Dissertation

submitted to the
Combined Faculties for the Natural Sciences and Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by
Dipl.-Biochem. Birgit Pfalz
Born in: Görlitz, Germany
Oral examination: June 7th, 2016

Comparing bacterial gene networks based on high-throughput phenomics

Referees:

Prof. Dr. Lars Steinmetz

Prof. Dr. Victor Sourjik

"A mind
that is stretched
by a new experience
can never go back
to its old dimensions."

(Oliver Wendell Holmes Jr.)

To my loved ones.

Acknowledgements

First of all, I would like to thank my supervisor Dr. Nassos Typas for accepting me as his first PhD student. It was a great experience to see the lab grow and to be part of what by now is such a vivid group. I benefitted immensely from your extensive knowledge, constant support and truly never ending enthusiasm. Thank you Nassos for trusting me with this project and for caring so much! I really appreciate the time and effort you put into mentoring me.

I would like to thank my thesis advisory committee Prof. Dr. Victor Sourjik, Prof. Dr. Lars Steinmetz and Dr. Marko Kaksonen for their support, encouragement and suggestions and I would furthermore like to thank Prof. Dr. Lars Steinmetz and Prof. Dr. Victor Sourjik for accepting to review my thesis.

I wish to thank our collaborators Prof. Dr. Helene Andrews-Polymenis, Prof. Dr. Michael McClelland and Dr. Steffen Porwollik for providing the *Salmonella* knock out library prior to publication and for providing constant updates on the curation of the library.

As I fear we all remember this time, I would like to thank Nassos, Alex, Christina, Matylda and Lucía for participating in single colony purifying the entire library…

A huge thank you goes to Anja Telzerow. Anja spent countless hours pouring, moving, pinning and photographing plates and put up with my "new perfect system" on how to arrange plates every single time. For this and her well placed sarcastic comments that made me laugh even in the most stressful screening days I would like to say thank you.

Along those lines I would like to thank the staff from the media and cleaning kitchens, in particular Nadja Nepke, who was the most diligent aid for pouring plates we could have hoped for.

I owe my deepest gratitude to Dr. Marco Galardini, who led the computational implementation of the comparative analysis and helped out on many other ends. Thank you Marco for your time, diligence and motivation! Countless times I came to the lab with the words "So, Marco did some magic yesterday", because you had delivered all I asked for over night. Also, thank you for your patience in explaining how the magic works.

I would also like to thank Dr. Pedro Beltrao for helpful discussions and feedback on the comparative analysis and Prof. Dr. KC Huang for his great interest in the project, encouragement and input on the screen analysis.

I would like to acknowledge Shu-Yi Su for her help with Matlab, the E-MAP toolbox and for guiding me so patiently through my first baby steps in programming. Also thank you to Dr. Robert Nichols for answering all my questions about the *E. coli* screen and for providing scripts.

A big thank you goes to Alison Waller for establishing the ortholog matching, benchmarking pipeline and basically "fixing all our lists". Surprisingly, she never lost her patience with the (lab) kids asking for all these little things.

For fruitful discussions and lots of useful feedback I would like to thank "the screening people" of the lab, Lucía Herrera, Manuel Banzhaf and Ana Rita Brochado. I would also like to thank George Kritikos for putting so much thought and time into perfecting IRIS and even more so for not running away upon hearing "Georgiiiiino…I have a question…".

I would like to say thank you to Matylda Zietek for making all the follow up experiments possible in such a short time. Matylda spent many hours preparing strains and inoculating well plates, never loosing her smile and optimism and always ready for my next idea. She also helped me in the last huge round of screening.

I furthermore wish to thank Alexandra Koumoutsi for her helpful suggestions and feedback on the metformin follow up. Also, I would like to thank Mihaela Pruteanu for providing strains and showing me how to tame the anaerobic chamber. Furthermore I would like to acknowledge Lucía Herrera for providing *E. coli* chemical genomics data and help in plotting with R, Ana Rita Brochado for her neat growth curve analysis script and providing MIC data, and Lisa Maier for providing overexpression data.

I am extremely grateful to Lisa Maier, who supported me so much in writing this thesis. Her positive attitude, diligence and kindness made it so much easier. Thank you Lisa for supervising me, proofreading and always sharing your chocolate!

Most importantly I would like to thank all past and present members of the Typas lab including our beloved adopted postdocs - it's been a blast! I truly think I'm not exaggerating when I say that the atmosphere in our lab is unparalleled. I watched this lab grow over the past four years and every new member contributed to what I experienced as a fun, kind and supportive atmosphere. So thank you all

for countless coffee break chats, always helping out and the fun times in- and outside the lab!

I would like to highlight Christina, who helped me with everything when I first started. She shared her extensive knowledge and gave me lots of encouragement. I must thank Lucía, who became an amazing friend. I will keep tons of great memories from our time at EMBL. Thank you Lucy for always finding a new angle to look at things when I was frustrated, for the great times travelling together and for being there and listening sooooo many times! Furthermore, I wish to thank the Frenchies, Morgane and Bachir, for fun times partying, the most hilarious coffee breaks and teaching me some (arguably) useful French.

At this point I would like to once again thank Nassos (and Alex of course) for nourishing the great atmosphere in the lab by caring so much and enabling awesome lab retreats and get togethers.

One thing that certainly kept me sane during my PhD (although it was pointed out to me that I'm crazy about it) was dancing. Therefore, I would like to thank Philipp, Antje, Kris, Ali and Pieter Aart for amazing times and for supporting me in teaching. It gave me new perspectives on delivering any message and I believe this influenced how I approach presenting in science.

Of course I am deeply grateful to all my friends, the many news ones I made in Heidelberg as well as the ones from the good old days. I am particularly thankful to Henry, Lucía and Noorie. They have been there from the very beginning, have seen me at my best and worst, always found the right words and reminded me of what matters when I needed it.

I wish to thank Max, who had the courage to support my decision in the first place and who kept on supporting me no matter what. I would also like to thank his family for their kindness, encouragement and support.

I am so grateful to my parents, who supported all my decisions and always trusted I will figure things out. Truth is, I wouldn't be here without you. Thank you so much! I also want to thank my sister and her family as well as the rest of my family for their constant encouragement and support.

Lastly, I am deeply grateful to Philippe. Thank you for daring to "join the concert at the end" as you would say, it made a really big difference. Thank you for always listening and being there for me!

Abstract

Novel genes are being discovered at constantly increasing rates by sequencing bacterial genomes and bacterial communities. Gene function discovery has been lagging behind, but recent technological advances allow us to apply reverse genetics approaches on a genome wide scale.

In this study I profile the growth of more than 3800 gene deletion mutants of the pathogen *Salmonella* Typhimurium in more than 550 perturbations including physical stresses, nutrient limitation, antibiotics and host defense molecules. Analysis of gene-drug interaction scores reveal significant phenotypes for 75% of the tested mutants. The data set provides a number of novel biological inferences, linking genes of unknown function to known pathways and providing insights into drug mode-of-action, uptake and efflux.

Using similar high-throughput data available for *E. coli*., I provide the first comprehensive cross-species comparison of genetic networks in bacteria. Correlation analysis and detection of functional modules reveals broad conservation of cellular pathways and drug responses between *Salmonella* and *E. coli*. However, I also find intriguing cases of network rewiring and investigate how species-specific genes connect to conserved modules.

Lastly, I investigate the highly different resistance levels of *Salmonella* and *E. coli* to the type 2 diabetes drug metformin and determine the *Salmonella*-specific efflux pump SmvA as the major component conferring drug resistance. Furthermore, I identify more transporters capable of exporting metformin and examine their wiring into the cellular networks of *E. coli* and *Salmonella*. This analysis reveals that many enterobacteria may have the potential to develop resistance against metformin leading to important implications for diabetic patients.

Zusammenfassung

Sowohl durch die Genomsequenzierung einzelner Organismen als auch durch das Sequenzieren von bakteriellen Ökosystemen werden ständig neue Gene entdeckt, die Bestimmung ihrer Funktion hängt jedoch hinterher. Jüngste technologische Fortschritte erlauben es uns nun aber Methoden der reversen Genetik genomweit anzuwenden.

In der vorliegenden Studie habe ich das Wachstum von mehr als 3800 verschiedenen Deletionsmutanten des Pathogens *Salmonella* Typhimurium unter mehr als 500 verschiedenen Bedingungen charakterisiert, eine Methode bekannt als *chemical genomics*. Die getesteten Bedingungen beinhalteten unter anderem physikalische Stressoren, die Limitierung von verfügbaren Nährstoffen, Zugabe von Antibiotika oder Molekülen der Immunabwehr.

Signifikante Phänotypen wurden für 75% der getesteten Mutanten gefunden. Das Datenset liefert Einblicke in eine Großzahl bisher unbekannter biologischer Zusammenhänge und erlaubt es Gene unbekannter Funktion mit bekannten zellulären Reaktionswegen oder Proteinkomplexen zu verknüpfen. Für die hier untersuchten Medikamente können außerdem Rückschlüsse auf Aufnahme und Export sowie zelluläre Targets geschlossen werden.

Unter Zuhilfenahme eines ähnlichen, für *E. coli* verfügbaren Datensets präsentiere ich außerdem den ersten bakteriellen speziesübergreifenden Vergleich von *chemical genomics* Daten. Korrelationsanalyse und die Bestimmung funktioneller Genmodule offenbarten eine weitreichende Konservierung zellulärer Signalwege. Dennoch fanden sich auch interessante Beispiele von Neuvernetzung. Ich habe außerdem untersucht wie spezies-spezifische Gene mit den konservierten Modulen verbunden sind.

Zuletzt habe ich die stark unterschiedlichen Resistenzniveaus von *Salmonella* und *E. coli* gegenüber dem Typ 2-Diabetesmedikament Metformin erforscht. Ein *Salmonella*-spezifischer Effluxtransporter namens SmvA wurde als entscheidende Komponente für die erhöhte Metforminresistenz identifiziert. Außerdem habe ich weitere zelluläre Pumpen entdeckt, die in der Lage sind Metformin zu exportieren. Die Ergebnisse zeigen, dass viele Enterobakterien das Potenzial haben Resistenz gegen Metformin zu entwickeln, was wichtige Auswirkungen für Diabetespatienten haben könnte.

Table of contents

List of Figures

List of Tables

1 Introduction

1.1 Gene function discovery is lagging behind

The number of bacterial genomes and communities sequenced has exploded in the last decade. Currently, almost 5,000 prokaryotic genomes are completely sequenced and assembled. Furthermore, >50,000 are in progress (NCBI, Jan 2016) and this number is constantly growing with the vast majority stemming from metagenomics studies sampling microbes in complex environments such as the gut or the ocean. With this overwhelming gain in sequence information comes the discovery of novel genes with sequences unseen before, with no similarity to any known gene or domain.

This poses the question how researchers shall investigate the function of completely unknown genes and demands the implementation of high-throughput approaches able to match the pace of incoming sequence information. However, strictly speaking this problem is not a new one. To date even 1,600 *E. coli* genes (30% of its genome) remain of unknown function (Biocyc, Jan 2016).

Thus, utilizing cutting-edge technology to discover gene function in a high-throughput manner is imperative and should not only be applied to new, less known but also to well-studied organisms. By exploiting the extensive knowledge and experimental setups we have for model organisms, we will be able to establish new techniques faster. After all, filling the gaps of knowledge in our lab strains will allow for inferring the function of similar genes or their domains in other organisms.

1.2 High-throughput approaches to understanding gene function

Bacterial gene function has been classically investigated using forward or reverse genetics approaches. Forward genetics determine the underlying genetic basis for an observed phenotype, such as a fitness defect, the inability to infect or utilize certain nutrients. Once the responsible gene is found, the gene product can be expressed, isolated and biochemically characterized. This approach can be turned around by deliberately mutating (e.g. deleting, modifying, overexpressing)

a gene of interest to examine the resulting phenotypic changes, a method referred to as reverse genetics.

Whereas forward genetics are not suitable to address the directed search for function of all the newly discovered genes, advances in genetic engineering, our ability to track many strains simultaneously (e.g. by genetic barcoding) and improvements in high-throughput acquisition and quantification of biological readouts allows adaptation of reverse genetics approaches to address this issue.

However, modulating an uncharacterized gene often does not result in any severe phenotype. The difficulty often lies in redundancy and/or robustness of cellular networks. Bacteria live in extremely complex and often constantly changing environments. They need to adapt quickly and often more than one gene product is capable of fulfilling a specific task, as they cannot afford a single mutation to be detrimental. Furthermore, many genes carry out niche-specific tasks, only needed under rather extreme conditions, which the bacteria will not encounter when grown in standard laboratory conditions.

Several ways to overcome this hurdle can be considered. For example, knowledge of the gene sequence allows for recombinant expression and subsequent biochemical characterization *in vitro*. Also, the search for direct, known interaction partners can be a very powerful tool to link a gene product of unknown function to a known cellular pathway. Alternatively, cellular back up systems can be overcome by challenging the organism with multiple physical, chemical or genetic perturbations (or combinations of those) to evoke a phenotype.

1.2.1 High-throughput characterization of gene products

With the advent of whole genome sequencing the opportunities to explore gene expression, protein localization and function increased drastically due to the possibility to systematically tag (and thereby purify or visualize), overexpress, and mutate genes. Combined with constantly improving robotic platforms tremendous advances have been made in systematically studying gene expression (both on RNA and protein level), protein and RNA localization and biochemical activity of gene products in various organisms.

For instance, next-generation sequencing fueled the discovery of tracrRNA (*trans*-activating CRISPR RNA) as the third most abundant transcript class after rRNA and tRNA in *Streptococcus pyogenes* [1]. This was the missing piece in the

puzzle to finally determine the mechanism of adaptive immunity against viral infections in bacteria using CRISPRs (clustered regularly interspaced palindromic repeats).

In another example knowledge of the genome sequence was used to construct a GFP-tagged library of yeast proteins to determine subcellular localization of more than 4,000 proteins, representing 75% of the yeast protein [2]. Later on this atlas was expanded to investigate the dynamics of 5330 proteins under different external stresses [3].

Even biochemical activities of proteins can be investigated in a high-throughput fashion. McKellar et al. recently determined the ligand binding profiles of three chemoreceptors from the kiwifruit pathogen, *Pseudomonas syringae* pathovar *actinidiae* (*Psa*). The strain used encodes 43 predicted chemoreceptors, none of which had been characterized before. In a high-throughput fluorescence-based thermal shift assay the authors identified which of 95 signal molecules tested are recognized by each recombinantly expressed ligand binding domain [4].

Of course also computational efforts to reliably annotate the biochemical activity of unknown gene products based on sequence and structure similarity, or using active side prediction are constantly improving [5, 6]. This will allow for rapidly transferring newly gained knowledge to similar protein domains or orthologous genes in other organisms.

The studies above, despite representing just a few selected examples, impressively prove how information on mere existence, expression, localization or biochemical activity of gene products helps us understand their function. However, the interplay of proteins within the cellular network is complex. Understanding the biochemical function of a single protein is not enough to understand its true role in the dynamic system of a living cell. Regulation, feedback and interaction with other pathways have to be taken into account to gain a more detailed understanding of cellular processes.

1.2.2 Physical interaction mapping

Many cellular processes are performed by machineries composed of multiple proteins. Therefore, identification of direct protein-protein interactions (PPIs) is a powerful tool for studying gene function.

Two-hybrid techniques are commonly used to determine protein-protein interactions. The concept relies on bringing two fragments of a modular protein together resulting in measurable functionality, e.g. enzymatic activity or induction of GFP expression. One fragment of the modular protein is fused to the prey, whereas the second fragment is fused to the bait. Upon PPI between bait and prey the fragments complete each other, which can be detected in the chosen read-out. Early studies utilized two-hybrid strategies to systematically map protein-protein interactions in several viruses [7-9]. Later, assembly of the budding yeast genome resulted in extensive genome-wide studies to map protein-protein interactions using the same method [10-13].

Furthermore, two landmark studies made use of affinity purification followed by mass spectrometry (AP-MS) to reveal modularity of the yeast protein complex landscape [14-16]. A bait protein is coupled to an AP tag and co-purified with interacting partners, which are subsequently detected by MS. The advantage of this method is that proteins are expressed at endogenous levels and associated proteins can be identified as well. However, this method will miss transient interactions. Both, two-hybrid as well as AP-MS strategies, have been successfully applied on a global scale in *E. coli* [17, 18] and other bacteria [19].

An extremely powerful method to detect also transient and even conditionally induced PPIs is FRET (fluorescence resonance energy transfer). The technique measures interaction of two proteins both fused to different fluorescent proteins. Detections of the PPI is due to the energy transfer between the excited donor fluorophore and the acceptor fluorophore when brought into close proximity by the protein-protein interaction. Kentner and Sourjik applied this concept for measuring all interactions in the chemotaxis network of *E. coli* and investigating their response to chemotactic stimulation [20]. Although the scale in which FRET has been applied so far is comparably small [20, 21], improvements in automated microscopy platforms promise up-scaling in the near future.

Measuring PPIs gives insights into the structure and composition of protein complexes and can thereby be a useful tool to associate gene products of unknown function with known proteins. However, how protein complexes work together to build cellular pathways, crosstalk between those pathways and their responses to environmental cues remain largely elusive.

1.2.3 Genetic interaction mapping

As previously mentioned, reverse genetics are a classical approach to address gene function discovery. Genetic perturbations comprise deletion, disruption (by insertions), overexpression or point mutation of a gene, all of them potentially resulting in gain or loss of function. However, a single genetic perturbation rarely results in a severe phenotypic change [22]. The gene of interest might be expressed only under specific conditions, the product of a related gene might cover for the loss or feedback regulation might normalize the effect.

Thus, applying additional stress, e.g. a chemical or second genetic perturbation, might evoke a phenotype revealing the importance of the mutated gene, maybe even conditional essentiality. Before explaining how this reverse genetics strategy of investigating gene-gene, gene-drug or even drug-drug interactions facilitates gene function discovery, I will first outline the different kinds of interactions observed and how we can measure them in a high-throughput fashion.

Neutral, positive and negative interactions

A genetic interaction between two loci suggests that the resulting gene products influence each other's function and can be determined by measuring how the phenotype upon double perturbation differs from the expected combination of individual perturbations.

The assumption is that the fitness upon double perturbation is equal to the product of fitness upon each single perturbation. For example, inhibition of protein 4 in Figure 1 using a drug and simultaneous deletion of gene 1 affects independent cellular pathways. If the drug treatment by itself results in a growth inhibition down to 70% of wild-type growth and the gene deletion mutant reaches 60%, the expected fitness of the double perturbation would be 42% of wild-type growth. In this case we would speak of a neutral gene-drug interaction.

However, if the targeted processes are connected in the cellular network, aggravating or alleviating interactions can be expected. For instance, deletion of gene 2 in Figure 1 will have no further effect if the downstream protein 4 is fully inhibited by a drug. The combination of the two perturbations will not result in the expected multiplicative fitness. Instead, the cells will behave just like upon drug treatment alone, thereby actually better than expected, which we refer to as a positive (or alleviating) interaction. Several studies have shown that positive

interactions frequently indicate gene products acting in the same cellular complex or pathway assuming that removal of the second component has no additional effect [23, 24].

In contrast, when two targeted pathways lead to a common product (e.g. chemically inhibiting protein 4 and deleting gene 3 in Figure 1 the effect on the cell will be more severe than expected, which we refer to as a negative (or aggravating) interaction. If the common product is essential and there are no other pathways in place to yield this product the combination of perturbations will even result in synthetic lethality.

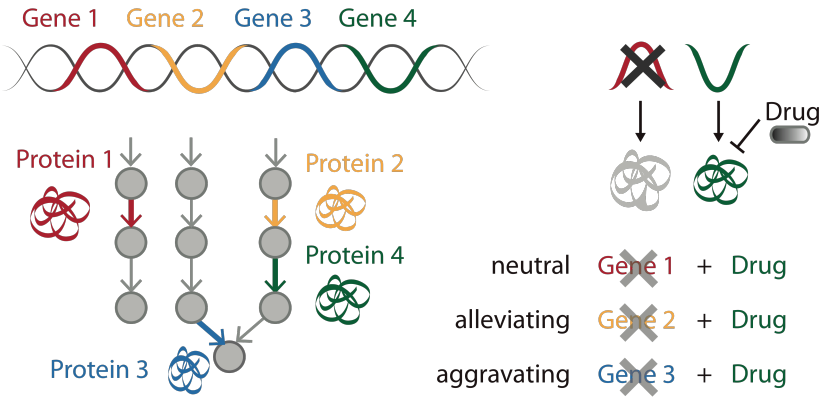

Figure 1: Neutral, alleviating and aggravating gene-drug interactions.
Deletion of a gene upon simultaneous inhibition of a second gene product (protein 4) using a drug can result in different gene-drug interactions depending on the functional dependencies between both genes. Typically, when independent pathways are targeted the interaction is neutral. In contrast, perturbation of the same pathway usually leads to alleviating (positive), perturbation of parallel pathways to aggravating (negative) interactions. Adapted from Brochado and Typas [25].

Positive and negative interactions follow the principles of a long known phenomenon called epistasis. Fisher described "deviations from the expected quantitative combination of independently functioning genes as epistacy". Another typical use of the word epistasis is described by Bateson as "situations in which the activity of one gene masks effects of another locus, allowing inferences about the order of gene action", so only referring to positive interactions (for both quotes see Box 1 in C. Boone's review on Exploring genetic interactions and networks with yeast [26]).

Measuring large scale genetic interactions in high throughput

When the first complete systematic gene deletion libraries became available in yeast in 2002 [22], measuring genetic interactions was moved to a genome-wide scale pioneering tremendous advances in gene function discovery. The techniques

to measure fitness in high-throughput were largely developed in studies investigating gene-gene interactions but are applicable to other designs as well.

In a landmark study, Tong et al. describe the development of synthetic genetic array (SGA), a technology to generate haploid double mutants and screen for synthetic lethality [27]. The screen was later expanded to cross 132 query mutants with the entire viable yeast deletion collection, so ~4700 mutants [28]. In SGA the phenotype of each mutant arrayed on a solid agar surface is assessed individually, however this approach was focused on detection of synthetic sickness or lethality (so negative interactions).

The next step came with the introduction of the E-MAP (epistatic mini-array profile) strategy [29, 30]. In this method the colony size of mutants arrayed in a high-density format is measured precisely, thereby allowing for detection of negative as well as positive genetic interactions in a high-throughput manner.

In parallel to SGA, an alternative method called dSLAM (diploid-based synthetic lethal analysis with microarrays) was developed to assess the fitness of mutants in competitive growth experiment. Each mutant is labeled with a genetic barcode and therefore the mutant's abundance in a pool can be detected by microarray. The advantage of this method is the enormous throughput that can be reached due to the pooling approach and by now the analysis pipelines have improved to provide a quantitative interaction score [31, 32].

All methods have been extensively used to study genetic and chemical genetic interactions in yeast [28, 30, 33, 34]. The E-MAP approach was afterwards adapted for use in *E. coli*, resulting in pioneering studies to build the first large-scale interaction networks of bacterial cells [35-40].

Genetic interactions provide insights into bacterial biology

In principle, all kinds of perturbations can be combined to reveal gene-gene, gene-drug, or drug-drug interactions, all of them giving unique insights into cellular biology. Work in this field has been pioneered in yeast, but recently also cellular networks of bacteria are being explored using the same approach.

When by 2006 the first systematic deletion libraries were available in *E. coli* [41, 42], two methods, GIANT-coli and eSGA, have been reported to construct double deletion mutants in this bacterium [35, 37]. In both methods Hfr conjugation is employed to transfer a mutation from a donor strain to a recipient strain carrying a second, differently marked mutation. The fitness of each strain is assessed by measuring growth on a solid agar surface.

A first genetic interaction map was published in 2011 and focused on interactions of genes involved in cell envelope biogenesis [38]. Considering that membrane proteins are usually difficult to purify and capture in PPI studies, this was a particularly interesting focus for the first genetic interaction study and indeed the authors report new players involved in outer membrane integrity, lipopolysaccharide (LPS) transport and formerly unknown regulatory circuits [38]. However, construction of targeted double deletions in high throughput is challenging in other bacteria lacking the Hfr mating system. One approach is to create random transposon insertion libraries in the background of a single deletion, but then throughput is greatly diminished [43]. However, the recently developed CRISPRi technology promises advances in combinatorial power in the near future [44].

A much simpler, yet arguably equally powerful method is to combine a gene deletion with a chemical perturbation, also referred to as chemical genomics and following the principles of gene-drug interactions discussed above. Chemical genomics studies have been reported in a number of bacteria, including *E. coli, Staphylococcus aureus, Streptococcus pneumonia, Francisella novicida* and *Shewanella oneidensis* to name a few [36, 45-50]. All studies led to gene function or even pathway discovery, e.g. the identification of phosphofructokinase and an alternative proline biosynthesis pathway in *F. novicida* [49].

Chemical genomics furthermore bring the advantage that not only gene function and wiring but also drug mode of action can be investigated. Amongst others the mechanism of kibdelomycin, a drug efficient against multidrug-resistant *S. aureus* (MRSA), and the mechanism of synergy between trimethoprim and sulfa drugs have been discovered using the data sets mentioned above [36, 50]. This aspect of chemical genomics is especially relevant considering studies conducted in bacterial pathogens. Another advantage is the possibility to probe complex stresses and mimic less defined, natural habitats in order to assess a gene's importance in context of the multiple environmental cues encountered by the bacterium. Once the stress applied becomes more complex, the simple concept of neutral, positive and negative gene-drug interactions to map relationships between the targeted pathways seems less applicable, because the target of a complex stress cannot be defined. However, comparing interaction profiles of different mutants across conditions can reveal functional relation.

The power of genetic interaction profiles

All large-scale genetic interaction studies mentioned above make use of the same powerful concept referred to as guilt by association. As all mutants (typically gene deletions) are sampled against a variety of chemical stresses or second genetic perturbations a profile across conditions, a so-called phenotypic signature, is obtained for each mutant. It turns out that genes with similar phenotypic signatures often act in the same cellular pathway, functional module or even protein complex [28, 36]. Therefore, it is possible to associate genes of unknown function to known cellular pathways, functional modules or even protein complexes. Typas et al. discovered two novel regulators of peptidoglycan synthesis aided by hierarchical clustering of *E. coli* chemical genomics data, to just name one example in which the guilt by association concept has been successfully applied [51].

Focusing on interaction profiles rather than single interactions also allows for comparison between cellular complexes and pathways on a network level in order to probe for crosstalk and regulatory relations. Furthermore, building networks based on chemical genomics can be especially useful to investigate drug uptake, targets and efflux. Considering that phenotypic signatures of drugs span across thousands of genes, whereas mutant signatures usually only range over hundreds of conditions, collapsing the long list of genes into functional modules and building networks based on their interaction with the tested drugs can positively reduce complexity in order to reveal higher order processes [36].

A more global visualization of interaction data will furthermore facilitate cross-species comparison.

1.3 Comparing large-scale interaction data from different organisms reveals network rewiring

Comparing genetic interaction data from different species has begun to shed light on how changes in DNA sequence result in changes in the cellular network and consequently phenotypic variation.

In a landmark study Dixon et al. compare genetic interactions between budding and fission yeast (*Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, respectively) [52]. The two organisms share about 75% of their genes, but are separated by one billion years of evolution [53]. One distinct difference is a

genome duplication event in *S. cerevisiae,* resulting in paralogs for many genes, where *S. pombe* has only singletons. Thus, crucial differences in important cellular pathways can be observed. For instance, as the names suggest the division mechanism is different in both organisms. Also, *S. cerevisiae* as opposed to *S. pombe* is missing complexes involved in crucial regulatory tasks such as the pre-mRNA splicing or the RNA interference machinery. Hence, substantial differences in cellular network organization between the two organisms are expected. Focusing on synthetic sick and lethal interactions Dixon et al. report 29% conservation of genetic interactions [52]. This means the majority of interactions observed are species-specific confirming network rewiring and revealing interesting species-specific differences.

In parallel, the investigation of E-MAP studies from orthologous gene pairs in *S. cerevisiae* and *S. pombe* led to the notion that although negative interactions between genes involved in parallel cellular processes show a certain level of conservation, the overall phenotypic signatures and especially positive interactions between genes coding for physically interacting proteins were even more conserved [24]. Thus, conservation at the level of functional modules is rather high (even human genes have been shown to be able to replace their yeast counterparts [54]), whereas cross-talk between those modules is less conserved (compare Figure 2).

This observation suggests that not only loss, gain or change of function of certain genes is responsible for evolutionary changes and speciation events, but far more so the changes in how genes interact, how they are wired into the cellular network and how entire functional modules get repurposed [55]. Intuitively, this idea also explains the vast phenotypic variation that can be achieved even by little changes in DNA sequence.

Supporting this concept Tischler et al. used RNA interference to perturb hundreds of *C. elegans* gene pairs orthologous to those showing synthetic lethality in *S. cerevisiae.* The conservation of essential genes between those organisms is comparably high, however, the authors estimated that maximum 5% of synthetic lethal interactions between non-essential genes are conserved [56].

Taken together, previous studies suggest that comparison of large-scale genetic interaction data has the power to reveal important features of cellular network organization, species-specific gene functions and will also give us a better understanding of how even highly conserved core processes interact and adapt to environmental changes.

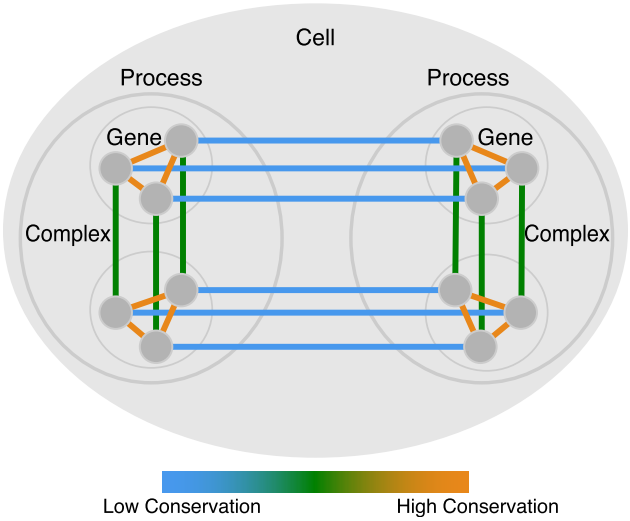

Figure 2: Conservation between functional modules is higher than conservation between cellular processes.
Genetic interactions between gene pairs whose products act in the same complex show the highest level of conservation (orange), whereas interactions between genes involved in the same biological process are less conserved (green). Interactions between genes acting in distinct biological processes are poorly conserved (blue). Adapted from Ryan et al. [55]

1.4 *Salmonella* and *E. coli* – cousins or worlds apart?

Comparing large-scale interaction data for bacteria has been lagging behind, although it is of crucial interest. As bacteria are haploid and in many cases easily genetically modified, creation of e.g. chemical genomics data sets is comparably easy. Since a single gene deletion results in complete removal of the gene product it is possible to assess conditional essentiality and find an essential common core genome between organisms. Furthermore, considering that we frequently find evolutionary related pathogenic and non-pathogenic bacterial strains (sometimes even within the same species), comparison of large-scale interaction might give insights into how rewiring and repurposing of cellular pathways and regulatory circuits contributes to the phenotypic variation resulting in such distinct life-styles. It will help us understand how e.g. horizontally acquired virulence factors integrate in the cellular network, but also how existing pathways adapt to new environments like the infected host. Lastly, comparing chemical genomics in particular, allows for investigating differences in drug response between bacteria and might therefore contribute to the design of more narrowly targeted anti-bacterial therapies, which is necessary to keep the development of multi-drug resistant strains to a minimum.

The most comprehensive bacterial chemical genomics data set exists for *E. coli* [36]. *Salmonella* is one of *E. coli's* closest relatives and a gene deletion library became recently available [57] allowing for creation of a comparable chemical genomics data set in this organism (this study). Whereas the species *E. coli* consists of commensal and pathogenic strains, *Salmonella* is almost purely pathogenic with the distinct feature of being able to survive and replicate inside of host cells. As the two bacteria exhibit such significant differences in life-style despite their comparably close relation, comparing their cellular networks might shed new light into adaptation of cellular processes to new niches.

1.4.1 Evolution of *Salmonella enterica* subspecies *enterica*

The genera *Salmonella* and *Escherichia,* both part of the gamma-proteobacteria class and *Enterobacteriaceae* family, represent Gram-negative, rod-shaped bacteria with a facultative anaerobe life style. *Salmonella* and *E. coli* diverged from a common ancestor approximately 100 million years ago [58].

Since the divergence of *Salmonella* from *E. coli*, virulence has evolved in three phases [59, 60].

First came the acquisition of *Salmonella* pathogenicity island 1 (SPI-1) by horizontal gene transfer. SPI-1 encodes a set of virulence factors involved in invasion of epithelial cells in the intestine, recruitment of neutrophils and intestinal fluid secretion. Therefore, as demonstrated by experiments in mice, mutations in SPI-1 genes attenuate virulence of Salmonella upon oral infection but have no attenuating effect after intraperitoneal injection. The presence of SPI-1 distinguishes all lineages of the genus *Salmonella* from *E. coli.*

The acquisition of *Salmonella* pathogenicity island 2 (SPI-2) by horizontal gene transfer represents the second major evolutionary event forming two distinct species, *Salmonella enterica* and *Salmonella bongori*. *S. bongori* lacking SPI-2 is a pathogen of cold-blooded animals, most commonly found in reptiles. In *S. enterica* SPI-2 genes are involved in systemic spread of the infection and intracellular survival. Thus, mutations lead to attenuated virulence even upon injection of the bacteria into the peritoneal cavity of mice.

The species *S. enterica* is further divided into several subspecies: *enterica* (subspecies I), *salamae* (II), *arizonae* (IIIa), *diarizonae* (III), *houtenae* (IV), and *indica* (VI). Whereas most subspecies harbor bacteria associated with cold-blooded animals, the formation of *Salmonella enterica enterica* resulted in a

drastic host range expansion. Strains of this subspecies are most frequently isolated from avian and mammalian hosts. This implies that divergence of *S. enterica enterica* involved the evolution of factors allowing the bacteria to overcome a more complex host immune system.

Members of the subspecies *enterica* are further classified into more than 2300 serological variants, so called serovars. The classification is based on the presences of antigens following the Kauffman-White scheme [61]. Interestingly, the host range varies widely between serovars. For instance, *Salmonella* Typhi (i.e. *S. enterica* subspecies *enterica* serovar Tyhpi) is restricted to higher primates and humans, whereas *Salmonella* Typhimurium can infect a broad variety of avian and mammalian hosts, including mice and humans [59].

1.4.2 *Salmonella* Typhimurium pathogenesis

Non-typhoidal *Salmonella* (NTS) strains, such as *Salmonella* Typhimurium and *Salmonella* Enteriditis are the most common cause of food-borne disease worldwide with case numbers estimated as high as 1.3 billion per year including 3 million deaths [62].

In humans *Salmonella* Typhimurium (STm) causes acute gastroenteritis characterized by fever, diarrhea, intestinal cramping and neutrophil infiltrates. In otherwise healthy individuals this is a self-limiting disease. In immunocompromised, very young or elderly patients, however, the infection can spread to a systemic level and result in life-threatening bacteremia. Especially in sub-saharan Africa NTS strains have emerged as the dominant cause of bloodstream infections in adults and children, with the most prominent risk factors being HIV infection, malaria and malnutrition. All patients present with severe fever; hepatosplenomegaly and respiratory symptoms are common as well, whereas intestinal symptoms are rare. Case fatality has been estimated to 20-28% and many multi-drug resistant strains are emerging [63-65].

The infectious cycle of STm has been the subject of excellent reviews [66-70]. In brief, after ingestion of contaminated food or water, *Salmonella* is able to survive the acidic pH of the stomach and proceed to the small intestine. There it passes the mucous layer and invades the intestinal epithelium. The ability to enter non-phagocytic epithelial cells is mediated by a type III secretion system (T3SS) encoded on SPI-1. The needle-like complex allows the bacteria to inject effector

proteins into the host cell, where they cause major actin rearrangements resulting in membrane ruffling and uptake of the bacteria [70].

Once inside the host cell *Salmonella* resides in a membrane-bound compartment, the *Salmonella* containing vacuole (SCV). Effectors secreted through the vacuolar membrane by a second T3SS encoded on SPI-2 prevent fusion with lysosomes allowing for intracellular survival and replication [66]. After crossing the epithelial barrier the bacteria reside in phagocytic cells such as macrophages, dendritic cells and neutrophils.

Interactions with the host cells trigger a series of events resulting in gut inflammation. Infected host cells secrete cytokines such as interleukin (IL) 18 and 23 which help to amplify the innate immune response by activating T cells to secrete interferon-gamma, IL-17 and IL-22, which in turn stimulate antimicrobial responses in macrophages and epithelial cells. For instance, IL-17 triggers the secretion of chemokines by epithelial cells resulting in recruitment of neutrophils to the inflammation site. In sub-epithelial tissues neutrophils prevent extracellular growth of *Salmonella* and therefore help avoiding systemic dissemination. IL-22 induces the secretion of antimicrobial peptides such as calprotectin and lipocalin-2, designed to deprive luminal bacterial pathogens from zinc and iron, respectively.

In consequence to the initiated immune response the bacteria are finally cleared from sub-epithelial tissues. Furthermore, during the course of inflammation the intestinal lumen is rendered a hostile environment due to the induced presence of reactive oxygen species (ROS), and antimicrobial peptides as well as the diminished nutrient availability caused by diarrhea. All of these represent general antibacterial defenses, therefore it is not surprising that the resident microbiota is damaged. Recent reports demonstrate that *Salmonella* Typhimurium, however, has developed several strategies to overcome the antibacterial defenses and exploit the created niche to outgrow the impaired microbiota (reviewed in [69]). For example, most commensal enterobacteria produce enterobactin, an iron chelator inhibited by the antimicrobial peptide lipocalin-2 produced during inflammation. *Salmonella* produces an additional iron chelator, salmochelin, which is not bound by lipocalin-2 and is therefore not affected by the iron deprivation [71]. Additionally, tetrathionate is produced in the gut lumen by oxidation of the naturally occurring thiosulfate during inflammation. Studies demonstrated how *Salmonella* can utilize tetrathionate as an electron acceptor for anaerobic respiration and how this allows utilization of

ethanolamine as a carbon source not accessible for other bacteria [72, 73]. This confers an advantage to *Salmonella*, because most other bacteria are restricted to comparably less efficient fermentation.

In conclusion, a subset of the *Salmonella* bacteria arriving in the terminal ileum invades the intestinal epithelium resulting in acute inflammation. In turn this gives an advantage to the bacteria staying behind in the lumen, allowing them to thrive and outcompete the resident microbiota, thereby promoting the transmission via the fecal/oral route.

1.4.3 Genomic and phenotypic comparison of *S.* Typhimurium and *E.coli*

In this study I create a chemical genomics data set for *Salmonella* Typhimurium using the recently published single gene deletion collections [57], which were kindly provided by Helene Andrews-Polymenis and Michael McClelland prior to publication. The data set is supposed to serve as a resource to investigate gene function as well as drug mode of action but also to allow for comparison to large-scale interaction data in *E. coli.* The available chemical genomics data set in *E.* coli is large and both organisms are well studied. Especially for *E. coli* annotation and experimental evidence for gene function are extensive allowing for benchmarking and ensuring comparability of the data sets.

Furthermore, recent comparative genome studies provide insights into genes shared between strains of the *E. coli* and *S. enterica* species determining a core- and pangenome. For instance, in 2011, Jacobsen et al. compared 35 *S. enterica* genomes alongside two different *E. coli* genomes. The authors report a *Salmonella* core genome of approximately 2,800 gene families shared by all *Salmonella* strains investigated. The complement of gene families detected across all strains accounts to 10,000 and is referred to as the *Salmonella* pangenome. In a similar study Gordienko et al. demonstrate that despite extensive horizontal gene transfer between *E. coli* and *Salmonella*, the latter maintains a stable species-specific gene pool, whereas *Shigella* and *E. coli* share a common gene pool consistent with previous phylogenetic analyses [58].

