Title
Vulnerability and robustness in the essential gene complement of two bacterial species, profiled with CRISPRi

Author
Silvis, Melanie

Publication Date
2020

Vulnerability and robustness in the essential gene complement of two bacterial species, profiled with CRISPRi

by
Melanie Silvis

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Biochemistry and Molecular Biology

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Carol Gross
7AFBE9046E0744C...

Carol Gross
Chair

DocuSigned by:
Hana El-Samad
D7...4DC...

Hana El-Samad

DocuSigned by:
Alexander Johnson
2...745B...

Alexander Johnson

DocuSigned by:
Oren Rosenberg
97...4ED...

Oren Rosenberg

DocuSigned by:
JOSEPH BONDY-DENOMY
0B105738279D4A2...

JOSEPH BONDY-DENOMY

Committee Members

DEDICATION AND ACKNOWLEDGEMENTS

This work has been the culmination of many years of effort, on my part but also on the part of my many mentors and collaborators. I benefitted very early in my scientific career from the faith and curiosity of many classmates, teachers, mentors, and bosses, without which I would have never thought to seriously pursue applying to graduate school much less becoming a scientist. Throughout graduate school I was lucky enough to have been afforded control over my time, which allowed me to pursue independent projects with many exciting scientists within and nearby UCSF, through which I gained access to many interesting biological fields and different ways of approaching complex problems. To synthesize the advice and opinions of so many scientists has been its own challenge and privilege, and I learned so much in the process.

I was lucky to have been raised by very stubborn women, including my mother and grandmother, to whom education represented both a practical and symbolic asset to the family. Valuing my education got me part of the way through this PhD, but undoubtedly it was the stubbornness that allowed me to finish. To all my families that I've accumulated over the years, this one is for you.

CONTRIBUTIONS

Chapter 1 is reprinted as a manuscript in preparation:

Silvis, M.R., Gross, C.A. CRISPR screens for gene function.

Chapter 2 is reprinted as a manuscript in preparation:

Silvis, M.R., Rajendram, M., Shi, H., Osadnik, H., Gray, A., Peters, J.M., Hearne, C.C., Kumar, P., Huang, K.C., Gross, C.A. Functional analysis of *E. coli* essential genes using CRISPRi

MRS, MR, KCH and CAG primarily conceived and designed the study; MRS, MR, HS, HO, CCH, PK performed experiments; MRS built software; MRS, MR, HS, HO analyzed data; MRS, MR, KCH, CAG wrote and/or edited the manuscript.

Chapter 3 is reprinted as a manuscript in preparation:

Hawkins, J.S.*, Silvis, M.R.*, Koo, B.-M., Peters, J.M., Jost, M., Hearne, C.C., Weissman, J.S., Todor, H., Gross, C.A. Mismatch-CRISPRi reveals conserved expression-fitness relationships in bacteria

JSH, MRS, BMK, JMP, HT and CAG designed the study; JSH, MRS, BMK, and CCH performed experiments; JSH, MRS, HT, MJ built software; JSH, MRS, BMK, HT analyzed data; JSH, MRS, BMK, JMP, HT, MJ, JSW and CAG wrote and/or edited the manuscript.

*JSH and MRS contributed equally to this work

Vulnerability and robustness in the essential gene complement of two bacterial species, profiled with CRISPRi

Melanie R. Silvis

ABSTRACT

Bacterial essential genes contribute to the most fundamental processes of cellular life. The study of their functions *in vivo* has long been intractable to systematic genetic approaches, which are fundamental to understanding pathway level connections that govern cellular life and are a requirement for dissecting the complex cellular processes to which essential genes contribute. In Chapter 1 of this work I review recent advances in mapping gene-phenotype relationships in bacteria using the CRISPR-based technology, CRISPR interference (CRISPRi) for titratable gene knockdowns, focusing on their applications to the studies of essential genes, the exploration of chemical-genetic interactions, and the prospects for disentangling complex phenotypes in diverse bacterial species. In Chapter 2 I describe my analysis of the essential gene functions in the model Gram-negative bacterium *Escherichia coli* and the model Gram-positive *Bacillus subtilis* using datasets from paired chemical-genetic screens. In this work I identify both shared and Gram-negative specific mechanisms of collateral sensitization to antibiotic action. In Chapter 3 I investigate a fundamental property of essential genes, which is the relationship between their expression level and the cellular growth rate. Here, further developing CRISPRi tools in bacteria to predictably titrate knockdown efficacy, I interpret the knockdown-fitness relationships of each essential gene in *E. coli* and *B. subtilis*, discovering broad conservation of constraints setting and maintaining expression levels across these diverged species.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

INTRODUCTION

Bacteria must protect their genetic material from invasive DNA elements such as phages and plasmids, and they accomplish this using diverse immune systems (Bernheim and Sorek 2019) that are tasked with both recognition and response to these molecularly simple foreign agents. The mechanisms of these two activities—recognition and the protective responses—have provided a bountiful source of inspiration for technological development with applications in diverse areas including basic research, diagnostics, and therapeutics (Knott and Doudna 2018; Pickar-Oliver and Gersbach 2019; C. H. Huang, Lee, and Doudna 2018; Y. Li et al. 2019). CRISPR immunity systems, arguably the category of bacterial immune systems with the most penetrance in recent technological development, couple programmable sequence-specific recognition with specified nuclease activities, using either multi-protein or single-effector complexes (Hille et al. 2018). Importantly, these systems are highly modular, enabling them to be easily ported into different contexts while retaining their targeting and nuclease activities, and are often amenable to combination with other enzymes or mutation in order to expand their functionalities.

A major driver of CRISPR-based technology development is their application in functional genomics: coupling genome-scale genetic perturbations with high-throughput phenotypic assays to systematically define gene-phenotype relationships. Microbial genomes are a major source of new gene content and functionalities. As the pace of bacterial whole-genome sequencing increases, higher-throughput functional assays must be developed to link the identification of novel genes and their products with their cellular roles. CRISPR-based tools are poised to bridge the gap between genotype and phenotype, and this review will focus on CRISPR interference (CRISPRi) approaches for bacterial gene knockdown. CRISPRi offers three specific features that will enable rapid genotype-phenotype associations: (1) it enables rapid construction of genome-scale barcoded libraries of pre-programmed gene knockdowns, (2) CRISPRi can reduce gene expression to intermediate levels (i.e. not “all or nothing”),

allowing the investigation of intermediate phenotypes with specific implications for the essential genes, and (3) CRISPRi tools have now been demonstrated to function in diverse microbial species, allowing gene functions to be studied across large evolutionary timespans.

