALDI-TOF	Matrix-assisted laser desorption ionization-time of flight mass spectrometry
ML	Machine learning
PCA	Principal Component Analysis
kNN	k-Nearest-Neighbours
SVM	Support Vector Machine
ANN	Artificial Neural Network
SGD	Stochastic Gradient Descent
SGD+N	Stochastic Gradient Descent with Nestrov Momentum
CNN	Convolutional Neural Network
RCNN	Region-based Convolutional Neural Network
FPN	Feature Pyramid Network
FoV	Field(s) of View
DAPI	4,6-diamidino-2-phenylindole
IoU	Intersection over Union
FISH	Fluorescence In Situ Hybridization
LoD	Limit of Detection

Omnium rerum principia parva sunt.

The beginnings of all things are small.

— Marcus Tullius Cicero, *de Finibus Bonorum et Malorum*

1

Introduction

Contents

1.1 Bacterial Infections and Modern Antibiotics 1
  1.1.1 Host-Pathogen Interactions in Bacterial Disease 1
  1.1.2 History and Modes of Action of Modern Antibiotics . . 5
1.2 The Rise of Antimicrobial Resistance 11
  1.2.1 Biochemical Mechanisms of Resistance 11
  1.2.2 Current Timeline 15
  1.2.3 Drivers of Resistance 17
1.3 Strategies to tackle Antimicrobial Resistance 17
  1.3.1 Novel Antibiotics 17
  1.3.2 Public Health Initiatives 18
  1.3.3 Antimicrobial Susceptibility Testing - Current Approaches 19
  1.3.4 Antimicrobial Susceptibility Testing - Novel Developments 22
1.4 Introduction to Deep Learning 26
  1.4.1 A Brief Overview of Machine Learning 26
  1.4.2 Artificial Neural Networks for Supervised Learning . . . 29
  1.4.3 Neural Networks in Computer Vision 35

1.1 Bacterial Infections and Modern Antibiotics

1.1.1 Host-Pathogen Interactions in Bacterial Disease

One of the greatest drivers in the epic odyssey of human history is the evolutionary journey we are locked in with the untold billions of species of microscopic organisms with whom we share the Pale Blue Dot - intertwining cycles of competition and co-dependence that have shaped our biology, society and technology to define human existence across ages.

The scale of diversity of this microscopic world is vast, with considerable differences in the level of organisation of these single-cellular organisms. On the smallest scale, viruses are little more than protein-encapsulated nucleic acids able only to divide inside another host organism - prompting basic questions whether they can be considered to be alive. On the largest and most complex scale we would find species of fungi - cellular organisms with multiple layers of sub-cellular organisation, often capable of switching between unicellular and multi-cellular lifestyles. Between these two extremes, we would find bacteria - unicellular organisms with a defined cell membrane and cytoskeleton, and with a degree of sub-cellular structure, but nevertheless lacking the diversity of organisation found in eukaryotes.

Much can be written about the beneficial co-dependence we enjoy with bacteria that enable our survival [1, 2]; in this thesis we concern ourselves instead with the cycle of competition with pathogenic bacteria. Only a small fraction of bacteria can be considered pathogenically infectious; of this small percentage, a small percentage of infections will progress to clinically apparent disease [3]. The infection is typically acquired from interaction with a bacterial community in a reservoir - the modes of transmission include physical contact, airborne and droplet spread, animal bites, or ingestion of contaminated food or water.

Once transmission has occurred, the first objective is the colonisation of the host surface - typically covered with mucus [4, 5] containing antimicrobial peptides, antibodies and commensal bacterial species seeking to exclude enteric pathogens by competition. Adaptations to cross the mucus are varied [6], and include metabolic adaptations to out-compete commensal species, secretion of proteases to degrade the mucus, or the expression of flagella to aid motility. Following the mucus, the next challenge is adhesion to host cells [7] - typically achieved by specific ligand-receptor interactions with protein or glycoconjugate targets on the host surface and the expression of multifunctional adhesins called pilli. Beyond facilitating attachment, these adaptations also allow the formation of a biofilm [8] - a polymer-matrix-enclosed bacterial community that facilitates advantageous conditions for growth, and resists mechanical, chemical and immunological clearance by the host.

Once colonisation is achieved, the second objective is typically the infiltration of the host cells themselves - this grants further protection and enables access to the nutrient-rich cellular microenvironment. This usually proceeds by the manipulation of the host actin cytoskeleton reorganisation mechanisms [9] - membrane reorganisation has the intended side-effect of engulfing the small bacteria into the cytoplasm. This can occur either passively, when ligand-receptor interactions between bacterial and host membrane components induce host membrane reorganisation (the so called 'zipper' mechanism), or actively, where the bacterium injects into the host (eg. via the aforementioned pilli) effector proteins that induce membrane reorganisation (the 'trigger' mechanism).

The third objective of a now intracellular bacterium is naturally to replicate; replication occurs either freely in the cytoplasm, or more commonly in distinct vacuoles [10]. The pathogen will typically remodel the vacuole by altering their proteomic or lipidomic contents to better suit their metabolism, traffic-in essential nutrients or evade degradation.

The final frontier is naturally cell-to-cell spread of the infection, facilitating systemic infection and further increasing the pathogens's growth opportunities. Here, the possibilities are even more varied than before. Pathogens that replicate in the cytoplasm often hijack the host actin machinery to assemble actin 'tails'; the force of assembly propelling them around the cytoplasm and eventually towards other cells [11]. Other exit strategies [12] include induced host cell lysis, extrusion of pathogen-containing vacuoles, of the formation of protrusions to form bridges towards uninfected cells - all mechanisms usually involving actin remodelling through hijacking of host pathways.

So far we have briefly revisited the topic of infection from the pathogen's point of view - the journey is naturally complicated at every step by the host passive and active immune system. As mentioned, mucosal and interstital layers contain passive defences, and are also patrolled actively by macrophages and sentinel cells, high-turnover of epithelial cells encourages mechanical clearance, and bacterial vacuoles and cytosolic bacteria are targets of lysosomal degradation. When these defences are insufficient and the infection proceeds to clinically apparent disease, the common first reaction is inflammation of affected tissues - a non-specific but complex response [13] that involves increased permeability of blood vessels leading to influx of plasma. This contains antimicrobial compounds such as lysozyme, antibodies, and clotting factors, and facilitates the recruitment of phagosomic leukocytes to the site. Non-specific inflammation is followed by specific [14] adaptive response, involving lymphocytes producing targeted antibodies towards the infection source (B cells), or directly inducing apoptosis of diseased host cells (T cells).

Of the many possible symptoms experienced by an infected individual, some can be directly attributed to the (beneficial) immune response - such as the redness and swelling caused by inflammation; or fever, also linked to inflammation. Other symptoms can be directly attributed to the infection - sometimes simple growth is enough to disrupt normal tissue function but commonly the pathogens produce toxins as a by-product of growth that have systemic effects - destroying metabolic machinery, killing cells outright, or overstimulating the immune response to the point of toxicity (as is thought to be the case, for example, with the recent COVID-19 pandemic, still ongoing at the time of writing of this thesis). While most bacterial infections are naturally cleared by the immune system given time, sometimes medical intervention is required to alleviate symptoms and reduce the risk of long-term complications and improve patient outcomes.

The premier weapon of choice in the clinic is the arsenal of antibiotics - natural and synthetic chemicals that exhibit specific antimicrobial activity that directly kill pathogenic bacteria or at least inhibit their growth by targeting specific bacterial pathways. Antibiotics have been used since ancient times, and modern formulations have revolutionised 20th century medicine. Easy access to antibiotics has eroded their effectiveness however, by applying a strong selection pressure on pathogens to evolve resistance mechanisms towards them [15]. Antimicrobial Resistance

(AMR) is now a global health emergency, on course to undo a century of progress across both the developed and developing world. To address this crisis, composite, global solutions are needed [16], involving changes to medical policy and practice, further pharmaceutical innovation, as well as technological developments in areas such as diagnostics. In this thesis we concern ourselves with the last point - the development of novel and rapid diagnostic technology that may help conserve our existing antibiotic arsenal.

This brief review naturally cannot hope to do justice to the fields of cell biology of infection and immunology. However, it is the author's hope that it provides a useful, high-level primer to the true subject of the thesis, which is the field of antibiotics, antibiotic-induced cellular damage resulting in distinct phenotypes of antibiotic action, and finally, usage of such phenotypes in a diagnostic capacity as a proxy for antibiotic susceptibility on a single cell level. We begin our journey with a survey of the antibiotic landscape.

1.1.2 History and Modes of Action of Modern Antibiotics

While the advent of modern antibiotics is universally credited to Alexander Fleming's chance discovery of penicillin whilst experimenting with cultures of *Staphylococcus aureus* in 1927 [17], the first modern antibiotic was actually arsphenamine [18] - another chance find by Paul Ehrling following exhaustive systematic screening of compounds active against *Treponema pallidum* - the cause of syphilis.

Penicillin G is the world's first β-lactam antibiotic; its bactericidal effects rely on disrupting cell wall formation [19]. Specifically - the bacterial cell wall is composed of polysaccharide chains crosslinked by peptide chains (forming the so-called peptidoglycan), where the cross-linking is catalyzed by essential penicillin-binding proteins (PBPs, also known as DD-transpeptidases). β-lactams are structural analogues of the D-Ala-D-Ala, the terminal residues of the nascent peptide cross-link, and irreversibly acylate the serine residue in PBP active site. The PBPs have no eukaryotic analogous, which contributes the the relative non-toxicity of the treatment. Beyond inhibiting further growth and division, a slow-down of the peptidoglycan cross-linking creates a build-up of precursor substrates, which is linked with the activation of autolytic hydrolases that actively digest existing peptidoglycan, eventually leading to cell lysis.

From this basis, further generations of penicillins were developed by modifications to the side-chain [18]. First-generation penicillins (1950s) featured simple carboxylic acids in the side chain and were most effective against Gram-positive cocci, while the second-generation compounds such as ampicillin and amoxicillin (aminopenicillins, 1960s) introduced an amino group - which increased the resistance to the acidic environment of the stomach (and thus bio-availability of the medication) and broadened the activity spectrum to some Gram-negative species. Third-generation compounds like carbenicillin introduced instead acidic residues to further broaden the spectrum, while fourth-generation compounds (1970s) instead added bulky sidechains to both widen the spectrum and increase steric resistance against inactivation by β-lactamases - the major driver or resistance to β-lactams which we will encounter later in more detail. Since the 1970s, further improvement has come from combining penicillins with β-lactamase inhibitors - a prominent example being the combination of amoxicillin (2nd generation penicillin) and clavulanic acid (β-lactamase inhibitor), marketed as co-amoxiclav or augmentin.

Penicillins are not the only β-lactams; cephalosporin was isolated from a fungal culture in the 1940s, and shows similar activity and mode of action to the penicillins due to shared stereo-chemistry. The important functional difference is the replacement of the penicillin penam backbone with a cephem backbone, with the option for two variable side-chains instead of one [20] - giving more options for tuning the activity spectrum and bio-availability. Clinically, cephalosporins show higher resistance to β-lactamases than penicillins and result in better clinical outcomes in selected cases [21]. Like the penicillins, their historical development has occurred in distinct generations, with subsequent generations widening the spectrum of activity towards Gram-negative pathogens. Further variations on the original penam backbone have yielded carbapenems (1970s), which feature the broadest spectrum and highest resistance to β-lactamases [22]. Monobactams (1980s) are unique amongst β-lactams as the β-lactam moiety is not fused to another ring, and these compounds are uniquely specific against Gram-negative species.

Parallel with the development of β-lactams, Paul Ehrling's success with arsphenomine inspired further systematic screening that lead to the discovery of sulphonamides - the first synthetic broad-spectrum antibiotics (1940s). Sulphonamides target folic acid production [23], which is important in nucleic acid synthesis and thus essential to growth. Folic acid is synthesised in bacteria from para-amino-benzoic acid (PABA) by dihydropteroate synthetase - sulphonamides are structural analogous of PABA and thus disrupt folic acid synthesis, leading to bacteriostatic effects. In contrast, while human cells require folic acid for analogous reasons, they do not synthesise it themselves, rather relying on absorption from external sources. Further development of sulphonamides has continued in the decades that followed and several formulations are still in regular use today in specialist applications - however they have 'lost' the race to β-lactams, which showed fewer side effects.

The next major family of antibiotics to enter use were the aminoglycosides [24] (1940s onwards) - a diverse family of hydrophilic amino sugars with particular affinity towards nucleic acids. Examples of aminoglycosides include gentamicin, and neomycin. 4 different families of aminoglycosides exist, depending on their structure - but all target the A-site of the 16S rRNA of the 30S ribosomal subunit. Aminoglycosides vary in their affinity for various parts of the A-site, and their exact mechanisms, but all induce conformational changes of the 30S subunit, thus promoting tRNA-codon mismatch and thus protein mistranslation. Some aminoglycosides can further target translation by inhibiting translation initiation or elongation. Although aminoglycosides are highly specific to prokaryotic rRNA, rRNA is broadly conserved between prokaryotes and eukaryotes - it is thought this similarity may account for some of the side effects observed clinically [25]. Aminoglycosides have broad-spectrum activity against aerobic Gram-negative and Gram-positive species, and are currently experiencing renewed scientific interest due to rising resistance against other broad-spectrum antibiotics that had replaced them in clinical practice due to lower toxicity.

Simultaneous with the development of aminoglycosides, another major family targeting the 30S ribosomal subunit was discovered - the tetracyclines. Unlike aminoglycosides, tetracyclines consist of 4 fused hydrocarbon chains, with a variety of side-chains [26]. Like aminoglycosides, tetracyclines are active against both Gram-positive and Gram-negative species and target the A-site of the 16S rRNA; unlike aminoglycosides, they also exhibit activity towards eukaryotic ribosomes, however tetracyclines do not access eukaryotic cells readily, reducing toxicity. The rise of AMR has diminished the clinical effeciveness of tetracyclines, although they remain in use.

Another broad-spectrum translation inhibitor of the same vintage is chloramphenicol - this time reversibly binding to the 50S subunit A-site [27], blocking the action of peptidyl transferase and thus preventing the condensation reaction that forms the polypeptide chain. Eukaryotic ribosomes are not a target, but chloramphenicol passes biological membranes readily and causes mitochondrial stress in human cells [28] (mitochondrial ribosomes resemble bacterial 70S ribosomes). Chloramphenicol is not longer in regular use in the developed world over toxicity concerns, but remains in use in the developing world.

Following these early successes, the next major family introduced were the macrolides (starting from erithromycin, 1950s), named as such due to the macrocyclic lactone ring that they all share. They are primarily highly active against Gram-positive aerobic bacteria but suffer from poor bioavilability and acid stability - improved with later generations such as azithromycin. Macrolides are also translation inhibitors and also bind to the 50S subunit, obstructing the exit tunnel and interrupting translation elongation, although recent research in uncovering more nuance in this mechanism [29].

The 1950s also brought another family of cell-wall-targeting antibiotics - the glycopeptides; glycosylated polycyclic peptides which, similarly to β-lactams, disrupt peptidoglycan formation by targeting the D-Ala-D-Ala terminus. [30] Unlike β-lactams, glycopeptides directly bind to D-Ala-D-Ala and thus inhibit crosslinking, and are most active against Gram-positive infections.

The last major innovation of the 1950s were the ansamycins - a broad family featuring an aromatic moiety bridged by an aliphatic chain. An important category of ansamycins are rifamycins - premier treatment for mycobacterial infections, including tuberculosis and leprosy [31]. Uniquely amongst the families we have discussed so far, rifamycins block transcription, by binding to the β-subunit of bacterial RNA polymerase (RNAP) and sterically blocking the elongation of RNA. [32] In contrast, eukaryotic polymerases are not targeted. Uncommonly amongst antibiotics, rifamycins show strong drug-drug interactions with other common medications, complicating multi-drug therapies.

The streptogramins (1960s) are unique, in that a treatment course always contains two distinct, unrelated compounds that act synergistically and show cooperative binding. Type-A streptogramins are polyunsaturated cyclic macrolactones that bind both the A and P sites of the 50S ribosome subunit and thus disrupt the polypeptide condensation reaction; Type B streptogramins are cyclic hepta-or hexadepsipeptides and bind the exit tunnel of the 50S subunit (similar to the macrolides), blocking translation elongation. [33]. The binding of type-A streptagramins induces conformational changes in the ribosome that further expose the type B binding site. Individually, each compound has bacteriostatic effects; together, bacteriocidal. Streptogramins are used for the treatment of aerobic and anaerobic Gram-positive organisms.

The 1960s have also brought another major family of antibiotics that has since risen to prominence as one of the most commonly used classes - the quinolones. Quinolone antibiotics target bacterial type-II topoisomerases: gyrase and topoisomerase-IV. Gyrase and topoisomerase-IV control DNA torsion and degree of supercoiling, especially in support of DNA replication and RNA transcription complexes. They achieve this by introducing double-stranded breaks in the DNA, cleaving and ligating staggered breaks 4 nucleotides apart in opposite strands - creating so-called 'cleavage complexes'. [34] Quinolones bind at the active site of the enzymes, sterically inhibit the ligation step and stabilise the enzyme-DNA complex at the break site. Over time, this causes an accumulation of cleavage complexes, which become permanent double-stranded breaks. These trigger DNA repair pathways and the SOS response - cell death occurs when these processes are overwhelmed by the DNA damage. Subtle differences between human and bacterial type-II topoisomerase binding sites spares human cells from damage. Initially used for the treatment of Gram-negative organisms and showing poor bioavailability, later generations of quinolones (including the very popular ciprofloxacin) enhanced the spectrum to some Gram-positive species and improved the distribution within human tissues.

With the end of the 1960s, we are reaching the end of the golden age of antibiotic discovery, and have only two more major families left to cover. The 1980s brought us the oxazolidinones - a fully synthetic class of broad spectrum translation inhibitors used against a variety of Gram-positive infections, and named after the oxazolidone ring in the structure. [35]. Oxazolidones bind at the A-site of the bacterial 50S ribosomal subunit, which we have discussed before, although unlike chloramphenicol, oxazolidones inhibit translation initiation, not elongation by peptidyl transferase. The effect is bacteriostatic in nature. The pharmacokinetic properties of oxazolidinones and their bioavilability make them suitable to prolonged administration time, making them promising candidates for conditions such as tuberculosis or osteomyelitis.

The last family we will discuss are the lipopeptides - a broad and novel family of natural product derivatives active against Gram-positive species, and common drugs of last resort for severe infections resistant to other classes of antibiotics. Lipopeptides actively disrupt multiple cell membrane functions, leading to rapid cell death. Daptomycin (2000s), the class progenitor, actively inserts itself into the cell membrane and leads to rapid depolarisation [36, 37] - the details of this process are an active area of research. Unfortunately they show poor oral bioavilaibility and need to be administered parenterally. The lipopeptides bear resemblance to the antimicrobial peptides secreted by the immune system, that we briefly discussed before.

Comparing the major families, we can see emerging trends. Antibiotics target highly conserved processes that are crucial to the bacterial lifecycle - peptidoglycan formation, translation, transcription and DNA replication. The modes of action are

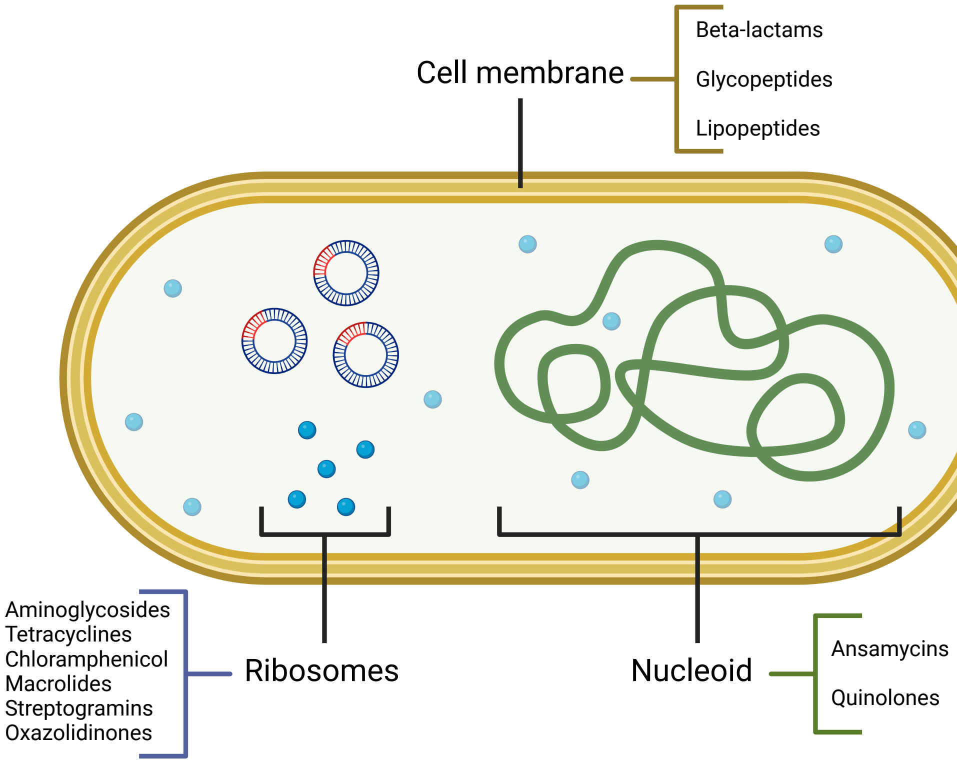


Figure 1.1.1: Antibiotic targets in a model bacterial cell Antibiotics target key lifecycle processes such as transcription, translation or cell wall formation, which all have defined sites within the model bacterial cell. These sites include the nucleoid, ribosomes and cell membrane. Figure created with Biorender.

similarly conserved and highly specific; antibiotics are monomodal. These modes of actions have specific sites, such as the nucleoid, ribosomes and cell membranes (Fig 1.1.1) Antibiotic therapy is also often a monotherapy - the effectiveness hinges on the inhibition of one specific biochemical step of one specific mechanism. This creates evolutionary pressure on the pathogen that is both strong, and highly specific - creating the perfect conditions for the development of resistance.

1.2 The Rise of Antimicrobial Resistance

1.2.1 Biochemical Mechanisms of Resistance

The modes of resistance to antibiotics are highly specific to the antibiotic, but all modes share one of 3 core mechanisms: mutation of the antibiotic binding site to reduce antibiotic specificity and cytotoxicity, active inactivation of the antibiotic, or reduction in the intracellular concentration of the antibiotic - either by preventing entry, or active transport outside the cell.

The predominant mode of resistance to β-lactams is the production of inactivating β-lactamases, particularly in Gram-negative bacteria [38]. These can be broadly divided into serine-based β-lactamases and zinc metallo-β-lactamases. All serine-based β-lactamases employ an acetylation mechanism to perform a nucleophilic attack on the amide carbonyl carbon of the β-lactam moiety, hydrolysing it; mechanisms of the metallo-enzymes are more varied and still objects of active study. β-lactamases are often carried on plasmids and mobile genetic elements and trafficked between species, furthering AMR spread. By contrast, in Gram-positive species, the main mode of resistance is modification to the PBP active site. [39]. Of particular clinical concern is the spread of Enhanced-Spectrum Beta Lactamases (ESBL) that inactivate many sub-classes of β-lactams [40] as well as carbapenems, given these drugs status as premier β-lactams in current use, and their intrinsic resistance to hydrolysis [41].

Resistance to sulphonamides is primarily driven either by mutations of dihydropteroate synthase, or horizontal acquisition of alternative systems [42].

In contrast, resistance to aminoglycosides in more multifaceted. [24, 25, 43]. Target site mutation is not seen often - likely because most bacteria encode multiple copies of rRNA, and because of the intricacy and importance of the translation machinery - however has been observed in the case of *Mycobacterium tuberculosis*, the cause of tuberculosis. Enzymatic target site modification has been seen however, consisting of 16S rRNA methylation by rRNA methyltransferases, in a way such that translation activity is not inhibited, but aminoglycoside binding is. Aminoglycoside modifying enzymes (AMEs) are commonly carried on plasmids (sometimes together with β-lactamases hinting at a shared origin), and can be divided into 3 categories based on their interaction with amino or hydroxyl groups of the aminoglycoside: aminoglycoside N-acetyltransferases acetylate, phosphotransferases phosphorylatre and nucleotidyltransferases adenylate. Low-level resistance to aminoglycosides has been observed due to efflux mechanisms actively pumping out the chemical.

Resistance to tetracyclines follows the trends of aminoglycosides, but in addition features the presence of ribosome protection proteins (RPP) - in this specific case, the Tet family of proteins catalyse the unbinding of tetracyclines from rRNA, allowing translation to continue in their presence. [44].

Macrolide resistance is driven heavily by efflux pumps of the Mef family and RPPs of the Msr family, which operate in a similar manner to the Tet proteins of tetracycline [29]. rRNA methyltransferases (including the Cfr, Erm and Ksg families) have also been observed, as well as point mutations in both rRNA and ribosomal proteins. Macrolide inactivation is carried out by macrolide esterases of the Ere family which hydrolyse the lactone ring, or by phosphotransferases of the Mph family; both of which are found commonly on mobile genetic elements.

In contrast with the β-lactams that also target peptidoglycan synthesis, the major mode of resistance to glycopeptides is not enzymatic inactivation of the antibiotic, but alternative peptidoglycan synthesis - replacing the D-Ala-D-Ala pathway with D-Ala-D-Lac or D-Ala-D-Ser [30].

Resistance to rifamycins is mainly driven by point mutations in the *rpoB* gene that encodes the β-subunit of RNAP [45]. RNAP is highly conserved; consequently the spectrum of known mutations is narrow and conserved between different species. Isolated instances of efflux and enzymatic inactivation have been observed, but there is no consensus if these are clinically relevant.

When considering the streptogramins we have to consider both types separately. [33] For type-A, enzymatic modification of either the drug or target are most common; RNA methylation is carried out by the Cfr protein family, Vat proteins acetylate the streptogramins themselves, and Lsa and Vga RPPs can displace bound antibiotics. In type-B, rRNA methylation by is carried out by the Erm, Cfr and Ksg proteins. There is considerable cross-resistance between type-B streptogramins and macrolides, owning to the shared mode of action. Although resistance to streptogramins has been observed clinically, incidence remains low relative to other antibiotic classes.

Unfortunately, quinolone resistance is becoming a widespread clinical issue. Quinolones themselves are partially to blame - the induction of the SOS response by

DNA damage increases the bacterial cell's mutation rate, increasing the likelihood of successful adaptation [46]. Consequently, mutations of the gyrase and topoisomerase-IV are common; and almost universally occur at one of the two residues that have been found to coordinate quinolone binding to the enzyme. Plasmid-mediated forms of resistance are also common - the *Qnr* family of proteins decrease the binding affinity of gyrase and topoisomerase-IV for DNA and thus reduce the amount of cleavage complexes in the steady state, and also inhibit the binding of quinolones to the complexes. Other plasmid-borne proteins include acetyltransferases and efflux pumps, that work analogously to the cases we've discussed already.

Incidences of resistance to oxazolidinones remain low, largely due to their relative novelty, and a fully synthetic nature with no biological analogues. However, 3 different mechanisms have been observed so far [35] - a point mutation close to the binding site in 23S rRNA, 2 different mutations in 50S ribosomal proteins, and the expression of the *Cfr* family rRNA-methyltransferases, that we've already seen inhibiting 50S binding antibiotics such as macrolides and streptogrammins.

Clinical resistance to daptomycin and other lipopeptides has a low incidence; nevertheless, there are several mechanisms worth discussing. [36] There are occurrences of hydrolytic enzymes targeting daptomycin in soil bacteria and non-pathogenic strains - although these have not been observed clinically, there are concerns over horizontal acquisition of such systems by pathogenic strains. Phenotypically, resistance is associated a thickened cell envelope and an increase in surface charge, leading to decreased affinity of daptomycin for the cell wall. Multiple mechanisms have been implicated or suggested to mediate this; these remain an area of active research.

Once again, we are in a good position to reflect, and compare and contrast the mechanisms of resistance. Many antibiotics share targets, which leads to the occurrence of simultaneous cross-resistance to multiple antibiotic families - particularly in the cases of target site modification or mutation. Inactivating enzymes are often carried on mobile genetic elements, and trafficked between different species. Resistance is naturally more common in older antibiotics - however even for novel antibiotics that have not seen significant amounts of resistance in the clinic, there exist known mechanisms of resistance. The development of resistance to these novel ‘antibiotics of last resort’ is but a matter of time - and history teaches us that the timescales are frighteningly short. Equipped with the knowledge of biochemical resistance mechanisms, we can now turn our attention to consider the timeline of resistance development.

1.2.2 Current Timeline

Regardless of resistance mechanism or antibiotic origin, the average indicative timegap between the clinical introduction of an antibiotic and the first observed cases of resistance varies between 2-15 years - although gaps in accurate monitoring make precise measurements of this impossible. Of particular concern is the rise of Multi Drug Resistant organisms that carry resistance to several classes of antibiotics used to treat them.

As AMR has been developing, so has the global perception and response to the crisis. [47] The early days of the 1940s, 1950s and 1960s were marked by optimism, but also occasional concern about reports of Penicillin and Methicillin resistance in *Staphylococcus*, which began to waver following the discovery of horizontal transmission of AMR genes in the mid 1960s. Unfortunately, collective and coordinated action did not manifest until the 1980s - including WHO working groups and non-profit organisations such as the Alliance for the Prudent Use of Antibiotics. Lamentably, significant, multinational attention and funding did not manifest until the 2010s - only in 2013 did Dame Sally Davies, the Chief Medical Officer for England, declare AMR ‘as important as climate change’. At the time of writing, up to 1.2 million deaths worldwide are directly attributed to AMR [48], with this number predicted to rise to 10 million by 2050, if left unchecked.

Figure 1.2.1 illustrates selected key events [49] in the development of antimicrobial resistance.

We are now in good position to consider some of the drivers of AMR.

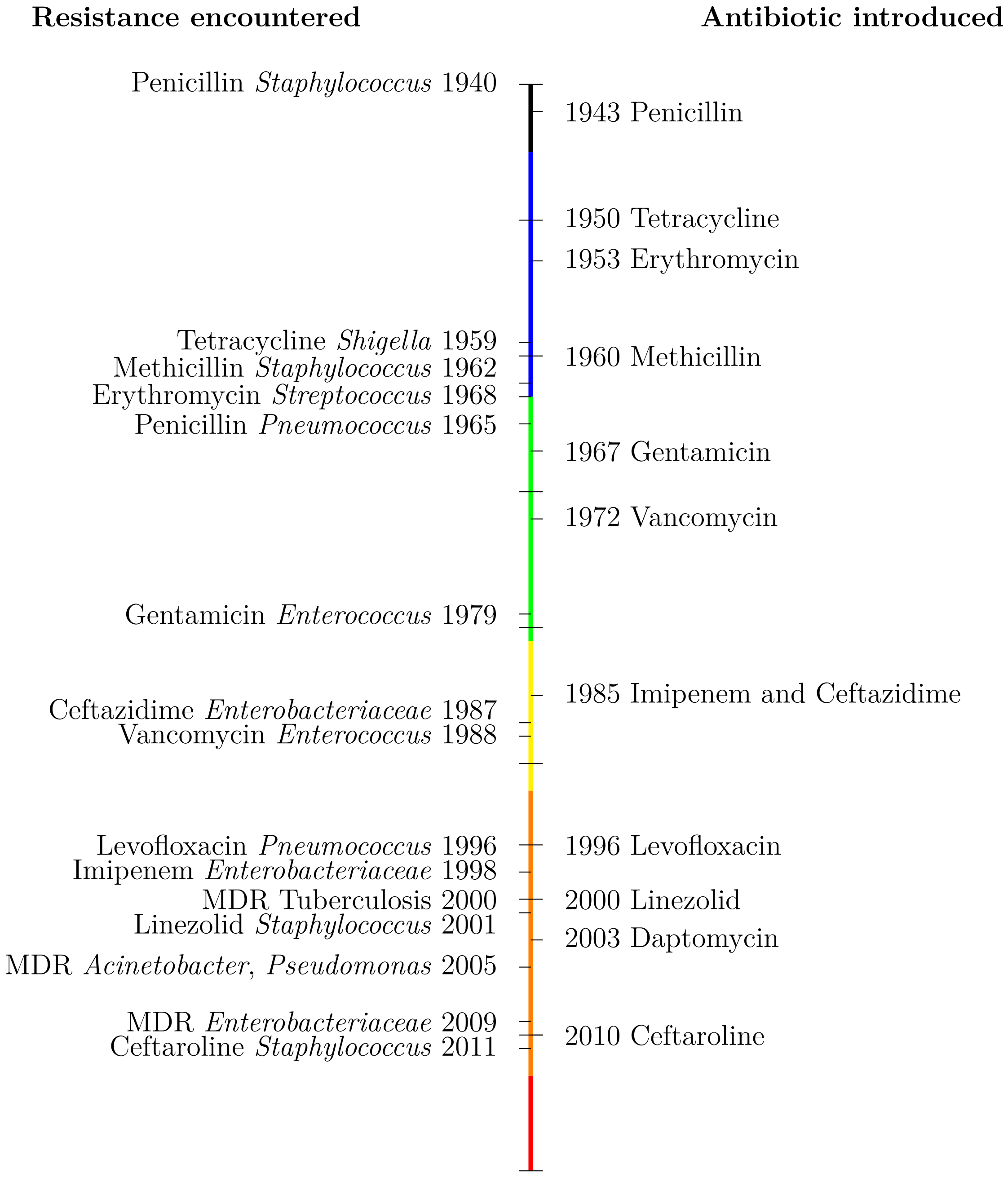


Figure 1.2.1: Timeline of resistance development and AMR response severity, including selected key events. On the left-hand side: antibiotic to which resistance was encountered, genus in which resistance was observed, and date. MDR - Multi Drug Resistant. On the right-hand side: date of introduction of various representative antibiotics. Events referenced from Ventola [49]. Colour codes correspond to increasing awareness of AMR.

1.2.3 Drivers of Resistance

By far the major driver of the rise of AMR is unnecessary over-prescription [50] of antibiotics, especially in ambulatory settings [51]. Beyond overuse in clinical settings, overuse in agriculture and food production (as growth enhancers, or pre-emptive prophylaxis) are also significant drivers [52]. Other important contributions come from changing societal patterns, especially increases in human migration, tourism and urbanisation [53]. Antibiotics accumulate in the environment (including wastewater [54] and soil [55]), which facilitates spread of resistance between different microbial species by horizontal gene transfer. Similar effects are achieved more directly by insufficient sanitation (leading to, for example, faecal–oral transmission) and disinfection (for example, contamination of health-workers' gloves). A full socio-economic analysis of the complex factors that drive AMR spread is beyond our scope - we can nevertheless proceed to the strategies used to manage it, bringing us closer to the topic of this thesis.

1.3 Strategies to tackle Antimicrobial Resistance

1.3.1 Novel Antibiotics

The rate of antibiotic discovery is slowing down, with few truly novel antibiotics brought to market in the last 20 years. Part of the reason is the exhaustion of the discovery platforms that have fuelled much of the discovery pipeline so far [56, 57]. Much of the early Golden Age was powered by the Waksman Platform - systematic, experimental screening of soil bacteria (specifically actinomycetes) for lead compounds active against common pathogens - later improved and optimised by semi-synthetic chemistry to improve efficacy and pharmacokinetics. However, the rate of discovery of novel lead compounds has decreased exponentially, leading to the conclusion that actinomycetes have been overmined. The discovery of fully-synthetic antibiotic families, such as the quinolones has fuelled hopes for a modern target-centric platform - utilising proteomics and genomics to identify viable targets and targeting them via high-throughput synthetic chemistry. Although this approach has proven successful in many areas, it failed for antimicrobials [58] - largely due to the impermeable nature of Gram-negative cell envelopes. There are concentrated efforts to revive the target-centric platform by focusing on cell envelope permeation rules [59], and investigate new sources of natural produces [60], but the impact of those measures remains to be seen. Nevertheless, the discovery of novel compounds only postpones the problem of AMR rather than providing a more composite solution - resistance will inevitably occur.

The interpid reader will correctly deduce that the problem could be addressed better with sufficient financial incentives - unfortunately, novel antibiotics suffer from a unique profitability challenge [61]. Unlike chronic diseases, antibiotic treatments are short, and thus bring limited profits. The problem of AMR means any novel antibiotic will be prescribed sparingly; and due to the rise of multi-drug resistance and few unique antibiotic targets, the novel antibiotic will invariably be rendered less-effective with time, limiting profits. Instead, the main focus of alleviating the AMR crisis has been antimicrobial stewardship and public health initiatives - which we will discuss next.

1.3.2 Public Health Initiatives

The main public health initiative to fight the rise of AMR has been conservation of the existing antibiotic arsenal - often called ‘Antimicrobial Stewardship’ (AMS), which focuses on the reduction of inappropriate prescription. Evidence [62] suggests these measures are effective, and are reducing hospital length of stay, healthcare costs, rates of resistance, and hospital-acquired infections. No universally agreed definition of AMS exists, but certain elements are common [63]. These include multi-level governance, education of physicians, involvement of nursing and pharmacy staff, and monitoring of antibiotic prescription. Most of the antibiotic prescribing happens in ambulatory care; evidence suggests a major failing of AMS is the inappropriate prescription of antibiotics for self-limiting conditions [51]. Effective implementation of AMS is challenging - professional boundaries and hierarchies between healthcare staff have been identified as one of the most common barriers across all settings [64], although low and middle-income countries additionally struggle with the cost of specialist antibiotics, and appropriate physician education [65]. Although generally accepted as effective, the exact impact of AMS is hard to measure, and the quality of studies in the area remains poor [66]. Regardless of local guidance, the cornerstone of successful AMS policy is ensuring the correct antibiotic is administered, at the correct time, in the correct dose - and then correctly de-escalated. Crucial to this strategy is rapid and specific diagnosis. Antimicrobial Susceptibility Testing (AST) is an umbrella term that covers a range of diagnostic techniques and technologies that aim to identify the pathogen and infer or quantify its degree of resistance - enabling correct antibiotic treatment. Let us review current AST approaches now.

1.3.3 Antimicrobial Susceptibility Testing - Current Approaches

The first step towards a clinical AST is culture, isolation and identification. A patient sample is acquired (consisting of blood, urine, or other fluid) and cultured to increase the Colony Forming Unit (CFU) counts for subsequent analysis. This is followed by isolation - usually patient presentation combined with the operator's experience will narrow the search to the tractable volume. Selective culture discriminating between the organism's Gram status, growth conditions (aerobic vs. anaerobic), growth rate, or specific biochemistry (salt tolerance, carbon sources etc) achieves isolation and identification of the pathogen. Additional information may be gained from selective staining and microscopy. The isolation step is labour-and-time-consuming - it requires a skilled microbiologist and on average hours to days to achieve.

Current ASTs [67–72] can be broadly divided into two categories based on the type of information they provide.

Phenotypic ASTs provide categorical or quantitative output on the amount of antibiotic needed to inhibit bacterial growth (the resistance phenotype) - which is quantified by the Minimum Inhibitory Concentration (MIC). Invariably, this is achieved by measuring the growth of the isolate in the presence of an antibiotic, thus requiring another laborious and time-consuming culture step. Classic methods for this include semi-quantitative agar plate disk or gradient diffusion, where the size of the inhibition zone is standardised against the MIC. The reference methods of choice are broth or agar microdilution, which measure growth as a function of antibiotic concentration (and potentially time), and allow accurate quantification of the MIC. The readout is either growth on an agar plate, or an increase in Optical Density (OD, OD_{600}) of a broth culture. Current efforts to increase the throughput and speed of ASTs focus on automating broth microdilution (BD Phoenix® by BD, Vitek® by Biomerieux, Microscan® by Beckman Coulter, Sensititre® by Thermo Fisher Scientific), or couple automated microscopy to detect agar plate colonies earlier (Growth Direct System® by Rapid Micro Biosystems). These can alleviate manpower issues, but fundamentally do not remove the time taken for the re-culture step.

The relationship between the MIC and the pathogen's classification as susceptible or resistant, and the ongoing course of action, is governed by guidelines from expert panels, such as the European Committee on Antimicrobial Susceptibility Testing (EUCAST) [73].

The other class of ASTs are marker-based ASTs - instead of measuring the resistance phenotype directly, they measure a biomarker associated with resistance. Common here are genetic methods that screen for known resistance genes, involving either qPCR (amplification) or microarrays (hybridisation). Unlike phenotypic methods, their specificity means many of them can operate without pathogen isolation requirement. These methods are faster, remove the AST culture step, and highly specific, and also are well suited to rapid pathogen identification. However, they suffer notable disadvantages - the presence of a genetic marker does not generally correlate with the presence of resistance, and screening is limited to a small list of *a priori* defined targets. The former issue may be addressed by using reverse-transcriptase qPCR to look at the transcriptome instead, while the latter may be addressed by whole genome sequencing - however, translation to clinical practice has been slow. Further, genetic methods cannot provide an MIC, or a straightforward mapping to one.

Apart from genetic methods, the other major marker-based AST is matrix-assisted laser desorption ionization-time of flight mass spectrometry (MALDI-TOF MS) - which searches for signatures of proteins from a bacterial culture lysate with rapid turn-around time. Although generally used for bacterial identification from proteomics spectra, MALDI-TOF MS can also search for proteins that mediate resistance (such as inactivating enzymes) or antibiotics that have been hydrolysed or enzymatically modified - measuring the resistance phenotype directly. However, this still requires extensive optimisation of target fragments for each antibiotic-species pair.

The modern clinical laboratory will focus on the phenotypic assays as the primary procedure, and complement it with marker methods, depending on local guidelines and resources. Although usually separate, some recent advancements have combined phenotypic and marker assays and have recently entered the market - a good example is the Accelerate Pheno ®system by Accelerate Diagnostics [74], which combines growth monitoring of individual cells under an automated microscope and nucleic acid probe hybridisation.

We can summarise the current AST landscape. The primary indicator of choice is the MIC of an isolate. This can be provided by phenotypic growth assays, which are robust, can be automated, and require little *a priori* knowledge of likely targets - however, they are fundamentally limited by the time required for 2 successive culture steps, which could stretch into days. Secondary indicators can be obtained much faster by looking for specific biochemical markers associated with resistance - but they require calibration, screen pre-defined targets, and not always correlate with the primary indicator.

There is considerable demand for methods that either accelerate the phenotypic screen, or find universal biomarkers that need less *a priori* calibration. There are also secondary pressures to increase the cell-specificity of ASTs (most methods can only operate on entire bacterial cultures), and bring them closer to the Point of Care (most methods require expensive equipment, operators and laboratory space).

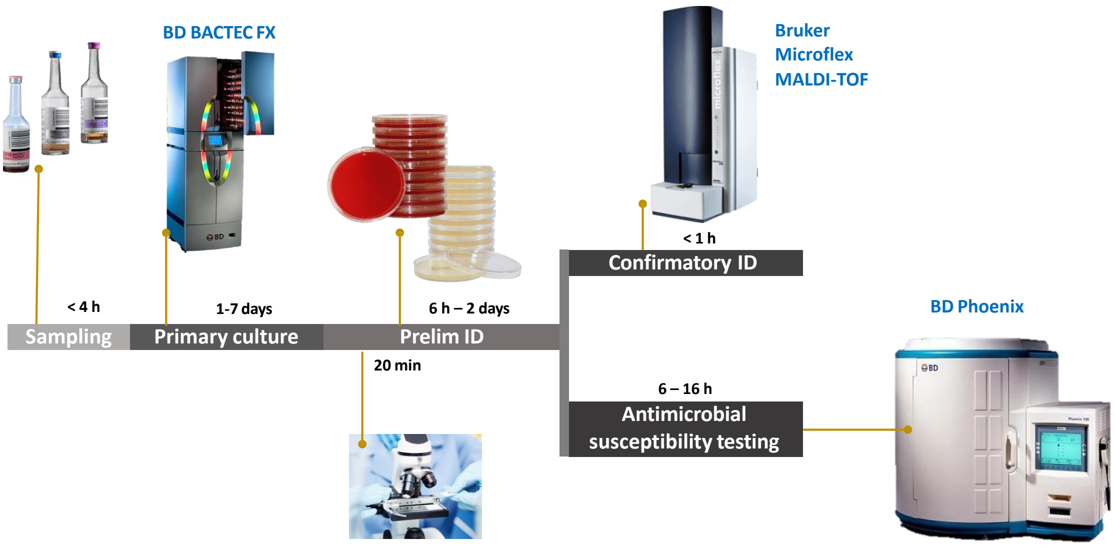


Figure 1.3.1: An representative gold-standard phenotypic AST pipeline A patient sample undergoes primary culture to increase CFU density for further processing. This is followed by isolation and preliminary identification through selective culture, potentially including microscopy. The pathogen identity may be verified through MALDI-TOF. Simultenously, a phenotypic growth-based AST can be used to quantify the MIC of the isolate to a panel of antibiotics by directly measuring the threshold concentration needed to inhibit culture growth. Figure adapted from a presentation by Dr. Conor Feehily.

We can now proceed to consider novel ASTs that aim to mitigate these shortcomings.

1.3.4 Antimicrobial Susceptibility Testing - Novel Developments

One of the main current directions in AST development is the development of novel platforms which may house new assays. One of such platforms is microfluidics. Such devices offer a controlled micro-environment in a portable package with multiplexing capabilities and minimal sample volume, and are well suited to bringing complex biochemical assays closer to the Point of Care. Another platform is flow cytometry - a mature technology well suited for single cell analysis. Regardless of platform of choice, emerging ASTs may again be categorised into either phenotypic or biomarker assays (Figure 1.3.2, with similar advantages and disadvantages as outlined above.

Currently the platform of choice for metabolic markers is MALDI-TOF MS which

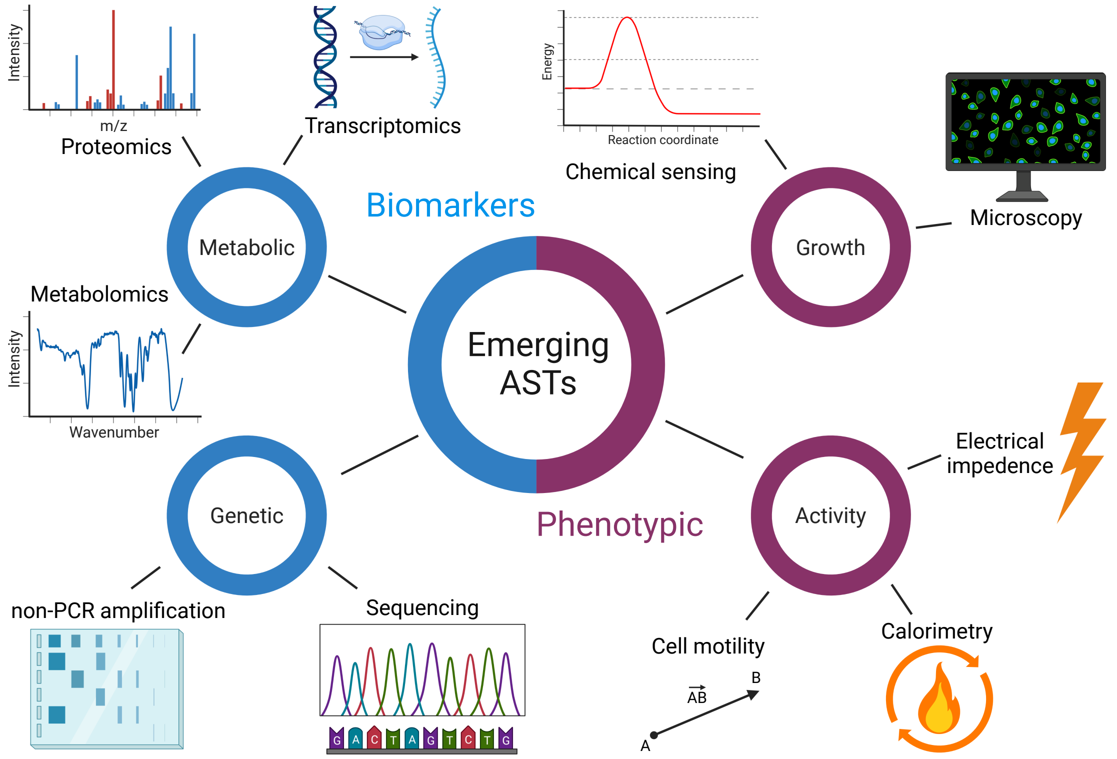


Figure 1.3.2: Emerging ASTs by mode and reporting outcome. Emerging ASTs can be divided into phenotypic assays that measure changes directly caused by antimicrobial action; and biomarker assays, that measure secondary markers associated with resistance. Figure created with Biorender.

offers proteomic output, but other platforms are in development. One approach is to shift the detection earlier in the metabolic cycle - looking at the transcriptome instead of the proteome. The transcriptome is sensitive to antimicrobial action [75], and has already been used to create candidate assays [76, 77]. Several recent examples use Raman spectroscopy with single cell resolution [78, 79] or whole culture lysate [80] to distinguish treated and untreated susceptible and resistant cells (available as Seraspec® by Seraph Biosciences). Other label-free system measures the total metabolic output by calorimetry (calScreener® by Symcel). Promising, but not yet commercialised markers include glucose metabolism [81]; recently, tandem liquid chromatography and mass spectrometry were used to identify a library of metabolites which were predictive of the MIC [82]. These techniques are promising, but the high equipment cost and operating requirements make the barrier-to-entry high - standard immunohistochemistry approaches can be used to track specific proteomic targets of high concern using lateral flow tests [83, 84] Chen *et al* provides an excellent, technical and recent overview of novel metabolic markers [85].