From studies like these we can derive information about genes specific to the genus, species, subspecies, serovar and even strain level. This will aid the investigation of how evolutionary recent or ancient genes contribute to phenotypic responses to external stresses. One assumption may be that

evolutionary young players are frequently involved in very niche-specific tasks and hence relevant only under very specific conditions (resulting in rare observations of significant phenotypes for those genes in chemical genomic screens). Furthermore, it will be interesting to test whether genes of the core genome shared between *Salmonella* and *E. coli* vary in their phenotypic response. This would indicate species-specific rewiring or repurposing of core cellular processes to adapt to a new environment, e.g. the *Salmonella* containing vacuole, or to respond to the acquisition of new genes. Differential regulation of homologous genes also brings implications for the clinical setting. For example, it has been shown to be involved in LPS modifications conferring resistance to the antibiotic polymyxin B to *Salmonella* [74]. In general, the cell envelope is of particular interest as it constitutes communication interface and at the same time barrier to the external environment, allowing interactions with e.g. the host or other microbes while conferring protection from unwanted molecules such as antibiotics. However, due to their biochemical properties envelope components are difficult to study. Genetic interactions of genes coding for envelope components or their synthesis machineries are therefore of great interest.

1.5 Aim of the study

The aim of my PhD study was to create an extensive gene-drug interaction map for the pathogen *Salmonella* Typhimurium to be used as a valuable resource for studying gene function, drug mode of action and wiring of cellular processes.

In the following chapters I describe how I acquired the chemical genomics data set with more than 4 Mio gene-drug interactions and how it can be utilized to investigate gene function and drug mode of action. Furthermore, I describe how comparison of chemical genomics data from two different organisms can reveal network conservation as well as rewiring. Lastly, I investigate an intriguing difference in *Salmonella's* and *E. coli's* capacity to deal with the widely prescribed type 2 diabetes drug metformin.

The completion of this project was of course a highly collaborative effort. Therefore, I included a disclaimer at the end of every chapter to ensure clarity on contributions of colleagues and collaborators.

2 Creating an extensive gene-drug interaction map for *Salmonella* Typhimurium

In the following chapter I describe the creation of the chemical genomics data set for *Salmonella* Typhimurium. I used targeted single-gene deletion mutants arrayed on a solid agar surface containing sub-inhibitory concentrations of various chemicals. Colony size served as a proxy for fitness to determine gene-drug interaction scores. After extensive quality control and curation of the mutant library the data set yields more than 4 Mio gene-drug interactions, providing significant phenotypes for many genes of unknown function. I furthermore show that comparing interaction profiles between mutants gives insights into their functional relation.

2.1 Study design

2.1.1 Method selection

Chemical genomics are highly flexible in kind and scale of perturbations to be tested. Furthermore, they did not require any additional manipulation of the existing *S.* Typhimurium knock out strains. Therefore, I preferred this approach to measuring genetic interactions in double deletion mutants (i.e. gene-gene interactions), which would have required introduction of a high-throughput double mutant generation pipeline. Furthermore, the aim of this study was to probe the role of each gene in the great variety of conditions *Salmonella* encounters in its natural habitat and the infected host, which can be more easily reflected by physical and chemical perturbations as compared to a second genetic perturbation.

As robotics were available I preferred the arrayed setup to a pooled approach to avoid biases due to initial strain abundance and fitness [25]. Growing the bacteria on a solid surface allowed me to work in a high-throughput manner by selecting a high-density format of 1536 mutants per plate, whereas in liquid culture the maximum reasonable format would have been 384 mutants per plate.

Also, the higher throughput not only significantly reduces the time necessary to perform the experiments, but also costs.

As a proxy of fitness I measured colony size after 12-14 h. This has been proven to be an accurate measure of growth as time course and end point measurements of colony size correlate (George Kritikos, unpublished data). Based on this end point colony size I calculated an interaction score for each gene-drug combination in order to obtain the chemical genomics data set.

2.1.2 Genomic space probed

The two single-gene deletion (SGD) libraries used in this study were provided by Helene Andrews-Polymenis (Texas A&M) and Michael McClelland (UCI) prior to publication. By design the collection does not include essential genes and amongst others excluded the following classes: rRNAs, tRNAs, ~100 structural elements of active lysogenic phage and protein coding sequences under 100 aa in size if not annotated in the genome of *S.* Typhimurium strain LT2 [57].

The mutants were created in the background of *Salmonella enterica* servovar Typhimurium 14028s, referred to as wild-type *Salmonella* in the following. Replacing the gene of interest with an antibiotic resistance cassette using the classical lamba-Red recombinase system yielded the desired mutants [75]. Primers were designed to preserve the first and last 30 bases of each targeted sequence in order to reduce polar effects on neighboring genes. The kanamycin (Kan) marker was placed co-directional with the deleted gene whereas the chloramphenicol (Cm) cassette was placed in opposite direction [57]. Both cassettes carry promoters at the end to ensure transcription of the downstream genes. Notably, as bacterial genes are organized in operons, this might result in polar effects due to transcription of the downstream gene at unphysiological levels or production of antisense RNA against the upstream gene in case of the oppositely placed Cm cassette.

We single colony purified the two collections to remove any potential contaminants and ensure an isogenic population. Next, we rearranged them on a minimal number of 96 well plates so all mutants finally fit on five plates in the desired 1536 array format. The final data set presented here consists of 3834 unique mutants covering more than 80 % of the *S.* Typhimurium genome (*S.* Typhimurium strain LT2, which is closely related to the 14028s strain used here but better annotated, has 4672 genes, Biocyc database, Dec 2015). More than 3000

mutants are present in both libraries, leaving around 400 mutants unique to each collection.

2.1.3 Chemical space probed

I selected the chemicals used in this study to impact a large variety of cellular processes and mimic different environments encountered by *Salmonella* (Figure 3A).

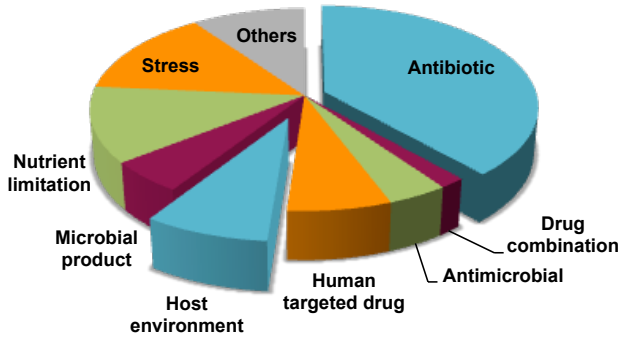

Figure 3: Chemical space probed.
Grouping of 175 unique stresses subjected to further analysis.

Antibiotics target central cellular processes such as replication, gene expression and cell wall synthesis. Thus, challenge with sub-inhibitory concentrations of antibiotics will reveal genes involved in these essential pathways and furthermore result in strong phenotypes for mutants of crucial players. As gene-drug interactions are generally rare, those strong phenotypes will provide power for cluster and network analysis.

Chemical genomics proved useful to study not only gene function but also drug mode of action, because comparison of phenotypic signatures across genes gives insights into common or distinct secondary targets, uptake and efflux mechanisms. Thus, besides including a wide range of antibiotic classes, I often test members of different generations of the same class. Furthermore, synergies and antagonisms between drugs, or between drugs and components of the patient's diet are often poorly understood. Keeping this in mind and taking into account other projects ongoing in the lab, I included a few antibiotic combinations as well as combinations of antibiotics with the common food additive vanillin.

During the passage through different environments *Salmonella* encounters stresses like nutrient limitation, changes in osmolarity, pH (stomach acid) and temperature (e.g. inside host compared to outside). Thus, I included general stresses, such as changes in salt concentration, pH and temperature. I also test

minimal media with different carbon sources. Minimal media conditions allow us to assess nutrient requirements for the tested mutants. Furthermore, strong phenotypes for auxotrophs will add power when clustering algorithms are used to visualize chemical genomics data.

Inside the host bacteria are furthermore exposed to other microbes (also relevant outside the host), the compounds they produce as well as host hormones, metabolites, components of the host defense mechanisms, and even drugs consumed by the human host. To sample these more complex stresses I included conditioned media derived from other bacteria, host hormones, human targeted drugs as well as minimal media conditions designed to mimic the environment in the *Salmonella* containing vacuole [76].

Initially I attempted to test 200 unique chemicals, many of them in up to four different concentrations resulting in 750 single conditions. A number of conditions had to be excluded due to solubility limits or insufficient growth of *Salmonella.* Finally, 175 unique stresses (Figure 3A) represented by 585 single conditions were subjected to further analysis.

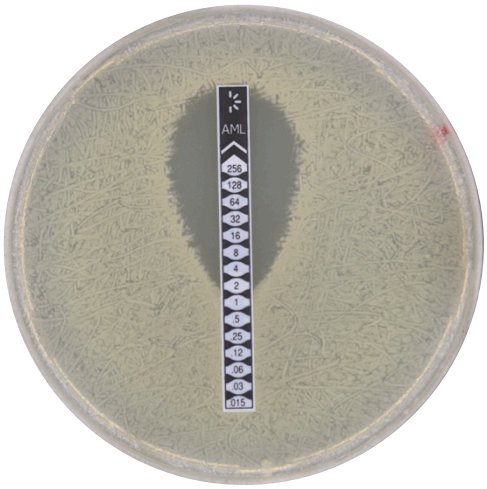

Figure 4: MIC test strip for Amoxicillin.
Wild-type *S.* Typhimurium was spread as a lawn and the MIC test strip was placed on top immediately. After over night incubation at 37°C the MIC can be read at the intersection of the halo with the MIC test strip, in this case ~1.5 µg/ml. The highest concentration screened for Amoxicillin was 0.5 µg/ml.

As every substance had to be applied in sub-inhibitory concentrations my first step was to determine the minimum inhibitory concentration (MIC) for each chemical. For most antibiotics MIC test strips were commercially available (Figure 4). For the remaining compounds I determined the MIC by literature research (also taking into account data on related organisms, e.g. concentrations tested in the chemical genomics study in *E. coli* [36]) or manual testing. For most

chemicals I screened four different concentrations, the highest one usually not exceeding 50% MIC.

2.1.4 Experimental layout

In the following text I will use the terms Kan library and Cm library for the part of the single gene deletion collection carrying the kanamycin or chloramphenicol resistance marker, respectively. I arranged both libraries on a total of five different 1536 plates, the first two containing only Kan mutants, the second two containing only Cm mutants and the last one consisting of half Kan and half Cm mutants. I refer to these plates as the five library plates.

A single condition describes one defined perturbation, e.g. one specific concentration of a chemical, one specific carbon source in minimal media or one specific temperature tested. I measured four replicates for every library plate under every condition.

In the results below I will therefore frequently refer to three different dimensions: mutants (on five library plates), conditions and replicates.

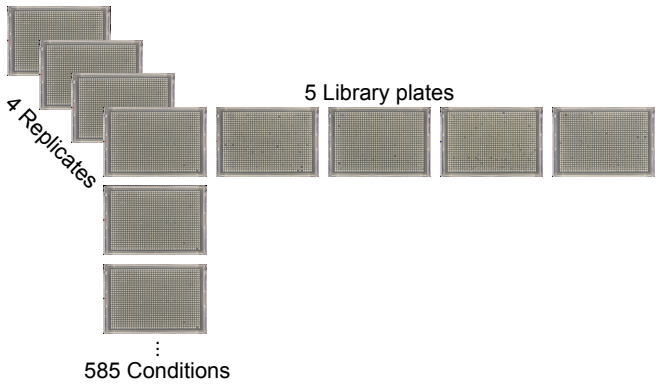

Figure 5: Screen layout.
All mutants are combined on five different library plates in 1536 format. Every single condition was measured in four replicates.

2.2 Data acquisition

2.2.1 Workflow

To obtain the test plates we added the dissolved chemical to 2% LB-agar prior to pouring the plates. At this step some chemicals had to be discarded due to solubility limits.

Glycerol stocks of the mutants were kept in 384 well plates at -80°C and all of the transfer steps described below were carried out using the Singer Rotor. To arrange the library plates in 1536 format we transferred the mutants from a thawed well plate to an agar plate and in a second step combined four different 384 plates into one 1536 plate. These 1536 plates served as sources for transferring the bacteria to the test plates and were maintained for up to one month by re-arraying on LB agar (see also Maintenance of library plates, page 90). In the following I will use the term batch for all conditions tested from one initial arrangement of frozen stocks into 1536 format (Figure 6).

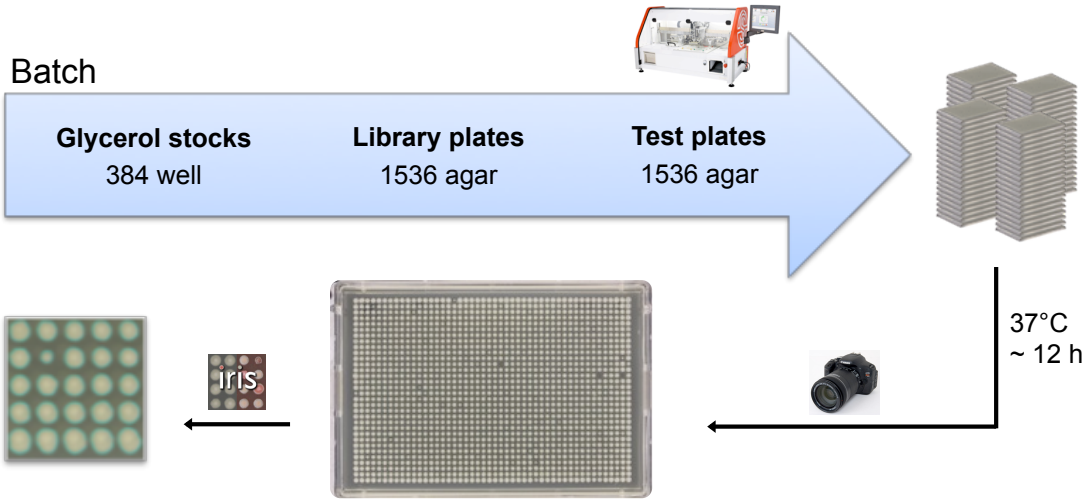

Figure 6: Data acquisition workflow.
Mutants were stored as glycerol stocks in 384 well plates at -80°C. Using the Singer Rotor the bacteria were transferred to agar plates and arranged in 1536 format. Next, the mutants were transferred to agar plates containing the chemical perturbation. All conditions tested from the same initial arrangement of five library plates in 1536 format belong to the same batch. After incubation at 37°C a high-resolution picture was obtained and the colony size was measured as area covered by the colony in pixels using the Salmonella growth profile of IRIS (developed by George Kritikos, unpublished data).

Again the Singer Rotor was used to transfer the libraries from source to test plates containing the chemical perturbation. After incubation at 37°C for an average duration of 12 h a high-resolution picture was taken. When the bacteria were growing generally slow in a specific condition a longer incubation time was allowed. However, this time cannot be extended to more than 24 h, because plates dry out and even if kept wet the growth dynamics between outer and inner part of the plate will differ too much to be suitable for the analysis pipeline used. Thus, some conditions had to be discarded at this step due to insufficient growth, majorly a series of carbon sources in M9 minimal medium and anaerobic conditions.

Using the pictures I measured the colony sizes using the Salmonella growth profile of IRIS (software developed by George Kritikos, manuscript in preparation) as the area covered by the colony in pixels.

2.2.2 Troubleshooting

After a small pilot screen of 40 conditions I calculated general parameters such as distribution of median plate colony sizes, plate interquartile ranges (IQRs) and replicate correlation. It became evident that poor replicate correlations usually involved the first plate pinned from each source (see correlations involving replicate A in Figure 7). In the first transfer from a new source plate the plastic pad carries a lot of material to the target plate influencing the inoculum strongly.

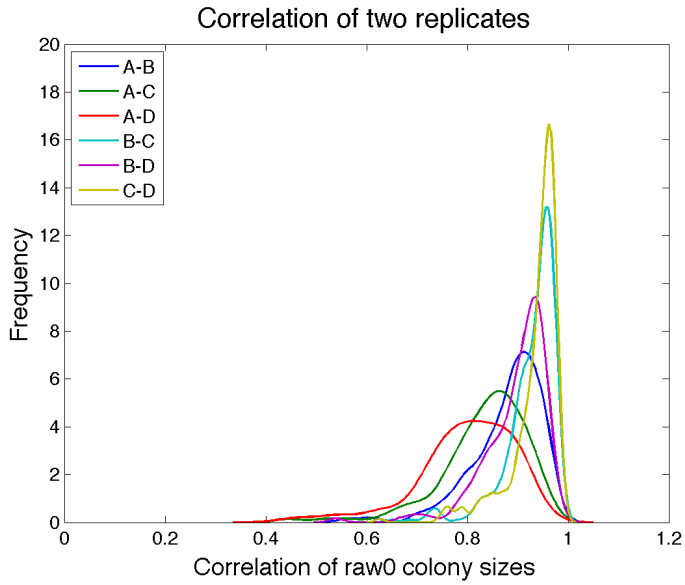

Figure 7: The first plates pinned from a new source show an inoculum effect.
Pairwise Pearson replicate correlation across conditions and library plates. Pairs involving the first plate pinned (A-B, A-C, A-D) show lower correlation than the other pairs.

In a test array I investigated correlation between consecutively arrayed plates with the premise of four replicates, so calculating overall correlation of four consecutive plates starting on the first, second, third plate and so on. No significant differences could be detected from the third plate onwards, meaning that plates 3-6 will yield similar correlations as plates 4-7 and so on. Plates 1-4 and 2-5 show an overall decreased correlation indicating that the first two plates pinned from a fresh source show the inoculum effect (data not shown). Thus, I adapted the procedure by pinning two dummy plates (plain LB-agar) before each set of test plates and discarded the data from this first batch.

Initially, replicates for each condition were pinned from the same source plate to keep logistics simple and to avoid difference in storage time of the chemical plates. However, during the course of experiments, reproducible technical biases proved to be a confounding factor in the analysis. If replicates come from the same source the risk of reproducing technical artifacts and therefore detecting them as gene-drug interaction is higher. Thus, after becoming aware of this problem, I spread the array of replicates across different sources and days.

2.2.3 Quality control

First, I inspected all plates visually. Plates with insufficient overall growth, or with severe pinning errors (uneven agar plates can result in incomplete pinning) were removed immediately. The remaining plates were subjected to a thorough quality control procedure, for which I set up an automated pipeline applicable and used for other screen data in the lab.

Quality control measure	Threshold applied
Replicate correlation	< 0.8
Median library plate correlation	< 0.7
Colony size median	< 2300 px, > 3800 px
Normalized colony size IQR	> 300 px
Circularity median	< 0.83
Circularity standard deviation	> 0.07

Table 1: Quality control parameters and their corresponding thresholds.

I chose all thresholds for the measures discussed below based on the overall distribution in the data set as demonstrated for a few examples in Figure 8. For a complete overview of all quality control parameters and their corresponding thresholds see Table 1.

I removed plates with poor replicate correlation or poor overall correlation to the same library plate across conditions, as this can be an indicator of technical errors such as irregular pinning of the agar surface or systematical errors such as faulty arrangement of the 1536 plate, accidental turning of the plate etc.

In order to calculate a meaningful control size for the interaction scoring all plates have to be brought to the same level of overall growth (see 2.2.5). As this is achieved by a multiplicative normalization step (see 2.2.4), plates with extremely high or low median growth have to be removed otherwise small differences in

colony sizes would be artificially inflated or compacted due to extreme multiplication factors.

As the growth of most mutants is unaffected by the genetic perturbation and interactions with the chemicals are rare due to the sub-inhibitory concentrations, the colony sizes are largely uniform. Therefore, a high IQR of colony sizes in one plate can indicate different sources of noise such as irregular pinning, strong stresses affecting many mutants, fluctuations in oxygen availability in center and edges of the plates. An uneven agar surface for example can result in locally higher pinning pressure, thereby larger colonies and eventually artificially high interaction scores. This bias can only be detected after normalizing for increased growth in outer rows and columns. Therefore, I excluded plates with a normalized colony size IQR higher than 300 px from the analysis.

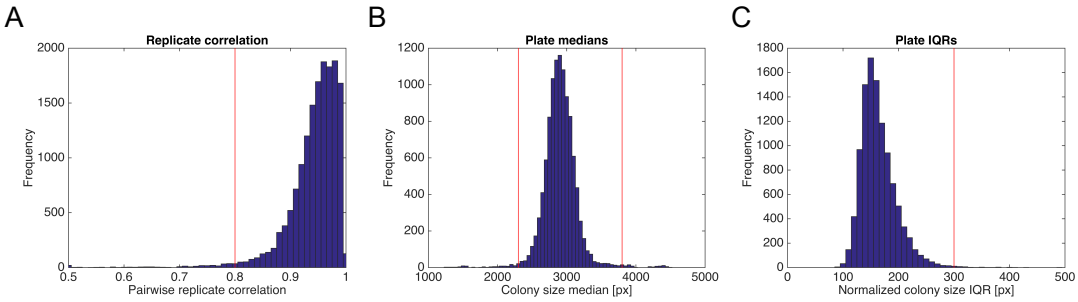

Figure 8: Overall distribution of selected quality control measures and thresholds applied.
For every parameter the threshold was chosen based on the data set in order to remove only plates with very extreme values and is indicated by the red line. Minimal media data was excluded from this analysis. A) All pairwise Pearson replicate correlations of raw colony sizes. B) Plate medians of raw colony sizes. C) Plate interquartile ranges (IQRs) of normalized colony sizes.

Furthermore I sometimes observed sliding of the pad on the agar surface during the transfer from the source to the test plate. This results in oval colonies and artificial colony sizes. I excluded plates with very poor overall colony circularity (median < 0.83) as well as plates with very diverse colony circularity values (standard deviation > 0.07) to account for local effects.

Plates based on minimal media show widely different distributions in the investigated parameters. Auxotroph mutants cause an increased number of small colony sizes skewing median colony size and IQR. The extreme growth defects result in strong phenotypes and as the interaction scores are normalized by condition in a final step, it is unlikely that small changes caused by technical errors have a stronger impact than the actual phenotypes. Thus, I kept all minimal media conditions without applying the quality control pipeline.

2.2.4 Normalization

Before using the colony sizes as a proxy of fitness for the interaction scoring a series of normalization steps are necessary. All of them are implemented in an adapted version of the EMAP toolbox for MATLAB [30].

In brief, a second order surface correction is applied to every 1536 plate to correct for pressure biases in the robotics used. Next, the two outermost rows and columns are normalized to the growth in the center of the plate. This is necessary as the colonies in the outer rows and columns grow to bigger sizes due to a higher availability of space, nutrients and oxygen. Lastly, every plate is brought to the same overall growth by a multiplicative correction. This step is crucial for the calculation of the control size representing the unperturbed growth of each mutant, which I will explain in more detail in 2.2.5.

To account for small differences introduced in the initial arrangement of 1536 plates from frozen stocks all normalization steps were carried out per batch (for definition see 2.2.1).

2.2.5 Scoring and clustering

The gene-drug interaction scores are calculated by comparing the normalized colony size of a mutant in a specific condition (mean across replicates in this condition) to the overall growth of that same mutant (median across all replicates and all conditions) in a modified t-test. If the mutant has a growth defect in the condition of question this will be reflected in a negative s-score, if the mutant grows better than usual it will yield a positive s-score (Figure 9).

A minimum of two replicates per library plate and condition is required for the score calculation (note that although four replicates were measured for each plate, some of them may have been removed in the quality control process).

As the mutants should be affected by only few conditions the median across many conditions should be representing the unperturbed growth of the mutant. This is only valid if the overall growth is similar across the tested conditions. Therefore, every plate is corrected for differences in overall growth (due to fluctuations in temperature, aeration, incubation time etc.) during the normalization.

The modification lies in a minimum bound for the experimental and control standard deviation. Reproducibility can occasionally be extremely high resulting in artificially strong s-scores, which do not reflect the strength of the gene-drug

interaction proportionally but are caused by extremely low variance values in the denominator of the equation. Therefore, a minimum bound based on the expected variance for mutants with similar growth phenotypes was introduced.

Finally, the s-score distribution of each condition was rescaled to fit the spread of a normal distribution by adjusting the IQR to 1.35, thereby ensuring comparability of conditions before applying clustering algorithms. The normalized s-score matrix was then subjected to two-dimensional, hierarchical clustering using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/) and the resulting gene-drug interaction map was visualized using Java TreeView (http://jtreeview.sourceforge.net/).

$$S = \frac{\mu_{Exp} - \mu_{Cont}}{\sqrt{\frac{s_{Var}}{n_{Exp}} + \frac{s_{Var}}{n_{Cont}}}} \qquad s_{Var} = \frac{var_{Exp} \cdot (n_{Exp} - 1) + var_{Cont} \cdot (n_{Cont} - 1)}{n_{Exp} + n_{Cont} - 2}$$

μ_{Exp} = Mean across replicates of normalized colony sizes for one specific mutant under one specific condition
var_{Exp} = Variance of the above, with a minimum bound
n_{Exp} = Number of replicates (usually 4)

μ_{Cont} = Median of normalized colony sizes for a specific mutant across all conditions (considering all replicates)
var_{Cont} = Variance of the above, with a minimum bound
n_{Cont} = Typical number of replicates (determined by the mode of number of replicates across all conditions)

Figure 9: Calculation of gene-drug interaction scores.
To calculate the gene-drug interaction a modified t-test is applied to yield the s-score. The s-score compares the colony size of a mutant upon one specific perturbation against the overall growth of that same mutant across all conditions. Minimum bounds on the experimental and control variance avoid artificially strong s-scores.

2.3 Determining sources of noise

2.3.1 Batch effects

While accumulating data I visually inspected the growing clustergram after every batch. Sometimes I noticed strong positive s-scores for a large set of mutants in one set of conditions and the same mutants having strong negative phenotypes in another set of conditions (Figure 10). There was neither obvious similarity between the chemicals nor between the mutants. However, the data for the chemicals was often obtained on consecutive days. Changes in colony size can occur while maintaining the library source plates on agar and those changes are consequently transferred to the chemical test plates. When this happens a mutant will have a bimodal distribution of colony sizes across the condition of the current

batch. For example, some mutants might take a longer time to recover after being transferred from frozen stocks to agar plates or they might be affected by keeping the plates at 4°C, when sources are not used the same day. Also, these errors can be introduced while transferring the mutant from one agar plate to the next, either due to cross-contamination or uneven inoculum (e.g. pad has uneven length of pins or touches some colonies only at the edge, transferring very little bacteria). As the control size of each mutant is equal to the median across conditions, a bimodal distribution will lead to a false control size and all the conditions with smaller colony sizes (left peak in the distribution) will result in negative scores, whereas all the conditions with bigger colony sizes will result in positive scores.

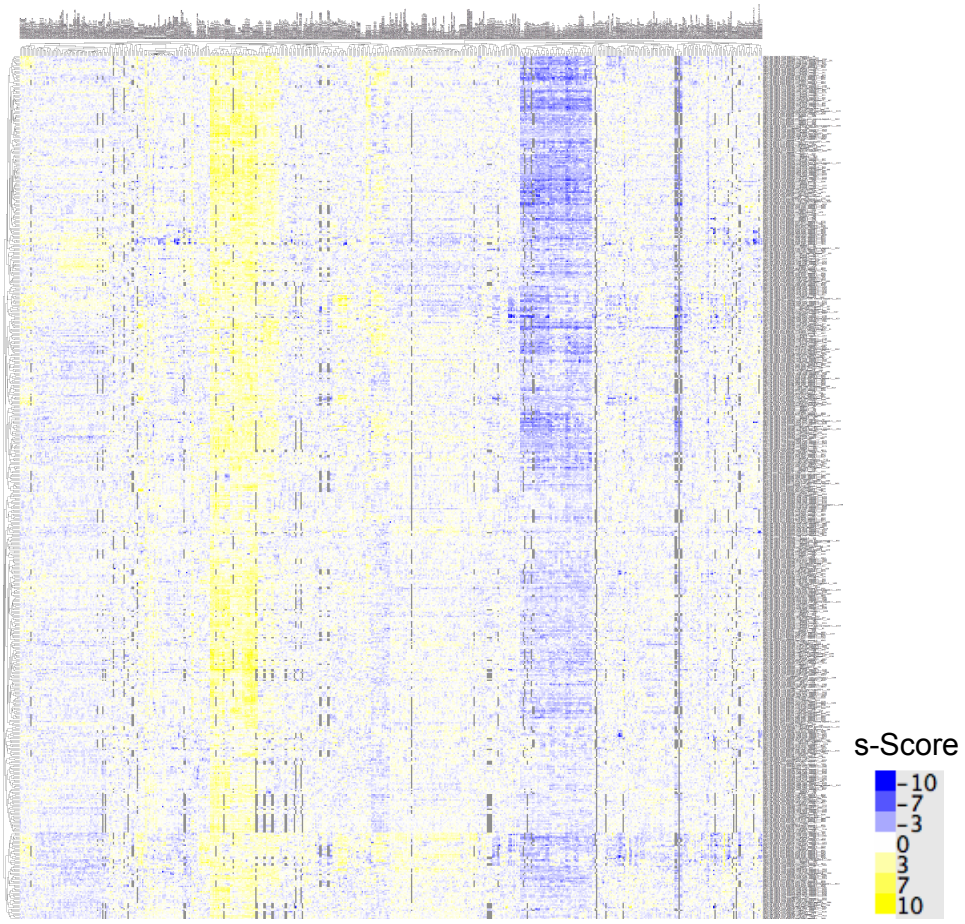

Figure 10: Sub-batch effects.
Fragment of the entire clustergram across 585 conditions (x-axis) and ~650 mutants (y-axis). The cluster of conditions with almost all negative scores (blue stretch) are all conditions tested in the beginning of batch 4 whereas conditions belonging to the yellow stretch were all tested at the end of batch 4.

Thorough inspection of the conditions behaving similar always revealed a time dependency. To correct for this effect I introduced sub-batches when

necessary, meaning that data from the same batch was split dependent on the days they were obtained and processed separately for normalization and scoring.

2.3.2 High mutant correlation can be caused by technical and intrinsic noise

After correcting for sub-batch effects, I still observed the phenomenon of many mutants behaving extremely similar across conditions. However, this time there was no obvious relation between those conditions, neither in respect to chemical structure of the compounds tested, nor in acquisition time. Furthermore, there was no functional relation between the mutated genes. A similar observation was made in the *E. coli* data set for mutants of the KEIO collection separating from small RNA mutants [36]. The small RNA mutants were created in a different genetic background resulting in distinct base level sensitivity to the tested chemicals. Therefore, the mutants in this sub-population were more similar to each other than to any other mutant in most of the conditions.

To investigate whether the effect in my data set is similar to the observation in *E. coli*, I calculated pairwise correlations of s-scores across all conditions and clustered the resulting data. Indeed there was one subset of mutants with extremely high correlations (see sub-cluster C1 in Figure 11A and B), which corresponded to the mutants described above. Furthermore, I observed additional clusters with underlying, yet much milder, correlations.

Location effects contribute to underlying mutant correlation

First, I determined the location of the mutants in all sub-clusters (Figure 11C). Surprisingly, sub-clusters C2 and C5 showed clear location effects. For example, mutants in cluster C5a are all located towards the plate corners. Several explanations can be considered. As the biases seem geometric it can be a systematic effect introduced by the robotic transfer. Furthermore, plate corners are the first place that would be affected by evaporation (plates drying out causes dips in the agar, especially in the corners), the plate center might be more deprived of oxygen etc. In any case, the mild surface correction implemented in EMAP might not be sufficient to correct for these biases. Alternatively, a simple solution is to normalize each subset to the same overall growth separately. This way the correcting factors are only applied to sets of mutants with similar underlying behavior, thereby avoiding the creation of bimodal distributions and

imprecise control sizes. As small underlying effects will not disturb the clustering whenever 'real' phenotypes occur, we only applied the separate normalization to the sub-clusters with high internal correlation (median correlation above 0.1 in Figure 11B), namely C2b, C5a and C5c.

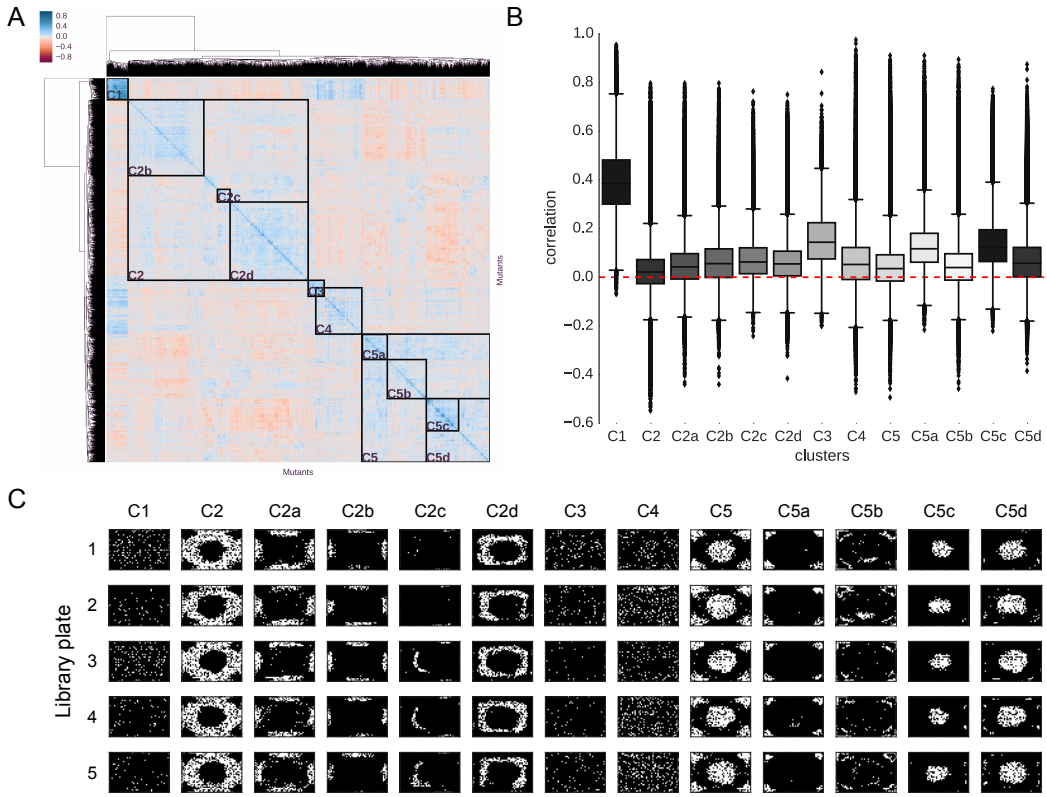

Figure 11: Position effects partly account for underlying correlations between mutants.
A) Clustergram of all pairwise Pearson correlations of s-scores across all conditions (except minimal media conditions). B) Distributions of Pearson correlations within each selected sub-cluster. C) Location of mutants in each selected sub-cluster across the five library plates. White spots indicate the presence of the mutant located in this position of the plate in the respective sub-cluster. Sub-clusters C2 and C5 show clear position effects, which are broken down further when investigating smaller sub-clusters whereas the mutants in sub-clusters C1, C3 and C4 seem to be randomly distributed across the five library plates.

However, mutants belonging to sub-clusters C1, C3 and C4 showed random distribution across the library plates, indicating that there was neither a technical nor a systematic effect during creation or re-array of the library. Notably, the correlation of mutants in C3 and C4 to those in C1 was comparably high. Therefore, I combined data from only C1, C3 and C4 and clustered them separately to obtain a better separation for the respective mutants (Figure 12A). Again, I selected fragments of this clustergram for further analysis. The distributions showed that it is desirable to separate C*1a, C*1b and C*4 for normalization as overall correlation is well above 0.2 (Figure 12B). However, C*1b

and C*4 consist of a rather low number of mutants. If distributed unevenly across library plates this can cause problems, as the normalization factor will be calculated on a very low number of mutants. Combining mutants from C*1b and C*4 results in a bimodal distribution of score signature correlations indicating different behavior of these two sets of mutants (Figure 12C). Thus, normalizing them together might result in artifacts. Therefore, I combined C*1b and C*1c as well as C*3 and C*4 to increase the number of mutants in each group. As shown in Figure 12C, combining these groups does not result in bimodal distributions indicating that their behavior is sufficiently similar.

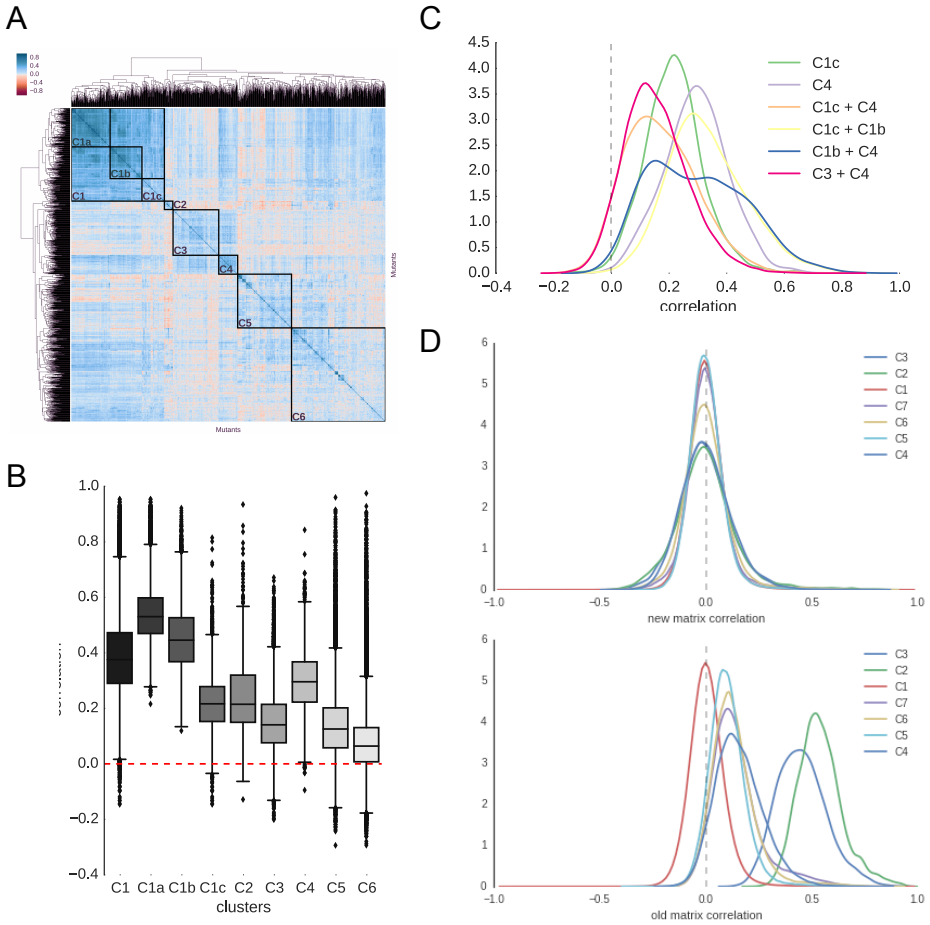

Figure 12: Grouping genes with high underlying correlation for separate normalization.
Note that all sub-clusters shown in this figure are referred to as C* in the main text. A) Mutants from sub-clusters C1, C3 and C4 were combined, pairwise correlations calculated and clustered. Several sub-clusters were selected for further analysis. B) Distribution of s-score correlations within the selected sub-clusters. C) Distributions of correlations across sub-clusters and combinations of those. D) All pairwise correlations of mutants within selected normalization groups before and after separate normalization.