In this review we will discuss recent developments in the application of CRISPRi to generate libraries of unique genetic perturbations in bacterial systems of interest, and their applications in characterizing gene function. In the process, we will highlight both technological and conceptual advances in functional genomics that are afforded by these new scalable technologies.

CRISPR interference (CRISPRi) for targeted gene knockdown in bacteria

The twin roles of CRISPR immunity systems—recognition and nuclease activity—are productively uncoupled in CRISPR interference (CRISPRi) approaches, that use nuclease-inactive effector proteins which retain the ability to stably bind their target DNA, enabling their use as programmable transcriptional repressors. In this section we will review the initially described features of bacterial CRISPRi and how recent comprehensive analyses have provided a more detailed picture that will influence how large-scale functional studies are designed and interpreted.

To date, the majority of bacterial CRISPRi applications have built on early work using a catalytically inactive mutant of the Type II-A Cas9 system from *Streptococcus pyogenes* (dCas9^{Spy}) for programmable repression in *Escherichia coli* (Qi et al. 2013; Bikard et al. 2013). Natural Type II systems consist of a single effector protein (Cas9) which is targeted to DNA by the sequence of the spacer region in a two-RNA duplex. In these first concurrent demonstrations of bacterial CRISPRi, the two-RNA duplex was supplied using either the natural configuration of crRNA:tracrRNA (Bikard et al. 2013) or using a simplified single chimeric sgRNA (Qi et al. 2013), in either case the spacer region being easily modified to determine the targeted locus. Binding of the dCas9-sgRNA complex to DNA involves a two-part recognition

mechanism, one carried out by each member of the complex (Sternberg et al. 2014; Szczelkun et al. 2014). First, dCas9 recognizes a short, double-stranded DNA sequence called the protospacer adjacent motif (PAM) which is identified by 1D diffusion scanning along DNA stretches. Second, the DNA duplex adjacent to the PAM is unwound as a DNA:RNA hybrid is formed between the sgRNA spacer region and the target DNA strand. Once the dCas9-sgRNA-DNA complex and 20bp R-loop is formed it is extremely stable, exhibiting off-rates on the order of ~6hr in mammalian cells (Richardson et al. 2016) and the complex is hypothesized to only undergo dissociation events in bacterial systems following DNA replication cycles (Jones et al. 2017).

In the initial work establishing bacterial CRISPRi, several important properties of CRISPRi were characterized that have, in most cases, been borne out in subsequent comprehensive studies. Each of these properties influence the design and interpretation of CRISPRi libraries for use in functional genomics.

First, the initial work outlined two mechanisms by which CRISPRi can function in bacteria. When targeted to promoter regions, dCas9-sgRNA complexes prevent RNAP recognition and therefore transcription initiation. Alternatively, when targeted within gene open reading frames (ORFs), dCas9-sgRNA binding is sufficient to prevent RNAP elongation, as assayed by NET-seq (Churchman and Weissman 2011) in now two species (Qi et al. 2013; Peters et al. 2016). The latter mode is more commonly employed in bacterial CRISPRi screens to date, and has the advantages of not requiring TSS annotation (Lee et al. 2019) and having increased target space within ORFs as opposed to promoter regions.

Efficacy of knockdown when targeted within ORFs shows clear strandedness. sgRNAs designed to base pair with the non-template (coding) strand are more effective than those designed to base pair with the template (non-coding) strand, and this has been repeatedly shown across species (Peters et al. 2016; Lee et al. 2019; Cui et al. 2018). The molecular mechanism explaining this strandedness is not known, however recent *in vitro* work has shown

that template-strand targeted dCas9-sgRNAs are more easily surpassed by both bacterial multi-subunit RNAPs and by phage single-subunit RNAPs than non-template targeted complexes (Widom et al. 2019), suggesting this pass-through property is common across diverse RNAPs. Again, most functional genomics approaches using CRISPRi in bacterial systems now exclusively use non-template targeting sgRNAs, though some benefits to exploring weak perturbations using template-targeting sgRNAs have been discussed (Rousset et al. 2018).

Although early work with small numbers of sgRNAs suggested that efficacy of repression by targeting dCas9-sgRNA complexes within ORFs was location-dependent—targeting 5' within the ORF having a stronger effect than targeting towards the 3' end (Qi et al. 2013)—more comprehensive analysis with larger numbers of sgRNAs has found no significant difference in repression, by assaying either expression (protein level) (Hawkins et al. 2019) or indirectly by assaying fitness of essential gene knockdowns (Wang et al. 2018). The relatively consistency of effects from targeting different loci within the same gene (Hawkins et al. 2019) have lent support to analytical approaches that use a synthesis of the measurements from multiple sgRNAs targeting the same gene (discussed further below).

Bacterial genes are often organized into clusters (operons) that are co-transcribed, and because CRISPRi acts by interfering with transcription elongation the potential to interfere with the expression of neighboring genes is a central concern that continues to be explored. Early work conceptualized these effects in two distinct categories: the expression of genes downstream of the targeted gene was anticipated to be reduced as elongating RNAP were stalled before reaching their ORFs (CRISPRi "forward polarity"), and by an incompletely understood mechanism genes upstream of the targeted gene were frequently observed to be repressed (CRISPRi "reverse polarity"). Both forward and reverse polarity were shown to function in *B. subtilis* using a synthetic *rfp-gfp* operon, with forward polarity having a stronger effect than reverse polarity (Peters et al. 2016). These data support the conclusion that bacterial CRISPRi can provide operon-level functional classification.