One trend in genetic marker ASTs is to switch away from targeted assays towards whole genome sequencing using Next Generation Sequencing - taking the genetic to the genomic level [86–89], however, at present, suffers issues such as limited genotypic-phenotypic concordance and the lack of database and methods coverage and standardisation; the current evidence for using genomic analysis in clinical decision making is insufficient [90]. Another trend is to improve cost and accessibility by relaxing the requirement for thermal cycling in conventional PCR genetic testing by using alternative amplification technology such as loop mediated isothermal amplification [91, 92], and to improve the performance of hybridisation tests by using non-natural nucleic acid probes [93].

There exist concentrated efforts to develop novel phenotypic indicators of susceptibility. These include electrical impedance impedence [94] and fluorescent reporters [94, 95] (FASTinov® by FASTinov), all delivered via flow cytometry. Other indicators include bacterial vibrations measured on a resonant crystal [96], or by Atomic Force Microscopy [97], although these have not been commercialised.

Cellular growth remains a common phenotypic indicator, although the ways to report it are evolving, with improving sensitivity, specificity and time-to-result. Recent examples of new ways of measuring growth include light scatter [98], field-effect enzymatic detection [99], fluorogenic redox indicators [100], culture oxygen changes [101], 16S rRNA quantification by Nanopore sequencing [102], ATP bioluminescence [103], or measuring volatile compounds emitted during growth (Reveal®, Specific Diagnostics). A promising, microfluidics-compatible platform for new growth assays is wide-field microscopy [104–108] - which offers the simultaneous interrogation and tracking of many cells on a single-cell level. A close relative of cellular growth are changes to cellular morphology [109], and motion patterns [110, 111]. These approaches have already resulted in systems in various stages of commercialisation (dRAST® by QuantaMatrix, ASTar® by Q-linea, QuickMIC® by Gradientech, oCelloScope® by BioSenseSolutions). Further information can be gained nucleic acid hybridisation techniques to provide both phenotypic and genomic marker output [112] (Accelerate Pheno® by Accelerate Diagnostics as already discussed). Van Belkum provides a further technical analysis [113].

The approach closest to our work is to evaluate the effect of antibiotics on sub-cellular structures under an automated microscope, especially those directly attacked. These structures include the cell membrane, nucleoid, or ribosomes (antibiotic targets of section 1.2.1). Antibiotics induce characteristic, phenotypic changes to these structures in susceptible cells, creating phenotypes of antimicrobial susceptibility. The readout is phenotypic, since it quantifies directly the antimicrobial effect. The difference is the timescale - the phenotypic change is visible on the timescale of a single bacterial lifecycle, rather than several (prerequisite for growth). A range of such phenotypes have been characterised experimentally [114–118] for a range of antibiotics and target organisms. Explicit, hand-engineered features extracted from such phenotypes have been used to distinguish between methicillin-susceptible and resistant *Staphylococcus aureus* clinical isolates [119], and a deep-learning classifier has been used to distinguish phenotypes without manual feature engineering in synthetic fields of view [120] - but without demonstrating that this analysis is sensitive to the actual susceptibility of clinical isolates.

Regardless of underlying technology, novel rapid AST systems have to overcome several barriers [121] before deployment, including regulatory, cost-effectiveness, demonstrated clinical utility, and ongoing quality control.

In this work, we present advances towards adapting sub-cellular phenotypes of the nucleoid and cell membrane for a rapid AST, using cultured clinical isolates of *Escherichia coli*. Depending on the antibiotic of choice, our assay can deliver results in 30 minutes, and can provide direct mapping to the MIC of the isolate. To provide rapid and robust analysis of the phenotypes, we use deep learning and computer vision techniques, which we will review now, prior to proceeding to the scientific contributions of this thesis.

1.4 Introduction to Deep Learning

1.4.1 A Brief Overview of Machine Learning

If machine learning (ML) as a field ever had a founding moment, it was the Dartmouth Summer Research Project on Artificial Intelligence [122] - a brainstorming conference held in 1956 at Dartmouth College in Hanover, New Hampshire, organised by John McCarthy and attended by other early luminaries such as Claude Shannon and Marvin Minsky. McCarthy's stated objective was:

> *...to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves...*
>
> — John McCarthy, Marvin Minsky, Nathaniel Rochester & Claude Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955

True to this form, the distinguishing difference between the intrinsically linked fields of applied statistics and ML is the focus - statistics focuses on the properties of relationships within data; while ML focuses on learning and improving abstract but generalisable patterns that enable predictions on unseen data.

To enable such learning from data, a mathematically-encoded feedback mechanism is required - taking the form of an objective function that needs to be minimised or maximised (and as such, linking ML to the field of mathematical optimisation). Unlike in statistics, the available data are divided into training and testing sets - the training set is used maximally and iteratively to train an algorithm, while the test set is used as an unbiased estimator. The different approaches used to implement such feedback can be used to divide ML approaches into broad, separate paradigms:

1. **Supervised learning**: the algorithm is provided raw data and human-curated desired outputs. The job of the algorithm is to match raw data to the correct outputs. In this case, the objective function is usually an accuracy metric

that quantifies how well the raw data are matched by the algorithm. This approach is commonly used for tasks such as classification or regression.

2. **Unsupervised learning**: the algorithm is provided raw data only, with no desired outputs. The objective function instead is a metric that quantifies agreement with a certain distribution - either empirically discovered by the algorithm or imposed *a priori*. Such algorithms are used to discover or verify relationships within data; the classic example is cluster analysis.
3. **Reinforcement learning**: instead of feeding data to an algorithm, the algorithm (now called an 'agent') is provided with a set of actions with which it can change the state of its environment. The objective function (now called 'reward') is a function of the environment. This approach easily supports a dynamic objective function without assuming or requiring an exact mathematical model (which may be unfeasible), and is commonly used for to model complex systems such as protein folding, games and autonomous vehicles.

Regardless of paradigm, a certain sequence of steps are common in the ML pipeline. Real-world data are usually not 'clean' - before further work, the ML practitioner needs to consider outliers, errors, and data formatting issues. The data needs to be split into training and testing sets - and care needs to be taken to ensure both are similarly distributed, such that the training set provides learning applicable to the real-world scenario; and the test set, an unbiased estimator. The training data are often further subdivided into training and validation sets - the validation data are further withheld and used to monitor for overfitting. The topic of ML bias is a major issue in every area where ML is deployed and worthy of discussion [123–125] far beyond the scope of this thesis. For supervised techniques, the training set needs human annotation - usually a laborious process.

The next step is pre-processing. This involves feature selection, or engineering - selecting or constructing from the raw data individual, measurable, informative, independent and discriminating properties that the algorithm will actually use.

In a way, this process constructs intermediate representations of the raw data that are more compact and informative. This is usually the hardest and most domain-specific part of the pipeline - we will review the techniques relevant to the subject of this thesis in the next section. This process may also involve normalisation or standardisation; putting features on the same scale is a hard or soft requirement for the correct performance of many algorithms. Feature selection also commonly involves dimensionality reduction - this avoids the ‘curse of dimensionality’ (well discussed by Bellman[126]), which in the context of ML increases the volume of data needed exponentially with increasing dimensionality. Dimensionality reduction techniques may involve linear transformations such as Principal Component Analysis (PCA) (decomposition onto variance-maximising basis set by factorisation of the covariance matrix), Linear Discriminant Analysis (linear mapping to a lower dimensional space that best separates labelled classes), Singular Value Decomposition (decomposition by factorisation of the data itself). It may also involve non-linear transformation, such as kernel-PCA (projection to higher dimensional space where data becomes separable) or isometric mapping (distance-preserving projection).

Following the pre-processing, the practitioner will select and train an algorithm - the choice will depend on the degree of structure in the data, its quantity and complexity[127]. Heterogeneity of chosen features is important - certain algorithms require features to be scaled to similar ranges (using normalisation, like we just discussed) and too much redundancy (correlated features) can lead to numerical instabilities. It is important to match the representational capacity of the algorithm and the complexity of the problem. A model with not enough capacity will not learn the patterns in the data and perform poorly, while a model with too much capacity will simply memorise all the training example, and generalise poorly to unseen date (overfitting).

Beyond optimising performance, there are also other secondary considerations. One is model explainability. Explainability is a broad term covering our ability to understand the model’s decision making. This includes the ability to quantify the influence individual features have on the final result, understanding the effect of outliers and edge cases, and quantification of the confidence we can place in the results. This is a crucial quality control step in the pipeline, allowing the practioner to assess the model's performance prior to deployment on unseen data. Explainability generally decreases with increasing model complexity.

Once a model is trained, it is evaluated using the chosen performance metrics on the test set. Most algorithms feature hyperparameters - parameters that control the learning process that are experimentally optimised. Most algorithms benefit from hyperparameter optimisation, although to different extents[128]. The techniques for picking the optimal hyperparameters range from a fixed grid search to genetic algorithms and Bayesian approaches. In modern practice, much of algorithm selection and optimisation can be taken care of by automatic ML platforms [129, 130].

With a trained model, the practitioner's job is not yet done. The principal assumption of applied ML is that the training and testing data are unbiased and accurate representation of the data the model will be deployed on - that is, they are similarly distributed. This assumption almost always fails in practice - over time, data experiences 'drift', where its distribution changes over time. A deployed model needs to be continuously monitored and improved as necessary to maintain performance.

In this thesis we present work that relies heavily on supervised learning in the context of computer vision - applying ML to digital images and videos. We have just discussed the generic ML pipeline common to every project, including this one, and mentioned the major challenges and considerations. We can now launch ourselves into a discussion of supervised learning techniques of most relevance to this thesis.

1.4.2 Artificial Neural Networks for Supervised Learning

The two classic tasks of supervised learning are classification and regression, and many ML problems can be reduced to one of those tasks. The main difference between the tasks is the output target variable - continuous (regression) or categorical (classification). As such, most algorithms can be adapted for either purpose.

Some of the simplest (and most explainable) supervised techniques are regression models, which attempt to explicitly model the mathematical relationship between the independent (features) and dependent (outcomes) variables. Regression models find best-fit parameters $\mathbf{W}$ for a *a priori* specified model f by mathematical minimisation of a 'goodness of fit' metric (also called 'loss', $L(\mathbf{y}, \hat{\mathbf{y}})$) between the dependent variable observables y_i and the fitted predicted values $\hat{y}_i = f(\mathbf{W}, x_i)$; that is, solving $\arg\min_{\mathbf{W}} L(\mathbf{y}, f(\mathbf{W}, \mathbf{x}))$, where $\mathbf{y}$ and $\mathbf{x}$ are the vectors of dependent and independent variable observables $\mathbf{y} = [y_1, y_2 ... y_i]$ and $\mathbf{x} = [x_1, x_2 ... x_i]$. The loss chosen may vary, but will commonly include a likelihood function (which will be maximised, fitting the Maximum Likelihood Estimation framework), or a least-squares formulation. The loss will then be optimised; either in closed-form or numerically by an iterative (finitely terminating and ideally convergent) algorithm. [131]. For regression tasks (predicting a continuous variable), the classic models of choice are linear (that is, $f(x_i) = \sum_i w_i x_i + \epsilon_i$) and non-linear regression, and classification tasks are well served by logistic regression (that is, $f(x_i) = \frac{1}{1 + exp(-\sum_i w_i x_i)}$, which has a probabilitic interpretation.) Regression models are simple and highly explainable, but perform poorly on when the data are noisy, collinear or high-dimensional. The supervision in regression comes from pairwise labelling - the imposed association of observables into distinct combinations (x_i, y_i).

Unlike regression, instance-based models focus on local relationships (again, a distance metric) between individual observables, on the assumption that 'similar' observables exist 'close' together in some space. k-Nearest Neighbours (kNN) uses a weighed vote of k nearest points when deciding the outcome of a test point. Support Vector Machines (SVM) instead find a maximum-margin (ie. separation) hyperplane that separates classes, either in the original space or a higher-dimensional one. Instance-based models scale well with data dimension, but poorly with data quantity - making them useful for complex problems with limited amounts of training data needed.

Another major family of supervised algorithms are decision trees - a family of techniques that aim to partition the feature space such that 'similar' observables are grouped together, and from such grouping, inferring a sequence of binary decisions. Decision trees are highly explainable and non-parametric, but mathematically the problem of finding the optimum tree is non-deterministic - this can make model optimisation challenging in some cases.

The family of algorithms of most relevance to us in this thesis are the Artificial Neural Networks (ANN), and we will devote most time to those. An ANN consists groups of individual neurons (width) connected into layers (depth) - most networks are unidirectional (feed-forward) through the layers, although cyclic connections create Recurrent Neural Networks, suitable for learning time-series relationships. Each neuron implements a weighted summation of its inputs, applies an activation function to the sum, and outputs the result. Figure 1.4.1 shows the functional representation of a neuron, mathematically summarised by Equation 1.4.1:

$$x_{i,j+1} = \phi(\sum_i w_{i,j} x_{i,j} + b) \tag{1.4.1}$$

where ϕ is the activation function, $x_{i,j}$ is the feature i of layer j, w the weight and b a linear bias term (in actual practice, the bias term is rolled up as an extra dimension of w). Equation 1.4.2 shows the usual formulation in matrix form:

$$\mathbf{X_{i+1}} = \phi(\mathbf{W X_i}) \tag{1.4.2}$$

ANNs are not inherently novel - their recent rise to stardom has been fuelled by the simultaneous rise of cheap compute from consumer GPUs (and their open-source drivers such as CUDA [132]) and access to large datasets, as well as access to open-source libraries which implement efficient tensor operations and automatic differentiation that power backpropagation (such as TensorFlow [133] or PyTorch [134]), and high-level APIs that accelerate prototyping (such as Keras [135]).

At this point we can recognise that by using the identity activation $\phi(x) = x$ we recover the case of linear regression; while applying a sigmoid $\phi(x) = \frac{1}{1+exp(-x)}$ recovers logistic regression. It can also be shown that both the kNN and SVM can be reproduced in ANN form. This trend continues - it has been shown that an ANN of arbitrary width and/or depth can approximate any continuous function on

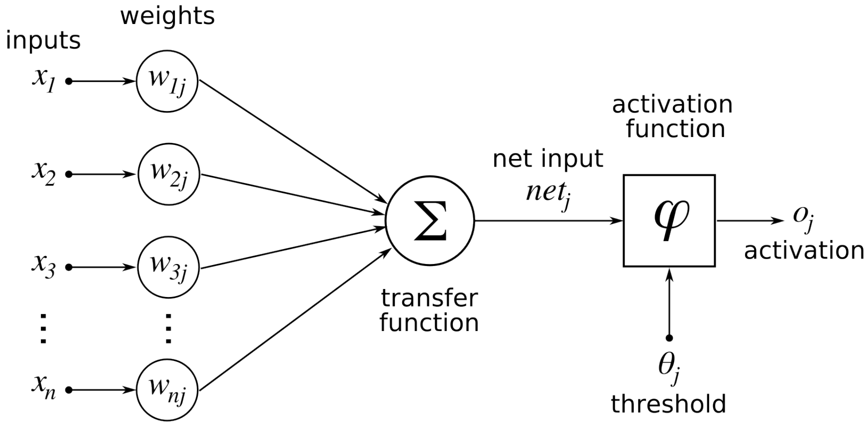


Figure 1.4.1: Functional representation of an artificial neuron Every individual neuron in an ANN integrates a weighed sum of its inputs, applies an activation function to the sum, and outputs the result. *Figure reproduced from: Chrislb, CC BY-SA 3.0 http://creativecommons.org/licenses/by-sa/3.0, via Wikimedia Commons*

a Euclidian space to any degree of precision. [136]. This property requires depth [137] and a non-linear activation function, but results have been extended to ANNs of bound width [138] and different architectures [139]. [1] Conversely, this means any continuous function has a numerical ANN representation.

Like any other technique, we need to attach a loss function to the output against which to optimise. The choice depends on the computational task. For classification tasks, a common choice is the cross-entropy:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = -\frac{1}{m}\sum_{i}^{m} y_i \, log(\hat{y}_i) \tag{1.4.3}$$

where m is the number of observables, y is the observable value, and $\hat{y}$ is the predicted value. It can be shown that minimising cross-entropy is equivalent to maximising the likelihood of a Bernoulli distributed variable - explaining the application in discrete classification tasks.

The ANN is optimised numerically by gradient descent - minimising $L(\mathbf{y}, \hat{\mathbf{y}})$ with respect to the weights $\mathbf{W}$, to find the optimum weights. In a wide and deep network, $L(\mathbf{y}, \hat{\mathbf{y}})$ is high dimensional. Gradient descent amounts to evaluating the

---

[1] ANNs are not the only techniques with universality properties. For example, well-constructed SVMs also have this property [140]. However, this only guarantees the existence of weights, not that it is feasible to learn them. SVMs, unlike ANNs scale poorly with data volume which makes them poorly suited for large-scale tasks that ANNs excel at.

vector of partial derivatives $\nabla_{\mathbf{W}} L(\mathbf{y}, \hat{\mathbf{y}})$ with respect to the weights, and updating the weights in its direction by step size h - also called the 'learning rate'. Our networks have multiple layers - the 'forward pass' computation graph is a composite function the form of Equation 1.4.2. $\nabla_{\mathbf{W}} L(\mathbf{y}, \hat{\mathbf{y}})$ may be now decomposed into a summation of partial derivative products using the chain rule along the forward pass graph. Naive computation of this may not be feasible. However, it is possible to compute the derivative with the same time complexity as the function itself by using backpropagation, which computes the partial derivatives recursively from the end of the graph - the derivatives of layer j are completely specified by the derivatives of layer $j+1$. This works for any differentiable function.

It might not be practical to perform the gradient descent step (also called 'parameter update') using all available training data - in practice the available training data are divided into batches ('batch size' is one of the key hyperparameters), and a parameter update happens after each batch (this is sometimes called Mini-batch Gradient Descent). An 'epoch' consists of all batches.

We see the 'online' nature of Mini-batch Gradient Descent means it scales well with increasing volume of data (in fact, showing best possible, linear scaling), backpropagation will handle any continuous loss function, and the universality conditions allow ANNs to model any continuous function. Together, these properties make ANNs uniquely well suited for a wide range of large scale modelling tasks.

The loss function formulates the optimisation problem, and gradient descent provides a framework for solving it. We need to introduce one last piece - the actual algorithm [141] which will walk down the loss landscape: turning $\nabla_{\mathbf{W}} L$ into $d\mathbf{W}$. One could naively partition the training samples between batches with each epoch. ($d\mathbf{W} = \frac{h}{|B|} \nabla_{\mathbf{W}} \sum_{\mathbf{y}, \hat{\mathbf{y}} \in B} L(\mathbf{y}, \hat{\mathbf{y}})$ - the gradient of the loss function evaluated and normalised over batch B, multiplied by learning rate h). However, this creates noise, and does not allow the algorithm to escape local extrema in the high-dimensional loss landscape (where $\nabla_{\mathbf{W}} L = 0$). Instead of using the derivative, we introduce a weighed average of previous gradients, called 'momentum'. To do so, we replace the derivative with a 'velocity', which accumulates with previous gradients. The strength of the accumulation is controlled by a new hyperparameter β (we're also switching to $\mathbf{g_t}$ for convenience):

$$\mathbf{v}_t = \beta\,\mathbf{v}_{t-1} + \frac{1}{|B|}\,\nabla_{\mathbf{W}} \sum_{\mathbf{y},\hat{\mathbf{y}} \in B} L(\mathbf{y}, \hat{\mathbf{y}}) = \beta\,\mathbf{v}_{t-1} + \mathbf{g}_t \tag{1.4.4}$$

$$d\mathbf{W} = h\,\mathbf{v}_t \tag{1.4.5}$$

Random partitioning without replacement with added momentum creates the Stochastic Gradient Descent optimizer (SGD) [2]. Further refinement to momentum can be done by calculating the gradient in the velocity term with respect to anticipated parameters at $t+1$ rather than t and then correcting the difference [142], creating Stochastic Gradient Descent with Nestrov Momentum (SGD+N). It is good practice to decay the learning rate during training, which aids in convergence. However, in practice many feature spaces are quasi-sparse - certain features occur more frequently than others in the training data. Uniform decay of h across all parameters would then lead to poor convergence for the less-frequent features. This observation leads us to a family of adaptive algorithms. Rescaling h by the vector of the sum of the squares of the gradients seen up to step t creates Adagrad, decaying the sum of squares like momentum does with gradients creates RMSProp; taking Adagrad and estimating h dynamically from $d\mathbf{W}$ creates Adadelta. Combining RMSProp with momentum creates Adam, addition of Nestrov momentum creates NAdam. In Adam, addition to velocity, we also accumulate the sum of squares $\mathbf{s}_t$; β_1 and β_2 control the accumulation (note also the renormalization of $\mathbf{v}_t$ and $\mathbf{s}_t$) :

$$\mathbf{v}_t = \beta_1\,\mathbf{v}_{t-1} + (1-\beta_1)\,\mathbf{g}_t \tag{1.4.6}$$

$$\mathbf{s}_t = \beta_2\,\mathbf{s}_{t-1} + (1-\beta_2)\,\mathbf{g}_t^2 \tag{1.4.7}$$

$$d\mathbf{W} = h\,\frac{\mathbf{v}_t}{1-\beta_1^t}\,\frac{1}{\sqrt{\frac{\mathbf{s}_t}{1-\beta_2^t}}+\epsilon} \tag{1.4.8}$$

We make heavy use of both Adam and SGD in this thesis.

---

[2] Exact nomenclature varies.

Before we move onto computer vision applications, we should address some secondary points. Deep ANNs have a lot of representational capacity and are thus prone to overfitting - modern ANNs feature regularisation mechanisms that penalise complexity, which include norm constrains or penalties on $\mathbf{W}$, and layers such as Dropout [143].

1.4.3 Neural Networks in Computer Vision

The classic tasks for computer vision include classification and segmentation - understanding the composition of scenes captured by digital cameras. The challenge is the data - which is highly dimensional and heterogeneous. Objects of interest are local, occupying patches close together in pixel-space, delineated by edges or textures, but can vary in their position (translation), rotation, perspective and illumination. As an illustrative example - a single 4k Ultra High Definition image in full colour (the standard TV display resolution at the time of writing) contains $3840 * 2160 * 3 = 24883200$ pixels, or independent dimensions. This makes feature extraction and engineering difficult. Traditional approaches to feature engineering include the construction of filters that recognise edges or geometrical shapes, attempt to identify and extract significant keypoints in a structured manner, or template-match against a reference - a full review is beyond our scope, but is well covered by Szeliski [144] and other textbooks.

The revolution came in the shape of Convolutional Neural Networks (CNNs) - a class of ANNs that feature a hierarchy of learnable convolutional filters [3]. These filters are now part of $\mathbf{W}$ and learnt during training. This means the CNN in training engineers complex features across all scales of the image, perfectly suited for the task at hand (as defined by $L(\mathbf{y}, \hat{\mathbf{y}})$). CNNs maintain the universality conditions of ANNs [145]; further, the convolution operation is translation equivariant and local, which addresses many of the problems we mentioned. It is worth mentioning that nothing we said so far cannot be achieved with vanilla ANNs - the true power of CNNs is not that the convolution is uniquely powerful at feature engineering,

---

[3] Strictly speaking, CNNs generally implement a cross-correlation and not a convolution, but the difference is trivial.

but rather that it is *exceptionally* efficient parameter-wise. An equivalent vanilla ANN would be too large to handle efficiently.

It is time for some mathematics. From an input $X(i, j, C_{in})$ (which could be an image with C_{in} channels), a convolution layer constructs a bank of feature maps C_{out} deep. Successive layers are parameterised such that C_{out} increases and the spatial size decreases, creating a hierarchy of features.

$$X_{i+1}(x, y, c_2) = \phi\,(\sum_{c_1\, \in \{C_{in}\}} \sum_{(n,m) \in K} w(n, m, c_1, c_2)\, X_i(x + n, y + m, c_1) + b(c_1)) \tag{1.4.9}$$

where (x, y) are the spatial coordinates, c_1 is an input channel is the set of input channels $\{C_{in}\}$, c_2 an output channel in set $\{C_{out}\}$, (n, m) the spatial coordinates of filter kernel K, and ϕ the nonlinear activation function.

Beyond convolutional layers, modern CNNs also feature pooling operations which reduce the dimensionality and provide a degree of translation invarience, and batch normalisation, which re-normalise inputs between layers and accelerate convergence.

The first modern CNN that substantially outperformed classic computer vision approaches in a classification task was AlexNet [146], followed closely by the VGG [147] series of models that used repeating convolutional blocks followed by fully connected layers, and Network in Network [148] models that disposed of fully connected layers in favour of 1x1 convolutions. Then, Inception concatenated multi-branch convolutions with different kernel sizes [149]. The next big improvement came from residual models, which introduce identity skip connections to modify the forward pass to $\mathbf{X}_{i+1} = \phi(\mathbf{W}_i\mathbf{X}_i + \mathbf{X}_i)$ between layers that help gradients propagate through deeper networks. This created the ResNet [150] series and later ResNext [151]. The generalization of the residual concept, and also our primary feature encoder of choice, is the DenseNet [152], which introduces concatenation of features between blocks. This has the advantage of allowing feature reuse - each layer i has access to both the composite features encoded until i, and each intermediate feature map separately. The disadvantage is the additional computing and memory requirements in the feature reuse - which in our context we judged to be

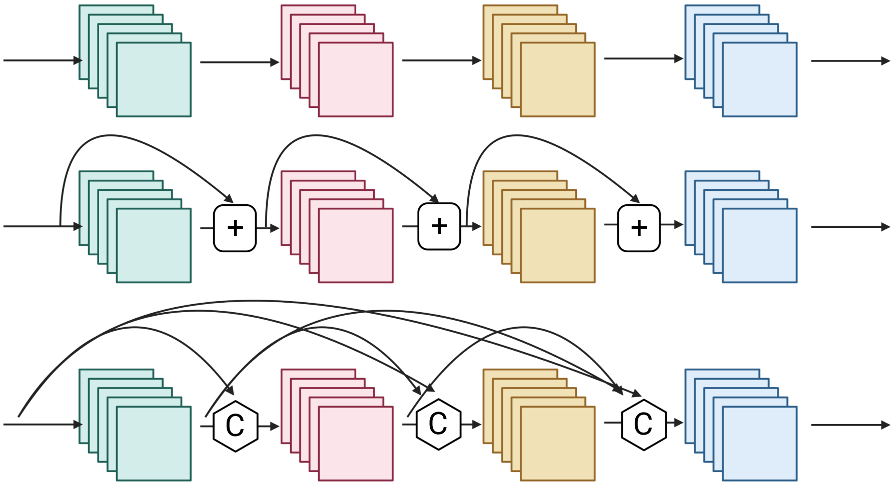


Figure 1.4.2: CNN architectures visualised *(Top)* A classic, linear CNN, including AlexNet, VGG, Inception. The feature maps are propagated linearly through the network, layer by layer. *(Middle)* A residual network with skip connections which implement addition. Examples include ResNet and ResNext. Skip connections still propagate feature maps linearly, but allows bypassing of groups of layers. *(Bottom)* A generalised residual network with concatenation operations. Examples include DenseNet. Concatenation connections allow non-linear propagation. Figure created with Biorender.

a non-issue. Figure 1.4.2 shows an illustrative cross-section of the main connectivity patterns. Further developments have focused on automatic architecture discovery, and reducing the memory and computing footprint without sacrificing performance - notable results of this include MobileNet [153] and EfficientNet [154].

With great power comes a great electricity bill [4], and also a great responsibility. Since the CNN will engineer features such that the loss function is minimised, great care must be taken that the features are meaningful, and not external bias. Methods for CNN explainability include approximating the decision making with a simpler model, perturbation approaches, and tracing backpropagation steps - Ibrahim provides a detailed overview [155]. This is particularly important in our context - both the clinican and regulator needs detailed knowledge of why and how clinically-relevant decisions are taken. In this thesis we use a technique called

---

[4] This is only a partially a joke. Recent analysis by OpenAI shows the computing requirements for training models are currently doubling every 3.4 months. *https://openai.com/blog/ai-and-compute/*

saliency mapping - we calculate the gradient of the final output with respect to the input image, thus highlighting regions that contributed most to the output.

By attaching a cross entropy loss to the DenseNet (Equation 1.4.3), we can create a classifier - which is one of our main tools in this thesis.

To create a segmenter, we need to be more creative - the challenge of segmentation is coupling classification with localisation of objects in the image, outputting bounding boxes of object instances. One approach is to first propose suitable bounding boxes and then classify them. This two-stage approach has resulted in the family of Region-based Convolutional Neural Networks (RCNN). RCNN [156] operated an explicit search algorithm in image space followed by a CNN classifier; Fast-RCNN [157] operated it in feature space to save time. Faster-RCNN [158] introduced a Region Proposal Network for further acceleration. Instead of a single feature map, it is helpful to operate on a pyramid of features - a network with bottom-up downscaling, top-down upscaling and horizontal connections creates a Feature Pyramid Network (FPN).

It is possible to do object detection in a single stage - prominent examples include YOLO [159] and DETR [160]. Single-stage models are much faster and less computationally intensive, however, they generally offer poorer performance. Since we are not constrained by either time or resources in this work, we chose to work with a two-stage architecture.

It is helpful for us to localise objects by computing bounding boxes; it is more helpful to fully segment them by delineating their precise edges. This produces a localisation map with pixel-accuracy. This can be achieved by adding a regressor layer that regresses the mask within the bounding box output. Doing this to Faster-RCNN yields Mask-RCNN [161], similar modifications can be done to single-stage networks. We use a customised Mask-RCNN as our segmenter of choice in this thesis.

Before we move on, we should take a second to consider where the field of computer vision is currently going. Just as CNNs provided the main research direction over the past decade, the field is currently undergoing another transformation by a class of models called transformers [5]. Originally a family of sequence-to-sequence encoder-decoder models used in Natural Language Processing, transformers utilise a self-attention mechanism in the encoding and decoding pathways, as well as between them. This enables the model to learn complex relationships between individual features in both the input and target spaces, and are thus well suited for learning context. Unlike the CNN convolution kernel, this process is not forced to be local, providing greater freedom of association of features; further, the same transformer architecture is able to handle simultaneous input of different types of features. In a way, the transformer can be seen as yet another step along the way to a universal feature engineering paradigm, fitting well with the advances and needs for multimodal AI systems. The exact impact on computer vision remains to be seen; at the time of writing, Vision Transformers have begun to outperform state-of-the-art CNNs in standard tasks. [162]

---

[5] In fact, DETR is one of the first examples of transformers in computer vision.

Some people worry that artificial intelligence will make us feel inferior, but then, anybody in his right mind should have an inferiority complex every time he looks at a flower.

— Alan Kay

2

Deep Antimicrobial Susceptibility Phenotyping (DASP) assay design, development and training.

Contents

2.1 DASP Concept 41
2.2 DASP Design and Training 43
  2.2.1 Phenotypes 43
  2.2.2 Training and Computational Experiments 48
2.3 DASP Validation Results 53
  2.3.1 Segmentation 53
  2.3.2 Phenotype Classification 56
2.4 Classifier Explainability 59
2.5 Experimental Methods 63
2.6 Discussion 70
  2.6.1 Segmentation and classification of phenotypes 70
  2.6.2 Experimental design 72
  2.6.3 Future directions in optimising phenotypes 74
  2.6.4 Generalisability of findings 75
2.7 Conclusions 76

2.1 DASP Concept

To address some of the issues we identified in the current AST landscape, we designed a candidate AST that uses sub-cellular phenotypes of the bacterial nucleoid and membrane, revealed by wide-field fluorescence microscopy and selective fluorescent staining of the subcellular structures. In susceptible organisms, antibiotic action induces characteristic and reproducible changes in the organisation and morphology of the nucleoid - creating a distinct, single cell-specific ‘susceptibility’ phenotype. By contrast, in resistant organisms, these changes either do not occur, or are distinctly weaker - the cells retain an ‘untreated’ appearance. The ‘resistant’ phenotype is thus similar to the ’untreated’ phenotype.

Wide-field fluorescence microscopy is able to capture large fields of view (FoV) which contain numerous individual bacteria displaying such phenotypes. Automating the microscopy (already available commercially as standard in most high-end optical microscope setups) allows automated acquisition of large volumes of data.

To develop an AST, we combined automated wide-field fluorescence microscopy with a 2-stage CNN pipeline that automatically segments individual bacteria in every FoV, and then performs 2-class, binary classification into either the ‘resistant’ or ‘susceptible’ class, based on their phenotypic response to antimicrobial action. Treated cells that are susceptible show the expected phenotypic response and are thus classified as ‘susceptible’; conversely, resistant cells do not show the reponse and resemble untreated cells, and are classified as ‘resistant’. Taken together, this achieves Deep Antimicrobial Susceptibility Phenotyping (DASP) of single cells by the specific effect of an antibiotic on subcellular structures - its targets.

The AST takes as an input bacterial cultures grown to a particular OD. Samples of the culture are then treated with antibiotics at a pre-determined concentration and time, after which the cultures are inactivated by chemical fixation with formaldehyde - this kills the bacteria whilst preserving their internal structure (and thus the phenotype). To reveal the phenotype, the cell membranes are permeabilised by incubation with alcohol, followed by fluorescent staining of the DNA (thus revealing the nucleoid) and phospholipids (thus revealing the cell membrane). Such prepared

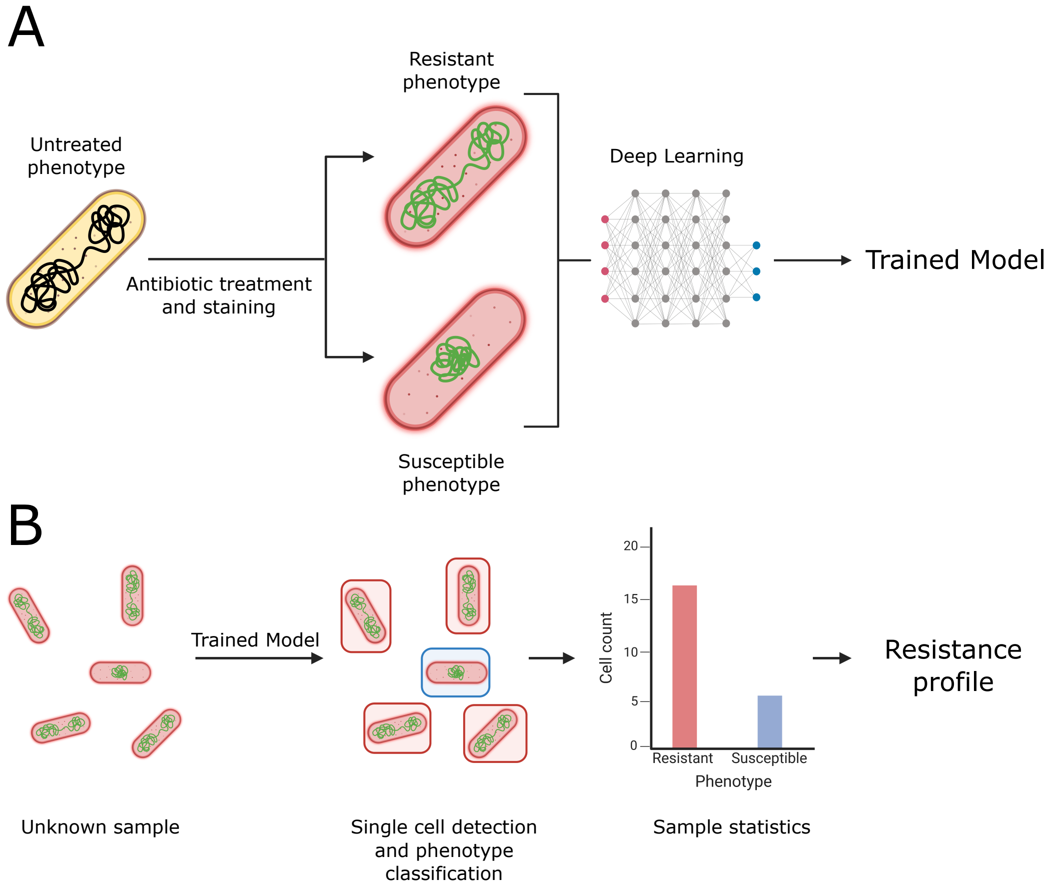


Figure 2.1.1: DASP: an AST assay based on single-cell phenotyping and deep learning.
A. Live *E.coli* cells are treated with an antibiotic, which induces changes in the organization of nucleoid and membrane. Suceptible cells show strong phenotypic changes, while resistant cells are not affected and resemble untreated cells. Cells are fixed and the phenotype is revealed by fluorescent staining. A deep learning pipeline is trained to distinguish the susceptible phenotype from the resistant (untreated) phenotype, with single cell resolution.
B. An unknown sample can be processed and fed into the trained model, which classifies the phenotypes on a single cell level to produce sample-wide classification statistics. These statistics can then be used to obtain information on the resistance of the entire sample.

cells are then imaged under a fluorescent microscope in a semi-automated fashion; the 2-stage CNN pipeline can then operate directly on the acquired micrographs to segment and classify phenotypes without further processing.

Figure 2.1.1 illustrates the concept we just outlined above, while Figure 2.1.2 shows the basic experimental flow.

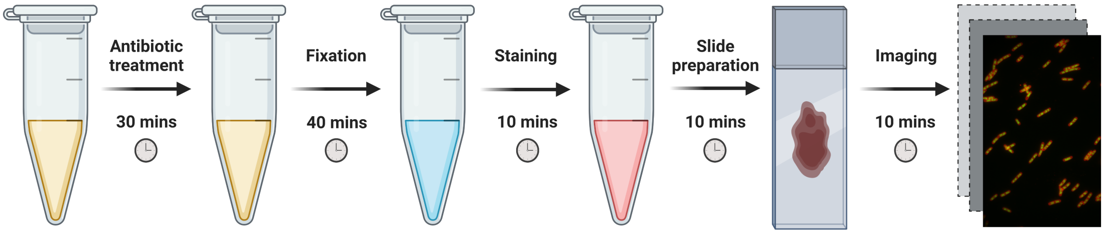


Figure 2.1.2: DASP experimental wetlab process. Live bacterial cultures are treated with an antibiotic of choice at a pre-determined concentration and time, fixed with formaldehyde and permeabilised with ethanol. Fixed cells are fluorescently stained with DAPI (nucleoid) and Nile Red (membrane). The solution is placed on an agarose pad slide and imaged on a widefield fluorescence microscope. Figure created with Biorender.

2.2 DASP Design and Training

2.2.1 Phenotypes

To implement the concept outlined above, we characterised a series different susceptible phenotypes to different antibiotics in the lab-reference *E.coli* strain MG1655, as well as the untreated phenotype (which we used as proxy for the resistant phenotype). We used 4 clinically relevant antibiotics: ciprofloxacin (CIP), gentamicin (GENT), rifampicin (RIF) and co-amoxiclav (COAMOX). For further details and experimental conditions, see *Experimental methods* section 2.5 and figure 2.5.1. We used 4,6-diamidino-2-phenylindole (DAPI; green signals in Figure 2.2.1 and beyond) to stain the nucleoid, and Nile Red (NR; red signals in Figure 2.2.1 and beyond) to stain the cell membrane, which revealed the phenotypes under a fluorescent microscope.

Figure 2.2.1 shows the phenotypes we are working with in this thesis.

In the untreated phenotype (2.2.1 A), we could observe distinct copies of the chromosome in each cell - exact number depends on culturing conditions, in our case cells were cultured in rich medium leading to a high chromosome count. The chromosomes were organised into heterogenous macrodomains by nucleoid-associated proteins - we could see distinct heterogeneity in the nucleoid structure.

In contrast, treatment with the fluoroquinolone ciprofloxacin (CIP) produced a compaction of the chromosome towards the centre of the cell (or occasionally one

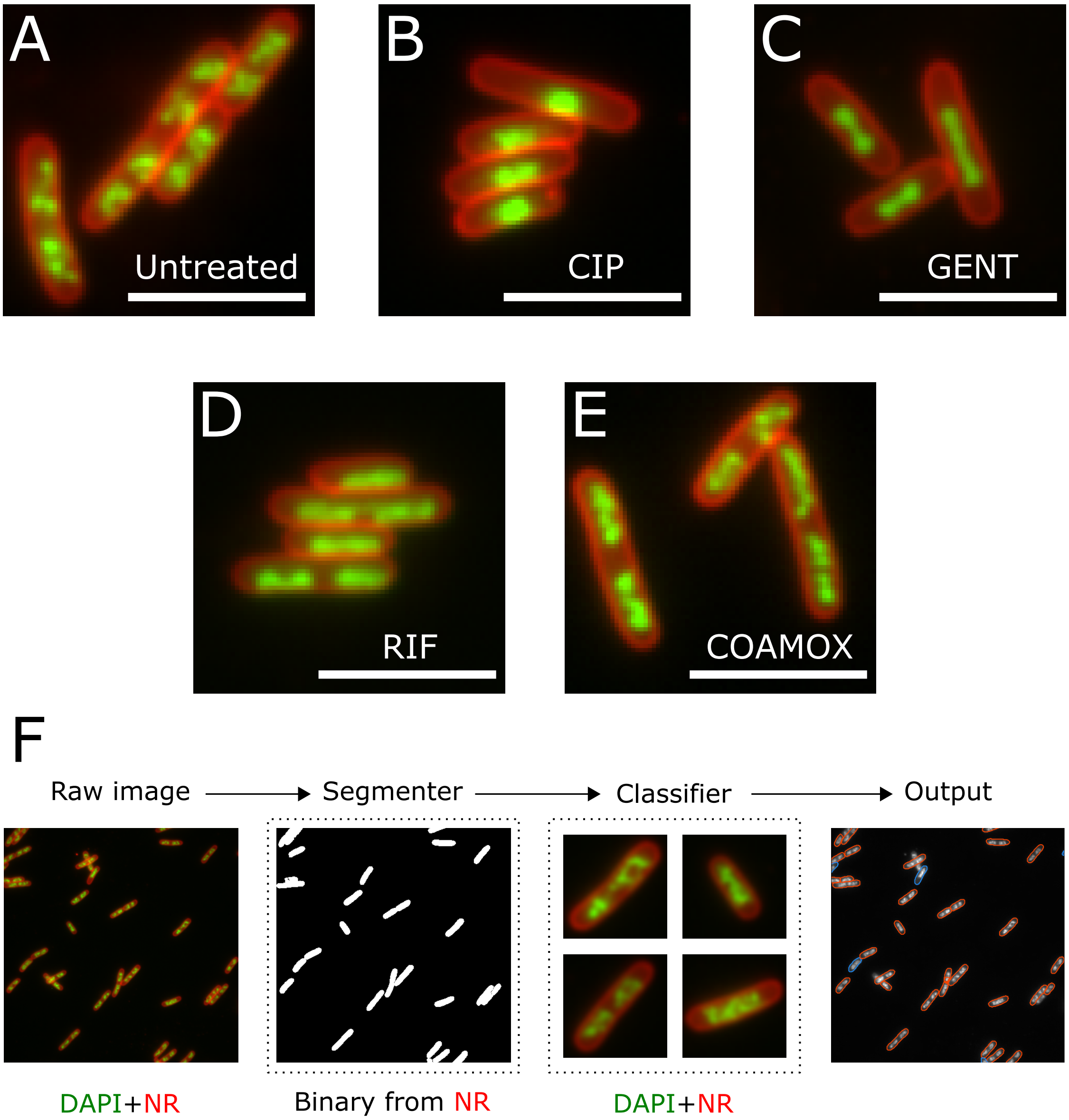


Figure 2.2.1: Phenotypes, and analysis pipeline.
A. Untreated phenotype of MG1655, which resembles the resistant phenotype. Scale bar indicates 5µm.
B. Ciprofloxacin susceptible phenotype.
C. Gentamicin susceptible phenotype.
D. Rifampicin susceptible phenotype.
E. Co-amoxiclav susceptible phenotype. Note, this phenotype is morphologically similar to the untreated/resistant phenotype.
F. A multichannel image containing both the DAPI (green, nucleoid) and Nile Red (red, membrane) signals is split into individual channels. A Mask-RCNN segmenter segments single-instance binary masks from the Nile Red channel, which are then used to isolate individual cells. A phenotype classifier classifies individual cells into either the resistant or susceptible class, using both channels. Taken together, the output of the segmenter and classifier allows precise localisation of single cells within the original FoV by their phenotype.

of the poles) due to DNA damage accumulation and gyrase and topoisomerase-IV inhibition (2.2.1 B). Similarly, treatment with aminoglycoside gentamicin (GENT), which binds to the 30S ribosomal subunit, also lead to nucleoid compaction - although in this case the chromosomes did not fully merge into one spot (2.5.1 C). Treatment with the ansamycin rifampicin (RIF), which targets RNA Polymerase, lead to a decompaction of the nucleoid (2.5.1 D); we could still distinguish individual chromosomes, but the heterogeneity of organisation evident in the untreated phenotype was lost. Finally, we also treated cells with co-amoxiclav (COAMOX) (2.5.1 E) - the β-lactam amoxicillin with a β-lactamase clavulonic acid - interestingly, this time the phenotypic response was weak, and the susceptible phenotype resembled the untreated/resistant phenotype.

Figures 2.2.2 and 2.2.3 show representative FoVs showing the various phenotypes we have discussed.

To develop the phenotypes, we experimentally optimised the treatment conditions (antibiotic concentration and treatment time) to minimize treatment time - preliminary experiments showed a concentration of 20x EUCAST breakpoint (Figure 2.5.1) to be a good value. Nevertheless, we should reasonably expect different treatment conditions to produce different phenotypes. For example, Nonejuie et al [118] treated *E.coli* with 0.025 mg/L of ciprofloxacin for 2 hours (5x MIC for their specific strain), compared to our treatment conditions of 10 mg/L for 30 minutes (∼1000x MIC for our strain). In their phenotype, they observed nucleoid compaction too, but it was accompanied by dramatic cell elongation, with cells continuing to grow to over 10x their usual size, but without cell division. Since we are aiming our work towards a diagnostic endpoint (where we do not have *a priori* access to the MIC), we chose the EUCAST guidelines to anchor our treatment conditions, instead of the MIC. To expand the applicability of our platform, further experiments will be required to optimise the treatment conditions of multiple antibiotics for multiple possible pathogen species.

Our assay began with a culture step to a defined OD, which had two purposes. Firstly, this controlled the number of cells that entered the pipeline. If too few cells

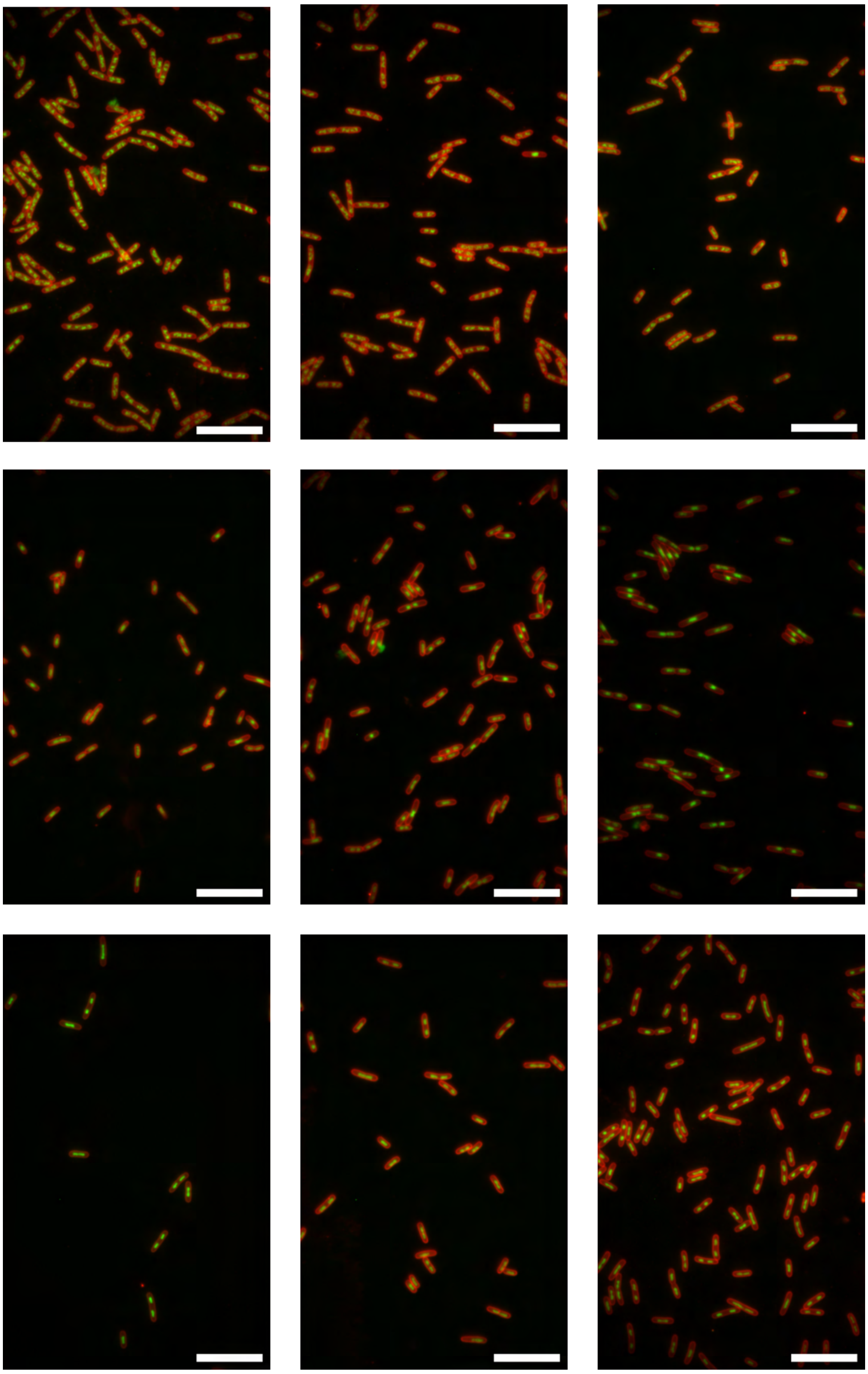


Figure 2.2.2: Representative FoVs of untreated, ciprofloxacin-treated and gentamicin-treated cells. *Top row* 3 representative FoVs showing the untreated phenotype, where no antibiotic treatement was used. *Middle row* As above, but for ciprofloxacin treated cells. *Bottom row* As above, but for gentamicin treated cells. Scale bar indicates 5µm.