Changes in genetic background might account for high mutant correlation

While separate normalization can solve the effect, the cause of this is still under investigation. First, I performed whole genome sequencing on seven

problematic mutant strains alongside several control mutants as well as the *S.* Typhimurium wild-type strain from two different stock aliquots (one of them stemming from the library plates). No common mutation between all problematic mutants but absent in the control group could be detected. However, many genetic variations could lead to a common phenotype. Indeed, several mutations observed might impair envelope integrity (data not shown) and consequently alter drug permeability. As the initial sample size was small and sequencing coverage for some not ideal, we decided to experimentally confirm genetic background alterations as a cause for the phenotypic similarity between those mutants before sequencing more strains. To achieve this we re-transduced 30 mutants into wild-type background to loose any genetic background alterations that may have occurred. We are currently testing the sensitivities of the library strains and their re-transduced equivalents in a subset of conditions that displayed strong interactions in the original screen.

Separate median normalization resolves underlying mutant correlations

To summarize, the following groups have been separated for the multiplicative plate-to-plate normalization: sub-clusters C2b, C5a and C5c from the initial score signature correlation clustering due to strong position effects caused by the technique as well as C*1a, C*1b+C*1c and C*3+C*4 from the subsequent correlation clustering. In the last step of the normalization procedure a separate factor is applied to bring these mutant sets to the same median growth thereby accounting for difference in base line sensitivity or growth (see also 6.1.6 Normalization, interaction scoring and clustering in the Materials and Methods section). Note, that strong phenotypes (very big or small colony sizes) are not affected by this procedure.

2.4 Data quality

Finally, after applying the quality control pipeline and correcting for underlying effects I present a gene-drug interaction map for 7200 mutants in 558 conditions yielding more than 4 Mio interaction scores (Figure 13A).

2.4.1 Responses for 75% of the tested mutants

As shown in Figure 13B the s-score distribution of the entire data set is centered around zero, indicating that gene-drug interactions are rare (by design due to the sub-inhibitory concentrations used). Nevertheless, 75% of the tested mutants show at least one significant phenotype considering a false discovery rate of 0.05. Considering gene redundancy this shows that the set of conditions tested was sufficiently broad to target many different processes. In a similar study conducted in *E. coli* around 50% of the tested mutants showed at least one phenotype in 324 single conditions. I could not observe any differences in phenotype numbers between the Kan and Cm library (Figure 13C).

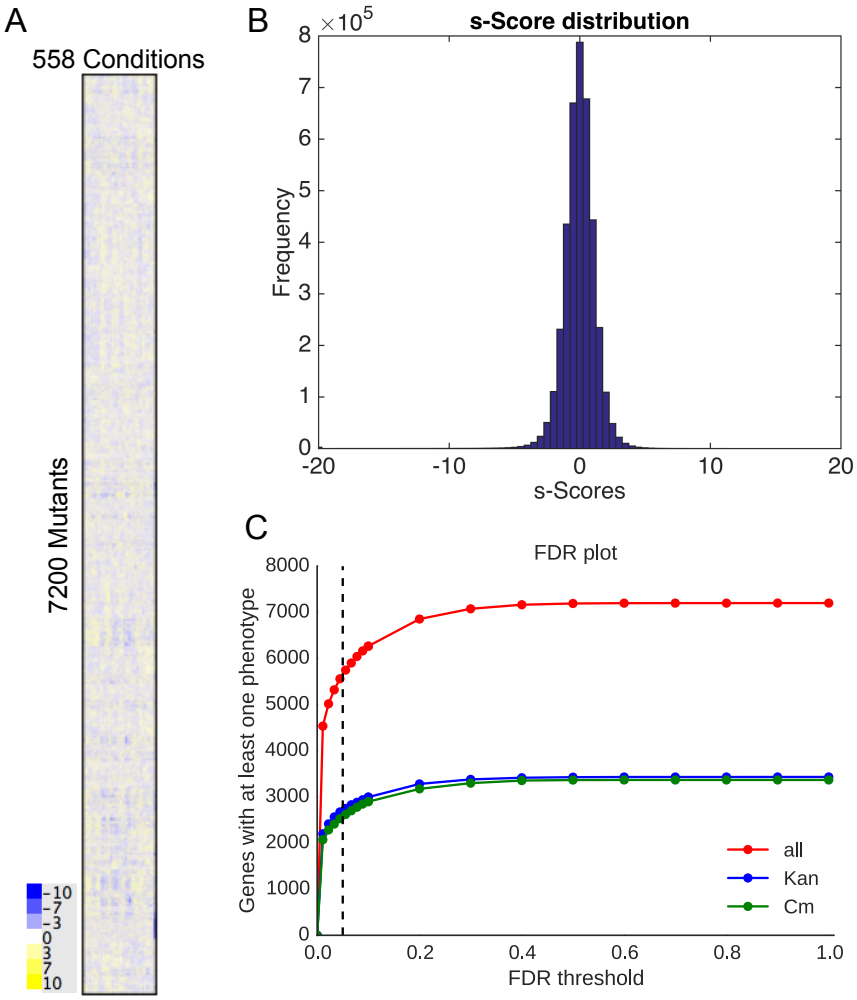

Figure 13: Chemical genomics for 7200 *Salmonella* Typhimurium mutants in 558 conditions reveals phenotypes for three quarters of the tested mutants.
A) Gene-drug interaction map of the entire data set. Shown are the s-scores two-dimensionally clustered across all 7200 mutants and 558 conditions. B) The distribution of s-scores shows that gene-drug interactions are rare (by design due to the sub-inhibitory concentrations tested). C) With a false discovery rate of 0.05 ~75% of the mutants show at least one phenotype. The Kan and Cm library behave the same.

2.4.2 Clone correlation

As mutants of the same gene should exhibit the same behavior testing the correlation between clones is a good quality check. Therefore, we first calculated correlation of s-score signatures between clones present multiple times in the same library (introduced when re-arranging the library after colony purification and referred to as intra-library correlation). As shown in Figure 14A and B most clones behave similar across conditions, which becomes especially evident when strong growth defects (negative s-scores) are observed. The seemingly low overall correlation value can be explained by the low number of data points and the fact that the majority of scores is neutral (so around zero), a phenomenon explained in more detail below.

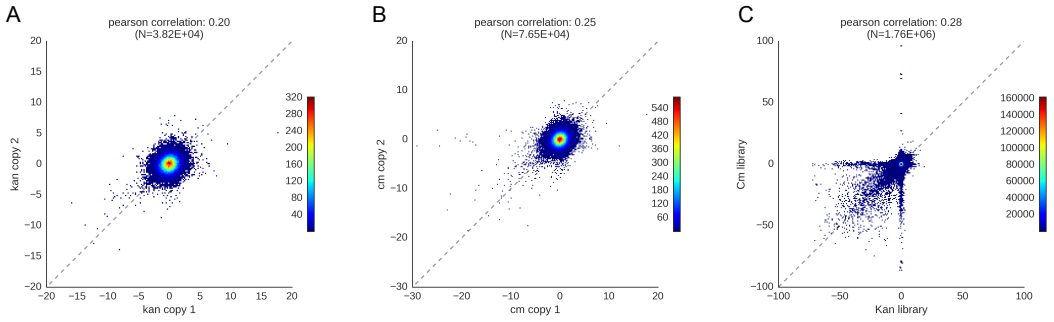

Figure 14: Intra- and inter-library clone correlation.
One data point represents the s-scores of one clone pair in one condition. A) Scatter plot of s-scores across conditions for all clone pairs of mutants with multiple clones present within the Kan library B) Same as A) for the Cm library. C) Mutants of the same gene derived from the Kan or Cm library are compared.

However, for the Cm library there are a few cases when one clone shows significant phenotypes whereas its replicate exhibits s-scores around zero in the same condition (e. g. data points along the horizontal in Figure 14B). Note that there is one point per clone pair and condition, so several data points could be derived from the same clone pair. Indeed, all points along the horizontal were derived from just one clone pair. Possible explanations are cross-contamination from adjacent wells or one clone having acquired a secondary mutation. We removed the disagreeing pair from further analysis.

Additionally, we examined the correlation between mutants of the same gene derived from the Kan or Cm library (inter-library correlation, Figure 14C).

Here we observe that many more clone pairs show diverging behavior. This is first of all due to the overall much higher number of pairs (few hundred in the intra-library comparison, but almost every mutant is present in both libraries,

resulting in over 3000 pairs tested in the inter-library comparison). Also, as the clones come from two different libraries, there are more possible explanations besides cross-contamination and secondary mutations mentioned above. Because the two clones carry different resistance cassettes with slightly different design (co-directional vs. opposite placement), different polar effects may account for differences in behavior. Another possibility is that one mutant is simply not correct or was misplaced at any step of the library assemble or subsequent rearrangements.

In total, 429 Kan-Cm pairs showed significantly different behavior. To decide which clone to keep for further analysis we manually curated this list based on total number of phenotypes for each mutant, conditions in which those occur and most correlated genes. For example, sometimes we will find genes of the same operon correlating highly with one but not the other clone. This approach allowed us to keep the more likely correct clone in most cases. Whenever we were unable to make an informed decision based on the data, we excluded the Kan-Cm pair completely.

We excluded a total of 223 strains (126 Kan and 97 Cm) from further analysis. This represents only 3% of the tested mutants, leaving almost 7000 mutants subjected to further analysis. Intra- and inter-library correlations after removal of these noisy strains are shown in Figure 15.

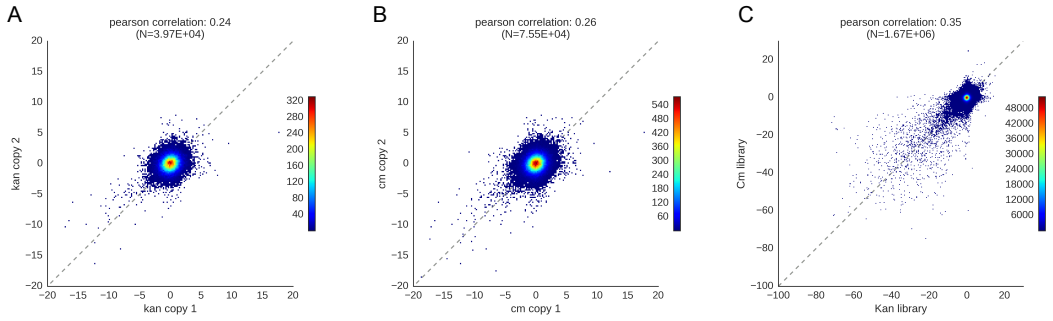

Figure 15: Intra and inter-library clone correlation after removal of noisy strains.
One data point represents the s-scores of one clone pair in one condition. A) Scatter plot of s-scores across conditions for all clone pairs of mutants with multiple clones present within the Kan library B) Same as A) for the Cm library. C) Mutants of the same gene derived from the Kan or Cm library are compared.

Only few strains present multiple times within a library were removed, therefore the change in overall intra-library correlation is marginal, resulting in a correlation of 0.24 for Kan clone pairs and 0.25 for Cm clone pairs. The curation applied for Kan-Cm clone pairs resulted in removal of strongly disagreeing pairs,

which is reflected in the highly reduced number of data points along the horizontal and vertical of Figure 15C (after curation) compared to Figure 14C (before curation). Overall inter-library correlation improved from 0.28 to 0.35. Note that the number of data points in Figure 15A and B are several orders of magnitudes smaller than for Figure 15C explaining why the overall correlation value is smaller for the intra-library correlation compared to the inter-library correlation.

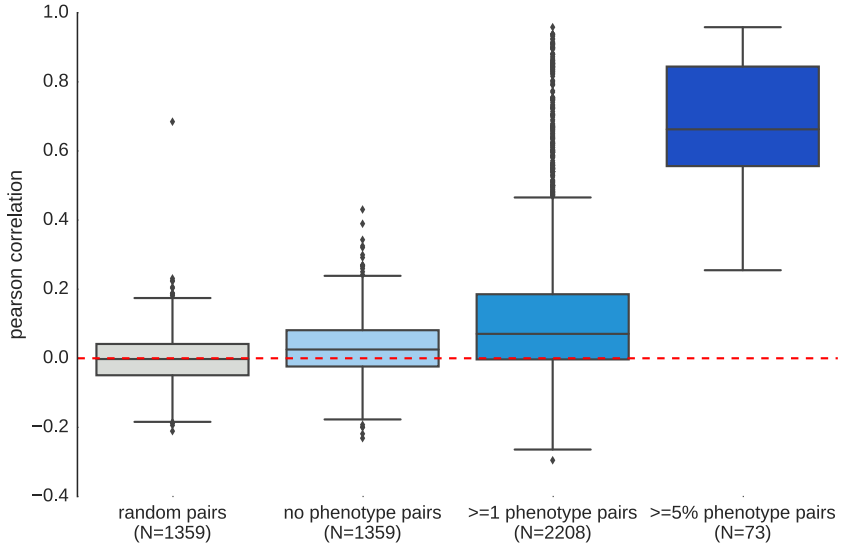

Figure 16: The presence of significant phenotypes is crucial for calculation of meaningful correlations. Distribution of Kan-Cm clone correlations after removal of noisy strain considering (from left to right): random Kan-Cm pairs (mutants of different genes), Kan-Cm pairs with no phenotypes, Kan-Cm pairs with the requirement of one significant phenotype for at least one clone, or Kan-Cm with the requirement of one clone showing significant phenotypes in at least 5% of the tested conditions.

The overall correlations may appear low but are expected as most s-scores are centered around zero (note the density in Figure 14 and Figure 15). When most data points are scattered around zero, correlation will be accordingly low and only a sufficient amount of strong negative or positive s-scores is able to drive correlation. This is demonstrated in Figure 16. Correlations between Kan and Cm clones of the same gene barely differ from a control set of random pairs (different genes!) when none of the clones show a significant phenotype. However, when one or more significant phenotypes are observed, the distribution clearly shifts to higher correlations. This effect is even stronger when at least one clone shows significant s-scores in at least 5% of the tested conditions.

This phenomenon indicates that lacking correlation is not necessarily an indicator for lacking relation between mutants. The number of observed phenotypes has to be considered. However, high correlation between mutant pairs indicates relation (compare the two rightmost boxes to the random set in Figure 16). This is a concept that can be expanded to functional relation and is applied in the benchmarking discussed below.

2.5 Benchmarking

In order to associate genes of unknown function with known pathways or complexes I make the assumption that similar behavior across conditions and thereby high s-score signature correlations indicate functional relation. To test this hypothesis I investigated the s-score correlation of genes with known interaction on e.g. operon, pathways or protein complex level.

2.5.1 Defining benchmarking sets

Known direct protein-protein interactions (PPIs) are the most valuable benchmarking set. In many cases removal of a single protein complex component will render the complex unfunctional. Therefore, gene deletions of different components of the same complex are expected to show highly similar behavior.

Most bacterial protein complexes have been experimentally validated in *E. coli* and are conserved in *Salmonella.* Thus, experimental evidence for protein-protein interactions in *Salmonella* Typhimurium is scarce. In fact, less than 20 PPIs experimentally verified in *S.* Typhimurium can be found in databases. However, functional modules are highly conserved across organisms [24, 54, 56, 77] and surely this observation holds true for players of crucial cellular pathways between the closely related *E. coli* and *Salmonella.*

Therefore, we took a list of known pairwise protein-protein interactions of *E. coli* from the Biocyc database and tested all pairs with orthologs in *Salmonella* for their score signature correlation. Furthermore, we included data for *Salmonella* from Biocyc pathways and operons, KEGG modules and pathways as well as STRING experimental data.

2.5.2 Mutants of functionally related genes show high score signature correlations

First, we calculated all possible pairwise correlations between mutants across all conditions, which are represented by the black lines in all panels of Figure 17. As expected, most mutants are uncorrelated, so the distribution is sharply centered around zero.

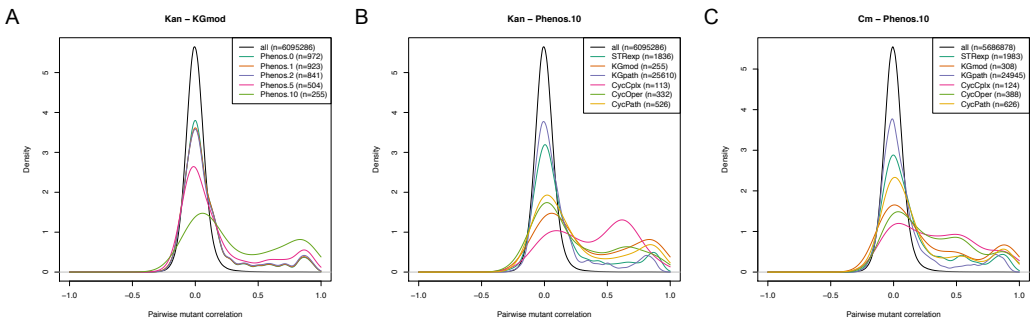

Figure 17: Benchmarking - high correlation of phenotypic signatures indicates functional relation.
A) All possible pairwise mutant correlations within the Kan library (black line) and correlation of mutant pairs present in KEGG modules with no restriction on phenotype number (Phenos.0) or requiring at least 1, 2, 5 or 10 significant phenotypes to be present in each mutant of the pair. B) All possible pairwise mutant correlations within the Kan library (black line) and correlation of mutant pairs present in different benchmarking data sets (also see Table 5 on page 94 in the Materials and Methods section) requiring at least 10 significant phenotypes to be present in each mutant of the pair. C) Same as B) for the Cm library.

However, if I limit the calculation to include only pairs of mutants with known interactions I observe a second peak at high correlations. Figure 17A demonstrates this on the example of mutant pairs of the Kan library present in KEGG modules. As gene-drug interactions are rare, it has to be considered that meaningful correlations can only be calculated if the mutants exhibit at least a few significant phenotypes. If no significant phenotypes are detected all s-scores will be centered around zero and correlation will be poor even if the deleted genes interact (in this case the chemical stresses were not of the right kind or strength). To account for this, I next restrained the calculation to only consider mutant pairs with at least 1, 2, 5, or 10 significant phenotypes present in each mutant. As shown in Figure 17A, the peak around zero decreases with increasing number of phenotypes whereas the peak at high correlations increases. Note, that the total number of mutant pairs considered decreases due to the minimum bound for phenotypes. The number of considered pairs is indicated in the legend.

The same trend is observed for all other benchmarking sets used, both for the Kan and Cm library (Figure 17 B and C). Interestingly the Cm library seems to

benchmark better against gene pairs present in the same operon. As the Cm resistance cassette is placed in the opposite direction of the original gene, the produced antisense RNA against the upstream gene can cause polar effects. Due to these polar effects a mutant of the Cm library might behave similar to a double mutant of the intended and the upstream gene. Thus, the likelihood of showing similar behavior to a mutant in the same operon (e.g. the actual mutant of the upstream gene) would be increased.

In conclusion, mutants of functionally related genes show highly similar reactions across the conditions tested. Therefore, a high score signature correlation serves as an indicator of functional relation and will allow to link genes of unknown function to known cellular pathways or complexes.

2.6 Merging the Kan and Cm library

For the interspecies comparison described in the following chapter, robustness of the chemical genomics data set is of crucial importance. As we screened two different knock-out libraries, tested their correlation and removed disagreeing strains (see 2.4.2 Clone correlation), we have the opportunity to combine these two data sets to yield a more robust s-score matrix.

To approach this we averaged the s-scores derived from Kan or Cm mutants of the same gene. If multiple Kan or Cm clones were present one was randomly selected to calculate the Kan-Cm average. If a mutant was present in only one library a pseudo-averaging strategy as described by Collins [30] was applied unless the mutant was present multiple times within that library. In this case two clones of the same library were averaged.

In Figure 18 we compare benchmarking performance of the separate libraries to the averaged data set. The presented ROC (receiver operator characteristics) curves describe how well a set of gene-pairs with known interaction can be distinguished from a random set of the same size based on their score signature correlation. In brief, true positive rates are plotted against false positive rates for different thresholds of s-score signature correlation (assuming above the threshold the pair should be interacting). The larger the area under the curve, the better the distinction of the two sets. For each individual benchmarking set correlations derived from the combined libraries yield a larger area under the curve than correlations derived from each single library, indicating that the combined data set is indeed more robust.

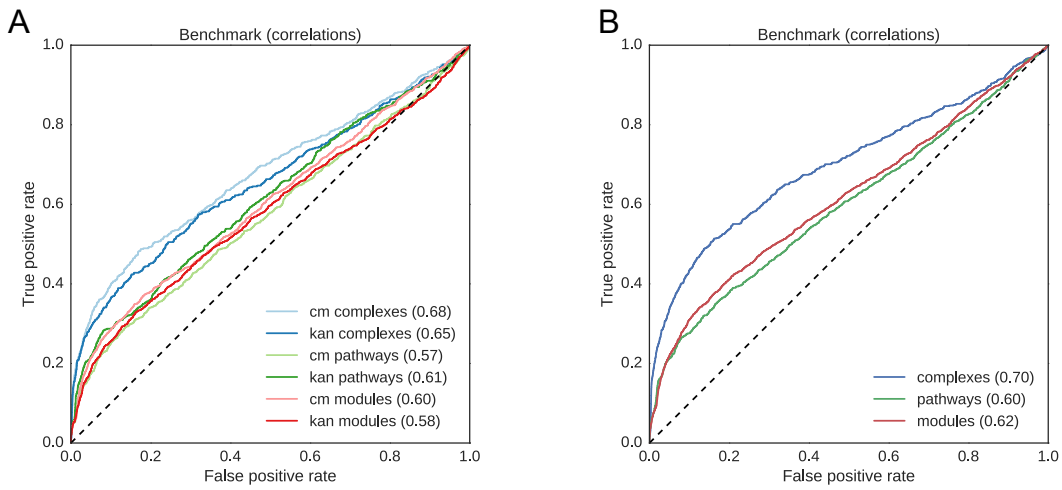

Figure 18: Combining the Kan and Cm libraries yields a more robust data sets with better benchmarking performance.
ROC (receiver operating characteristics) curves of mean s-score signature correlation of gene pairs belonging to the indicated annotation compared to a random set of the same size. Number in parenthesis is the area under the curve. A) Benchmarking against Biocyc protein complexes (inferred from *E. coli*), Biocyc pathways and KEGG modules for the separate libraries. B) Same as A) after averaging clones of the same gene from the different libraries.

2.7 Robust phenotypes for genes of unknown function

Using the chemical genomics map of merged, robust s-scores I next wanted to determine whether we are able to detect significant phenotypes for genes of unknown function (orphans). As a joined effort in the lab we recently curated the functional annotations of all *E. coli* genes resulting in 1585 confirmed orphans. For 915 of those orthologous genes exist in *Salmonella,* 760 of which are represented as mutants in my data set. We can assume that the great majority of genes present but not annotated in *E. coli* are likewise orphans in *Salmonella.* Thus we divided all *Salmonella* mutants in the following three categories: mutants orthologous to an *E. coli* orphan, mutant orthologous to an *E. coli* gene of known function and mutants with no ortholog in *E. coli.*

In Figure 19 we count the number of mutants with no or at least one significant phenotype in each category and observe significant phenotypes for more than 500 orphan genes. The conditions a particular orphan is sensitive or resistant to will provide a starting point for follow up analysis. Additionally, ~10% of the orphans show high correlation to a gene present in the combined benchmarking sets of BioCyc complexes, pathways and KEGG modules (and therefore of known function, data not shown). As the benchmarking analysis proofed that high score signature correlations indicate functional relation, we will

be able to link the orphans to known cellular complexes and pathways providing further information on their potential function. In the next chapter I expand this idea to linking orphans to robust modules of functionally related genes predicted from chemical genomics data.

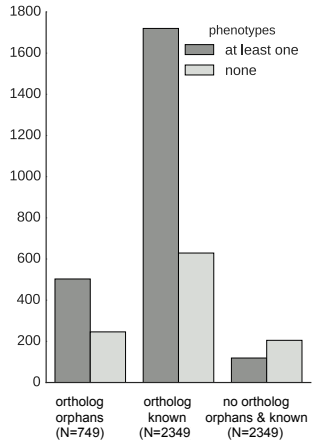

Figure 19: More than 500 orphan genes show significant phenotypes.
We divided all *Salmonella* mutants in the following three categories: mutants orthologous to an *E. coli* orphan ('ortholog orphan'), mutant orthologous to an *E. coli* gene of known function ('ortholog known') and mutants with no ortholog in *E. coli,* which can be of known or unknown function ('no ortholog, orphan & known'). For each set we count the number of mutants with no phenotype and the number of mutants with at least one significant phenotype (considering a false discovery rate of 0.05).

In contrast to orthologous genes, mutants of *Salmonella*-specific genes (panel 'no ortholog' in Figure 19) show a higher proportion of mutants with no phenotypes. Species-specific genes might fulfill specialized tasks, which may be important for growth only under niche-specific conditions, which are not easily mimicked by chemical perturbations. Therefore mutants of species-specific genes might be generally less responsive than conserved genes, an idea explored in more detail in the following chapter.

2.8 Conclusions

In summary, I created a chemical genomics data set for 7200 *Salmonella* Typhimurium mutants in 558 conditions yielding more than 4 million gene-drug interaction scores.

In the course of obtaining this data set I setup an automated quality control pipeline and discovered potential genetic background changes in part of the utilized mutant collection, which we are currently investigating. Furthermore, we subjected the two subsets of the collection (Kan and Cm library) to a detailed

comparative analysis including manual curation of Kan-Cm pairs with diverging phenotypes. This allowed us to average the data in order to obtain a robust gene-drug interaction map. Benchmarking against sets of known interactions from different sources proved high correlation of phenotypic signature as an indicator of functional relation.

From this data set we can derive significant phenotypes for more than 500 genes of unknown function. Thus, it will serve as a valuable resource for exploring gene function and drug mode of action for both, our laboratory as well as the entire microbiology community.

2.9 Contribution disclaimer

I designed concept, acquisition and analysis of the presented chemical genomics screen in *S.* Typhimurium. The gene deletion libraries were provided by Helene-Andrews-Polymenis (Texas A&M) and Michael McClelland (UCI) prior to publication. Steffen Porwollik (UCI) provided updates on the library annotation. Data acquisition was partly aided by Anja Telzerow, Nadja Nepke and Matylda Zietek (all EMBL Heidelberg). George Kritikos (EMBL Heidelberg) developed the image analysis software used in this study (Iris, unpublished software). Shu-Yi Su (former EMBL Heidelberg, currently Max Planck Institute of Psychiatry, Munich) contributed to adaptations in the EMAP toolbox used for normalization and scoring. Alison Waller (EMBL Heidelberg) implemented the benchmarking pipeline. Marco Galardini (EBI Hinxton) co-designed and implemented the detailed mutant analysis (correlation cluster analysis, clone correlations and removal of noisy strains).

3 Inter-species comparison reveals conservation and rewiring of cellular networks

In the following chapter I compare the chemical genomics data set created for *Salmonella* Typhimurium to a previously published *E. coli* data set [36]. First, I determine a set of shared conditions and investigate whether conserved genes also show conserved phenotypic responses. Furthermore, I gain insights about the conservation level of highly responsive mutants and their spatial distribution across the chromosome. Next I investigate modules of functionally related genes and drugs predicted from each data set, their overlap and differences. From this analysis we can derive a number of inferences for orphan gene function.

3.1 Determining a shared data set

3.1.1 Condition selection

In 2011, Nichols et al. presented a chemical genomics data set for *E. coli*, measuring the growth of the KEIO collection of single gene deletions in 324 conditions covering 114 unique stresses [36]. A total of 91 unique stresses overlap with the study presented here. Both studies include physical perturbations as well as chemical stresses, most of them tested in several concentrations. Whereas physical stresses (temperature, pH, UV exposure) and minimal media conditions will be compared directly, we employed correlation analysis to determine the most comparable concentrations for each chemical stress.

Assuming a certain gene-drug interaction, we expect this interaction to be visible in the highest drug concentration tested. Following this rationale, one approach was to pick the highest concentration of a given chemical present in each data set. However, the minimal inhibitory concentrations (MICs) of a given drug can differ significantly between *E. coli* and *Salmonella* due to differences in drug uptake and efflux, or primary target sensitivity. Even if each concentration would represent e.g. 20% MIC for the respective organism, the drug effect might still differ depending on the shape of the dose response curve. In one organism the dose response might be very sharp, meaning that a slight increase in drug

concentration can make the difference between full growth and complete inhibition. If the other organism has a more linear drug response, the tested concentrations, even if based on percent MIC, will have widely different effects. Therefore, we wanted to choose concentration pairs that behave similarly.

To decide on the most comparable concentrations we took into account the overall amount of significant phenotypes observed for each organism, the fraction of common genes sharing phenotypes, as well as the overall correlation between s-scores for orthologs. To outline this strategy, Figure 20 demonstrates one example.

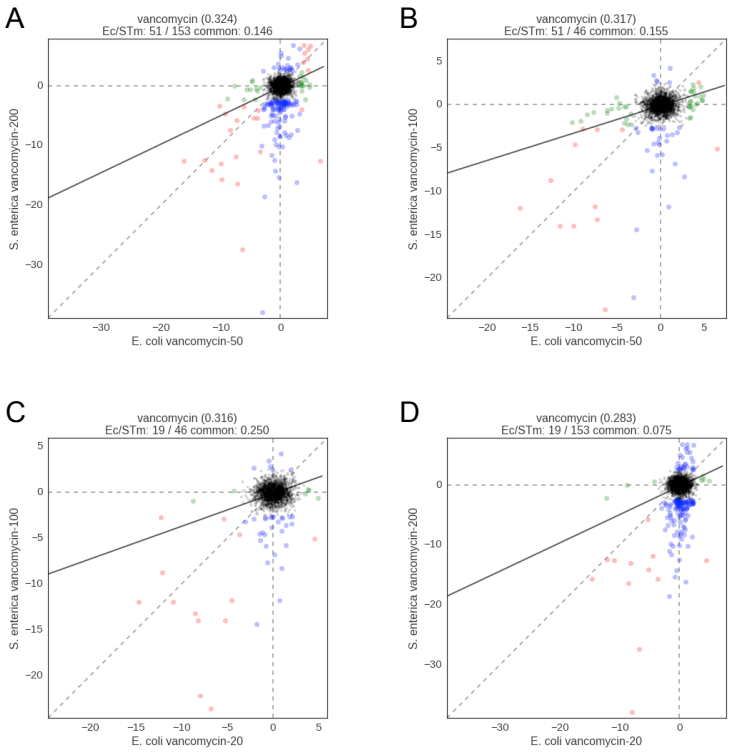

Figure 20: Overall correlation, total and common number of phenotypes were used to determine the most comparable concentration pair for a given chemical.
Different combinations of vancomycin concentrations (in µg/ml) tested in *E. coli* and *Salmonella* are shown rank ordered by overall correlation (A-D). Indicated are the overall correlation across orthologs (in parentheses), the total number of phenotypes observed in each organism as well as the fraction of phenotypes in common (orthologous genes showing a phenotype, red points) as opposed to phenotypes present only in *E. coli* (green) or only in *Salmonella* (blue).

It shows scatter plots comparing vancomycin s-scores derived from *E. coli* mutants or their corresponding *Salmonella* ortholog for different concentration pairs. Here I show the four concentration pairs with the highest overall correlation (rank ordered A-D). Figure 20A, for example, shows the interaction scores of *E. coli* mutants grown at 50 µg/ml vancomycin and their *S.* Typhimurium ortholog grown at 200 µg/ml vancomycin. This is the concentration pair with the highest

overall correlation (amongst all possible combinations, having tested *E. coli* in 20 and 50 µg/ml and *Salmonella* in 50, 100 and 200 µg/ml). However, for *Salmonella* we detect a total of 153 significant phenotypes in this condition, while only 51 are detected for *E. coli*, suggesting that 200 µg/ml vancomycin generally affect *Salmonella* more than 50 µg/ml affect *E. coli.* Lowering the concentration used for *Salmonella* to 100 µg/ml results in a more similar number of phenotypes between the two organisms (Figure 20B). In this combination we find common genes with similar phenotypes, as well as species-specific phenotypes while overall correlation is still comparably high, so we choose it for further analysis. The same strategy for selecting the most comparable concentration pair was applied to all other chemicals. If several concentration pairs showed similar behavior, we selected the higher concentrations.

In a few cases the tested concentration range and/or sensitivity to the tested chemical differed too much between the two species. This is reflected in the residual number of phenotypes: For each selected concentration pair and each ortholog pair we substract the *E. coli* s-score from the *Salmonella* s-score resulting in a residual. The mean of these residuals across ortholog pairs is shown in Figure 21. Conditions with an absolute mean of residuals greater than 50 were excluded from further analysis. This is the case for azidothymidine (more/stronger phenotypes in *E.coli*) and triclosan (more/stronger phenotypes in *Salmonella*).

3.1.2 Mutant selection

In the following comparative analysis I either refer to the complete data set for each organism or a shared set of mutants. The shared set is based on orthologous genes predicted using OrthoMCL [78].

In the *E. coli* chemical genomics map the s-score of each mutant is derived from the average of two clones carrying the Kan cassette [36]. The s-score used for *Salmonella* is derived from the average of the Kan and Cm clone as described in 2.6 Merging the Kan and Cm library (page 39).

In conclusion, we compare two robust data sets with s-scores derived from multiple clones in a set of conditions tested for comparability in strength and similarity of effect.

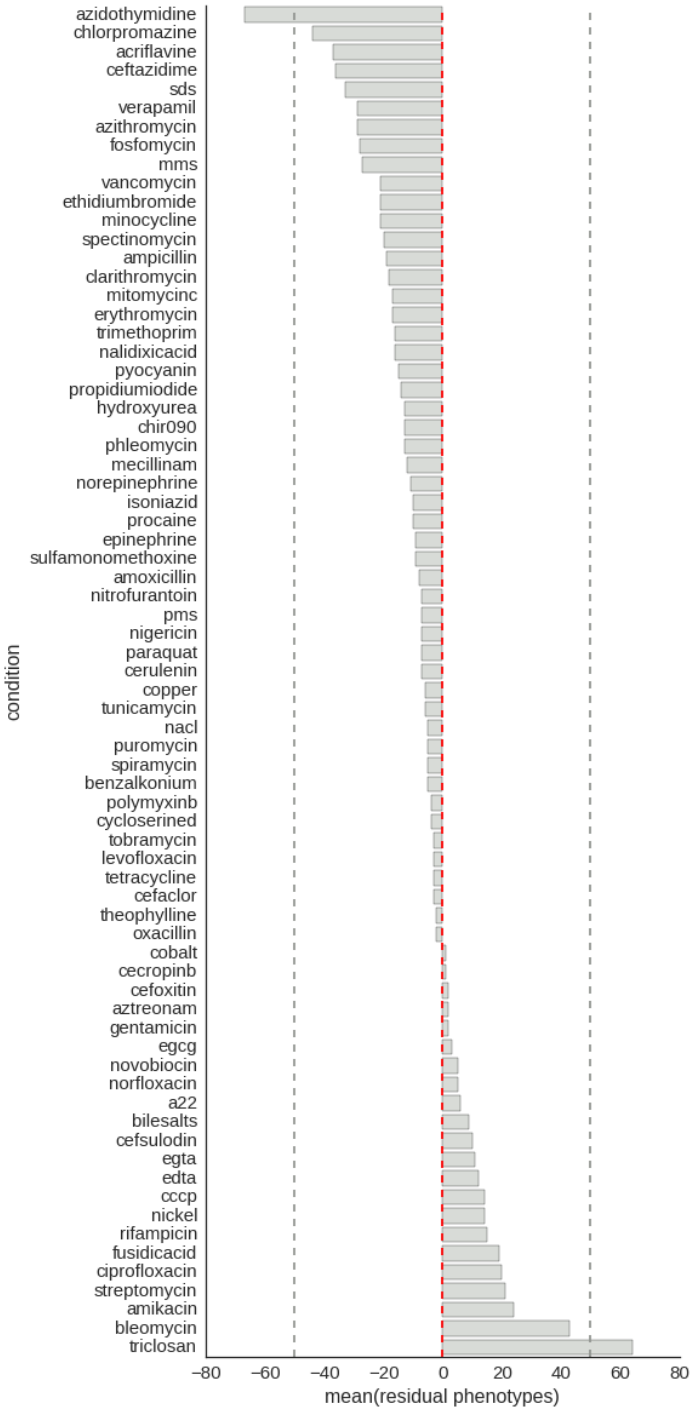

Figure 21: Comparing conditions from the *E. coli* and *Salmonella* data set.
For each chemical stress the selected concentration pair was tested for differences in overall effect in *E. coli vs Salmonella*. The s-score from *E. coli* in the respective condition was substracted from the *Salmonella* s-score for each pair of orthologous genes. The mean of these residuals across ortholog pairs is shown.

3.1.3 The shared conditions recover a large proportion of responsive mutants

For *E. coli* Nichols et al. reported at least one significant phenotype for 50% of the tested mutants (considering a false discovery rate of 0.05). As described in chapter 2.4.1 (page 33) I find at least one significant phenotype for 75% of all *Salmonella* mutants tested. Figure 22 demonstrates that this holds true when testing the s-scores averaged between Kan and Cm clones. The discrepancy between the number of responsive mutants in each organism may be explained by the bigger and broader range of conditions tested in *Salmonella.* If this is the case, the numbers should be more similar when considering only shared conditions.

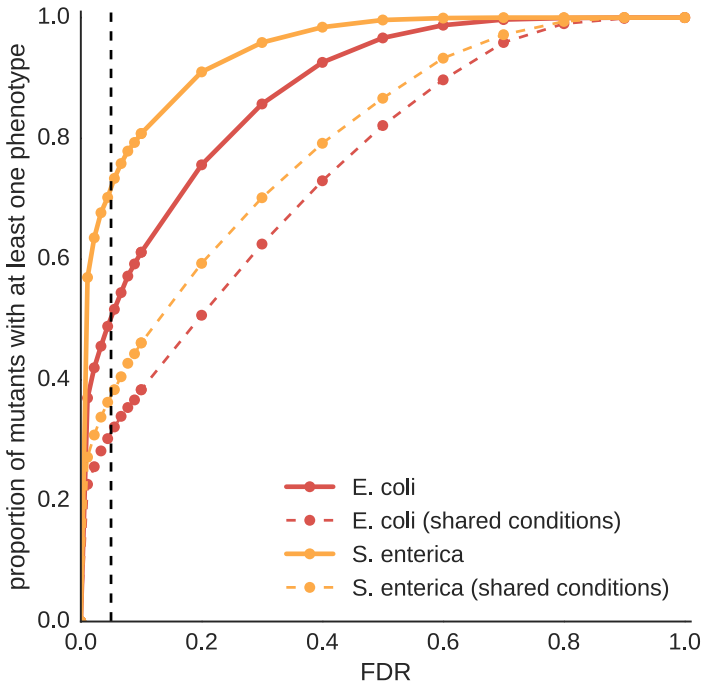

Figure 22: A large proportion of responsive mutants is recovered based on phenotypes detected in shared conditions.
Number of mutants with at least one significant phenotype considering different false discovery rates for the EC and STM data set, based either on all conditions or on shared conditions.

The total amount of mutants with at least one significant phenotype expectedly drops for both organisms when considering only shared and therefore much less conditions. Despite reducing the number of conditions more than 3.5-fold for *E.coli* and more than 6-fold for *Salmonella*, we recover the majority of responsive mutants. The absolute number remains slightly higher for *Salmonella* than for *E. coli* (dotted lines in Figure 22), but the difference between the two

organisms is significantly reduced. One possible explanation for the residual difference is a greater sensitivity in colony size detection for the *Salmonella* screen caused by a higher camera resolution and improved image analysis software. Higher precision in the colony size detection results in a bigger dynamic range for detecting phenotypes and allows the detection of subtle differences as significant phenotypes. Second, higher reproducibility in the *Salmonella* data set would lead to increased absolute s-scores (see s-score formula on page 27, the lower the variance the higher the s-score). This would distinguish true interactions better from noise and thus lead to the detection of more significant phenotypes. Lastly, higher levels of gene redundancy in *E. coli* would explain this effect. Upon deletion of one gene the paralog is still available to fulfill the tasks, therefore no phenotype would be detected.