Both forward and reverse polarity have been challenging to employ as assumptions in functional analysis. First, forward polar effects from targeting non-essential genes upstream of essential genes have comprised a relatively low percentage of false positives in essential gene assignments using CRISPRi in *E. coli* (Wang et al. 2018; Rousset et al. 2018), highlighting incomplete knowledge of transcription unit organization even in well-studied species. Second, comprehensive descriptions of reverse polar effects have been similarly case-by-case, with a more general position-dependent effect when targeting within 50bp of the ends of upstream genes, and a small number of cases seeming to cause reverse polarity for targeting at any position (Wang et al. 2018).

Almost certainly there are multiple layers occluding the underlying phenomenon in these reports. The first is that it has been traditionally easiest to measure the effects of targeting up- or down-stream of essential genes on fitness as a proxy for polar effects on transcription, or by using synthetic operons of fluorescent proteins. Each has caveats that may prevent generalization. Furthermore, while the (or any) mechanism of reverse polarity is incompletely understood, species-specific effects remain a likely source of discrepant results.

A third property suggested by early bacterial CRISPRi applications is that targeting in bacterial genomes was highly specific. Using RNA-seq, sgRNAs targeting a non-endogenous gene (*rfp*) have been shown to not interfere with endogenous gene expression (Qi et al. 2013; Peters et al. 2016; Qu et al. 2019), arguing that targeting specificity is theoretically high in relatively small bacterial genomes that have limited off-target sequence potential. At odds with these demonstrations, evidence in support of relaxed targeting constraints was also explored early on, suggesting that targeting required the PAM sequence and at least 12bp of PAM-proximal sequence (Qi et al. 2013; Bikard et al. 2013). More recent phenotypic screens have further characterized the impact of these relaxed targeting constraints on phenotype assignment, identifying multiple cases where off-target effects are a source of false positives, for example in identification of genes contributing to fitness (Rousset et al. 2018; Cui et al. 2018).

These studies have characterized 9 or 11bp of sequence identity to promoter targeting or non-template strand targeting (respectively) as sufficient to interfere with expression at an off-target site. These findings highlight the requirement of multiple sgRNAs per target gene in functional screens to avoid false positives, as targeting specificity is not ensured by the uniqueness of 23bp (spacer + PAM) in the genome.

Reconceptualizing these relaxed targeting requirements as an asset to CRISPRi approaches, modifications to sgRNA spacers that impact targeting efficacy are emerging as key levers by which to predictably tune knockdown and to therefore explore intermediate phenotypes. The majority of bacterial CRISPRi approaches have controlled the cellular concentrations of dCas9-sgRNA complex to modulate knockdown efficacy, usually by regulating the expression of either or both complex members using defined inducible promoters (Qi et al. 2013; Peters et al. 2016; X. Li et al. 2016; Cui et al. 2018; Rock et al. 2017). This has enabled, for example, the exploration of fitness-promoting genes by constructing CRISPRi libraries in permissive conditions (dCas9 or sgRNA is not expressed) and experimentation in non-permissive conditions (dCas9 or sgRNA expression is maximally induced). However, as detailed above, perfect complementarity between the 20bp spacer region of the sgRNA and the target DNA is not required for gene knockdown, as was appreciated early in the discovery of Type II systems (Jinek et al. 2012), and early evidence suggested that intermediate knockdown effects could be achieved by deliberately introducing mismatched bases into spacers at precise locations.

The underlying logic of the impacts on dCas9-sgRNA binding by single mismatches have now been evaluated in a number of assays, with results in general agreement with a few bacterial CRISPRi-specific lessons. A comprehensive analysis of all single-mismatch variants targeting *gfp* in two bacterial species in a CRISPRi assay has shown that the impacts of single-mismatches can be reliably and precisely predicted by a simple model using only sequence features of the spacer region (Hawkins et al. 2019). This comprehensive dataset captured the

full range of intermediate efficacies and found that—in agreement with early reports (Qi et al. 2013)—mismatches in the seed (PAM-adjacent) region had the largest impacts on efficacy, but that even PAM-distal single mismatch variants retained measurable activity. Two recent large studies of singly mismatched sgRNAs are in general agreement with the results from this analysis, highlighting the universal nature of the CRISPRi mechanism. A large dataset of dCas9-sgRNA association rates *in vitro* to diverse DNA substrates (Boyle et al. 2017) showed the same seed-region sensitivity to mismatches, but found that sgRNAs were completely tolerant of mismatches at the very distal end of the spacer (positions 19-20) in this assay. Similarly, in a recent mammalian CRISPRi study, singly mismatched sgRNAs were used to repress essential genes and achieve intermediate growth rates (Jost et al. 2019). This dataset—relating many sequence features of the sgRNA spacer and mismatch identity to growth rate—was used to develop an elastic net linear regression model capturing a majority of the variance in the dataset (r^2=0.52 between predicted and measured growth rates). Again, this model describes a system in which mismatches at the distal sgRNA end are completely tolerated, in agreement with the *in vitro* association rate measurements. One interpretation of these data is that (1) bacterial systems are a very sensitive assay for dCas9-sgRNA binding, in which very slow association rates can be captured, perhaps because the effects are amplified by multiple dCas9-sgRNA-DNA interactions per cell, and (2) that mechanistic differences in the mammalian CRISPRi systems which use KRAB fusions to recruit stable, silencing chromatin modifications again reduce sensitivity to small changes in sgRNA efficacy. If true, bacterial CRISPRi systems may represent a novel assay for fundamental properties of dCas9-sgRNA-DNA interactions, for example in how anti-CRISPR proteins interact with them to prevent binding (Rauch et al. 2017).