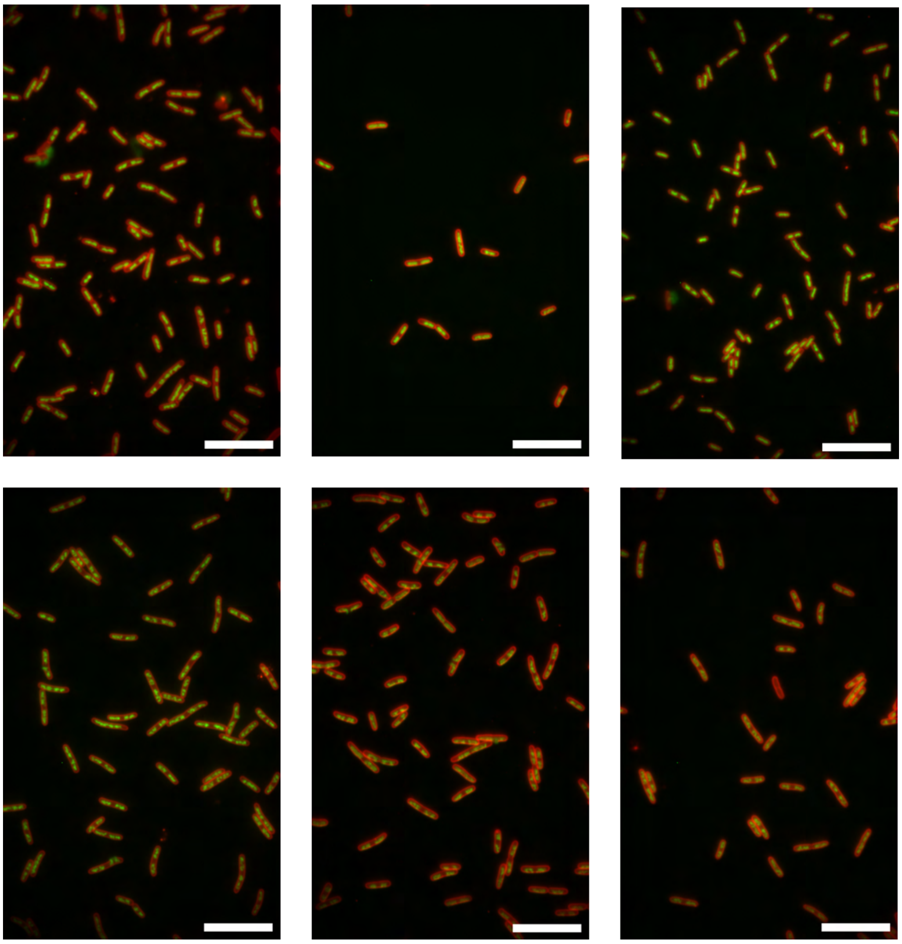


Figure 2.2.3: Representative FoVs of rifampicin-treated and co-amoxiclav treated cells. *Top row* 3 representative FoVs showing the rifampicin susceptible phenotype, *Bottom row* As above, but for co-amoxiclav treated cells. Scale bar indicates 5µm.

had entered, we would not have sufficient single-cell statistics to draw conclusions. Conversely, if too many cells had entered, cells could have aggregated, preventing us from obtaining high resolution imaging of single cell phenotypes in monolayers; this would have also affected our ability to segment them correctly. Secondly, this also ensured all cells were at the same stage of growth. As a bacterial culture grows, the cells actively undergo epigenetic changes - cells in exponential phase differ in their gene expression from cells in stationary phase. By culturing to an OD of 0.2, we ensured all cells were in the exponential phase, where their metabolic activity is the highest - coincidentally, this is also the phase where antibiotics (metabolic inhibitors) are considered most active.

Following the culture step and experimentally-determined antibiotic treatment, the cells underwent fixation and permeabilisation. These steps were needed to stop the antibiotic action at the desired stage, inactivate the sample, preserve the phenotype and allow efficient staining of the structures that were of interest to us.

Following those steps, we stained the nucleoid (DNA) with DAPI, and cell-membrane (peptidoglycan) with Nile Red. The exact choice of dyes was determined by our microscope setup,

To actually image our stained samples, we used a semi-automatic microscope with a single, fixed, oil-immersion objective. Samples were mounted one-at-a-time on agarose pads (Figure 2.5.2 and imaged inverted through the top cover slip - immobilisation of the bacteria was thus achieved through mechanical pressure of the agarose on top of the bacteria (from the point of view of the objective). This allowed us to obtain high-resolution images of the phenotypes with minimum optical distortion, and the mounting method was agnostic to bacterial species, as the bacteria were mechanically held against the clean top cover slip without relying on any specific chemical interactions between the bacterium and cover slip surface.

2.2.2 Training and Computational Experiments

Figure 2.1.1 F shows the 2-step computation pipeline we designed to achieve rapid and robust detection of the different phenotypes. We performed 7 independent repeat imaging experiments (See Figure 2.2.4 A; each following the outline of Figure 2.1.1; see *Experimental Methods* section 2.5 for further details), which produced 7 independent collections of micrographs, containing cells that are untreated, and also those showing the susceptible phenotype of each of the antibiotics we chose to work with.

As mentioned before, we used a customised Mask-RCNN on a ResNet50 backbone as the first stage, and trained it on Nile Red (membrane) signals from all phenotypes coming from the first 6 repeat experiments. All FoVs from those first 6 experiments were divided into the training and validation set randomly without replacement, creating a training set of 459 FoVs containing 29297 individual

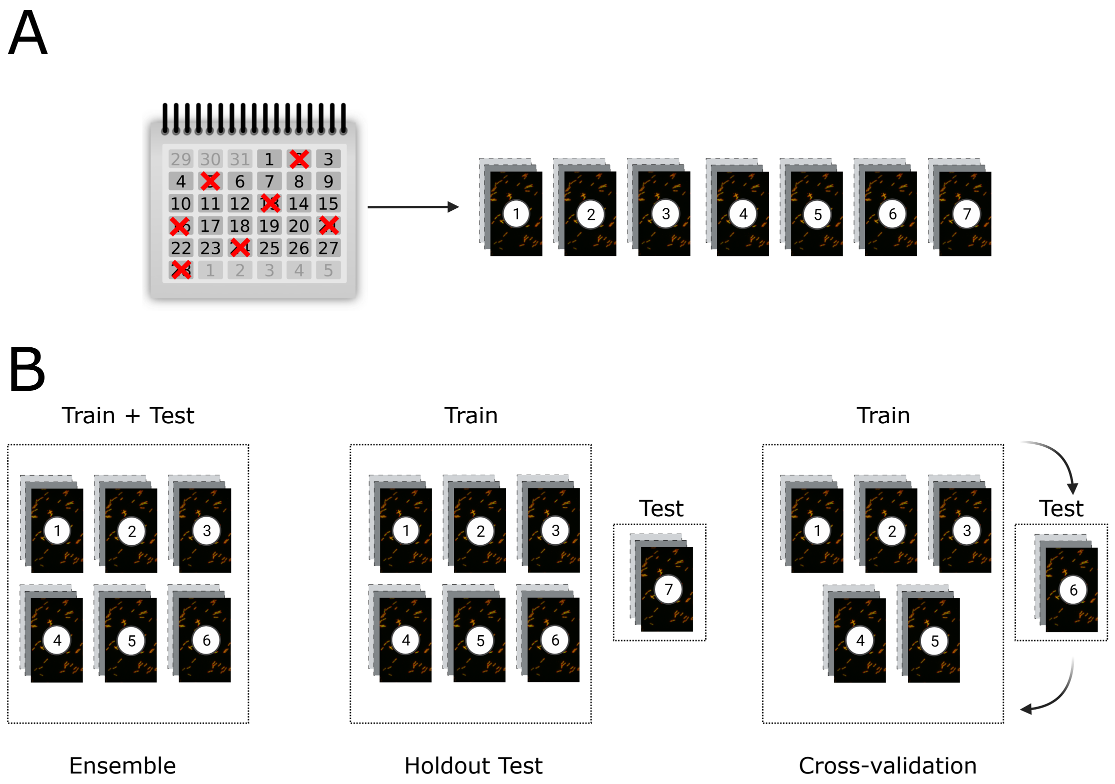


Figure 2.2.4: Setup of computational experiments
A. 7 independent repeat wetlab experiments were carried out on 7 different days, each producing a collection of micrographs of cells showing different phenotypes.
B. Experiments were divided into a range of computational setups. In the 'ensemble' setup (*Left*), all cells from the first 6 experiments were put together into a single dataset, from which the training, testing and validation sets were drawn randomly without replacement. This quantifies the expected upper bound of classifier performance - where the model had the chance to consider all experiments. In the 'holdout test' (*Middle*), the model was trained on a class-balanced dataset containing the same number of cells from each of the 6 experiments, and evaluated on a class-balanced dataset from the 7th, holdout experiment. The holdout experiment was not used in training or hyperparameter optimization, making it an unbiased estimator of model performance on unseen data. In the 'cross-validation' (*Right*), a model was trained on a class-balanced dataset from 5 of the 6 experiments, and tested on the 6th. The test experiment was rotated around; the final reported value is the average over 6 different models trained and tested on different permutations of the experiments. This setup estimates model robustness against experimental variation.

cells, and a validation set of 115 FoVs with 9044 cells. To provide an unbiased estimate of segmentation quality, the segmenter was evaluated on FoVs coming from the 7th, holdout experiment, from which no data or insight was used in either training or hyperparameter optimisation.

The second stage consisted of a DenseNet121 classifier, which we trained on both the DAPI (nucleoid) and Nile Red signals in a range of computational experiments to evaluate its robustness and performance in various contexts. For each antibiotic, we trained binary classifiers using the untreated phenotype (standing in as proxy) for the resistant class, and the antibiotic specific susceptible phenotype for the susceptible class (therefore for example, a ciprofloxacin classifier was trained on untreated and ciprofloxacin-susceptible phenotypes).

First, we again took all cells from all FoVs in the first 6 experiments into one dataset, which we divided into training, validation and testing sets by random partition without replacement - this we called the 'Ensemble' setup (Figure 2.2.4 B *Left*). For each antibiotic, a model was trained and validated on the training and validation sets, and tested on the tested set. This set up provides the expected upper bound of model performance, allowing it to see training data from all experiments.

Secondly, we trained a model on a class and experiment balanced dataset from the first 6 experiments, selecting cells randomly without replacement, and tested on a balanced dataset from the 7th, holdout experiment - from which no data was used in training or optimization (Figure 2.2.4 B *Middle*). As an illustrative example - a model was trained by selecting randomly 500 untreated and 500 susceptible phenotypes from each of the 6 experiments (total of 3000 untreated and susceptible phenotypes), and tested on 500 untreated and 500 susceptible phenotypes from the 7th experiment. Drawing equal numbers of cells from each experiments ensures there is no experimental bias - where one training experiment contributes more than any other. Drawing equal numbers of cells from each class ensures there is no class-bias, where final results are distorted simply because more examples of a class were available during training. Testing on a separate, hold-out experiment protects against human bias and provides a true test set.

Lastly, we created an experimental 'cross-validation' (Figure 2.2.4 B *Right*). Here, we again trained models on a class-and-experiment-balanced dataset taken from 5 of the 6 experiments, and tested on an equivalently balanced dataset from the 6th experiment (note, the 7th, holdout experiment was not used here). The test experiment was rotated around, producing 6 different models trained and tested on different permutations of the 6 training experiments - the final value reported is the average result over those 6 models. With this set up, we could examine the robustness of the classifier and its hyperparameters to experimental bias - seeing if changing the training data drastically impacts performance.

For further details of training, see *Experimental Methods* section 2.5. Given the volume of models trained, it is impractical to provide in this thesis all of the training and diagnostic data. However, for the reader's peace of mind and completeness of argument, we provide a short selection, focusing on the holdout test models, since these are used in later chapters to evaluate most of our key findings.

Figure 2.2.5 shows the training logs of the holdout classifiers. In all cases, we observe a decrease of the cross-entropy loss as a function of increasing epoch, reaching a saturation. Continuous validation on the validation set is used to monitor for over-fitting; the hallmark of overfitting simultaneous decrease in training loss and increase in validation loss - indicating the model is memorising training examples and not generalising well. In all cases, we observe validation loss consistently decreases with training loss, in fact, outpacing it. This is likely because the data augmentation (applied only to training data, not validation or testing data) is too harsh - nevertheless, model retains good performance and there is no cause for concern.

Figure 2.2.6 shows histograms of weights of different types of layers inside the holdout classification models. In most circumstances, a well-behaved deep ANN model will show a quasi-normal distribution of weights around their expected parameters (normally either 0 or 1, depending on the layer type). Over-or-under-determined models deviate from this - for example, an over-determined model will make its decision on a single feature, and thus might place a large weight on only one parameter, skewing the distribution away from normal.

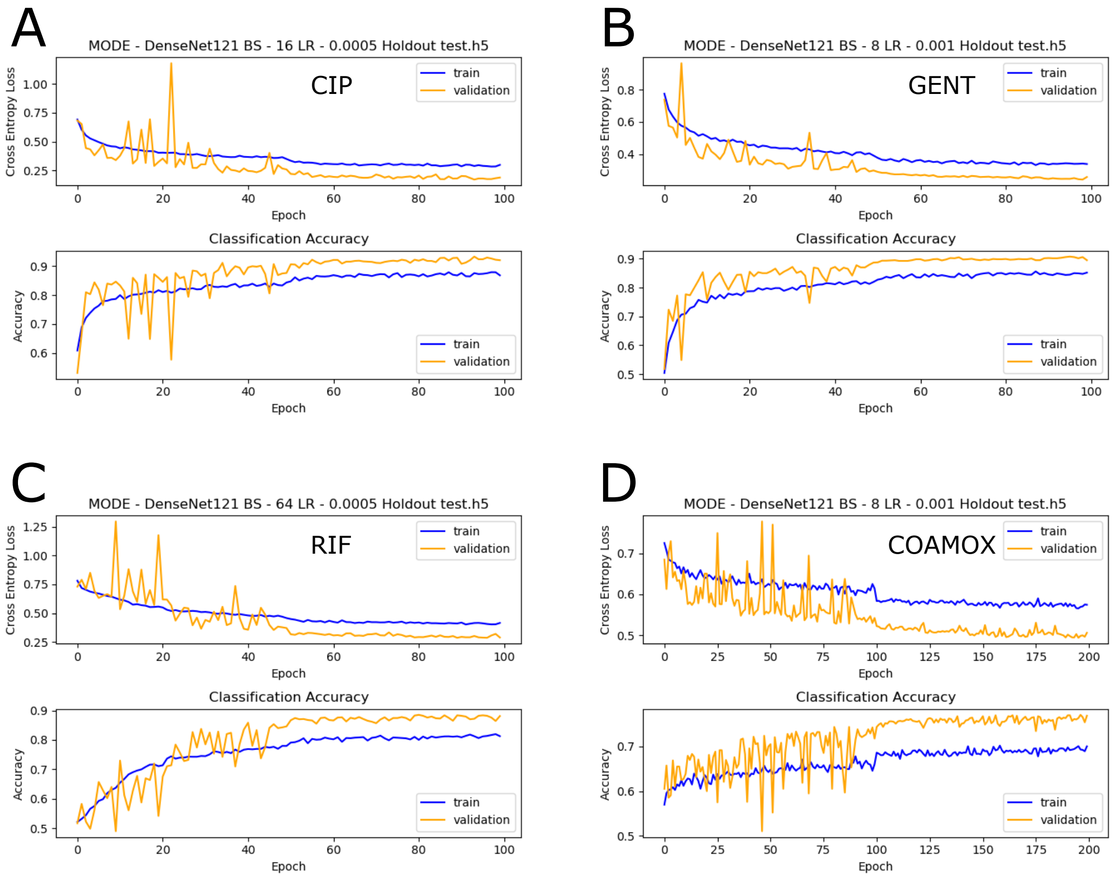


Figure 2.2.5: Example training logs of holdout classifiers.
A. *Top* Cross-Entropy training loss of the ciprofloxacin classifier. Training loss in orange, validation loss in blue. *Bottom* Corresponding classification accuracy.
B. As above, but for the gentamicin classifier.
C. As above, but for the rifampicin classifier.
D. As above, but for the co-amoxiclav classifier.

The top part of the panels shows the evolution of the weight histogram of convolution layers with increasing network depth. On the x-axis, we placed the weight histogram bins, and the network depth (index of layer in the network) on the y-axis. The colour intensity corresponds to the log frequency density in the bin. In this way, we could visualise the evolution of the weights as a function of network depth - in the case of the convolutional layers we see a rough normal distribution centred on 0, which is to be expected. In the middle panels, we did the same, but with batch normalisation layers - here we see a bimodal roughly normal distribution centered on 0 and 1. These correspond to the normalisation mean and bias (offset) of the layers - thus indicating our features are correctly scaled throughout the network. Lastly, in the bottom panels, we computed the (single) weight histogram of the fully-connected layer at the end - all preceding layers can be interpreted as the feature engineering encoder, while the final layer uses the features to compute the final classification decision. Here, we again see a quasi-normal distribution centred on 0, which is to be expected - our models are not over-focusing on any particular feature.

2.3 DASP Validation Results

2.3.1 Segmentation

To objectively quantify the quality of the segmentation from the segmenter, we evaluated our segmenter (trained on the first 6 experiments as described above) on the 7th, holdout experiment - a total test dataset of 155 FoVs containing 13,247 ground-truth, annotated cells showing various phenotypes. From this dataset, the segmenter detected 12,147 cells, resulting in total maximum recall of 92%.

To evaluate the bounds of segmenter perfomance, we computed the Precision-Recall curves at different Intersection over Union thresholds (IoU) (Figure 2.3.2 *Top*, see *Experimental methods* section 2.5 for further details). We see the same pattern at every IoU thershold - the classifier maintains strong precision as recall increases, before experiencing a sharp drop towards zero, before full recall is reached. This drop-off likely occurs due to the inherent limitations of Mask-RCNN - for every true ground-truth object, the segmenter has a tendency to produce multiple candidate regions of various quality. To filter and select from this set the single candidate region that (most likely) provides the best possible localisation, the segmenter implements non-maximum suppression - an explicit algorithm that filters candidate regions based on their mutual IoU criterion. This step is common in computer vision segmentation algorithms - since most algorithms are trained for multi-class segmentation in heterogeneous images, objects of the same class are rarely in close proximity. In our case, we trained Mask-RCNN for 2-class segmentation (cell and background); in very dense FoVs (Figure 2.3.1 *Left*), the non-maximum suppression filters struggled to distinguish between spurious candidate regions generated for

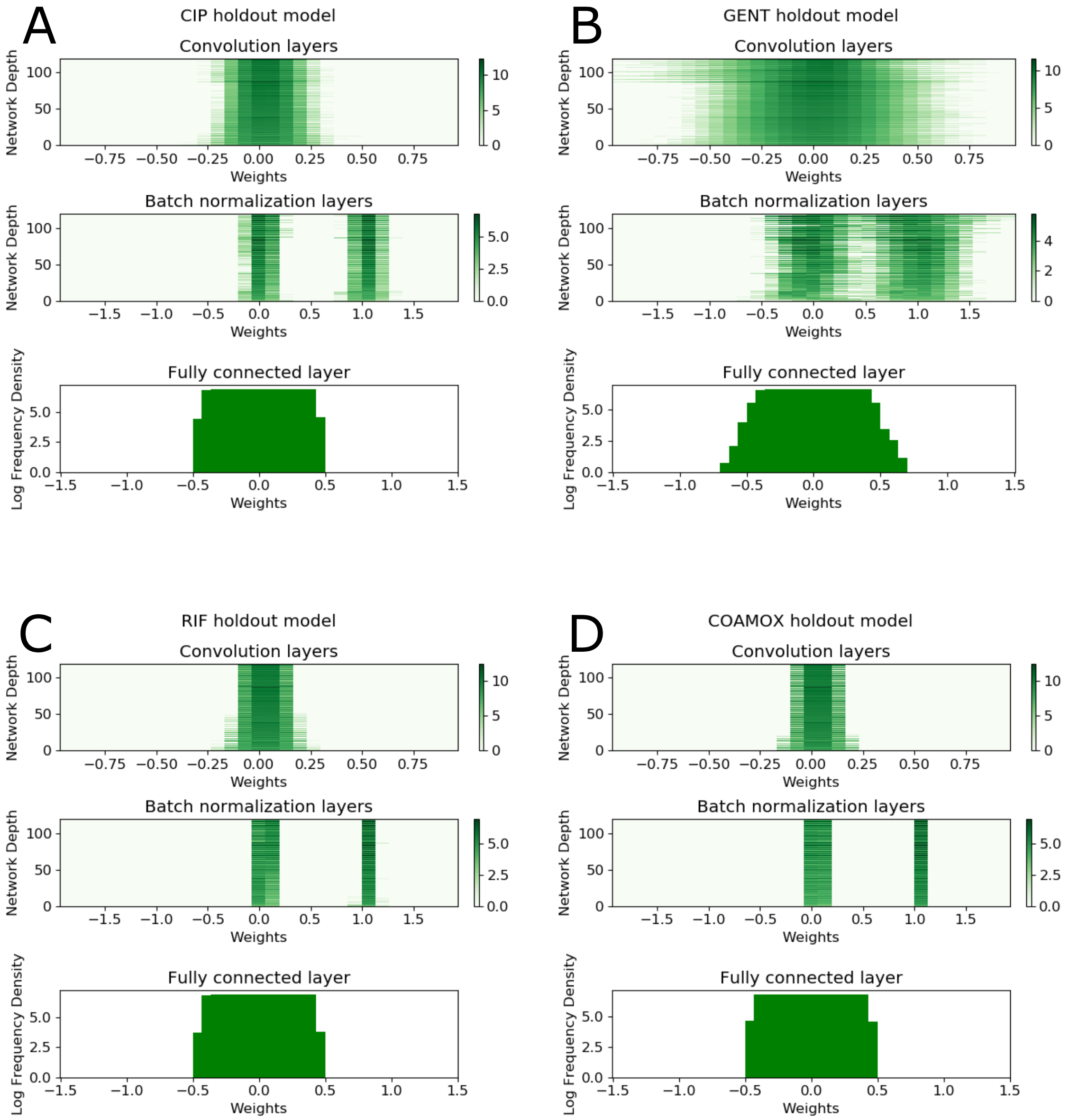


Figure 2.2.6: Example weight histograms of holdout classifiers

A. *Top* Stacked histogram of weights of convolution layers (bins on x-axis and bin log frequency density in green colour) as a function of network depth (layer index on the y-axis). *Middle* As above, but for batch normalisation layers. *Bottom* Single histogram of the fully connected layer weights at the end of the network.
B. As above, but for the gentamicin holdout model.
C. As above, but for the rifamicin holdout model.
D. As above, but for the co-amoxiclav holdout model.

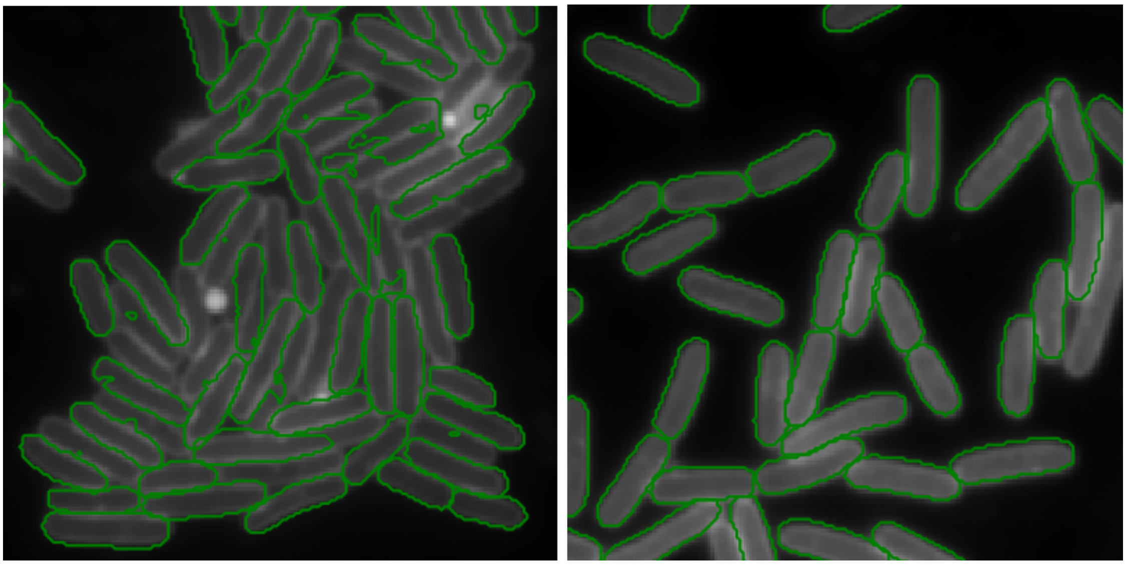


Figure 2.3.1: Segmentation in dense and sparse FoVs Single-channel membrane stained FoVs as presented to the segmenter. Green overlays indicate raw segmentation results. (*Left*) In dense FoVs, the non-maximum suppression removes correct predictions in dense clusters and fails to remove some spurious regions, leading to segmentation inaccuracies. (*Right*) In sparse FoVs, sufficient distance between cells exists for correct functioning of non-maximum suppression.

each ground-truth cell object, and correct candidate regions generated when ground-truth cell objects are closely packed. As a result, the filter removed some correct regions in neighbouring cells whilst keeping some spurious regions - creating false negatives, segmentation inaccuracies, and lowering total recall. As the IoU criterion got more stringent (and thus, the segmentation task becomes harder), this drop off became more apparent, as fewer correct detections survived. By contrast, this is not an issue in more sparse FoVs (2.3.1 *Right*), where sufficient distance between cells exists for the filter to function correctly.

To complement the above analysis, we also computed the Confidence-Recall curve corresponding to the Precision-Recall curve (Figure 2.3.2 *Bottom*). The confidence gives a numerical measure (on a scale 0-1, interpreted as a probability) of how confident the segmenter is in its own predictions. Here, we again saw the same pattern - the segmenter maintains high confidence in all of its predictions, right until the final recall value.

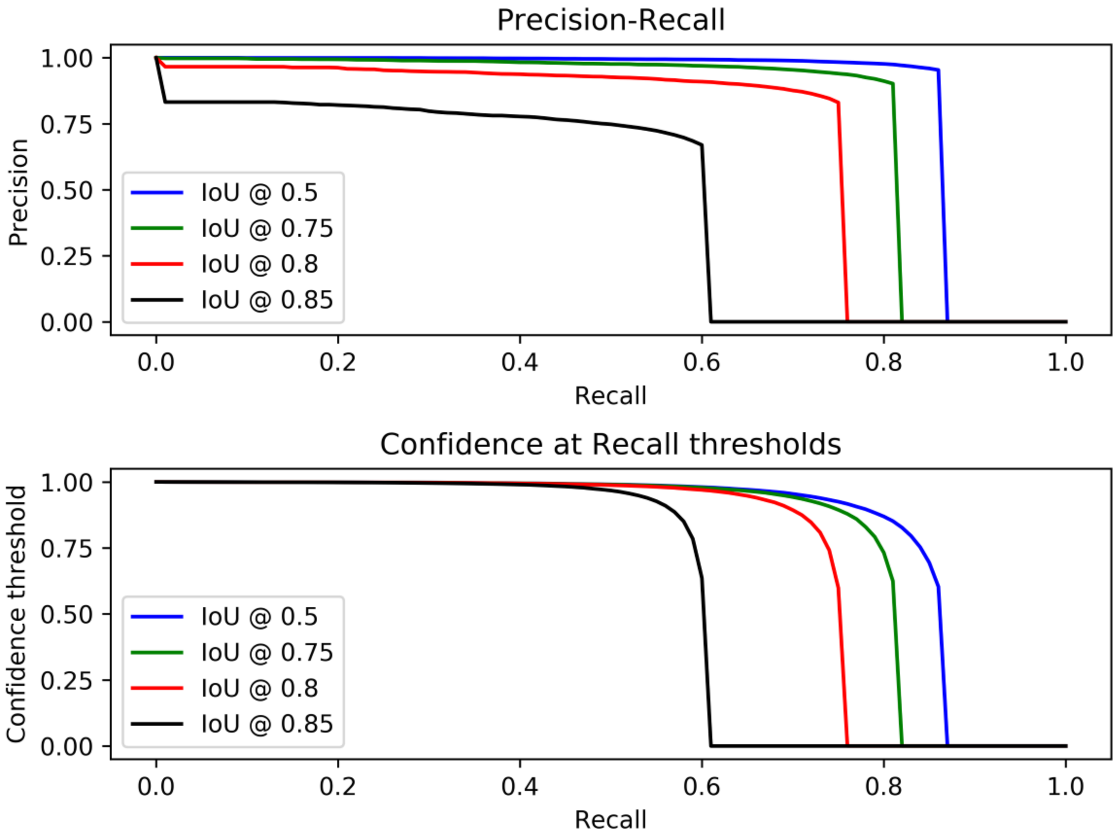


Figure 2.3.2: Precision-Recall and Confidence-Recall curves of the segmenter, evaluated on the holdout test experiment. *Top* Precision-Recall curve at various IoU thresholds, evaluated on the holdout experiment. *Bottom* Confidence-Recall curve at various IoU thresholds, evaluated on the holdout experiment.

2.3.2 Phenotype Classification

To test the classifier, we carried out the computational experiments we described. We ran the holdout experiment using 500 randomly selected cells per class and experiment for all antibiotics, except for co-amoxiclav, where we could only select 400. All antibiotic susceptible models correctly segmented untreated cells and classified them as resistant (Figure 2.3.3 A-D, red overlays). In antibiotic treated susceptible cells, the specific models trained to recognise those susceptible phenotypes correctly classified them as well (Figure 2.3.3 E-H, blue overlays). Considering the overlays, we could see the models occasionally make mistakes (blue overlays in untreated FoVs, red overlays in treated FoVs). One driver of this is cell-cell heterogeneity - in treated FoVs, not all cells might respond to the antibiotic on the same timescale, leading to their classification as resistant (ie. non-responding). Cells that are less metabolically active could be expected to react slower - thus had not yet converted to the susceptible phenotype by the end of the incubation time. Conversely, in untreated FoVs, cells might be classified as susceptible if they were captured at a particular cell-cycle stage that looks like the resistant phenotype - the models have no information about the treatment or metadata of the sample, and are making their classification decision only on the visual appearance of individual, segmented cells. The quality of the segmentation can also influence results - the classifier is provided only with the output of the segmenter, and no additional filtering step exists that guarantees that cells were segmented correctly. If incorrectly segmented cells are passed to the classifier, this could lead to anomalous decisions - although fortunately these instances are rare enough that they do not affect our results. Across the entire holdout experiment, we achieved single cell classification accuracies of above 80% - meaning over 80% of untreated cells were correctly classified in the resistant class, and over 80% of treated, susceptible cells were classified as susceptible. (Figure 2.3.3 I-L).

Beyond the holdout test, we also carried out the other experiments we described (Figure 2.3.4). Our best and most consistent results came from the 'ensemble' results, and they consistently scored several percentage points of accuracy above other setups (for example, ciprofloxacin ensemble achieved a per-cell accuracy of 93%, above the cross-validation 90% and holdout 89%). This is to be expected, and by design - in the ensemble experiment, the model in training has access to the broadest possible dataset; the aim of the setup was to test the upper bound of performance with the dataset at hand. In the 'cross-validation', we systematically withheld one of the experiments for testing, and class-balanced the datasets between experiments to ensure no experiment contributes more data than any other. Cross-validation accuracies were consistently lowered than the ensemble by average of 6% - this is to be expected, as these models had access to less training data. Despite this, they maintained consistently high performance, with an average of 81.5% accuracy across the antibiotics (vs. 87.5% for the ensemble). The relationship between the holdout test and the upper bound ensemble performance was more complex - in

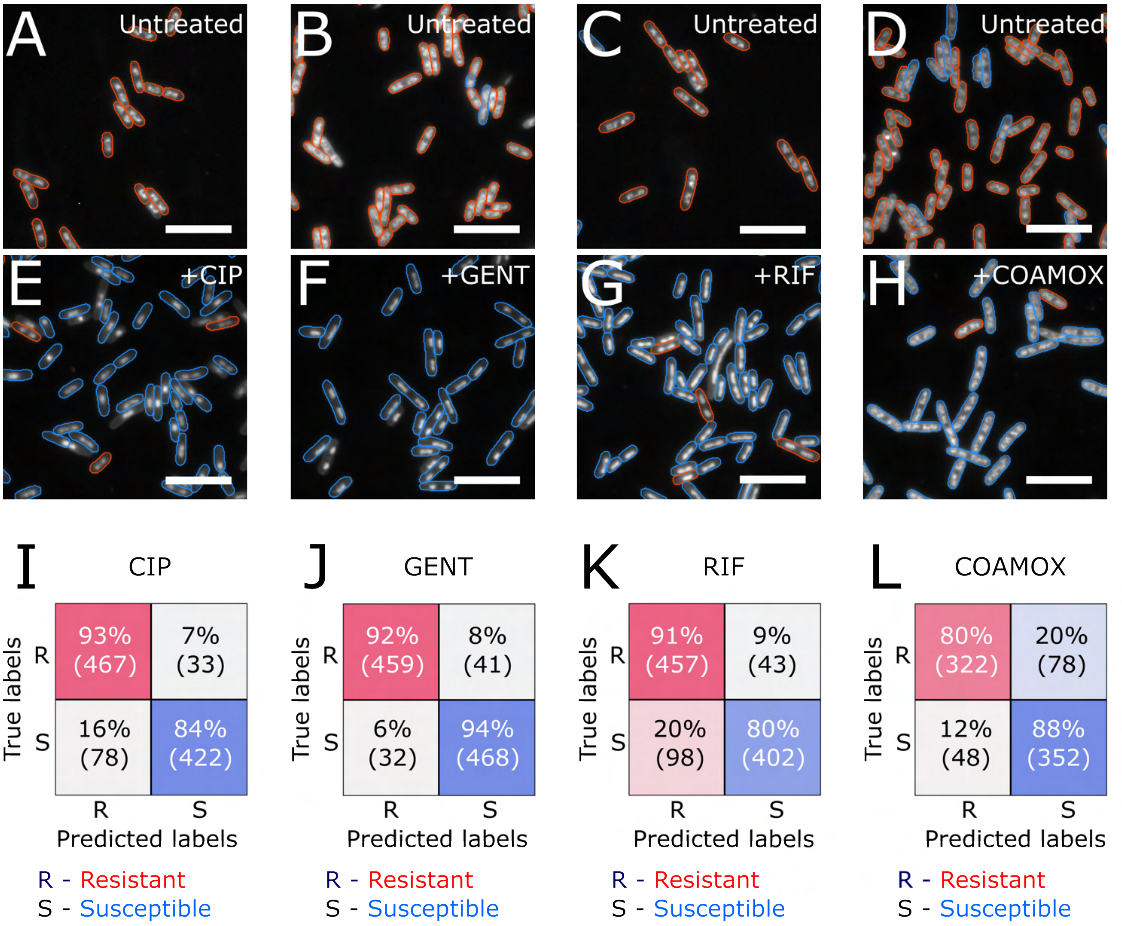


Figure 2.3.3: Binary classification of resistant and susceptible phenotypes in the holdout experiment.

A. Representative FoV of untreated cells from the holdout experiment, in grayscale, and an overlay showing phenotype detections. Resistant classifications shown in red, susceptible ones in blue. Ciprofloxacin resistant/susceptible classifier. Scale bar indicates 5µm.

B-D. As **A**, but with the gentamicin classifier (**B**), the rifampicin classifier (**C**) and the co-amixiclav classifier (**D**)

E. As **A**, but with ciprofloxacin-treated cells, showing the ciprofloxacin susceptible phenotype.

F. As **B**, but with gentamicin-treated cells, showing the gentamicin susceptible phenotype.

G. As **C**, but with rifampicin-treated cells, showing the rifampicin susceptible phenotype.

H. As **D**, but with co-amoxiclav treated cells, showing the co-amoxiclav susceptible phenotype.

I. Holdout test performance of the ciprofloxacin classifier, evaluated on class-balanced, randomly sampled 1000 cells from the holdout experiment. Percentage and absolute class counts shown in both the resistant (R) and susceptible (S) classes.

J. As **I**, but for the gentamicin classifier.

K. As **I**, but for the rifampicin classifier.

L. As **I**, but for the co-amoxiclav classifier, and with a dataset of 800 cells.

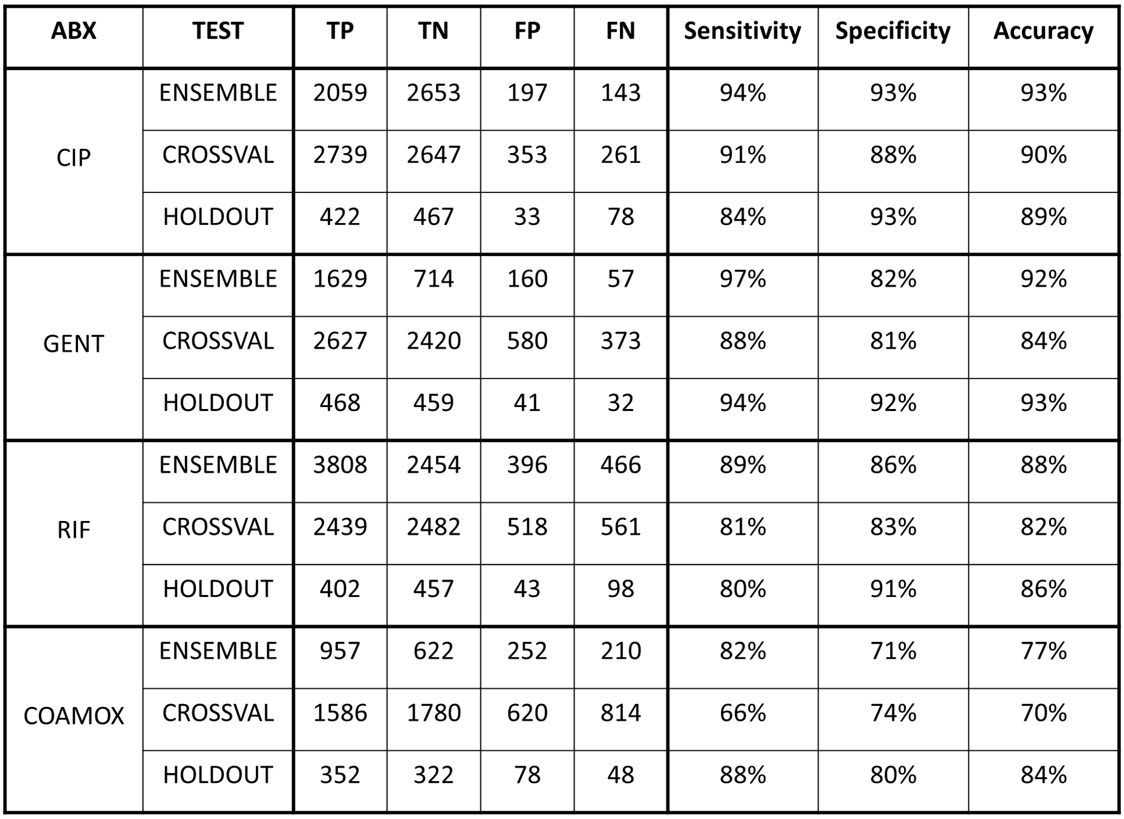
some cases the holdout outperformed the ensemble (co-amoxiclav 7% increase), although more commonly the holdout suffered a few percentage points (for example, ciprofloxacin 4% drop, or rifampicin 2% drop). This was largely due to experimental variation between experiments - the differences were on the same order of magnitude as the relative performance loss in the cross-validation setting. The sources of experimental variation can here can include reagents, microscope alignment, and minor clonal differences between separate colonies taken from the agar plate for processing - each experiment was done on a different day from a separate colony.

Comparing antibiotics against each others, we saw the antibiotics with susceptibility phenotypes that condense the nucleoid (ciprofloxacin and gentamicin) produced the most consistent classifier performance across all experiments (92% ensemble upper bound accuracy). This was a strong, consistent phenotype, and we anticipated that it would be easy to learn. The performance decreased as the phenotypes got more nuanced. For example, the nucleoid decompaction of rifampicin, in which the heterogeneous macrodomains of the nucleoid were lost, produced an ensemble accuracy of 88%. The most nuanced phenotype was co-amoxiclav, which was visually hard to distinguish from the untreated phenotype - but even here, our models achieved an ensemble accuracy of 77%.

Across the board, we concluded our models can reliably segment and classify different susceptiblity phenotypes.

2.4 Classifier Explainability

We found it encouraging to see that our models can indeed segment and classify different phenotypes correctly. However, deep ANN models such as the CNNs we are using are notoriously hard to interpret and explain. In order to further develop the technique, we needed to gain insight into the decision making process of the CNN. From the troubleshooting point of view, this was necessary to allow us to verify the models were not learning spurious features when classifying phenotypes - such as background pixels. From the technique development point of view, this was necessary to build end-user confidence in the decisions, and allow understanding of



ABX	TEST	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
CIP	ENSEMBLE	2059	2653	197	143	94%	93%	93%
	CROSSVAL	2739	2647	353	261	91%	88%	90%
	HOLDOUT	422	467	33	78	84%	93%	89%
GENT	ENSEMBLE	1629	714	160	57	97%	82%	92%
	CROSSVAL	2627	2420	580	373	88%	81%	84%
	HOLDOUT	468	459	41	32	94%	92%	93%
RIF	ENSEMBLE	3808	2454	396	466	89%	86%	88%
	CROSSVAL	2439	2482	518	561	81%	83%	82%
	HOLDOUT	402	457	43	98	80%	91%	86%
COAMOX	ENSEMBLE	957	622	252	210	82%	71%	77%
	CROSSVAL	1586	1780	620	814	66%	74%	70%
	HOLDOUT	352	322	78	48	88%	80%	84%

Figure 2.3.4: Full binary classification results of resistant and susceptible phenotypes across all computational experiments. Ensemble, cross-validation (cross-val) and holdout test results are presented for all 4 antibiotics (abx) used in this thesis. True positive (TP) is the number of treated cells correctly classified as susceptible. True negative (TN) is the number of untreated cells correctly classified in the resistant class (showing no response due to being untreated). False positive (FP) and false negative (FN) are the numbers of misclassified untreated cells and treated cells respectively.

contributing factors. Finally, from the research point of view, this was necessary to help us understand the phenotypes themselves, by highlighting features that can discriminate between them.

To understand CNN classification decisions, we used guided saliency mapping. In short (see *Experimental methods* section 2.5 for further details), we fed example inputs to the classifier, then calculated the gradient of the decision with respect to original input pixels. Overlaying the propagated gradient over the input image produced an attention heatmap, highlighting regions that contributed most to the classification decision. We carried out this procedure for randomly selected single cells of the susceptible phenotypes, with the correct antibiotic susceptible classifier, that were classified correctly (Figure 2.4.1).

Examining the heatmaps, we saw the patterns we expect. In ciprofloxacin treated cells (Figure 2.4.1 A), we saw from the heatmaps the classifier indeed focused strongly on a single point of nucleoid compaction; the model learned features we expected it to learn. The same could be said for the other nucleoid compactor - gentamicin (Figure 2.4.1 B). Here again we saw heatmaps focusing onto single points, but not as strongly as with ciprofloxacin - matching the differences between the phenotypes. In the nucleoid decompactor rifampicin (Figure 2.4.1 C), the heatmaps were instead diffuse over the entire nucleoid structure, once again matching our expectations. The results from co-amoxiclav (Figure 2.4.1 D) were particularly interesting. The classifier appeared to focus on particular, selected points towards the edges of the nucleoid and cell, but little attention is paid to the cell membrane. This was unexpected, since the mode of action of co-amoxiclav is peptidoglycan formation disruption (membrane). We could speculate there are perhaps secondary effects of co-amoxiclav that have an effect on the nucleoid. Co-amoxiclav slows peptidoglycan formation, leading to a buildup of precursor products - perhaps their accumulation had an effect on nucleoid structure. Unfortunately, we have no evidence to take these speculations further.

We just looked at examples of correctly classified cells. Further insight could be gained by looking at examples of cells that were misclassified (Figure 2.4.2) - where similar patterns emerged. In ciprofloxacin treated susceptible cells misclassified as resistant (Figure 2.4.2 A), we see the heatmaps point at two objects instead of one - the cells were misclassified because the nucleoids were not condensed enough. This could be due to cell-cell heterogeneity: not all cells in a culture will necessarily respond on the exact same timescale due to genomic and environmental factors. The same conclusion could be reached in the gentamicin case (Figure 2.4.2 B). It was hard to gain much insight from the rifampicin case (Figure2.4.2 C) - the nucleoid appeared just as decompacted as the correctly-classified examples, and the heatmaps did not produce much insight. With the co-amoxiclav examples (Figure2.4.2 D), again it was hard to tell - the models were looking at the entire nucleoid. Interestingly, it

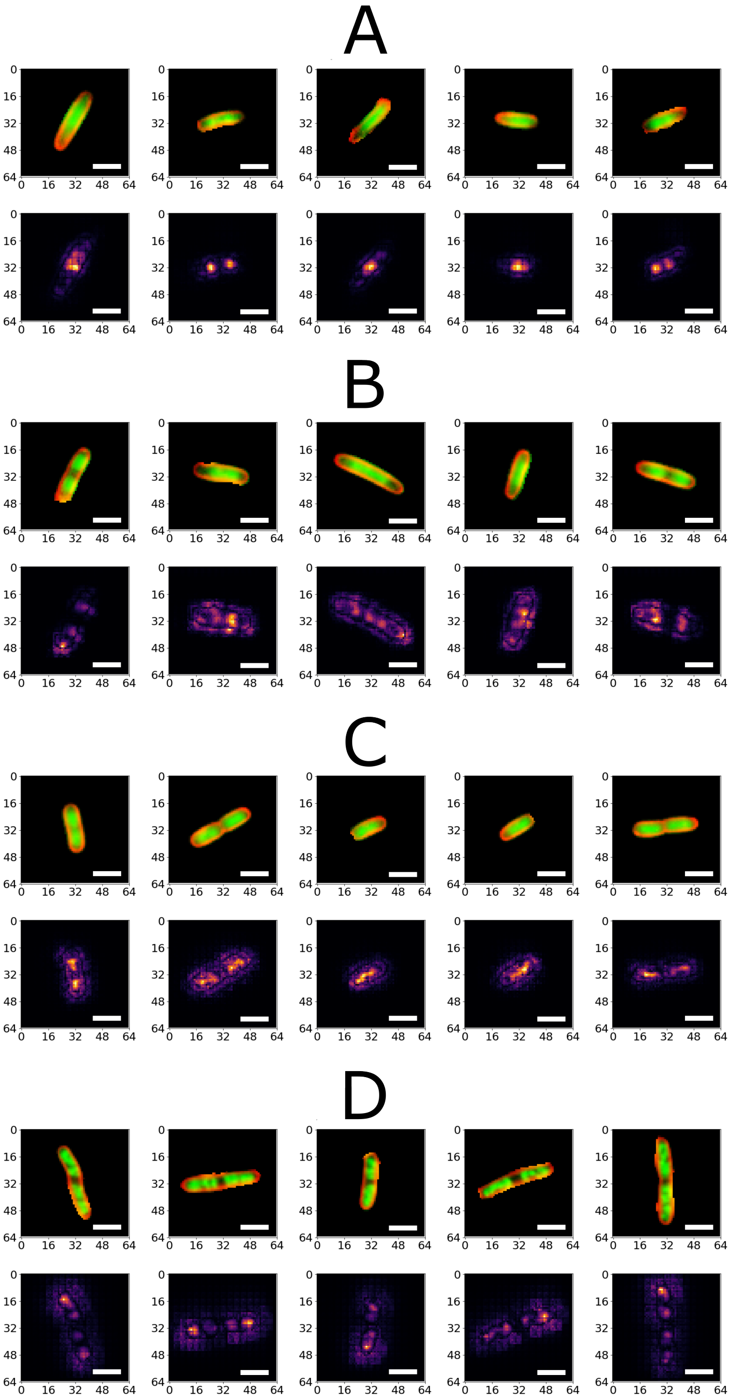


Figure 2.4.1: Saliency mapping of randomly selected, correctly classified susceptible cells..

A. (*Top*) Randomly selected ciprofloxacin susceptible cells, after pre-processing steps, as presented to the classifier. All these cells were correctly classified as susceptible (*Bottom*) Guided saliency maps corresponding to selected single-cell phenotypes, absolute value. Bright regions highlight areas that contributed most to the classification decision. Scale bar indicates 5µm. Pseudocolour scale ranges from 0 to 1 for the absolute value of the gradient, as shown.

B. As above, but for gentamicin.

C. As above, but for rifampicin.

D. As above, but for co-amoxiclav

appeared here that the misclassified cells were shorter than the correctly classified cells - although again we did not have sufficient evidence to pursue this claim.