3.2 Conserved genes are more responsive and show conserved phenotypic signatures

3.2.1 Orthologous mutants behave similar across conditions

The species *Escherichia coli* and *Salmonella enterica* are closely related and many of the chemicals tested in each study are antibiotics targeting core cellular processes. Given the close relation of the two organisms, I expect genes involved in these pathways to respond similar to treatment with sub-inhibitory concentrations of antibiotics. Also the response to physical stresses should be largely conserved.

To investigate this hypothesis, we determined the similarity in the phenotypic responses of orthologous mutants by comparing their s-score signature correlation.

As mentioned before, the calculation of meaningful correlations depends on the presence of significant phenotypes. Thus, it is not surprising that correlations derived from orthologous pairs with no significant phenotypes are centered around zero (Figure 23, panel 'blank'). However, when requiring at least one significant phenotype for each mutant, the correlation distribution shifts to higher values (note the tail towards high correlations in Figure 23, panel 'phenotype'). This effect is strongly increased when at least one mutant of each pair is very responsive and presents significant phenotypes in at least 10% of the shared

conditions. In this case we observe a clear shift towards high correlations when compared to random mutant pairs from the same subset (compare panel 'important' to panel 'random important' in Figure 23).

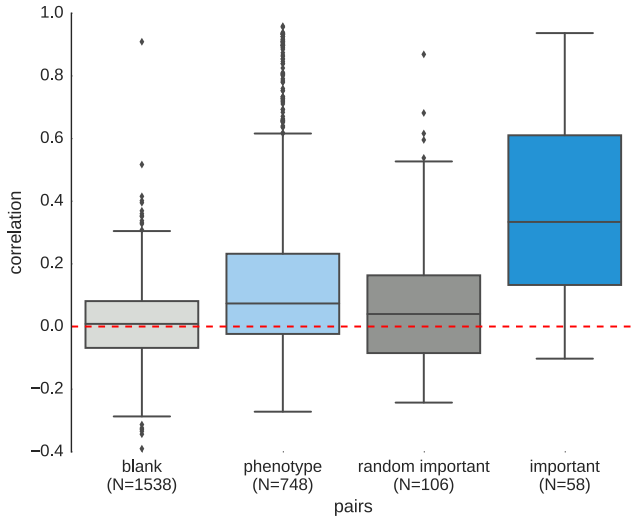

Figure 23: Mutants of orthologous genes behave similar across conditions.
S-score signature correlation (Pearson) of orthologous gene pairs with no phenotypes (blank), or with the requirement of one mutant having at least one phenotype (phenotype), or with the requirement of one mutant having a significant phenotype in more than 10% of the shared conditions (important) compared to a set of random correlations from the same mutant subset (random important).

Nevertheless, a few ortholog pairs remain with low correlation despite being wired deeply into the cellular networks. I plan to investigate these cases in the future. Possible explanations are either biological (one organism may have repurposed the gene, or has a redundant player) and/or technical (polar effects in one organism confounding the phenotypic signature of the gene, incorrect mutants on one side etc.).

In conclusion, I find mutants of orthologous genes central to the cellular network (many phenotypes) to behave largely similar across conditions.

3.2.2 Highly responsive genes are often highly conserved

Genes conserved across organisms are often involved in central cellular pathways. In contrast, species-specific genes and/or those acquired recently in evolution are believed to be involved in niche-specific tasks. Thus, one can expect that species-specific genes are only relevant for fitness in special conditions. Highly conserved genes on the other hand would be expected to respond to a variety of perturbations targeting central cellular pathways.

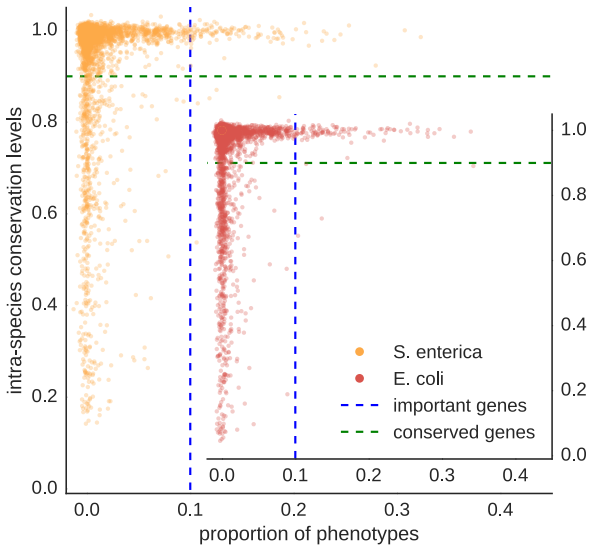

Figure 24: Highly responsive genes are often highly conserved.
Highly conserved genes are wired deeper into the cellular network and are more likely to exhibit a high number of interactions with chemicals. For each gene the proportion of significant phenotypes observed in the shared conditions is plotted against its intra-species conservation level (large-scale blast score ratio).

In Figure 24 we plot the proportion of conditions in which a mutant shows significant phenotypes against its conservation level within the species. Note that for each organism the respective complete data set is considered. Indeed, I observe that only highly conserved genes are wired deeply into the cellular network and exhibit a great amount of gene-drug interactions, while poorly conserved genes usually show only very little significant phenotypes.

Interestingly, there are a few exceptions. For *E. coli* we find the mutants of genes involved in LPS biosynthesis (*waaQ, waaP*) and one prophage element (*ylcG*) as highly responsive despite poor conservation. For *Salmonella* we find mutants of the *rfb* cluster encoding genes for O-antigen biosynthesis (*rfbDIFGJ* and *rfbUNP*). The *rfb* gene cluster is one determinant used for serotyping of *Salmonella* strains as it is known to be very diverse between serovars. In line with this we observe low intra-species conservation for these genes.

Disruptions in LPS and O-antigen biosynthesis will alter outer membrane properties of bacteria. The outer membrane constitutes the major barrier against chemical perturbations and disruption of its composition impacts entry of many compounds [79].

3.2.3 Responsive mutants show a similar distribution across the chromosome

The above observations are also reflected in the distribution of responsive mutants (phenotypes in at least 10% of the shared conditions) across the chromosome of each organism.

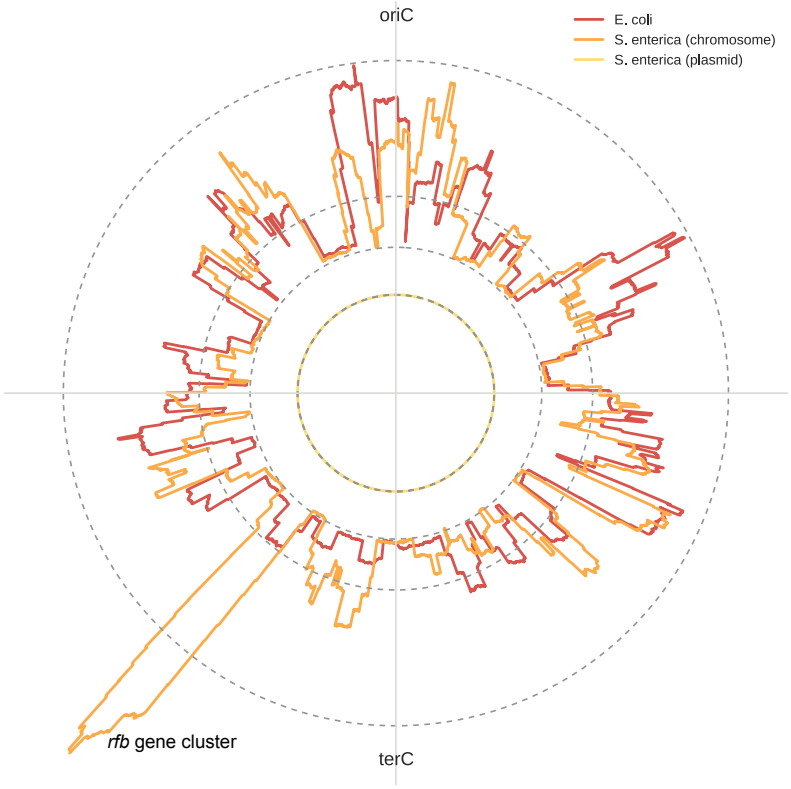

Figure 25: Distribution of responsive mutants across the *E. coli* and *Salmonella* chromosomes.
For each organism the circular plot depicts the chromosome with coordinates adjusted to start at the origin of replication (oriC = 0 bp). We plotted the spatial enrichment for genes exhibiting phenotypes in at least 10% of the shared conditions in a 100 kb sliding window. The middle dashed line indicates zero enrichment, positive enrichment is plotted towards the outside with the outer and inner dashed line representing the maximum and minimum permutation thresholds, respectively (for details see page 96 in the Materials and Methods section). The Salmonella plasmid is represented in a separate trace in the center of the plot and a 10 kb sliding window was used for its analysis.

In Figure 25 we show chromosomal regions enriched for highly responsive genes (peaking towards the outside). Regions enriched for or reduced in responsive genes largely overlap between *E. coli* and *Salmonella* with small shifts accounting for insertions or deletions in one or the other organism (so the responsive genes are actually the same). This reflects the broad conservation of phenotypic responses observed between orthologs.

Consistent with the previous notion of the highly responsive *rfb* gene cluster in *Salmonella,* we find the chromosomal region in which it is encoded strongly enriched in responsive mutants (strong peak in lower left quadrant of Figure 25). The *E. coli* K12 lab strain is known to lack production of O-Antigen due to mutations in the *rfb* region [80]. In fact, a large part of the *rfb* gene cluster is missing explaining the lack of phenotypes in this region and the subsequent shifts between *E. coli* and *Salmonella* peaks.

As previously observed by Nichols et al. for *E. coli* [36], the region around the origin of replication is enriched for responsive mutants while the terminus region is void. This is congruent with the idea of very important (and therefore responsive) genes often being located near the origin of replication so their expression can benefit from the higher gene dosage during replication [81]. In contrast, the plasmid majorly encoding for virulence determinants shows no region enriched for responsive mutants. This is expected considering that the shared conditions do not target virulence-related processes.

In conclusion, comparing single genes across *E. coli* and *Salmonella* revealed that conserved genes often show conserved phenotypic signatures. Furthermore we find that highly responsive genes are often wired deeply into the cellular network.

3.3 Conservation and rewiring of functional modules

After investigating the phenotypic responses of single mutants shared between *E. coli* and *Salmonella*, we next wanted to test how well their interactions, so the similarity between different mutants is conserved across the two organisms.

3.3.1 Predicting conserved functional modules based on phenotypic signatures

As demonstrated in 2.5 Benchmarking a high correlation between s-score signatures of two mutants is indicative of functional relation. Thus, we can use s-score signature correlation to predict clusters of related mutants in each organism but also to identify conserved functional modules.

Identification of conserved clusters

To identify gene modules conserved between *E. coli* and *Salmonella* we used a variant of a method applied in a similar study investigating conservation of functional modules between *S. cerevisiae* and *S. pombe* [55, 82]. In brief, we considered each ortholog pair between *E. coli* and *Salmonella* as one entity. Next, a merge score was computed for each possible combination of entities m1 and m2:

$$\frac{1}{|m1|+|m2|}\sum_{a\epsilon m1}\sum_{b\epsilon m2} r_{ecoli}(ab) + \frac{1}{|m1|+|m2|}\sum_{a\epsilon m1}\sum_{b\epsilon m2} r_{salmonella}(ab)$$

where r is the Pearson correlation between gene a and gene b across the shared conditions. If the sum of correlations in *E. coli* or *Salmonella* was below zero, the merge score was set to zero for this pair of entities. A high merge score was obtained when s-score signatures of genes a and b correlate strongly in both organisms.

Once the merge scores had been computed for each possible entity pair, we sorted them and started merging the pairs with the highest merge score into modules. Each time a pair was merged, the merge score between this expanded module and all other entities was recomputed and all the scores were sorted again. This procedure was repeated until the merge score fell under a defined threshold, which was chosen as 0.4 based on the benchmarks. The benchmarking performance for the selected threshold is shown in Figure 26.

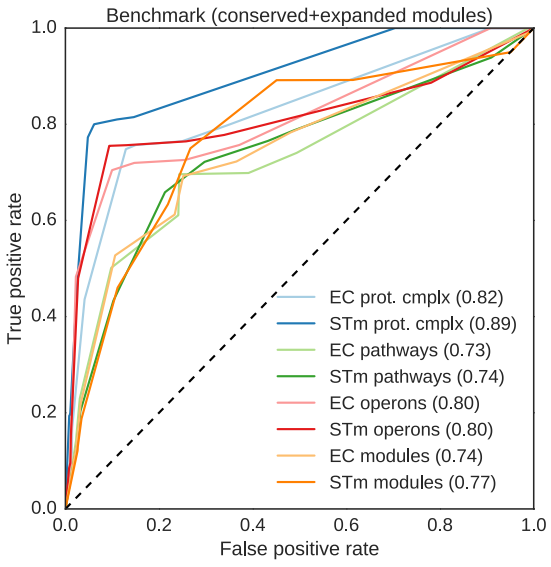

Figure 26: Benchmarking of conserved modules and their expansions.
Mutant pairs present in the conserved modules predicted based on a merge score of 0.4 are tested against gene pairs present in Biocyc complexes and pathways as well as DOOR operons for the respective organisms (STm complexes inferred from EC, for database references see Table 5 on page 94 of the Materials and Methods section).

Conserved modules represent various cellular processes

Within the conserved modules we often find protein complexes, cellular pathways or entire operons. Figure 27 gives a few, selected examples and demonstrates that we find conservation in all kinds of cellular processes. For instance, we find components of the Rcs signal transduction system to behave similar in both organisms. The module also includes *wcaF,* a gene involved in colanic acid biosynthesis, which is regulated by the Rcs system [83]. LpoB was discovered as the regulator of peptidoglycan synthesis factor PBP1B (encoded by *mrcB*) during follow up of *E. coli* genetic interaction studies [51]. This interaction is also reflected in the cellular network of *Salmonella.* Additionally, *dedD* and *yajG* are associated with this cluster. DedD is involved in cell division [84], which has to be closely linked to the synthesis of new cell wall material. *YajG* is located upstream of *ampG* and harbors the promoter region for *ampG* in its open reading frame. Thus, deletion of *yajG* likely abolishes expression of *ampG,* which is involved in recycling of cell wall material [85]. Furthermore we find many clusters involved in metabolism such amino acid biosynthesis, sugar uptake and breakdown. One example is represented in Figure 27 showing genes involved in maltose uptake and breakdown clustering into a functional module.

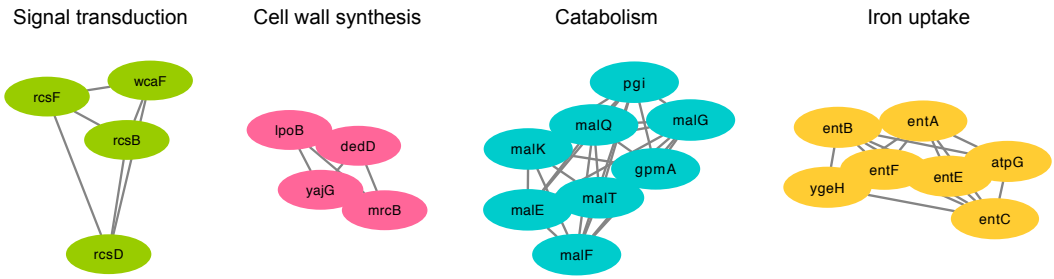

Figure 27: Examples of functional modules conserved between *E. coli* and *Salmonella.*
Nodes represent ortholog pairs, the *E. coli* gene name was chosen for display. Edges represent the merge score.

Interestingly, we also find genes of unknown function within conserved functional modules. This presents a great starting point for follow up studies as the connections are extremely robust due to the combination of data sets from different organisms obtained independently in different labs.

One example is the *E. coli* orphan gene *ygeH,* found in one module with genes involved in enterobactin synthesis (Figure 27). Enterobactin is a siderophore necessary for iron uptake. *YgeH* is a predicted transcriptional regulator located within the remnants of a pathogenicity island fully present in pathogenic *E. coli*

and *Shigella* strains, where it has been shown to be involved in virulence [86, 87]. Interestingly, it appears to be orthologous to the *Salmonella hilA* gene (passing the threshold for calling orthologs but genomic context is different). HilA is a transcriptional regulator activating invasion genes and has been indicated to be regulated by iron availability [88, 89]. Limited iron availability is a typical signal for the bacterium to be inside the human host. Therefore, it can be assumed that *hilA* and *ygeH* are activated upon iron limitation and in turn positively regulate the expression of enterobactin to ensure sufficient iron uptake.

As a next step I plan to systematically detect orphan genes within conserved modules and investigate their phenotypic signatures in more detail. This will aid creating hypotheses about their function and designing targeted validation experiments.

However, this analysis only includes orthologous pairs with conserved phenotypic signatures. In order to investigate the role of species-specific genes and those behaving differently in the two organisms we expanded the functional modules by testing their correlation to the remaining genes.

3.3.2 Expanding conserved modules with species-specific genes reveals their wiring into the cellular network

Expansion of conserved clusters with species-specific genes

We expanded the conserved modules to include genes with behavior specific to either species. For each organism we collected all genes not already represented in a conserved module. Next, we calculated the median Pearson correlation of the sole gene with all other genes within a given conserved module. The median value was put equal to zero if any of the correlations with the genes present in the module were found to be below zero. If the median correlation met a certain minimum we added the sole gene to the highest correlated conserved module. The benchmarking performance of the resulting expanded modules is shown in Figure 28A. The minimum required median correlation was chosen as 0.3 based on ROC curve benchmarking only considering gene pairs between the expanded and conserved modules. An example demonstrating the benchmarking performance based on this threshold for the EC specific expansion is shown in Figure 28B.

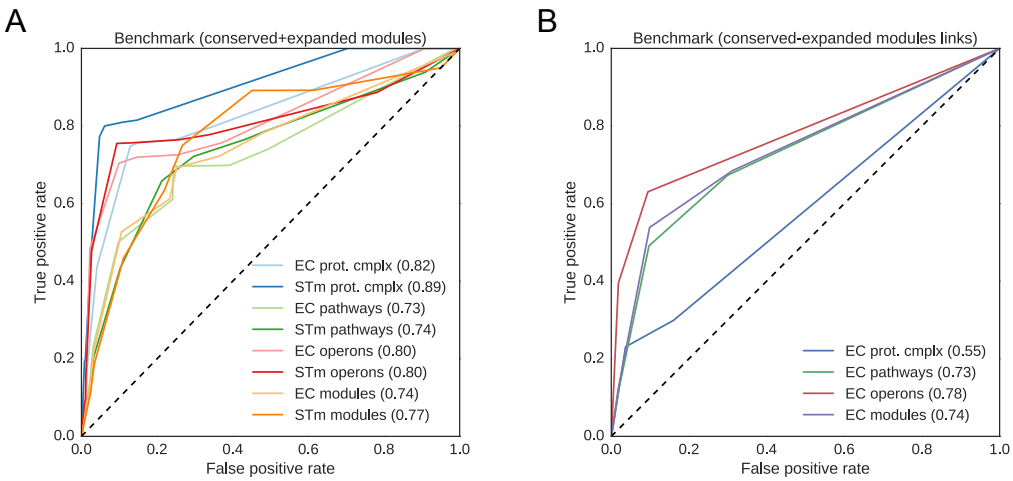

Figure 28: Benchmarking of expanded gene modules.
A) Mutant pairs present in the expanded modules were tested against gene pairs present in several benchmarking sets. The minimum correlation required to add a gene to a conserved module was 0.3. B) Same as A) considering only gene pairs where one is present in a conserved module whereas the other is a species-specific addition.

Figure 29A illustrates the diversity of functional modules. While some modules remain entirely conserved, others have a few species-specific additions or are composed almost entirely of genes with distinct responses in both organisms.

One specific example is shown in Figure 29B, where I show the expansions for the previously discussed conserved module containing the peptidoglycan synthesis gene *mrcB* and its regulator *lpoB*.

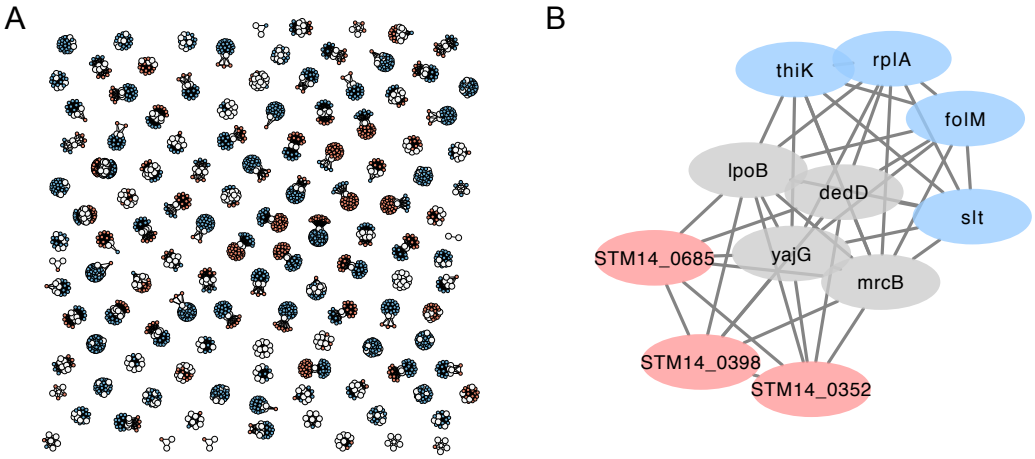

Figure 29: Functional modules predicted based on phenotypic signature similarity show varying levels of conservation.
Conserved functional modules predicted based on similarity of phenotypic signatures and their species-specific expansion. White/grey nodes represent orthologs of *E. coli* and *Salmonella* found to be in conserved modules, edges between them are based on their merge score. Blue nodes represent *E. coli* specific genes, red nodes represent *Salmonella* specific genes. Edges involving species-specific genes are based on s-score signature correlation across the shared conditions.

In *E. coli* we find the soluble lytic transglycosylase (*slt*) involved in peptidoglycan recycling associated with the module [90]. The corresponding *Salmonella* gene was not represented by a mutant in the library and can therefore not be connected to the module. It would be interesting to investigate why *folM* (tetrahydrofolate biosynthesis), *thiK* (thiamin kinase), *rplA* (ribosomal protein of the 50S subunit) and *ybdZ* (STM14_0685, orphan gene) are connected to the cluster. For this purpose the first step would be to examine the conditions in which the phenotypes overlap with the conserved part of the module.

This analysis furthermore allows us to investigate how species-specific genes are linked to conserved modules. In the example depicted in Figure 29B we find two *Salmonella*-specific genes *safA* (STM14_0352) and STM14_0398 (putative inner membrane protein). SafA is the major subunit of *Salmonella* atypical fimbriae, a virulence factor with largely unexplored function discovered by Folkesson et al. [91, 92]. Expression of fimbriae has to be tightly coordinated with rearrangements in the cell wall and, interestingly, the same group later on described the impact of peptidoglycan recycling components on invasion and intracellular survival [93]. Thus it would be interesting to investigate this link further and the second *Salmonella*-specific gene should be considered as potential mediator.

In conclusion, the expanded module analysis allows us to identify shared genes with differing responses. However, the presence of a mutant in each data set has to be considered. Therefore it would be useful to overlay this information with the network for easier exploration. Furthermore, softening the correlation threshold might identify cases detected as species-specific because one ortholog is just above the threshold while the other one is just below. Allowing additions where at least one ortholog is above the correlation threshold and the other one has the same tendency (low residual between correlations) would tackle this problem. The big benefit of this analysis approach is that we can see the role of genes present in only one species.

3.4 Drug response conservation

In chemical genomics studies we not only obtain a phenotypic signature for each mutant, but also for each drug. Significant phenotypes can indicate genes involved in the uptake or efflux of the drug as well as cellular processes targeted. Furthermore, similarities between phenotypic signatures of conditions across

genes can indicate shared (primary or secondary) targets, or mechanisms of uptake and efflux. Using the same principles as applied for gene modules we calculated a merge score for shared conditions between *E. coli* and *Salmonella* and grouped drugs with related responses into modules.

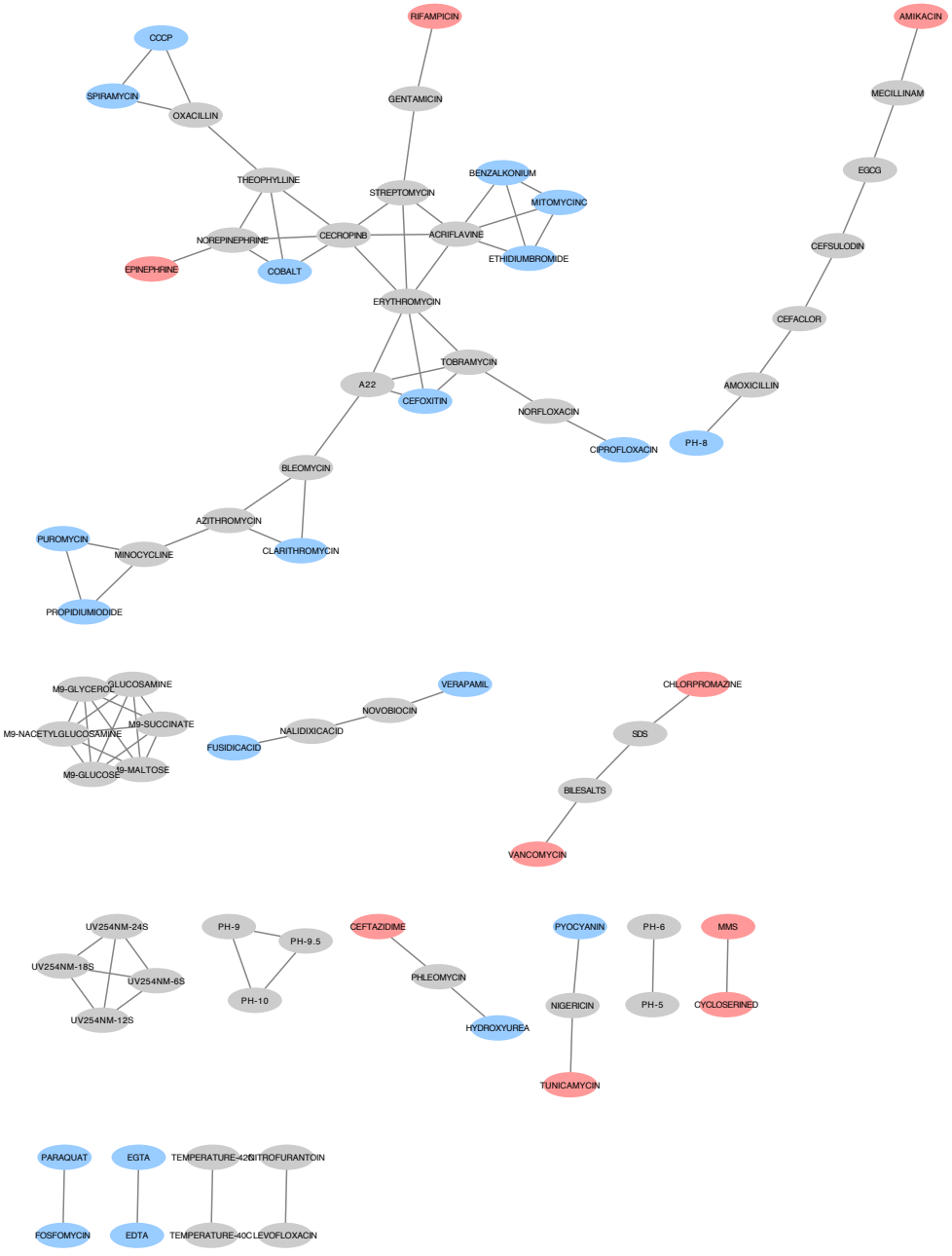

Figure 30: Network of drugs with conserved phenotypic signatures and species-specific correlated drugs. Drugs with conserved phenotypic responses in both organisms are shown as grey nodes and the edges between them are based on the merge score. Species-specific additions are shown in blue for *E. coli* and in red for *Salmonella.* Edges involving species-specific genes represent s-score correlation with the connected drug within the organism.

Figure 30 demonstrates that physical stresses evoke highly conserved responses. Conditions falling in the same category, such as exposure to UV, high temperature, low or high pH cluster into modules with phenotypic signatures conserved in both organisms (grey nodes). The same can be observed for minimal media conditions with differing carbon sources. The responses to the majority of antibiotics are conserved. Furthermore, drugs targeting the same process are frequently connected in the network, e.g. nitrofurantoin and levofloxacin, cefaclor and cefsulodin or bile salts and SDS.

Notably, many of the perturbations with differing responses in both organisms evoke envelope stress (benzalkonium, vancomycin, EDTA to name just a few). It would be interesting to test, whether this difference is due to the missing O-antigen in *E. coli.* Generally, we plan to examine differences between the two organisms as follows: Within a module of drugs we can apply gene ontology (GO) enrichment analysis for each drug in each organism to identify the cellular processes targeted. We can then compare drug responses based on both, the merge score (overall similarity) and the overlap in GO processes targeted. This way differences and commonalities between the two organisms can be related to distinct biological processes related to uptake, efflux or target of the drug.

3.5 Conclusions

Here I present the first cross-species comparison of chemical genomics data in bacteria. We established that conserved genes frequently show conserved phenotypic signatures. Furthermore we find that responsive genes are usually highly conserved and wired deeply into the cellular network whereas species-specific genes, with some exceptions, show only few responses.

We implemented a strategy do detect modules of functionally related genes based on their chemical genomics profile. We find that many cellular processes are reflected in modules conserved in *E. coli* and *Salmonella.* Next, we expanded these modules to detect connected genes with species-specific responses. Both strategies provide insights into orphan gene function. In the future we would like to probe the conservation of connections between modules, as this could reveal conserved and rewired cross-talk between pathways.

Lastly, I compare drug signatures between the two organisms. Whereas many stress responses are conserved, we get hints that certain envelope perturbations evoke distinct responses.

3.6 Contribution disclaimer

The comparative analysis presented in this chapter was designed in close collaboration with Marco Galardini (EBI, Hinxton). Marco Galardini led the implementation of the analysis pipeline, but all steps were first extensively discussed between the two of us and revised for refinements by me.

4 The *Salmonella* efflux pump SmvA confers resistance to the diabetes drug metformin

In this chapter I will describe the role of a rare efflux pump, SmvA, in conferring resistance to the widely prescribed type 2 diabetes (T2D) drug metformin. The drug was included in the chemical genomics due to a previously observed high resistance level of *Salmonella* compared to other organisms (Ana Rita Brochado, personal communication). The obtained gene-drug interaction map revealed the *Salmonella*-specific SmvA as a major player in drug efflux. Introducing the pump into the much more metformin sensitive *E. coli,* thereby conferred resistance to metformin. Furthermore, I explored the wiring of other pumps capable of exporting metformin into the cellular networks of *E. coli* and *Salmonella.*

4.1 Background

4.1.1 *Salmonella* shows high levels of resistance to metformin

I included the T2D drug metformin in my study due to an interesting observation made by my colleague Ana Rita Brochado. For her project she was determining the MIC of a broad range of drugs for a variety of bacteria. Interestingly, *Salmonella* Typhimurium showed the highest level of resistance to metformin amongst the Gram-negative bacteria she tested (Figure 31). It was even more resistant than *Pseudomonas aeruginosa,* which is known for its multi-drug resistant properties [94].

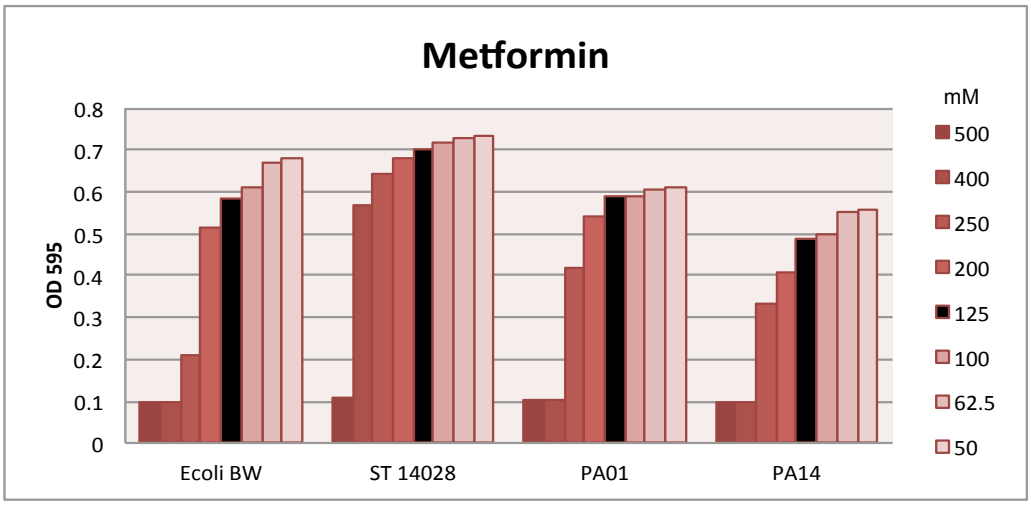

Figure 31: Metformin dose responses of *E. coli* BW25113, *S.* Typhimurium 14028s and *P. aeruginosa* PA01 and PA14.
Salmonella is the most resistant among the three organisms with only mild growth inhibition at 400 mM metformin, a concentration at which all other bacteria are completely inhibited. Figure provided by Ana Rita Brochado.

4.1.2 Metformin reduces hepatic glucose production

Metformin, a biguanide drug, is currently the first-line oral treatment against hyperglycemia in T2D. The use of this drug is extensive with prescriptions for at least 120 million patients worldwide. It was shown to reduce the risk of complications in overweight diabetic patients and, compared to insulin or sulfonylureas, result in fewer hypoglycemic attacks and less extensive weight gain [95, 96].

Although the drug was introduced in the clinic many years ago, the complete mechanism of action remains elusive. However, it is generally accepted that a decrease in hepatic glucose production and consequently reduction in blood glucose levels are the main beneficial result of the treatment [95].

Under physiological conditions metformin appears in a positively charged form explaining both, the preferential action on hepatocytes, as those express comparably high levels of the organic cation transporter 1 (OCT1, Figure 32), as well as the accumulation of the drug in active mitochondria. Furthermore, in 2000, two independent reports showed a mild but specific inhibition of the respiratory chain complex I by metformin. The resulting reduction in cellular energy and the associated increase in the AMP/ATP ratio explain the observed activation of AMPK (AMP activated protein kinase). Activation of AMPK switches the cell from an anabolic to a catabolic state, resulting in increased glucose uptake and fatty acid oxidation to restore the cellular ATP pool, while

ATP-consuming biosynthetic pathways are shut down. Here, the key aspect in metformin's beneficial effect on hyperglycemia is the inhibition of gluconeogenesis [95, 97]. Additionally, metformin has been shown to increase insulin sensitivity, adding to the improvement of metabolic parameters in diabetic patients [98].

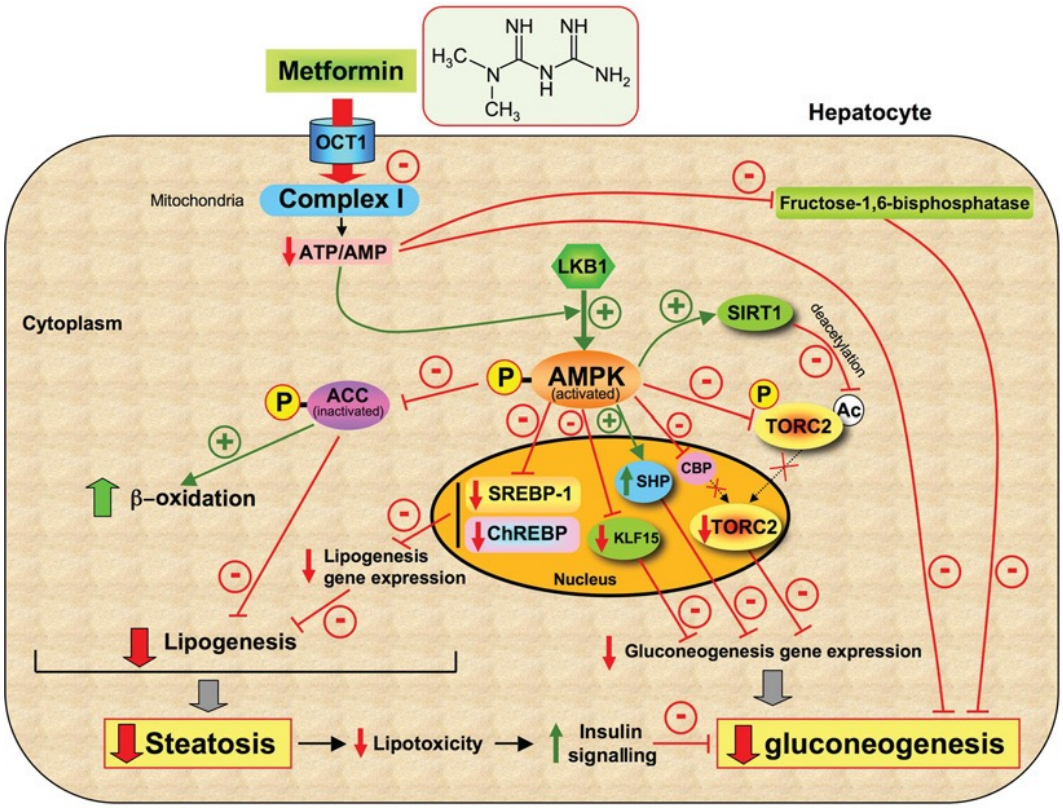

Figure 32: Proposed mechanism of metformin action in hepatocytes.
Metformin is taken up by the organic cation transporter 1 (OCT1) and exerts an inhibiting effect on the respiratory chain complex 1 in mitochondria. The resulting decrease in cellular energy levels and thereby increase of relative AMP levels yield AMPK activation. Consequently gluconeogenesis is inhibited by reduced gluconeogenesis gene expression as well as allosteric inhibition of involved enzymes. Furthermore, lipogenesis is decreased (reducing the risk of complications due to steatosis) and insulin sensitivity is increased contributing to the control of glucose excretion by the liver. Adapted from Viollet et al. [95].

Taken together, the discussed effects of metformin on the hepatocyte metabolism explain its anti-hyperglycemic properties.

Notably, new mechanisms of action are being discovered constantly [99, 100]. For example, metformin has been proposed to have anti-neoplastic properties and to protect against several diabetic complications, both, in AMPK-dependent and –independent pathways [95, 99, 101, 102]. Furthermore, the drug is widely used off-label for the treatment of polycystic ovary syndrome, steatohepatitis and HIV-associated metabolic abnormalities [103] and is the first drug in clinical trials

addressing longevity (MILES – Metformin in Longevity Study, Identifier: NCT02432287, TAME – Targeting Aging with Metformin, currently waiting for approval, [104])

4.1.3 Metformin affects the gut microbiota

Metformin is only effective upon oral administration. In fact, already in 1984, Bonora and colleagues reported that metformin has no acute, hypoglycemic effect on non-diabetic subjects when administered intravenously [105]. This poses the question whether metformin affects the human gut microbiota and whether this could contribute to the beneficial effect of the drug.

A recent study conducted in mice on a high-fat diet showed that metformin treatment besides improving glucose homeostasis indeed shifted the microbiota composition. Interestingly, the microbial composition shifted towards a community found in mice consuming a regular chow diet [106]. *Akkermansia*, bacteria capable of using mucin as a nutrient source dominated after metformin treatment (also reported in [107]). The authors furthermore show that oral administration of *Akkermansia* was able to improve glucose homeostasis and insulin sensitivity in mice on a high-fat diet [106]. A similar study showing the reverting effect of metformin on a high-fat diet as well as changes in the microbiota was recently conducted in rats [108]. The above results suggest that the beneficial effect of metformin is at least partially due to changes in the microbiome.