The potential to predictably tune knockdown via the deliberate introduction of single mismatches into sgRNAs opens up many experimental opportunities to systematically characterize gene function. Because mismatches are introduced into the spacer which can be used as a strain barcode in next-generation sequencing, multiple intermediate levels of target

gene knockdown can be assessed within the same pool and the same experiment. Compared to titration of knockdown by inducing dCas9-sgRNA complex concentration, some experimental and theoretical work has made the case that tuning knockdown efficacy in the latter way is prone to noise and dependence on native expression levels of the targeted genes (Vigouroux et al. 2018). Instead, the authors provide evidence that tuning knockdown using sgRNAs with mismatches between their spacers and the target DNA experiences less noise. While the proposed mechanism of this distinction—that RNAP dislodges dCas9-sgRNA complexes when sgRNAs are mismatched—remains to be fully supported, *in vitro* it has been shown that *E. coli* RNAP passes through even fully matched dCas9-sgRNA complexes on DNA ~30% of the time and does not dissociate when it reaches the blockade (Widom et al. 2019). This is consistent with persistent attempts to pass through even fully matched dCas9-sgRNA-DNA complexes, possibly suggested by the observed NET-seq signals of pausing (Qi et al. 2013; Peters et al. 2016).

Finally, multiple reports have noted the unsuitability of specific CRISPRi system variants for certain bacterial species and characterized some important issues that should be incorporated into the design and analysis of CRISPRi screens for gene function. Toxicity caused by dCas9 expression has now been identified by multiple groups (Qu et al. 2019; Rock et al. 2017) as an sgRNA-independent phenomenon, although the mechanism remains unclear. In some cases, toxicity is specifically related to the dCas9 variant used (Rock et al. 2017) and manifests as generalized sensitivity to stresses. In other cases, toxicity is a function of expression level (Qu et al. 2019). High levels of expression of dCas9-sgRNA complexes have also been shown to have sequence-specific toxicities in *E. coli* (Cui et al. 2018), suggesting that it is not the translation load alone that can interfere with normal cell physiology. Users of CRISPRi to investigate genotype-phenotype relationships must clearly avoid and control for possible non-specific phenotypes, which can be addressed, for example, by evaluating large numbers of non-targeting sgRNAs.

CRISPRi and assays for gene function

CRISPRi has to date been used to explore bacterial gene function in two primary modes, and almost exclusively in model species. In the first mode, relatively small arrayed CRISPRi libraries (<500 sgRNAs) are constructed and deeply phenotyped using relatively low throughput quantitative assays. This has been optimally applied for “systems biology approaches”: simultaneously capturing the networks of functional interactions between targeted genes by quantifying large numbers of independent phenotypes. In the second type of approach, advances in parallel oligo synthesis strategies enable the construction of large pooled libraries (>30,000 sgRNAs) which are phenotyped in a more limited manner. These works have been more in line with “functional genomics”: addressing targeted questions about the contributions of individual genes to a specific process, usually one that is linked to growth or fitness. Forward-looking strategies will likely capitalize on the benefits afforded by each (including development of bespoke phenotypic assays for pathways of interest), aim to increase the throughput for deep phenotyping of larger libraries, and extend CRISPRi screens for gene function to diverse species.

Arrayed CRISPRi approaches to identify gene function

Chemical-genetic approaches query large numbers of genetic perturbations for their ability to influence sensitivity or resistance to small molecule inhibitors of growth (reviewed in (Cacace, Kritikos, and Typas 2017)). By profiling growth quantitatively across large numbers of chemical conditions, these “phenotypic signatures” have provided an important means to identify functional interactions between genes and even to identify functions for uncharacterized genes using “guilt by association” logic. In bacteria and in yeast, chemical-genetic screening technologies have embraced arrayed library formats, enabling collection of chemical phenotypes of bacterial genome-wide deletion libraries (Nichols et al. 2011; Shiver et al. 2016; Kritikos et al. 2017) and heterozygous and homozygous yeast deletion libraries (Hillenmeyer et

al. 2008). Importantly, the ability to screen yeast genes in a haploid state (heterozygous gene deletions) extended the surveyable genetic space to essential genes—genes for which a cell cannot tolerate the loss of both copies. The profiling of bacterial essential genes has now recently moved into the same theoretical domain, using CRISPRi libraries that partially reduce essential genes' expression. This has enabled the first characterization of a bacterial essential gene interaction network in *B. subtilis*, importantly using phenotypic signatures to link one essential gene of unknown function (*ylaN*) to known pathways in iron-sulfur cluster biogenesis (Peters et al. 2016).

Chemical-genetic profiling is reliant on the number, diversity, and specificity of conditions available in which to screen for phenotypes in order to resolve the functions of cellular pathways. Therefore, the use of small molecule antibiotics has been primary. Importantly, there is a secondary outcome of comprehensively screening for the determinants of antibiotic resistances, which is a more complete picture of the intrinsic and acquirable mechanisms of resistance that bacteria possess (Shiver et al. 2016). The ability to screen essential genes for their contribution to these mechanisms may be profound, as essential genes are usually the direct targets of drugs (Peters et al. 2016) and are largely conserved across diverse species (Koo et al. 2017; Grazziotin, Vidal, and Venancio 2015). Datasets of chemical-genetic interactions for bacterial essential genes may also support predictions of synergistic antibiotic combinations, which have been found to be rare (Brochado et al. 2018). As more species are screened and phenotyped in this manner, the generalizability of chemical-genetic interactions—and therefore antibiotic resistance and sensitization mechanisms—can be assessed.

Morphological variation is another source of complex and quantifiable phenotypes with which to understand gene function (K. C. Huang 2015; Campos et al. 2018). Multiple points can be made to support the value of capturing morphological phenotypes from arrayed CRISPRi libraries: first, cell morphology is deeply tied to fitness and takes inputs from diverse cellular processes; second, cell to cell variability is a meaningful morphological property; third, the

dynamics of morphological change over time are informative; and finally, capturing strain phenotypes individually can also support more fine-grained analysis including protein localization (Kuwada, Traxler, and Wiggins 2015) and subcellular properties such as nucleoid structure (Nonejuie et al. 2013).

CRISPRi libraries specifically can aid in these approaches by offering the ability to titrate gene knockdown. Whereas strong perturbations might converge on one catastrophic morphological outcome, slight perturbations may provide more diverse outcomes related to specific gene function. Strong knockdown of essential genes, on the other hand, has also been useful in ascribing general gene function (Peters et al. 2016; Veening and Liu 2017). The microscopy hardware, automated image capture software, and analysis tools have all made great strides towards accommodating the capture of many strain phenotypes in parallel (Ursell et al. 2017; Shi, Colavin, Lee, et al. 2017; Campos et al. 2018).