Altogether, using saliency mapping allowed us to verify that the strong features we expected to be heavily discriminative (such as nucleoid compaction of ciprofloxacin and gentamicin), and less discriminative but still obvious (nucleoid decompaction of rifampicin) were indeed learnt as expected. Further, it also demonstrated to us that models are able to engineer their own complex features that are not obvious, even to the expert observer - such as in the case of co-amoxiclav. We learnt the models are not learning spurious features, and instead base their decisions on the phenotypes themselves. Together, this helped to increase our confidence in the numerical results we achieved, setting us up well for the clinical deployment of the assay.

2.5 Experimental Methods

Sample preparation

The strain of *E.coli* used in training and validation was MG1655. In all experiments, individual colonies of MG1655 were cultured overnight in 5ml of Lysogeny Broth (LB) at 37°C, then diluted 1:100 by volume in 5 ml of EZ Rich Defined Medium (RDM; Teknova). These subcultures were cultured at 37°C until reaching an OD of 0.2 in a shaking incubator. From this subculture, 1ml aliquots were either treated with an antibiotic of choice to produce cells showing the susceptible phenotype, or left untreated to produce the resistant phenotype. Figure 2.5.1 shows the treatment conditions used to generate the susceptible phenotypes. The cells were then fixed by addition of 2.5% formaldehyde and incubation for 30 minutes. The fixed cells were washed 3 times with Phosphate Buffered Saline (PBS) via centrifugation at 4500 RCF for 3 minutes, and then incubated with 100% ethanol for 10 minutes to achieve membrane permeabilisation. Following this, cells were resuspended in PBS and stained by addition of 10 µg/L of DAPI (GeneTex, catalogue number GTX16206) and 1 µg/L Nile Red (Fisher Scientific, catalogue number 10464311),

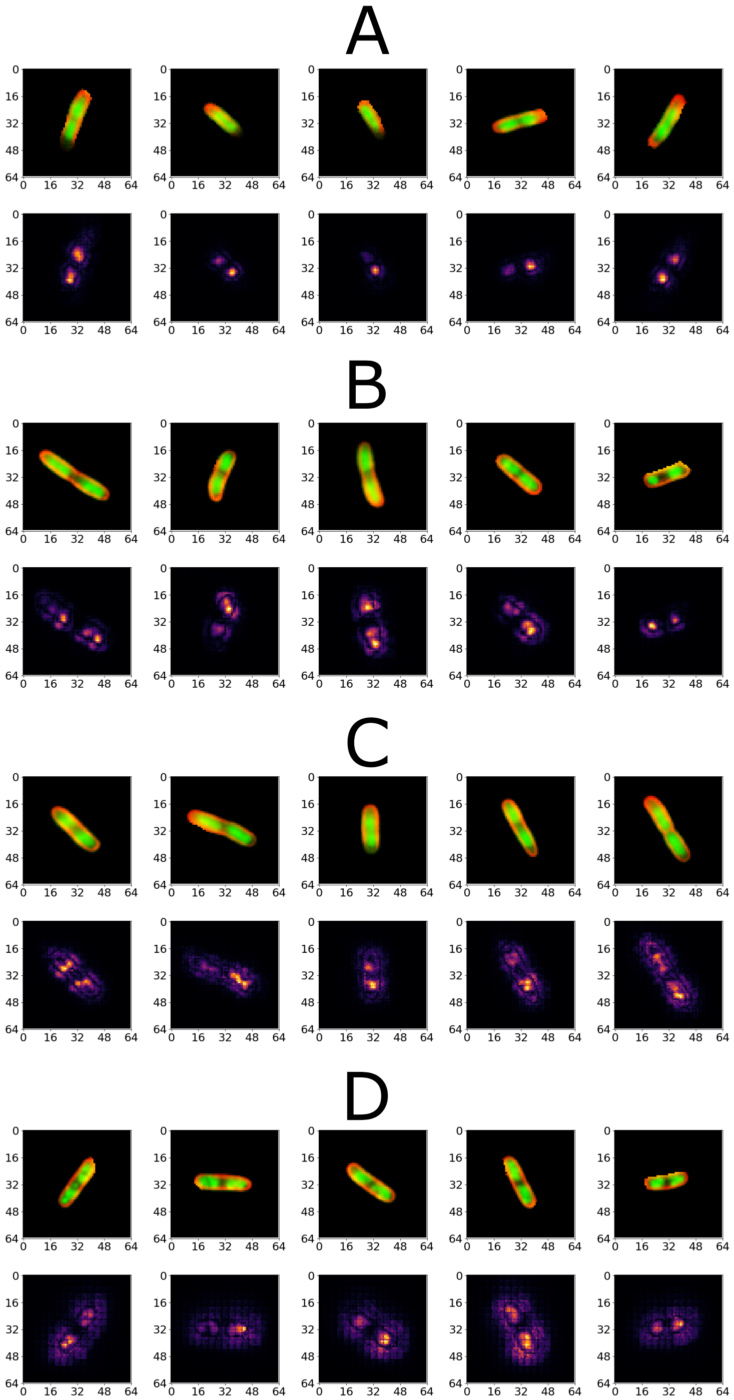


Figure 2.4.2: Saliency mapping of randomly selected, misclassified susceptible cells.

A. (*Top*) Randomly selected ciprofloxacin susceptible cells, after pre-processing steps, as presented to the classifier. All these cells were misclassified classified resistant (*Bottom*) Guided saliency maps corresponding to selected single-cell phenotypes, absolute value. Bright regions highlight areas that contributed most to the classification decision. Scale bar indicates 5µm. Pseudocolour scale ranges from o to 1 for the absolute value of the gradient, as shown.

B. As above, but for gentamicin.

C. As above, but for rifampicin.

D. As above, but for co-amoxiclav

Antibiotic	Code	EUCAST breakpoint (mg/L)	Treatment concentration (mg/L)	Treatment time (min)
Ciprofloxacin	CIP	0.5	10	30
Gentamicin	GENT	2	40	30
Rifampicin	RIF	-	100	30
Co-amoxiclav	COAMOX	8	160	60

Figure 2.5.1: Treatment conditions and abberviations used to generate susceptible phenotypes. Antibiotic column contains the standard name of the antibiotic. The code refers to the abbreviation used in places in this thesis. The EUCAST breakpoints are clinical-standard MIC thresholds used for classifying resistant and susceptible isolates. No breakpoint has been established for rifampicin in *E.coli*. The treatment concentration and time are the treatment conditions used.

and incubated at room temperature for 10 minutes. Cells were then washed twice with PBS and finally re-suspended in a small volume of 10-20 µL.

Such prepared samples were imaged on agarose pads. Briefly, these were prepared by melting 1% high-purity agarose (Bio-Rad, catalogue number 1613101) in half-concentration of PBS and depositing 100 µL of agarose on a microscopy cover slip. As the agarose was cooling, a second slip was placed on top, creating an agarose 'sandwich'. The second slide had been burned in a plasma cleaner at 500 °C for 60 minutes.

Imaging

Agarose mounted samples were imaged inverted (through the second cover slip) on a Nanoimager-S fluorescence microscope (Oxford Nanoimaging). Briefly, a blue (405 nm) and a green (532 nm) laser were combined using a dichroic mirror and coupled into a fibre optic cable. The output was focused into the back focal plane of a 100x oil immersion objective, and the fluorescence emission was collected by the objective and separated into two emission channels onto a sCMOS camera (Orca flash V4, Hamamatsu). To make best use of the camera dynamic range, the DAPI emission was imaged using 405 nm excitation and Nile Red was imaged using 532 nm excitation, and both signals were acquired separately and consecutively for each FoV. Laser powers were kept constant at 1.5 kW/m^2. For each signal and FoV, a stack of 30 frames was acquired at 30 ms exposure (33 Hz switching

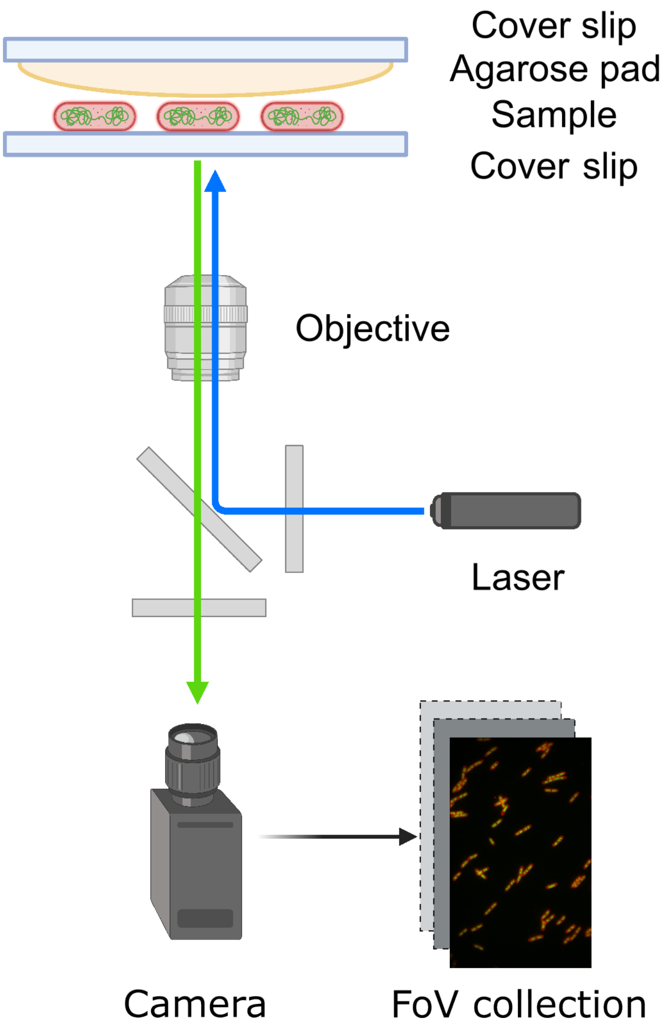


Figure 2.5.2: Microscopy of samples on agarose pads. Stained bacteria were mounted on an agarose pad sandwiched between two cover slips, and imaged inverted on a widefield epifluorescence microscope. Figure created with Biorender.

frequency). The multiple-acquisition and autofocus capabilities of the Nanoimager were used to reduce human bias. Figure 2.5.2 illustrates the microscopy setup and the agarose mounted samples.

Deep Learning: segmentation

The training data for the Mask-RCNN segmenter was generated using only the Nile Red (membrane) signal of every FoV acquired. FoVs with both treated and untreated cells were used. The 30 frames in each FoV were averaged, and expanded to RGB space by replicating the image in each colour channel. During training, the images were augmented on-the-fly by random cropping to a size of 256 by

256 pixels, followed by a random sequence of transformations, which included horizontal and vertical flips and translations, rotations, cutout [163] and Gaussian blurring. The corresponding ground-truth annotations were similarly transformed - these were generated by manual data annotation followed by cycles of model boot-strapping. The internal parameters of Mask-RCNN were optimised to perform well on a 2-class segmentation task in dense FoVs, consisting of modifications to the RPN and Non-Maximum Suppression parameters, from a standard implementation [164]. Hyperparameter tuning of the learning rate and batch size was achieved by grid-search, maximising segmentation accuracy on validation data.

In the end, the best performing model was trained using a learning rate of 0.003 and batch size of 2 with the Adam optimizer, using a momentum parameter of 0.9 and weight decay per epoch of 0.001, applied to all weights. The model was trained in 4 consecutive steps, starting from initial weights trained on the MS COCO dataset that are publicly available [165]. In the first step, the model 'top', consisting of the RPN and second stage classifier and mask regressors were trained for 50 epochs. In the second step, the entire network was trained together, including the feature-encoding FPN, for 50 epochs. In the 3rd step, the 'top' was fine-tuned for 50 epochs using 10% of the initial learning rate. Finally, the entire network was fine-tuned again for 50 epochs, using the reduced learning rate.

Deep Learning: classification

Training data for the classifier was generated by using both channels of every FoV acquired (thus both membrane and nucleoid signals). To train the 'resistant' class, untreated cells were used. To train the 'susceptible' class, treated cells were used. Both channels were averaged separately, and the DAPI channel was registered automatically to the Nile Red signal using cross-correlation to correct for any drift. Individual cells were extracted from assembled images using the ground-truth mask annotations, and all cells were resized to a common size of 64 by 64 pixels by zero-padding in either dimension if below the target size, or interpolating down to the target size if above. Histogram equalization was applied to every cell independently for each channel, within the segmentation mask only. During training, cells were augmented on-the-fly using a random sequence of affine transformations and colour intensity augmentations, which included unsharp masking, random brightness modifiers in HSB colour space, Gaussian noise addition, random Gaussian blurring and intentional channel misalignment (acheived by channel-wise random translation). Classifier learning rate, batch size and optimiser of choice were optimised by grid search against validation accuracy. The classifier was implemented in Keras [135] version 2.2.4, and trained in two steps with cross-entropy loss and ReLU activations in all layers except the final one - in the first step, training occurred at the full learning rate for 50 epochs, and in the second, at 10% of the learning rate for further 50 epochs. To produce the final classification from a vector of per-class scores, the greatest per-class score was chosen, which corresponds to the output seen as most likely by the classifier.

Deep Learning: saliency mapping

Attention heatmaps over example classification inputs were computed by calculating the gradient of the output category with respect to the input. Positive (>0) gradients for positive activations were propagated (guided backpropagation) [166]; to produce the final output we visualised the absolute value of the gradient.

Segmentation and classification metrics

Classification performance of the binary resistant-susceptible classifiers was analysed as a confusion matrix, which displays the True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) counts for each class. From these, secondary metrics such as sensitivity, specificity (also known as recall) and accuracy can be calculated as follows:

$$Sensitivity = \frac{TP}{TP + TN} \tag{2.5.1}$$

$$Specificity = \frac{TN}{TN + FP} \tag{2.5.2}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.5.3}$$

To quantify the quality of the segmentation, we used Precision-Recall curves using bounding boxes of detections and corresponding annotations, at various IoU thresholds.

Precision is a measure of how well relevant objects were retrieved as a function of all objects returned:

$$Precision = \frac{TP}{TP + FP} \tag{2.5.4}$$

Unlike in classification, the segmentation tasks also has to consider *how well* the segmentation was done (the output is no longer a categorical variable). To put a bound on the amount of overlap that counts as a successful detection (to give a True Positive, or else a False Negative), we considered the Intersection over Union (IoU) between candidate detection regions (D) and the corresponding ground truth annotation (GT):

$$IoU(D, GT) = \frac{|D \cap GT|}{|D \cup GT|} \tag{2.5.5}$$

As the main metric for the segmenter, we calculated precision as a function of recall (the Precision-Recall curve), at different IoU thresholds - this allows us to see how well objects are segmented as the segmenter reaches the bounds of it's performance, whilst IoU scaling allows us to do the same as the criterion for counting a successful detection becomes more stringent.

2.6 Discussion

2.6.1 Segmentation and classification of phenotypes

Our first achievement in this chapter was segmentation of cells in complex FoVs. To achieve this, we adapted a general purpose CNN segmenter Mask-RCNN and trained it for a 2-class segmentation task (cell/background). To train it, we used only the Nile Red signal (membrane stain) - we did so to simplify the task and prevent the segmenter from becoming confused by the nucleoid signal. We chose a generic segmenting architecture - this shows the approach is general and versatile, although it also revealed certain limitations of segmentation performance, where the non-maximum suppression filter limited the upper-bound segmentation performance. In further iterations of this work, to avoid this problem, we would consider switching away from a generic architecture towards one specialised for cell segmentation - a promising example introduced into the literature since the project began is the UNet derivative Cellpose [167].

After the segmentation, we demonstrated the second part of the pipeline - phenotype classification. Here, we demonstrated consistently high single-cell classification performance across a range of antibiotics from different familieis with different modes of action (quinolone ciprofloxacin targetting DNA gyrases, aminoglycoside gentamicin targeting translation, ansamycin rifampicin targeting transcription, and β-lactam co-amoxiclav targetting peptidoglycan formation), and different visual phenotypes. There appeared to be a correlation between the apparant visual strength of the phenotypes and performance - ciprofloxacin and gentamicin consistently outperformed the other antibiotics. Across all antibiotics, we observed a high degree of robustness to experimental variation - limiting access to data in the ‘cross-validation’ setup resulted in only 6% loss relative to upper bound performance. The unseen holdout tests were also consistent, on the same order of magnitude as upper bound performance. We performed 6 repeat wetlab experiments which we used for training and validation. We arrived at this count empirically, after considering the variation within the experiments; we stopped at the point where we performed enough experiments to encompass the natural variation of the system and provide sufficient evidence to support our claims. Naturally, while 6 experiments proved enough, results could be further improved with access to more training data, and more experiments, and further verified with access to more antibiotics. We selected 4 representative antibiotics to test with our platform - nevertheless, similar phenotypes have been experimentally characterised for a large amount of antibiotics and species [115–117] - we can reasonably expect our approach to scale beyond the antibiotics we chose.

Using saliency mapping, we verified the models indeed engineer their own complex features to achieve the classification task. Gradient based saliency has the advantage of simplicity and ease of interpretation, but using it with ReLU activations like we do leads to a saturation problem - in the forward pass negative activations are capped at 0, and thus cannot be propagated back. Thus, such a heatmap is at best an approximation of *some* of the total gradient. To address that, we used guided-backpropagation, which addresses the ReLU nonlinearity by propagating positive gradients for positive activations only (guided-backpropagation), which we judged to be the best compromise between representing the actual gradient, and producing useful explanations. However, there are some concerns about this approach - most notably, that it is most sensitive to lower (closer to the input) layer weights. As a result, the explanation is skewed towards input features - putting too much trust in it can lead to confirmation bias. At the time of writing, no definite conclusion has been reached about the exact impact of such factors on this and other methods - as a result, whilst we were encouraged by our results, we have to limit our conclusions.

In further iterations of this work, it will be important to go beyond guided-backpropagation and consider a collection of methods that, taken together, may provide further insight. Good candidates for this are perturbative methods, since these can potentially provide pixel-accurate explanations whilst remaining model-agnostic (and can thus avoid the ReLU non-linearity backpropagation issues that we discussed). Another direction could be to sacrifice pixel-level explainability and consider co-training simpler models (such as regression models) on broader features soley to provide explanations to more complex models, that are more suitable for the end user.

In our 2-step pipeline, we chose to work with CNN architectures in both steps. This has multiple advantages - like we discussed theoretically in the *Introduction* and observed experimentally in this chapter, both the segmenter and classifier can engineer their own features, making the pipeline reusable across antibiotics with minimal human oversight. To complement this study, it would however be useful to compare our results against the classical approach - manually engineering features (such as explicit morphology statistics or fluorescent intensity values) and using a classical ML algorithm instead of a CNN. This is also the closest existing approach in literature [119], although pursued for *S.aureus* rather than *E.coli*. Due to relative lack of novelty and time constraints of the work, we chose not to pursue this approach, and focused ourselves on other tasks.

The curious reader will at this point notice that our 2-step pipeline could in-fact be reduced further to a single-step pipeline - we could train the Mask-RCNN to both segment and classify different phenotypes at the same time. After some preliminary experiments, and our experience with similar projects in other domains, we chose not to pursue this approach further, and instead adopted the 2-step pipeline. Notably, this allowed us to fully customise both parts of the pipeline - the segmenter and classifier could be fully tuned independently, and fed different parts of the total data that we deemed most useful for their task (membrane stain for segmentation, both stains for classification).

2.6.2 Experimental design

For our experimental design, we actively picked the best conditions where a phenotype of susceptibility could be seen. Further work is necessary to establish if this phenotyping approach can scale beyond logarithmic phase, and mixed-phase samples. We also expect culture conditions to play a major role in the displayed phenotype - we cultured cells in rich media in both the overnight and sub-culture stages (both LB and RDM are similarly rich, however RDM is less auto-fluorescent, which works in our favour during phenotype imaging). Cells grown in poor media will have lower metabolic rates, and a different untreated phenotype (for example, cell length and chromosome count increases with medium richness). This has obvious implications for the appearance of the phenotype, and also for the conditions needed to generate it. For example, we would expect cells grown in a poor medium to take longer to respond to the antibiotic - applying the treatment optimised for rich medium conditions would then show a no response, creating a false negative result. This problem also propagates to the decision pipeline. The classifier makes its decision *solely* on the visual appearance of the phenotype after a fixed treatment time - it has no access to experimental metadata that it could use to understand the phenotype in context. To resolve this problem, two approaches are possible. We could simply expand the training data to include cells cultured in different conditions, allow the classifier to learn the relationship, and then verify we included enough training data to handle real-world variation (much like we did with the computational experiments so far presented). A better approach would be to expand the classifier itself to handle multi-modal input - the phenotype and its metadata. At the moment, this is beyond our scope, but a promising and needed future direction for the project.

The other steps of our protocol, such as fixation and staining were experimentally optimised for *E.coli*, and may not work for other species. This will be particularly important - in the diagnostic capacity, we would have no *a priori* knowledge of the pathogen. A universal fixation and permeabilisation protocol that could handle any sample is unlikely to exist; preliminary work in the group on other aspects of this project beyond this thesis has already revealed major difficulties with the permeabilisation of certain Gram-positive species, for example. At the very least, a battery of treatments will be necessary to handle different species.

A potential solution to many of the problems we identified could be to abandon fixed-cell imaging entirely, and image live cells as they are responding to antimicrobial attack. This solves the problem of *a priori* optimisation of treatment conditions, as cells could be exposed directly to therapeutically-relevant concentrations of antibiotics, and their effects quantified with time. Further, this also eliminates the need for optimised fixation and permeabilisation, since these steps are no longer needed. The only step that remains is optimised and universal staining of live cells - naturally, this presents its own problems. We present some advances in this direction in later chapters of this thesis, and reserve further discussion for that time.

We used agarose pad immobilisation for our samples, which relied on mechanical pressure rather than specific surface functionalisation for immobilisation - and thus proved highly efficient and species-agnostic. However, this method significantly reduced our experimental throughput. Agarose-mounted samples could only be mounted and imaged one-at-a-time, and are incompatible with existing high-throughput microscopy platforms that utilise multi-well plates. In our preliminary work, we investigated a range of alternative approaches, including chemical functionalisation of a cover slip surface with chemicals that bind to the outer surface of bacteria (such as poly-l-lysine, or chitosan). This would increase throughput, and transition better to existing high-throughput microscopy platforms. Unfortunately, we could not achieve consistent, reproducible and high-quality imaging in this way. The surface treatment was uneven, which resulted in bacteria being immobilised at a range of angles to the surface normal (whereas with agarose all cells were immobilised horizontally due to mechanical pressure), which made the imaged phenotype highly inconsistent. Further, the treatment was also rough, which scattered light at the interface, leading to an appreciable loss of resolution. Most critically, we observed the efficiency of immobilisation varied with different genotypes of *E.coli*. We speculated this was due to different outer surface receptors and their interactions with the surface-treatment - nevertheless, since our aim was to develop an AST where no *a priori* knowledge of the pathogen is available, we concluded this could not be a suitable approach to take forward.

2.6.3 Future directions in optimising phenotypes

We derived our susceptibility phenotypes from the structure of the nucleoid and membrane, and used the membrane further to allow accurate cell segmentation.

These are useful, but not the only choices for phenotyping. Another prominent target is the ribosomes, which are targeted more directly and by more antibiotics than the nucleoid. Membrane staining allowed us to segment cells, but this could have also been achieved by brightfield, darkfield or phase-contrast imaging - these cannot provide the fine-structure of the membrane that we achieved with Nile Red, but all are label-free and more universal between species than any fluorescent staining. Assessing our guided-backpropagation results, it appeared the membrane structure did not contribute as much to the phenotype as we had anticipated, even for the membrane-acting co-amoxiclav, and was perhaps an unnecessary choice.

The changes in the structures we chose provides a direct phenotypic output of bacterial susceptibility. Widefield fluorescence microscopy however also opens up the possibility of interrogating genotypic targets via Fluorescence In Situ Hybridization (FISH) probes that target genes of interest - such as known resistance genes, down to single-molecule levels. To circumvent the difficulties with genotypic assays that we discussed (such as lack of straightforward concordance with phenotypic outputs), the same approach could also be used to target transcriptomic targets, measuring expression of resistance genes during antimicrobial attack, not just their presence. Going up one step in the gene expression ladder, proteomic targets could also be addressed with immunofluorescence. Nevertheless, widefield fluorescence microscopy does not solve some of the fundamental problems, such as the need to identify a limited panel of such targets *a priori* - in contrast, the phenotypic changes we see are specific to the antibiotics used, but the structures themselves are agnostic to the antibiotic and strain. Nevertheless, all of the above are promising further directions in the project.

2.6.4 Generalisability of findings

Finally, all the results we developed and presented in this chapter came from the same strain of susceptible *E.coli* - the lab reference strain. The aim of the project is to develop a *clinical* AST. Thus, our models need to be able to generalise to different strains of *E.coli* and especially to clinical isolates - which may carry any and all resistance mutations we discussed, and further genomic adaptations specific to their infection site. These may complicate the phenotypes. In order for the models to generalise, the phenotypic response itself must be homogeneous between different isolates, and a function only of the treatment conditions and isolate MIC - a major key assumption taken so far. Further, so far we have only used untreated cells (and thus the untreated phenotype) as proxies for the resistant phenotype - another key assumption which we have not yet provided any evidence for.

We chose to work with 4 different antibiotics, and observed, unsurprisingly, that those which produce the most visually striking phenotypes (such as ciprofloxacin) produce more consistent results than the visually weaker phenotypes (such as co-amoxiclav). Speculatively, it seems our model performance decreased the further away we went from the translation and transcription targets. Further experiments will be required to assess the performance loss on other classes of antibiotics that do not target translation or transcription, such as trimethoprim (folic acid synthesis), or colistin (direct membrane damage).

True validation of the technique can thus only come from deployment of our trained models on a range of clinical isolates of *E.coli*, with varying degrees of susceptibility to the antibiotics we used. Alas, this is exactly what we shall discuss in the next chapter.

2.7 Conclusions

In this chapter we have introduced the 2-stage DASP phenotyping platform, and examined the phenotypes of *E.coli* MG1655 to 4 different classes of antibiotics with different modes of action. We trained both parts of the platform on the susceptible, lab-standard MG1655 strain, and validated the accuracy in a range of computational experiments - where we learnt our models were robust against experimental variation and human bias. Examining the decisions of the classifier, we concluded the classifier indeed engineers its own complex features to achieve the classification objective, as we had intended.

Ὁ βίος βραχύς, ἡ δὲ τέχνη μακρή, ὁ δὲ καιρὸς ὀξύς,
ἡ δὲ πεῖρα σφαλερή, ἡ δὲ κρίσις χαλεπή

Life is short, and Art long; the crisis fleeting; experience perilous, and decision difficult.

— Hippocrates of Kos, *Aphorisms*

3

DASP Deployment on Clinical Isolates

Contents

3.1 Phenotype Heterogeneity in Clinical Isolates 77
  3.1.1 Ciprofloxacin-susceptible phenotypes 79
  3.1.2 Ciprofloxacin-resistant phenotypes 81
3.2 Segmentation and Classification of Clinical Isolates . . 82
  3.2.1 Clinical isolates at a fixed treatment point 83
  3.2.2 Mixed samples and the Limit of Detection 92
  3.2.3 Antibiotic titration allows MIC quantification 97
3.3 Experimental Methods 101
3.4 Discussion 105
  3.4.1 Deployment of models on clinical isolates 105
  3.4.2 Sources of errors in the classifier 107
  3.4.3 Generalisability of findings 109
  3.4.4 Future directions 110
3.5 Conclusion 112

3.1 Phenotype Heterogeneity in Clinical Isolates

In the previous chapter, we described the development and validation of DASP on the lab-reference, susceptible MG1655 strain of *E.coli.* We concluded that true validation could only come from deploying our models trained on the lab-reference strain on clinical isolates of *E.coli.* To this end, we obtained access to a range of clinical isolates routinely collected and stored by the John Radcliffe Hospital

Microbiology laboratory for diagnostic purposes. Upon re-use for this work, the isolates were pseudoanonymised using codes EC1-6, and these codes will be in use in this thesis. We chose to continue working with ciprofloxacin as the model antibiotic - both due to its clinical relevance in the treatment of *E.coli*-caused infections, and due to its easily recognisable phenotype of susceptibility. We experimentally measured the MICs of a range of isolates (measurement performed by Dr. Conor Feehily as part of wider project family), and all isolates had already undergone whole-genome sequencing as part of a previous study [168], which assigned resistance genotypes to them (Figure 3.3.1, see *Experimental methods* section 3.3). We then cultured, treated, stained, processed and imaged the samples of clinical isolates using the same methods (Figures 2.1.2, 2.5.2) we used for the MG1655 samples used to train our models. From each culture, samples were either left untreated, or treated with ciprofloxacin at the same concentration as MG1655 samples (Figure 2.5.1) - producing collections of micrographs of cells from each. We identified two isolates that were highly resistant to ciprofloxacin (EC5 and EC6) - resistant to the concentration we used to develop the susceptible phenotypes in the susceptible lab strain. We further identified 4 isolates with varying degrees of susceptibility - 3 with high susceptiblity (EC1-3), and and one with intermediate susceptibility (EC4), where the MIC value approached the treatment concentration. Together, this allowed us to examine the relationship between the phenotype, treatment conditions, MIC and isolate variation, addressing some of the fundamental assumptions we took in the previous chapter.

All of our clinical isolates were clinical isolates of *E.coli*, which made them highly compatible with the experimental conditions and setup we optimised for our *E.coli* training strain MG1655. We did not observe statistically significant differences in the quality of antibiotic treatment, staining, fixation or microscopy. Further, since our desired endpoint was diagnostic (where no prior knowledge of the isolate behaviour could be assumed at the point of deployment), this solidified our choice not to alter any of the conditions - processing isolates identically to the training strain.

3.1.1 Ciprofloxacin-susceptible phenotypes

Looking at the susceptible isolates allowed us to validate our first fundamental assumption from the previous chapter - the homogeneity of the susceptible phenotype response across different genotypes of *E.coli.* If the response was not homogeneous, our models would not have been able to generalise as a clinical AST.

Across all isolates, the untreated phenotype broadly resembled that which we had seen in the training strain: two distinct copies of the chromosomes could be seen, exhibiting some heterogeneity of organisation and taking up most of the cell volume. (Figure 3.1.1 A-D). In limited cases, we observed a degree of nucleoid compaction even in untreated cells (Figure 3.1.1 C, rightmost cell), which we attributed to natural cell cycle processes and variation, or cellular stress factors beyond our experimental control. Nevertheless, this could introduce a source false-positive errors, as the classifier made its decisions solely on the visual appearance of the cell.

To generate the susceptible phenotypes, we treated the isolates with the same concentration of ciprofloxacin that we used to develop the susceptible phenotype in the training strain - a concentration of above the MICs of all isolates (treatment concentration of 10mg/L for 30 minutes, MICs of 0.008 mg/L, 0.03mg/L, 0.5mg/L and 8mg/L for EC1-4 respectively). In EC1-3, the highly susceptible isolates, we saw a phenotype of susceptibility that closely matched that of the training strain - a nucleoid condensed towards either the centre, or cell pole, validating our fundamental assumption. In contrast, in the intermediate susceptibility EC4 (Figure 3.1.2 H), we saw a mixed response. After the 30 minutes treatment, only a fraction of the cells had responded (leftmost 2 cells in panel), while others still showed the untreated phenotype (rightmost cell in panel). This hinted to us that the phenotypic response depends not just on treatment conditions, but the MIC relative to the treatment conditions - isolates that are more susceptible respond stronger and faster. Armed with this insight, we turned our attention to resistant isolates to investigate if the MIC dependency trend continues.

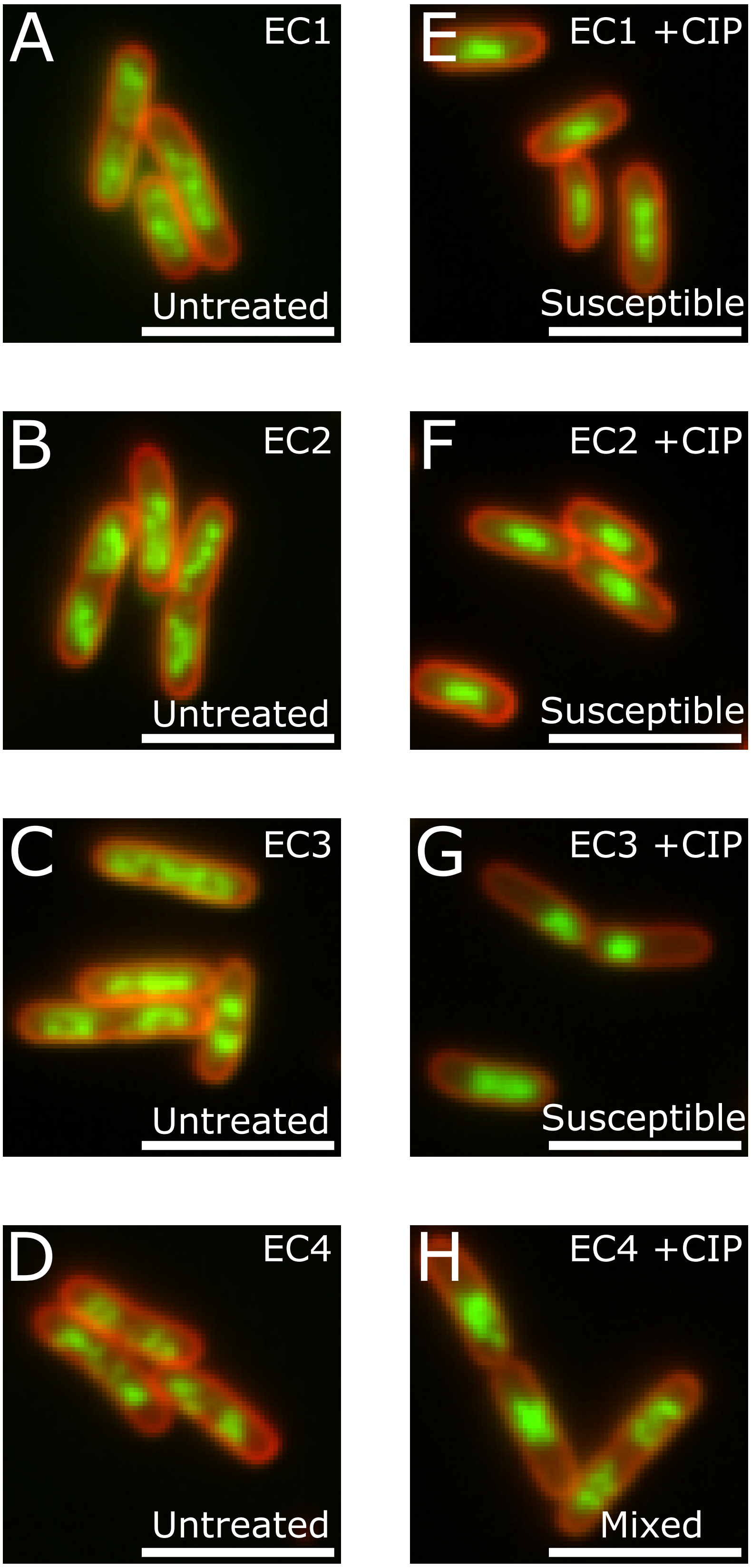


Figure 3.1.1: Untreated and susceptible phenotypes in clinical isolates
A-D. The untreated phenotypes of EC1-EC4. Scale bar 5 µm.
E-H. The ciprofloxacin susceptible phenotypes of EC1-EC4. Cells treated with 10mg/L of ciprofloxacin - over the MIC value.

3.1.2 Ciprofloxacin-resistant phenotypes

Access to resistant isolates allowed us to revisit our second fundamental assumption - the similarity of the untreated, and treated resistant phenotype in a resistant strain. Further, we could also explore the heterogeneity of those phenotypes in clinical isolates, going beyond the single training strain we used in the previous chapter.

In both EC5 and EC6, the untreated phenotype again broadly resembled that of the training strain (Figure 3.1.2 A-B). The major difference appeared to be the separation of the chromosomes - in EC5, the chromosomes appear pushed towards the poles of the cells, in EC6 they were found together quite close to the cell middle, while the training strain (Figure 2.2.1 A) appeared to be the intermediate case between those. The reasons for that were unclear, we could speculate the genomic differences between the strains could manifest themselves in subtle shifts in chromosome organisation.

To generate the resistant phenotypes of EC5 and EC6 (Figure 3.1.2 C-D), we treated them with the same concentration of ciprofloxacin we used to develop the susceptible phenotypes, which was significantly below the MICs of these isolates (treatment concentration of 10 mg/L against EC5 and EC6 MICs of 72 mg/L and 108 mg/L, for 30 minutes as before). This concentration was not enough to inhibit cell growth over any timescale, producing the resistant phenotype. Comparing the untreated and resistant phenotypes side-by-side (Figure 3.1.2 A,C for EC5, B,D for EC6), we saw little qualitative difference between them. In both cases, we could still distinguish individual chromosomes with their heterogeneity, and their relative separation appeared unchanged. We concluded that, qualitatively, and in the case of ciprofloxacin and *E.coli*, our fundamental assumption held - the untreated phenotype was a close proxy for the resistant phenotype. Further, the MIC dependency of the phenotypic response continued, building the necessary foundation for a clinical assay.

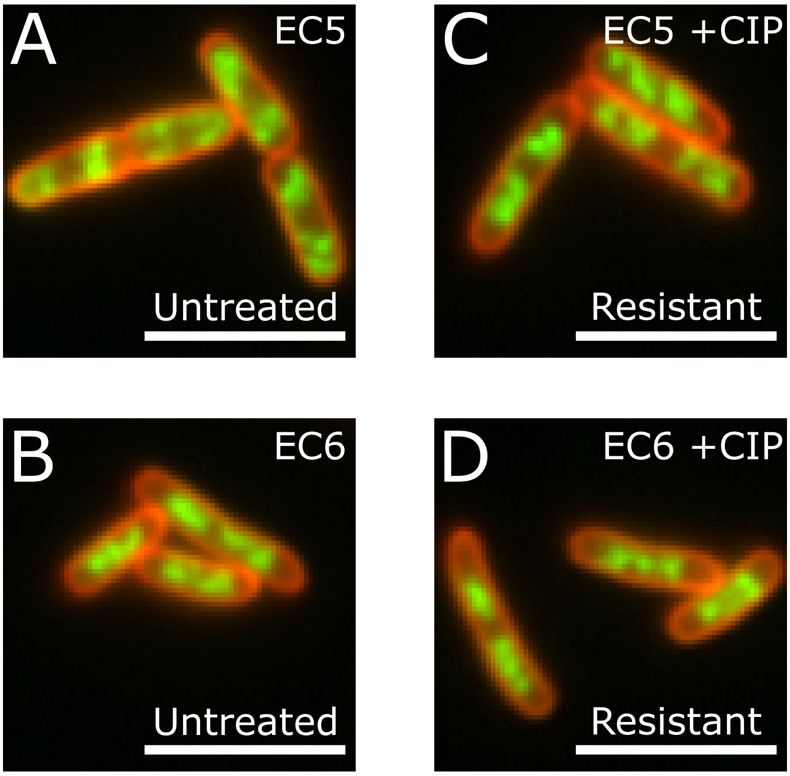


Figure 3.1.2: Untreated and resistant phenotypes in clinical isolates.
A-B. The untreated phenotypes of EC5 and EC6, two highly ciprofloxacin resistant *E.coli* (EC5 MIC = 72 mg/L, EC6 MIC = 108 mg/L). Scale bar 5 µm.
C-D. The ciprofloxacin resistant phenotypes of EC5 and EC6. Cells treated with 10mg/L of ciprofloxacin - below the MIC value.

3.2 Segmentation and Classification of Clinical Isolates

By examining the phenotypes of clinical isolates, we managed to find evidence for our underlying assumptions from the previous chapter: the untreated phenotype is a good proxy for the resistant phenotype, the resistant and susceptible phenotypes show high homogeneity between different clinical isolates, and that the phenotypic response appears to be a function of both treatment conditions and MIC. Armed with this, we proceeded to the next question - does the response of our model show similar homogeneity when applied to clinical phenotypes?

To proceed, we chose to use the same segmentation model we used in the previous chapter, and the holdout ciprofloxacin-susceptible classifier. Both models were applied on the FoVs of clinical isolates. The holdout classifier was trained on a balanced dataset, and used to evaluate our metric for performance on unseen data - now, we chose to extend it to unseen clinical isolates. The models underwent no fine-tuning or transfer learning from the clinical isolate data, and made all decisions purely on what it learnt on the single susceptible training strain.

3.2.1 Clinical isolates at a fixed treatment point

For each of the clinical isolates EC1-6, we performed the DASP wetlab experiment in triplicate (Figure 2.1.2, see also *Experimental methods* sections 3.3 and 2.5), producing 3 independent collections of FoVs; each collection included FoVs with treated cells, and FoVs with untreated cells. We applied the segmentation and classification models to each individual FoV from each of the isolates. In FoVs containing untreated cells (Figures 3.2.1 A,C; 3.2.2 A,C; 3.2.3 A,C) we saw the models classified the majority of cells as resistant (red overlays) - this was expected as the cells had not been treated, thus could show no reaction. Examining the false-positive errors (susceptible detections in untreated FoVs; blue overlays), we saw 3 categories of errors. Firstly, certain untreated cells showed a degree of nucleoid compaction (like we discussed in the previous section), and thus were classified as susceptible - errors caused by biological heterogeneity of the sample and factors beyond experimental control. Secondly, errors in segmentation, including both incomplete and spurious segmentations (Figure 3.2.2 A contains a good example) - errors originating in the first stage of the pipeline and propagating upwards. Thirdly, some cells were classified as susceptible for no visually apparent reason - true false-positive errors generated by the pipeline.

Next we examined the treated FoVs of the isolates. In the highly susceptible isolates EC1-3 (Figures 3.2.1 B,D; 3.2.2 D), the classifier correctly classified the majority ($> 80\%$) of the cells as susceptible. Of the cells that were not, many did not show nucleoid compaction and resembled the resistant phenotype (and thus were classified as such).

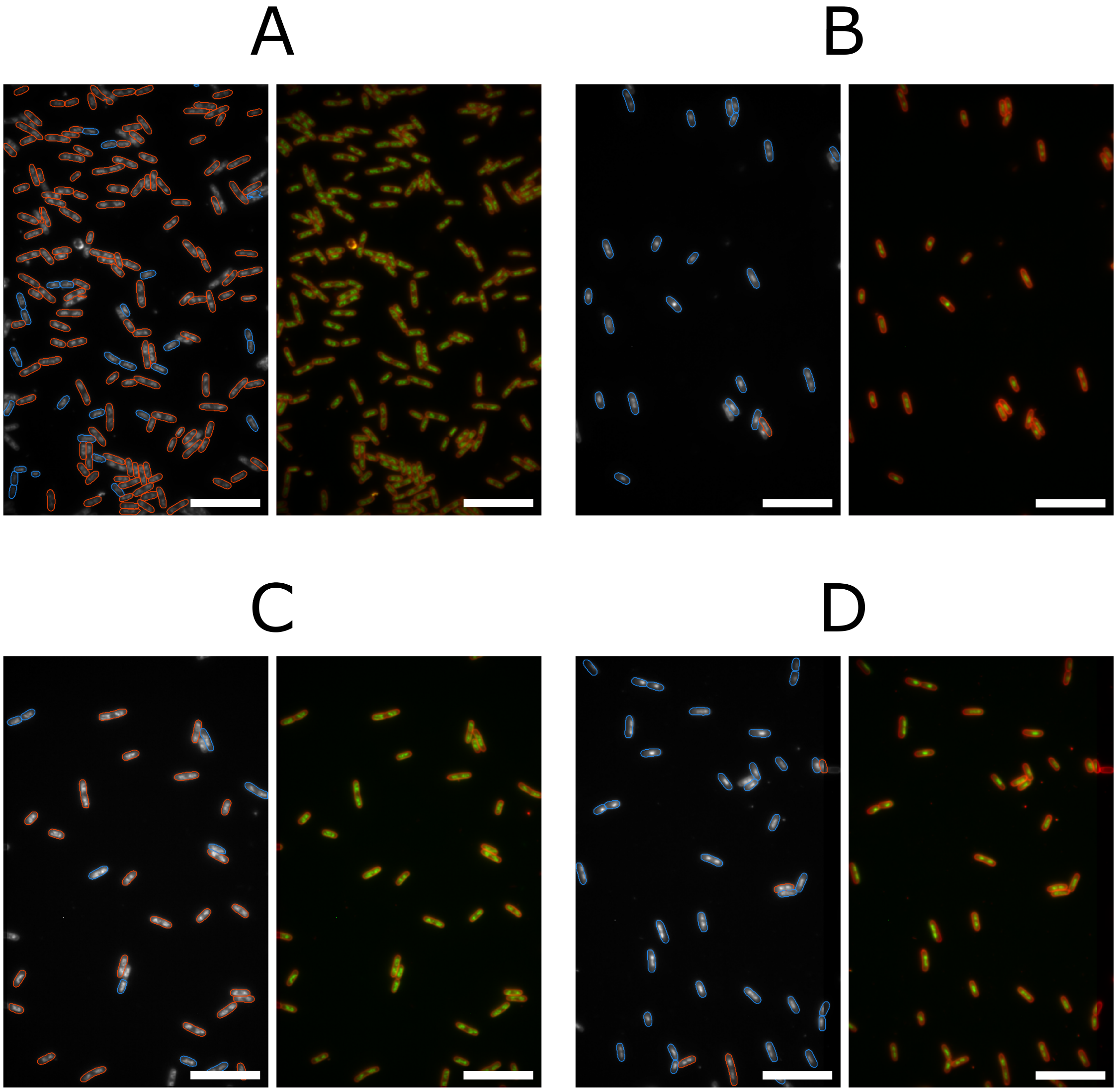


Figure 3.2.1: Representative FoVs and phenotype detections of ciprofloxacin-susceptible clinical isolates EC1 and EC2, in both untreated and ciprofloxacin treated conditions.

A. (*Left*) Detection overlay of a field of view of untreated EC2. Red detections are classified as resistant/untreated, blue detections are classified as susceptible. Scale bar indicates 5µm. (*Right*) Corresponding raw field of view.

B. As **A**, but for a ciprofloxacin treated EC2 field of view.

C. As **A**, but for untreated EC1.

D. As **B**, but for ciprofloxacin treated EC1.

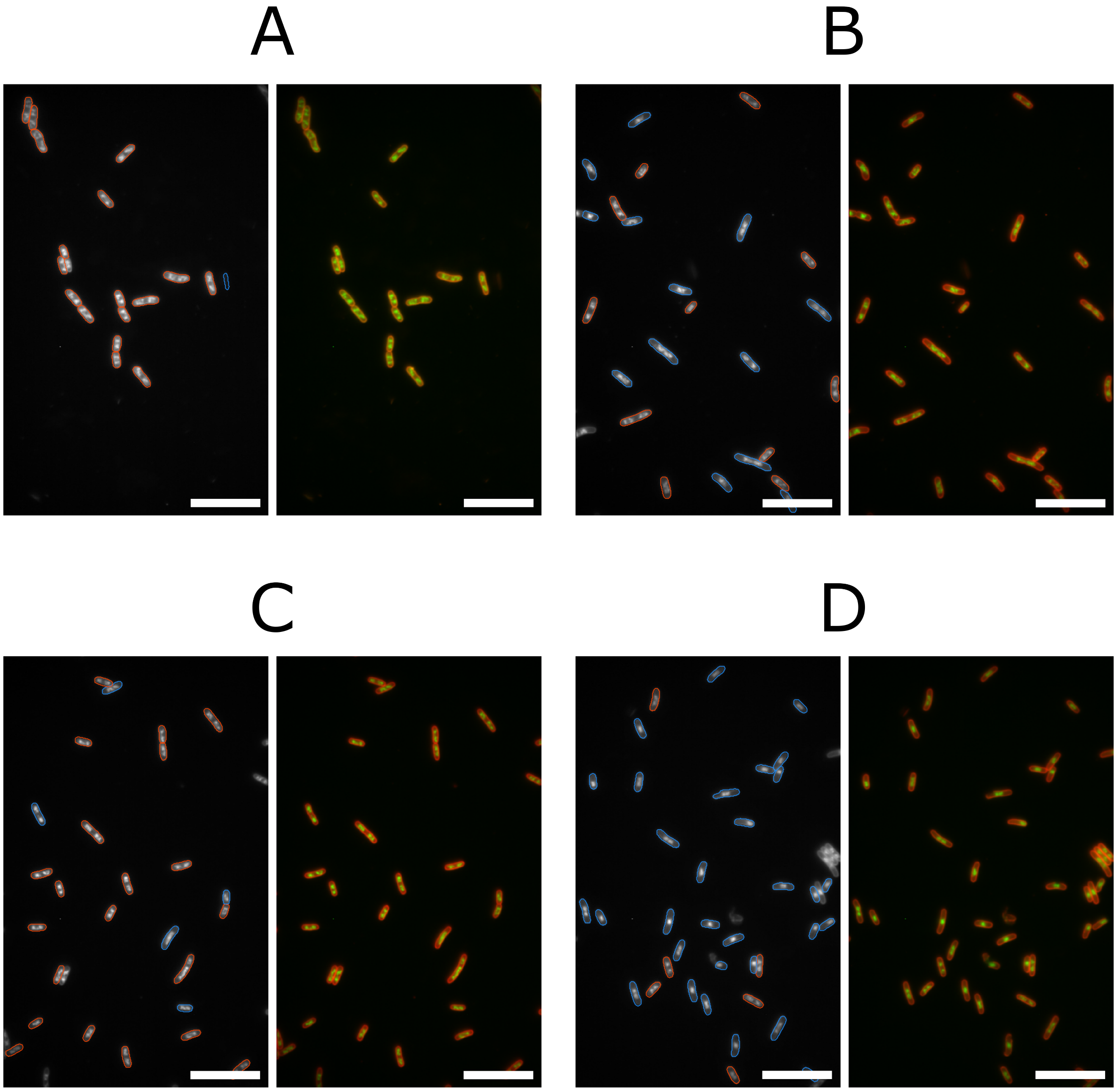


Figure 3.2.2: Representative FoVs and phenotype detections of ciprofloxacin-susceptible clinical isolates EC3 and EC4, in both untreated and ciprofloxacin treated conditions.