Furthermore, Forslund et al. recently found a significant influence of metformin treatment on changes in the gut flora of diabetic patients. The authors use metagenomics to integrate patient data from multiple countries and compare non-diabetic controls to diabetic patients with or without metformin treatment. Significant differences between the untreated diabetic patients and the non-diabetic controls were observed, accounting for the dysbiosis caused by the diabetic condition itself. However, the authors also detect significant differences within the diabetic patients dependent on their metformin treatment status. Upon metformin treatment they report an increase in *Escherichia* subspecies as well as a decrease in *Intestinibacter* subspecies compared to untreated diabetic patients. Also, consistent with previous reports in animals, in some of the used data sets metformin treatment reverted *Akkermansia* abundance towards levels found in non-diabetic controls [109].

The ability of metformin to affect host health status indirectly has been furthermore demonstrated in a study in *Caenorhabditis elegans*. A positive effect of metformin on *C. elegans* longevity, reported in 2010, was later shown to be caused by alterations in the metabolism of *E. coli* used as a food source for the worm [110, 111]. The authors observe that metformin disrupts the folate metabolism of *E. coli.* Consequently the amount of folate and methionine available to the worm is reduced mimicking a beneficial diet restriction. The dose-dependent effect of metformin was not observed when the worm was fed on metformin-resistant bacteria [110]. Taken together these results suggest that alterations in bacterial metabolism can contribute to how drugs affect the host.

4.1.4 Motivation

Taken together, the above results suggest that metformin might have an extremely complex mechanism of action influencing host metabolism as well as microbiota composition, potentially allowing the creation of a niche for enterobacteria.

Therefore, I was very interested in investigating the source of the high resistance level of the pathogenic enterobacterium *Salmonella* Typhimurium to metformin. Potential implications on susceptibility of diabetic patients to *Salmonella* infection are discussed at the end of this chapter.

4.2 Results

4.2.1 The efflux pump SmvA contributes to resistance against the type II diabetes drug metformin

To identify players contributing to the high resistance level of *Salmonella* Typhimurium to the T2D drug metformin, I first determined gene deletion mutants exhibiting sensitivity to the drug in my chemical genomics data set. As described in 2.2.5 Scoring and clustering (see page 26) increased susceptibility of a mutant towards a specific drug will result in negative s-scores.

Table 2 shows the top ten most sensitive mutants at the highest metformin concentration tested. At the top of the list are both deletion mutants of the gene *ndh* coding for an NADH dehydrogenase (NDH). *Salmonella* has two distinct NDHs, NDH-1 encoded by the *nuo* cluster of genes and NDH-2 encoded by *ndh.*

NDHs constitute complex I of the respiratory chain and *E. coli* has been shown to utilize both NDHs during glucose limited aerobic growth [112].

Gene Name	Locus Tag	Library	S-score at 100 mM Metformin
ndh	STM14_1385	Cm	-26.05
ndh	STM14_1385	Kan	-21.89
smvA	STM14_1900	Cm	-13.46
sixA	STM14_2936	Kan	-9.60
dam	STM14_4196	Kan	-8.88
smvA	STM14_1900	Kan	-8.02
	STM14_2111	Cm	-7.71
sapA	STM14_2042	Cm	-7.53
ylbF	STM14_0621	Cm	-7.33
hisA	STM14_2570	Cm	-7.02

Table 2: Metformin sensitive mutants.
The top ten most sensitive mutants to 100 mM metformin and their corresponding s-scores are shown in rank order.

As metformin inhibits complex I of the respiratory chain in mitochondria it is likely that the bacterial complex I is inhibited as well. Interestingly, the mutants of both NDHs show distinct phenotypes upon treatment with metformin in *Salmonella.* Whereas *ndh* deletion leads to strong negative interaction, deletion of any *nuo* component has the opposite effect (Figure 33A). The same trend was observed in a similar *E. coli* chemical genomics data set (data not shown).

Deletion of one NDH forces the bacterium to use the other one. The positive interactions observed for the *nuo* mutants mean that metformin treatment is irrelevant upon deletion of NDH-1, therefore suggesting it as the drug target. In line with this premise we observe a synthetic sick interaction when deleting NDH-2 and simultaneously inhibiting NDH-1 by treatment with metformin.

NDH-2 exclusively uses NADH and flow of electrons to ubiquinone does not result in an electrochemical gradient [113]. If metformin uptake depends on the proton motive force (pmf), this might also explain the discrepancy in phenotypes. Upon knock-out of NDH-1, *Salmonella* is forced to use NDH-2 resulting in reduced pmf and potentially reduced uptake of metformin explaining the positive gene-drug interaction. In contrast, deletion of NDH-2 might result in a stronger pmf leading to increased metformin uptake and synthetic sickness.

In addition to *ndh*, the *smvA* deletion mutants of both libraries exhibited dose-dependent negative interactions in all metformin concentrations tested and

were amongst the top ten negative scores in the highest concentration (Figure 33 B and C as well as Table 2).

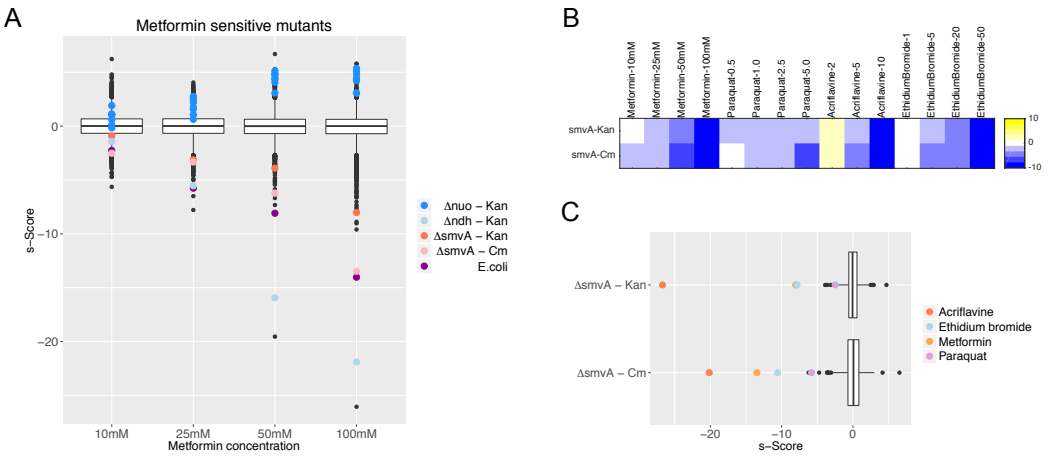

Figure 33: Mutants of *smvA* show high sensitivity to metformin and other substances.
A) Distributions of s-scores across all mutants for all metformin concentrations tested. Mutants of the respiratory complex I show positive interactions with metformin (all *nuo* mutants present in the Kan library are shown, namely *nuoABEGHJKN*), whereas a respiratory complex II mutant (*ndh*) interacts negatively with metformin. Both *smvA* mutants as well as *E.coli* show high sensitivity to metformin reflected in strongly negative s-scores with increasing metformin concentration.B) s-Scores of the *smvA* Kan and Cm clone in selected conditions. C) Distribution of s-scores across all conditions for both *smvA* deletion mutants. S-scores in the highest concentration of chemicals exported by the SmvA pump are indicated.

SmvA is an inner membrane transporter of the major facilitator superfamily (MFS) [114]. In *S.* Typhimurium SmvA has been reported to be responsible for efflux of paraquat, also called methyl viogen, hence the name (deletion resulting in **s**ensitivity to **m**ethyl **v**iogen) [114-116]. Furthermore, SmvA is involved in the efflux of paraquat, acriflavine and ethidium bromide. Indeed, both *smvA* mutants tested show negative interactions with these drugs and their highest concentrations result in the most negative s-scores observed for *smvA* mutants across all conditions (Figure 33 B and C).

Already when the *smvA* nucleotide sequence was first reported, it was pointed out that genomic organization in the surroundings of the gene differ between *Salmonella* and *E. coli* [115], and indeed the pump is not present in the closely related bacterium. In fact, SmvA is only present in selected *Salmonella, Klebsiella* and *Acinetobacter* subspecies amongst a few exceptions (STRING database, Feb 2016 and [114]). These bacteria are not directly related, suggesting that *smvA* was acquired multiple times independently, likely by horizontal gene transfer.

Interestingly, wild-type *E. coli,* which I had included in the library array, behaves similar to the *smvA* deletion strains, scoring very negative in all metformin concentrations (Figure 33A). This is consistent with the greater

sensitivity compared to *Salmonella* observed before (Figure 31) and supports the hypothesis that the SmvA pump not present in *E. coli* might play a major role in metformin export.

To investigate the role of SmvA in metformin resistance in more detail, I measured the growth of wild-type *E.coli* (EC), wild-type *Salmonella* (STM) and the *Salmonella* Δ*smvA*::kan mutant in increasing concentrations of metformin. First however, both *smvA* mutants were confirmed by PCR and re-transduced into wild-type background to exclude secondary mutations. Both mutants behaved the same, before and after re-transduction (data not shown), therefore I only use the Δ*smvA*::kan mutant (in the following referred to as STM Δ*smvA*).

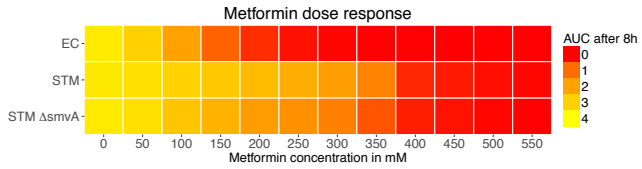

Figure 34: Metformin dose response for wild-type *E. coli* (EC) and wild-type *Salmonella* (STM) and a *Salmonella smvA* deletions strain (STM ΔsmvA).
OD (578nm) was measured every half an hour upon incubation in LB with different concentrations of metformin, shaking at 37°C. Area under the curve (AUC) after 8 h is displayed.

As shown in Figure 34 *E. coli* is very sensitive to metformin, with a significant drop in growth (represented by area under the curve after 8h) at as little as 100 mM metformin. Wild-type STM on the other hand is rather resistant and severely impacted only at 400 mM. Confirming the results obtained in the high-throughput screen (see Figure 33), the *smvA* deletion mutant is more sensitive than wild-type STM. However, sensitivity does not reach the level of EC suggesting there are other factors contributing to the difference in resistance between EC and STM.

4.2.2 Efficient drug efflux may be critical for resistance to biguanide drugs

Although differences in drug uptake and primary target sensitivity may well contribute to the difference in resistance between EC and STM Δ*smvA*, I focused on the potential role of other efflux pumps, because preliminary results from experiments investigating the response of gut bacterial species alongside *E. coli* and *Salmonella* to metformin and phenformin suggested that efficient drug export might be the key to *Salmonella's* high resistance levels. Phenformin is a compound

closely related to metformin but capable of passing the bacterial envelope more easily due to a large hydrophobic side group (Figure 35).

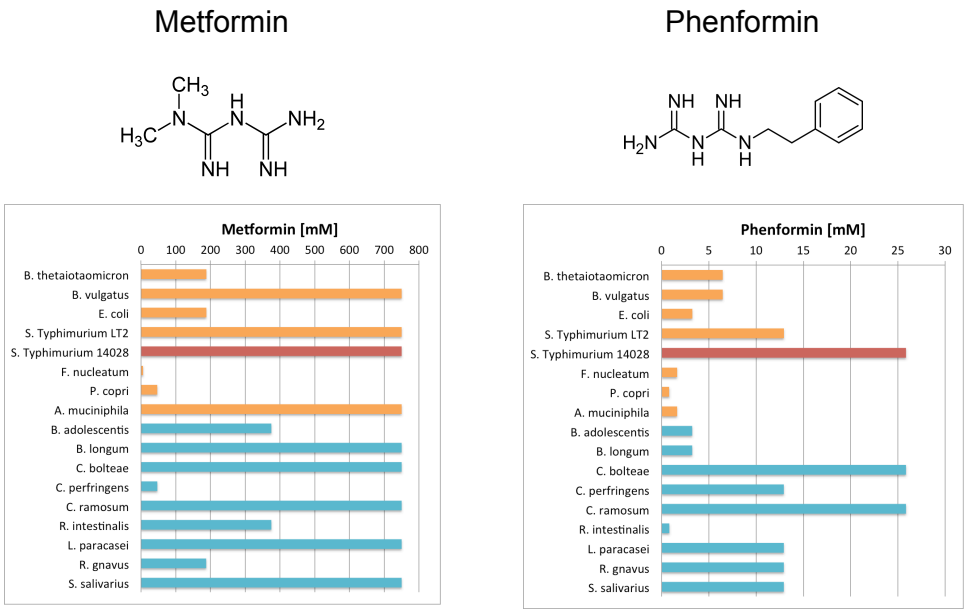

Figure 35: Dose response of gut bacterial species to metformin (left panel) and phenformin (right panel). The bacteria were grown anaerobically in MGAM, at 37°C in a two-fold dilution series of the indicated drug and OD (578 nm) was measured every 30 min. Shown is the lowest concentration at which no growth was observed. Gram-negative species are shown in orange, with the *Salmonella* wild-type strain used in this study highlighted in red. Gram-positive organisms are shown in blue. Numbers are retrieved from two within plate replicates (cultures were grown in 384 well plates).

Figure 35 shows the lowest deadly concentration of metformin (left panel) and phenformin (right panel) for several gut bacterial species in comparison to *Salmonella* and *E. coli.* Amongst the Gram-negative bacteria tested, *Salmonella, Bacteroides vulgatus* and *Akkermansia muciniphila* show high resistance to metformin. This is worth noting, as an increase in *Akkermansia* subspecies was reported for metformin treated mice and humans [106, 109].

In phenformin however, the two *Salmonella* strains tested are clearly the most resistant among the Gram-negatives. As phenformin enters the cell more easily, efficient drug efflux is probably critical for decreasing intracellular drug concentration and thereby avoiding detrimental effects.

With this in mind the role of other efflux pumps should be considered for explaining the remaining difference between EC and STM Δ*smvA* metformin sensitivity as deletion of *smvA* might induce the expression of another pump by feedback regulation.

First, *Salmonella* may possess more species-specific efflux pumps involved in metformin efflux. Although no other pumps are detected in the metformin top

hits of the chemical genomics data, induction of a second pump upon *smvA* deletion might explain this phenomenon. In this case a single deletion of the second pump would not result in a phenotype.

Second, a pump shared by both organisms but with differential expression levels might be involved. If expressed higher in *Salmonella* this would explain the discrepancy between EC and STM Δ*smvA*. Both, different base line expression levels as well as different levels of induction upon stress should be considered.

4.2.3 Multiple efflux transporters present in both organisms are capable of exporting metformin

To explore the possibility of other pumps involved in bacterial metformin efflux, I examined other large-scale data sets created in the lab, including an EC chemical genomics data set, provided by Lucía Herrera, and an EC overexpression screen, provided by Lisa Maier.

In the EC chemical genomics the KEIO collection of single gene deletions [41] was tested against a variety of stresses following the same method used in this study.

The overexpression screen was performed using a library of ASKA barcoded deletion mutants complemented with their corresponding TransBac plasmid (Hirotada Mori, unpublished resource). The growth of this library was determined at 0 and 100 mM metformin and using different IPTG concentrations to induce expression of the deleted gene from the single copy plasmid. For each IPTG concentration a t-test was applied to compare the relative growth of each strain between drug and no drug condition.

The first thing I noticed was that deletion of the *mdtK* gene resulted in the strongest negative interaction at 40 mM metformin in the EC chemical genomics (Figure 36A). This implied that the corresponding gene product MdtK, a multidrug efflux transporter of the MATE (multidrug and toxic compound extrusion) family, is involved in metformin efflux in EC. Due to technical reasons no data for this mutant is available for higher concentrations. This result was confirmed in the EC overexpression screen. Overexpression of *mdtK* resulted in better growth at 100 mM metformin compared to other strains (Figure 36B).

In the STM chemical genomics presented in this study, however, I could not detect a significant phenotype for *mdtK* deletion mutants (Figure 36C), which can be explained by the presence of *smvA* in these strains. I only observed a mild

negative interaction at the highest concentration. This might hint towards a contribution of MdtK to metformin efflux once the SmvA pump is saturated.

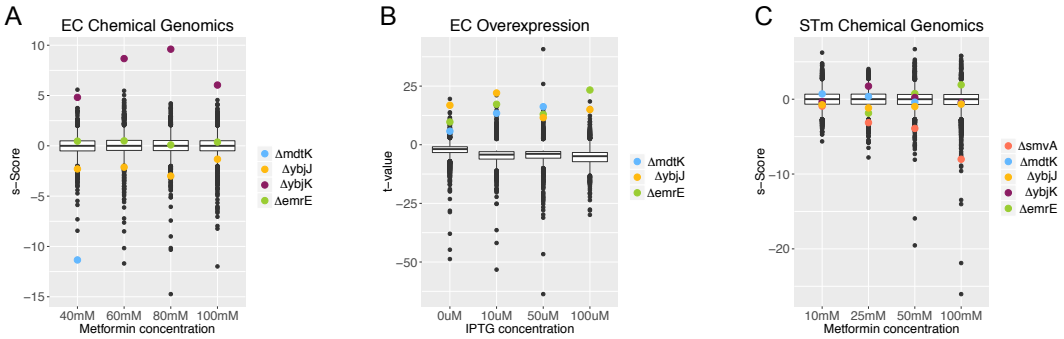

Figure 36: Additional chemical genomics and overexpression data reveal other pumps capable of exporting metformin.
A) EC chemical genomics (Lucía Herrera). S-score distribution across all EC deletion mutants of the KEIO collection in 40, 60, 80 and 100 mM metformin. Due to technical reasons the *mdtK* deletion mutant was only present in 40 mM metformin. B) EC overexpression screen (Lisa Maier). T-value distribution of all complemented deletion mutants using 0, 10, 50 or 100 µM IPTG to induce expression of the deleted gene from the TransBac plasmid. The t-test compared growth of each strain compared to all other strains in drug (100 mM metformin) *vs* no drug. C) STm chemical genomics (this study). S-score distribution of all STM deletion mutants in 10, 25, 50 and 100 mM metformin. Data points for one representative clone of the *smvA, mdtK, ybjJ* and *emrE* mutants are highlighted if present in the respective data set.

The second outstanding observation was the positive interactions of *ybjK* mutants with metformin in the EC chemical genomics (Figure 36A). Positive interactions indicate increased resistance to the chemical perturbation. Interestingly, YbjK (also RcdA) is the repressor of YbjJ [117], a putative MFS transporter. Deletion of *ybjK* would therefore result in enhanced expression of *ybjJ,* which in turn would explain the increased resistance to metformin assuming that YbjJ is capable of exporting metformin.

This result was confirmed in the EC overexpression screen. Overexpression of *ybjJ* facilitated growth at 100 mM metformin (Figure 36B). In contrast, deletion of *ybjJ* does not result in any significant phenotypes in the EC or STM chemical genomics (Figure 36 A and C). Therefore it can be assumed, that YbjJ does not play a role in metformin efflux under physiological conditions when *smvA* and *mdtK* are present in STM and EC, respectively. However, YbjJ appears to be capable of metformin efflux. Hence, it is a potential candidate to explain the discrepancy between EC and STM Δ*smvA*, if induced upon *smvA* deletion or if induced upon metformin stress only in STM but not in EC.

Lastly, the EC overexpression screen revealed EmrE as another multidrug efflux protein able to export metformin. *EmrE* overexpression rescued growth at 100 mM metformin (Figure 36B). Interestingly, EmrE has a very broad substrate

spectrum and overexpression of the pump has been reported to confer resistance to paraquat, ethidium bromide and acriflavine amongst others [118-120]. This profile is similar to the substrate spectrum of SmvA.

Nevertheless, deletion of *emrE* does not result in any significant phenotypes in the EC chemical genomics (Figure 36A). Also, deletion of an *emrE* paralog found in *Salmonella* (STM14_1998) does not yield significant phenotypes in the STM chemical genomics (Figure 36C). Taken together, the above results suggest that *emrE,* just as *ybjJ,* may not be expressed under normal conditions.

However, induction of all three pumps (MdtK, YbjJ, EmrE) upon deletion of *smvA* might be responsible for the discrepancy in metformin resistance observed between EC and STM Δ*smvA*.

4.2.4 SmvA and MdtK are the major players in metformin efflux in *E. coli* and *Salmonella,* respectively

To investigate the role of MdtK, YbjJ and EmrE in metformin efflux, I first confirmed the previously observed sensitivities by measuring the dose response of the corresponding EC and STM deletion strains to metformin.

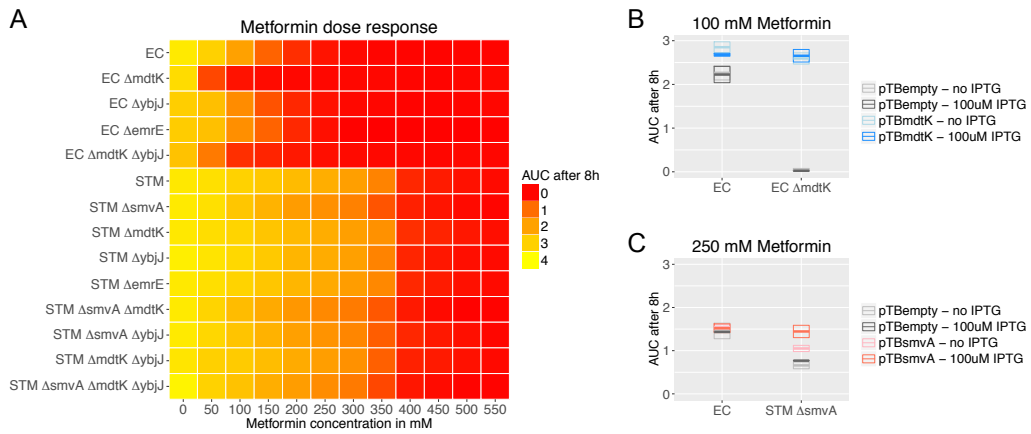

Figure 37: SmvA and MdtK are the major players in metformin export in STM and EC, respectively.
A) Heatmap displaying growth of the indicated strains as area under the curve (AUC) after 8h of growth at 37°C, shaking with increasing concentrations of metformin. Shown is the average of six replicates, for EC emrE the average of two replicates. B) Complementation of EC and EC mdtK with pTBmdtK (+/- IPTG) at 100 mM metformin compared to empty vector. C) Complementation of STM and STM ΔsmvA with pTBsmvA (+/- IPTG) at 250 mM metformin compared to empty vector. B and C) Shown is the average (thick line) growth as area under the curve after 8h +/- standard error (box outlines) of two replicates.

As shown in Figure 37A deletion of *mdtK* in EC indeed results in increased metformin sensitivity compared to wild-type EC. This effect can be rescued by providing *mdtK* on a plasmid (TransBac, unpublished resource, H. Mori) as

demonstrated for growth at 100 mM metformin in Figure 37B. Notably, leaky expression of *mdtK* is sufficient to rescue growth (compare Figure 37B, pTBmdtK – no IPTG).

In contrast, neither deletion of *ybjJ* nor *emrE* result in any changes in sensitivity to metformin in *E. coli.* Similarly, the EC Δ*mdtK* Δ*ybjJ* double mutant behaves like the single deletion mutant of *mdtK.* (Figure 37A). Altogether these results suggest, that MdtK is the major player in *E. coli* metformin efflux, whereas YbjJ and EmrE are likely not expressed under physiological conditions.

As previously discussed, deletion of *smvA* in STM increases susceptibility to metformin. This effect can be reverted by complementation with *smvA* expressed from a plasmid. For this purpose we cloned the *Salmonella smvA* gene into the TransBac plasmid background. Figure 37C demonstrates the IPTG-dependent rescue of STM Δ*smvA* growth at 250 mM metformin when complemented with pTBsmvA.

Deletion of no other pump (*mdtK, ybjJ, Salmonella emrE* paralog) results in any noticeable change in STM sensitivity to metformin. This can be explained by the presence of SmvA in those strains.

To determine whether these pumps might be induced upon deletion of *smvA,* thereby explaining the different behavior of EC and STM Δ*smvA* we constructed several double and triple deletion mutants. If any of the pumps take over metformin efflux upon deletion of *smvA,* additional deletion of the second pump should abolish resistance to metformin. However, neither the STM Δ*smvA* Δ*mdtK* nor the STM Δ*smvA ΔybjJ* double mutants show higher sensitivities than the *smvA* single mutant. The same is true for an STM Δ*smvA* Δ*mdtK* Δ*ybjJ* triple mutant (Figure 37A). Together these results suggest, that induced expression of *mdtK* and *ybjJ* upon *smvA* deletion cannot be the explanation for the high residual metformin resistance.

Notably, the STM Δ*mdtK* Δ*ybjJ* double mutant has a mild fitness defect in the no drug condition (see first column in Figure 37A), which seems to be absent upon additional knock out of *smvA* (see triple mutant). One possible explanation is induction of another pump compensating for the lack of MdtK and/or YbjJ as a result of feedback regulation upon *smvA* deletion. Thus, we are currently testing a double mutant of *smvA* and the *Salmonella emrE* paralog.

In conclusion, MdtK appears to be the major player in metformin efflux in EC, whereas STM uses the species-specific pump SmvA, despite the presence of MdtK and YbjJ.

4.2.5 Investigating the global cellular response to metformin

To gain a more global understanding of each organism's response to metformin treatment, we are currently investigating transcriptomic changes upon metformin treatment. EC, STM and STM Δ*smvA* were grown to exponential phase and treated with a pulse of 250 mM metformin or a no drug control for 20 min before harvesting for RNA extraction.

We will determine differentially expressed genes in a) STM *vs* STM Δ*smvA* in absence of metformin to understand changes induced by deletion of *smvA,* e.g. the expression of another pump compensating for loss of *smvA,* b) EC drug *vs* no drug as opposed to STM drug *vs* no drug to understand differences in the response to metformin treatment between the two organisms and c) the differences in metformin response between STM and STM Δ*smvA*.

4.2.6 Conferring metformin resistance to *E. coli*

As SmvA appears to be the major player in metformin efflux in *Salmonella,* transfer of the *smvA* gene to *E. coli* should result in increased resistance to metformin. To test this hypothesis we recombinantly expressed the *Salmonella smvA* gene in *E. coli* using the IPTG-inducible low copy vector pTBsmvA.

As presented in Figure 38 expression of *smvA* indeed results in increased resistance of wild-type *E. coli* to metformin. At 200 mM metformin the growth of wild-type EC or EC carrying the empty vector is fully inhibited, whereas upon expression of *smvA* growth is only partially affected.

Notably, overexpression of the pumps MdtK, YbjJ and EmrE, also confers increased resistance to metformin. This is in agreement with the observations of the EC overexpression screen (Figure 36B) and with the notion that complementation of EC mdtK with pTBmdtK resulted in better than wild-type growth (Figure 37B).

In conclusion, the above results suggest that despite *E. coli's* sensitivity to metformin the organisms harbors the potential to acquire resistance by altering expression or activity levels of intrinsic pumps.

Notably, none of the pump overexpressions renders EC as resistant as STM, suggesting the organisms do not only differ in drug efflux. Therefore, after taking into account results from the ongoing RNAseq analysis, we will also investigate drug uptake and molecular target sensitivity to determine the cause of the distinct drug responses.

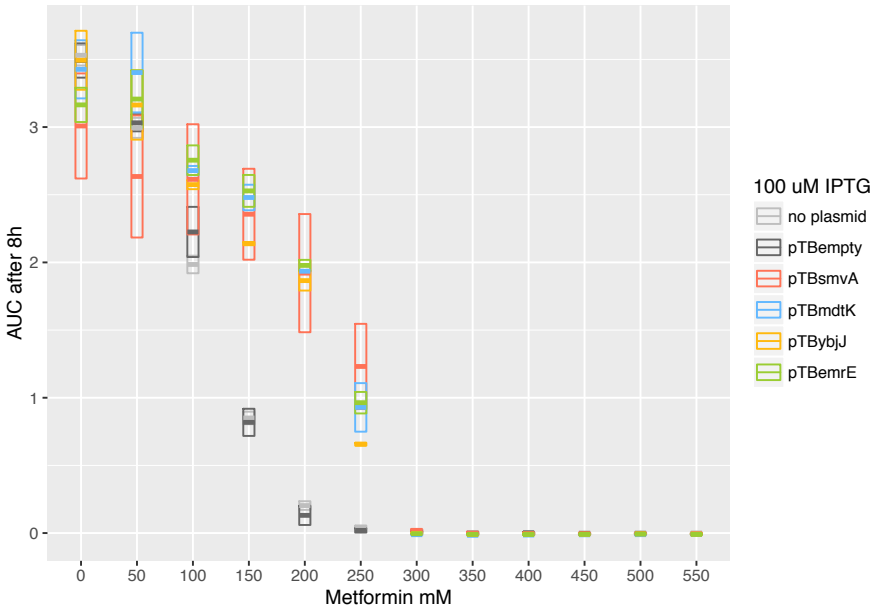

Figure 38: Expression of the *Salmonella* Typhimurium gene *smvA* or overexpression of an intrinsic pump confers metformin resistance to *E. coli.*
Wild-type *E. coli* as well as *E. coli* carrying the empty TransBac vector show complete inhibition at 200 mM metformin. When *E. coli* is complemented with plasmids carrying either *smvA, mdtK, ybjJ* or *emrE*, susceptibility to metformin is decreased and complete inhibition is only reached at 300 mM metformin. Shown is the average (thick line) growth as area under the curve after 8h +/- standard error (box outlines) of two replicates.

4.3 Conclusions and discussion

In summary, I confirmed a high resistance level of *Salmonella* Typhimurium to the T2D drug metformin and identified SmvA as the efflux transporter responsible for metformin export in *Salmonella*.

E. coli does not carry the *smvA* gene and is much more sensitive to the drug. Although deletion of *smvA* renders *Salmonella* more susceptible to metformin, it is still able to outperform *E. coli.* Thus, I investigated the potential role of other efflux pumps, namely MdtK, YbjJ and EmrE, which were previously shown to increase resistance to metformin upon overexpression.

Whereas YbjJ and EmrE do not seem to be involved in metformin efflux in any of the two organisms, MdtK is the pump responsible for metformin efflux in *E. coli.* However, deletion of *mdtK* in a *Salmonella smvA* deletion background has no additional effect on metformin sensitivity. This indicates that differential regulation of MdtK does not account for the discrepancy in metformin susceptibility in *E. coli* compared to the *Salmonella smvA* deletion mutant. To gain a better understanding of the underlying cause for this difference, we are

currently investigating the global changes in gene expression upon metformin stress in both strains.

In further experiments I show that complementation of wild-type *E. coli* with the *Salmonella smvA* gene results in increased resistance to metformin. However, the same can be achieved by overexpressing other pumps present in *E. coli* and capable of exporting metformin (including MdtK responsible for metformin efflux in *E. coli*) suggesting that *E. coli* carries the potential for developing resistance to metformin.

This is particularly interesting considering the observed increase in *Escherichia* subspecies in the microbiota of metformin treated diabetic patients [109] and the notion that MdtK, YbjJ and EmrE are rather specific to *Enterobacteriaceae* (Table 3).

Pump	**Highest conservation level in subspecies of the genera**
mdtK	*Escherichia, Enterobacter, Salmonella, Citrobacter, Klebsiella*
ybjJ	*Escherichia, Enterobacter, Salmonella, Citrobacter, Klebsiella, Shigella*
emrE	*Escherichia, Citrobacter, Shigella,* paralog present in *Salmonella*

Table 3: MdtK, YbjJ and EmrE are conserved in *Enterobacteriaceae.*
For each pump the genera with the highest conservation level are shown (STRING database, Feb 2016).

Taken together with the high metformin resistance of *A. muciniphila,* also showing elevated abundance in metformin treated patients, this poses the question whether constant exposure of gut bacteria to metformin in diabetic patients constitutes a selective pressure favoring metformin resistant bacteria. During the course of experiments we interestingly observed *E. coli* mutants involving an *mdtK* deletion to develop secondary mutations rendering them metformin resistant quite quickly. This suggests that the cellular network indeed tries to adjust to the selective pressure. Characterization of these suppressor mutants would give additional insights in the resistance potential.

Furthermore, enteric side effects are common upon metformin treatment, but can be diminished by gradually increasing the dose instead of starting on the desired dose directly [121]. This could be because more time is allowed for the microbiota to adjust to the constant exposure to metformin.

Mutations leading to elevated expression or activity of any of the investigated pumps, such as loss of function mutation in *ybjK* (repressor of *ybjJ*) could render *Escherichia* resistant to the drug. Thus, we are currently collaborating with

Kristoffer Forslund (Bork lab, EMBL) to examine the published metagenomics study for mutations in any of the pumps carried by *Escherichia.*

Furthermore, metformin treatment may confer an advantage to *smvA* carrying species. Interestingly, SmvA is a rare, rather pathogen-specific pump, reported to be present only in *Salmonella enterica, Klebsiella pneumoniae* and *Acinetobacter baumannii*. It has to be considered that metformin treated diabetic patients may be more susceptible to these pathogens. Whereas *Salmonella* Typhimurium infections are comparably rare in the western world (where T2D is most prevalent), *A. baumannii* and *K. pneuomoniae* have emerged as important multi-drug resistant nosocomial pathogens. Interestingly, diabetes has been reported to be a common risk factor [122-124], but metformin treatment status is often not mentioned in the relevant literature. As records of *A. baumannii* and *K. pneuomoniae* infections should be available from hospitals, we started a collaboration with Lars Juhl Jensen (University of Copenhagen) to compare whether diabetic patients on metformin treatment present this complication more often than diabetic patients on other treatment.

Overall the increased resistance of the pathogen *Salmonella* Typhimurium to the orally administered T2D drug metformin brings implications for long-term treatments. The effects of non-antibiotic drugs on the human microbiome are more and more considered, and we may even be able to exploit beneficial shifts in microbial composition. However, the creation of new niches, which may be occupied by pathogens should be a concern, especially in long-term treatments.

4.4 Contribution disclaimer

I designed concept, acquisition and analysis of the presented experiments. The deletion strains were created by Matylda Zietek (EMBL Heidelberg) and myself. Acquisition of growth curves was carried out together with Matylda Zietek. Mihaela Pruteanu (EMBL Heidelberg) provided the gut bacterial species and I obtained anaerobic growth curves for them. I analyzed all data aided by a script for growth curve analysis provided by Ana Rita Brochado (EMBL Heidelberg). She furthermore provided Figure 31. Lucía Herrera and Lisa Maier (both EMBL Heidelberg) provided data for Figure 36A and Figure 36B, respectively.

5 Discussion and perspectives

Salmonella Typhimurium (STM) is a Gram-negative pathogen causing acute gastroenteritis and in immune-compromised, very young or elderly individuals infection can result in life-threatening bacteremia [68]. Especially in the African continent, invasive non-typhoidal *Salmonella* strains are emerging as a major source of multi-drug resistant bacteria [63]. Although STM is an extensively studied model pathogen, the function on many genes remains unknown.

In the presented study I used a high-throughput reverse genetics approach to gain insights into the roles of non-essential genes and simultaneously monitor drug responses. I profiled the growth of more than 3800 unique STM mutants in more than 550 single conditions ranging from simple physical perturbations, such as changes in pH and temperature, over treatment with sub-inhibitory concentrations of antibiotics, targeting specific cellular pathways, to highly complex stresses like the exposure to conditioned media stemming from other microbes. This yielded a chemical genomics map with more than 4 Mio gene-drug interactions.

In the course of the experiments I developed a quality control pipeline applicable and used for other data sets of similar kind. Furthermore, I determined important characteristics of the utilized deletion collection. In particular we find evidence for a higher proportion of polar effects in the Cm library, which was anticipated due to the antisense transcripts created from this cassette. Additionally, we find a set of mutants with extremely high underlying correlation. We are currently investigating changes in genetic background, which might have occurred during library construction, as a possible explanation for this phenomenon.

5.1 Potential for directed follow up studies

5.1.1 Investigating gene function and drug mode of action

I detected significant phenotypes for 75% of the tested mutants and demonstrated that correlation of phenotypic signatures indicates functional

relation. Thus, we can link genes of unknown function to known protein complexes or cellular pathways.

Indeed, after merging the libraries to ensure robust s-scores we detect a phenotype for more than 500 confirmed orphan genes. For many of those we find high correlation to a known gene (present in benchmarking set) and/or can draw connections to modules of functionally related genes. Thus, the presented data set is rich in information content to drive future follow up studies.

Chemical genomics data are not only useful for investigating gene function but also give insights into drug mode of action. Mutants showing strong phenotypes within a specific condition can be involved in uptake or efflux of the compound or be (connected to) the primary or secondary targets. In the follow up study on metformin efflux I identify primary target specificity as well as the drug efflux pump for the type 2 diabetes drug metformin based on its gene-drug interaction profile.

Strong interactions with uptake or efflux mechanisms are particularly interesting as they might indicate that the compound does not reach high intracellular concentration in the wild-type. Hence, combining those compounds with adjuvants altering membrane integrity could render them effective.

Furthermore, similarities of phenotypic signatures between two drugs can indicate common modes of uptake and efflux as well as shared primary or secondary targets. Based on this data, it would be possible to propose drug combinations likely acting synergistically. It would be particularly interesting to investigate this for combinations of outdated antibiotics or the combination of an antibiotic and drugs not designed for targeting bacteria. Especially in the light of emerging multi-drug resistant *Salmonella* strains the directed search for efficient treatment strategies is indicated. Furthermore, this knowledge might be transferable to other bacteria.

Of course, the follow up potential of the presented data set is by far not exhausted in this thesis. However, previous studies of similar kind have demonstrated the long-term usage of chemical genomic maps to drive hypothesis-based projects ([51, 125-128] to name just a few mechanistic studies utilizing the *E. coli* chemical genomics data). Therefore, the gene-drug interactions detected in this study will continue to allow researchers to generate hypotheses or confirm results obtained for genes and drugs they are investigating.

5.1.2 Exploring evolution

I also presented the to my knowledge first comparison of bacterial phenotypic landscapes. After thorough curation we compared the STM data set created in this study to a similar *E. coli* chemical genomic map published by Nichols et al. [36].

Interestingly, mutants of conserved genes are more likely to be responsive to multiple conditions and their phenotypic signature is frequently conserved between STM and the related *E. coli* (EC). Less conserved genes often show only few or no significant phenotypes congruent with the idea that evolutionary distinct genes may be involved in rather niche-specific tasks and thereby relevant for growth in only few conditions.

From pan-genome studies conducted in STM and EC we can retrieve information about whether a gene is shared between species or only present within one species, subspecies or even strain-specific [58, 129]. It will be interesting to test whether there is a correlation between the responsiveness of a mutant and the time since acquisition. Presumably more recently acquired genes will be wired less deeply into the cellular network and thus exhibit less phenotypes. Furthermore, this analysis can be expanded to elements of horizontal gene transfer such as prophages.

Analysis of functional modules predicted based on the phenotypic signatures in each data set shows a large conservation of core cellular processes. As a next step it would be interesting to probe interaction between those modules and their conservation. In eukaryotes it has been proposed that within module conservation is high whereas between module conservation is poor suggesting rewiring of cross-talk between pathways in order to adjust to new environmental conditions [55].

Adding genes with species-specific responses or those present in only one organism gives insights how they wire into the existing cellular network. In fact, many clear species-specific differences observed throughout the whole study stem from components known to have high subspecies or strain specificity (LPS, O-antigen, virulence factors). This is not surprising, as the strain classification is based on phenotypic differences and differential antigen expression. However, this brings the implication that obtaining chemical genomics profiles for natural isolates and overlaying this data with sequence similarity information might provide new insights into evolution.

5.2 Expanding the cellular network

For every bacterium a large proportion of genes is not needed for growth in laboratory conditions, but only important upon certain environmental cues. While I detected phenotypes for 75% of the mutants, a similar study conducted in *E. coli* reported phenotypes for 50% of the tested mutants [36]. The chemical space probed in my study was bigger and more diverse. Furthermore, I included complex stresses such as media conditioned by other microbes and conditions mimicking the infected host environment. This demonstrates that expansion of the chemical space can increase the number of insights we gain. Growing the mutants in presence of other bacteria or within the host environment would be ideal, however, increased technical challenges, costs and diminished throughput have to be considered.