The future prospects for high-content phenotype capture for arrayed bacterial CRISPRi libraries are exciting. In some cases, it will remain easier to design and create arrayed CRISPRi libraries than other forms of libraries used for reverse genetic screens, in particular where essential gene phenotypes are important. Some complex phenotypes will be best explored using assays of separate strains, in particular those phenotypes for which multiplexing is not possible (Fuhrer 2016; Breinig et al. 2015), where cell-to-cell heterogeneity is important to quantify, or where cross-complementation between different genetic backgrounds is an issue. Because arrayed libraries are intrinsically barcoded by their sgRNA spacers, arrayed libraries can be pooled before assaying if barcode counting by next-generation sequencing is available. This approach ensures an even distribution of each mutant, which can aid in sensitivity for enrichment or depletion assays (Peters et al. 2019). Arrayed libraries might also provide a helpful starting position for the investigation of genetic interactions.

Genome-wide, pooled CRISPRi libraries to characterize specific phenotypes

In the second mode of CRISPRi screens for gene function, pooled library construction achieves high complexity, allowing comprehensive analysis of many genes with high statistical power. This has already been applied to the identification of genes that promote a fundamental bacterial property: growth in rich media. While these examples of genome wide tiling CRISPRi libraries in *E. coli* (Wang et al. 2018; Cui et al. 2018) have explored the sources of false positives in depletion based screens, their use in an understudied Vibrio species (Lee et al. 2019) has clearly demonstrated their value in identifying new biology.

The contributions of genes to growth can themselves be dissected further in these pooled assays. For example, titrating knockdown of essential genes using single-mismatch sgRNAs has allowed the characterization of “knockdown-fitness curves” representing the relationship between attempted knockdown and growth rate (Hawkins et al. 2019). This relationship encompasses the ways in which bacteria may be robust to CRISPRi knockdown—expression of essential genes in excess of what they’re required to support growth, or feedback on expression to maintain optimal levels (Rousset et al. 2018)—and can identify points of vulnerability, where no robustness to knockdown is observed. These features feed into the reconceptualization of essentiality as a quantitative trait with many facets, including evolvability and environmental dependence (Rancati et al. 2018). Large, complex libraries also support the evaluation of essentiality with respect to genetic and environmental contexts, for example by screening a common pool of sgRNAs in differing genetic backgrounds, or by testing depletion in different conditions. The use of simple enrichment or depletion screens to identify genes contributing to a specific process of interest is already being explored. For example, tolerance to the presence of industrial chemicals is a desirable feature for bioproduction (Wang et al. 2018), or in an interesting dissection of a more complex phenotype, genes supporting the infection and reproduction of diverse phage (Rousset et al. 2018).

With regards to large pooled CRISPRi libraries, a few technical points continue to arise and have been addressed using distinct methodologies from different groups. First and foremost is the quantification of per-strain fitness in optimal growth conditions. Following work in mammalian systems with CRISPR/CRISPRi libraries, some recent studies (Hawkins et al. 2019; Wang et al. 2018) have favored the use of growth metrics that use steady-state growth assumptions to calculate effective doubling time relative to an estimate of wild-type doubling time (Gilbert et al. 2014; Kampmann, Bassik, and Weissman 2013; Jost et al. 2019). This approach requires accurate estimation of wild-type doubling, which can be done using large numbers of unique non-targeting sgRNAs and provides an effective means of distinguishing slight growth phenotypes from noise. It also requires, however, that steady-state growth assumptions are reasonable given the experimental design, for example by maintaining exponential growth through back-dilution. In other cases, the experimental question does not rely on differentiating intermediate phenotypes, and rather seeks to categorize genes as (for example) growth promoting or not. In these cases, recent work has avoided the use of non-targeting control sgRNAs and instead estimated significance using the statistical framework from common RNA-seq approaches implemented in DESeq2 (Love, Huber, and Anders 2014) or using approaches to maximize sensitivity to small effect sizes using maximum likelihood estimates and combining the information from multiple sgRNAs targeting the same gene (Lee et al. 2019; W. Li et al. 2014).

In the future, we anticipate the development of further assays for combination with CRISPRi libraries in a few specific directions. First, screens for complex phenotypes that support follow up screens for each step of the biological process of interest. Second, the use of fitness-independent enrichments and depletions. FACS-based separation of complex libraries using differences in fluorescence (Hawkins et al. 2019) or even morphological characteristics like single-cell width (Shi, Colavin, Bigos, et al. 2017) is a facile method well-suited to genome-scale questions. Fluorescence can be tied to pathway activation (i.e. signaling, stress response

activation, differentiation processes) for a genome scale library. Finally, the growing application of CRISPRi tools to diverse bacterial species (Peters et al. 2019) will provide access to more complex and relevant phenotypes from non-model species, including pathogenesis (Qu et al. 2019), but also provide an exciting new lens with which to view gene function. The cross-species comparison of gene function, in particular for largely conserved sets of genes such as the essential genes, using directly comparable CRISPRi tools, will provide an unprecedented view of the fundamental requirements of bacterial cell physiology and where adaptation has allowed new properties to emerge.

REFERENCES

Bernheim, Aude, and Rotem Sorek. 2019. “The Pan-Immune System of Bacteria: Antiviral Defence as a Community Resource.” *Nature Reviews Microbiology*. https://doi.org/10.1038/s41579-019-0278-2.

Bikard, David, Wenyan Jiang, Poulami Samai, Ann Hochschild, Feng Zhang, and Luciano A Marraffini. 2013. “Programmable Repression and Activation of Bacterial Gene Expression Using an Engineered CRISPR-Cas System.” *Nucleic Acids Research* 41 (15). https://doi.org/10.1093/nar/gkt520.

Boyle, Evan A., Johan O.L. Andreasson, Lauren M. Chircus, Samuel H. Sternberg, Michelle J. Wu, Chantal K. Guegler, Jennifer A. Doudna, and William J. Greenleaf. 2017. “High-Throughput Biochemical Profiling Reveals Sequence Determinants of DCas9 off-Target Binding and Unbinding.” *Proceedings of the National Academy of Sciences of the United States of America* 114 (21): 5461–66. https://doi.org/10.1073/pnas.1700557114.