A. (*Left*) Detection overlay of a field of view of untreated EC4. Red detections are classified as resistant/untreated, blue detections are classified as susceptible. Scale bar indicates 5µm. (*Right*) Corresponding raw field of view.

B. As **A**, but for a ciprofloxacin treated EC4 field of view.

C. As **A**, but for untreated EC3.

D. As **B**, but for ciprofloxacin treated EC3.

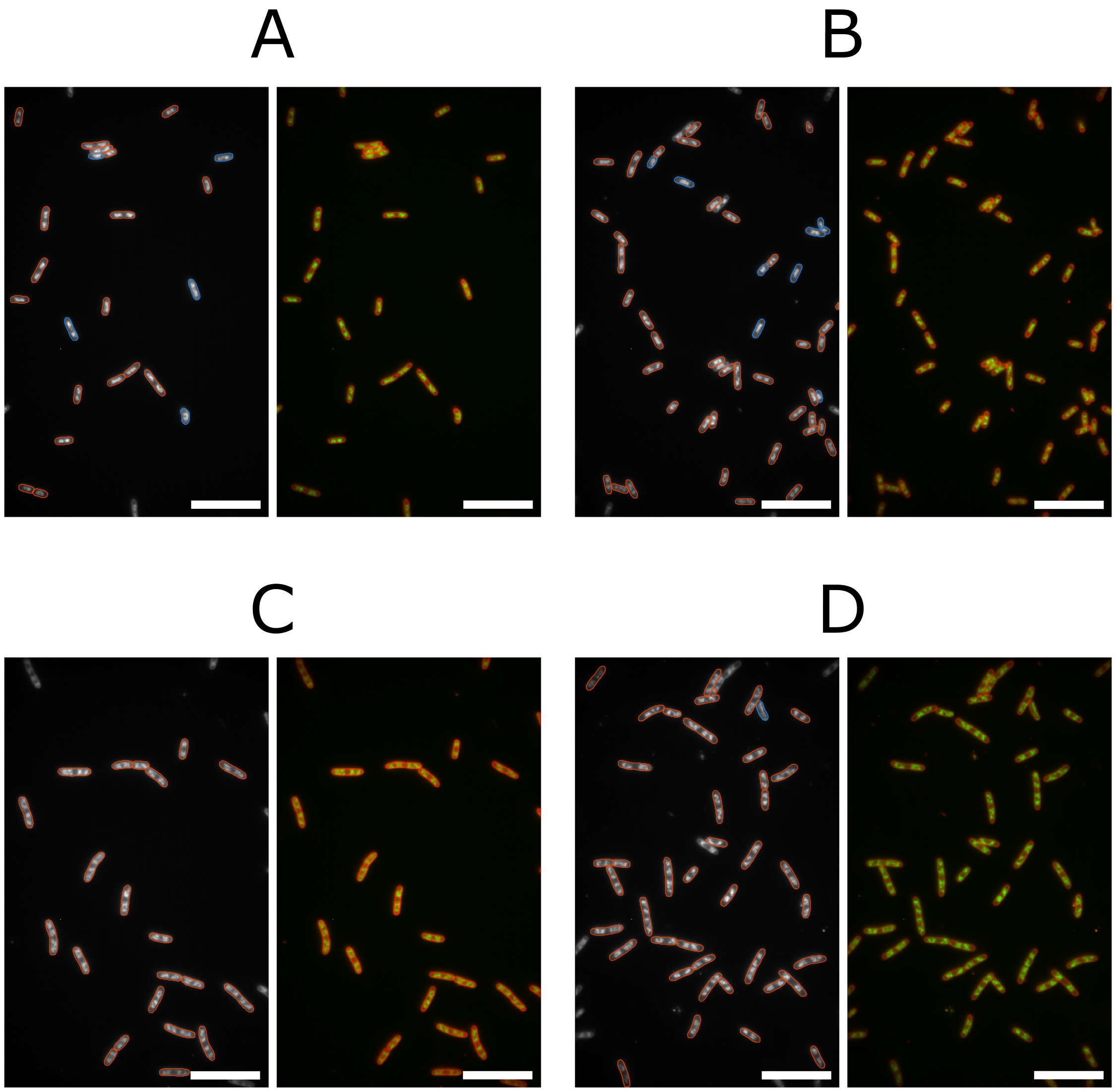


Figure 3.2.3: Representative FoVs and phenotype detections of ciprofloxacin-susceptible clinical isolates EC5 and EC6, in both untreated and ciprofloxacin treated conditions.

A. (*Left*) Detection overlay of a field of view of untreated EC6. Red detections are classified as resistant/untreated, blue detections are classified as susceptible. Scale bar indicates 5µm. (*Right*) Corresponding raw field of view.

B. As **A**, but for a ciprofloxacin treated EC6 field of view.

C. As **A**, but for untreated EC5.

D. As **B**, but for ciprofloxacin treated EC5.

Again, this was likely due to biological heterogeneity of cells in culture, as not all cells would respond to the antibiotic on the same timescale (a source of error analogous to the first type we discussed above).

In the case of ciprofloxacin, the nucleoid condensation susceptible phenotype is the result of the DNA fragmentation (creation of double-stranded breaks) induced by gyrase and topoisomerase IV inhibition [169], which the cell attempts to repair by homologous recombination [170] of both copies of the nucleoid in the centre of the cell. The fragmentation happens when a DNA-bound topoisomerase is inhibited by ciprofloxacin, preventing it from ligating the double-stranded break it has created. Thus, both the rate of damage accumulation and the rate of damage repair are dependent on the metabolic activity rate of the cell. It is thus possible that cells that are less metabolically active would have a slower rate of phenotype conversion; metabolic heterogeneity inside the culture thus leading to a phenotype conversion heterogenity. Such metabolic heterogeneity has been observed for E.coli [171].

The proportion of cells classified as resistant increased in the intermediate susceptibility case of EC4 (Figure 3.2.2 B); fewer cells still appeared to be responding. This trend continued when we considered the resistant isolates EC5-6 (Figure 3.2.3 B,D), with very few cells responding and the majority classified as resistant. As before, a small proportion of cells did respond and were classified as susceptible - however, given that we could also observe this in untreated samples, it was hard to take this further to definite conclusions. The other two sources of error we observed in untreated samples were similarly present in treated samples.

It was instructive to consider individual FoVs, but to make decisions on the susceptibility of an entire culture, we had to look beyond and consider the distribution of classifications in each of the triplicate experiments. From each experiment, we analysed every FoV in the collection, and pooled together the single-cell classifications separately for treated and untreated conditions.

Without any antibiotic treatment, the distribution of cells was indeed skewed heavily towards the resistant class - for example, we obtained 77% and 96% resistant classifications for EC2 and EC5 respectively in a single representative experiment

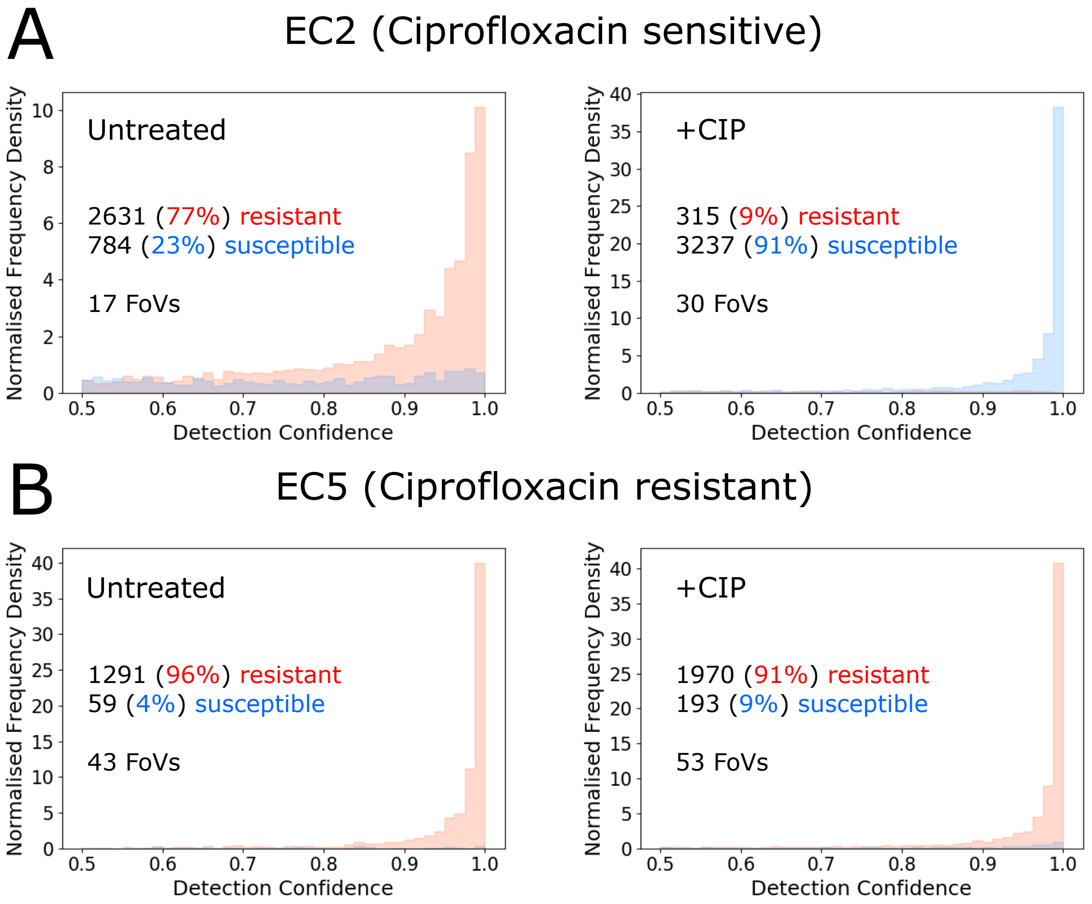


Figure 3.2.4: Distribution of single-cell classifications is sensitive to the resistance of clinical isolates.
A. (*Left*) Histogram of single-cell classifications from a representative imaging experiment of untreated, susceptible E. coli isolate EC2 cells, normalised to the total sum of detections, as a function of classification confidence. (*Right*) As before, but for ciprofloxacin treated EC2 cells.
B. As **A**, but for ciprofloxacin resistant EC5.

(Figure 3.2.4 A left; B left). On the other hand, in the treated condition, susceptible EC2 showed 91% susceptible classifications, while resistant EC5 showed only 9% susceptible. (Figure 3.2.4 A right; B right). High classifier confidence was retained in all cases.

This pattern was maintained across the library of all clinical isolates and across all repeat experiments. Next, we pooled together the distribution changes across all repeat experiments for each isolate, and again compared untreated and treated conditions. To compare pairwise means of treatment conditions, we carried out a Tukey's range test (essentially a t-test). In all isolates susceptible to ciprofloxacin at the level used (EC1-4), we observed a statistically significant ($p<0.001$) difference in the ratio of cells classified as susceptible between the untreated and treated conditions (Figure 3.2.5). In concordance with previous observations, the size of the ratio appeared to correlate with the difference between the isolate MIC and the treatment concentration. In highly susceptible EC1-3 85% cells were classified as susceptible in treated samples, whereas in intermediate EC4, only 38%. Further, in resistant isolates EC5-6, although the exact ratio of cells classified as susceptible varied, there was no statistically significant difference between the untreated and treated samples. The number of cells detected in each experiment and treatment condition that contributed varied between 500 and 3500 cells (Figure 3.2.6); providing solid statistics even within each experiment, which contributed to the relatively small ($n=3$) amount of repeat experiments needed.

Thus, we discovered that not only the phenotype response of the cell was homogeneous between isolates and a function of treatment conditions and MIC, but also that our models' responses indeed scaled similarly and were consistent with our results from the previous chapter. We demonstrated this holds on the level of individual FoVs, individual repeat experiments drawn from the same colony and imaged together, and multiple repeat experiments drawn from multiple colonies and imaged separately. At this level, we demonstrated the changes in phenotype distributions between untreated and treated cells could be used to inform of resistance around a fixed treatment point, providing a binary decision.

We however saw our results could be complicated by biological heterogeneity of the samples - cells not responding as planned in treated samples, or appearing to match the susceptible phenotype even in untreated samples. Further - all results so far presented had come from experiments on pure monocultures, where the reaction of all cells to the antibiotic treatment was governed by the same underlying distribution (however unknown to us at this time). To further consider the potential clinical utility of the assay, we needed to establish how the DASP response scales with the ground-truth ratio of resistant cells in the imaged population - which we could not achieve in monocultures. When cultured in liquid medium, we would

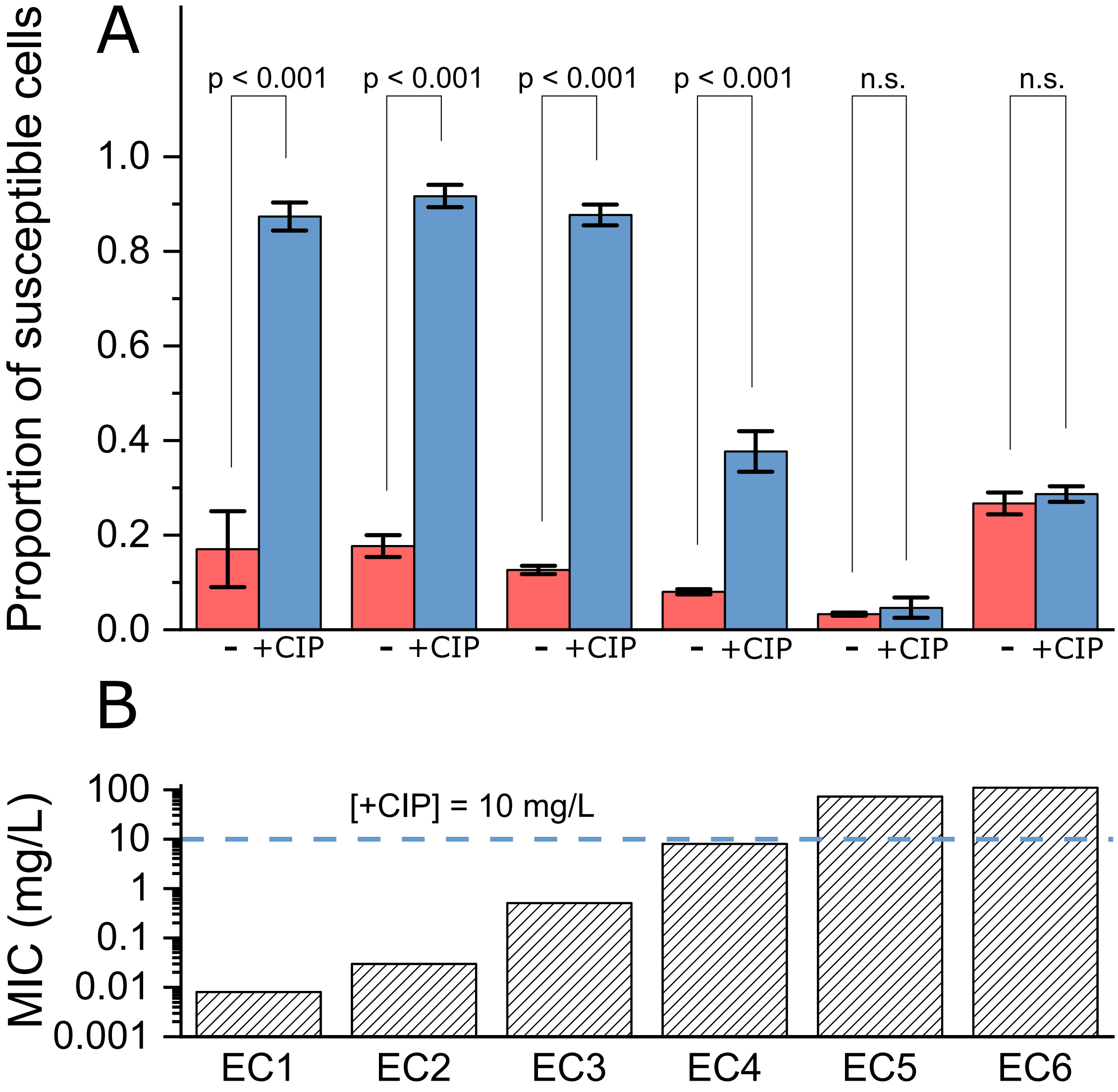


Figure 3.2.5: Changes in detection distribution upon treatment correlate with the degree of resistance in clinical isolates.

A. Ratio of susceptible phenotype detections across different clinical isolates, in untreated (-) and ciprofloxacin treated (+CIP) samples. Average of 3 biological replicates, error bars show the standard error of the mean. Overlay shows Tukey range test p-values, carried out pairwise between corresponding untreated and treated sets of repeats (n=3), test carried out at significance level of 0.05, "n.s." indicates not significant.

B. Ciprofloxacin MICs of the clinical isolates. Horizontal line indicates the treatment concentration used in the treated samples.

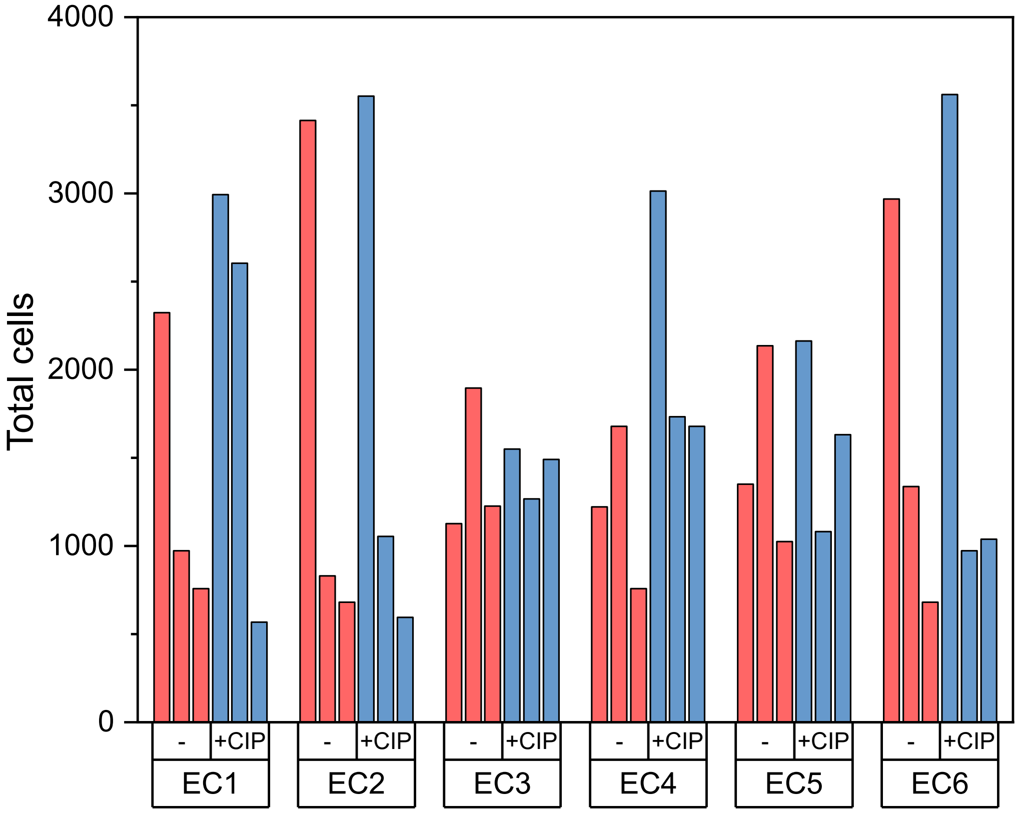


Figure 3.2.6: Total numbers of cells detected in each of the biological replicate experiments of clinical isolates. (*top axis*) Treatment condition. (*bottom axis*) Isolate code.

expect cells to grow and react to the antimicrobial individually and separately - the response of any cell is not affected by the response of any other. From this, we thus expected DASP to show linear scaling with changing ratio of resistant cells. An assay that shows a uniform response (reporting the same value regardless of the ground-truth concentration of cells) clearly achieves no sensitivity towards single cells. Conversely, an assay showing a non-linear response would indicate either that it is sensitive to something else other than single cells, or the assumption about the independence of single cell phenotypic response was incorrect.

Our next direction was to examine precisely this, by looking at mixed samples containing treated resistant and susceptible isolates in different proportions, and calculating the smallest possible concentration of resistant cells DASP could pick up. Beyond technique validation, this also had even more direct clinical motivation - missing resistant cells has greater patient impact than missing susceptible cells.

Further, many infections can feature ‘persister’ cells [172], which slow down their metabolism in response to stress, only to ‘fire up’ again afterwards, restarting the infection. To proceed, we thus repeated the above analysis, but this time in polymicrobial, mixed samples, which we discuss next.

3.2.2 Mixed samples and the Limit of Detection

To accomplish our aim of evaluating performance on mixed samples, we picked two of our isolates - susceptible EC3 and resistant EC5. We cultured colonies of the two isolates separately (separate monocultures of EC3 and EC5) using the same protocol we had followed so far, to the same OD. This was repeated for 3 biological replicates, creating 3 monocultures each of EC3 and EC5. To verify that our separate monocultures at the same OD indeed had the same CFU density prior to mixing, we carried out colony count measurements on the grown monocultures (see *Experimental methods* section 3.3, colonies counted by Dr. Conor Feehily) across a serial dilution in excess medium, using 3 samples per monoculture per dilution step - we confirmed that no statistically significant difference in the CFU density existed between any of the monocultures (Figure 3.2.7 *Right*).

All of the separate monocultures were then treated with ciprofloxacin at the concentration and time used so far (Figure 2.5.1 for reference), and fixed with formaldehyde. These monocultures were then used to prepare a 2-fold serial dilution of resistant monocultures (resistant culture ratio of 1) in susceptible monocultures (see *Experimental methods* section 3.3), creating 3 independent biological replicates for each dilution step. The experimental pipeline was then continued from this point: samples were incubated with ethanol, stained and imaged, producing 3 independent collections of FoVs for each dilution step.

To provide a theoretical prediction for the number of resistant cells in mixed samples, we developed a linear model (see *Experimental methods* section 3.3, Equation 3.3.1) and used results obtained from phenotyping monocultures (Figure 3.2.5), producing $p(EC3) = 0.12$, $p(EC5) = 0.95$ $\sigma_{p(EC3)} = \sigma_{p(EC5)} = 0.038$. As the CFU density between repeats was measured to be not statistically significant, we did not include it as an additional explanatory variable in the linear model or error propagation.

To match this with experimental data, we ran our segmentation and classification pipeline on the collections of FoVs in each dilution step, this time reporting the proportion of resistant cells, and plotted the averages in each dilution step together with the corresponding theoretical prediction (Figure 3.2.7 *Left*). In the theoretical model, we observe a perfect linear relationship across the entire range - notably, the proportion of resistant cells in purely susceptible culture (R=0) is non-zero - this reflects the presence of cells that are either non-responding, or miss-classified. Similarly, in purely resistant cultures (R=1), the ratio does not saturate at 1. Together, this models the effect of the biological heterogeneity and classifier error that we observed and discussed earlier in this chapter. Comparing experimental data to the theoretical prediction, we observe the experimental relationship to be still linear as expected, but flatter (theoretical gradient of 0.83, against experimental measurement of 0.49 ± 0.03), indicating a loss of sensitivity to resistant cells that is statistically significant (theoretical prediction lies outside the 95% confidence band of experimental fit). Figure 3.2.7 shows the number of cells identified in each of the replicate experiments.

To proceed, we carried out residual analysis on the linear fit to experimental data. First, we checked that our residuals are normally distributed by comparing against the expected normal distribution percentiles (Figure 3.2.10) We observed a roughly linear relationship, qualitatively in agreement with the normality assumption. We saw possible signs of sigmoidal kurtosis which may be indicative of an underlying uniform distribution - however, we did not have enough data to investigate this. Further, our fits were carried out to averages of experimental repeats - the sampling distribution of the mean of any underlying distribution tends to the normal distribution anyway [1]. Thus, we concluded that any sign of kurtosis was an artefact of limited (n=3) experimental repeats. Inspecting the residuals, we checked them for outliers by externally studentising the residuals (see *Experimental methods*

---

[1] Central limit theorem.

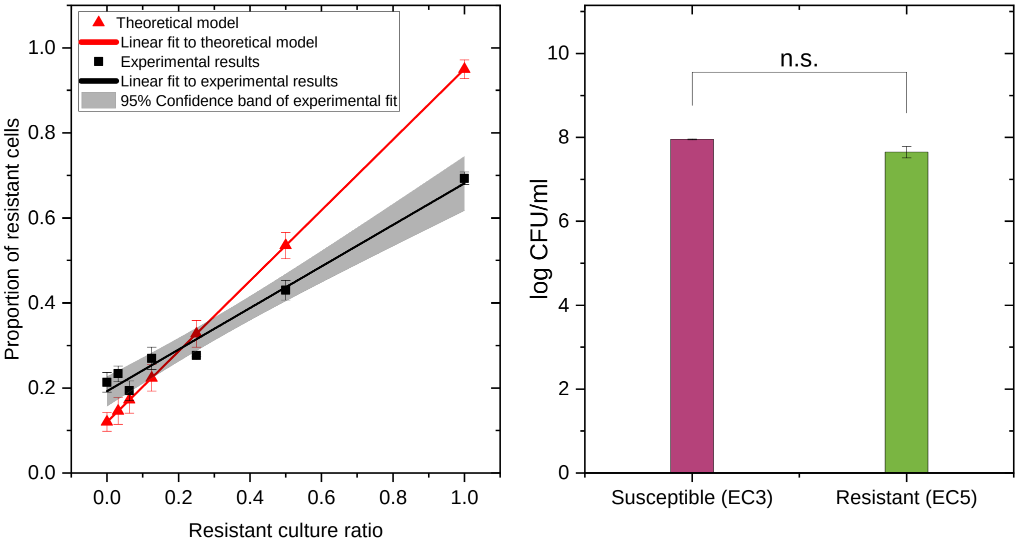


Figure 3.2.7: Proportion of resistant classification is a linear function of dilution of resistant culture in susceptible cells. (*Left*). The proportion of resistant classifications as a function of the dilution of resistant cells in susceptible cells. The resistant culture ratio represents the fraction of the resistant culture in the total sample In black, average of 3 biological replicates, with a linear regression trendline, error bars show the standard error of the mean. In red, theoretical prediction for the experimental results, with a linear fit, and error bars showing expected standard error for a mean of 3 replicates.(*Right*). Colony counts for both resistant (EC5) and susceptible (EC3) cultures, prior to treatment and mixing. Overlay shows the the Tukey range test p-values, carried out between mean CFU counts of both isolates (n=3) at significance level of 0.05, "n.s." indicates not significant.

section 3.3) and hypothesis-checking them in a two-tailed test at a significance level of 0.05 (Figure 3.2.9 *Bottom*). This analysis revealed no outliers, and no data points were excluded from further analysis.

To proceed to the limit of detection (LoD), we calculated the standard deviation of the residuals (Figure 3.2.9 *Middle*). Finally, we calculated the LoD as a function of the residual standard deviation (σ_{Res}) and slope of the linear regression line (m), according to the guidelines of the the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) [173]:

$$LoD = \frac{3.3\,\sigma_{Res}}{m} = 0.16 \tag{3.2.1}$$

Interpreting our result, we saw our LoD was around 15% of resistant culture mixture. However, it was hard to interpret this further - we did not have sufficient

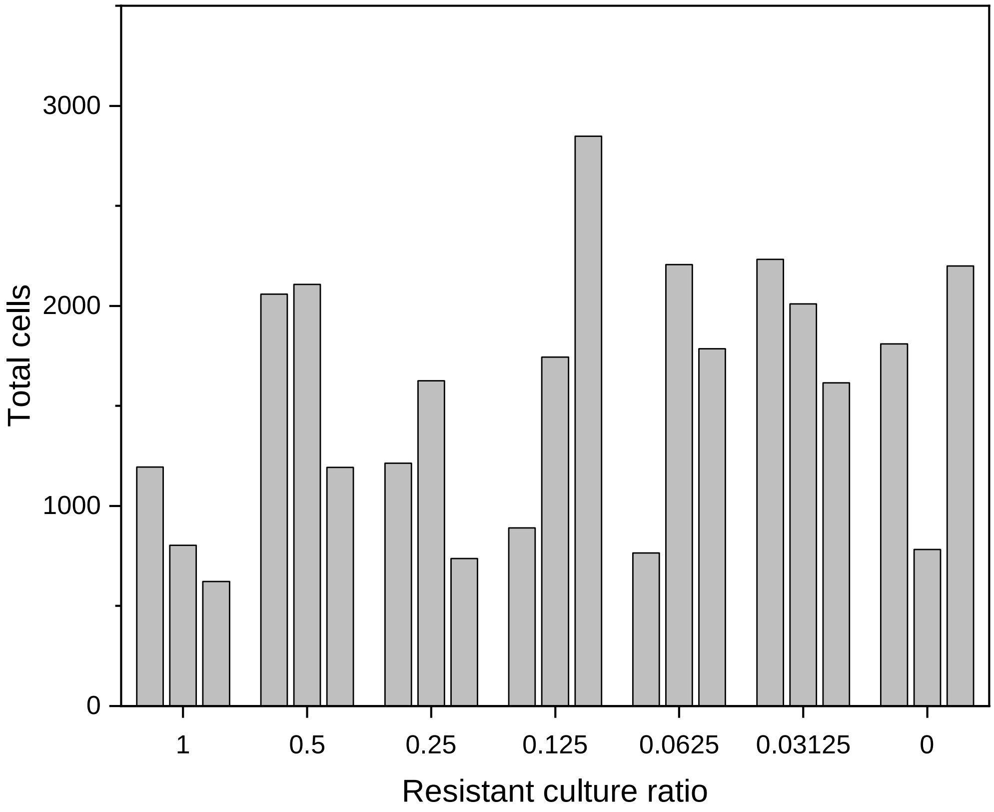


Figure 3.2.8: Total numbers of cells detected in each of the biological replicate experiments of the mixed samples, in each dilution step.

information to quantify precisely how many cells classified as resistant in either of the monocultures were actually resistant, or were mistakes made by the classifier. More concisely, we had no information with which we could parameterise the distribution that actually governs that phenotypic response of single cells as a function of treatment. Thus, it was difficult to infer from this LoD the actual LoD of resistant cells - the true measure of interest.

Nevertheless, we were reassured to see a linear scaling of the DASP response with decreasing ratio of resistant cells. By doing so, we indeed verified that DASP is sensitive to the phenotypic response of single cells, and that this response is independent for each cell.

So far in this chapter, we have demonstrated that the DASP response scaled homogenously with the phenotypic response of the cells, is a function of MIC and treatment conditions, and that it is sensitive the concentration of single-cells showing

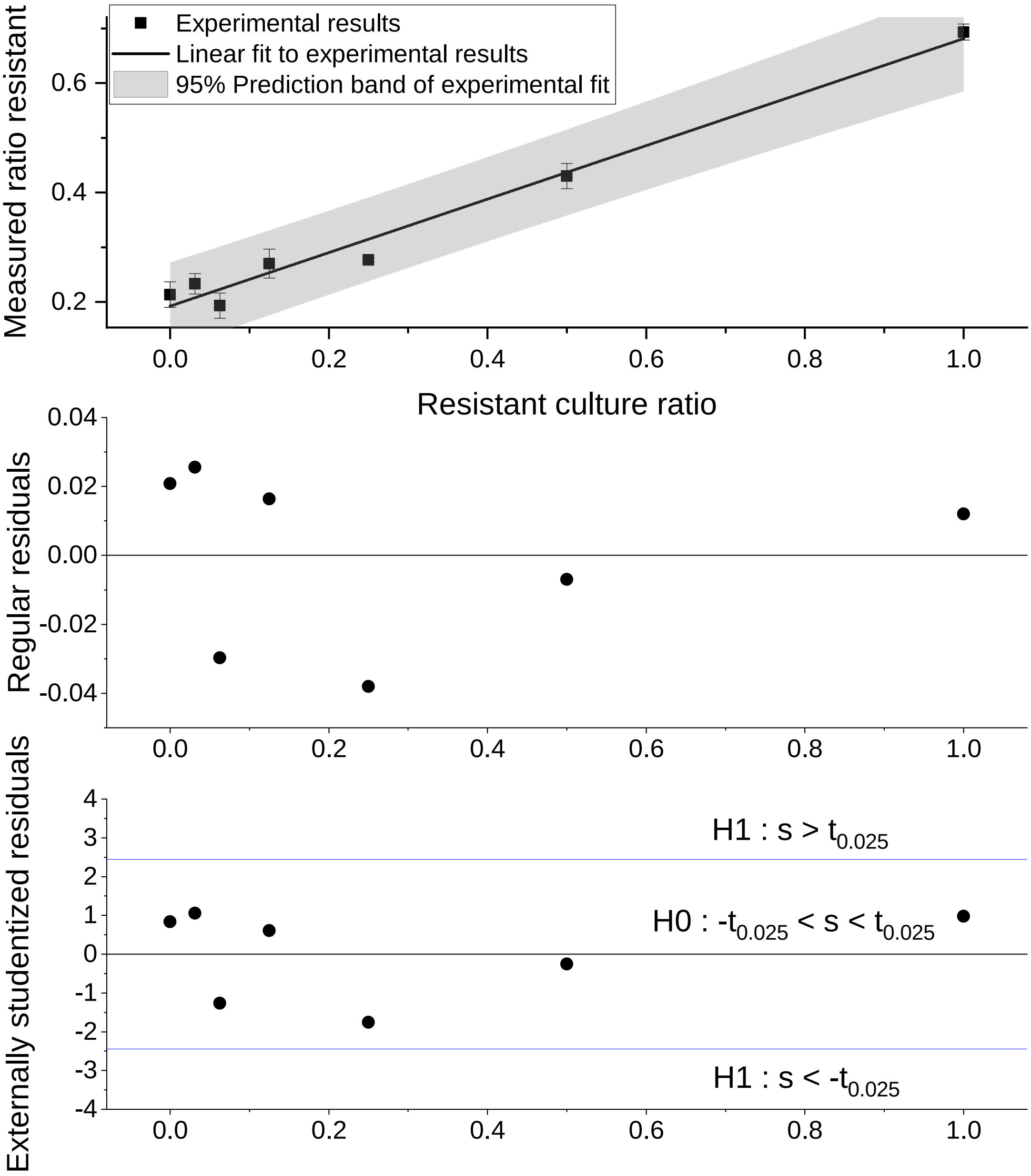


Figure 3.2.9: Residual analysis and limit of detection of resistant cells. (*Top*) Experimentally measured proportion of resistant classifications as a function of the dilution of the resistant cells in susceptible cells. Average of 3 biological replicates, error bars indicate standard error of the mean. 95% prediction interval indicated. (*Middle*) Linear fit residuals. (*Bottom*) Externally studentized linear fit residuals. Blue horizontal lines indicate thresholds of significance for a 2-tailed t-test at significance level of 0.05.

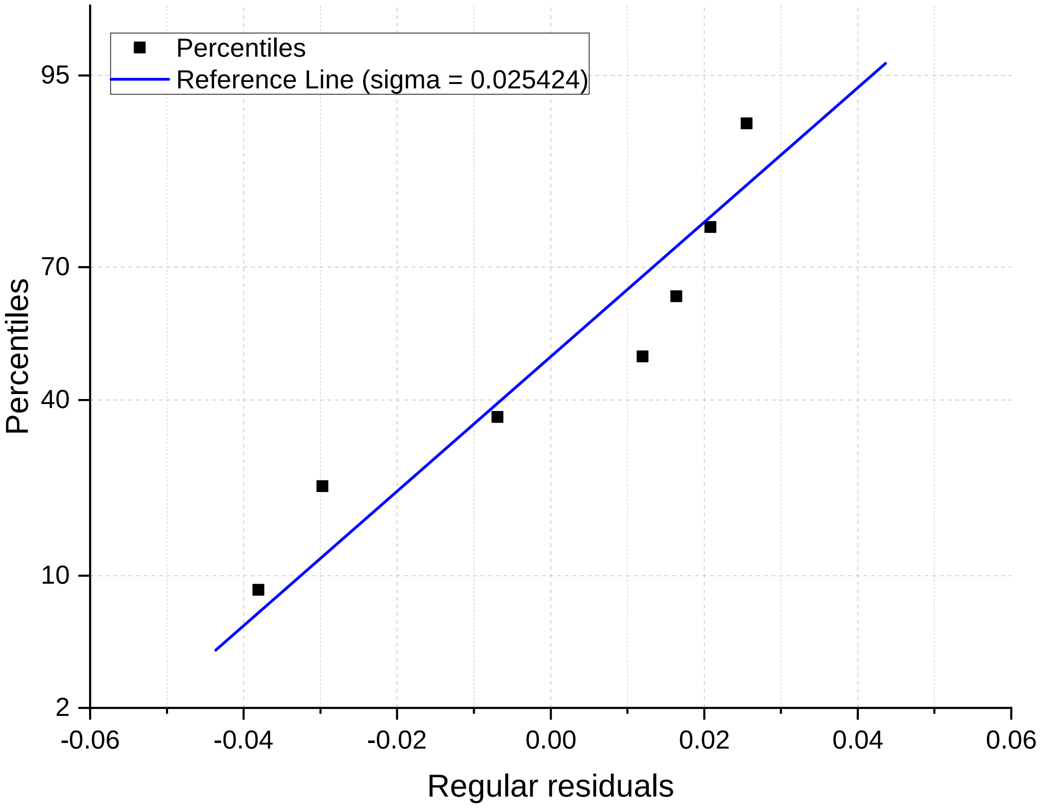


Figure 3.2.10: Normal probability plot of the linear fit residuals. Straight line indicates fit to a normal distribution.

different phenotypes. The last piece of the puzzle was to quantify the relationship between the DASP response and the MIC and treatment concentrations of isolates, which we could already see qualitatively in Figure 3.2.5. So far, all treated samples were treated at the same concentration (Figure 2.5.1). To proceed, we relaxed that condition, and treated different isolates with different concentrations of ciprofloxacin.

3.2.3 Antibiotic titration allows MIC quantification

Again, we picked a subset of our clinical isolates with diferent MICs to ciprofloxacin - the highly susceptible EC1 (MIC=0.008 mg/L), intermediate EC3 (MIC=0.5 mg/L), and resistant EC5 (MIC=72 mg/L). Once again, separate cultures of each of the isolate were prepared, but instead of treating with the same, fixed concentration of ciprofloxacin (10 mg/L) for 30 minutes, cultures were treated with a range of different concentrations, but also for 30 minutes. We anticipated the development of the phenotype to be a function of isolate MIC and both treatment time and concentration - thus kept the treatment time constant to investigate just the concentration dependence. To minimize dilution errors and provide a wide concentration range, the dilutions were prepared as 2-fold, and later 10-fold serial dilutions of the same ciprofloxacin stock solution. Following treatment, cultures were once again fixed, stained and imaged using the same protocol as before. Once again, samples were collected and processed in biological triplicate - therefore producing for each isolate and treatment concentration 3 independent collections of FoVs, which were then ingested into the phenotyping computational pipeline. In line with previous experiments, our main reporting measure was the proportion of susceptible cells as classified by the classifier.

Across all isolates (Fig 3.2.11 A), cells treated at sub-MIC concentration did not show a significant shift away from the resistant phenotype - the cells did not respond to a sub-MIC concentrations of ciprofloxacin, did not display the susceptible phenotype, and thus the proportion susceptible cells as decided by the classifier remained close to the baseline (of untreated cells). Conversely, at high treatment concentrations in excess of the MIC, we observed a strong response, with over 90% of cells classifier as susceptible. At intermediate values between these two asymptotes, the response varied logistically. In order to quantify this relationship, we proceeded to fit dose-response models (see *Experimental methods* section 3.3) for further details. In all repeat experiments, we maintained > 500 analysed cells (Fig 3.2.12), contributing to robust statistics.

DASP provides a phenotypic dose-response relationship after as little as 30 minutes of treatment. As we discussed previously, the current gold-standard assay quantifies culture growth over longer periods of time as a function of treatment concentration, not single-cell phenotypic responses like DASP. To compare our assay against the gold standard, we measured growth curves of the isolates as a function of both ciprofloxacin treatment concentration (same range as used for evaluating DASP) and time, by measuring the OD of growing cultures over 24 hours (see *Experimental methods* section 3.3, raw data collected by Dr. Conor Feehily). From

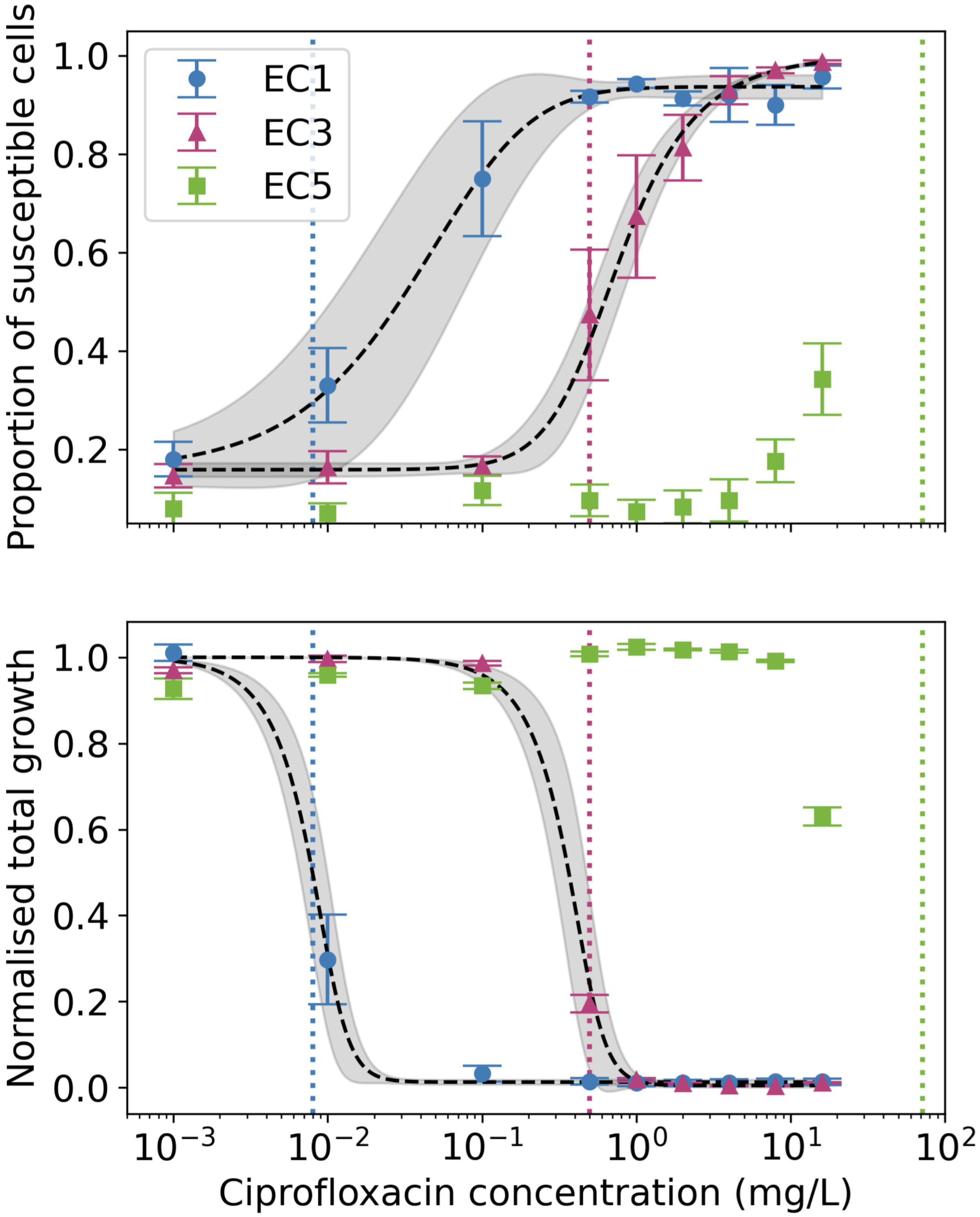


Figure 3.2.11: DASP provides equivalent information to a 24-hour growth assay, but within 30 minutes.

A. Ratio of susceptible detections as a function of ciprofloxacin treatment concentration. Average of three biological replicates, error bars show the standard error of the mean. Vertical coloured lines show the ciprofloxacin MICs of the isolates used. Black lines show the dose-response fit, grey regions show the 95% confidence band of the fit.

B. Total bacterial growth in liquid culture over 24 hours in the presence of ciprofloxacin, normalised to the growth of untreated cells. Measured MICs of the isolates are: EC1 – 0.008 mg/L, EC3 – 0.5 mg/L, and EC5 – 72 mg/L.

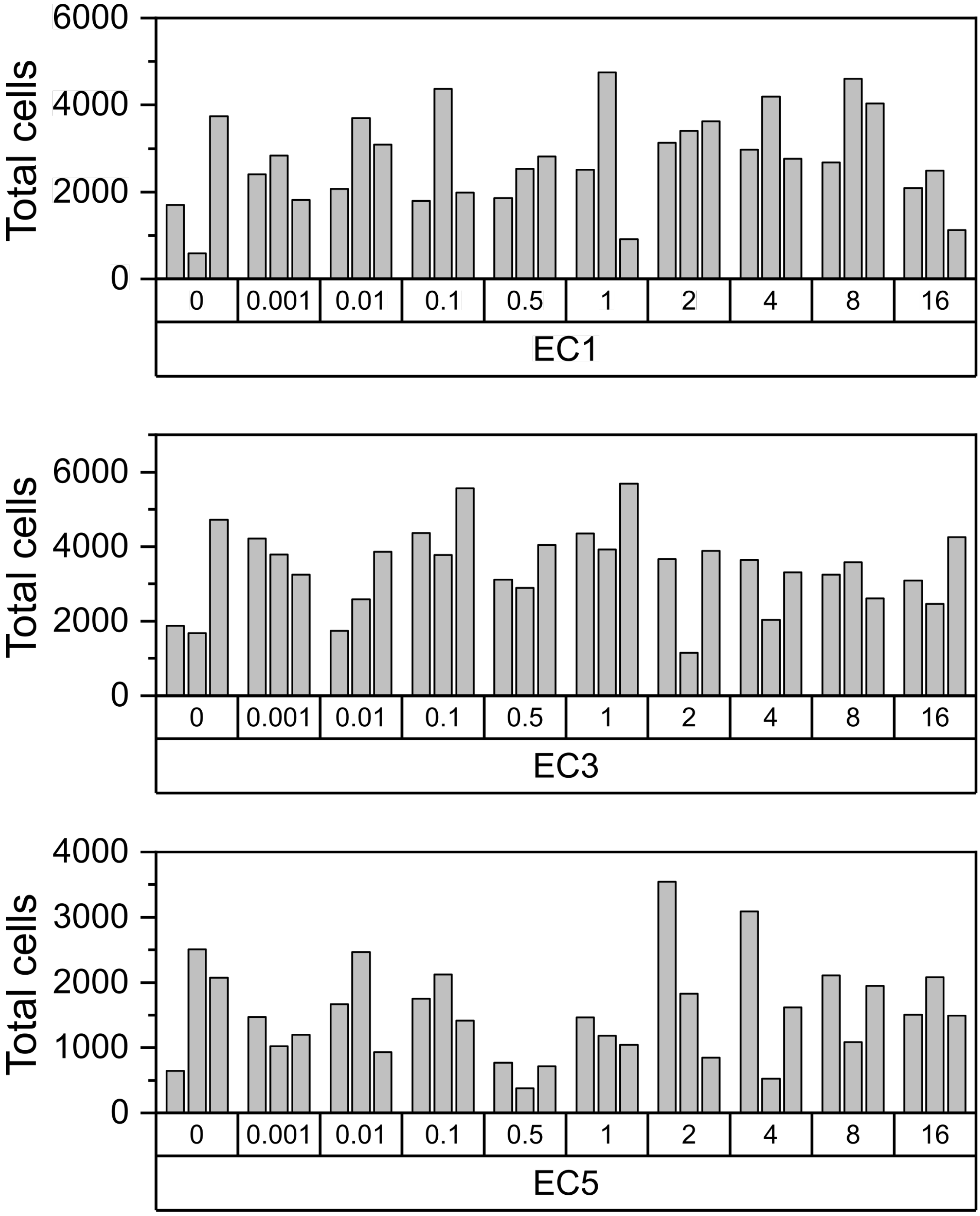


Figure 3.2.12: Total numbers of cells detected in each of the biological replicate experiments of ciprofloxacin titrations, in every isolate and ciprofloxacin concentration (*Top axis*) Ciprofloxacin treatment concentration in mg/L. "0" indicates no antibiotic treatment. (*Bottom axis*) Clinical isolate code.

this, we proceeded to calculate the total cell growth after 24 hours by numerical integration of the time-resolved signal in time. This was then normalised to the total growth of untreated cells, hereby creating a culture-wide ratio of total cell growth as a function of treatment concentration. Across all isolates and treatment concentrations, the total cell growth followed a relationship reciprocal to the DASP response. At sub-MIC concentrations, cell growth was not inhibited, and the cell growth ratio was close to 1 (indicating cells were growing like the untreated controls). At high treatment concentrations, cells did not grow, and the ratio was close to 0. Again, between the asymptotes cell growth varied logistically, and we fit appropriate dose-response models.