5.2.1 Addressing genes with no phenotype

A fraction of the mutants presenting no phenotype will be comprised of mutants of redundant genes. Upon deletion of one gene the other one will take over the tasks so that no phenotype will be detected. Simultaneous application of multiple stresses or creating multiple deletions can address this issue.

As an alternative to increasing the degree of perturbation, read-outs other than growth should be considered. While growth or survival may not be affected by certain mutations (under any condition), deletion of the gene in question may well affect the bacterium's ability to attach to a surface, swim, form biofilm or infect the host to name just a few. All of these processes are crucial for bacterial survival throughout its life cycle but will not be captured when measuring growth.

The expansion of read-outs is currently addressed in the lab. In an ongoing study the proportion of *E. coli* mutants, for whom a phenotype was detected, increased from 50% to 82% by combining information from a growth and biofilm formation read-out (personal communication, Lucía Herrera). The biofilm assay used in this study can be easily applied for other bacteria.

Furthermore, we are testing the entire *Salmonella* Kan deletion collection for its ability to infect and replicate in epithelial cells as well as macrophages (personal communication, Bachir El Debs). Processes such as epithelial cell invasion, intracellular survival and replication are extremely complex and rely on a constant communication between bacterial and host cell factors in the correct spatial and temporal resolution. As it is impossible to recreate this complex interplay *in vitro,*

I expect that we will be able to retrieve phenotypes for so far silent mutants from this infection screen.

5.2.2 Integration of other large-scale data sets

Another strategy to increase the information content is the integration of multiple data types as orthogonal information can complete the global picture of the cellular network. Pioneering work in yeast sought to interpret genetic interaction networks by taking into account physical interaction data and revealed insights into pathway redundancy as well as gene essentiality [82, 130-133].

Physical interaction data for STM proteins is scarce and inferring the interactions from other organisms might lead to faulty conclusions. However, there are a number of *Salmonella* specific data sets that can be used.

For example, co-expression of genes is often associated with functional relation. STM gene expression has been extensively studied in the context of infection and small RNA discovery [76, 134]. Identifying genes co-expressed in multiple studies and/or different conditions would provide robust co-expression data eliminating biases caused by differing sample treatments, sequencing platforms or analysis pipelines. This information can be subsequently used to complement the gene-drug interaction network. The big advantage of using e.g. expression data is that we can include information on essential genes which are by design not represented in studies utilizing deletion mutants.

Furthermore, it will be highly interesting to implement data from the *Salmonella* infection screen ongoing in our lab. The expression of virulence genes is often activated only upon detection of certain host signals. Thus, their influence on core cellular processes is difficult to determine *in vitro*. Combining gene-drug interaction and infection related data might be able to bridge this gap. After all, it is intuitive that bacteria have to wire virulence factors into existing cellular pathways to efficiently counter host defenses and to successfully adapt to replicate in a hostile host environment such as the acidic *Salmonella* containing vacuole.

5.3 Considerations for long-term drug treatments

The follow up study on metformin efflux presented in this thesis is just one example of how we can combine knowledge from different large-scale interaction data sets to accelerate and focus hypothesis-driven research. Additionally to

discovering distinct mechanisms of efflux in *E. coli* and *Salmonella,* I report two more silent efflux pumps capable of exporting this drug. This implies a higher potential for resistance in bacteria carrying these genes.

The great impact of the human microbiota on the health of the host is becoming more and more evident. Thus, consequences of disrupting the microbiota must be considered. Although awareness of microbiota-related side effects of antibiotic treatment is rising, we are just at the beginning of understanding the manifold disruptions our microbial system may be exposed to in daily life.

Gastroenteric side effects are amongst the most common consequences of any drug treatment, also for metformin. Even if not designed to inhibit functions of the bacterial cell, many drugs will nevertheless influence them. This may not result in inhibition of bacterial growth, but it can shift the balance in the microbiota, if some bacteria are more affected than others. These shifts in microbial compositions can subsequently create new ecological niches. Prebiotics even exploit this phenomenon by favoring bacteria beneficial for the host. Other treatments might create niches for harmful bacteria or even pathogens. However, we must consider that any compound, drug or potential prebiotic, may have mixed effects on the microbiome, which might furthermore depend on the exact composition of the resident microbiota before treatment. Metformin, the drug investigated in this study, illustrates the potential for opposing effects. The increase in *Akkermansia* subspecies and alterations in bacterial metabolism seem to have a positive influence on host metabolism and health. However, prolonged treatment might also increase the risk of severe infections with multi-drug resistant bacteria. Also, the effect on funghi and viruses present in the flora has been largely unexplored.

5.4 Perspectives

The investigations on metformin efflux revealed that *E. coli* and *Salmonella* use distinct pumps to export the drug. *E. coli* uses MdtK whereas *Salmonella* appears to be rewired to use the species-specific pump SmvA despite the presence of *mdtK* in the *Salmonella* genome. Furthermore, both organisms encode other pumps capable of exporting metformin. However, even upon deletion of the major players these pumps remain unused.

This example nicely demonstrates a few remaining questions and challenges in understanding bacterial biology. First, it shows that presence of a gene does not necessarily imply usage. Likewise, absence of a gene should not be blindly correlated to absence of function. This is especially important in the light of bacteria living in communities. For example, in the human gut bacteria utilize each other's metabolites, horizontal gene transfer allows for acquiring certain gene functions, degradation of extracellular compounds may be carried out by enzymes secreted from one bacterium, whereas the breakdown products may be used by others as well.

Therefore, we have to ask ourselves how complete the picture we gain of the cellular network by studying single bacteria can possibly be. We use gain and loss of genes as a measure for speciation events and expect accordingly differential behavior of the diverging strains. However, we must consider that genes may be lost because their function is not needed in the presence of other bacteria. Whereas this loss may alter the outcome of experimental investigations in the laboratory the role in the natural environment may be much less impacted (or much more for the matter).

In conclusion, information from large-scale genetic, chemical and physical interaction data has accelerated our understanding of cellular processes and dynamics and will continue to spur hypothesis-driven research. With recent advances in technology we have the means to manipulate any variable at any scale. Genetic perturbations can range from point mutations to deleting entire operons, the condition space can include single, highly defined perturbations or very complex stresses. We must consider interaction partners depending on whether we measure single cells, cultures, communities or even interactions with the host. Effective integration and presentation of the accumulating data will certainly be one of the next big challenges in the field.

6 Materials and Methods

6.1 Creating a chemical genomics data set for *Salmonella* Typhimurium

6.1.1 Bacterial strains

The single gene deletion collections used in this study are based on *Salmonella enterica* serovar Typhimurium 14028s [57]. They were provided by Helene Andrews-Polymenis (Texas A&M) and Michael McClelland prior to publication.

In brief, the deletion mutants were created by replacing the gene of interest by either a kanamycin (Kan) or a chloramphenicol (Cm) resistance cassette using the lamba-Red recombinase system [75]. To reduce polar effects the first and last 30 bases of each targeted sequence were preserved. The kanamycin cassette was placed co-directional to the deleted gene, whereas the chloramphenicol cassette was placed in opposite direction.

We single colony purified and re-arrayed the library, so in this study I present 7200 mutants of 3834 unique genes. For more than 3000 genes a clone is present in each library, for a number of genes duplicates of the same clone were included to fill up plates. Furthermore we included *E. coli* BW25113 (referred to as wild-type *E. coli*), *Salmonella* Typhimurium LT2 and *Salmonella* Typhimurium 14028s (ATCC14028, referred to as wild-type *Salmonella*) in the re-array.

6.1.2 Chemical perturbations

Tested conditions

For a complete list of perturbations tested in this study see Appendix B Tested conditions (page 113).

MIC testing

Minimal inhibitory concentrations were tested using MIC test strips (Oxoid or Liofilchem) if available. Wild-type *Salmonella* (ATCC14028) was spread as a lawn on a regular LB Lennox agar plate. Next, the MIC test strip was carefully placed on top ensuring complete contact with the agar surface. After over night

incubation at 37°C the MIC was read at the intersection of the resulting halo with the MIC test strip. If no halo was observed the MIC was determined as greater than the highest concentration on the strip.

For manual testing chemical solutions were added in a concentration series to liquid LB Lennox agar before pouring the plates reflecting the method used for the chemical genomics study itself. After the agar was solidified *Salmonella* was streaked out on the agar surface and inspected for growth after over night incubation at 37°C.

Test plate preparation

For all steps involving agar plates Singer Plus Plates were used. The plates were prepared using a pump (PM05, AES Laboratoire) dispensing a fixed volume of 43 ml media into every plate at 500 rpm. All plates contained 2% (w/v) agar and unless otherwise indicated the base media was LB Lennox. Chemicals were added to the liquid agar from higher concentrated stock solutions to reach the desired final concentration. For the detailed overview of base media used for each condition see Appendix B Tested conditions (page 113) and for the base media recipes Appendix A Base media (page 113).

Conditioned media (CM) conditions were created by mixing 2x LB Lennox agar with equal amounts of conditioned media (filter sterilized) and water to reach 1x LB Lennox agar. Therefore, the base media is still full LB Lennox and the conditioned media is diluted by the factor 4. Growth conditions were: For *Pseudomonas aeruginosa* (PA01 and PA14): LB Miller, aerobically, 37°C, shaking, over night. For *Akkermansia muciniphila*: MGAM (GAM Broth_Modified**,** HyServe), anaerobically, 37°C, standing, over night. A media control treated exactly like the culture was included for both cases.

SPI2-inducing and non-inducing conditions were based on conditions tested by Kröger et al. when creating a transcriptomic atlas of *Salmonella* in infection-relevant conditions [76]. For the low magnesium modification, final concentration of $MgSO_4$ was lowered to 0.01 mM. For the low iron modification I added the iron scavenger bipyridyl at a final concentration of 0.1 mM.

UV exposure

For UV treatment the mutant collections were arrayed on LB Lennox agar plates and exposed to UV (254 nm) using a UV Stratalinker 2400 (Stratagene) for 6, 12, 18 or 24 seconds prior to incubation at 37°C.

6.1.3 Robotic procedures

All robotic transfers were carried out using the Singer Rotor (Singer Instruments).

Transfer from glycerol stock to agar plate

The mutant libraries were stored as glycerol stocks in 384 well plates at -80°C (15% (v/v) glycerol). First, all plates (10 Kan, 10 Cm) were thawed completely and centrifuged at 800 rpm for 2 min. The Singer Rotor was used to transfer the bacteria from the well plates to LB Lennox, 2% (w/v) agar plates containing either 30 µg/ml kanamycin or 10 µg/ml chloramphenicol. The only exception was 384 Kan plate 7 as it contained *E. coli* and *Salmonella* wild-type strains as well as *S.* Typhimurium LT2, therefore it was arrayed onto LB agar without antibiotic.

In brief, the procedure selected from the standard methods of the Singer rotor was the following:

- Source: Multi-Well 384, Target: PlusPlate 384, Pad: Long Pin 384
- Program: Spot Many
- Settings: 2 plates per source, wet mix for the source plate and default pinning pressure for the target plate.

The resulting agar plates were incubated over night at 37°C.

Arrangement of library plates in 1536 format

Next day, the 20 agar plates in 384 format were used to assemble the five library plates in 1536 format using the following Singer Rotor procedure:

- Source: PlusPlate 384, Target: PlusPlate 1536, Pad: Short Pin 384
- Program: 1:4 Array
- Settings: No offset and default pinning pressure for the source plate as well as the target plate.
- Assembly: see Table 4

The resulting plates were incubated at 37°C for 4-6 hours and re-arrayed using the procedure described below (see Transfer to test plate).

1536 plate number	384 plates used	Antibiotic used
1	Kan 1-4	Kan
2	Kan 5-8	None (includes wt strains)
3	Cm 1-4	Cm
4	Cm 5-8	Cm
5	Kan 9/10 and Cm 9/10	None

Table 4: Arrangement of library plates in 1536 format from 384 format source plates.

Maintenance of library plates

Library plates were maintained for up to one month before going back to glycerol stocks and arranging them freshly. During this time the plates were kept at 4°C unless they were used the same day, but never for longer than one week. Instead they were regularly re-arrayed using the same procedure as described below for transferring mutants to the test plates containing chemical perturbation.

Transfer to test plates

To transfer the bacteria from source plates to test plates containing the chemicals a Singer Rotor procedure with the following parameters was used:

- Source: PlusPlate 1536, Target: PlusPlate 1536, Pad: Short Pin 1536
- Program: Replicate many
- Settings:
    - Recycle: Full, revisiting source (up to 20 plates were pinned from one source plate using the same pad)
    - Source: no offset, pinning pressure 50%
    - Target: pinning pressure 70%

From each source I first pinned 2-4 times on LB Lennox agar (with Kan and Cm as applicable). Those plates were later used as new source plates. Only then I transferred mutants from the same source to test plates. All test plates were incubated at 37°C for approximately 12 hours unless otherwise indicated (e.g. room temperature or high temperature conditions). If growth was poor the incubation time was extended.

6.1.4 Colony size measurement

Picture acquisition

Pictures were taken using a robotic setup (SPImager, S&P Robotics Inc.) fixing the positions of camera (Canon EOS Rebel T3i) and plate on a black background. The camera settings used are: shutter speed 1/100, aperture F2.8, ISO 400. Plates with severe pinning errors (e.g. incomplete pinning due to uneven agar surface) were discarded right away.

IRIS

All pictures were subjected to analysis using the *Salmonella* growth profile of Iris version 0.9.4.57. In brief, the software crops the picture, converts it to gray scale, detects the format and divides the picture in the corresponding number of tiles. In each tile the largest circular object is detected as the colony using an intensity threshold. The colony size corresponds to the number of pixels in the area covered by the colony.

6.1.5 Quality control

The quality control applied is described quite extensively in the main text. In brief, all parameters (see Table 1 on page 24) were calculated for the entire initial data set and thresholds for extreme values decided based on the overall distributions. File names of plates exceeding the thresholds were automatically detected using a Matlab script and combined in one text file. This text file can be provided when using the Matlab EMAP toolbox to normalize and score the data. If provided data from the listed files are automatically blanked for the analysis.

6.1.6 Normalization, interaction scoring and clustering

All normalization and scoring steps were carried out using the EMAP toolbox for Matlab [30].

The normalization steps include a second order surface correction, an outer frame correction and a plate-to-plate normalization. The outer frame correction brings mutants of the two outer-most rows and columns to the overall growth of the plate center. The plate-to-plate correction brings the plate middle mean (mean of middle 10% of colony sizes around the median) to the same value (calculated based on the entire data set) by multiplying every colony size on a 1536 plate with

the same factor. This procedure was adapted to take into account the different normalization groups described in chapter 2.3.2 High mutant correlation. For every subgroup the factor is calculated to bring the median colony size of this subgroup to the desired plate middle mean. The subsequent interaction scoring was not modified and a description can be found in the EMAP toolbox manual.

Prior to using cluster analysis to visualize the data it is necessary to bring conditions to a comparable level. Therefore, we rescaled the s-score distribution of each condition (across all mutants) to the spread of a standard normal distribution (IQR=1.35) as described by Nichols et al. [36].

Two-dimensional hierarchical clustering analysis (similarity metric: correlation (centered), clustering method: complete linkage) was performed using Cluster 3.0 (http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm) and the resulting dendrogram was visualized using Java TreeView (http://jtreeview.sourceforge.net/).

6.2 Further analysis of the *Salmonella* chemical genomics data set

All the computational analysis carried out on the chemical genomics data after s-score rescaling were based on the Python programming language. The following python software libraries have been used for data analysis and visualization:

- Python v2.7.11
- NumPy v1.10.2 [135]
- SciPy v0.16.1 [136]
- Pandas v0.17.1 [137]
- Fastcluster v1.1.13 [138]
- Scikit-learn v0.17 [139]
- Networkx v1.9.1 [140]
- BioPython v1.65 [141]
- Matplotlib v1.5.1 [142]
- Seaborn v0.7.0 [143]
- Jupyter v1.0.0 [144]

6.2.1 Mutant correlations for detecting normalization groups

Mutant-mutant correlation matrices were constructed computing all versus all Pearson's correlations of s-scores across conditions. The correlations were then subjected to hierarchical clustering (distance metric euclidean and linkage method average). Sub-clusters for further analysis were selected manually and within cluster correlation recalculated.

6.2.2 Defining significant phenotypes

The s-scores of all *Salmonella* mutants were subjected to an FDR correction, in order to highlight true phenotypes from experimental noise, using the same approach described by Nichols et al. [145]. Phenotype rarefaction curves were derived from counting the number of strains exhibiting at least one significant phenotype upon the respective threshold and a final FDR threshold of 5% was chosen.

6.2.3 Exclusion of noisy strains and library averaging

Noisy clone pairs were determined by their residual s-scores in each condition (i.e. s-score clone A minus s-score clone B). Mutants were labeled noisy if a) the sum of residuals across all conditions was <-100 or >100 or b) they showed particularly high residuals (greater than mean of all residuals +/- three times the standard deviation of all residuals) in 10 or more conditions. Noisy pairs were manually revised based on total number of phenotypes, conditions in which they occur and top correlated genes. If no informed decision on which clone is more likely correct could be made the complete pair was removed from further analysis.

To yield the Kan-Cm combined data set, we averaged the s-scores derived from Kan or Cm mutants of the same gene. If multiple Kan or Cm clones were present one was randomly selected to calculate the Kan-Cm average. If a mutant was present in only one library a pseudo-averaging strategy as described by Collins [30] was applied. In brief, the single s-score was averaged together with the median s-score of all mutants in the same condition and from the other library with similar s-score; an s-score was considered similar if the residual with the mutant was below half the s-score of the mutant. This strategy was not used if the mutant was present multiple times within that library. In that case two clones of the same library were averaged.

6.2.4 Benchmarking

To determine whether a high s-score signature correlation between two mutants is indicative of functional relation we benchmarked all pairwise mutant correlations against sets of genes known to interact, participate in the same cellular pathway or sharing genomic context. The various sets used as benchmarks are listed in Table 5.

Name	Database	Organism	Reference
CycCplx	Biocyc	*S.* Typhimurium 14028s + inferred from *E. coli* MG1655	[146]
CycPath	Biocyc	*S.* Typhimurium 14028s	[147]
CycOper	Biocyc	*S.* Typhimurium 14028s	[147]
DOOR Operons	DOOR	*S.* Typhimurium 14028s	[148]
KGmod	KEGG	*S.* Typhimurium LT2	[149, 150]
KGpath	KEGG	*S.* Typhimurium LT2	[149, 150]
STRExp	STRING	*S.* Typhimurium 14028s	[151-154]

Table 5: Benchmarking data sets used.

For each benchmarking data set we broke all entries down into binary combinations, e.g. a protein complex ABC, was broken down into the combinations A-B, A-C and B-C. Next, homodimeric combinations (A-A) were removed.

The Pearson correlation of s-scores across all conditions for all possible pairwise mutant combinations was calculated and the distribution plotted in a density curve. Next, only correlations of mutant pairs present in the corresponding benchmarking data set were added as a separate density curve. The same was repeated requiring a minimum number of significant phenotypes to be present in each mutant to be considered.

To compare benchmarking performance before and after merging the Kan and Cm library ROC (receiver operating characteristic) curves were drawn. Mutant pairs derived from the respective benchmarking set were used as true positive set. The negative set was derived from all possible mutant combinations between genes present in the benchmarking sets but not interacting, as we cannot infer the relationship status between genes not present in the annotations sets. The s-score signature correlation threshold was then changed to measure the changes

in true and false positives rates. An additional test was carried out comparing the average correlation inside each benchmarking set as compared to the same number of random mutant pairs.

6.3 Interspecies comparison

The *E. coli* data set used in this comparative analysis was published in 2011 by Nichols et al. [145]. Significant phenotypes were defined using a false discovery rate as described in the publication and as applied also for this study (see 6.2.2 Defining significant phenotypes).

6.3.1 Selection of shared conditions

The *E. coli* and *Salmonella* chemical genomics data sets overlap in 91 unique stresses. The original s-score matrices were restricted to those conditions to enable direct comparisons between the two species. Conditions related to cellular stresses (pH, temperature, exposure to UV) and minimal media conditions were paired directly. For chemical stresses the different concentrations used in the two species might not be directly comparable, therefore requiring a strategy to highlight the most comparable concentrations for each chemical. We reasoned that in the optimal concentrations the orthologs in the two species should show some degree of correlation in their s-scores, exhibit a comparable number of phenotypes, and that some orthologs might show similar significant phenotypes in both species. For each single chemical condition shared by the two species, concentration pairs were sorted by their residual ratio in the number of phenotypes shown by orthologs in both species. To avoid local minimum, conditions pairs with residual ratios up to three times the lowest ones were considered. The selected concentrations pairs were then sorted by the ratio of phenotypes shared between the two species, again allowing pairs with up to three times the minimum shared phenotypes ratio. If more than one concentration pair satisfied those requirements, we chose the pair with higher concentrations, as it would show more significant phenotypes and thereby facilitate the detection of correlations between mutants and conditions. If the resulting paired concentrations still showed a residual in the number of phenotypes higher than 50, the chemical was excluded entirely. This was the case for azidothymidine and triclosan.

The resulted shared conditions matrices then comprised 89 single conditions and are listed in Appendix C Shared conditions for interspecies comparison.

6.3.2 Ortholog correlation and intra-species conservation

The following genome sequences were used (NCBI RefSeq entries):

- *E. coli:* NC_000913.3 (*Escherichia coli* str. K-12 substr. MG1655)
- *S.* Typhimurium: NC_016856.1 / NC_016855.1 (*Salmonella enterica* subsp. *enterica* serovar Typhimurium str. 14028S)

Orthologous genes between the two species (as well as paralogs) were predicted using OrthoMCL [78].

The correlation between orthologs in the shared conditions was measured by considering only the 1:1 (one-to-one) orthologous groups found by OrthoMCL, thus excluding any groups including paralogs. The orthologs were then divided by their number of phenotypes: an ortholog pair was either defined “blank” (both mutants do not show any phenotype across all shared conditions), “phenotype” (both mutants show at least one phenotype in the shared conditions) and “important” (at least one mutant shows more a phenotype in at least 10% of the shared conditions).

Gene conservation levels inside the two species were assessed by measuring the BSR (blast score ratio) [155] of each gene inside all the available closed genome sequences available for *E. coli* and *S. enterica*; the genomes (61 for *E. coli* and 41 for *S. enterica)* were downloaded from the NCBI FTP database. The average BSR across all genomes was used as a measure of gene conservation inside the species; genes with an average BSR score lower than 0.9 were considered poorly conserved in the species.

6.3.3 Chromosomal location of responsive genes

Spatial enrichment of responsive genes showing significant phenotypes in at least 10% of the shared conditions was inspected with an enrichment strategy similar to the one used for E. coli [36]. For both, the *Salmonella* and *E. coli* chromosome, regions of 100kb with a 1kb sliding window were used, while regions of 10kb were used for *Salmonella* plasmid. Genomic positions were adjusted so that position zero is occupied by the origin of replication (oriC). For *E. coli* the position was derived from the refseq entry NC_000913.3 (feature "rep_origin"). For *Salmonella's* chromosome the position of oriC was determined

from the DoriC database [156, 157], while for the plasmid the position of the repC gene was used as a starting point.

6.3.4 Detecting clusters of mutants and conditions

Only mutants/conditions with at least one significant phenotype were considered. Mutant correlation matrices were constructed computing all pairwise Pearson correlations of s-score signatures across the shared conditions. The correlations were then subjected to hierarchical clustering (distance metric euclidean and linkage method average). The same was done for calculating s-score signature correlations of conditions across mutants. Clusters of genes and conditions were defined by requiring the average within cluster correlation to exceed a certain threshold. For mutants clustering the selected correlation threshold was 0.3 for both species, while for conditions clustering a correlation threshold of 0.2 and 0.15 was used for *E. coli* and *Salmonella*, respectively. This was necessary due to the different overall conditions correlation profiles in the two species.

6.3.5 Conserved and expanded gene modules

To identify conserved gene modules between the two species we used a variant of a method adopted in a similar study between the budding and fission yeast [55, 82]. Each ortholog pair between *E. coli* and *Salmonella* was arranged in separate modules, such that each module contains one ortholog pair. A merge score was then computed for each pair of modules m1 and m2:

$$\frac{1}{|m1|+|m2|}\sum_{a\epsilon m1}\sum_{b\epsilon m2} r_{ecoli}\,(ab) + \frac{1}{|m1|+|m2|}\sum_{a\epsilon m1}\sum_{b\epsilon m2} r_{salmonella}\,(ab)$$

where r is the Pearson correlation between gene a and gene b across the shared conditions. If the sum of correlations in *E. coli* or *Salmonella* was below zero, the merge score was set to zero for that modules pair.

Once the merge scores had been computed for each possible module pair, we sorted them and started merging the pairs with the highest merge score, until the merge score fell under a defined threshold. Each time a pair is merged, the merge score between the expanded module and all other modules was recomputed and all the scores were sorted again. Additionally, modules were not merged if they contained a gene that was already featured in another module with size greater than one (due to the presence of paralogs). A merge score threshold of 0.4 was

chosen for the creation of conserved modules, using the ROC curve benchmarks described above.

The conserved modules created with this strategy were expanded to include genes specific to either species clustering with the module. For each gene not already present in a module with size greater than one we computed the median Pearson's correlation with all genes in each conserved module. The median value was put equal to zero if any of the correlations with the genes present in the module were found to be below zero. We then sorted all the median correlations and added each gene to the conserved module for which they showed the highest correlation, until a certain threshold was met. We used a ROC curve benchmark considering only the gene pairs between the expanded and conserved modules to pick a correlation threshold of 0.3.

6.4 Investigating *Salmonella* Typhimurium metformin resistance

6.4.1 Bacterial strains, plasmids and antibiotics

Bacterial strains

All *S.* Typhimurium deletion strains were derived from the single gene deletion collection used in this study. All mutants were re-transduced into ATCC14028s wild-type background by P22 transduction and verified by PCR.

The *E. coli* strains were isolated from the KEIO collection [41], re-transduced into BW25113 wild-type background by P1 transduction and the mutation was likewise verified by PCR.

In all cases, when multiple clones were present, both were tested and one was kept for further experiments.

Double mutants were constructed by P1/P22 transduction. If necessary, the antibiotic cassette was removed from one strain using transient expression of flippase from the temperature-sensitive pCP20 plasmid as described by Datsenko and Wanner [75]. A list of all strains used in chapter 4 can be found below.

Label	Background	Mutation
STM	*S.* Typhimurium 14028s	---
STM ΔsmvA	*S.* Typhimurium 14028s	Δ*smvA*::kan
STM mdtK	*S.* Typhimurium 14028s	Δ*mdtK*::cat
STM ybjJ	*S.* Typhimurium 14028s	Δ*ybjJ*::cat
STM emrE	*S.* Typhimurium 14028s	Δ*emrE*::kan
STM ΔsmvA mdtK	*S.* Typhimurium 14028s	Δ*smvA*::kan Δ*mdtK*::cat
STM ΔsmvA ybjJ	*S.* Typhimurium 14028s	Δ*smvA*::kan Δ*ybjJ*::cat
STM mdtK ybjJ	*S.* Typhimurium 14028s	Δ*mdtK*::cat Δ*ybjJ*
STM ΔsmvA mdtK ybjJ	*S.* Typhimurium 14028s	Δ*smvA*::kan Δ*mdtK*::cat Δ*ybjJ*
EC	*E. coli* BW25113	---
EC mdtK	*E. coli* BW25113	Δ*mdtK*::kan
EC ybjJ	*E. coli* BW25113	Δ*ybjJ*::kan
EC emrE	*E. coli* BW25113	Δ*emrE*::kan
EC mdtK ybjJ	*E. coli* BW25113	Δ*mdtK*::kan Δ*ybjJ*

Table 6: Bacterial strains used in chapter 4 The *Salmonella* efflux pump SmvA confers resistance to the diabetes drug metformin.

Plasmids

Except for pTBsmvA all plasmids were derived from the TransBac library, a collection of self-transmissible low copy expression plasmids of *E. coli* genes (unpublished).

pTBsmvA was created by amplifying *smvA* from wild-type *S.* Typhimurium with primers containing regions homologous to the TransBac empty vector (5'-attcattaaagaggagaaacgagctcGAAGGGAGAGTTATGTTTCGTCAGTG-3' and 5'-ggccgcataggccggcccccgcatgcTTATCGGCGTTGGGCTTTTG-3'). The empty TransBac vector was cut using AvrII and CutSmart Buffer (NEB). Gibson assembly was used to join the fragments (Gibson Assembly Master Mix, NEB, 1:2 vector to insert ratio). Subsequent electroporation into electro-competent wild-type *E.coli,* selection on tetracycline and MidiPrep (Qiagen) yielded the pTBsmvA plasmid.

TSS transformation was used to transform all plasmids into all *E. coli* strains used here. As the TransBac library is available in BW38029 Hfr background we used conjugation to introduce the plasmids into the *Salmonella* strains tested in this study. I used repeated streak outs on non-selective medium to derive

BW38029 Hfr from BW38029 Hfr carrying pTBmdtK. After confirming that the strain had lost the plasmid conferring tetracycline resistance I used TSS transformation to introduce pTBsmvA, which was not part of the TransBac library, into the same background.

Label	Source	Resistance
pTBempty	TransBac plasmid library	Tetracycline
pTBsmvA	TransBac plasmid library	Tetracycline
pTBmdtK	TransBac plasmid library	Tetracycline
pTBybjJ	TransBac plasmid library	Tetracycline
pTBemrE	This study	Tetracycline

Table 7: Plasmids used in chapter 4 The *Salmonella* efflux pump SmvA confers resistance to the diabetes drug metformin.

Antibiotics

Antibiotics were used in the following final concentrations: 30 μg/ml kanamycin (Kan), 10 μg/ml chloramphenicol (Cm), 10 μg/ml tetracycline (Tet).

6.4.2 Previous studies analyzed

E. coli chemical genomics

The *E. coli* chemical genomics data set I examined was created by Lucía Herrera (unpublished) using the KEIO collection of *E. coli* single gene deletions [41] and following the same principles and analysis pipeline as outlined for this study. Here I use s-scores obtained for all mutants in four different concentrations of metformin: 40, 60, 80 and 100 mM.

E. coli overexpression screen

The overexpression screen was performed by Lisa Maier (unpublished) using a library of *E. coli* barcoded deletion mutants (ASKA, background BW38029) complemented with their corresponding TransBac plasmid (Hirotada Mori, unpublished resource). The growth of this library was determined at 0 and 100 mM metformin employing different IPTG concentrations to induce expression of the deleted gene from the low copy plasmid. For every strain and every IPTG concentration a t-test was used to compare the relative growth of each strain in drug to no drug condition.

6.4.3 Growth experiments

Unless otherwise indicated, all growth curves for EC and STM strains were obtained in LB Lennox, at 37°C, shaking with an initial OD (578nm) of 0.01. The strains were grown in 384 well plates with a total volume of 50 µl and OD at 578 nm was measured every 30 min using a BioTek Synergy HT plate reader. The resulting growth curves were objected to base line correction and final OD as well as area under the curve (AUC) after 8 h were calculated (script provided by Ana Rita Brochado).

Metformin dose responses and complementation experiments

The dose response to metformin was tested by growing all bacterial strains in increasing concentrations of metformin, namely 0 (no drug control), 50, 100, 150, 200, 250, 300, 350, 400, 450, 500 and 550 mM.

All complementation experiments were performed across the metformin concentration range mentioned above, either without IPTG (the plasmids exhibit leaky expression) or with 100 µM IPTG.

Metformin and phenformin dose response of gut bacterial species

Growth curves for the following bacteria were determined under anaerobic conditions: *Salmonella enterica* Typhimurium 14028s, *Salmonella enterica* Typhimurium LT2, *Escherichia coli* BW25113, *Akkermansia muciniphila* (DSM 22959), *Bacteroides thetaiotaomicron* (DSM 2079), *Bacteroides vulgatus* (DSM 1447), *Bifidobacterium adolescentis* (DSM 20083), *Bifidobacterium longum* (DSM 20088), *Clostridium bolteae* (DSM 15670), *Clostridium perfringens* (DSM 756), *Clostridium ramosum* (DSM 1402), *Fusobacterium nucleatum* (DSM 15643), *Lactobacillus paracasei* (ATCC SD5275), *Prevotella copri* (DSMZ 18205), *Roseburia intestinalis* (DSM 14610), *Ruminococcus gnavus* (ATCC 29149), *Streptococcus salivarius* (DSMZ 20560).

The bacteria were grown anaerobically in 384 well plates, at 37 °C. MGAM medium (GAM Broth_Modified, HyServe) was used and metformin and phenformin added in a two-fold dilution series (starting at 750 mM and 26 mM, respectively). A no drug control was included. The bacteria were grown statically and shaken for 30 sec before measuring OD (578 nm) every 30 min in a BioTek Eon plate reader.

References

1. Deltcheva, E., et al., *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.
2. Huh, W.K., et al., *Global analysis of protein localization in budding yeast.* Nature, 2003. **425**(6959): p. 686-91.
3. Breker, M., et al., *LoQAtE--Localization and Quantitation ATlas of the yeast proteomE. A new tool for multiparametric dissection of single-protein behavior in response to biological perturbations in yeast.* Nucleic Acids Res, 2014. **42**(Database issue): p. D726-30.
4. McKellar, J.L., J.J. Minnell, and M.L. Gerth, *A high-throughput screen for ligand binding reveals the specificities of three amino acid chemoreceptors from Pseudomonas syringae pv. actinidiae.* Mol Microbiol, 2015. **96**(4): p. 694-707.
5. Lee, D., O. Redfern, and C. Orengo, *Predicting protein function from sequence and structure.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 995-1005.
6. Mills, C.L., P.J. Beuning, and M.J. Ondrechen, *Biochemical functional predictions for protein structures of unknown or uncertain function.* Comput Struct Biotechnol J, 2015. **13**: p. 182-91.
7. Bartel, P.L., et al., *A protein linkage map of Escherichia coli bacteriophage T7.* Nat Genet, 1996. **12**(1): p. 72-7.
8. McCraith, S., et al., *Genome-wide analysis of vaccinia virus protein-protein interactions.* Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4879-84.
9. Flajolet, M., et al., *A genomic approach of the hepatitis C virus generates a protein interaction map.* Gene, 2000. **242**(1-2): p. 369-79.
10. Fromont-Racine, M., J.C. Rain, and P. Legrain, *Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens.* Nat Genet, 1997. **16**(3): p. 277-82.
11. Fromont-Racine, M., et al., *Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins.* Yeast, 2000. **17**(2): p. 95-110.
12. Uetz, P., et al., *A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.* Nature, 2000. **403**(6770): p. 623-7.
13. Ito, T., et al., *A comprehensive two-hybrid analysis to explore the yeast protein interactome.* Proc Natl Acad Sci U S A, 2001. **98**(8): p. 4569-74.
14. Gavin, A.C., et al., *Proteome survey reveals modularity of the yeast cell machinery.* Nature, 2006. **440**(7084): p. 631-6.
15. Krogan, N.J., et al., *Global landscape of protein complexes in the yeast Saccharomyces cerevisiae.* Nature, 2006. **440**(7084): p. 637-43.
16. Gavin, A.C., et al., *Functional organization of the yeast proteome by systematic analysis of protein complexes.* Nature, 2002. **415**(6868): p. 141-7.

17. Butland, G., et al., *Interaction network containing conserved and essential protein complexes in Escherichia coli.* Nature, 2005. **433**(7025): p. 531-7.
18. Rajagopala, S.V., et al., *The binary protein-protein interaction landscape of Escherichia coli.* Nat Biotechnol, 2014. **32**(3): p. 285-90.
19. Typas, A. and V. Sourjik, *Bacterial protein networks: properties and functions.* Nat Rev Microbiol, 2015. **13**(9): p. 559-72.
20. Kentner, D. and V. Sourjik, *Dynamic map of protein interactions in the Escherichia coli chemotaxis pathway.* Mol Syst Biol, 2009. **5**: p. 238.
21. Alexeeva, S., et al., *Direct interactions of early and late assembling division proteins in Escherichia coli cells resolved by FRET.* Mol Microbiol, 2010. **77**(2): p. 384-98.
22. Giaever, G., et al., *Functional profiling of the Saccharomyces cerevisiae genome.* Nature, 2002. **418**(6896): p. 387-91.
23. Collins, S.R., et al., *Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map.* Nature, 2007. **446**(7137): p. 806-10.
24. Roguev, A., et al., *Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast.* Science, 2008. **322**(5900): p. 405-10.
25. Brochado, A.R. and A. Typas, *High-throughput approaches to understanding gene function and mapping network architecture in bacteria.* Curr Opin Microbiol, 2013. **16**(2): p. 199-206.
26. Boone, C., H. Bussey, and B.J. Andrews, *Exploring genetic interactions and networks with yeast.* Nat Rev Genet, 2007. **8**(6): p. 437-49.
27. Tong, A.H., et al., *Systematic genetic analysis with ordered arrays of yeast deletion mutants.* Science, 2001. **294**(5550): p. 2364-8.
28. Tong, A.H., et al., *Global mapping of the yeast genetic interaction network.* Science, 2004. **303**(5659): p. 808-13.
29. Schuldiner, M., et al., *Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile.* Cell, 2005. **123**(3): p. 507-19.
30. Collins, S.R., et al., *A strategy for extracting and analyzing large-scale quantitative epistatic interaction data.* Genome Biol, 2006. **7**(7): p. R63.
31. Pierce, S.E., et al., *A unique and universal molecular barcode array.* Nat Methods, 2006. **3**(8): p. 601-3.
32. Decourty, L., et al., *Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles.* Proc Natl Acad Sci U S A, 2008. **105**(15): p. 5821-6.
33. Pan, X., et al., *A robust toolkit for functional profiling of the yeast genome.* Mol Cell, 2004. **16**(3): p. 487-96.
34. Hillenmeyer, M.E., et al., *The chemical genomic portrait of yeast: uncovering a phenotype for all genes.* Science, 2008. **320**(5874): p. 362-5.
35. Typas, A., et al., *High-throughput, quantitative analyses of genetic interactions in E. coli.* Nat Methods, 2008. **5**(9): p. 781-7.
36. Nichols, R.J., et al., *Phenotypic Landscape of a Bacterial Cell* Cell, 2011. **144**(1): p. 143-156.

37. Butland, G., et al., *eSGA: E. coli synthetic genetic array analysis.* Nat Methods, 2008. **5**(9): p. 789-95.
38. Babu, M., et al., *Genetic interaction maps in Escherichia coli reveal functional crosstalk among cell envelope biogenesis pathways.* PLoS Genet, 2011. **7**(11): p. e1002377.
39. Babu, M., et al., *Quantitative genome-wide genetic interaction screens reveal global epistatic relationships of protein complexes in Escherichia coli.* PLoS Genet, 2014. **10**(2): p. e1004120.
40. Kumar, A., et al., *Conditional Epistatic Interaction Maps Reveal Global Functional Rewiring of Genome Integrity Pathways in Escherichia coli.* Cell Rep, 2016. **14**(3): p. 648-61.
41. Baba, T., et al., *Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.* Mol Syst Biol, 2006. **2**: p. 2006 0008.
42. Kitagawa, M., et al., *Complete set of ORF clones of Escherichia coli ASKA library (A Complete Set of E. coli K-12 ORF Archive): Unique Resources for Biological Research.* DNA Res, 2005. **12**(5): p. 291-9.
43. van Opijnen, T., K.L. Bodi, and A. Camilli, *Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms.* Nat Methods, 2009. **6**(10): p. 767-72.
44. Hawkins, J.S., et al., *Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi).* Methods Mol Biol, 2015. **1311**: p. 349-62.
45. Deutschbauer, A., et al., *Evidence-based annotation of gene function in Shewanella oneidensis MR-1 using genome-wide fitness profiling across 121 conditions.* PLoS Genet, 2011. **7**(11): p. e1002385.
46. Donald, R.G., et al., *A Staphylococcus aureus fitness test platform for mechanism-based profiling of antibacterial compounds.* Chem Biol, 2009. **16**(8): p. 826-36.
47. Huber, J., et al., *Chemical genetic identification of peptidoglycan inhibitors potentiating carbapenem activity against methicillin-resistant Staphylococcus aureus.* Chem Biol, 2009. **16**(8): p. 837-48.
48. van Opijnen, T. and A. Camilli, *A fine scale phenotype-genotype virulence map of a bacterial pathogen.* Genome Res, 2012.
49. Enstrom, M., et al., *Genotype-phenotype associations in a nonmodel prokaryote.* MBio, 2012. **3**(2).
50. Phillips, J.W., et al., *Discovery of kibdelomycin, a potent new class of bacterial type II topoisomerase inhibitor by chemical-genetic profiling in Staphylococcus aureus.* Chem Biol, 2011. **18**(8): p. 955-65.
51. Typas, A., et al., *Regulation of peptidoglycan synthesis by outer membrane proteins* Cell, 2010. **143**(7): p. 1097-1109.
52. Dixon, S.J., et al., *Significant conservation of synthetic lethal genetic interaction networks between distantly related eukaryotes.* Proc Natl Acad Sci U S A, 2008. **105**(43): p. 16653-8.
53. Hedges, S.B., *The origin and evolution of model organisms.* Nat Rev Genet, 2002. **3**(11): p. 838-49.