Breinig, Marco, Felix A Klein, Wolfgang Huber, and Michael Boutros. 2015. “A Chemical–Genetic Interaction Map of Small Molecules Using High-throughput Imaging in Cancer Cells.” *Molecular Systems Biology* 11 (12): 846. https://doi.org/10.15252/msb.20156400.

Brochado, Ana Rita, Anja Telzerow, Jacob Bobonis, Manuel Banzhaf, André Mateus, Joel Selkrig, Emily Huth, et al. 2018. “Species-Specific Activity of Antibacterial Drug Combinations.” *Nature* 559 (7713): 259–63. https://doi.org/10.1038/s41586-018-0278-9.

Cacace, Elisabetta, George Kritikos, and Athanasios Typas. 2017. “Chemical Genetics in Drug Discovery.” *Current Opinion in Systems Biology* 4: 35–42. https://doi.org/10.1016/j.coisb.2017.05.020.

Campos, Manuel, Sander K Govers, Irnov Irnov, Genevieve S Dobihal, François Cornet, and Christine Jacobs-Wagner. 2018. “Genomewide Phenotypic Analysis of Growth, Cell Morphogenesis, and Cell Cycle Events in Escherichia Coli.” *Molecular Systems Biology* 14 (6): e7573. https://doi.org/10.15252/msb.20177573.

Churchman, L Stirling, and Jonathan S Weissman. 2011. “Nascent Transcript Sequencing Visualizes Transcription at Nucleotide Resolution.” *Nature* 469 (7330): 368–73. https://doi.org/10.1038/nature09652.

Cui, Lun, Antoine Vigouroux, François Rousset, Hugo Varet, Varun Khanna, and David Bikard. 2018. “A CRISPRi Screen in E. Coli Reveals Sequence-Specific Toxicity of DCas9.” *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-04209-5.

Fuhrer, Tobias. 2016. “Genomewide Landscape of Gene-Metabolome Associations in Escherichia Coli.” *Mol Syst Biol* 12 (888): 1–14. https://doi.org/10.15252/msb.

Gilbert, Luke A, Max A Horlbeck, Britt Adamson, Jacqueline E Villalta, Yuwen Chen, Evan H Whitehead, Carla Guimaraes, et al. 2014. “Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation.” *Cell* 159: 647–61. https://doi.org/10.1016/j.cell.2014.09.029.

Grazziotin, Ana Laura, Newton M. Vidal, and Thiago M. Venancio. 2015. “Uncovering Major Genomic Features of Essential Genes in Bacteria and a Methanogenic Archaea.” *FEBS Journal* 282 (17): 3395–3411. https://doi.org/10.1111/febs.13350.

Hawkins, John S, Melanie R Silvis, Byoung-Mo Koo, Jason M Peters, Marco Jost, Cameron C Hearne, Jonathan S Weissman, Horia Todor, and Carol A Gross. 2019. “Modulated Efficacy CRISPRi Reveals Evolutionary Conservation of Essential Gene Expression-Fitness Relationships in Bacteria.” *BioRxiv*, 805333. https://doi.org/10.1101/805333.

Hille, Frank, Hagen Richter, Shi Pey Wong, Majda Bratovi, Sarah Ressel, and Emmanuelle Charpentier. 2018. “The Biology of CRISPR-Cas: Backward and Forward.” *Cell* 172: 1239–59. https://doi.org/10.1016/j.cell.2017.11.032.

Hillenmeyer, Maureen E., Eula Fung, Jan Wildenhain, Sarah E. Pierce, Shawn Hoon, William Lee, Michael Proctor, et al. 2008. “The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes.” *Science*. https://doi.org/10.1126/science.1150021.

Huang, Chun Hao, Ko Chuan Lee, and Jennifer A Doudna. 2018. “Applications of CRISPR-Cas

Enzymes in Cancer Therapeutics and Detection." *Trends in Cancer*. https://doi.org/10.1016/j.trecan.2018.05.006.

Huang, Kerwyn Casey. 2015. "Applications of Imaging for Bacterial Systems Biology." *Current Opinion in Microbiology* 27: 114–20. https://doi.org/10.1016/j.mib.2015.08.003.

Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A Doudna, and Emmanuelle Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21. https://doi.org/10.1126/science.1225829.

Jones, Daniel Lawson, Prune Leroy, Cecilia Unoson, David Fange, Vladimir Ćurić, Michael J Lawson, and Johan Elf. 2017. "Kinetics of DCas9 Target Search in Escherichia Coli." *Science* 357 (6358): 1420–24. https://doi.org/10.1126/science.aah7084.

Jost, Marco, Daniel A Santos, Reuben A Saunders, Max A Horlbeck, John S Hawkins, Sonia M Scaria, Thomas M Norman, et al. 2019. "Titrating Gene Expression with Series of Systematically Compromised CRISPR Guide RNAs." *BioRxiv*, 717389. https://doi.org/10.1101/717389.

Kampmann, Martin, Michael C Bassik, and Jonathan S Weissman. 2013. "Integrated Platform for Genome-Wide Screening and Construction of High-Density Genetic Interaction Maps in Mammalian Cells." *Proceedings of the National Academy of Sciences* 110 (25): E2317–26. https://doi.org/10.1073/pnas.1307002110.

Knott, Gavin J, and Jennifer A Doudna. 2018. "CRISPR-Cas Guides the Future of Genetic Engineering." *Science*. https://doi.org/10.1126/science.aat5011.

Koo, Byoung-Mo, George Kritikos, Jeremiah D. Farelli, Horia Todor, Kenneth Tong, Harvey Kimsey, Ilan Wapinski, et al. 2017. "Construction and Analysis of Two Genome-Wide Deletion Libraries for Bacillus Subtilis." *Cell Systems*. http://www.cell.com/cell-systems/pdf/S2405-4712(16)30447-1.pdf.