In both dose response models, the approximate inflexion points of the curves (corresponding to free parameter c in the fit) were measured to be close to the true MIC values of our isolates, as measured and presented in previous sections. Specifically, for EC1, the measured MIC was 0.008 mg/L, with a corresponding DASP response inflexion at 0.17 ± 0.11 mg/L; and for EC3, the measured MIC was 0.5 mg/L, and the inflexion occurred at 0.49 ± 0.18 mg/L. In the comparative case of cell growth, the EC1 inflexion took place at 0.011 ± 0.002 mg/L, and EC3 at 0.49 ± 0.02 mg/L. We saw in the case of EC3, our results were in high agreement - the inflexion points of both the DASP response and the cell growth models matched the true MIC. In the case of EC1, there was some divergence in the DASP response. Overall, we concluded our results demonstrated that DASP provide valuable MIC-related information for different clinical isolates, faster than established methods. However, further work is required to establish the effect of isolate-to-isolate heterogeneity on the precise location of the MIC along the dose-reponse model.

3.3 Experimental Methods

Clinical isolates

Clinical isolates used were blood culture isolates of *E. coli* processed routinely for diagnostic purposes and stored by the Microbiology Laboratory of the Oxford University Hospitals NHS Foundation Trust, Oxford, UK. Upon processing and

Isolate name	Genotype	MIC (mg/L)
EC1	*marR*^{N3}	0.008
EC2	*marR*^{N3}	0.03
EC3	*gyrA*^{L83} *parC*^{I80}	0.5
EC4	*gyrA*$^{L83, N87}$ *parC*^{I80}	8
EC5	*gyrA*$^{L83, N87}$ *parC*$^{I80, V84}$ *parE*^{L529}	72
EC6	*gyrA*$^{L83, N87}$ *parC*^{I80} *parE*^{A458}	108

Figure 3.3.1: Genotypes and MICs of clinical isolates. Isolate name refers to the pseudorandom code used for the isolate throughout this work. The genotype column contains identified mutations that are considered associated with resistance to ciprofloxacin, and the MIC column lists the experimentally derived ciprofloxacin MIC values of the isolates.

storage, the isolates were pseudoanonymised and all patient-identifiable information was removed from records. Upon re-use for this work, the isolates underwent a second round of pseudoanonymisation, and assigned codes EC1-6, in order of decreasing susceptibility to ciprofloxacin.

Isolates were cultured, treated, processed and imaged in the same way as MG1655 samples, described previously. For ciprofloxacin titration assays, the antibiotic was serially diluted across a concentration range from 16 mg/L to 0.001 mg/L and co-incubated with 1 ml aliquots of bacteria as described previously.

All isolates in the study had been whole-genome-sequenced on the Illumina platform as described previously [168], and AMR genotypes were assigned using the ResFinder [174] database with Abricate v0.9.8 46 (–min-id 95 –min-cov 95). The MICs of the clinical isolates were calculated empirically by E-test strip, or where the MIC exceeded the range of the strip, by a broth dilution.

Ethics

We analysed bacterial isolates routinely stored by the John Radcliffe Hospital Microbiology laboratory; no sampling was specifically undertaken for the purposes of this study and no patient-related information was accessed. In the UK, bacterial isolates routinely cultured from human clinical samples do not require ethical approval for analysis under the provisions of the Human Tissue Act as they do not contain any material considered to be human tissue.

Mixed samples

Monocultures of EC3 and EC5 were prepared as described previously, treated with ciprofloxacin at 10mg/L and fixed in formaldehyde. An aliquot was kept of each of the pure fixed monocultures. Further, pairs of monocultures were mixed in equal proportion by volume, creating a 50:50 equally mixed culture (corresponding to a resistant culture ratio of 0.5). These mixed cultures were then serially diluted by half in excess treated susceptible monoculture. Together with the monoculture aliquotes, this created a 2-fold dilution series of the resistant monoculture spanning a resistant culture ratio of 1 (pure resistant monoculture) down to 0 (pure susceptible monoculture).

Colony counting

From each monoculture of EC3 and EC5, a small aliquot was taken prior to ciprofloxacin treatment, and serially diluted in a 10-fold dilution in excess RDM growth medium, on a multiwell plate, across 8 orders of magnitude. From each dilution step of each monoculture, 3 separate 10 µl samples were taken, dropped individually and separately onto an LB agar plate, which was then incubated at 37°C overnight. Following the incubation, each plate was examined for microbial growth in any of the 3 samples per dilution step per monoculture. The greatest dilution that still provided individually countable colonies in all samples was chosen for each monoculture. For each monoculture, the mean colony count across the 3 samples at the chosen dilution step was calculated, and used together with the dilution and original sample volume of 10 µl to calculate the CFU density in the undiluted monoculture.

To analyse the differences in the means CFU density between separate monocultures of EC3 and EC5, a Tukey range test was performed between means per monoculture at a significance level of 0.05.

Theoretical prediction for mixed samples

We expected the proportion of resistant cells (R) to be a linear combination of contributions from both mixed culture components at every resistant dilution step (r), with a corresponding bound on the standard deviation:

$$\begin{gathered} R = r\,p(EC5) + (1-r)\,p(EC3) \\ \delta_R = \sqrt{\sigma^2_{p(EC5)} + \sigma^2_{p(EC3)}} \end{gathered} \tag{3.3.1}$$

where $p(X)$ is a discrete function of the isolate X that returns the ratio of resistant cell classifications upon treatment, σ_R is the standard deviation of R, and $\sigma_{p(X)}$ the standard deviation of $p(X)$. σ_R could then be used to compute the expected standard error for 3 biological replicates: $SE_R = \frac{\sigma_R}{\sqrt{3}}$.

Outlier detection in mixed samples

From a linear fit to experimental data (that is, a model of form $\mathbf{y} = \mathbf{X}\beta + \epsilon$), regular residuals r_i were computed as the difference between model predicted value $\hat{y}_i$ and the measured observable y_i:

$$r_i = y_i - \hat{y}_i \tag{3.3.2}$$

The studentized residual is then:

$$s_i = \frac{r_i}{\hat{\sigma}\sqrt{1-h_{ii}}} \tag{3.3.3}$$

where $\hat{\sigma}$ is the residual standard deviation and h_{ii} is the leverage score of the i^{th} observable $\mathbf{x_i}$ computed from the model fit:

$$h_{ii} = \mathbf{x_i}^T(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x_i} \tag{3.3.4}$$

To externally studentize the residual of the i^{th} observable, the observable itself is excluded from its variance estimation:

$$\hat{\sigma}_i^2 = \frac{1}{n-m-1} \sum_{j=1, j\neq i}^{n} r_j^2 \tag{3.3.5}$$

where n is the number of observables, and m the number of model parameters.

For a 2-tailed significance test for outlier detection at significance level of 0.05, any studentised residual that exceeded $s_i > t_{0.025}$ or $s_i < -t_{0.025}$ was excluded, where t_α is the t-statistic at significance level α and 6 (7 observables - 1) degrees of freedom.

Dose-response models

To model the ratio of cells classified as susceptible as a function treatment concentration, and also the total growth of cells, we use non-linear least squares to fit a generalised logistic function:

$$f(x) = d + \frac{a-d}{(1+(\frac{x}{c})^b)^g} \tag{3.3.6}$$

where a and d are the lower and upper asymptotes, b is the scale parameter, c is the x-coordinate of the inflexion point, and g is the asymmetry parameter. The confidence bands (CB) of the fit can be calculated directly from the covariance matrix of the fit:

$$CB(x) = \hat{y}(x) \pm t_{\alpha/2,\nu} \sqrt{\chi_\nu^2 \sum_{j,k=0}^{n} \frac{\partial f(x)}{\partial p_j} \frac{\partial f(x)}{\partial p_k} C_{jk}} \tag{3.3.7}$$

where $\hat{y}(x)$ is the best fit estimate at x, $t_{\alpha/2,\nu}$ is the upper $\alpha/2$ critical value for the t-distribution with N-n degrees of freedom, ν is the degrees of freedom, χ_ν^2 is the reduced chi-square of the fit, C is the covariance matrix, p are the best fit parameters and $f(x)$ is the generalised logistic function.

3.4 Discussion

3.4.1 Deployment of models on clinical isolates

Our first objective in this chapter was to verify our assumption about the homogeneity of the phenotype response between our training strain, and clinical isolates. By using the same treatment protocol we had used for the training strain, we indeed showed that the resistant phenotype resembles the untreated phenotype, and that susceptible phenotype of susceptible isolates resembles the susceptible phenotype of the training strain - within the limits of the data available to us.

Following this, we deployed our models on the clinical isolates, where we showed our models revealed a statistically significant difference in the proportion of susceptible cells between untreated and treated susceptible isolates - but crucially no such difference in resistant isolates. This showed the distribution of single cell classifications could indeed be used to infer information about the resistance of a sample, around a fixed treatment point.

Then, we used mixed-culture samples to attempt to quantify the limit of detection of our method. We saw the relationship between DASP reporting outcomes and the concentration of resistant cells was linear as expected, but flatter than the theoretical model conditioned on our prior experiments. In mixed cultures, DASP overestimated the proportion of resistant cells in high dilution factors, and underestimated in in low dilution factors. The likely cause of this was random or systematic error in the serial dilution. Further, we chose not to propagate the error arising from the CFU density, as the differences between the susceptible and resistant monocultures were measured to be not statistically significant. This simplified our linear model, however, likely underestimated the errors - in future iterations, it will be important to keep track of all sources of error.

Following our work with mixed samples, we proceeded to quantify the relationship between the DASP reporting outcomes, the isolate MIC and treatment concentration. Comparing to the gold-standard clinical assay measuring growth, we saw signs of concordance between the dose-response models, however, further work will be required to map out this relationship more precisely. DASP promises an potentially more detailed definition of the MIC. Currently, MIC is only established on a colony-wide level - if concordance between the cellular growth and DASP measure continues, DASP outcomes could be used to define the MIC on a cellular level - a more sensitive definition than the current culture-wide MIC.

3.4.2 Sources of errors in the classifier

Throughout all the results presented in this chapter, one of the overarching themes was the source of errors that affect the numerical results. As discussed, we saw two major sources of error.

Firstly, we saw genuine classification errors, arising either as in the classifier, or in the segmentation and propagating upwards. There are many ways of mitigating those errors - an obvious solution is to include a greater range of training data. All of our models were trained on only a single susceptible strain of *E.coli*, and applied to clinical isolates that showed meaningful genomic differences from the training strain - training on a range of susceptible strains should noticeably improve performance. Other improvements could come from refining the pipeline itself to generalise better. Our binary classifiers were trained using binary cross-entropy, which allows a probabilistic interpretation of classifier per-class scores - for each cell, the final classification as reported was simply the most probable. We speculate we could achieve lower error rates by switching to hinge loss [2] for binary labels $y_i \in \{-1, 1\} \forall i$:

$$L(y_i, \hat{y_i}) = max(0, 1 - y \cdot \hat{y}) \tag{3.4.1}$$

Unlike cross-entropy, hinge loss stops penalising correct predictions once the stipulated separation from the margin is reached. This could improve generalisability of our models by removing the incentive currently imposed by cross-entropy to learn insignificant and highly specific cellular features to continue minimizing the loss even once robust separation of classes is achieved (in effect, overtraining). In doing so, we would lose the useful probabilistic properties of cross-entropy scores (hinge loss scores have no such interpretation), although arguably we did not make use of these propertFies in our analysis. More importantly for us, hinge loss shows poorer (linear) loss scaling away from the objective than cross-entropy (logarithmic), in effect penalising gravely incorrect predictions less - the switch would likely lead to slower convergence during training. Another improvement could be a switch to

---

[2] With L2 regularisation, avoided here for brevity of argument.

ensemble methods, combining several weaker classifiers; or the addition of limited feature engineering to complement the features derived by the model. All of our models were trained on a single susceptible strain, and used the untreated phenotype as a proxy for the resistant phenotype. Clearly, this was a viable approach, and our results demonstrate our models generalise readily to clinical isolates. However, with access to clinical isolates, we could improve error rates further by training or fine-tuning the resistant class on a library of treated, resistant isolates; and the susceptible class on a library of treated, susceptible isolates. At the time of writing, we did not have a large enough library of isolate data to implement this, but this remains an active objective for future work. Beyond improving the details of training, we could also improve the details of model performance evaluation. However, unlike in the previous chapter and the training data, we did not annotate the clinical data - thus no ground truth exists against which metrics could be computed.

Secondly, we saw biological heterogeneity of cells and the cell cycle that produces cells that exhibit the visual phenotype with little apparent biological basis - this included susceptible phenotypes in untreated samples, and resistant phenotypes in susceptible, treated samples. In this case, our options were more limited, but this observation merits careful discussion. Our training data of susceptible cells was annotated by treatment condition, not by visual inspection of the phenotype - all untreated cells were used to generate the resistant class, for example. We did this deliberately to avoid human bias, and allow the models to learn naturally the heterogeneity in the samples. The visual phenotype was strongly associated with antimicrobial action, but not unique to it. As an illustrative example, while the ciprofloxacin susceptible phenotype is caused by ciprofloxacin induced DNA fragmentaton, the same effect can be generated by a range of causal factors (for example, UV radiation) in the absence of ciprofloxacin. This thus may account for cells showing the susceptible phenotype in untreated samples. Conversely, susceptible cells showing the resistant phenotypes in cultures treated could be seen potentially as metabolically inactive cells (either through their particular epigenetics, or perhaps simply because they are dead). At any rate, these ‘errors’ could not be interpreted as true errors of the methods, but rather a limitation of the phenotypes themselves, which were evidently not the injective function of antimicrobial susceptibility we had assumed them to be at the beginning of this thesis.

Much of this could however be improved with advances in the mechanistic understanding of the phenotypes themselves. All our results relied on visual assessment of the phenotypes (even if quantified by a deep-learning model). However, the visual appearance was only an observable of the underlying 3D structure of the nucleoid, its interactions with its environment, and response to perturbations caused by the antibiotics - we did not model any of the interactions. An appealing prospect could be to create a numerical biophysical model of the nucleoid and its key interactions - this could allow *in-silico* studies of antibiotic effects, and provide predictions which could then be experimentally verified by our data. As discussed in the previous chapter, we deliberately did not engage in feature engineering - however, extracting explicit features (such as nucleoid volume, for example) would allow us to quantify the phenotypes better, and build simple mathematical models based on their known 3D structures. Doing so goes beyond the scope of this thesis, but remains a valuable future research direction.

We just discussed ways of understanding and quantifying the phenotypic response on a biophysical level in single cells. However, most of our results in this chapter came from looking at populations of cells. The heterogeneity we observed and discussed in populations makes it clear that the parameters of the hypothetical mathematical models must come from some distribution (or else all cells would react identically). Again, attempting to parameterise these distributions is beyond the scope of available methods; we considered this worthwhile to mention, but promptly move onto more fruitful parts of this discussion.

3.4.3 Generalisability of findings

All results presented in this and the previous chapter relied on bacteria grown in liquid culture to a specified OD. The medium we chose was rich, which facilitated a high metabolic rate, and thus rapid growth and a quick phenotypic response. In the future, it will be important to repeat this work across a range of growth media, including especially poor media - where both the timescale and final phenotype presentation may change. Liquid culture is also not the only way to culture bacteria, solid substrates such as agar plates are also very common. Diagnostically, working with bacteria taken from agar plates also accelerates the process - if bacteria could be lifted off a plate and phenotyped directly, this would save the time taken for the culture step at the beginning of the current DASP experimental workflow. Nevertheless, this could present several challenges: solid substrates are poorer media than liquid formulations (leading to problems mentioned above), and bacteria grow in colonies and biofilms. Bacteria growing together in biofilms exhibit different patterns of gene expression than bacteria growing separately in liquid culture. Further, the high rate of cell-cell signalling in biofilms breaks one of the assumptions of our analysis - that the cells and their phenotypic responses are independent.

3.4.4 Future directions

The long term aim of this work is to develop a clinical assay that reduces as far as possible the time to result - maximum time savings could be achieved if the pathogen isolation step could be avoided entirely, and bacteria captured and phenotyped directly from a patient sample. An ideal candidate technology for this is microfluidics. As a platform, microfluidics could facilitate patient sample cleanup (for example, removing blood components from blood), bacteria isolation, enrichment and immobilisation through customised filters, and then all the treatment steps of the DASP pipeline through laminar flow over immobilised bacteria. Such treated bacteria could then be imaged and phenotyped directly on the microfluidic chip. Again however, this is not without challenges. Aside from the technical difficulties of implementing what we just discussed, from the phenotyping perpective, eliminating the culture step also removes the standardisation of samples we have worked with so far - and thus challenge our models' ability to generalise. While attractive and within technological reach at the time of writing, these developments fall outside the scope of the current work, and remain highly promising future research directions.

Although our results are promising, it is important not to overstate the impact just yet. In this chapter, we worked only with one antibiotic, and 6 clinical isolates of one species. Within this limited set, 4 isolates (EC3-EC6) shared a very similar set of *gyrA* and *parC* mutations (Figure 3.3.1), which encode DNA gyrase and topoisomerase IV respectively - the targets of ciprofloxacin (see *Introduction* for a more in-depth discussion). Two of the isolates (EC3,EC4) have only these mutations, and these were our intermediate resistance isolates (MIC=0.5 mg/L, 8 mg/L). EC5 and EC6 further featured *parE* mutations (also topoisomerase IV) in addition - and these accounted for our highly resistant isolates (MIC=72 mg/L, 108 mg/L). In future work, it will be important to expand in all of these dimensions - more genotypes, more species, more antibiotics, and seeing how well the relationships we identified in this chapter reproduce themselves.

A crucial bottleneck in addressing this, and many other points, is experimental throughput. To generate this work, we processed and imaged each sample individually, in a workflow that took several hours for each sample. To proceed further, it is clear to us that a high-throughput platform capable of multiplexed sample preparation and imaging is necessary. Several such platforms exist, allowing automated imaging in well-plate format. Unfortunately, our technique is not compatible with commercial solutions. Our bacteria are immobilised mechanically by an agarose layer from above (Figure 2.5.2), which renders it incompatible with standard well plates. By contrast, we experimented with surface immobilisation techniques (functionalising the well-plate surface chemically with "sticky" compounds such as chitosan or poly-l-lysine) - unfortunately we could not achieve the same quality of imaging as we did with agarose. Development of a custom high-throughput imaging system is a top priority in the future directions of the project.

Although it will be important to replicate our findings in more experimental conditions, another question outstanding is the generalisability of our models between conditions - for example, applying models trained on *E.coli* to another Gram-negative bacillus such as *K. pneumoniae*, or applying ciprofloxacin models to the similar phenotype of gentamycin. Although it is hard to envisage eventual deployment of models in this context, this work would help establish the robustness of our models in different contexts.

We have discussed much, we are in a good position to reflect. We have demonstrated robustness of our models to phenotypic variation of clinical isolates, validated our underlying assumptions from the previous chapter, and mapped out the relationship between treatment concentration and MIC to demonstrate that DASP can provide equivalent information to a gold standard growth assay, but in a fraction of the time. However, we lack a deeper understanding of the phenotypes. This includes understanding from the mechanistic, biophysical perspective, but also from the developmental one - the phenotypes are acquired at a single, experimentally optimised timepoint. In doing so, we are insensitive to temporal effects - allowing us to see the progression of the phenotype would not only help us understand it, but potentially further reduce the time to result, as the population could be interrogated earlier.

In the next chapter, we venture out towards the missing piece of the puzzle - the evolution of the phenotype in time as a function time and treatment conditions.

3.5 Conclusion

In this chapter, we took phenotyping models we developed in the previous chapter, and successfully applied them to clinical isolates. In doing so, we verified our previous assumptions about the homogeneity of the phenotypic response between the training lab strain and isolates, the concordance between the untreated phenotype and treated resistant phenotype, and the homogeneity of the model response when applied to isolates. Later, we calculated the limit of detection, and showed DASP is indeed sensitive to single resistant cells, as expected. Finally, we mapped out the relationship between the DASP response and the MIC and treatment concentrations of the isolates. Comparing that to the gold-standard growth assay, we demonstrated that DASP can provide equivalent information, but in a fraction of the time. Although our results were promising, an important future direction will be to repeat this work in different genotypes, species and antibiotics.

Very well; time will explain.

— Jane Austen, *Persuasion*

4

DASP Time Evolution

Contents

4.1 Phenotypes as a function of time 113
4.2 DASP as a function of time 123
4.2.1 DASP as a function of single-cell phenotype development 123
4.2.2 Aggregate DASP response as a function of time 129
4.3 Experimental Methods 130
4.4 Discussion 133
4.4.1 Deployment of models on timelapses 133
4.4.2 Future directions 134
4.5 Conclusions 136

4.1 Phenotypes as a function of time

In the previous chapter, we considered the deployment of DASP on clinical isolates of *E.coli*. We saw that both the phenotypic response of the isolates, and the DASP model response to them, were largely homogeneous. We observed that indeed the untreated phenotype was a good proxy for the resistant phenotype, showed that the DASP response is indeed sensitive to single cells, and mapped out the relationship between the response, treatment concentration and isolate MIC, which allowed us to beat gold-standard clinical assays by an order of magnitude. However, our work lacked a deeper, mechanistic and developmental understanding of both the phenotypes themselves, and the DASP response, as a function of time. This is important. From the basic science perspective, the phenotype is a readout of antimicrobial action - from the timecourse of the response, we could infer the dynamics of antimicrobial uptake and metabolism on a single-cell level, controlling for other factors that may influence the result, such as antimicrobial uptake rate. From the technique development perspective, understanding how the DASP response changes as a function of temporal phenotype development would give insight into the specific single-cell features the models are sensitive too, beyond the gradient backpropagation we carried out in Chapter 2. Further, the timelapse also promises to be a good tool for establishing appropriate treatment concentrations and durations, by allowing easy visualisation of the phenotype end state.

In the work presented so far, we had worked only with fixed cells. This offered us several advantages. Firstly, all cellular processes were paused, and chemically preserved by the fixative, which facilitated high-resolution imaging. Secondly, this allowed us to permeabilise the membranes, which allowed consistent uptake of the nucleoid stain DAPI. Thirdly, this considerably reduced the technical requirements on the imaging - stained and fixed cells could be imaged at leisure, without concern for the cytotoxic effects of laser illumination, atmospheric and temperature control of the microscope setup, or the provision of appropriate growth medium in the sample solution.

In moving to high-resolution live-cell imaging, we had to address all of the above mentioned constraints. The primary concern was staining live nucleoids, the key readout contributing to our phenotypes. DNA dyes such as DAPI that we had used are DNA-intercalators and have cytotoxic effects - creating additional cellular stress that may affect nucleoid morphology and dynamics independently of the antimicrobial-induced effects, and thus breaking one of the fundamental assumptions we took when creating DASP - that the susceptibility phenotype is a direct reporter of antimicrobial action alone. Further, in live cells, access to the nucleoid is complicated by the intact cell membrane, which has to remain intact for the cell to remain alive. This limits dye uptake (relative to permeabilised

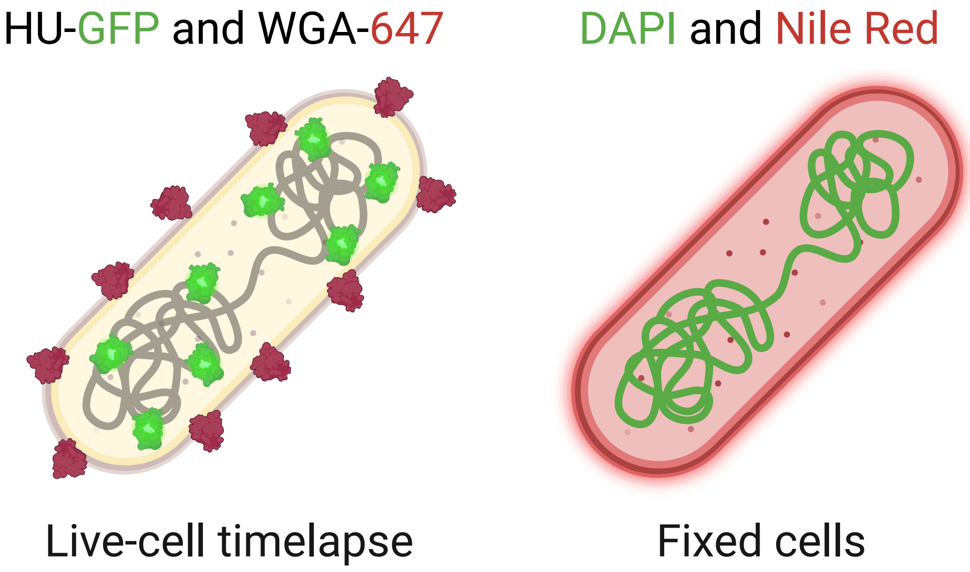


Figure 4.1.1: Comparison of staining methodology for live and fixed cell imaging. (*Left*) Live-cell staining. HU is a highly expressed and active nucleoid associated protein. When fused with GFP, it provides high-resolution staining of the nucleoid. A fluorescent derivative of wheat germ agglutenin binds sialic acid residues in the plasma membrane, and is used as a membrane stain. Staining depends entirely on fluorescent proteins. (*Right*) Fixed-cell staining. DAPI and Nile Red are small-molecule chemical dyes that are highly specific against DNA and membrane phospholipids respectively.

cells). Lastly, the variable cell state of live cells dynamically alters the dye's ability to stain its intended target - introducing heterogeneity in staining quality that was not present in fixed cells. The membrane stain presents similar concerns, although to a lesser extent, as the membrane is more available to the outside environment. Unfortunately, the choice of nucleoid dyes that meet all 3 of the aforementioned criteria (non-perturbative, cell-permeable, insensitive to cell state) is highly limited. Literature [114] has revealed a small selection of "dead cell" stains (used primarily to detect dead cells with compromised membranes) that are sufficiently non-perturbative to be used for this task, especially the SYTOX ™ family of dyes. Unfortunately, despite a range of optimization experiments, we could not obtain a sufficient signal-to-noise ratio with them to proceed. We attributed this to their poor permeability.

To proceed, we resulted to using a genetically engineered system, using *E.coli* mutants that expressed a GFP-tagged, histone-like protein HU (see *Experimental* *methods* section 4.3 for details). HU is a highly-expressed nucleoid-associated protein which plays a role in DNA maintenance and repair. In previous work we used DAPI; HU co-localises with DAPI [175]. Further, its concentration is maintained at a constant level throughout the cell cycle [176]. Thus, imaging the HU-GFP signal provides by proxy a consistent nucleoid stain comparable to that used in our fixed cell protocol (and on which our models were trained). Further, in contrast to external, chemical dyes, HU is continuously produced in live cells, reducing concerns over photobleaching and increasing suitability for long-term imaging. To complement it, we used Wheat Germ Agglutinin functionalised with Alex Fluor ™ 647 (WGA-647) as the membrane stain. WGA-647 binds to N-Acetylglucosamine in the bacterial cell well, providing a membrane readout. It shows strong preference for Gram-positive cell walls [177], however, we were able to work with it in the Gram-negative *E.coli* by increasing the incubation concentration. Together, the combination of HU-GFP and WGA-647 provided a good analogue to our fixed-cell staining protocol, that was compatible with live cell imaging (Fig 4.1.1).

The cells were imaged on agarose pads again, in a reservoir containing growth medium and WGA-647 - this allowed continuous staining and replenishment of the fluorescent signal, much like the genetic incorporation of GFP into HU. Together, this created a system suitable for long-term live-cell imaging that closely resembled the fixed cell protocols we had been working with until then. The HU-GFP strain we worked with was derived from the same K-12 original lab strain as our training strain MG1655, and showed similar susceptibility and phenotype dynamics.

To be a suitable system for our use, the system and experimental setup had to be non-cytotoxic. To test that, we cultured cells without any antibiotic treatment (Figure 4.1.2, where we observed that nucleoid replication and cell division occurs approximately once every 20 minutes, a close analogue to the expected replication time in RDM medium (Figure 4.1.3).

With that in mind, we proceeded and treated the cells with a range of antibiotics, at 10x EUCAST breakpoint concentration (half of the values previously reported in Figure 2.5.1). This concentration represented a departure from our previous

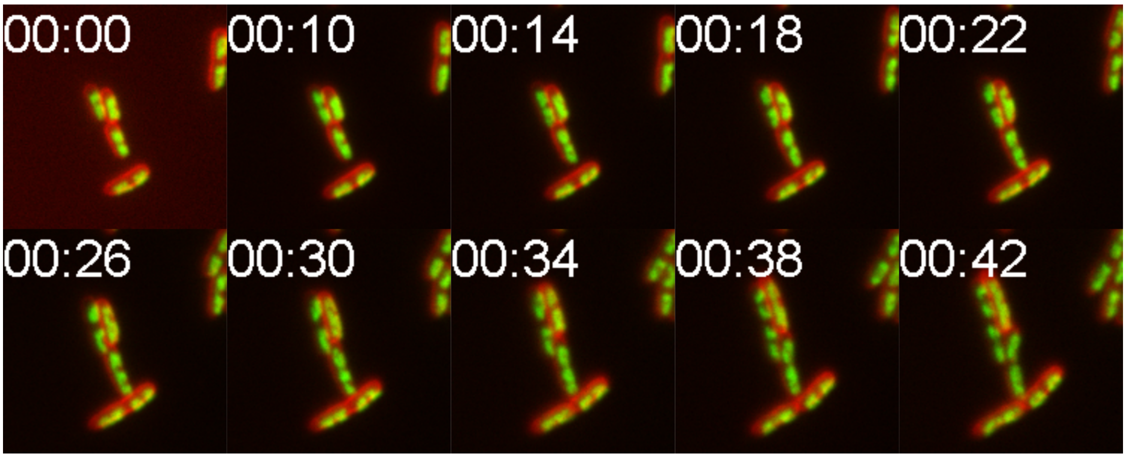


Figure 4.1.2: Timelapse of untreated HU-GFP reveals expected cell division and nucleoid replication. Timelapse fluorescence imaging of cell division in HU-GFP *E.coli*. GFP signals in green, WGA-647 signals in red.

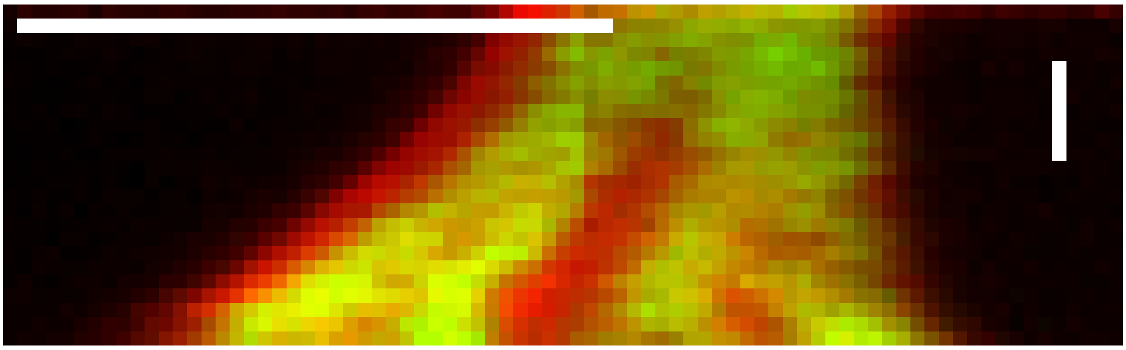


Figure 4.1.3: Kymograph of a single, representative HU-GFP *E.coli* cell with no antibiotic treatment, undergoing cell division.. Kymograph taken along the long axis of the cell and across all timepoints. Horizontal bar represents 5 µm, vertical bar represents 20 minutes. GFP signals in green, WGA-647 signals in red. Nucleoid replication and cell division are both clearly visible.

studies, which were carried out at 20x EUCAST concentration - this was due to the limited solubility of some of the antibiotics in the agarose pads that we needed for live cell imaging. When treating with ciprofloxacin (Figure 4.1.4), we observed rapid nucleoid condensation towards the middle of the cell, and a pause of cell division (Figure 4.1.5). Condensation occurred rapidly, within 15 minutes of the start of the experiment.

Similarly, treatment with gentamicin (Figure 4.1.6) also produced nucleoid compaction, although the individual nucleoids did not merge into one at the cell centre, which we had observed with ciprofloxacin. Again, compaction occurred

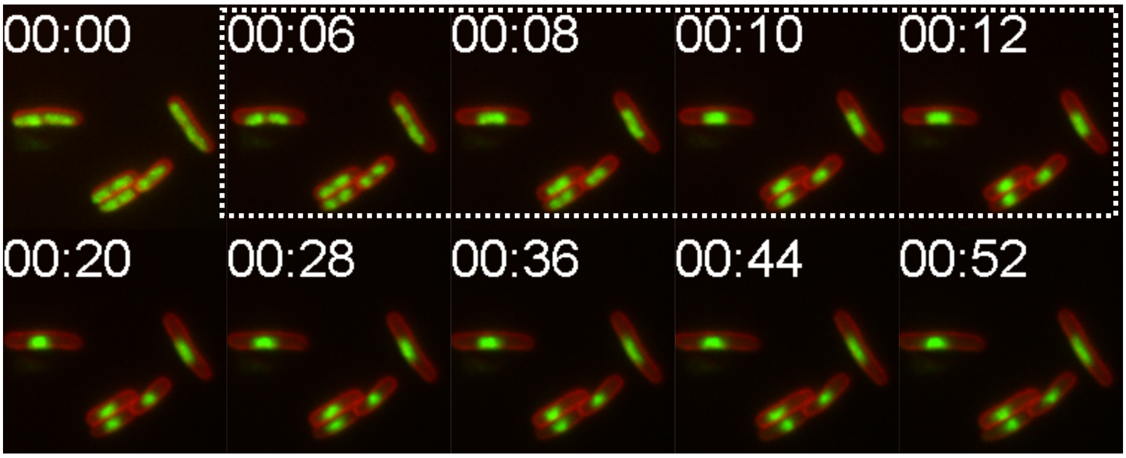


Figure 4.1.4: Timelapse of ciprofloxacin-treated HU-GFP reveals reveals a halt of cell division and condensation of both nucleoids towards cell centre. Timelapse fluorescence imaging of ciprofloxacin-induced inhibition of gyrase and topoisomerase-IV leads to nucleoid compaction in *E.coli.* GFP signals in green, WGA-647 signals in red. White box indicates most dynamic region.

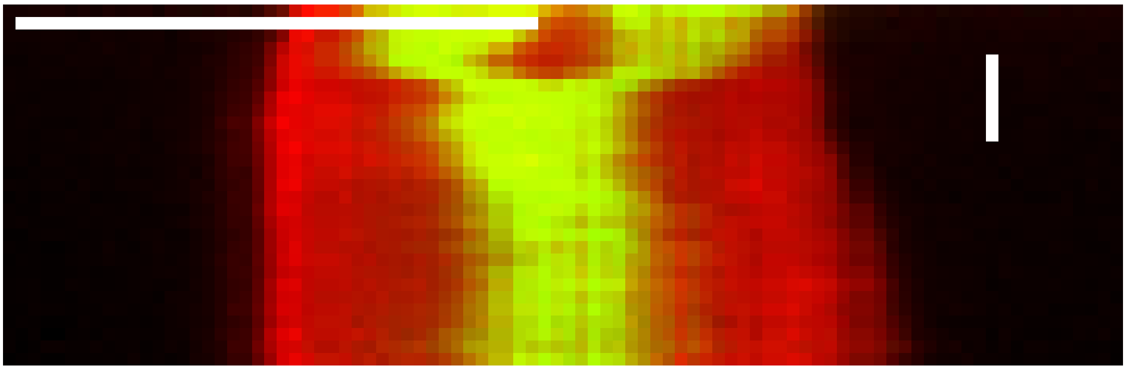


Figure 4.1.5: Kymograph of a single, representative HU-GFP *E.coli* cell under ciprofloxacin treatment.. Kymograph taken along the long axis of the cell and across all timepoints. Horizontal bar represents 5 µm, vertical bar represents 20 minutes. GFP signals in green, WGA-647 signals in red. Nucleoid condensation towards cell centre clearly visible. No cell division.

rapidly, within 25 minutes. Unlike ciprofloxacin however, in this case the compaction does not appear to be the final state of the phenotype. After 50 minutes, some of the nucleoids developed a "doughnut-shaped" appearance, with a distinct hole in the middle.

As before, treatment with rifampicin (Figure 4.1.8 led to nucleoid decompaction (Figure 4.1.9, occuring rapidly within minutes of treatment start. In this timelapse, we observed some misalignment between the channels that our alignment algorithm could not compensate for. This was likely a limitation of the microscope - since

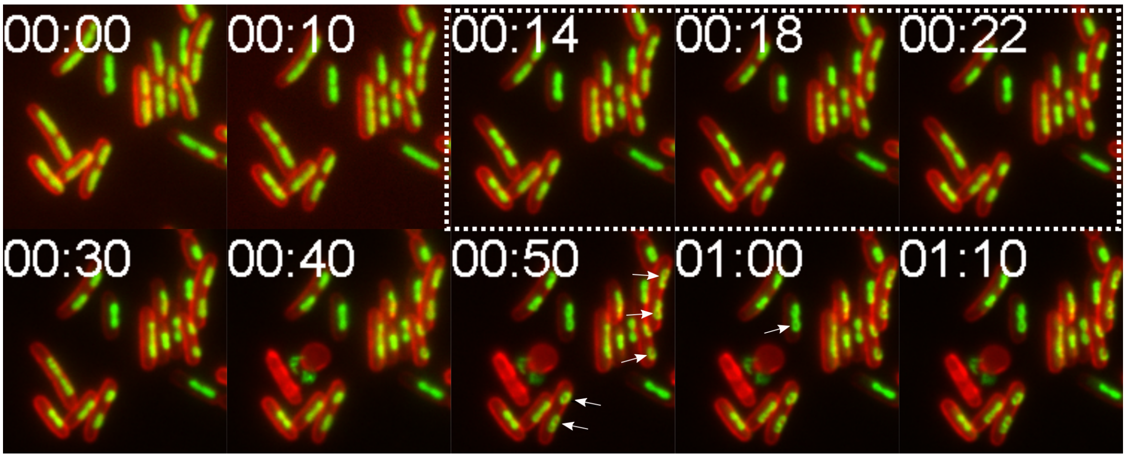


Figure 4.1.6: Timelapse of gentamicin-treated HU-GFP reveals reveals a halt of cell division and condensation of both nucleoids, but without merging. Timelapse fluorescence imaging of gentamicin-induced inhibition of ribosomes leads to compaction of individual nucleoids, but without a merger. *E.coli.* GFP signals in green, WGA-647 signals in red. White box indicates most dynamic region. White arrows indicate a unique 'doughnut-shaped' appearance some condensed nucleoids develop over a longer timescale.

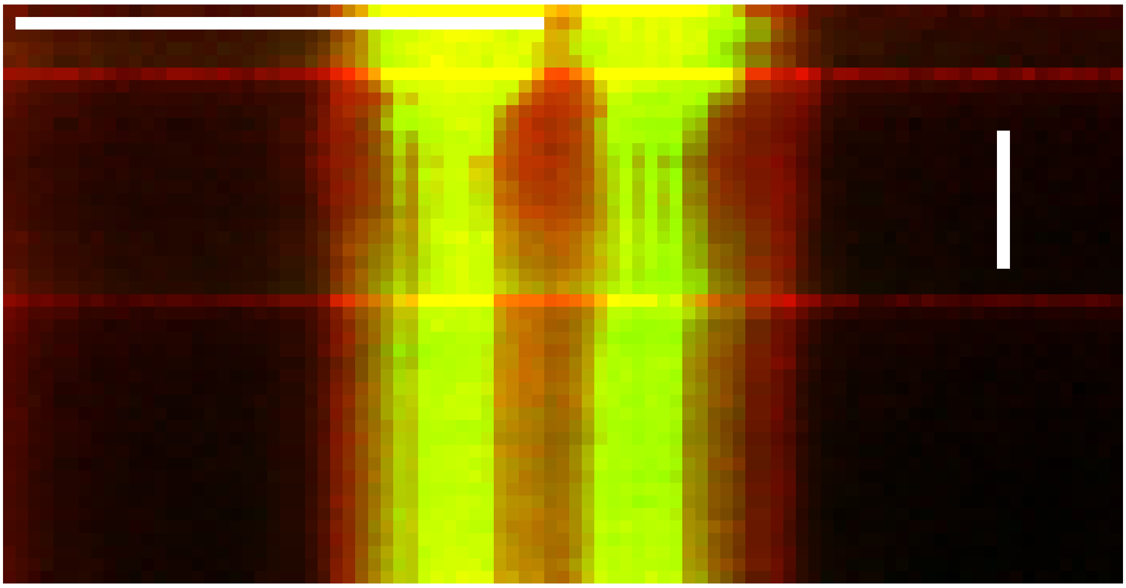


Figure 4.1.7: Kymograph of a single, representative HU-GFP *E.coli* cell under gentamicin treatment.. Kymograph taken along the long axis of the cell and across all timepoints. Horizontal bar represents 5 µm, vertical bar represents 20 minutes. GFP signals in green, WGA-647 signals in red. Condensation of individual nucleoids visible, but no merger. No cell division.

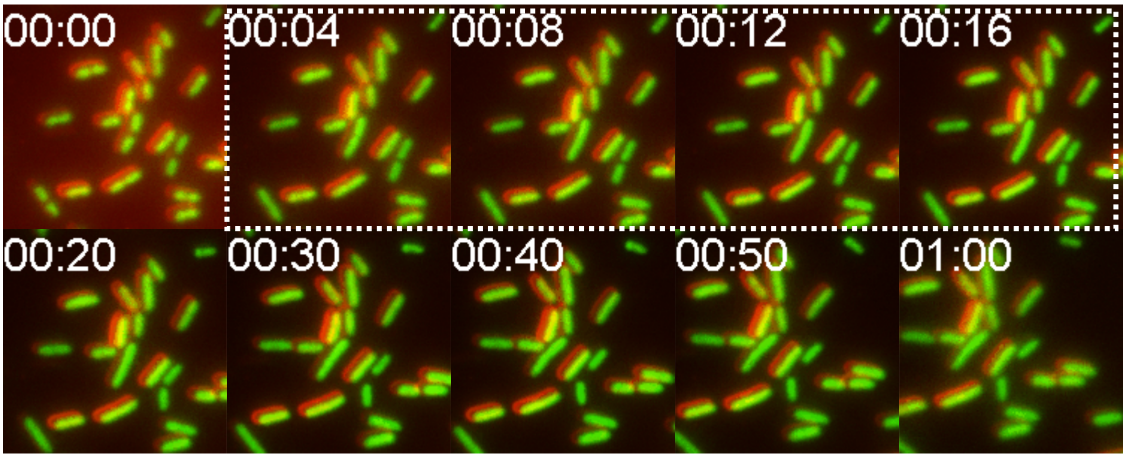


Figure 4.1.8: Timelapse of rifampicin-treated HU-GFP reveals a halt of cell division and decompaction of both nucleoids. Timelapse fluorescence imaging of rifampcin-induced inhibition of RNA-polymerase leads to decompaction of individual nucleoids. *E.coli.* GFP signals in green, WGA-647 signals in red. White box indicates most dynamic region.

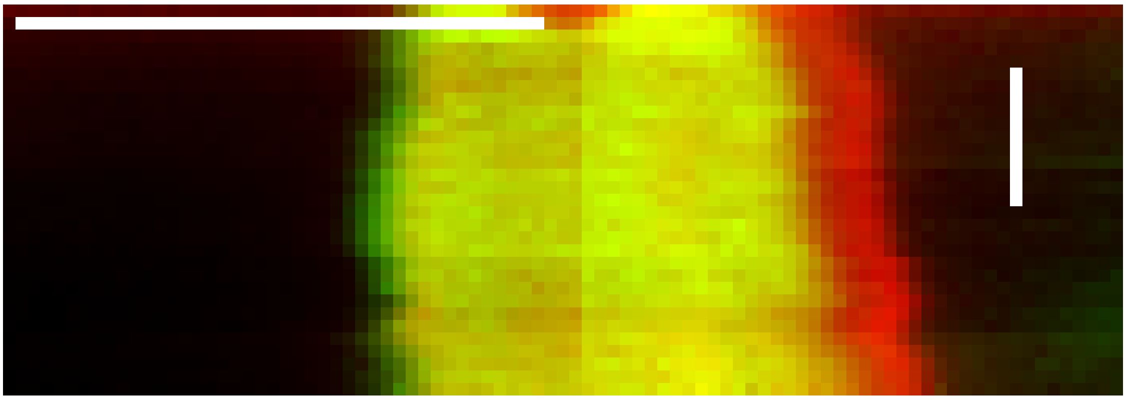


Figure 4.1.9: Kymograph of a single, representative HU-GFP *E.coli* cell under rifampicin treatment.. Kymograph taken along the long axis of the cell and across all timepoints. Horizontal bar represents 5 µm, vertical bar represents 20 minutes. GFP signals in green, WGA-647 signals in red. Decompaction of individual nucleoids visible. No cell division.

the timelapse was acquired continuously without stage movement between frames, it is unlikely to be caused by drift, and is instead an optical aberration.

Examining the effect of co-amoxiclav (Figure 4.1.10) provided us with by far the most excitement. Unlike in fixed cells (Figure 2.2.1 E), we did not observe the subtle loss of nucleoid heterogeneity - likely because the cells were live, and cell-processes masked any such subtle change. On the other hand, we did observe a loss of membrane integrity, leading to a loss of the nucleoid (and wider cell shape), and thus

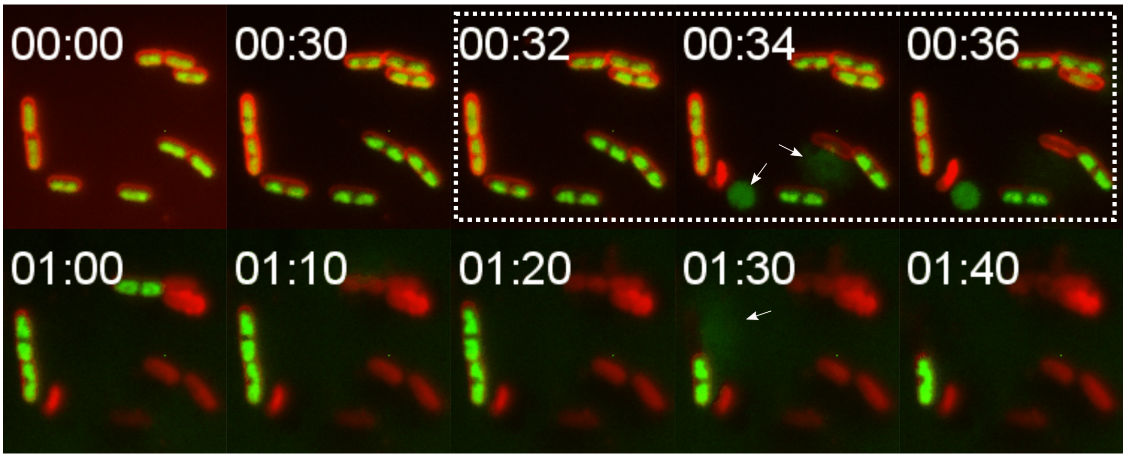


Figure 4.1.10: Timelapse of co-amoxiclav treated HU-GFP reveals reveals loss of membrane integrity. Timelapse fluorescence imaging of co-amoxiclav attack on live cells, showing loss of membrane integrity and cell death. *E.coli.* GFP signals in green, WGA-647 signals in red. White box indicates most dynamic region. White arrows show individual cell-death events.

cell death. Examining the fluorescence intensity of both the nucleoid and membrane and comparing to untreated cells (Fig 4.1.11), we see a gradual loss of membrane intensity in the buildup to the lysis event. This could potentially be a readout of membrane integrity - as we discussed in the introduction, blocking of peptidoglycan synthesis can induce a breakdown of existing cell wall, and thus a reduction in WGA-647 binding sites. Further work and controls for photobleaching will be required to pursue this hypothesis. Unlike the previous antibiotics, the timescale of this processes was much more variable, with first events observed 30 minutes after the start of the treatment, and the process continuing for at least another 60 minutes.