54. Kachroo, A.H., et al., *Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity.* Science, 2015. **348**(6237): p. 921-5.
55. Ryan, C.J., et al., *Hierarchical modularity and the evolution of genetic interactomes across species.* Mol Cell, 2012. **46**(5): p. 691-704.
56. Tischler, J., B. Lehner, and A.G. Fraser, *Evolutionary plasticity of genetic interaction networks.* Nat Genet, 2008. **40**(4): p. 390-1.
57. Porwollik, S., et al., *Defined single-gene and multi-gene deletion mutant collections in Salmonella enterica sv Typhimurium.* PLoS One, 2014. **9**(7): p. e99820.
58. Gordienko, E.N., M.D. Kazanov, and M.S. Gelfand, *Evolution of pan-genomes of Escherichia coli, Shigella spp., and Salmonella enterica.* J Bacteriol, 2013. **195**(12): p. 2786-92.
59. Baumler, A.J., et al., *Evolution of host adaptation in Salmonella enterica.* Infect Immun, 1998. **66**(10): p. 4579-87.
60. Groisman, E.A. and H. Ochman, *How Salmonella became a pathogen.* Trends Microbiol, 1997. **5**(9): p. 343-9.
61. Lan, R., P.R. Reeves, and S. Octavia, *Population structure, origins and evolution of major Salmonella enterica clones.* Infect Genet Evol, 2009. **9**(5): p. 996-1005.
62. Coburn, B., G.A. Grassl, and B.B. Finlay, *Salmonella, the host and disease: a brief review.* Immunol Cell Biol, 2007. **85**(2): p. 112-8.
63. Feasey, N.A., et al., *Invasive non-typhoidal salmonella disease: an emerging and neglected tropical disease in Africa.* Lancet, 2012. **379**(9835): p. 2489-99.
64. Kariuki, S., et al., *Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission?* J Med Microbiol, 2006. **55**(Pt 5): p. 585-91.
65. Morpeth, S.C., H.O. Ramadhani, and J.A. Crump, *Invasive non-Typhi Salmonella disease in Africa.* Clin Infect Dis, 2009. **49**(4): p. 606-11.
66. Abrahams, G.L. and M. Hensel, *Manipulating cellular transport and immune responses: dynamic interactions between intracellular Salmonella enterica and its host cells.* Cell Microbiol, 2006. **8**(5): p. 728-37.
67. Haraga, A., M.B. Ohlson, and S.I. Miller, *Salmonellae interplay with host cells.* Nature Reviews Microbiology, 2008. **6**(1): p. 53-66.
68. LaRock, D.L., A. Chaudhary, and S.I. Miller, *Salmonellae interactions with host processes.* Nature Reviews Microbiology, 2015. **13**(4): p. 191-205.
69. Santos, R.L., et al., *Life in the inflamed intestine, Salmonella style.* Trends Microbiol, 2009. **17**(11): p. 498-506.
70. Zhou, D. and J. Galan, *Salmonella entry into host cells: the work in concert of type III secreted effector proteins.* Microbes Infect, 2001. **3**(14-15): p. 1293-8.
71. Raffatellu, M., et al., *Lipocalin-2 resistance confers an advantage to Salmonella enterica serotype Typhimurium for growth and survival in the inflamed intestine.* Cell Host Microbe, 2009. **5**(5): p. 476-86.

72. Winter, S.E., et al., *Gut inflammation provides a respiratory electron acceptor for Salmonella.* Nature, 2010. **467**(7314): p. 426-9.
73. Thiennimitr, P., et al., *Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota.* Proc Natl Acad Sci U S A, 2011. **108**(42): p. 17480-5.
74. Winfield, M.D. and E.A. Groisman, *Phenotypic differences between Salmonella and Escherichia coli resulting from the disparate regulation of homologous genes.* Proc Natl Acad Sci U S A, 2004. **101**(49): p. 17162-7.
75. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proc. Nat. Acad. Sci. USA, 2000. **97**: p. 6640-6645.
76. Kroger, C., et al., *An infection-relevant transcriptomic compendium for Salmonella enterica Serovar Typhimurium.* Cell Host Microbe, 2013. **14**(6): p. 683-95.
77. Beltrao, P., G. Cagney, and N.J. Krogan, *Quantitative genetic interactions reveal biological modularity.* Cell, 2010. **141**(5): p. 739-45.
78. Li, L., C.J. Stoeckert, and D.S. Roos, *OrthoMCL: identification of ortholog groups for eukaryotic genomes.* Genome research, 2003. **13**(9): p. 2178-2189.
79. Delcour, A.H., *Outer membrane permeability and antibiotic resistance.* Biochim Biophys Acta, 2009. **1794**(5): p. 808-16.
80. Stevenson, G., et al., *Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster.* J Bacteriol, 1994. **176**(13): p. 4144-56.
81. Couturier, E. and E.P. Rocha, *Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes.* Mol Microbiol, 2006. **59**(5): p. 1506-18.
82. Bandyopadhyay, S., et al., *Functional maps of protein complexes from quantitative genetic interaction data.* PLoS Comput Biol, 2008. **4**(4): p. e1000065.
83. Majdalani, N. and S. Gottesman, *The Rcs phosphorelay: a complex signal transduction system.* Annu Rev Microbiol, 2005. **59**: p. 379-405.
84. Gerding, M.A., et al., *Self-enhanced accumulation of FtsN at Division Sites and Roles for Other Proteins with a SPOR domain (DamX, DedD, and RlpA) in Escherichia coli cell constriction.* J Bacteriol, 2009. **191**(24): p. 7383-401.
85. Park, J.T., *Why does Escherichia coli recycle its cell wall peptides?* Mol Microbiol, 1995. **17**(3): p. 421-6.
86. Ren, C.P., et al., *The ETT2 gene cluster, encoding a second type III secretion system from Escherichia coli, is present in the majority of strains but has undergone widespread mutational attrition.* J Bacteriol, 2004. **186**(11): p. 3547-60.
87. Ideses, D., et al., *A degenerate type III secretion system from septicemic Escherichia coli contributes to pathogenesis.* J Bacteriol, 2005. **187**(23): p. 8164-71.

88. Rishi, P., et al., *Salmonella enterica serovar Typhimurium hilA-lacZY fusion gene response to iron chelation or supplementation in rich and minimal media.* J Environ Sci Health B, 2004. **39**(5-6): p. 861-70.
89. Ellermeier, J.R. and J.M. Slauch, *Adaptation to the host environment: regulation of the SPI1 type III secretion system in Salmonella enterica serovar Typhimurium.* Curr Opin Microbiol, 2007. **10**(1): p. 24-9.
90. Cho, H., T. Uehara, and T.G. Bernhardt, *Beta-lactam antibiotics induce a lethal malfunctioning of the bacterial cell wall synthesis machinery.* Cell, 2014. **159**(6): p. 1300-11.
91. Folkesson, A., et al., *Multiple insertions of fimbrial operons correlate with the evolution of Salmonella serovars responsible for human disease.* Mol Microbiol, 1999. **33**(3): p. 612-22.
92. Folkesson, A., S. Lofdahl, and S. Normark, *The Salmonella enterica subspecies I specific centisome 7 genomic island encodes novel protein families present in bacteria living in close contact with eukaryotic cells.* Res Microbiol, 2002. **153**(8): p. 537-45.
93. Folkesson, A., et al., *Components of the peptidoglycan-recycling pathway modulate invasion and intracellular survival of Salmonella enterica serovar Typhimurium.* Cell Microbiol, 2005. **7**(1): p. 147-55.
94. Breidenstein, E.B., C. de la Fuente-Nunez, and R.E. Hancock, *Pseudomonas aeruginosa: all roads lead to resistance.* Trends Microbiol, 2011. **19**(8): p. 419-26.
95. Viollet, B., et al., *Cellular and molecular mechanisms of metformin: an overview.* Clin Sci (Lond), 2012. **122**(6): p. 253-70.
96. Group, U.K.P.D.S., *Effect of intensive blood-glucose control with metformin on complications in overweight patients with type 2 diabetes (UKPDS 34).* The Lancet, 1998. **352**(9131): p. 854-865.
97. Hundal, R.S., et al., *Mechanism by which metformin reduces glucose production in type 2 diabetes.* Diabetes, 2000. **49**(12): p. 2063-9.
98. Gunton, J.E., et al., *Metformin rapidly increases insulin receptor activation in human liver and signals preferentially through insulin-receptor substrate-2.* J Clin Endocrinol Metab, 2003. **88**(3): p. 1323-32.
99. Hur, K.Y. and M.S. Lee, *New mechanisms of metformin action: Focusing on mitochondria and the gut.* J Diabetes Investig, 2015. **6**(6): p. 600-9.
100. Napolitano, A., et al., *Novel gut-based pharmacology of metformin in patients with type 2 diabetes mellitus.* PLoS One, 2014. **9**(7): p. e100778.
101. Foretz, M., et al., *Metformin: from mechanisms of action to therapies.* Cell Metab, 2014. **20**(6): p. 953-66.
102. Foretz, M. and B. Viollet, *Therapy: Metformin takes a new route to clinical efficacy.* Nat Rev Endocrinol, 2015. **11**(7): p. 390-2.
103. Hajjar, J., M.A. Habra, and A. Naing, *Metformin: an old drug with new potential.* Expert Opin Investig Drugs, 2013. **22**(12): p. 1511-7.
104. Anisimov, V.N., *Metformin: do we finally have an anti-aging drug?* Cell Cycle, 2013. **12**(22): p. 3483-9.
105. Bonora, E., et al., *Lack of effect of intravenous metformin on plasma concentrations of glucose, insulin, C-peptide, glucagon and growth*

hormone in non-diabetic subjects. Curr Med Res Opin, 1984. **9**(1): p. 47-51.
106. Shin, N.R., et al., *An increase in the Akkermansia spp. population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice.* Gut, 2014. **63**(5): p. 727-35.
107. Lee, H. and G. Ko, *Effect of metformin on metabolic improvement and gut microbiota.* Appl Environ Microbiol, 2014. **80**(19): p. 5935-43.
108. Zhang, X., et al., *Modulation of gut microbiota by berberine and metformin during the treatment of high-fat diet-induced obesity in rats.* Sci Rep, 2015. **5**: p. 14405.
109. Forslund, K., et al., *Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota.* Nature, 2015. **528**(7581): p. 262-6.
110. Cabreiro, F., et al., *Metformin retards aging in C. elegans by altering microbial folate and methionine metabolism.* Cell, 2013. **153**(1): p. 228-39.
111. Onken, B. and M. Driscoll, *Metformin induces a dietary restriction-like state and the oxidative stress response to extend C. elegans Healthspan via AMPK, LKB1, and SKN-1.* PLoS One, 2010. **5**(1): p. e8758.
112. Calhoun, M.W., et al., *Energetic efficiency of Escherichia coli: effects of mutations in components of the aerobic respiratory chain.* J Bacteriol, 1993. **175**(10): p. 3020-5.
113. Matsushita, K., T. Ohnishi, and H.R. Kaback, *NADH-ubiquinone oxidoreductases of the Escherichia coli aerobic respiratory chain.* Biochemistry, 1987. **26**(24): p. 7732-7.
114. Villagra, N.A., et al., *SmvA, and not AcrB, is the major efflux pump for acriflavine and related compounds in Salmonella enterica serovar Typhimurium.* J Antimicrob Chemother, 2008. **62**(6): p. 1273-6.
115. Hongo, E., et al., *The methyl viologen-resistance-encoding gene smvA of Salmonella typhimurium.* Gene, 1994. **148**(1): p. 173-4.
116. Santiviago, C.A., et al., *The Salmonella enterica sv. Typhimurium smvA, yddG and ompD (porin) genes are required for the efficient efflux of methyl viologen.* Mol Microbiol, 2002. **46**(3): p. 687-98.
117. Shimada, T., et al., *A novel regulator RcdA of the csgD gene encoding the master regulator of biofilm formation in Escherichia coli.* Microbiologyopen, 2012. **1**(4): p. 381-94.
118. Purewal, A.S., I.G. Jones, and M. Midgley, *Cloning of the ethidium efflux gene from Escherichia coli.* FEMS Microbiol Lett, 1990. **56**(1-2): p. 73-6.
119. Morimyo, M., et al., *Cloning and characterization of the mvrC gene of Escherichia coli K-12 which confers resistance against methyl viologen toxicity.* Nucleic Acids Res, 1992. **20**(12): p. 3159-65.
120. Nishino, K. and A. Yamaguchi, *Analysis of a complete library of putative drug transporter genes in Escherichia coli.* J Bacteriol, 2001. **183**(20): p. 5803-12.
121. Hermann, L.S., *Metformin: a review of its pharmacological properties and therapeutic use.* Diabete Metab, 1979. **5**(3): p. 233-45.

122. Alsultan, A.A., et al., *High frequency of carbapenem-resistant Acinetobacter baumannii in patients with diabetes mellitus in Saudi Arabia.* J Med Microbiol, 2013. **62**(Pt 6): p. 885-8.
123. Dijkshoorn, L., A. Nemec, and H. Seifert, *An increasing threat in hospitals: multidrug-resistant Acinetobacter baumannii.* Nat Rev Microbiol, 2007. **5**(12): p. 939-51.
124. Han, S.H., *Review of hepatic abscess from Klebsiella pneumoniae. An association with diabetes mellitus and septic endophthalmitis.* West J Med, 1995. **162**(3): p. 220-4.
125. Oh, E., et al., *Selective ribosome profiling reveals the cotranslational chaperone action of trigger factor in vivo.* Cell, 2011. **147**(6): p. 1295-308.
126. Chong, Z.S., W.F. Woo, and S.S. Chng, *Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in Escherichia coli.* Mol Microbiol, 2015. **98**(6): p. 1133-46.
127. Parshin, A., et al., *DksA regulates RNA polymerase in Escherichia coli through a network of interactions in the secondary channel that includes Sequence Insertion 1.* Proc Natl Acad Sci U S A, 2015. **112**(50): p. E6862-71.
128. Yunck, R., H. Cho, and T.G. Bernhardt, *Identification of MltG as a potential terminase for peptidoglycan polymerization in bacteria.* Mol Microbiol, 2016. **99**(4): p. 700-18.
129. Jacobsen, A., et al., *The Salmonella enterica pan-genome.* Microb Ecol, 2011. **62**(3): p. 487-504.
130. Kelley, R. and T. Ideker, *Systematic interpretation of genetic interactions using protein networks.* Nat Biotechnol, 2005. **23**(5): p. 561-6.
131. Beyer, A., S. Bandyopadhyay, and T. Ideker, *Integrating physical and genetic maps: from genomes to interaction networks.* Nat Rev Genet, 2007. **8**(9): p. 699-710.
132. Ulitsky, I. and R. Shamir, *Pathway redundancy and protein essentiality revealed in the Saccharomyces cerevisiae interaction networks.* Mol Syst Biol, 2007. **3**: p. 104.
133. Zhang, L.V., et al., *Motifs, themes and thematic maps of an integrated Saccharomyces cerevisiae interaction network.* J Biol, 2005. **4**(2): p. 6.
134. Westermann, A.J., et al., *Dual RNA-seq unveils noncoding RNA functions in host-pathogen interactions.* Nature, 2016. **529**(7587): p. 496-501.
135. Van Der Walt, S., S.C. Colbert, and G. Varoquaux, *The NumPy array: a structure for efficient numerical computation.* Computing in Science & Engineering, 2011. **13**(2): p. 22-30.
136. Jones, E., T. Oliphant, and P. Peterson, *SciPy: Open source scientific tools for Python.* http://www. scipy. org/, 2001.
137. McKinney, W. and others, *Data structures for statistical computing in Python*, in *Proceedings of the 9th Python in Science Conference*. 2010. p. 51-56.
138. M\"ullner, D., *fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python.* Journal of Statistical Software, 2013. **53**(9): p. 1-18.

139. Pedregosa, F., et al., *Scikit-learn: Machine learning in Python.* The Journal of Machine Learning Research, 2011. **12**: p. 2825-2830.
140. Hagberg, A.A., D.A. Schult, and P.J. Swart, *Exploring network structure, dynamics, and function using NetworkX*, in *Proceedings of the 7th Python in Science Conference (SciPy2008)*. 2008: Pasadena, CA USA. p. 11-15.
141. Cock, P.J., et al., *Biopython: freely available Python tools for computational molecular biology and bioinformatics.* Bioinformatics, 2009. **25**(11): p. 1422-1423.
142. Hunter, J.D., *Matplotlib: A 2D graphics environment.* Computing in Science & Engineering, 2007: p. 90-95.
143. Waskom, M., et al. *seaborn: v0.6.0 (June 2015)*. 2015 June; Available from: 10.5281/zenodo.19108.
144. Perez, F. and B.E. Granger, *IPython: a system for interactive scientific computing.* Computing in Science & Engineering, 2007. **9**(3): p. 21-29.
145. Nichols, R.J., et al., *Phenotypic landscape of a bacterial cell.* Cell, 2011. **144**(1): p. 143-156.
146. Keseler, I.M., et al., *EcoCyc: fusing model organism databases with systems biology.* Nucleic acids research, 2013. **41**(D1): p. D605-D612.
147. Caspi, R., et al., *The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.* Nucleic acids research, 2008. **36**(suppl 1): p. D623-D631.
148. Mao, X., et al., *DOOR 2.0: presenting operons and their functions through dynamic and integrated views.* Nucleic acids research, 2014. **42**(D1): p. D654-D659.
149. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic acids research, 2000. **28**(1): p. 27-30.
150. Kanehisa, M., et al., *KEGG as a reference resource for gene and protein annotation.* Nucleic acids research, 2016. **44**(D1): p. D457-D462.
151. Snel, B., et al., *STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene.* Nucleic acids research, 2000. **28**(18): p. 3442-3444.
152. Szklarczyk, D., et al., *STRING v10: protein-protein interaction networks, integrated over the tree of life.* Nucleic acids research, 2014: p. gku1003.
153. von Mering, C., et al., *STRING 7--recent developments in the integration and prediction of protein interactions.* Nucleic Acids Res, 2007. **35**(Database issue): p. D358-62.
154. von Mering, C., et al., *STRING: known and predicted protein-protein associations, integrated and transferred across organisms.* Nucleic Acids Res, 2005. **33**(Database issue): p. D433-7.
155. Sahl, J.W., et al., *The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes.* PeerJ, 2014. **2**: p. e332.
156. Gao, F. and C.T. Zhang, *DoriC: a database of oriC regions in bacterial genomes.* Bioinformatics, 2007. **23**(14): p. 1866-7.

157. Gao, F., H. Luo, and C.T. Zhang, *DoriC 5.0: an updated database of oriC regions in both bacterial and archaeal genomes.* Nucleic Acids Res, 2013. **41**(Database issue): p. D90-3.

Appendix

A. Base media

LB Lennox

1% (w/v) Bacto tryptone
0.5 % (w/v) Bacto yeast extract
0.5 % (w/v) NaCl

LB no salt

1% (w/v) Bacto tryptone
0.5 % (w/v) Bacto yeast extract

M9 Minimal media

1x M9 salts
1 mM MgSO4
0.1 mM CaCl2
20% (w/v) glucose (or other carbon sources with equimolar C atom concentration)

M9 Minimal media complete

1x M9 salts
1 mM MgSO4
mM CaCl2
20% (w/v) glucose (or other carbon sources with equimolar C atom concentration)
1x AAA
1x BAA
1x FAA

PCN media

80 mM MES (pH 5.8) for InSPI2
80 mM MOPS (pH 7.4) for NonSPI2
4 mM Tricine
100 µM FeCl3
376 µM K2SO4
50 mM NaCl
0.4 mM K2HPO4/KH2PO4 pH 5.8 for InSPI2
25 mM K2HPO4/KH2PO4 pH 7.4 for NonSPI2
0.4 % glucose (22.2 mM)
15 mM NH4Cl
1 mM MgSO4
0.01 mM CaCl2
10 nM Na2MoO4
10 nM Na2SeO3
4 nM H3BO3
300 nM CoCl2
100 nM CuSO4
800 nM MnCl2
1 nM ZnSO4

B. Tested conditions

Label	Chemical	Concentration	Media background	Batch
4AMINOSALICYLICACID-10	4-Aminosalicylic acid	10 µg/ml	LB Lennox	7
4AMINOSALICYLICACID-100	4-Aminosalicylic acid	100 µg/ml	LB Lennox	7
4AMINOSALICYLICACID-200	4-Aminosalicylic acid	200 µg/ml	LB Lennox	7
4AMINOSALICYLICACID-50	4-Aminosalicylic acid	50 µg/ml	LB Lennox	7
5BROMODEOXYURIDINE-10	5-Bromodeoxyuridine	10 µg/ml	LB Lennox	7
5BROMODEOXYURIDINE-100	5-Bromodeoxyuridine	100 µg/ml	LB Lennox	7
5BROMODEOXYURIDINE-200	5-Bromodeoxyuridine	200 µg/ml	LB Lennox	7
5BROMODEOXYURIDINE-50	5-Bromodeoxyuridine	50 µg/ml	LB Lennox	7
5FLUOROURACIL-10UM	5-Fluorouracil	10 µM	LB Lennox	6
5FLUOROURACIL-1UM	5-Fluorouracil	1 µM	LB Lennox	6
5FLUOROURACIL-2UM	5-Fluorouracil	2 µM	LB Lennox	6
5FLUOROURACIL-5UM	5-Fluorouracil	5 µM	LB Lennox	6
A22-0.05	A22	0.05 µg/ml	LB Lennox	4

Label	Chemical	Concentration	Media background	Batch
A22-0.1	A22	0.1 µg/ml	LB Lennox	4
A22-0.25	A22	0.25 µg/ml	LB Lennox	4
A22-0.5	A22	0.5 µg/ml	LB Lennox	4
A22-2.0	A22	2 µg/ml	LB Lennox	7
ACRIFLAVINE-10	Acriflavine	10 µg/ml	LB Lennox	6
ACRIFLAVINE-2	Acriflavine	2 µg/ml	LB Lennox	6
ACRIFLAVINE-5	Acriflavine	5 µg/ml	LB Lennox	6
ADULTBOVINESERUM-10PCVV	Adult bovine serum	10 %(v/v)	LB Lennox	7
ADULTBOVINESERUMHI-10PCVV	Adult bovine serum - heat inactivated	10 %(v/v)	LB Lennox	7
AMIKACIN-0.05	Amikacin	0.05 µg/ml	LB Lennox	2
AMIKACIN-0.1	Amikacin	0.1 µg/ml	LB Lennox	2
AMIKACIN-0.25	Amikacin	0.25 µg/ml	LB Lennox	2
AMOXICILLIN-0.05	Amoxicillin	0.05 µg/ml	LB Lennox	3
AMOXICILLIN-0.1	Amoxicillin	0.1 µg/ml	LB Lennox	3
AMOXICILLIN-0.25	Amoxicillin	0.25 µg/ml	LB Lennox	3
AMPICILLIN-0.1	Ampicillin	0.1 µg/ml	LB Lennox	5
AMPICILLIN-0.25	Ampicillin	0.25 µg/ml	LB Lennox	5
AMPICILLIN-0.5	Ampicillin	0.5 µg/ml	LB Lennox	5
ARTICAINE-100	Articaine	100 µg/ml	LB Lennox	7
ARTICAINE-250	Articaine	250 µg/ml	LB Lennox	7
ARTICAINE-50	Articaine	50 µg/ml	LB Lennox	7
ARTICAINE-500	Articaine	500 µg/ml	LB Lennox	7
AZIDOTHYMIDINE-0.5	Azidothymidine	0.5 µg/ml	LB Lennox	7
AZIDOTHYMIDINE-1.0	Azidothymidine	1 µg/ml	LB Lennox	7
AZIDOTHYMIDINE-2.5	Azidothymidine	2.5 µg/ml	LB Lennox	7
AZIDOTHYMIDINE-5.0	Azidothymidine	5 µg/ml	LB Lennox	7
AZITHROMYCIN-0.01	Azithromycin	0.01 µg/ml	LB Lennox	5
AZITHROMYCIN-0.025	Azithromycin	0.025 µg/ml	LB Lennox	5
AZITHROMYCIN-0.05	Azithromycin	0.05 µg/ml	LB Lennox	5
AZITHROMYCIN-0.1	Azithromycin	0.1 µg/ml	LB Lennox	5
AZTREONAM-0.0005	Aztreonam	0.0005 µg/ml	LB Lennox	2
AZTREONAM-0.001	Aztreonam	0.001 µg/ml	LB Lennox	2
AZTREONAM-0.002	Aztreonam	0.002 µg/ml	LB Lennox	2
AZTREONAM-0.005	Aztreonam	0.005 µg/ml	LB Lennox	2
BENZALKONIUM-1	Benzalkonium	1 µg/ml	LB Lennox	4
BENZALKONIUM-10	Benzalkonium	10 µg/ml	LB Lennox	4
BENZALKONIUM-2	Benzalkonium	2 µg/ml	LB Lennox	4
BENZALKONIUM-5	Benzalkonium	5 µg/ml	LB Lennox	4
BERBERINE-100	Berberine	100 µg/ml	LB Lennox	7
BERBERINE-200	Berberine	200 µg/ml	LB Lennox	7
BERBERINE-25	Berberine	25 µg/ml	LB Lennox	7
BERBERINE-50	Berberine	50 µg/ml	LB Lennox	7
BILESALTS-0.05PCWV	Bile salts	0.05 %(w/v)	LB Lennox	5
BILESALTS-0.1PCWV	Bile salts	0.1 %(w/v)	LB Lennox	5
BILESALTS-0.25PCWV	Bile salts	0.25 %(w/v)	LB Lennox	5
BILESALTS-0.5PCWV	Bile salts	0.5 %(w/v)	LB Lennox	5
BILESALTS-1PCWV	Bile salts	1 %(w/v)	LB Lennox	7
BILESALTS-2PCWV	Bile salts	2 %(w/v)	LB Lennox	7
BIPYRIDIL-0.01MM	Bipyridil	0.01 mM	LB Lennox	5
BIPYRIDIL-0.025MM	Bipyridil	0.025 mM	LB Lennox	6
BIPYRIDIL-0.05MM	Bipyridil	0.05 mM	LB Lennox	5
BIPYRIDIL-0.1MM	Bipyridil	0.1 mM	LB Lennox	6
BLASTICIDINS-1	Blasticidin S	1 µg/ml	LB Lennox	7
BLASTICIDINS-10	Blasticidin S	10 µg/ml	LB Lennox	7
BLASTICIDINS-5	Blasticidin S	5 µg/ml	LB Lennox	7
BLEOMYCIN-0.01	Bleomycin	0.01 µg/ml	LB Lennox	4
BLEOMYCIN-0.025	Bleomycin	0.025 µg/ml	LB Lennox	4
BLEOMYCIN-0.05	Bleomycin	0.05 µg/ml	LB Lennox	4
BLEOMYCIN-0.1	Bleomycin	0.1 µg/ml	LB Lennox	4
BUTYRATE-10MM	Butyrate	10 mM	LB Lennox	7
BUTYRATE-20MM	Butyrate	20 mM	LB Lennox	7
CARBAMAZEPINE-100	Carbamazepine	100 µg/ml	LB Lennox	7
CARBAMAZEPINE-250	Carbamazepine	250 µg/ml	LB Lennox	7
CARBAMAZEPINE-50	Carbamazepine	50 µg/ml	LB Lennox	7
CCCP-0.1	CCCP (Carbonyl cyanide 3-chlorophenylhydrazone)	0.1 µg/ml	LB Lennox	4
CCCP-0.5	CCCP (Carbonyl cyanide 3-chlorophenylhydrazone)	0.5 µg/ml	LB Lennox	4
CCCP-1.0	CCCP (Carbonyl cyanide 3-chlorophenylhydrazone)	1.0 µg/ml	LB Lennox	4
CCCP-2.0	CCCP (Carbonyl cyanide 3-chlorophenylhydrazone)	2.0 µg/ml	LB Lennox	4
CECROPINB-0.05	Cecropin B	0.05 µg/ml	LB Lennox	6
CECROPINB-0.1	Cecropin B	0.1 µg/ml	LB Lennox	6
CECROPINB-0.25	Cecropin B	0.25 µg/ml	LB Lennox	6
CECROPINB-0.5	Cecropin B	0.5 µg/ml	LB Lennox	6
CEFACLOR-0.05	Cefaclor	0.05 µg/ml	LB Lennox	3
CEFACLOR-0.1	Cefaclor	0.1 µg/ml	LB Lennox	3
CEFACLOR-0.25	Cefaclor	0.25 µg/ml	LB Lennox	3
CEFACLOR-0.5	Cefaclor	0.5 µg/ml	LB Lennox	3
CEFACLOR-0.5REPEAT	Cefaclor	0.5 µg/ml	LB Lennox	7

Label	Chemical	Concentration	Media background	Batch
CEFACLOR+TOBRAMYCIN-0.5+0.5	Cefaclor + Tobramycin	0.5+0.5 µg/ml	LB Lennox	7
CEFOTAXIME-0.0025	Cefotaxime	0.0025 µg/ml	LB Lennox	3
CEFOTAXIME-0.005	Cefotaxime	0.005 µg/ml	LB Lennox	3
CEFOTAXIME-0.01	Cefotaxime	0.01 µg/ml	LB Lennox	3
CEFOTAXIME-0.02	Cefotaxime	0.02 µg/ml	LB Lennox	3
CEFOXITIN-0.05	Cefoxitin	0.05 µg/ml	LB Lennox	7
CEFOXITIN-0.1	Cefoxitin	0.1 µg/ml	LB Lennox	7
CEFOXITIN-0.25	Cefoxitin	0.25 µg/ml	LB Lennox	7
CEFOXITIN-0.5	Cefoxitin	0.5 µg/ml	LB Lennox	7
CEFSULODIN-1	Cefsulodin	1 µg/ml	LB Lennox	4
CEFSULODIN-10	Cefsulodin	10 µg/ml	LB Lennox	4
CEFSULODIN-2	Cefsulodin	2 µg/ml	LB Lennox	4
CEFSULODIN-5	Cefsulodin	5 µg/ml	LB Lennox	4
CEFTAZIDIME-0.01	Ceftazidime	0.01 µg/ml	LB Lennox	7
CEFTAZIDIME-0.02	Ceftazidime	0.02 µg/ml	LB Lennox	7
CEFUROXIME-0.1	Cefuroxime	0.1 µg/ml	LB Lennox	4
CEFUROXIME-0.25	Cefuroxime	0.25 µg/ml	LB Lennox	4
CEFUROXIME-0.5	Cefuroxime	0.5 µg/ml	LB Lennox	4
CEPHALEXIN-1	Cephalexin	1 µg/ml	LB Lennox	4
CEPHALEXIN-10	Cephalexin	10 µg/ml	LB Lennox	4
CEPHALEXIN-2	Cephalexin	2 µg/ml	LB Lennox	4
CEPHALEXIN-5	Cephalexin	5 µg/ml	LB Lennox	4
CEPHALOTHIN-0.05	Cephalothin	0.05 µg/ml	LB Lennox	4
CEPHALOTHIN-0.1	Cephalothin	0.1 µg/ml	LB Lennox	4
CEPHALOTHIN-0.25	Cephalothin	0.25 µg/ml	LB Lennox	4
CEPHALOTHIN-0.5	Cephalothin	0.5 µg/ml	LB Lennox	4
CERULENIN-0.5	Cerulenin	0.5 µg/ml	LB Lennox	7
CERULENIN-1.0	Cerulenin	1 µg/ml	LB Lennox	7
CERULENIN-2.5	Cerulenin	2.5 µg/ml	LB Lennox	7
CERULENIN-5.0	Cerulenin	5 µg/ml	LB Lennox	7
CHIR090-0.0025	CHIR-090	0.0025 µg/ml	LB Lennox	5
CHIR090-0.005	CHIR-090	0.005 µg/ml	LB Lennox	5
CHIR090-0.01	CHIR-090	0.01 µg/ml	LB Lennox	5
CHIR090-0.02	CHIR-090	0.02 µg/ml	LB Lennox	6
CHLORHEXIDINE-0.05REPEAT	Chlorhexidine	0.05 µg/ml	LB Lennox	5
CHLORHEXIDINE-0.1REPEAT	Chlorhexidine	0.1 µg/ml	LB Lennox	5
CHLORHEXIDINE-0.25REPEAT	Chlorhexidine	0.25 µg/ml	LB Lennox	5
CHLORHEXIDINE-0.5REPEAT	Chlorhexidine	0.5 µg/ml	LB Lennox	5
CHLORHEXIDINE-1	Chlorhexidine	1 µg/ml	LB Lennox	7
CHLORHEXIDINE-2.5	Chlorhexidine	2.5 µg/ml	LB Lennox	7
CHLORPROMAZINE-10UM	Chlorpromazine	10 µM	LB Lennox	6
CHLORPROMAZINE-1UM	Chlorpromazine	1 µM	LB Lennox	6
CHLORPROMAZINE-2.5UM	Chlorpromazine	2.5 µM	LB Lennox	6
CHLORPROMAZINE-5UM	Chlorpromazine	5 µM	LB Lennox	6
CINOXACIN-0.01	Cinoxacin	0.01 µg/ml	LB Lennox	7
CINOXACIN-0.05	Cinoxacin	0.05 µg/ml	LB Lennox	7
CINOXACIN-0.1	Cinoxacin	0.1 µg/ml	LB Lennox	7
CINOXACIN-0.25	Cinoxacin	0.25 µg/ml	LB Lennox	7
CIPROFLOXACIN-0.00025NGML	Ciprofloxacin	0.25 ng/ml	LB Lennox	3
CIPROFLOXACIN-0.0005NGML	Ciprofloxacin	0.5 ng/ml	LB Lennox	3
CIPROFLOXACIN-0.001NGML	Ciprofloxacin	1.0 ng/ml	LB Lennox	3
CIPROFLOXACIN-0.002NGML	Ciprofloxacin	2.0 ng/ml	LB Lennox	3
CLARITHROMYCIN-0.1	Clarithromycin	0.1 µg/ml	LB Lennox	5
CLARITHROMYCIN-0.25	Clarithromycin	0.25 µg/ml	LB Lennox	5
CLARITHROMYCIN-0.5	Clarithromycin	0.5 µg/ml	LB Lennox	5
CLARITHROMYCIN-1.0	Clarithromycin	1.0 µg/ml	LB Lennox	6
CLINDAMYCIN-1	Clindamycin	1 µg/ml	LB Lennox	4
CLINDAMYCIN-20	Clindamycin	20 µg/ml	LB Lennox	4
CLINDAMYCIN-5	Clindamycin	5 µg/ml	LB Lennox	4
CLINDAMYCIN-50	Clindamycin	50 µg/ml	LB Lennox	4
CLOFAZIMINE-0.1	Clofazimine	0.1 µg/ml	LB Lennox	5
CLOFAZIMINE-0.25	Clofazimine	0.25 µg/ml	LB Lennox	5
CLOFAZIMINE-0.5	Clofazimine	0.5 µg/ml	LB Lennox	5
CLOFAZIMINE-1.0	Clofazimine	1.0 µg/ml	LB Lennox	5
CLOFAZIMINE-10	Clofazimine	10 µg/ml	LB Lennox	7
CLOFAZIMINE-5	Clofazimine	5 µg/ml	LB Lennox	7
CM-AKKKERMANSIAMUCINIPHILA	Conditioned media Akkermansia muciniphila	n/a	LB Lennox	7
CM-LBCONTROL	Conditioned media LB control	n/a	LB Lennox	7
CM-MGAMCONTROL	Conditioned media MGAM control	n/a	LB Lennox	7
CM-PSEUDOMONASAERUGINOSA01	Conditioned media Pseudomonas aeruginosa PA01	n/a	LB Lennox	7
CM-PSEUDOMONASAERUGINOSA14	Conditioned media Pseudomonas aeruginosa PA14	n/a	LB Lennox	7
COBALT-0.05MM	CoCl2	0.05 mM	LB Lennox	5
COBALT-0.1MM	CoCl2	0.1 mM	LB Lennox	6
COBALT-0.3MM	CoCl2	0.3 mM	LB Lennox	5
COBALT-0.5MM	CoCl2	0.5 mM	LB Lennox	6
COLICINM-0.004PCVV	Colicin M	0.004 %(v/v)	LB Lennox	7
COLICINM-0.008PCVV	Colicin M	0.008 %(v/v)	LB Lennox	7