Kritikos, George, Manuel Banzhaf, Lucia Herrera-Dominguez, Alexandra Koumoutsi, Morgane

Wartel, Matylda Zietek, and Athanasios Typas. 2017. “A Tool Named Iris for Versatile High-Throughput Phenotyping in Microorganisms.” *Nature Microbiology* 2. https://doi.org/10.1038/nmicrobiol.2017.14.

Kuwada, Nathan J., Beth Traxler, and Paul A. Wiggins. 2015. “Genome-Scale Quantitative Characterization of Bacterial Protein Localization Dynamics throughout the Cell Cycle.” *Molecular Microbiology* 95 (1): 64–79. https://doi.org/10.1111/mmi.12841.

Lee, Henry H, Nili Ostrov, Brandon G. Wong, Michaela A. Gold, Ahmad S. Khalil, and George M. Church. 2019. “Functional Genomics of the Rapidly Replicating Bacterium Vibrio Natriegens by CRISPRi.” *Nature Microbiology*. https://doi.org/10.1038/s41564-019-0423-8.

Li, Wei, Han Xu, Tengfei Xiao, Le Cong, Michael I Love, Feng Zhang, Rafael A Irizarry, Jun S. Liu, Myles Brown, and X. Shirley Liu. 2014. “MAGeCK Enables Robust Identification of Essential Genes from Genome-Scale CRISPR/Cas9 Knockout Screens.” *Genome Biology* 15 (12): 554. https://doi.org/10.1186/s13059-014-0554-4.

Li, Xin-tian, Yonggun Jun, Michael J Erickstad, Steven D Brown, Adam Parks, Donald L Court, and Suckjoon Jun. 2016. “TCRISPRi : Tunable and Reversible , One-Step Control of Gene Expression.” *Nature Publishing Group*, 1–12. https://doi.org/10.1038/srep39076.

Li, Yi, Shiyuan Li, Jin Wang, and Guozhen Liu. 2019. “CRISPR/Cas Systems towards Next-Generation Biosensing.” *Trends in Biotechnology*. https://doi.org/10.1016/j.tibtech.2018.12.005.

Love, Michael I, Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15: 550. https://doi.org/10.1186/s13059-014-0550-8.

Nichols, Robert J, Saunak Sen, Yoe Jin Choo, Pedro Beltrao, Matylda Zietek, Rachna Chaba, Sueyoung Lee, et al. 2011. “Phenotypic Landscape of a Bacterial Cell.” *Cell* 144 (1): 143–56. https://doi.org/10.1016/j.cell.2010.11.052.

Nonejuie, Poochit, Michael Burkart, K. Pogliano, and Joe Pogliano. 2013. “Bacterial Cytological

Profiling Rapidly Identifies the Cellular Pathways Targeted by Antibacterial Molecules." *Proceedings of the National Academy of Sciences* 110 (40): 16169–74. https://doi.org/10.1073/pnas.1311066110.

Peters, Jason M., Alexandre Colavin, Handuo Shi, Tomasz L. Czarny, Matthew H. Larson, Spencer Wong, John S. Hawkins, et al. 2016. "A Comprehensive, CRISPR-Based Functional Analysis of Essential Genes in Bacteria." *Cell* 165 (6): 1493–1506. https://doi.org/10.1016/j.cell.2016.05.003.

Peters, Jason M, Byoung-Mo Koo, Ramiro Patino, Gary E Heussler, Cameron C Hearne, Jiuxin Qu, Yuki F Inclan, et al. 2019. "Enabling Genetic Analysis of Diverse Bacteria with Mobile-CRISPRi." *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0327-z.

Pickar-Oliver, Adrian, and Charles A Gersbach. 2019. "The next Generation of CRISPR–Cas Technologies and Applications." *Nature Reviews Molecular Cell Biology*. https://doi.org/10.1038/s41580-019-0131-5.

Qi, Lei S, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim. 2013. "Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression." *Cell* 152 (5): 1173–83. https://doi.org/10.1016/j.cell.2013.02.022.

Qu, Jiuxin, Neha K Prasad, Michelle A Yu, Shuyan Chen, Amy Lyden, Nadia Herrera, Melanie R Silvis, et al. 2019. "Modulating Pathogenesis with MOBILE-CRISPRI." *Journal of Bacteriology* 201 (22): 1–9. https://doi.org/10.1128/JB.00304-19.

Rancati, Giulia, Jason Moffat, Athanasios Typas, and Norman Pavelka. 2018. "Emerging and Evolving Concepts in Gene Essentiality." *Nature Reviews. Genetics*. https://doi.org/10.1038/nrg.2017.74.

Rauch, B.J., M.R. Silvis, J.F. Hultquist, C.S. Waters, M.J. McGregor, N.J. Krogan, and J. Bondy-Denomy. 2017. "Inhibition of CRISPR-Cas9 with Bacteriophage Proteins." *Cell* 168 (1–2). https://doi.org/10.1016/j.cell.2016.12.009.

Richardson, Christopher D, Graham J Ray, Mark A Dewitt, Gemma L Curie, and Jacob E Corn. 2016. “Enhancing Homology-Directed Genome Editing by Catalytically Active and Inactive CRISPR-Cas9 Using Asymmetric Donor DNA” 34 (3). https://doi.org/10.1038/nbt.3481.

Rock, Jeremy M., Forrest F. Hopkins, Alejandro Chavez, Marieme Diallo, Michael R. Chase, Elias R. Gerrick, Justin R. Pritchard, et al. 2017. “Programmable Transcriptional Repression in Mycobacteria Using an Orthogonal CRISPR Interference Platform.” *Nature Microbiology* 2 (February): 1–9. https://doi.org/10.1038/nmicrobiol.2016.274.

Rousset, Francois, Lun Cui, Elise Siouve, Christophe Becavin, Florence Depardieu, and David Bikard. 2018. “Genome-Wide CRISPR-DCas9 Screens in E. Coli Identify Essential Genes and Phage Host Factors.” *PLoS Genetics* 74 (4): 417–22. https://doi.org/10.1371/journal.pgen.1007749.