Putting together what we saw in this section, we can draw some conclusions. Our system appears to be non-perturbative and suitable for long-term imaging. With all antibiotics except co-amoxiclav, we saw the phenotype development was gradual - progressing from the unperturbed, untreated appearance, to some final state in a fashion that was linear with time. With co-amoxiclav, this was not the case - we did not observe the subtle loss of nucleoid structure we had seen in fixed cells, but instead sudden loss of membrane integrity, reading to rapid cell death. Unlike other phenotypes, this was not a linear process - the individual, single-cell timescale for this varied for over 1 hour (several cell cycles). Throughout

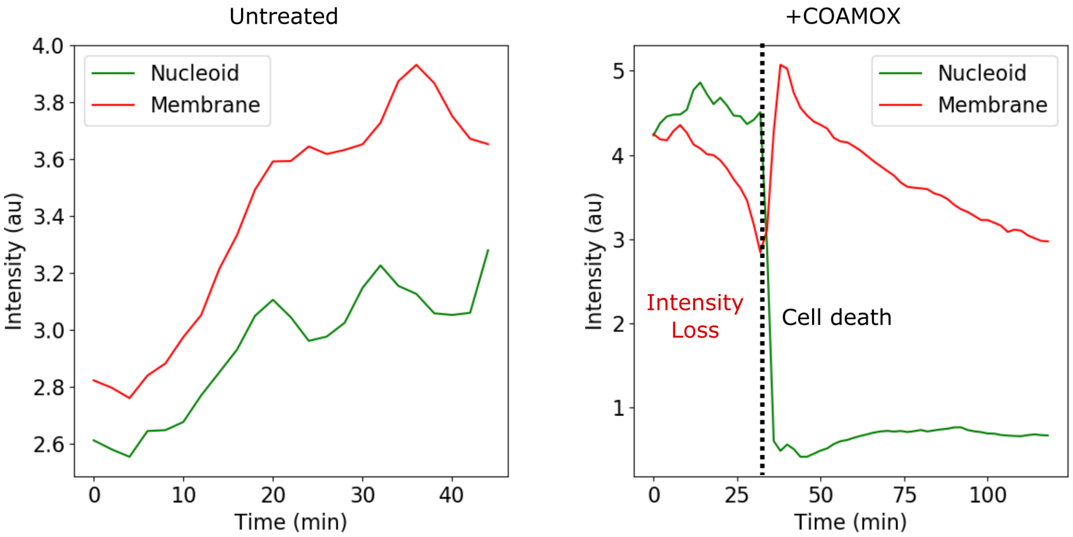


Figure 4.1.11: Co-amoxiclav treatment leads to a loss of membrane fluorescence intensity over time until catastrophic cell lysis occurs. (*Left*) Membrane and nucleoid fluorescence intensity changes in an example untreated HU-GFP cell, integrated over the cell area. Membrane intensity increases as the cell cycle progresses and the cell divides into two daughter cells. Nucleoid intensity oscillates during the replication cycle. (*Right*) Membrane and nucleoid fluorescence intensity changes in an example co-amoxiclav treated HU-GFP cell, integrated over the cell area. Membrane intensity gradually decreases until a cell lysis event, which is followed by a sharp increae in intensity, and gradual drop off due to diffusion and bleaching. Nucloid intensity is lost sharply at the lysis event.

the experiment series, we calculated the average amount of time elapsing between sample mounting (coinciding with the start of the antibiotic exposure,the true $t = 0$ of the timelapse) was around 5 minutes. This time was largely taken by the need to manually focus on the sample, and calibrate the autofocus. As a result, the beginning of the timelapse as we report it is not the true zero - as a result it is highly likely the development is accelerated, relative to what we saw with fixed cells.

Having seen the time evolution of our phenotypes, our next logical step was to run the DASP pipeline on them.

4.2 DASP as a function of time

4.2.1 DASP as a function of single-cell phenotype development

Like before, we used the same selection of models to run our analysis on timelapse data - the ensemble Mask-RCNN segmenter in the 1st stage, and the binary classifiers uesd to evaluate holdout experiments in the 2nd stage. To provide an internal reference point, each of the holdout binary classifiers was evaluated on both an appropriate treated FoV, and on an untreated FoV. The DASP pipeline was evaluated independently on each timepoint of each FoV. As a result, at each timepoint, the models made their decisions based solely on the phenotype presentation at that point. It is worth pointing out that testing our models in this fashion represents a significant domain-shift - not only the live nucleoid is less well-resolved than the static fixed one, but also there are fundamental differences in staining between the training and testing data.

In the case of ciprofloxacin, we observed a similarly strong response to what we had seen before. Observing first the ciprofloxacin binary classifier acting on an untreated FoV (Figure 4.2.1 A), we see as expected the majority of cells are classified as untreated/resistant at all time points. As the timelapse progresses, we do see a gradual decline in segmenter performance, with cells in clusters missing. As discussed in previous chapters, this is likely due to one of the fundamental limitations of Mask-RCNN, where the non-maximum supprression filter at the end incorrectly prunes densely packed detections in cell clusters, considering them artefacts of segmentation. Considering the ciprofloxacin-treated FoV (Figure 4.2.1 B), we see the classifier correctly classifying cells as susceptible. This occurs a much earlier time point than the full compaction of 15 minutes, instead occurring at the beginning of the process (around 5 minute mark). From this study, we concluded the ciprofloxacin classifier is sensitive to the first signs of nucleoid compaction, and remains sensitive to further compaction too.

Considering the gentamicin classifier in the same setup as above, we observed a higher noise floor in the untreated FoV (Figure 4.2.2 A), with more cells classifier as

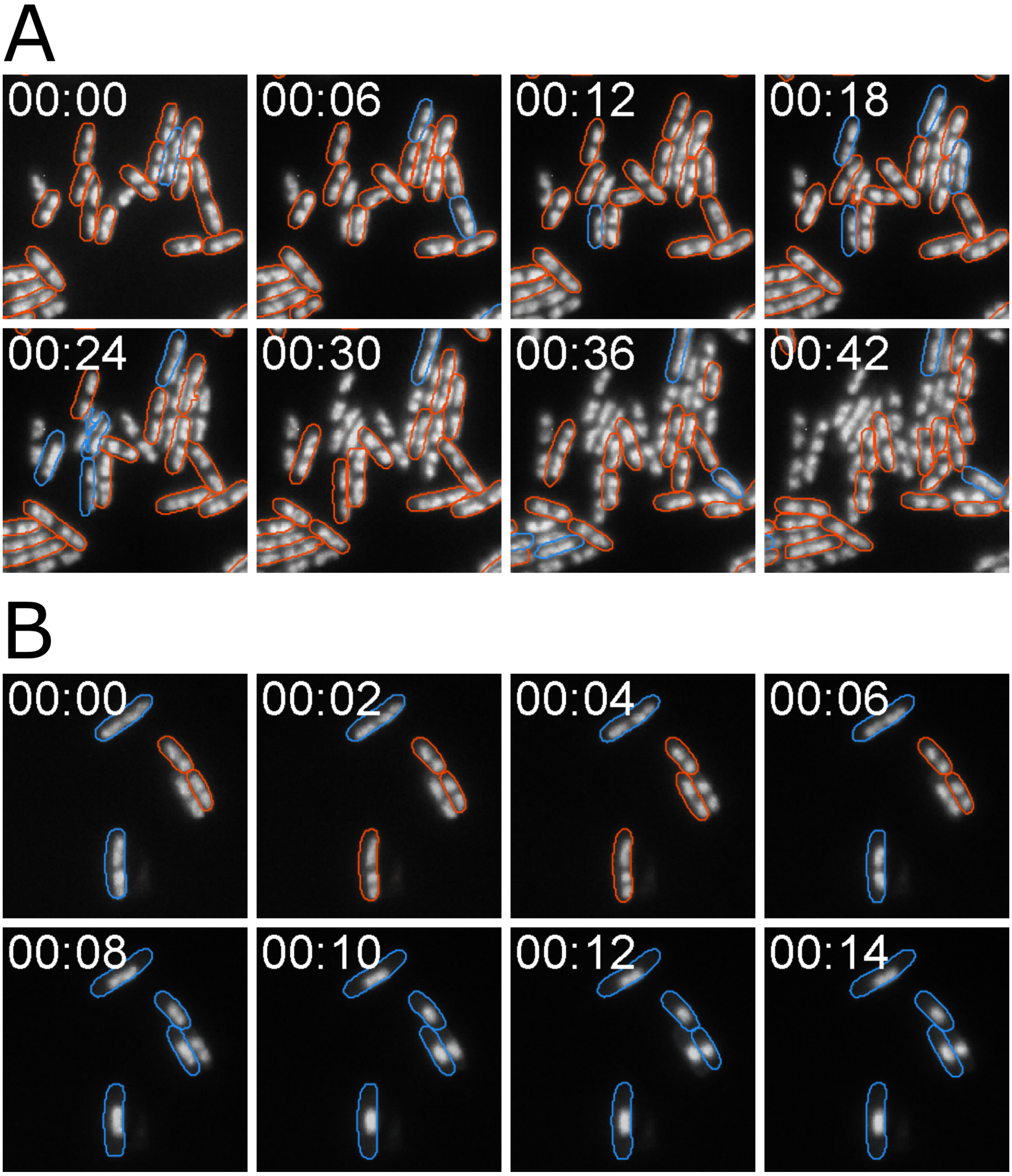


Figure 4.2.1: Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, ciprofloxacin-susceptible binary classifier. A. Phenotype detections as a function of time in an untreated FoV. Red overlays indicate untreated/resistant classifications. Blue overlays indicate ciprofloxacin-susceptible detections.
B. As A, but in a ciprofloxacin-treated FoV.

gentamicin-susceptible even in the absence of gentamicin - although the proportion of classifications did not appear to change with time. In the gentamicin-treated FOV (Figure 4.2.2 B), the proportion of gentamicin-susceptible detections did indeed change, and once again the models appeared to be sensitive to nucleoid compaction (as was hoped). Again, this seemed to occur on a faster timescale than expected, with consistent decisions made 8 minutes into the treatment course. Relative to other FoVs in this section, the gentamicin FoV is relatively denser. This likely contributes to a higher error rate - as we saw in Chapter 2, our segmenter struggles with dense FoVs. This then feeds inaccurate segmentations to the classifier.

Moving onto the rifampicin classifier and deploying it on the untreated FoV (Figure 4.2.3 A), the error rate is again high; by contrast, in the treated FOV (Figure 4.2.3 A), almost all cells are correctly classified as susceptible from almost the start of the timelapse. Interestingly, we noticed our algorithms seemingly coped well with the channel misalignment and this did not affect the analysis - this was likely because one of the data augmentation strategies we used was introducing deliberate channel misalignment to increase robustness against such a scenario.

Once again, the co-amoxiclav treatment brought us by far the most excitement. Observing the untreated FoV (Figure 4.1.6 A), we once again see a high amount of noise, with many spurious misclassification. However, in the co-amoxiclav treated FoV (Figure 4.2.4 B), we saw correct and consistent decision making, with a gradual progression towards susceptible detections as time went on. Interestingly, the models also showed themselves sensitive to cells that had burst after their membranes had been compromised - despite the lack of presence of such cells in the training data (such empty cell envelopes would not have survived the centrifugation and washing steps of the fixed cell protocol, and were not observed experimentally). Unlike the previous cases we examined, this time the models did not seem more sensitive than expected from the phenotypic change itself - cells were not classified as susceptible considerably ahead of their rupture. This was likely because this phenotype development was not as linear in time as the others, like we discussed.

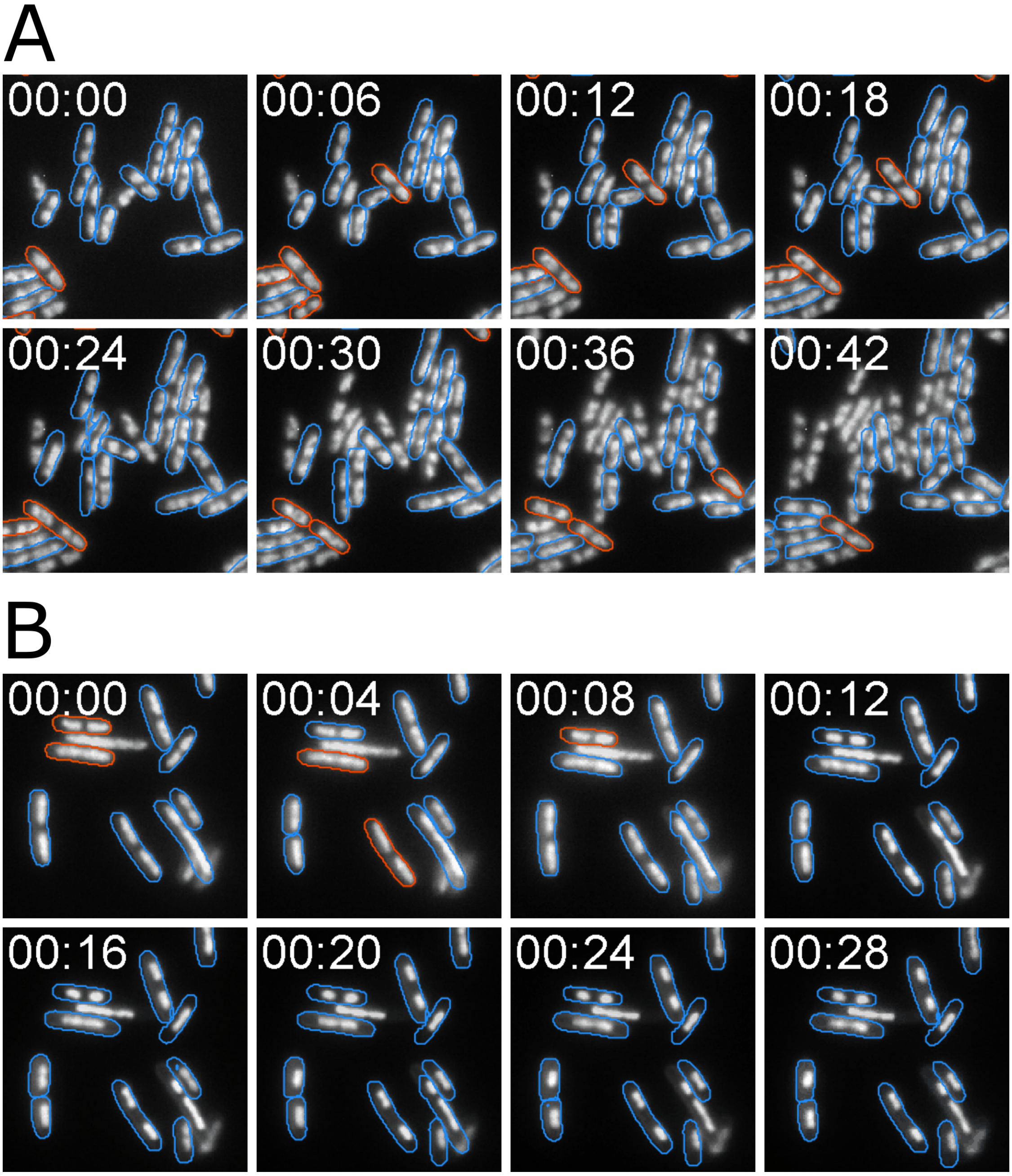


Figure 4.2.2: Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, gentamicin-susceptible binary classifier. **A**. Phenotype detections as a function of time in an untreated FoV. Red overlays indicate untreated/resistant classifications. Blue overlays indicate gentamicin-susceptible detections.
B. As A, but in a gentamicin-treated FoV.

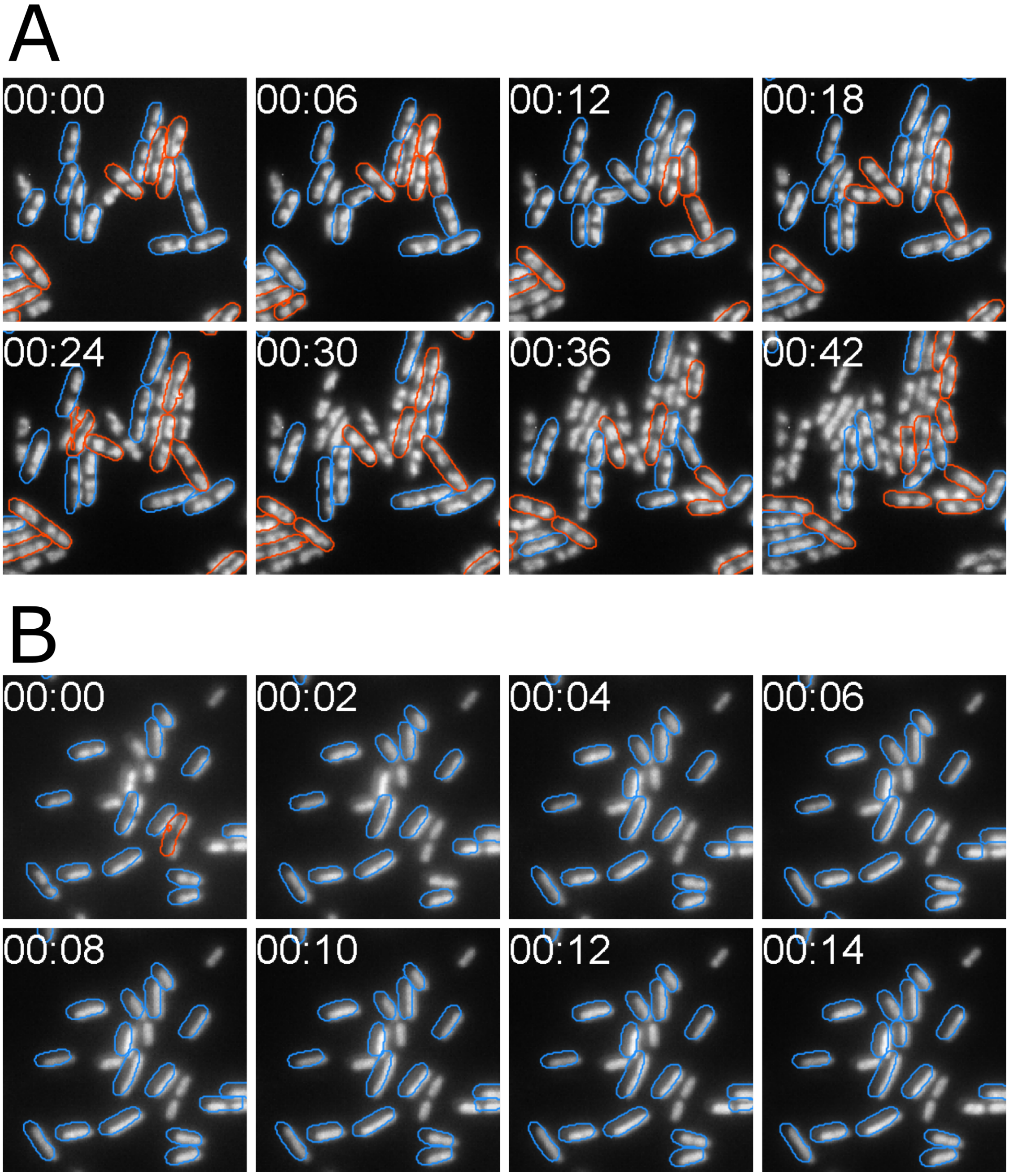


Figure 4.2.3: Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, rifampicin-susceptible binary classifier. A. Phenotype detections as a function of time in an untreated FoV. Red overlays indicate untreated/resistant classifications. Blue overlays indicate rifampicin-susceptible detections.
B. As A, but in a rifampicin-treated FoV.

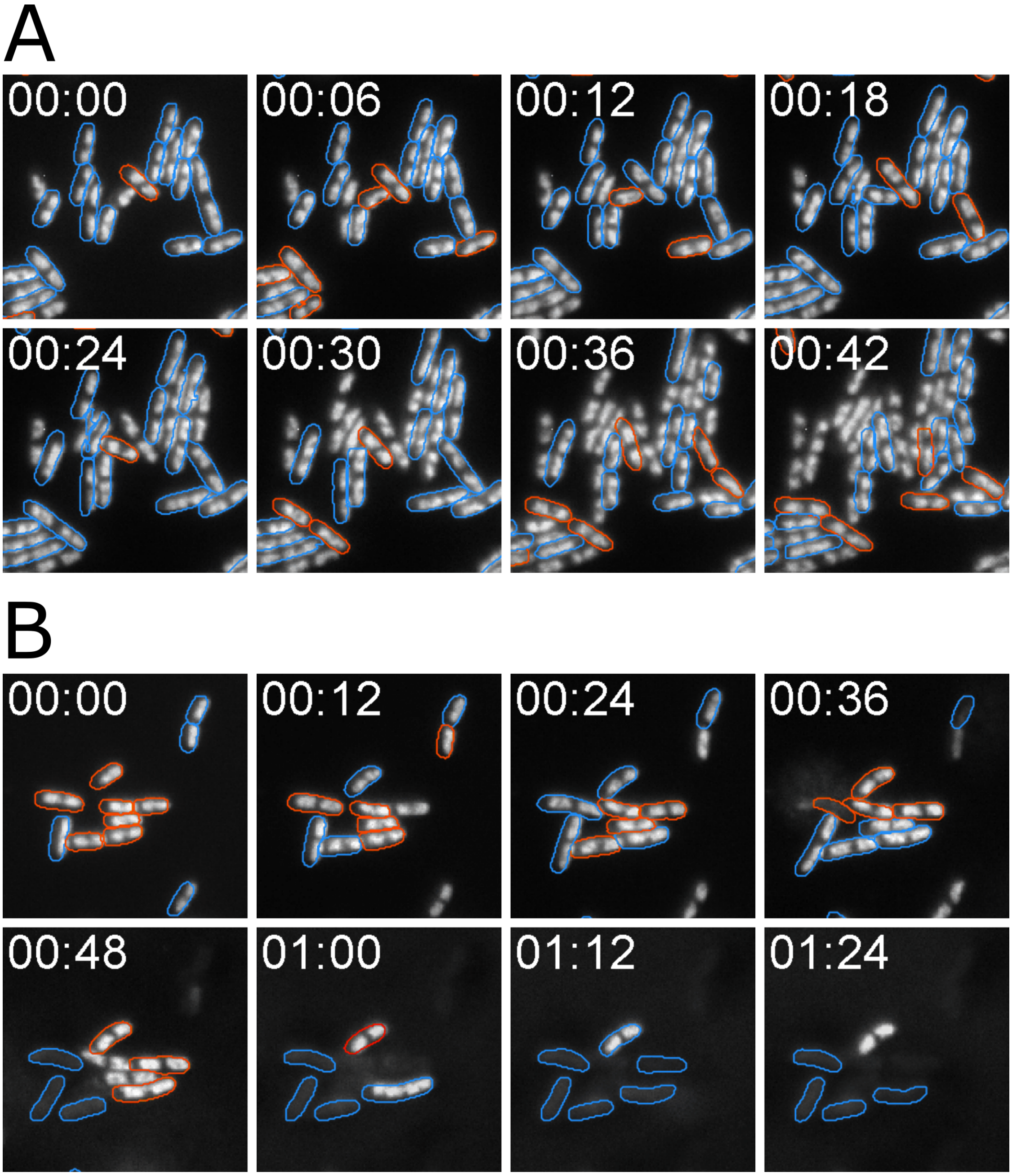


Figure 4.2.4: Phenotype detections as a function of time in a single FoV, in live HU-GFP *E.coli*, co-amoxiclav susceptible binary classifier. A. Phenotype detections as a function of time in an untreated FoV. Red overlays indicate untreated/resistant classifications. Blue overlays indicate co-amoxiclav susceptible detections.
B. As A, but in a co-amoxiclav treated FoV.

We are in good position to summarise. In this section we deployed models trained on fixed cells stained with DAPI and Nile Red, and deployed them on live-cell timelapses of HU-GFP with supplementary WGA-647. This lead to a high noise rate in untreated cells, with many cells misclassified as susceptible despite a lack of antibiotic. This was likely due to a significant domain shift relative to training data. As the untreated cells divided and the FoV became more dense, the segmentation performance decreased - although the approximate proportion of classifications did not change. This effect could be offset in time by lowering the density of cells at the start of the timelapse. In treated FoVs, the responses were more consistent. While in some cases, we did see a high error rate in the beginning of the treated timelapse (thus suggesting cells were susceptible before any reasonable time had elapsed), this could be explained by the short delay between the true start of the incubation and the timelapse start - some cells had indeed had time to react by then. In all cases, the trend in the treated FoVs was that the proportion of cells classified as susceptible increased with time, as expected. When we worked with fixed cells, we discovered that the diagnostically sensitive metric was the distribution of single-cell classifications in the population. Observing a similar pattern, we were led to hypothesise that perhaps the same wisdom will hold in live cells, and we should consider the overall distribution of cells.

4.2.2 Aggregate DASP response as a function of time

Aggregating single-cell responses to look at the population response, we observed generalisations of the patterns that had already became apparent from looking at the individual FoVs. With the ciprofloxacin classifier (Figure 4.2.5), we saw the model was highly specific to both untreated and treated samples, returning 100% susceptible cells in the treated sample after 10 minutes of treatment, and classifying over 80% of cells as untreated at all time points. Unfortunately, our gentamicin classifier (Figure 4.2.6) proved to be less specific to untreated cells - while it correctly classified over 95% of cells in the treated condition as susceptible, the error rate in the untreated control hovered around 80%, and rose with time. We had better fortunes

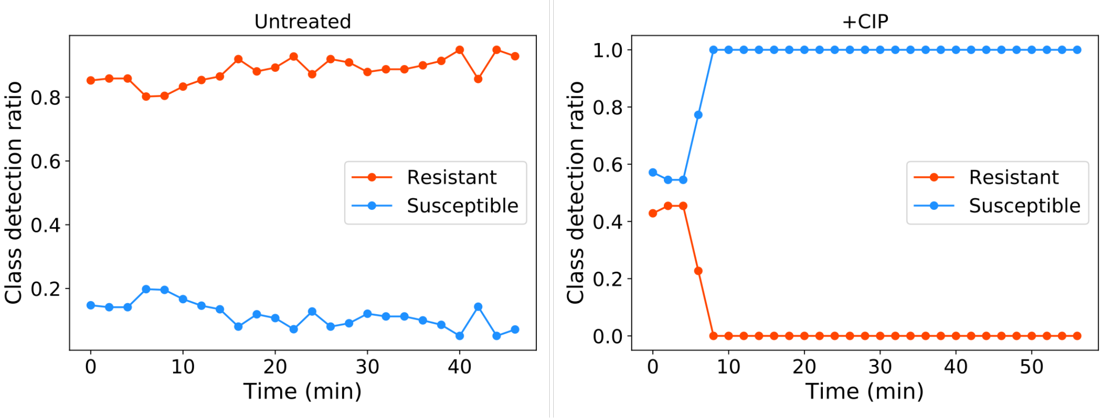


Figure 4.2.5: Population response to ciprofloxacin treatment in live HU-GFP *E.coli*. (*Left*) Phenotyping an untreated control. Majority of cells classified correctly as untreated/resistant across time. (*Right*) Phenotyping a ciprofloxacin-treated sample. Proportion of susceptible cells rises sharply to 100% within 10 minutes of treatment start.

with the rifampicin classifier (Figure 4.2.7) - again correctly classifying over 95% of cells as susceptible in the treated condition, and showing a decreasing error rate in the untreated control. Our results from co-amoxiclav (Figure 4.2.8) were again the most exciting. The classifier correctly predicted an increasing proportion of cells as susceptible in the treated condition, reaching a saturation of 95% after around 1 hour - which also qualitatively corresponded to the saturation of cell lysis in the sample (Figure 4.1.10). In the untreated control, the error rate increased and later fell on the timescale of around 30 minutes, which also qualitatively corresponded to the time for cell division in the medium and on the microscope (Figure 4.1.2).

4.3 Experimental Methods

Strain engineering

The NK9386 HU-GFP strain was was derived from CM735 *E.coli*, and expressed *hupA::GFP*, constructed as described in Marceau *et al* [178]. The strain was acquired as a gift from the laboratory of Olivier Espéli, Collège de France.

Sample preparation

Following the previously described protocols, individual colonies of the HU-GFP MG1655 strain were cultured overnight in LB liquid culture at 37°C, then diluted

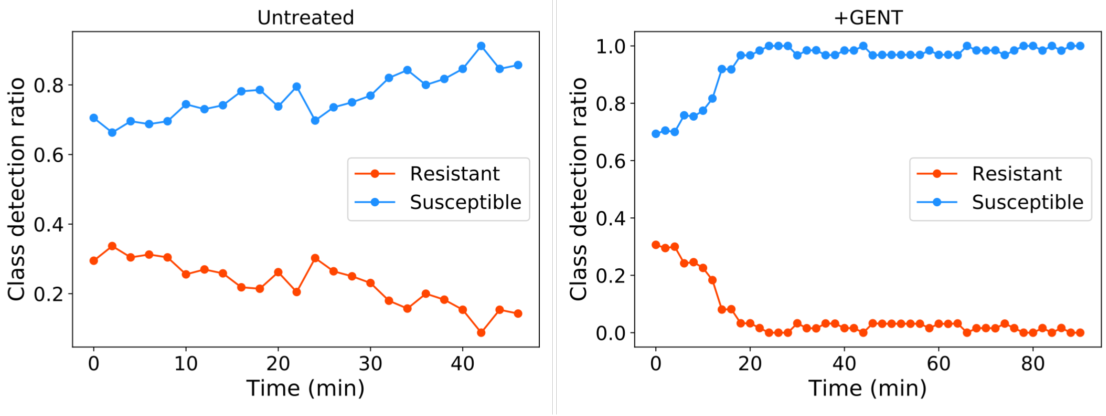


Figure 4.2.6: Population response to gentamicin treatment in live HU-GFP *E.coli.* (*Left*) Phenotyping an untreated control. Proportion of cells classified incorrectly as susceptible increases with time. (*Right*) Phenotyping a gentamicin-treated sample. Proportion of gentamicin susceptible cells rises sharply to 100% within 20 minutes of treatment start.

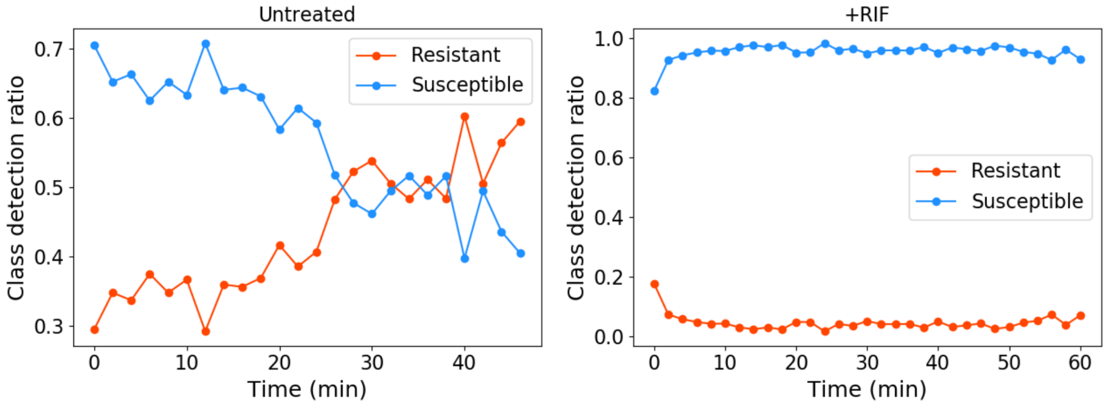


Figure 4.2.7: Population response to rifampicin treatment in live HU-GFP *E.coli.* (*Left*) Phenotyping an untreated control. Proportion of cells classified incorrectly as susceptible decreases with time. (*Right*) Phenotyping a rifampicin-treated sample. Almost all cells show the phenotype from an early timepoint.

1:100 by volume in 5 ml of RDM medium and grown until an OD of 0.2. Meanwhile, 2 further solutions were prepared. Firstly, a staining solution containing a 1x RDM medium and 20 µg/L wheat germ agglutinin functionalised with Alexa Fluor™ 647 (WGA-647, catalogue number W32466, ThermoFisher Scientific). Secondly, a treatment solution, containing again the medium and WGA-647, both at double concentration, but also with the addition of one of gentamicin, ciprofloxacin, rifampicin or co-amoxiclav at 20x EUCAST *E.coli* breakpoint concentration. Thirdly, an incubation solution, containing as above the medium and antibiotic, but no

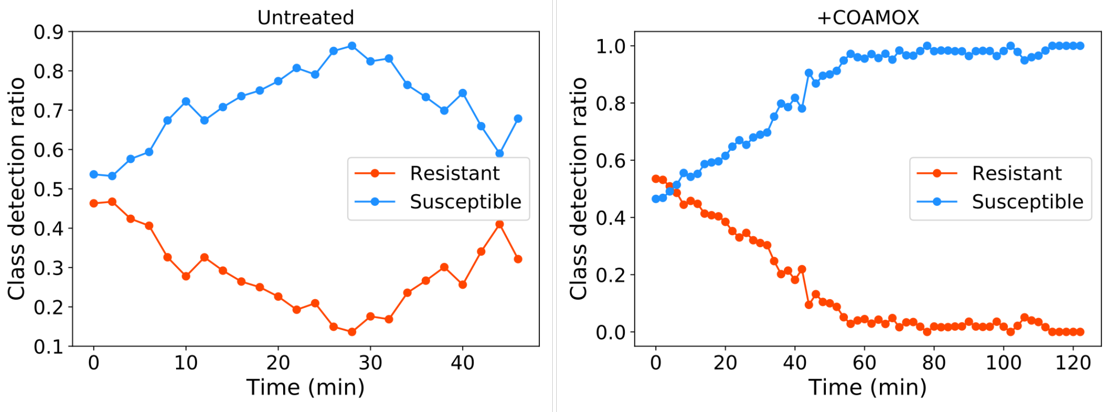


Figure 4.2.8: Population response to co-amoxiclav treatment in live HU-GFP *E.coli.* (*Left*) Phenotyping an untreated control. Proportion of cells classified incorrectly as susceptible rises and falls with time. (*Right*) Phenotyping a co-amoxiclav treated sample. Proportion of cells classified as susceptible rises steadily with time and saturates, even once all cells had undergone lysis.

WGA-647. The incubation solution was used to create agarose pads by 1:1 mixing with 2% high-purity agarose melted in PBS (thus creating the same pads as in previous work, except with the addition of RDM medium and antibiotic).

Once at the correct OD, cell cultures were spun down by centrifugation at 4500 RCF for 3 minutes and resuspended in the staining solution, and incubated in the dark at 37°C for 15 minutes. The cells were spun down again down to 20µL of PBS, and mixed 1:1 with the treatment solution. Immediately after mixing, an aliquot of this sample was placed on an agarose pad and imaged.

Timelapse microscopy

Prior to the experiment, the microscope was turned on, the empty agarose pad was placed on the stage and the stage heating control was set to 37°C for 1-2 hours, until the set temperature was reached. The prepared sample was placed on the agarose pad, and immediately afterwards the microscope was focused on the sample in the selected FoV, and the autofocus engaged permanently to keep the cells in focus. The GFP and Alexa Fluor™ signals were imaged using the 473 nm and 640 nm laser lines of the microscope, at 0.1 kW/m^2 and 0.5 kW/m^2 respectively. At each timepoint, a single frame with 30ms exposure was acquired, separately for the GFP and Alexa Fluor™ signals. The imaging was repeated every 2 minutes, and the autofocus was kept on between timepoints.

4.4 Discussion

4.4.1 Deployment of models on timelapses

In this chapter, we considered the development and time evolution of the phenotypes we have been working with. By running our DASP phenotyping platform frame-by-frame on the resulting timelapses, this allowed us to investigate the sensitivity and response of our models as a function of phenotype development. To do this, we used a genetically-modified strain of *E.coli* which carried a GFP tag on the nucleoid-associated protein HU - in effect providing regenerating, nucleoid staining analogous with the fluorescent dyes we had used in previous work. To complement it, we imaged the HU-GFP strain in a buffer containing growth medium, optional antibiotics, and a fluorescent derivative of wheat-germ agglutenin - which provided regenerating fluorescent membrane staining analogous with the membrane dyes we had used.

By looking at the growth us untreated cells, we characterised the system as minimally-perturbative, giving us high confidence that our findings would translate to the fixed cell diagnostic protocol. By looking at the development of the phenotypes in treated cells, we concluded the treatment timescales we had optimised in the previous chapters were appropriate, and our identified phenotypes (Figure 2.2.1) were indeed appropriate end states of the development. In the case of co-amoxiclav, our classifier was trained on cells before they had managed to lyse; in the other cases, the phenotype represented the end state of nucleoid compaction or de-compaction. By considering both the individual phenotype classifications and the aggregate population response, we saw in treated samples the models confidently classifying cells as susceptible in as little as 10 minutes for ciprofloxacin and rifampicin, 20 minutes for gentamicin and 60 minutes in co-amoxiclav. As we discussed in the results section, the experimental limitation of the setup (time taken for sample mounting and autofocus set up means the start of the timelapse does not correspond to the true start of the antibiotic incubation) made the precise optimal timings difficult to infer from our results - nevertheless, we concluded that the incubation times we had used could have been shortened further, scaling well with the aim of a rapid diagnostic assay. The individual sensitivities and specificities of the models in the time domain differed between models. We found the ciprofloxacin model by far the most specific and sensitive - correctly classifying untreated cells in untreated samples (high specificity) and susceptible cells in the treated samples (high sensitivity), across all timepoints. All other models retained high sensitivity in the treated samples (on the order of 95%), however their specificities varied. It is worth highlighting again the significant domain shift in the data fed to the models - these models were trained on fixed cells, where both the staining and visual appearance of the nucleoid differed significantly from live cells. This likely accounted for many of the errors we saw. The gentamicin model showed poor specificity, which only decreased as a function of time - which we found perplexing, as this phenotype was qualitatively similar to that of ciprofloxacin. In the rifampicin and co-amoxiclav models, the specificity rose and fell on the order of around 30 minutes. This was likely due to the nucleoid reorganisation going on during the cell cycle of dividing, untreated cells, mimicking the appearance of the phenotypes.

At this point, it is worth reinforcing that, while the phenotyping was carried out as a *function* of time, it was not carried out *in* time, as the models were fed individual frames at individual time points only. This is in concordance with the fixed-cell diagnostic protocol - our models were trained to operate on individual frames, not collections of frames simultaneously. As a result, they could not learn the time dependence of of the phenotypes, or the cell-cycle processes of untreated cells.

4.4.2 Future directions

A notable direction for future work will be to phenotype live cells *in* time - training on a sequence of frames (a video). This would further allow a shortening of the time-to-result, and improve the specificity of our models by allowing them to explicitly learn the time dependence of the phenotypes. Extending to the time domain would also allow tracking of live cells, in turn improving our ability to measure and respond to heterogeneities in the population.

To carry out this work, we used a genetic engineering to replace the nucleoid dye and allow unperturbative, long term imaging in a specially prepared strain. However, just like we saw before, the greatest utility of timelapse imaging lies in deploying it on clinical isolates. Unfortunately, the approach we took is naturally incompatible with clinical isolates - to proceed, we would need to revisit the original idea and find a highly-specific, non-perturbative and cell-permeable nucleoid dye. This remains an exciting direction for future research.

In common with the discussion in previous chapters, this area would also benefit from a higher throughput sample preparation and imaging system. Here, due to the timelapse requirements, the throughput is even lower than before - only a single sample and FoV can be simultaneously imaged, and the requirement for live cells makes batch processing and storage of samples impossible. This is an important bottleneck that will have to be addressed in future work. This will in turn allow us to examine more conditions, and treatments at difference concentrations - which may have both a different development timeline, and endpoints.

While in previous chapters we were able to control cell density to avoid large clusters that our segmenter struggled with, this was much harder to achieve here - particularly for the untreated cells, which were dividing and naturally forming dense clusters as proto-colonies. Further improvements to the pipeline will greatly benefit future live-cell experiments.

Lastly, in this chapter we studied the development of the phenotypes as a function of time, and the DASP response as a function of phenotype development, but our work remained purely observational. A useful research direction for the future will be to create predictive, biophysical [179, 180] models, which could generate hypotheses that could then be experimentally verified. This could include, for example molecular models of the nucleoid DNA and its interactions with the proteins that are targets of the antibiotics we used. More simplistically, explicit mathematical models of nucleoid geometry using splines could also be a good place to start. These would allow us to quantify the phenotypes better, and thus quantify the effects of antibiotics on the phenotypes. We did not pursue this approach in this thesis - our work had a diagnostic focus, and we used CNNs to automatically extract features and create implicit models of the phenotypes; we focused on proving robustness and reliability over interpretability of the phenotype models. Nevertheless, this is an important avenue for future research, and it feeds well into the diagnostic context too - explicit mathematical models of the phenotypes would allow us to benchmark our deep-learning models against simpler models trained on engineered features.

4.5 Conclusions

In this chapter we studied the development of the phenotypes as a function of time, and the response of our DASP platform as a function of phenotype development, by using a genetically-engineered strain. We discovered all of our models are highly sensitive, but show varying degrees of specificity to untreated controls. This was likely due to the large domain shift away from the training data, and because our models had no access to time information - thus were unable to learn time-dependent processes such as cell division, that were present in untreated controls. Further work in this area is needed to move beyond genetically-engineered strains, and towards clinical isolates.

No one will read this, Alex. Not to make you feed bad or anything, it might take away some of the pressure.

— Annique Loncq de Jong

5

Conclusions

Contents

5.1 The state of the field of Antimicrobial Susceptibility Testing . . . 137
5.2 The literature gap - diagnostic utility of CNN-derived features and subcellular phenotypes. . . . 139
5.3 Addressing the gap - scientific contributions of this thesis . . . 140
5.4 Extensions and Future work . . . 144
5.5 Closing remarks . . . 150

Socrates is believed [1] to have stated that the unexamined life is not worth living. In turn, we believe the unexamined thesis is not worth reading. At last, the time has come to critically examine the contributions and limitations of this very thesis, before we proceed to a graceful conclusion.

5.1 The state of the field of Antimicrobial Susceptibility Testing

We began this thesis with a survey of the field and problem of AMR. We reviewed the modes of action of modern antibiotics, and the biochemical modes of resistance bacteria have developed to them. In doing so, we discovered that known resistance

---

[1] Plato, *Apology of Socrates*, 38a5–6

mechanisms exist for nearly all known antibiotics, and the timeline between the clinical introduction of a novel antibiotic and the first occurrence of resistance can be as short as a few years. Coupled with the failure of the modern drug-discovery pipeline to materialise a sufficient amount of novel antibiotics to outpace resistance, the approach to manage the rise of AMR has focused on antimicrobial stewardship - the preservation of the existing antibiotic arsenal through conservative prescription.

Moving on, we then reviewed in detail the main cornerstone of antimicrobial stewardship - rapid and specific diagnosis provided by an array of AST tests. This array of pipelines provides pathogen identification, and quantifies its susceptibility to a range of antibiotics, providing clinically-relevant information for decision making. We discussed the how current gold-standard AST is a phenotypic assay focusing on culture growth in the presence of antibiotics. This assay provides an unbiased and highly interpretable readout of susceptibility, which is agnostic to the both the pathogen and antibiotic under test. Unfortunately, such as assay is also time-consuming, requiring several cycles of pathogen culture in the isolation and AST steps - stretching the time-to-result to over 36 hours, even for common pathogens. To complement the phenotypic assay and shorten the time-to-result, a range of tests focusing on genomic, genetic or metabolic biomarkers have been developed - but by testing only a predefined set of secondary markers associated with (but not necessarily causative of) resistance, the output of such tests is neither highly interpretable nor agnostic.

Naturally, much effort has been invested in solving the interpretability challenge of biomarkers - by investigating more general biomarkers, and more comprehensive analysis tools to make better use of existing biomarkers. While much progress has been made, at the time of writing, further work remains to be done to prove their general utility to the point where they may supersede phenotypic indicators. Thus, another line of enquiry in the field of ASTs has focused on reimagining the gold-standard phenotypic growth assay. The ways to measure growth are evolving; examples include direct measurement of metabolites associated with growth, or measures of culture density that are more sensitive than the OD. The platforms for growth measurement are similarly evolving. A common new platform for this is widefield microscopy, which allows simultaneous monitoring and tracking of a large number of live cells, improving the resolution of the gold-standard assay to better handle population heterogeneities. Unfortunately, the use of widefield microscopy for growth monitoring does not fundamentally remove the cell culture requirement, which contributes to a high time-to-result. A close proxy of measuring cellular growth is the measurement of sub-cellular or cellular features (such as morphology, or movement) as they react to antibiotics - offering in theory all of the advantages we outlined above for growth monitoring under widefield microscopy, while eliminating the time-consuming cell culture step. Unfortunately, this is also not without challenges. The major challenge of this approach is feature engineering - finding a good set of uncorrelated features from that are highly-specific to antimicrobial action, and not associated with normal cell-cycle processes such as cell division, or stress factors outside of antimicrobial action. If found, such features could then be used as phenotypic readouts of antimicrobial action on a single cell level, and thus direct phenotypic readouts of antimicrobial susceptibility.

5.2 The literature gap - diagnostic utility of CNN-derived features and subcellular phenotypes.

It is at this precise point that we identified a gap in existing literature. Previous work [114–118] has characterised a large range of phenotypes based on the sub-cellular morphology of the nucleoid and membrane, revealed by fluorescent staining and widefield microscopy, and showed they are indeed highly specific to antimicrobial action. It was then showed that hand-engineered features derived from these phenotypes were indeed quantitatively sensitive readouts of antimicrobial susceptibility on a single-cell level [119] - but hand-engineering of features did not make this a scalable approach. An independent line of enquiry later showed [120] that features derived automatically from such phenotypes by CNNs can be used to classify them - however no attempt was made to investigate such features in a diagnostic context. This left a gap in the literature - hand-engineered features had the potential to be used as the foundation of an AST, but lacked the necessary throughput for practical implementation. On the other hand, CNN-derived features had the necessary throughput, but could not be demonstrated to have the necessary sensitivity to susceptibility.

To bridge this gap, we again turned to CNNs, this time with a focus on diagnostic utility and the deployment of classifiers on clinical isolates. In doing so, we created our Deep Antimicrobial Susceptibility Phenotyping (DASP) platform, which we then tested against a sequence of fundamental assumptions that needed to be experimentally verified to prove the diagnostic utility of CNN-derived features. We then spent the reminder of this thesis proving these assumptions one-by-one.

5.3 Addressing the gap - scientific contributions of this thesis

Our first fundamental assumption concerned the experimental reproducibility and robustness of the derived features - whether models trained on one selection of experiments could maintain comparable performance when deployed on a replicate experiment. We dedicated the entire second chapter (2) of this thesis to this question. This was a crucial diagnostic assumption - a technique that could not reproduce its results reliably could never be used to inform decision making. To test this assumption, we repeated our diagnostic workflow on different days, and then trained and tested the DASP pipeline on different subsets of the total dataset, and in different computational setups. First, we pooled all training experiments into one dataset, from which training and testing sets were derived. This ‘ensemble’ setup provided the expected upper bound of performance. Against this, we cross-validated the experiments by systematic withholding of one experiment for testing, and further deployed models on specially prepared ‘holdout’ experiment, to guard against human bias in experiment selection. We evaluated four different antibiotics, representatives of their major antibiotic families in this manner. As a further challenge, we picked the four antibiotics to provide a wide range of unique nucleoid phenotypes - two of them focused of different modes of nucleoid compaction, one on the opposite process of nucleoid decompaction, while one (co-amoxiclav) was a membrane-acting antibiotic with a very subtle phenotype. Across all settings observed our models maintain a high accuracy, showing the results to be reproducible. To further complement our results, we investigated which cellular features the models were focusing on in their decision making. This was also an important step - if our models had been focusing on features we knew were of no importance (such as background noise outside the cell), this would indicate the models would likely not be reliable. Fortunately, we discovered our models had indeed been focusing on the structure of the nucleoid primarily, with some membrane contributions; this observation bolstered our confidence in our models.

Our second fundamental assumption concerned the homogeneity of the phenotypes themselves between different isolates of the same species, and the relationship between the phenotype, the treatment concentration and isolate susceptibility. In other words - do different isolates of similar susceptibility display similar phenotypes when treated? Again, this was a crucial diagnostic assumption - if different isolates produced different phenotypes of susceptibility and resistance, our models could not be relied on to accurately quantify susceptibility. To validate this assumption, we worked with a selection of clinical isolates of various susceptibility to ciprofloxacin - one of our original antibiotics. When we treated isolates with ciprofloxacin, we discovered indeed that isolates susceptible to the treatment concentration displayed the same susceptible phenotype that we had observed in the training strain. By contrast, resistant strains showed a phenotype similar to the untreated phenotype of the training strain. By doing so, we also validated our secondary assumption about the concordance of the untreated phenotype and the resistant phenotype. This was not a crucial milestone assumption though - we could have equivalently used the actual (treated) resistant phenotypes to train the resistant class. However, it is worth mentioning the limits of our evidence here - we tested only one antibiotic in 6 clinical isolates, which shared between themselves only a limited pool of genetic mutations associated with resistance. Further work in the future will be necessary to examine more genotypes and antibiotics to provide further evidence to back up our claims.

From our second fundamental assumption, we derived our third assumption - that our models would derive similar features from similar phenotypes but coming from different isolates. In other words - when different isolates display similar phenotypes, can our models maintain good accuracy when distinguishing between them? Again, this was a crucial assumption - our models would be expected to be deployed regularly on unseen data, and would thus need to generalise well to unseen clinical isolates. To test this assumption, we deployed models trained on a lab reference strain on clinical isolates. From running our models on collections of FoVs, we discovered the distribution of single-cell phenotype classifications was sensitive to the susceptibility of the isolate. Across all isolates, untreated samples yielded cells that were predominantly classified as resistant. In resistant isolates, treated samples similarly yielded predominantly resistant cells, and no statistically significant difference existed in the classification distribution between treated and untreated samples. In susceptible isolates however, this changed dramatically, and treated samples yielded predominantly susceptible classifications - revealing highly statistically significant differences in the distributions. In doing so, we showed the models were indeed able to generalise to different isolates. From this point, we decided to probe the relationship between the classification distribution, the isolate MIC and treatment concentration further, by treating a selection of our isolates with different concentrations of ciprofloxacin. In this way, we discovered the distribution followed a dose-response relationship with the treatment concentration, qualitatively centred on the isolate MIC. Further work will be needed to establish how to quantitatively map the isolate MIC from the dose-response curve. Importantly, this presented us with the opportunity to benchmark our approach against the current gold-standard growth assay. Here we discovered our dose-responses were reciprocal to those of the growth assay - thus providing analogous information, but without the time-consuming cell culture step.