Label	Chemical	Concentration	Media background	Batch
COLICINM-0.02PCVV	Colicin M	0.02 %(v/v)	LB Lennox	7
COLICINM-0.04PCVV	Colicin M	0.04 %(v/v)	LB Lennox	7
COLISTIN-0.05	Colistin	0.05 µg/ml	LB Lennox	4
COLISTIN-0.1	Colistin	0.1 µg/ml	LB Lennox	4
COLISTIN-0.25	Colistin	0.25 µg/ml	LB Lennox	4
COLISTIN-0.5	Colistin	0.5 µg/ml	LB Lennox	4
COPPER-0.5MM	CuCl2	0.5 mM	LB Lennox	6
COPPER-1MM	CuCl2	1 mM	LB Lennox	6
COPPER-2MM	CuCl2	2 mM	LB Lennox	6
COPPER-4MM	CuCl2	4 mM	LB Lennox	6
CTAB-16	CTAB (Cetyltrimethylammonium bromide)	16 µg/ml	LB Lennox	7
CTAB-32	CTAB (Cetyltrimethylammonium bromide)	32 µg/ml	LB Lennox	7
CTAB-4	CTAB (Cetyltrimethylammonium bromide)	4 µg/ml	LB Lennox	7
CTAB-8	CTAB (Cetyltrimethylammonium bromide)	8 µg/ml	LB Lennox	7
CYCLOSERINED-1.0	Cycloserine D	1.0 µg/ml	LB Lennox	6
CYCLOSERINED-2.5	Cycloserine D	2.5 µg/ml	LB Lennox	5
CYCLOSERINED-5.0	Cycloserine D	5.0 µg/ml	LB Lennox	6
CYCLOSERINED-7.5	Cycloserine D	7.5 µg/ml	LB Lennox	5
DEFENSINHNP1-16.67NGML	Defensin HNP-1 Human	16.67 ng/ml	LB Lennox	7
DEFENSINHNP2-16.67NGML	Defensin HNP-2 Human	16.67 ng/ml	LB Lennox	7
DLSERINEHYDROXAMATE-25	DL-Serine hydroxamate	25 µg/ml	LB Lennox	7
DLSERINEHYDROXAMATE-50	DL-Serine hydroxamate	50 µg/ml	LB Lennox	7
DTPA-10	DTPA (Diethylene triamine pentaacetic acid)	10 µg/ml	LB Lennox	7
DTPA-25	DTPA (Diethylene triamine pentaacetic acid)	25 µg/ml	LB Lennox	7
DTPA-5	DTPA (Diethylene triamine pentaacetic acid)	5 µg/ml	LB Lennox	7
DTPA-50	DTPA (Diethylene triamine pentaacetic acid)	50 µg/ml	LB Lennox	7
EDTA-0.1MM	EDTA	0.1 mM	LB Lennox	4
EDTA-0.5MM	EDTA	0.5 mM	LB Lennox	4
EGCG-1	EGCG (Epigallocatechin gallate)	1 µg/ml	LB Lennox	4
EGCG-20	EGCG (Epigallocatechin gallate)	20 µg/ml	LB Lennox	4
EGCG-5	EGCG (Epigallocatechin gallate)	5 µg/ml	LB Lennox	4
EGCG-50	EGCG (Epigallocatechin gallate)	50 µg/ml	LB Lennox	4
EGTA-0.1MM	EGTA	0.1 mM	LB Lennox	4
EGTA-0.5MM	EGTA	0.5 mM	LB Lennox	4
EGTA-1MM	EGTA	1.0 mM	LB Lennox	4
EPINEPHRINE-0.5MM	Epinephrine	0.5 mM	LB Lennox	7
ERYTHROMYCIN-0.5	Erythromycin	0.5 µg/ml	LB Lennox	6
ERYTHROMYCIN-1	Erythromycin	1.0 µg/ml	LB Lennox	6
ERYTHROMYCIN-2	Erythromycin	2.0 µg/ml	LB Lennox	6
ERYTHROMYCIN-4	Erythromycin	4.0 µg/ml	LB Lennox	6
ETHIDIUMBROMIDE-1REPEAT	Ethidium bromide	1 µg/ml	LB Lennox	7
ETHIDIUMBROMIDE-20REPEAT	Ethidium bromide	20 µg/ml	LB Lennox	7
ETHIDIUMBROMIDE-5REPEAT	Ethidium bromide	5 µg/ml	LB Lennox	7
ETHIDIUMBROMIDE-5REPEAT0	Ethidium bromide	50 µg/ml	LB Lennox	7
FOSFOMYCIN-0.005	Fosfomycin	0.005 µg/ml	LB Lennox	3
FOSFOMYCIN-0.01	Fosfomycin	0.01 µg/ml	LB Lennox	3
FOSFOMYCIN-0.025	Fosfomycin	0.025 µg/ml	LB Lennox	3
FOSFOMYCIN-0.05	Fosfomycin	0.05 µg/ml	LB Lennox	3
FUSIDICACID-1	Fusidic acid	1 µg/ml	LB Lennox	4
FUSIDICACID-20	Fusidic acid	20 µg/ml	LB Lennox	4
FUSIDICACID-5	Fusidic acid	5 µg/ml	LB Lennox	4
FUSIDICACID-50	Fusidic acid	50 µg/ml	LB Lennox	4
G418SULFATE-0.01	G418 sulfate	0.01 µg/ml	LB Lennox	7
G418SULFATE-0.05	G418 sulfate	0.05 µg/ml	LB Lennox	7
G418SULFATE-0.1	G418 sulfate	0.1 µg/ml	LB Lennox	7
G418SULFATE-0.25	G418 sulfate	0.25 µg/ml	LB Lennox	7
GENTAMICIN-0.01	Gentamicin	0.01 µg/ml	LB Lennox	2
GENTAMICIN-0.025	Gentamicin	0.025 µg/ml	LB Lennox	2
GENTAMICIN-0.05	Gentamicin	0.05 µg/ml	LB Lennox	2
GENTAMICIN-0.1	Gentamicin	0.1 µg/ml	LB Lennox	2
GRAMICIDIN-10	Gramicidin	10 µg/ml	LB Lennox	7
GRAMICIDIN-20	Gramicidin	20 µg/ml	LB Lennox	7
GRAMICIDIN-5	Gramicidin	5 µg/ml	LB Lennox	7
GRAMICIDIN-50	Gramicidin	50 µg/ml	LB Lennox	7
HYDROXYUREA-1	Hydroxyurea	1 µg/ml	LB Lennox	7
HYDROXYUREA-10	Hydroxyurea	10 µg/ml	LB Lennox	7
HYDROXYUREA-2.5	Hydroxyurea	2.5 µg/ml	LB Lennox	7
HYDROXYUREA-5	Hydroxyurea	5 µg/ml	LB Lennox	7
IMIPENEM-0.005	Imipenem	0.005 µg/ml	LB Lennox	3
IMIPENEM-0.005REPEAT	Imipenem	0.005 µg/ml	LB Lennox	7
IMIPENEM-0.01	Imipenem	0.01 µg/ml	LB Lennox	3
IMIPENEM-0.01REPEAT	Imipenem	0.01 µg/ml	LB Lennox	7
IMIPENEM-0.02	Imipenem	0.02 µg/ml	LB Lennox	3
IMIPENEM-0.025	Imipenem	0.025 µg/ml	LB Lennox	7
IMIPENEM-0.04	Imipenem	0.04 µg/ml	LB Lennox	3
IMIPENEM-0.05	Imipenem	0.05 µg/ml	LB Lennox	7
INSPI2-BASIC	InSPI2	n/a	PCN medium	7

Label	Chemical	Concentration	Media background	Batch
INSPI2-LOWIRON	InSPI2 low iron	n/a	PCN medium	7
INSPI2-LOWMAGNESIUM	InSPI2 low magnesium	n/a	PCN medium	7
ISONIAZID-100	Isoniazid	100 µg/ml	LB Lennox	7
ISONIAZID-20	Isoniazid	20 µg/ml	LB Lennox	7
ISONIAZID-5	Isoniazid	5 µg/ml	LB Lennox	7
ISONIAZID-50	Isoniazid	50 µg/ml	LB Lennox	7
KASUGAMYCIN-1.0	Kasugamycin	1.0 µg/ml	LB Lennox	6
KASUGAMYCIN-10.0	Kasugamycin	10.0 µg/ml	LB Lennox	6
KASUGAMYCIN-2.5	Kasugamycin	2.5 µg/ml	LB Lennox	5
KASUGAMYCIN-5.0	Kasugamycin	5 µg/ml	LB Lennox	5
LEVOFLOXACIN-0.005	Levofloxacin	0.005 µg/ml	LB Lennox	7
LEVOFLOXACIN-0.01	Levofloxacin	0.01 µg/ml	LB Lennox	7
LIDOCAINE-10	Lidocaine	10 µg/ml	LB Lennox	7
LIDOCAINE-100	Lidocaine	100 µg/ml	LB Lennox	7
LIDOCAINE-200	Lidocaine	200 µg/ml	LB Lennox	7
LIDOCAINE-50	Lidocaine	50 µg/ml	LB Lennox	7
LINEZOLID-1	Linezolid	1 µg/ml	LB Lennox	4
LINEZOLID-10	Linezolid	10 µg/ml	LB Lennox	4
LINEZOLID-2	Linezolid	2 µg/ml	LB Lennox	4
LINEZOLID-5	Linezolid	5 µg/ml	LB Lennox	4
LL37-0.67	LL37 Human	0.67 µg/ml	LB Lennox	7
LOPERAMIDE-100	Loperamide	100 µg/ml	LB Lennox	4
LOPERAMIDE-20	Loperamide	20 µg/ml	LB Lennox	4
LOPERAMIDE-5	Loperamide	5 µg/ml	LB Lennox	4
LOPERAMIDE-50	Loperamide	50 µg/ml	LB Lennox	4
LYSOSTAPHIN-3.33	Lysostaphin	3.33 µg/ml	LB Lennox	7
M9-ARABINOSE	M9Minimal - Arabinose	0.4 %(w/v)	M9 Minimal	4
M9-DMALICACID	M9Minimal - D-Malic acid	n/a	M9 Minimal	7
M9-FRUCTOSE	M9Minimal - Fructose	0.4 %(w/v)	M9 Minimal	4
M9-FUMARATE	M9Minimal - Fumarate	0.4 %(w/v)	M9 Minimal	4
M9-GALACTOSE	M9Minimal - Galactose	n/a	M9 Minimal	7
M9-GLUCOSAMINE	M9Minimal - Glucosamine	0.4 %(w/v)	M9 Minimal	5
M9-GLUCOSE	M9Minimal - Glucose	0.4 %(w/v)	M9 Minimal	4
M9-GLYCEROL	M9Minimal - Glycerol	0.4 %(w/v)	M9 Minimal	4
M9-LACTICACID	M9Minimal - Lactic acid	n/a	M9 Minimal	7
M9-LRHAMNOSE	M9Minimal - L-Rhamnose	n/a	M9 Minimal	7
M9-MALTOSE	M9Minimal - Maltose	0.4 %(w/v)	M9 Minimal	4
M9-METHYLPYRUVATE	M9Minimal - Methylpyruvate	n/a	M9 Minimal	7
M9-NACETYLGLUCOSAMINE	M9Minimal - N-Acetylglucosamine	0.4 %(w/v)	M9 Minimal	5
M9-PROLINE	M9Minimal - Proline	0.3 %(w/v)	M9 Minimal	4
M9-PYRUVATE	M9Minimal - Pyruvate	0.4 %(w/v)	M9 Minimal	5
M9-SORBITOL	M9Minimal - Sorbitol	0.4 %(w/v)	M9 Minimal	5
M9-SUCCINATE	M9Minimal - Succinate	0.4 %(w/v)	M9 Minimal	4
M9COMPLETE-MINUSAAA	M9Complete - minusAAA	n/a	M9 Complete	6
M9COMPLETE-MINUSBAA	M9Complete - minusBAA	n/a	M9 Complete	6
M9COMPLETE-MINUSFAA	M9Complete - minusFAA	n/a	M9 Complete	6
MECILLINAM-0.01	Mecillinam	0.01 µg/ml	LB Lennox	2
MECILLINAM-0.025	Mecillinam	0.025 µg/ml	LB Lennox	2
MECILLINAM-0.05	Mecillinam	0.05 µg/ml	LB Lennox	2
MEROPENEM-0.1NGML	Meropenem	0.1 ng/ml	LB Lennox	4
MEROPENEM-0.25NGML	Meropenem	0.25 ng/ml	LB Lennox	4
MEROPENEM-0.5NGML	Meropenem	0.5 ng/ml	LB Lennox	4
MEROPENEM-1.0NGML	Meropenem	1.0 ng/ml	LB Lennox	4
METFORMIN-100MM	Metformin	100 mM	LB Lennox	7
METFORMIN-10MM	Metformin	10 mM	LB Lennox	7
METFORMIN-25MM	Metformin	25 mM	LB Lennox	7
METFORMIN-50MM	Metformin	50 mM	LB Lennox	7
METRONIDAZOLE-1.0	Metronidazole	1 µg/ml	LB Lennox	2
METRONIDAZOLE-10.0	Metronidazole	10 µg/ml	LB Lennox	2
METRONIDAZOLE-30.0	Metronidazole	30 µg/ml	LB Lennox	2
METRONIDAZOLE-5.0	Metronidazole	5 µg/ml	LB Lennox	2
MINOCYCLINE-0.05	Minocycline	0.05 µg/ml	LB Lennox	3
MINOCYCLINE-0.1	Minocycline	0.1 µg/ml	LB Lennox	3
MINOCYCLINE-0.25	Minocycline	0.25 µg/ml	LB Lennox	3
MINOCYCLINE-0.5	Minocycline	0.5 µg/ml	LB Lennox	3
MITOMYCINC-0.005	Mitomycin C	0.005 µg/ml	LB Lennox	4
MITOMYCINC-0.01	Mitomycin C	0.01 µg/ml	LB Lennox	4
MITOMYCINC-0.025	Mitomycin C	0.025 µg/ml	LB Lennox	4
MITOMYCINC-0.05	Mitomycin C	0.05 µg/ml	LB Lennox	4
MMS-0.005PCVV	MMS (Methyl methanesulfonate)	0.005 µg/ml	LB Lennox	6
MMS-0.01PCVV	MMS (Methyl methanesulfonate)	0.01 µg/ml	LB Lennox	6
MMS-0.025PCVV	MMS (Methyl methanesulfonate)	0.025 µg/ml	LB Lennox	6
MMS-0.05PCVV	MMS (Methyl methanesulfonate)	0.05 µg/ml	LB Lennox	6
NACL-0MM	NaCl	0 mM	LB no salt	5
NACL-150MM	NaCl	150 mM	LB no salt	5
NACL-300MM	NaCl	300 mM	LB no salt	5
NACL-75MM	NaCl	75 mM	LB no salt	5

Label	Chemical	Concentration	Media background	Batch
NALIDIXICACID-0.1	Nalidixic acid	0.1 µg/ml	LB Lennox	4
NALIDIXICACID-0.25	Nalidixic acid	0.25 µg/ml	LB Lennox	4
NALIDIXICACID-0.5	Nalidixic acid	0.5 µg/ml	LB Lennox	4
NALIDIXICACID-1.0	Nalidixic acid	1.0 µg/ml	LB Lennox	4
NICKEL-0.1MM	NiCl2	0.1 mM	LB Lennox	7
NICKEL-0.5MM	NiCl2	0.5 mM	LB Lennox	7
NICKEL-1MM	NiCl2	1 mM	LB Lennox	7
NICKEL-2MM	NiCl2	2 mM	LB Lennox	7
NIGERICIN-0.1UM	Nigericin	0.1 µM	LB Lennox	6
NIGERICIN-0.5UM	Nigericin	0.5 µM	LB Lennox	6
NIGERICIN-1.0UM	Nigericin	1 µM	LB Lennox	6
NIGERICIN-5.0UM	Nigericin	5 µM	LB Lennox	6
NISIN-1	Nisin	1 µg/ml	LB Lennox	4
NISIN-10	Nisin	10 µg/ml	LB Lennox	4
NISIN-2	Nisin	2 µg/ml	LB Lennox	4
NISIN-5	Nisin	5 µg/ml	LB Lennox	4
NITROFURANTOIN-0.25	Nitrofurantoin	0.25 µg/ml	LB Lennox	7
NITROFURANTOIN-0.5	Nitrofurantoin	0.5 µg/ml	LB Lennox	7
NITROFURANTOIN-1.0	Nitrofurantoin	1 µg/ml	LB Lennox	3
NITROFURANTOIN-1.0REPEAT	Nitrofurantoin	1 µg/ml	LB Lennox	7
NITROFURANTOIN-10.0	Nitrofurantoin	10 µg/ml	LB Lennox	3
NITROFURANTOIN-2.0	Nitrofurantoin	2 µg/ml	LB Lennox	3
NITROFURANTOIN-2.0REPEAT	Nitrofurantoin	2 µg/ml	LB Lennox	7
NITROFURANTOIN-5.0	Nitrofurantoin	5 µg/ml	LB Lennox	3
NITROFURANTOIN-5.0REPEAT	Nitrofurantoin	5 µg/ml	LB Lennox	7
NONACTIN-0.5	Nonactin	0.5 µg/ml	LB Lennox	4
NONACTIN-1	Nonactin	1.0 µg/ml	LB Lennox	4
NONACTIN-2	Nonactin	2.0 µg/ml	LB Lennox	4
NONACTIN-20	Nonactin	20 µg/ml	LB Lennox	7
NONACTIN-4	Nonactin	4.0 µg/ml	LB Lennox	4
NONACTIN-8	Nonactin	8 µg/ml	LB Lennox	7
NONSPI2-BASIC	NonSPI2	n/a	PCN medium	7
NONSPI2-LOWIRON	NonSPI2 low iron	n/a	PCN medium	7
NONSPI2-LOWMAGNESIUM	NonSPI2 low magnesium	n/a	PCN medium	7
NOREPINEPHRINE-0.5MM	Norepinephrine	0.5 mM	LB Lennox	7
NORFLOXACIN-0.0025	Norfloxacin	0.0025 µg/ml	LB Lennox	3
NORFLOXACIN-0.005	Norfloxacin	0.005 µg/ml	LB Lennox	3
NORFLOXACIN-0.01	Norfloxacin	0.01 µg/ml	LB Lennox	3
NORFLOXACIN-0.02	Norfloxacin	0.02 µg/ml	LB Lennox	3
NOVOBIOCIN-0.5	Novobiocin	0.5 µg/ml	LB Lennox	4
NOVOBIOCIN-1	Novobiocin	1.0 µg/ml	LB Lennox	4
NOVOBIOCIN-10	Novobiocin	10 µg/ml	LB Lennox	7
NOVOBIOCIN-2.5	Novobiocin	2.5 µg/ml	LB Lennox	4
NOVOBIOCIN-20	Novobiocin	20 µg/ml	LB Lennox	7
NOVOBIOCIN-5	Novobiocin	5.0 µg/ml	LB Lennox	4
OXACILLIN-10	Oxacillin	10 µg/ml	LB Lennox	3
OXACILLIN-100	Oxacillin	100 µg/ml	LB Lennox	3
OXACILLIN-50	Oxacillin	50 µg/ml	LB Lennox	3
PARAQUAT-0.5	Paraquat	0.5 µg/ml	LB Lennox	4
PARAQUAT-1.0	Paraquat	1.0 µg/ml	LB Lennox	4
PARAQUAT-2.5	Paraquat	2.5 µg/ml	LB Lennox	4
PARAQUAT-5.0	Paraquat	5.0 µg/ml	LB Lennox	4
PAROMOMYCIN-0.1	Paromomycin	0.1 µg/ml	LB Lennox	7
PAROMOMYCIN-0.25	Paromomycin	0.25 µg/ml	LB Lennox	7
PAROMOMYCIN-0.5	Paromomycin	0.5 µg/ml	LB Lennox	7
PAROMOMYCIN-1	Paromomycin	1.0 µg/ml	LB Lennox	7
PENICILLING-0.05REPEAT	Penicillin G	0.05 µg/ml	LB Lennox	5
PENICILLING-0.1REPEAT	Penicillin G	0.1 µg/ml	LB Lennox	6
PENICILLING-0.25REPEAT	Penicillin G	0.25 µg/ml	LB Lennox	5
PENICILLING-0.5REPEAT	Penicillin G	0.5 µg/ml	LB Lennox	6
PENTAMIDINE-100	Pentamidine	100 µg/ml	LB Lennox	7
PENTAMIDINE-20	Pentamidine	20 µg/ml	LB Lennox	7
PENTAMIDINE-5	Pentamidine	5 µg/ml	LB Lennox	7
PENTAMIDINE-50	Pentamidine	50 µg/ml	LB Lennox	7
PH-10	Basic pH (TAPS)	10 pH	LB Lennox	5
PH-5	Acidic pH (HOMOPIPES)	5 pH	LB Lennox	5
PH-6	Acidic pH (MES)	6 pH	LB Lennox	5
PH-6.5	Acidic pH (MES)	6.5 pH	LB Lennox	6
PH-8	Basic pH (TAPS)	8 pH	LB Lennox	5
PH-8.5	Basic pH (TAPS)	8.5 pH	LB Lennox	6
PH-9	Basic pH (TAPS)	9 pH	LB Lennox	5
PH-9.5	Basic pH (TAPS)	9.5 pH	LB Lennox	6
PHLEOMYCIN-0.01	Phleomycin	0.01 µg/ml	LB Lennox	4
PHLEOMYCIN-0.025	Phleomycin	0.025 µg/ml	LB Lennox	4
PHLEOMYCIN-0.05	Phleomycin	0.05 µg/ml	LB Lennox	4
PHLEOMYCIN-0.1	Phleomycin	0.1 µg/ml	LB Lennox	4
PHOSPHOLIPASEA2-6.67	Phospholipase A2	6.67 µg/ml	LB Lennox	7

Label	Chemical	Concentration	Media background	Batch
PIPERACILLIN-0.05	Piperacillin	0.05 µg/ml	LB Lennox	3
PIPERACILLIN-0.1	Piperacillin	0.1 µg/ml	LB Lennox	3
PIPERACILLIN-0.25	Piperacillin	0.25 µg/ml	LB Lennox	3
PLUMBAGIN-100	Plumbagin	100 µg/ml	LB Lennox	7
PLUMBAGIN-50	Plumbagin	50 µg/ml	LB Lennox	7
PMS-0.01	PMS (Phenazine methosulfate)	0.01 µg/ml	LB Lennox	4
PMS-0.05	PMS (Phenazine methosulfate)	0.05 µg/ml	LB Lennox	4
PMS-0.1	PMS (Phenazine methosulfate)	0.1 µg/ml	LB Lennox	4
PMS-0.5	PMS (Phenazine methosulfate)	0.5 µg/ml	LB Lennox	4
POLYMYXINB-0.025	Polymyxin B	0.025 µg/ml	LB Lennox	2
POLYMYXINB-0.05	Polymyxin B	0.05 µg/ml	LB Lennox	2
POLYMYXINB-0.1	Polymyxin B	0.1 µg/ml	LB Lennox	2
POLYMYXINB-0.25	Polymyxin B	0.25 µg/ml	LB Lennox	2
PROCAINE-0.05MM	Procaine	0.05 mM	LB Lennox	3
PROCAINE-0.1MM	Procaine	0.1 mM	LB Lennox	3
PROCAINE-0.25MM	Procaine	0.25 mM	LB Lennox	3
PROCAINE-0.5MM	Procaine	0.5 mM	LB Lennox	3
PROMETHAZINE-1	Promethazine	1 µg/ml	LB Lennox	7
PROMETHAZINE-10	Promethazine	10 µg/ml	LB Lennox	7
PROMETHAZINE-25	Promethazine	25 µg/ml	LB Lennox	7
PROMETHAZINE-5	Promethazine	5 µg/ml	LB Lennox	7
PROPIDIUMIODIDE-1	Propidium iodide	1 µg/ml	LB Lennox	6
PROPIDIUMIODIDE-10	Propidium iodide	10 µg/ml	LB Lennox	6
PROPIDIUMIODIDE-20	Propidium iodide	20 µg/ml	LB Lennox	6
PROPIDIUMIODIDE-5	Propidium iodide	5 µg/ml	LB Lennox	6
PROPIONATE-10MM	Propionate	10 mM	LB Lennox	7
PROPIONATE-20MM	Propionate	20 mM	LB Lennox	7
PROPIONATE-40MM	Propionate	40 mM	LB Lennox	7
PSEUDOMONICACID-0.01	Pseudomonic acid	0.01 µg/ml	LB Lennox	4
PSEUDOMONICACID-0.025	Pseudomonic acid	0.025 µg/ml	LB Lennox	4
PSEUDOMONICACID-0.05	Pseudomonic acid	0.05 µg/ml	LB Lennox	4
PSEUDOMONICACID-0.1	Pseudomonic acid	0.1 µg/ml	LB Lennox	4
PUROMYCIN-1	Puromycin	1 µg/ml	LB Lennox	4
PUROMYCIN-10	Puromycin	10 µg/ml	LB Lennox	4
PUROMYCIN-2	Puromycin	2 µg/ml	LB Lennox	4
PUROMYCIN-5	Puromycin	5 µg/ml	LB Lennox	4
PVPI-100	PVP-I	100 µg/ml	LB Lennox	7
PVPI-20	PVP-I	20 µg/ml	LB Lennox	7
PVPI-5	PVP-I	5 µg/ml	LB Lennox	7
PVPI-50	PVP-I	50 µg/ml	LB Lennox	7
PYOCYANIN-0.1	Pyocyanin	0.1 µg/ml	LB Lennox	4
PYOCYANIN-0.5	Pyocyanin	0.5 µg/ml	LB Lennox	4
PYOCYANIN-1.0	Pyocyanin	1.0 µg/ml	LB Lennox	4
PYOCYANIN-4.0	Pyocyanin	4.0 µg/ml	LB Lennox	4
RAMOPLANIN-10	Ramoplanin	10 µg/ml	LB Lennox	7
RAMOPLANIN-100	Ramoplanin	100 µg/ml	LB Lennox	7
RAMOPLANIN-200	Ramoplanin	200 µg/ml	LB Lennox	7
RAMOPLANIN-50	Ramoplanin	50 µg/ml	LB Lennox	7
RESERPINE-1	Reserpine	1 µg/ml	LB Lennox	4
RESERPINE-20	Reserpine	20 µg/ml	LB Lennox	4
RESERPINE-5	Reserpine	5 µg/ml	LB Lennox	4
RESERPINE-50	Reserpine	50 µg/ml	LB Lennox	4
RIFAMPICIN-0.5	Rifampicin	0.5 µg/ml	LB Lennox	2
RIFAMPICIN-1.0	Rifampicin	1.0 µg/ml	LB Lennox	2
RIFAMPICIN-2.0	Rifampicin	2.0 µg/ml	LB Lennox	2
RIFAMPICIN-4.0	Rifampicin	4.0 µg/ml	LB Lennox	2
SANGUINARINE-16.67	Sanguinarine	16.67 µg/ml	LB Lennox	7
SDS-0.05PCWV	SDS	0.05 %(w/v)	LB Lennox	5
SDS-0.1PCWV	SDS	0.1 %(w/v)	LB Lennox	6
SDS-0.2PCWV	SDS	0.2 %(w/v)	LB Lennox	5
SDS-0.5PCWV	SDS	0.5 %(w/v)	LB Lennox	6
SDS+EDTA-0.1PCWV+0.1MM	SDS+EDTA	0.1/0.1 %(w/v)/mM	LB Lennox	6
SDS+EDTA-0.1PCWV+0.5MM	SDS+EDTA	0.1/0.5 %(w/v)/mM	LB Lennox	5
SDS+EDTA-0.5PCWV+0.1MM	SDS+EDTA	0.5/0.1 %(w/v)/mM	LB Lennox	5
SDS+EDTA-0.5PCWV+0.5MM	SDS+EDTA	0.5/0.5 %(w/v)/mM	LB Lennox	5
SPECTINOMYCIN-0.1	Spectinomycin	0.25 µg/ml	LB Lennox	5
SPECTINOMYCIN-0.5	Spectinomycin	0.5 µg/ml	LB Lennox	5
SPECTINOMYCIN-1.0	Spectinomycin	1 µg/ml	LB Lennox	5
SPECTINOMYCIN-2.5	Spectinomycin	2.5 µg/ml	LB Lennox	5
SPECTINOMYCIN-3.0	Spectinomycin	3 µg/ml	LB Lennox	7
SPECTINOMYCIN+VANILLIN-3.0+100	Spectinomycin+Vanillin	3+100 µg/ml	LB Lennox	7
SPIRAMYCIN-1	Spiramycin	1 µg/ml	LB Lennox	4
SPIRAMYCIN-10	Spiramycin	10 µg/ml	LB Lennox	4
SPIRAMYCIN-2	Spiramycin	2 µg/ml	LB Lennox	4
SPIRAMYCIN-5	Spiramycin	5 µg/ml	LB Lennox	4
STREPTOMYCIN-0.1	Streptomycin	0.1 µg/ml	LB Lennox	6
STREPTOMYCIN-0.25	Streptomycin	0.25 µg/ml	LB Lennox	6

Label	Chemical	Concentration	Media background	Batch
STREPTOMYCIN-0.5	Streptomycin	0.5 µg/ml	LB Lennox	6
STREPTOMYCIN-1.0	Streptomycin	1.0 µg/ml	LB Lennox	6
SUCROSE-10PCWV	Sucrose	10 %(w/v)	LB no salt	7
SUCROSE-1PCWV	Sucrose	1 %(w/v)	LB no salt	7
SUCROSE-2.5PCWV	Sucrose	2.5 %(w/v)	LB no salt	7
SUCROSE-5PCWV	Sucrose	5 %(w/v)	LB no salt	7
SULFAMONOMETHOXINE-1	Sulfamonomethoxine	1 µg/ml	LB Lennox	4
SULFAMONOMETHOXINE-20	Sulfamonomethoxine	20 µg/ml	LB Lennox	4
SULFAMONOMETHOXINE-5	Sulfamonomethoxine	5 µg/ml	LB Lennox	4
SULFAMONOMETHOXINE-50	Sulfamonomethoxine	50 µg/ml	LB Lennox	4
TBUTYLHYDROPEROXIDE-0.1MM	tert-Butyl hydroperoxide	0.1 mM	LB Lennox	7
TBUTYLHYDROPEROXIDE-0.5MM	tert-Butyl hydroperoxide	0.5 mM	LB Lennox	7
TBUTYLHYDROPEROXIDE-1MM	tert-Butyl hydroperoxide	1 mM	LB Lennox	7
TEICOPLANIN-1	Teicoplanin	1 µg/ml	LB Lennox	4
TEICOPLANIN-10	Teicoplanin	10 µg/ml	LB Lennox	4
TEICOPLANIN-2	Teicoplanin	2 µg/ml	LB Lennox	4
TEICOPLANIN-20	Teicoplanin	20 µg/ml	LB Lennox	7
TEICOPLANIN-5	Teicoplanin	5 µg/ml	LB Lennox	4
TEICOPLANIN-50	Teicoplanin	50 µg/ml	LB Lennox	7
TEMPERATURE-25C	Temperature	25 °C	LB Lennox	4
TEMPERATURE-30C	Temperature	30 °C	LB Lennox	4
TEMPERATURE-37C	Temperature	37 °C	LB Lennox	4
TEMPERATURE-40C	Temperature	40 °C	LB Lennox	5
TEMPERATURE-42C	Temperature	42 °C	LB Lennox	5
TETRACYCLINE-0.05	Tetracycline	0.05 µg/ml	LB Lennox	2
TETRACYCLINE-0.1	Tetracycline	0.1 µg/ml	LB Lennox	2
TETRACYCLINE-0.25	Tetracycline	0.25 µg/ml	LB Lennox	2
TETRACYCLINE-0.5	Tetracycline	0.5 µg/ml	LB Lennox	2
THEOPHYLLINE-1	Theophylline	1 µg/ml	LB Lennox	4
THEOPHYLLINE-100	Theophylline	100 µg/ml	LB Lennox	7
THEOPHYLLINE-20	Theophylline	20 µg/ml	LB Lennox	4
THEOPHYLLINE-200	Theophylline	200 µg/ml	LB Lennox	7
THEOPHYLLINE-5	Theophylline	5 µg/ml	LB Lennox	4
THEOPHYLLINE-50	Theophylline	50 µg/ml	LB Lennox	4
THIOSTREPTON-0.05	Thiostrepton	0.05 µg/ml	LB Lennox	7
THIOSTREPTON-0.1	Thiostrepton	0.1 µg/ml	LB Lennox	7
THIOSTREPTON-0.25	Thiostrepton	0.25 µg/ml	LB Lennox	7
THIOSTREPTON-0.5	Thiostrepton	0.5 µg/ml	LB Lennox	7
TIGECYCLINE-0.05	Tigecycline	0.05 µg/ml	LB Lennox	3
TIGECYCLINE-0.1	Tigecycline	0.1 µg/ml	LB Lennox	3
TIGECYCLINE-0.25	Tigecycline	0.25 µg/ml	LB Lennox	3
TIGECYCLINE-0.5	Tigecycline	0.5 µg/ml	LB Lennox	3
TOBRAMYCIN-0.025	Tobramycin	0.025 µg/ml	LB Lennox	3
TOBRAMYCIN-0.05	Tobramycin	0.05 µg/ml	LB Lennox	3
TOBRAMYCIN-0.075	Tobramycin	0.075 µg/ml	LB Lennox	3
TOBRAMYCIN-0.1	Tobramycin	0.1 µg/ml	LB Lennox	3
TOBRAMYCIN-0.1REPEAT	Tobramycin	0.1 µg/ml	LB Lennox	7
TOBRAMYCIN-0.25	Tobramycin	0.25 µg/ml	LB Lennox	7
TOBRAMYCIN-0.4	Tobramycin	0.4 µg/ml	LB Lennox	7
TOBRAMYCIN-0.5	Tobramycin	0.5 µg/ml	LB Lennox	7
TRICLOSAN-0.005	Triclosan	0.005 µg/ml	LB Lennox	6
TRICLOSAN-0.01	Triclosan	0.01 µg/ml	LB Lennox	6
TRICLOSAN-0.025	Triclosan	0.025 µg/ml	LB Lennox	6
TRICLOSAN-0.05	Triclosan	0.05 µg/ml	LB Lennox	6
TRIMETHOPRIM-0.1	Trimethoprim	0.01 µg/ml	LB Lennox	5
TRIMETHOPRIM-0.25	Trimethoprim	0.025 µg/ml	LB Lennox	5
TRIMETHOPRIM-0.5	Trimethoprim	0.05 µg/ml	LB Lennox	5
TUNICAMYCIN-0.5	Tunicamycin	0.5 µg/ml	LB Lennox	6
TUNICAMYCIN-1	Tunicamycin	1.0 µg/ml	LB Lennox	6
TUNICAMYCIN-3	Tunicamycin	3.0 µg/ml	LB Lennox	6
TUNICAMYCIN-7	Tunicamycin	7.0 µg/ml	LB Lennox	6
TYLOSIN-1	Tylosin	1 µg/ml	LB Lennox	7
TYLOSIN-20	Tylosin	20 µg/ml	LB Lennox	7
TYLOSIN-5	Tylosin	5 µg/ml	LB Lennox	7
TYLOSIN-50	Tylosin	50 µg/ml	LB Lennox	7
UV254NM-12S	UV	12 s	LB Lennox	7
UV254NM-18S	UV	18 s	LB Lennox	7
UV254NM-24S	UV	24 s	LB Lennox	7
UV254NM-6S	UV	6 s	LB Lennox	7
VANCOMYCIN-10	Vancomycin	10 µg/ml	LB Lennox	5
VANCOMYCIN-100	Vancomycin	100 µg/ml	LB Lennox	7
VANCOMYCIN-20	Vancomycin	20 µg/ml	LB Lennox	5
VANCOMYCIN-200	Vancomycin	200 µg/ml	LB Lennox	7
VANCOMYCIN-5	Vancomycin	5 µg/ml	LB Lennox	5
VANCOMYCIN-50	Vancomycin	50 µg/ml	LB Lennox	5
VANILLIN-100	Vanillin	100 µg/ml	LB Lennox	7
VANILLIN-100REPEAT	Vanillin	100 µg/ml	LB Lennox	7

Label	Chemical	Concentration	Media background	Batch
VANILLIN-200	Vanillin	200 µg/ml	LB Lennox	7
VANILLIN-50	Vanillin	50 µg/ml	LB Lennox	7
VERAPAMIL-1	Verapamil	1 µg/ml	LB Lennox	4
VERAPAMIL-20	Verapamil	20 µg/ml	LB Lennox	4
VERAPAMIL-5	Verapamil	5 µg/ml	LB Lennox	4
VERAPAMIL-50	Verapamil	50 µg/ml	LB Lennox	4

C. Shared conditions for interspecies comparison

E. coli	*Salmonella*
M9-GLUCOSAMINE	M9-GLUCOSAMINE
M9-GLUCOSE	M9-GLUCOSE
M9-GLYCEROL	M9-GLYCEROL
M9-MALTOSE	M9-MALTOSE
M9-NACETYLGLUCOSAMINE	M9-NACETYLGLUCOSAMINE
M9-SUCCINATE	M9-SUCCINATE
TEMPERATURE-25C	TEMPERATURE-25C
TEMPERATURE-40C	TEMPERATURE-40C
TEMPERATURE-42C	TEMPERATURE-42C
UV254NM-12S	UV254NM-12S
UV254NM-18S	UV254NM-18S
UV254NM-24S	UV254NM-24S
UV254NM-6S	UV254NM-6S
PH-10	PH-10
PH-5	PH-5
PH-6	PH-6
PH-8	PH-8
PH-9	PH-9
PH-9.5	PH-9.5
A22-0.5	A22-0.1
ACRIFLAVINE-10	ACRIFLAVINE-5
AMIKACIN-0.2	AMIKACIN-0.1
AMOXICILLIN-0.5	AMOXICILLIN-0.1
AMPICILLIN-4.0	AMPICILLIN-0.1
AZITHROMYCIN-1.0	AZITHROMYCIN-0.1
AZTREONAM-0.04	AZTREONAM-0.005
BENZALKONIUM-10	BENZALKONIUM-10
BILESALTS-2PCWV	BILESALTS-1PCWV
BLEOMYCIN-2.0	BLEOMYCIN-0.025
CCCP-2.0	CCCP-2.0
CECROPINB-0.1	CECROPINB-0.1
CEFACLOR-3.0	CEFACLOR-0.1
CEFOXITIN-1.0	CEFOXITIN-0.5
CEFSULODIN-12	CEFSULODIN-5
CEFTAZIDIME-0.05	CEFTAZIDIME-0.02
CERULENIN-2.0	CERULENIN-1.0
CHIR090-0.04	CHIR090-0.02
CHLORPROMAZINE-24UM	CHLORPROMAZINE-2.5UM
CIPROFLOXACIN-0.008	CIPROFLOXACIN-0.001NGML
CLARITHROMYCIN-5.0	CLARITHROMYCIN-0.25
COBALT-0.1MM	COBALT-0.1MM
COPPER-4MM	COPPER-2MM
CYCLOSERINED-16	CYCLOSERINED-5.0
EDTA-1.0MM	EDTA-0.5MM
EGCG-50	EGCG-20
EGTA-2MM	EGTA-1MM
EPINEPHRINE-0.25MM	EPINEPHRINE-0.5MM
ERYTHROMYCIN-0.1	ERYTHROMYCIN-0.5
ETHIDIUMBROMIDE-10	ETHIDIUMBROMIDE-5
FOSFOMYCIN-1.0	FOSFOMYCIN-0.05
FUSIDICACID-50	FUSIDICACID-50
GENTAMICIN-0.1	GENTAMICIN-0.025
HYDROXYUREA-10	HYDROXYUREA-2.5
ISONIAZID-1.0	ISONIAZID-20
LEVOFLOXACIN-0.002	LEVOFLOXACIN-0.005
MECILLINAM-0.03	MECILLINAM-0.01
MINOCYCLINE-1.0	MINOCYCLINE-0.5
MITOMYCINC-0.1	MITOMYCINC-0.05

MMS-0.05PCVV	MMS-0.01PCVV
NACL-600MM	NACL-300MM
NALIDIXICACID-2.0	NALIDIXICACID-1.0
NICKEL-1MM	NICKEL-2MM
NIGERICIN-0.1UM	NIGERICIN-0.1UM
NITROFURANTOIN-2.0	NITROFURANTOIN-2.0
NOREPINEPHRINE-1.0MM	NOREPINEPHRINE-0.5MM
NORFLOXACIN-0.02	NORFLOXACIN-0.005
NOVOBIOCIN-8	NOVOBIOCIN-5
OXACILLIN-0.5	OXACILLIN-50
PARAQUAT-5.0	PARAQUAT-1.0
PHLEOMYCIN-1.0	PHLEOMYCIN-0.1
PMS-0.05	PMS-0.5
POLYMYXINB-6.0	POLYMYXINB-0.1
PROCAINE-10	PROCAINE-0.5MM
PROPIDIUMIODIDE-20	PROPIDIUMIODIDE-20
PUROMYCIN-5	PUROMYCIN-5
PYOCYANIN-1.0	PYOCYANIN-0.5
RIFAMPICIN-2.0	RIFAMPICIN-1.0
SDS-4.0PCWV	SDS-0.5PCWV
SPECTINOMYCIN-6.0	SPECTINOMYCIN-3.0
SPIRAMYCIN-1	SPIRAMYCIN-10
STREPTOMYCIN-0.05	STREPTOMYCIN-0.5
SULFAMONOMETHOXINE-50	SULFAMONOMETHOXINE-1
TETRACYCLINE-0.5	TETRACYCLINE-0.1
THEOPHYLLINE-10	THEOPHYLLINE-1
TOBRAMYCIN-0.05	TOBRAMYCIN-0.05
TRIMETHOPRIM-0.4	TRIMETHOPRIM-0.1
TUNICAMYCIN-1	TUNICAMYCIN-0.5
VANCOMYCIN-50	VANCOMYCIN-100
VERAPAMIL-1	VERAPAMIL-50