Summarising, we validated our third assumption (again, within the limits allowed to us by our data), and established the distribution of single-cell classifications as a sensitive reporting outcome for CNN-derived features. In doing so however, we left ourselves with yet another assumption to prove - that the distribution of single-cell classifications was indeed sensitive to the ground-truth proportions of resistant cells inside the tested sample. Given how the distribution was simply the aggregate of results coming from individually phenotyped cells, this may seem like a trivial assumption. However, if our models were deriving their features from anything else other than single-cell phenotypes (such as image background, or cell features other than the fluorescent phenotype), this assumption would be false. Again, this was a crucial assumption - if our models could not be shown to be sensitive to the actual cell population, they could not be relied upon to inform of it. To test this assumption, we prepared mixed-culture, treated samples containing different proportions of susceptible and resistant cells. Phenotyping these samples, we discovered a linear relationship between the classification distribution and ground-truth cell proportion - the linear relationship could exist if and only-if this assumption was valid. Unfortunately, comparing against a theoretical model conditioned on our previous work with monocultures, we did observe the distribution was not as sensitive as expected - this was likely the result of experimental error in sample preparation. Further work will be necessary to repeat these experiments and gather more evidence in support of this claim. Proving these assumptions took us the bulk of the third chapter (3).

We should summarise. Across the two chapters, we proved that CNN features are experimentally reproducible, correlate well with observed phenotypic changes and generalise well across different isolates when the different isolates produce similar phenotypes. Further, we showed that both the phenotypes and features derived from them are strong, but not injective functions of isolate MIC and treatment concentration - for example, we could observe cells displaying a susceptible phenotype in untreated samples, with the derived features indeed classifying it erroneously as susceptible. Fortunately, we also discovered the rate of such events was low [2], which still enabled us to demonstrate the distribution of classifications

---

[2] Quantifying this is difficult, as this could only be achieved by visual inspection and grading of a large amount of single-cell phnoetypes, and comparing against model predictions - not viable for our dataset size and vulnerable to human bias.

from CNN-derived features was sensitive and robust to isolate MIC and treatment concentration. We probed this relationship further, and demonstrated equivalence to the gold-standard assay at a significant time saving. Finally, by examining mixed-culture samples, we demonstrated the distribution is indeed sensitive to the resistance of individual cells inside a sample.

Taken together, we conclude that we proved all the necessary assumptions to validate features derived by CNNs from subcellular phenotypes are indeed sensitive to antimicrobial susceptibility, and could thus form the basis of an AST assay. Thus, we successfully addressed the gap in the literature we had identified.

From this point, we spent the rest of this thesis extending our results to live-cell, timelapse imaging. Our objective here was to study CNN-derived features as a function of phenotype development - previously impossible in our single-timepoint, fixed-cell assay. We discovered the features were indeed highly sensitive to susceptible phenotypes in all antibiotics - for example, the ciprofloxacin model would confidently classify cells as susceptible as soon as nucleoid compaction occurred. Unfortunately, the features and models showed poor specificity, regularly reporting false-positives (i.e. false susceptible cells) in untreated, growing samples. This was likely a consequence of their training - the models were trained on fixed cells, and thus could never learn the dynamics of cell growth and how that affects the resistant phenotype. Nevertheless, the high sensitivity in the face of a large domain shift increased our confidence in our models, and allowed us to strengthen our scientific conclusions from the fixed-cell assay.

5.4 Extensions and Future work

One of the major directions for future work arising from this thesis will be to examine how well these findings can generalise. As discussed at multiple points throughout, we worked with only once bacterial species (*E.coli*), and developed models for only 4 antibiotics. When we deployed our models on clinical isolates, we chose to continue with only one of those antibiotics, and all 6 of our clinical isolates shared a common set of mutations associated with resistance. Future work will centre on repeating our analysis for a greater range of species, antibiotics and resistance genotypes; it is our hope that work presented herein provides a reference and template useful for that purpose. A crucial challenge that will need to be overcome to achieve that goal will be to increase the experimental throughput - both sample preparation and imaging. As discussed before, for this work, samples were prepared manually by hand, and imaged individually under a microscope, taking a considerable amout of man-hours. Increasing throughput will likely involve a redesign of the experimental pipeline and the development of a custom high-throughput imaging system.

Another direction for future work will involve benchmarking and optimisation of the computational pipeline to further improve performance. As we discussed, our Mask-RCNN segmenter struggles with segmenting dense FoVs due to fundamental architecture assumptions - more recent models designed specifically for dense FoV segmentation such as Cellpose [167] show considerable promise. Another optimization axis will be computational speed. The time taken for our current pipeline to process a single FoV is on the order of seconds. A promising goal would be to phenotype cells in real-time as the imaging data is acquired; this goal would further synergise well with increasing the experimental throughput as we have just discussed. Beyond benchmarking against other deep-learning approaches, it will also be important to benchmark against the conventional approach of hand-engineered features - which was beyond the scope of this thesis, but nevertheless remains an important quality control step that is yet to be taken. Similarly, we based our phenotypes on the morphology of the nucleoid and cell membrane, fluorescently stained. Naturally, these were not the only available choices. We could have equivalently picked other sub-cellular structures, such as ribosomes, and different imaging modalities, such as phase contrast. Once again, a proper evaluation of DASP will require benchmarking against multiple potential phenotyping targets.

In this work, we showed the utility of CNN-derived features in a diagnostic capacity. Nevertheless, our current DASP pipeline does not yet represent a fully autonomous, competitive AST. Firstly, the input to the experimental pipeline is an identified bacterial isolate, cultured in conditions optimal for that isolate. Thus, at this time, DASP does not remove the pathogen isolation and identification steps of the AST pipeline (both of which take significant time). The only time saving currently offered is the final culture step of the AST itself - a significant saving, but more remains to be done.

To achieve this diagnostic endpoint, we phenotyped cells by their reaction to antibiotic; but this is not the only utility of CNN-derived features. The line of thought can be inverted - instead of phenotyping cells by the phenotype they show, one could phenotype antibiotics by the phenotype they produce. This could support high-throughput mode-of-action studies and staging studies of novel antimicrobial compounds. This approach has been attempted in the literature with engineered features [118] - an extension to CNN-derived features appears to be an excellent research direction.

A promising direction towards actualisation of DASP as a AST is integration with microfluidics. A microfluidic device could facilitate all of the experimental wetlab stages 2.1.2 of DASP - antibiotic treatment, fixation, staining, as well as bacterial immobilisation suitable for imaging. This would offer an advantage in throughput; helpful as discussed above. Beyond throughput, a microfluidic platform could easily facilitate extensions of the basic assay. A promising direction here is offered by the addition of FISH probes to obtain genetic information from the sample - which could allow pathogen identification by interrogation of conserved sequences, or identification of genetic markers correlated with resistance (similarly to current PCR tests). In a similar fashion, incorporation of fluorescent reporters could allow access transcriptiomic (RNA-FISH), proteomic (immunostaining) or metabolomic (aptamers) biomarkers - thus bridging the gap between separate phenotypic and biomarker AST assays by bringing the best of both approaches under a singular offering. Further, a microfluidic device could potentially offer selective bacterial isolation directly from patient samples through a set of specialised filters - removing the requirement for pathogen isolation and culture, and offering another significant time saving. This presents a challenge not just for the microfluidics, but also for the phenotyping. In our work, our phenotypes were produced in cells cultured in rich media. Removing this culture step thus removes a source of phenotype standardisation. Further work will be necessary to investigate the effect of culture or capture conditions on the displayed phenotypes.

Lastly, we finished this thesis with an outlook to time-domain, live-cell imaging and phenotyping. This approach reveals much promise - making use of the temporal dimension promises to not only further cut down the diagnostic time, but also improve sensitivity to important sample heterogeneities, such as persister cells. An important barrier to generalising this further to clinical isolates is robust and non-perturbative fluorescent staining - although once this obstacle is overcome, we anticipate much of our methodology to translate over to that setting.

We just discussed the potential for future development of DASP as an AST; let us conclude this section with a broader discussion of AI-aided diagnosis and biomarker discovery - the parent theme of the work we presented herein. At the time of writing, AI-aided diagnosis is undergoing dramatic expansion, but at many challenges remain. [181]. Many of these are ubiquitous to any application of AI - these include data drift, model explainability, and the engineering challenge of making sure sufficient volumes of high-quality data are available to accurately sample the underlying distribution. Some, however, are specific to this field. AI-aided diagnosis systems often face a high trust and sociological barrier to entry. On one hand, patients generally trust human physicians more than automated systems, on the other hand, physicians trust their own training more than automated systems. Crucial to this trust building exercise is the ongoing work in AI explainability and safety - reducing the risks from black-box models; as well as careful product positioning - building products as ‘physician assistance devices’ rather than complete diagnostic solutions helps to build confidence on both sides of the fence. In the context of AMR, diagnostic systems such as DASP operate on patient-derived samples and provide clinically relevant information rather than a complete diagnosis and prognosis tailored to the patient, which will likely speed up implementation. Another general problem of AI amplified in this setting is the problem of bias - algorithms reflecting and propagating the biases present in their training data, which can cause poor performance when deployed. When dealing with models trained on highly sensitive and protected data (as is the case in this context), this poor performance can further exaggerate societal inequalities. Fundamental to the reduction of bias is proper robust regulation and quality control. Recently, major regulators such as the US Food and Drugs Administration have began introducing regulatory frameworks based on existing medical device regulation, focusing on commitments from manufacturers to periodically monitor systems for signs of bias and drift.

A crucial missing piece needed to further advance and implement AI-aided diagnosis in the coming decades is an objective, reliable and accessible framework for measuring and comparing performance of models and datasets. Metrics commonly used in machine learning often do not reflect clinical applicability. We saw an example of this in our very work, where we struggled initially to relate single-cell susceptibility classification decisions to clinically relevant information - only at the end of chapter 3 we managed to correlate the population average response to the clinically relevant MIC measure. This problem of metrics propagates down to the very loss functions used to train the models, and thus their outputs - much scope remains for defining metrics and objective functions that more directly relate to the ultimate clinical metric, which is whether the use (or not) of the model leads to an improvement in the quality of care that is perceptible to either the patient (in the form of better patient outcomes), or the healthcare system as a whole (in the form of improved efficiency without sacrificing patient outcomes). Only once such metric exists, a meaningful bench-marking of model performance can be made.

We touched briefly already on some of the ethical implications of AI-aided diagnosis. It is beyond our scope to provide a full discussion, let alone a guide, but we should provide an outline.

As we mentioned, AI models trained on biased data perpetuate the biases they learnt. In the context of healthcare, models trained on datasets that are imbalanced with respect to race, gender, age or any other protected characteristic will likely show poor performance on underrepresented categories, in effect discriminating against them. Solving this however is not as simple as balancing the dataset at train time, but requires a systematic review of the biases also inherent in collecting the data, and assessing whether the data as collected provides a fair and accurate representation of the entire population. Naturally, such an explicit approach can only work when the sources of bias are known - in practice this is quite certainly not true. In this case, synthetic data generation may provide a framework for automated discovery of unknown biases, by generating a wide range of realistic and intentionally but randomly biased datasets to serve as quality control.

Closely related to the problem of bias is the general problem of equality and equity. Given that biases exist, should the models be trained to provide equal treatment for everyone (equality), or actively aim to compensate for the inequalities to equalise outcomes (equity). In the author's view, no final decision on this front can ever be truly reached.

Another issue common in diagnosis is that of informed patient consent. This operates on two fronts. Firstly, training large models takes much personal data, and the field is only in its infancy. Whilst protocols for seeking consent for research purposes exist, the capabilities of models advance at such a rate that it is questionable to assume that any consent given can truly be informed. At the time of writing, such large models are generally only trained by unregulated commercial entities - unlike canonical medical research that happens either in publicly-owned research institutions, or in the highly-regulated pharmaceutical sector. Advancements in AI regulations that we discussed above will likely go some way towards addressing this.

While the processing of data for training may always be reasonably denied, a separate consent issue exists when interacting with such systems at diagnosis time - particularly as they become more integrated and more difficult to avoid. If such models improve patient outcomes, this can create pressure to comply with processing of additional personal data (that would otherwise not be required).

In the author's view, the greatest ethical challenge of AI-assisted diagnosis is the redefinition of the relationship between patient and physician. Physicians often refer to the 'science and art' of medicine. True to this form, the role of the physician is not only to digest 'data' from a patient to produce a diagnosis, but do so in an empathetic and compassionate manner - the human connection provides an essential contribution to the healing process. At the same time, AI-assisted diagnosis promises to not only improve outcomes by drawing on insight from data no individual physician can match, but also improve throughput. How much of the human connection in medicine can we afford to sacrifice in pursuit of better outcomes? As with all great questions, there is no clear answer.

5.5 Closing remarks

In this conclusion, we once again revisited the current state of the field of Antimicrobial Susceptibility Testing in the context of AMR, and highlighted the gap in the literature we had identified. We then reviewed the scientific contributions we made in this thesis to address the gap in the literature. Lastly, we considered the implications and limitations of our claims, and discussed some of the future directions arising from our contributions.

> *There is no real ending. It's just the place where you stop the story.*
> — Frank Herbert

References

[1] Taichi A. Suzuki et al. “Codiversification of gut microbiota with humans”. In: *Science* 377.6612 (2022), pp. 1328–1332.

[2] Emily R. Davenport et al. “The human microbiome in evolution”. In: *BMC Biology* 15.1 (2017), pp. 1–12.

[3] S Doron and S L Gorbach. *Bacterial Infections : Overview.* 2008.

[4] Michael A. McGuckin et al. “Mucin dynamics and enteric pathogens”. In: *Nature Reviews Microbiology* 9.4 (2011), pp. 265–278.

[5] Mark Zanin et al. “The Interaction between Respiratory Pathogens and Mucus”. In: *Cell Host and Microbe* 19.2 (2016), pp. 159–168.

[6] Jean Félix Sicard et al. “Interactions of intestinal bacteria with components of the intestinal mucus”. In: *Frontiers in Cellular and Infection Microbiology* (2017).

[7] Patrick Di Martino. “Bacterial adherence: much more than a bond”. In: *AIMS Microbiology* 4.3 (2018), pp. 563–566.

[8] Musa Hassan Muhammad et al. “Beyond Risk: Bacterial Biofilms and Their Regulating Approaches”. In: *Frontiers in Microbiology* 11.May (2020), pp. 1–20.

[9] Pascale Cossart and Craig Roy. “Manipulation of host membrane machinery by bacterial pathogens”. In: *Current Opinion in Cell Biology* 24.4 (2010), pp. 547–554.

[10] José Carlos Santos and Jost Enninga. “At the crossroads: Communication of bacteria-containing vacuoles with host organelles”. In: *Cellular Microbiology* 18.3 (2016), pp. 330–339.

[11] Marcia Goldberg. “Actin-based motility of intracellular pathogens”. In: *Microbiology and Molecular Biology Reviews* (2001), pp. 595–626.

[12] Punsiri M. Colonne, Caylin G. Winchell, and Daniel E. Voth. “Hijacking host cell highways: Manipulation of the host actin cytoskeleton by obligate intracellular bacterial pathogens”. In: *Frontiers in Cellular and Infection Microbiology* 6.SEP (2016), pp. 1–8.

[13] Linlin Chen et al. “Inflammatory responses and inflammation-associated diseases in organs”. In: *Oncotarget* 9.6 (2018), pp. 7204–7218. URL: `www.impactjournals.com/oncotarget/`.

[14] David Chaplin. “Overview of the immune response”. In: *Journal of Allergy and Clinical Immunology* 125.2 (2010), p. 345.

[15] J. Davies and Dorothy Davies. “Origins and evolution of antibiotic resistance.” In: *Microbiology and Molecular Biology Reviews* (2010), pp. 417–433.

[16] Francesca Prestinaci, Patrizio Pezzotti, and Annalisa Pantosti. "Antimicrobial resistance: A global multifaceted phenomenon". In: *Pathogens and Global Health* 109.7 (2015), pp. 309–318. URL: `http://dx.doi.org/10.1179/2047773215Y.0000000030`.

[17] Alexander Fleming. "On the antibacterial action of cultures of a penicillium, with special reference to their use in the isotion of B.Influenzae". In: *The British Journal of Experimental Pathology* 10.3 (1929), pp. 226–236.

[18] Søren Brøgger Christensen. *Drugs that changed society: History and current status of the early antibiotics: Salvarsan, sulfonamides, and β-lactams.* Vol. 26. 19. 2021.

[19] Karen Bush and Patricia A Bradford. "β-Lactams and β-Lactamase Inhibitors: An Overview". In: *Cold Spring Harbor Perspectives in Meddicine* Table 1 (2016), p. 22.

[20] Christopher J Harrison and Denise Bratcher. "Cephalosporins : A Review". In: *Pediatrics in review* 29.8 (2008).

[21] Janet R. Casey and Michael E. Pichichero. "Meta-analysis of cephabsporins versus penicillin for treatment of group a streptococcal tonsillopharyngitis in adults". In: *Clinical Infectious Diseases* 38.11 (June 2004), pp. 1526–1534. URL: `https://academic.oup.com/cid/article/38/11/1526/284524`.

[22] Krisztina M. Papp-Wallace et al. "Carbapenems: Past, present, and future". In: *Antimicrobial Agents and Chemotherapy* 55.11 (2011), pp. 4943–4960.

[23] Aben Ovung and Jhimli Bhattacharyya. "Sulfonamide drugs: structure, antibacterial property, toxicity, and biophysical interactions". In: *Biophysical Reviews* 13.2 (2021), pp. 259–272.

[24] Kevin M. Krause et al. "Aminoglycosides: An overview". In: *Cold Spring Harbor Perspectives in Medicine* 6.6 (2016), pp. 1–18.

[25] L. P. Kotra, J. Haddad, and S. Mobashery. "Aminoglycosides: Perspectives on mechanisms of action and resistance and strategies to counter resistance". In: *Antimicrobial Agents and Chemotherapy* 44.12 (2000), pp. 3249–3256.

[26] Ian Chopra and Marilyn Roberts. "Tetracycline Antibiotics: Mode of Action, Applications, Molecular Biology, and Epidemiology of Bacterial Resistance". In: *Microbiology and Molecular Biology Reviews* 65.2 (2001), pp. 232–260.

[27] Stefan Schwarz et al. "Molecular basis of bacterial resistance to chloramphenicol and florfenicol". In: *FEMS Microbiology Reviews* 28.5 (2004), pp. 519–542.

[28] Ching Hao Li et al. "Chloramphenicol causes mitochondrial stress, decreases ATP biosynthesis, induces matrix metalloproteinase-13 expression, and solid-tumor cell invasion". In: *Toxicological Sciences* 116.1 (2010), pp. 140–150.

[29] George P. Dinos. "The macrolide antibiotic renaissance". In: *British Journal of Pharmacology* 174.18 (2017), pp. 2967–2983.

[30] Hee Kyoung Kang and Yoonkyung Park. "Glycopeptide antibiotics: Structure and mechanisms of action". In: *Journal of Bacteriology and Virology* 45.2 (2015), pp. 67–78.

[31] W. Wehrli and M. Staehelin. "Actions of the rifamycins." In: *Bacteriological reviews* 35.3 (1971), pp. 290–309.

[32] Elizabeth A. Campbell et al. “Structural mechanism for rifampicin inhibition of bacterial RNA polymerase”. In: *Cell* 104.6 (2001), pp. 901–912.

[33] Sophie Reissier and Vincent Cattoir. “Streptogramins for the treatment of infections caused by Gram-positive pathogens”. In: *Expert Review of Anti-infective Therapy* (2021), pp. 1–13.

[34] Katie J. Aldred, Robert J. Kerns, and Neil Osheroff. “Mechanism of quinolone action and resistance”. In: *Biochemistry* 53.10 (2014), pp. 1565–1574.

[35] Claudia Foti et al. “Oxazolidinone antibiotics: Chemical, biological and analytical aspects”. In: *Molecules* 26.14 (2021).

[36] Truc Tran, Jose Munita, and Cesar Arias. “Mechanisms of Drug Resistance: Daptomycin Resistance”. In: *Annals of the New York Academy of Sciences* 1354 (2015), pp. 32–53.

[37] Judith N. Steenbergen et al. “Daptomycin: A lipopeptide antibiotic for the treatment of serious Gram-positive infections”. In: *Journal of Antimicrobial Chemotherapy* 55.3 (2005), pp. 283–288.

[38] Catherine L. Tooke et al. “β-Lactamases and β-Lactamase Inhibitors in the 21st Century”. In: *Journal of Molecular Biology* 431.18 (2019), pp. 3472–3500. URL: `https://doi.org/10.1016/j.jmb.2019.04.002`.

[39] Jed F. Fisher and Shahriar Mobashery. “β-Lactam resistance mechanisms: Gram-positive bacteria and mycobacterium tuberculosis”. In: *Cold Spring Harbor Perspectives in Medicine* 6.5 (2016), pp. 1–19.

[40] David Paterson and Robert Bonomo. “Extended spectrum beta lactamases: A clinical update”. In: *Clinical Microbiology Reviews* 18.4 (2005).

[41] Francis Codjoe and Eric Donkor. “Carbapenem Resistance: A Review”. In: *Medical Sciences* 6.1 (2017), p. 1.

[42] Dae Wi Kim et al. “A novel sulfonamide resistance mechanism by two-component flavin-dependent monooxygenase system in sulfonamide-degrading actinobacteria”. In: *Environment International* 127.January (2019), pp. 206–215. URL: `https://doi.org/10.1016/j.envint.2019.03.046`.

[43] Alisa W. Serio et al. “Aminoglycoside Revival: Review of a Historically Important Class of Antimicrobials Undergoing Rejuvenation”. In: *EcoSal Plus* 8.1 (2018).

[44] Rya Ero, Xin Fu Yan, and Yong Gui Gao. “Ribosome protection proteins—“new” players in the global arms race with antibiotic-resistant pathogens”. In: *International Journal of Molecular Sciences* 22.10 (2021).

[45] Beth P. Goldstein. “Resistance to rifampicin: A review”. In: *Journal of Antibiotics* 67.9 (2014), pp. 625–630.

[46] George A. Jacoby. “Mechanisms of resistance to quinolones”. In: *Clinical Infectious Diseases* 41.2 SUPPL. (2005).

[47] Scott H. Podolsky. “The evolving response to antibiotic resistance (1945–2018)”. In: *Palgrave Communications* 4.1 (2018). URL: `http://dx.doi.org/10.1057/s41599-018-0181-x`.

[48] Christopher JL Murray et al. “Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis”. In: *The Lancet* 399.10325 (Feb. 2022), pp. 629–655.

[49] C Ventola. “The Antibiotic Resistance Crisis”. In: *P&T* 40.4 (2015), pp. 277–283.

[50] Ramanan Laxminarayan et al. “Antibiotic resistance-the need for global solutions”. In: *The Lancet Infectious Diseases* 13.12 (2013), pp. 1057–1098.

[51] F. Christiaan K. Dolk et al. “Antibiotics in primary care in England: Which antibiotics are prescribed and for which conditions?” In: *Journal of Antimicrobial Chemotherapy* 73 (2018), pp. ii2–ii10.

[52] Alison H. Holmes et al. “Understanding the mechanisms and drivers of antimicrobial resistance”. In: *The Lancet* 387.10014 (2016), pp. 176–187.

[53] Peter Vikesland et al. “Differential Drivers of Antimicrobial Resistance across the World”. In: *Accounts of Chemical Research* 52.4 (2019), pp. 916–924.

[54] Ana Catarina Duarte et al. “Antibiotic Resistance in the Drinking Water: Old and New Strategies to Remove Antibiotics, Resistant Bacteria, and Resistance Genes”. In: *Pharmaceuticals* 15.4 (2022), pp. 1–22.

[55] Mariusz Cycoń, Agnieszka Mrozik, and Zofia Piotrowska-Seget. “Antibiotics in the soil environment—degradation and their impact on microbial activity and diversity”. In: *Frontiers in Microbiology* 10.MAR (2019).

[56] Bernardo Ribeiro da Cunha, Luís P. Fonseca, and Cecília R.C. Calado. “Antibiotic discovery: Where have we come from, where do we go?” In: *Antibiotics* 8.2 (2019).

[57] Kim Lewis. “The Science of Antibiotic Discovery”. In: *Cell* 181.1 (2020), pp. 29–45. URL: `https://doi.org/10.1016/j.cell.2020.02.056`.

[58] Ruben Tommasi et al. “ESKAPEing the labyrinth of antibacterial discovery”. In: *Nature Reviews Drug Discovery* 14.August (2015), p. 529. URL: `www.nature.com/reviews/drugdisc`.

[59] Erica Parker et al. “Implementation of permeation rules leads to a Fabl inhibitor with activity against Gram-negative pathogens”. In: *Nature Microbiology* 5.1 (2020), pp. 67–75.

[60] Alexander Crits-Christoph et al. “Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis”. In: *Nature* 558.7710 (2018), pp. 440–444. URL: `http://dx.doi.org/10.1038/s41586-018-0207-y`.

[61] Ilinca A. Dutescu and Sean A. Hillie. “Encouraging the development of new antibiotics: Are financial incentives the right way forward? A systematic review and case study”. In: *Infection and Drug Resistance* 14 (2021), pp. 415–434.

[62] Bethany A. Van Dort et al. “The impact of digital interventions on antimicrobial stewardship in hospitals: A qualitative synthesis of systematic reviews”. In: *Journal of Antimicrobial Chemotherapy* 77.7 (2022), pp. 1828–1837.

[63] Lesley Hawes, Kirsty Buising, and Danielle Mazza. “Antimicrobial stewardship in general practice: A scoping review of the component parts”. In: *Antibiotics* 9.8 (2020), pp. 1–13.

[64] Esmita Charani et al. “Investigating the cultural and contextual determinants of antimicrobial stewardship programmes across low-, middle- and high-income countries—A qualitative study”. In: *PLoS ONE* 14.1 (2019), pp. 1–20.

[65] Robert Rolfe et al. “Barriers to implementing antimicrobial stewardship programs in three low- and middle-income country tertiary care settings: findings from a multi-site qualitative study”. In: *Antimicrobial Resistance and Infection Control* 10.1 (2021), pp. 1–11. URL: `https://doi.org/10.1186/s13756-021-00929-4`.

[66] V. A. Schweitzer et al. *The quality of studies evaluating antimicrobial stewardship interventions: a systematic review.* May 2019.

[67] Antti Vasala, Vesa P. Hytönen, and Olli H. Laitinen. *Modern Tools for Rapid Diagnostics of Antimicrobial Resistance.* July 2020.

[68] Z. A. Khan, M. F. Siddiqui, and S. Park. “Current and Emerging Methods of Antibiotic Susceptibility Testing”. In: *Diagnostics* 9.2 (2019).

[69] Ina Gajic et al. “Antimicrobial Susceptibility Testing: A Comprehensive Review of Currently Used Methods”. In: *Antibiotics* 11.4 (2022), pp. 1–26.

[70] M. Benkova, O. Soukup, and J. Marek. “Antimicrobial susceptibility testing: currently used methods and devices and the near future in clinical practice”. In: *Journal of Applied Microbiology* 129.4 (2020), pp. 806–822.

[71] Ritu Banerjee and Romney Humphries. “Rapid Antimicrobial Susceptibility Testing Methods for Blood Cultures and Their Clinical Impact”. In: *Frontiers in Medicine* 8.March (2021).

[72] Thi Ngoc Diep Trinh and Nae Yoon Lee. “Nucleic acid amplification-based microfluidic approaches for antimicrobial susceptibility testing”. In: *Analyst* 146.10 (2021), pp. 3101–3113.

[73] EUCAST. *The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 12.0, 2022. http://www.eucast.org.* Tech. rep.

[74] Angella Charnot-Katsikas et al. *Use of the Accelerate Pheno System for Identification and Antimicrobial Susceptibility Testing of Pathogens in Positive Blood Cultures and Impact on Time to Results and Workflow.* Tech. rep. 2017. URL: `https://journals.asm.org/journal/jcm`.

[75] Xianghe Yan et al. “Comparative Transcriptome Analysis Reveals Differentially Expressed Genes Related to Antimicrobial Properties of Lysostaphin in Staphylococcus aureus”. In: *Antibiotics* 11.2 (2022).

[76] Marjan M Hashemi et al. “A Novel Platform Using RNA Signatures To Accelerate Antimicrobial Susceptibility Testing in Neisseria gonorrhoeae”. In: *Journal of Clinical Microbiology* 58.12 (2020), pp. 1–11.

[77] Tahmineh Khazaei et al. “RNA markers enable phenotypic test of antibiotic susceptibility in Neisseria gonorrhoeae after 10 minutes of ciprofloxacin exposure”. In: *Scientific Reports* 8.1 (2018), pp. 1–10. URL: `http://dx.doi.org/10.1038/s41598-018-29707-w`.

[78] Armelle Novelli Rousseau et al. “Fast Antibiotic Susceptibility Testing via Raman Microspectrometry on Single Bacteria: An MRSA Case Study”. In: *ACS Omega* (2021).

[79] Min Liu et al. “Single-Cell Identification, Drug Susceptibility Test, and Whole-genome Sequencing of Helicobacter pylori Directly from Gastric Biopsy by Clinical Antimicrobial Susceptibility Test Ramanometry”. In: *Clinical chemistry* 68.8 (2022), pp. 1064–1074.

[80] William John Thrift et al. “Deep learning analysis of vibrational spectra of bacterial lysate for rapid antimicrobial susceptibility testing”. In: *ACS Nano* 14.11 (2020), pp. 15336–15348.

[81] Maximilian Kittel et al. “Rapid susceptibility testing of multi-drug resistant Escherichia coli and Klebsiella by glucose metabolization monitoring”. In: *Clinical Chemistry and Laboratory Medicine* 57.8 (2019), pp. 1271–1279.

[82] Thomas Rydzak et al. “Metabolic preference assay for rapid diagnosis of bloodstream infections”. In: *Nature Communications* 13.1 (2022).

[83] Stephan Rösner et al. “Evaluation of a novel immunochromatographic lateral flow assay for rapid detection of OXA-48, NDM, KPC and VIM carbapenemases in multidrug-resistant Enterobacteriaceae”. In: *Journal of Medical Microbiology* 68.3 (2019), pp. 379–381.

[84] Hervé Volland et al. “Development and multicentric validation of a lateral flow immunoassay for rapid detection of mcr-1-producing Enterobacteriaceae”. In: *Journal of Clinical Microbiology* 57.5 (2019).

[85] Chen Chen and Weili Hong. “Recent development of rapid antimicrobial susceptibility testing methods through metabolic profiling of bacteria”. In: *Antibiotics* 10.3 (2021).

[86] Pedro Madrigal et al. “Machine learning algorithm to characterize antimicrobial resistance associated with the International Space Station surface microbiome”. In: *Microbiome* 10.1 (2022), pp. 1–12. URL: `https://doi.org/10.1186/s40168-022-01332-w`.

[87] Yunxiao Ren et al. “Prediction of antimicrobial resistance based on whole-genome sequencing and machine learning”. In: *Bioinformatics* 38.2 (2022), pp. 325–334.

[88] Ines Ferreira et al. “Species identification and antibiotic resistance prediction by analysis of whole-genome sequence data by use of ARESdb: An analysis of isolates from the unyvero lower respiratory tract infection trial”. In: *Journal of Clinical Microbiology* 58.7 (2020).

[89] Yuan Yue et al. “Whole-genome sequencing-based prediction and analysis of antimicrobial resistance in Yersinia enterocolitica from Ningxia, China”. In: *Frontiers in Microbiology* 13.July (2022), pp. 1–15.

[90] M. J. Ellington et al. “The role of whole genome sequencing in antimicrobial susceptibility testing of bacteria: report from the EUCAST Subcommittee”. In: *Clinical Microbiology and Infection* 23.1 (2017), pp. 2–22.

[91] Nathan G Schoepp et al. “Rapid pathogen-specific phenotypic antibiotic susceptibility testing using digital LAMP quantification in clinical samples”. In: *Science Translational Medicine* 9.410 (2017), pp. 1–28.

[92] Lalainasoa Odile Rivoarilala et al. “LAMP assays for the simple and rapid detection of clinically important urinary pathogens including the detection of resistance to 3rd generation cephalosporins”. In: *BMC Infectious Diseases* 21.1 (2021), pp. 1–10. URL: `https://doi.org/10.1186/s12879-021-06720-5`.

[93] Kathleen Mach et al. “Optimizing peptide nucleic acid probes for hybridization-based detection and identification of bacterial pathogens”. In: *Analyst* 144.5 (2019), pp. 1565–1574. URL: `file:///C:/Users/Carla%0ACarolina/Desktop/Artigos%0Apara%0Aacrescentar%0Ana%0Aqualifica%C3%A7%C3%A3o/The%0Aimpact%0Aof%0Abirth%0Aweight%0Aon%0Acardiovascular%0Adisease%0Arisk%0Ain%0Afile:///C:/Users/Carla%20Carolina/Desktop/Artigos%20para%20acrescentar%20na%20qualifica%C3%A7%C3%A3o/Th`.

[94] Kieran Mulroney et al. “Same-day confirmation of infection and antimicrobial susceptibility profiling using flow cytometry”. In: *eBioMedicine* 82 (2022), p. 104145. URL: `https://doi.org/10.1016/j.ebiom.2022.104145`.

[95] Takahiro Sawada et al. “Early detection of drug-resistant Streptococcus pneumoniae and Haemophilus influenzae by quantitative flow cytometry”. In: *Scientific Reports* 11.1 (2021), pp. 1–8. URL: `https://doi.org/10.1038/s41598-021-82186-4`.

[96] Ward L. Johnson et al. “Sensing bacterial vibrations and early response to antibiotics with phase noise of a resonant crystal”. In: *Scientific Reports* 7.1 (2017), pp. 1–12. URL: `http://dx.doi.org/10.1038/s41598-017-12063-6`.

[97] P. Stupar et al. “Nanomechanical sensor applied to blood culture pellets: a fast approach to determine the antibiotic susceptibility against agents of bloodstream infections”. In: *Clinical Microbiology and Infection* 23.6 (2017), pp. 400–405. URL: `http://dx.doi.org/10.1016/j.cmi.2016.12.028`.

[98] Randall T. Hayden et al. “Rapid antimicrobial susceptibility testing using forward laser light scatter technology”. In: *Journal of Clinical Microbiology* 54.11 (2016), pp. 2701–2706.

[99] Xuyang Shi et al. “Culture-free bacterial detection and identification from blood with rapid, phenotypic, antibiotic susceptibility testing”. In: *Scientific Reports* 8.1 (2018), pp. 1–11. URL: `http://dx.doi.org/10.1038/s41598-018-21520-9`.

[100] L. Jiang et al. “Digital antimicrobial susceptibility testing using the MilliDrop technology”. In: *European Journal of Clinical Microbiology and Infectious Diseases* 35.3 (2016), pp. 415–422.

[101] Petra Jusková et al. “Real-Time Respiration Changes as a Viability Indicator for Rapid Antibiotic Susceptibility Testing in a Microfluidic Chamber Array”. In: *ACS Sensors* 6.6 (2021), pp. 2202–2210.

[102] Yuto Kawai et al. “Development of an efficient antimicrobial susceptibility testing method with species identification by Nanopore sequencing of 16S rRNA amplicons”. In: *PLoS ONE* 17.2 February (2022), pp. 1–12. URL: `http://dx.doi.org/10.1371/journal.pone.0262912`.

[103] Tao Dong and Xinyan Zhao. “Rapid identification and susceptibility testing of uropathogenic microbes via immunosorbent ATP-bioluminescence assay on a microfluidic simulator for antibiotic therapy”. In: *Analytical Chemistry* 87.4 (2015), pp. 2410–2418.

[104] Özden Baltekin et al. “Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging”. In: *Proceedings of the National Academy of Sciences* 114.34 (2017), pp. 9170–9175.

[105] Kuangwen Hsieh et al. “Combating Antimicrobial Resistance via Single-Cell Diagnostic Technologies Powered by Droplet Microfluidics”. In: *Accounts of Chemical Research* 55.2 (2022), pp. 123–133.

[106] Hui Li et al. “Adaptable microfluidic system for single-cell pathogen classification and antimicrobial susceptibility testing”. In: *Proceedings of the National Academy of Sciences of the United States of America* 116.21 (2019), pp. 10270–10279.

[107] Jungil Choi et al. “Rapid antibiotic susceptibility testing by tracking single cell growth in a microfluidic agarose channel system”. In: *Lab on a Chip* 13.2 (2013), pp. 280–287.

[108] Donghui Song and Yu Lei. “Mini-review: Recent advances in imaging-based rapid antibiotic susceptibility testing”. In: *Sensors and Actuators Reports* 3.September (2021), p. 100053. URL: `https://doi.org/10.1016/j.snr.2021.100053`.

[109] Jungil Choi et al. “A rapid antimicrobial susceptibility test based on single-cell morphological analysis”. In: *Science Translational Medicine* 6.267 (2014).

[110] Rafael Iriya et al. “Rapid Antibiotic Susceptibility Testing Based on Bacterial Motion Patterns with Long Short- Term Memory Neural Networks”. In: *IEEE Sensors Journal* 20.9 (2020), pp. 4940–4950.

[111] Hui Yu et al. “Phenotypic Antimicrobial Susceptibility Testing with Deep Learning Video Microscopy”. In: *Analytical Chemistry* 90.10 (2018), pp. 6314–6322.

[112] Vinodh Kandavalli et al. “Rapid antibiotic susceptibility testing and species identification for mixed samples”. In: *Nature Communications* 13.1 (2022), pp. 1–8.

[113] Alex van Belkum et al. “Innovative and rapid antimicrobial susceptibility testing systems”. In: *Nature Reviews Microbiology* 18.5 (2020), pp. 299–311. URL: `http://dx.doi.org/10.1038/s41579-020-0327-x`.

[114] Somenath Bakshi et al. “Nonperturbative Imaging of Nucleoid Morphology in Live Bacterial Cells during an Antimicrobial Peptide Attack”. In: *Applied and Environmental Microbiology* 80.16 (2014), pp. 4977–4986.

[115] Sushmita Sridhar et al. “High-Content Imaging to Phenotype Antimicrobial Effects on Individual Bacteria at Scale”. In: *mSystems* 6.3 (June 2021).

[116] Lidia Araújo-Bazán et al. “Cytological profile of antibacterial FtsZ inhibitors and synthetic peptide MciZ”. In: *Frontiers in Microbiology* 7.OCT (Oct. 2016).

[117] Htut Htut Htoo et al. “Bacterial Cytological Profiling as a Tool To Study Mechanisms of Action of Antibiotics That Are Active against Acinetobacter baumannii”. In: *Antimicrobial Agents and Chemotherapy* 63.4 (Apr. 2019).

[118] Poochit Nonejuie et al. “Bacterial cytological profiling rapidly identifies the cellular pathways targeted by antibacterial molecules”. In: *Proceedings of the National Academy of Sciences of the United States of America* 110.40 (Oct. 2013), pp. 16169–16174.

[119] D. T. Quach et al. “Bacterial Cytological Profiling (BCP) as a Rapid and Accurate Antimicrobial Susceptibility Testing Method for Staphylococcus aureus”. In: *EBioMedicine* 4 (Feb. 2016), pp. 95–103.

[120] Christoph Spahn et al. “DeepBacs for multi-task bacterial image analysis using open-source deep learning approaches”. In: *Communications Biology* 5.1 (2022).

[121] Alex van Belkum et al. “Developmental roadmap for antimicrobial susceptibility testing systems”. In: *Nature Reviews Microbiology* 17.1 (2019), pp. 51–62. URL: `http://dx.doi.org/10.1038/s41579-018-0098-9`.

[122] Ronald R. Kline. “Cybernetics, automata studies, and the dartmouth conference on artificial intelligence”. In: *IEEE Annals of the History of Computing* 33.4 (2011), pp. 5–16.

[123] Wenlong Sun, Olfa Nasraoui, and Patrick Shafto. *Evolution and impact of bias in human and machine learning algorithm interaction.* Vol. 15. 8 August. 2020, pp. 1–39. URL: `http://dx.doi.org/10.1371/journal.pone.0235502`.

[124] Benjamin van Giffen, Dennis Herhausen, and Tobias Fahse. “Overcoming the pitfalls and perils of algorithms: A classification of machine learning biases and mitigation methods”. In: *Journal of Business Research* 144.February (2022), pp. 93–106. URL: `https://doi.org/10.1016/j.jbusres.2022.01.076`.

[125] Salem Alelyani. “Detection and evaluation of machine learning bias”. In: *Applied Sciences (Switzerland)* 11.14 (2021), pp. 1–17.

[126] Richard Bellman. *Adaptive control processes : a guided tour.* Princeton University Press, 1961.

[127] Iqbal H. Sarker. “Machine Learning: Algorithms, Real-World Applications and Research Directions”. In: *SN Computer Science* 2.3 (2021), pp. 1–21. URL: `https://doi.org/10.1007/s42979-021-00592-x`.

[128] Moshe Sipper. “High Per Parameter: A Large-Scale Study of Hyperparameter Tuning for Machine Learning Algorithms”. In: *Algorithms* 15.9 (2022).

[129] Takuya Akiba et al. “Optuna: A Next-generation Hyperparameter Optimization Framework”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2019), pp. 2623–2631.

[130] Haifeng Jin, Qingquan Song, and Xia Hu. “Auto-keras: An efficient neural architecture search system”. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2019), pp. 1946–1956.

[131] Mykel J Kochenderfer and Tim A Wheeler. *Algorithms for Optimization.* MIT Press, 2019. URL: `http://mitpress.mit.edu.`.

[132] David Luebke. “CUDA: Scalable parallel programming for high-performance scientific computing”. In: *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Proceedings, ISBI* (2008), pp. 836–838.

[133] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems.* 2015. URL: `https://www.tensorflow.org/`.

[134] Adam Paszke et al. “PyTorch: An imperative style, high-performance deep learning library”. In: *Advances in Neural Information Processing Systems* 32.NeurIPS (2019).

[135] Chollet François. *Keras*. 2015. URL: `https://keras.io`.

[136] G Cybenko. “Approximation by Superpositions of a Sigmoidal Function”. In: *Mathematics of Control, Signals, and Systems* 2 (1989), pp. 303–314.

[137] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. “Multilayer feedforward networks are universal approximators”. In: *Neural Networks* 2.5 (1989), pp. 359–366.

[138] Sejun Park et al. “Minimum Width for Universal Approximation”. In: *arXiv preprint arXiv:2106.11342:2006.08859* (2020), pp. 1–29. URL: `http://arxiv.org/abs/2006.08859`.

[139] Hongzhou Lin and Stefanie Jegelka. “ResNet with one-neuron hidden layers is a Universal Approximator”. In: *Advances in Neural Information Processing Systems* (2018), pp. 6169–6178.

[140] Barbara Hammer and Kai Gersmann. “A Note on the Universal Approximation Capability of Support Vector Machines”. In: *Neural Processing Letters* 17.1 (2002).

[141] Aston Zhang et al. “Dive Into Deep Learning”. In: *arXiv preprint arXiv:2106.11342* (2021). URL: `https://d2l.ai/index.html`.

[142] Aleksandar Botev, Guy Lever, and David Barber. “Nesterov’s accelerated gradient and momentum as approximations to regularised update descent”. In: *Proceedings of the International Joint Conference on Neural Networks* 2017-May.2 (2017), pp. 1899–1903.

[143] Nitish Srivastava et al. “Dropout: A simple way to prevent neural networks from overfitting”. In: *Journal of Machine Learning Research* 15 (2014), pp. 1929–1958.

[144] Richard Szeliski. *Computer Vision : Algorithms and Applications 2nd Edition*. 2021, p. 1196.

[145] Ding Xuan Zhou. “Universality of deep convolutional neural networks”. In: *arXiv preprint arXiv:1805.10769v2* (2018).

[146] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. “ImageNet Classification with Deep Convolutional Neural Networks”. In: *Advances in Neural Information Processing Systems* 25.2 (2012).

[147] Karen Simonyan and Andrew Zisserman. “Very deep convolutional networks for large-scale image recognition”. In: *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2015), pp. 1–14.

[148] Min Lin, Qiang Chen, and Shuicheng Yan. “Network in network”. In: *2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings* (2014), pp. 1–10.

[149] Christian Szegedy et al. “Going deeper with convolutions”. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 07-12-June (2015), pp. 1–9.

[150] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-Decem (2016), pp. 770–778.

[151] Saining Xie et al. "Aggregated residual transformations for deep neural networks". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-Janua (2017), pp. 5987–5995.

[152] Gao Huang et al. "Densely connected convolutional networks". In: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* 2017-Janua (2017), pp. 2261–2269.

[153] Howard Andrew et al. "Searching for MobileNetV3". In: *Proceedings of the IEEE International Conference on Computer Vision* (2019), pp. 1314–1324.

[154] Mingxing Tan and Quoc V. Le. "EfficientNet: Rethinking model scaling for convolutional neural networks". In: *36th International Conference on Machine Learning, ICML 2019* 2019-June (2019), pp. 10691–10700.

[155] Rami Ibrahim and M. Omair Shafiq. "Explainable Convolutional Neural Networks: A Taxonomy, Review, and Future Directions". In: *ACM Computing Surveys* (2022).

[156] Ross Girshick et al. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (2014), pp. 580–587.

[157] Ross Girshick. "Fast R-CNN". In: *Proceedings of the IEEE International Conference on Computer Vision* 2015 Inter (2015), pp. 1440–1448.

[158] Shaoqing Ren et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39.6 (2017), pp. 1137–1149.

[159] Joseph Redmon et al. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016-Decem (2016), pp. 779–788.

[160] Nicolas Carion et al. "End-to-End Object Detection with Transformers". In: *arXiv preprint arXiv:2005.12872v3* (2020).

[161] K He et al. "Mask R-CNN". In: *International Conference on Computer Vision.* 2017, pp. 2961–2969.

[162] Alexey Dosovitskiy et al. "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale". In: *arXiv preprint arXiv:2010.11929v2* (2021). URL: `http://arxiv.org/abs/2010.11929`.

[163] Terrance DeVries and Graham W. Taylor. "Improved Regularization of Convolutional Neural Networks with Cutout". In: *ArXiv* (2017). URL: `http://arxiv.org/abs/1708.04552`.

[164] Waleed Abdulla. *Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow.* 2017. URL: `https://github.com/matterport/Mask_RCNN`.

[165] Tsung Yi Lin et al. “Microsoft COCO: Common objects in context”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8693 LNCS.PART 5 (2014), pp. 740–755.

[166] Jost Tobias Springenberg et al. “Striving for simplicity: The all convolutional net”. In: *3rd International Conference on Learning Representations, ICLR 2015 - Workshop Track Proceedings* (2015), pp. 1–14.

[167] Carsen Stringer et al. “Cellpose: a generalist algorithm for cellular segmentation”. In: *Nature Methods* 18.1 (Jan. 2021), pp. 100–106.

[168] Samuel Lipworth et al. “Ten-year longitudinal molecular epidemiology study of Escherichia coli and Klebsiella species bloodstream infections in Oxfordshire, UK”. In: *Genome Medicine* 13.1 (2021), pp. 1–13.

[169] María Tamayo et al. “Rapid assessment of the effect of ciprofloxacin on chromosomal DNA from Escherichia coli using an in situ DNA fragmentation assay”. In: *BMC Microbiology* 9 (2009), pp. 1–11.

[170] Xuan Li and Wolf Dietrich Heyer. “Homologous recombination in DNA repair and DNA damage tolerance”. In: *Cell Research* 18.1 (2008), pp. 99–113.

[171] Emrah Şimşek and Minsu Kim. “The emergence of metabolic heterogeneity and diverse growth responses in isogenic bacterial cells”. In: *ISME Journal* 12.5 (2018), pp. 1199–1209.

[172] Kim Lewis. “Persister cells, dormancy and infectious disease”. In: *Nature Reviews Microbiology* 5.1 (2007), pp. 48–56.

[173] The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *Validation of Analytical Procedures: Text and Methodology (ICH Q2(R1)).* Tech. rep. 1995.

[174] Valeria Bortolaia et al. “ResFinder 4.0 for predictions of phenotypes from genotypes”. In: *Journal of Antimicrobial Chemotherapy* 75.12 (2020), pp. 3491–3500.

[175] Muriel Wery, Conrad L. Woldringh, and Josette Rouviere-Yaniv. “HU-GFP and DAPI co-localize on the Escherichia coli nucleoid”. In: *Biochimie* 83.2 (2001), pp. 193–200.

[176] Anteneh Hailu Abebe, Alexander Aranovich, and Itzhak Fishov. “HU content and dynamics in Escherichia coli during the cell cycle and at different growth rates”. In: *FEMS Microbiology Letters* 364.19 (2017), pp. 1–8.

[177] R. K. Sizemore, J. J. Caldwell, and A. S. Kendrick. “Alternate gram staining technique using a fluorescent lectin”. In: *Applied and Environmental Microbiology* 56.7 (1990), pp. 2245–2247.

[178] Aimee H. Marceau et al. “Structure of the SSB-DNA polymerase III interface and its role in DNA replication”. In: *EMBO Journal* 30.20 (2011), pp. 4236–4247. URL: `http://dx.doi.org/10.1038/emboj.2011.305`.

[179] Andrea M. Chiariello et al. “Polymer physics of chromosome large-scale 3D organisation”. In: *Scientific Reports* 6 (2016), pp. 1–8.

[180] John Marko. “Biophysics of protein-DNA interactions and chromosome organization John”. In: *Physica A: Statistical Mechanics and its Applications* 418 (2015), pp. 1–126. URL: `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3624763/pdf/nihms412728.pdf`.

[181] Christopher J. Kelly et al. “Key challenges for delivering clinical impact with artificial intelligence”. In: *BMC Medicine* 17.1 (2019), pp. 1–9.