Shi, Handuo, Alexandre Colavin, Marty Bigos, Carolina Tropini, Russell D. Monds, and Kerwyn Casey Huang. 2017. “Deep Phenotypic Mapping of Bacterial Cytoskeletal Mutants Reveals Physiological Robustness to Cell Size.” *Current Biology* 27 (22): 3419-3429.e4. https://doi.org/10.1016/j.cub.2017.09.065.

Shi, Handuo, Alexandre Colavin, Timothy K Lee, and Kerwyn Casey Huang. 2017. “Strain Library Imaging Protocol for High-Throughput, Automated Single-Cell Microscopy of Large Bacterial Collections Arrayed on Multiwell Plates.” *Nature Protocols* 12 (2): 429–38. https://doi.org/10.1038/nprot.2016.181.

Shiver, Anthony L, Hendrik Osadnik, George Kritikos, Bo Li, Nevan Krogan, Athanasios Typas, and Carol A Gross. 2016. “A Chemical-Genomic Screen of Neglected Antibiotics Reveals Illicit Transport of Kasugamycin and Blasticidin S.” *PLoS Genetics* 12 (6). https://doi.org/10.1371/journal.pgen.1006124.

Sternberg, Samuel H, Sy Redding, Martin Jinek, Eric C Greene, and Jennifer A Doudna. 2014. “DNA Interrogation by the CRISPR RNA-Guided Endonuclease Cas9.” *Nature* 507 (7490): 62–67. https://doi.org/10.1038/nature13011.

Szczelkun, Mark D, Maria S Tikhomirova, Tomas Sinkunas, Giedrius Gasiunas, Tautvydas Karvelis, Patrizia Pschera, Virginijus Siksnys, and Ralf Seidel. 2014. “Direct Observation of R-Loop Formation by Single RNA-Guided Cas9 and Cascade Effector Complexes.” *Proceedings of the National Academy of Sciences of the United States of America* 111 (27): 9798–9803. https://doi.org/10.1073/pnas.1402597111.

Ursell, Tristan, Timothy K. Lee, Daisuke Shiomi, Handuo Shi, Carolina Tropini, Russell D. Monds, Alexandre Colavin, et al. 2017. “Rapid, Precise Quantification of Bacterial Cellular Dimensions across a Genomic-Scale Knockout Library.” *BMC Biology* 15 (1): 1–15. https://doi.org/10.1186/s12915-017-0348-8.

Veening, Jan-Willem, and Xue Liu. 2017. “High-throughput CRISPRi Phenotyping Identifies New Essential Genes in Streptococcus Pneumoniae.” *Mol Syst Biol* 12 (888): 1–14. https://doi.org/10.15252/msb.

Vigouroux, Antoine, Enno Oldewurtel, Lun Cui, David Bikard, and Sven van Teeffelen. 2018. “Tuning DCas9’s Ability to Block Transcription Enables Robust, Noiseless Knockdown of Bacterial Genes.” *Molecular Systems Biology* 14 (3): e7899. https://doi.org/10.15252/msb.20177899.

Wang, Tianmin, Changge Guan, Jiahui Guo, Bing Liu, Yinan Wu, Zhen Xie, Chong Zhang, and Xin Hui Xing. 2018. “Pooled CRISPR Interference Screening Enables Genome-Scale Functional Genomics Study in Bacteria with Superior Performance-Net.” *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-04899-x.

Widom, Julia R., Victoria Rai, Christopher E. Rohlman, and Nils G. Walter. 2019. “Versatile Transcription Control Based on Reversible DCas9 Binding.” *RNA* 25 (11): 1457–69. https://doi.org/10.1261/rna.071613.119.

Chapter 2

Functional analysis of *E. coli* essential genes using CRISPRi

INTRODUCTION

Bacterial essential genes encode for the fundamental reactions of cellular life and are often the direct targets of antibiotics, but the challenges of manipulating them genetically have thus far precluded systematic approaches to understanding their roles *in vivo*. CRISPR interference (CRISPRi) for the titratable knockdown of bacterial genes (Qi et al. 2013; Bikard et al. 2013; Peters et al. 2016) is positioned to narrow the gap between essential genes and their cellular phenotypes, however few systematic phenotypic studies have been completed (Peters et al. 2016; Veening and Liu 2017). High-dimensional phenotyping approaches such as chemical-genetic screening (Nichols et al. 2011; Shiver et al. 2016) and morphological profiling (Veening and Liu 2017; Nonejuie et al. 2013; Peters et al. 2016; Campos et al. 2018) each present high-throughput modalities with which to capture large numbers of independent, quantitative phenotypes of arrayed bacterial libraries. For example, chemical-genetic screens of the *Escherichia coli* non-essential gene deletion library (Baba et al. 2006) have identified functional interactions on the basis of shared chemical signatures (Nichols et al. 2011) as well as surveying available cellular mechanisms of antibiotic resistance (Shiver et al. 2016). The extension of these screening technologies to the phenotypes of slight essential gene depletions has already yielded surprising cross-pathway functional interactions of essential genes and further insights into antibiotic action in the model Gram-positive *Bacillus subtilis* (Peters et al. 2016). Furthermore, the titratable nature of CRISPRi for perturbing essential gene levels allow both the sensitive probing of key regulators of cell morphology, but also exploration of “terminal phenotypes” in which essential genes are fully repressed, each reflecting important aspects of gene function (Peters et al. 2016).

As CRISPRi has been established as a facile tool with which to study essential genes diverse microbial species (Lee et al. 2019; Veening and Liu 2017; Rock et al. 2017; Peters et al. 2019), it presents a unique opportunity to compare the conservation of essential gene functions across diverged species (Hawkins et al. 2019). A major distinction in the bacterial domain is the

presence of the Gram-negative specific outer membrane across gammaproteobacterial and other major pathogens and commensal species. The outer membrane is a complex structure that defines a unique bacterial compartment (the periplasm) and provides protection against small molecule and other physical stressors (Rojas et al. 2018) in the environment. Despite its centrality in bacterial physiology, major gaps in our knowledge of how the barrier function of this structure is created, in part due to its essential nature in Gram-negative species. Finally, comparisons across species will fruitfully explore differences in the morphological space available to different species based on their intrinsic structural differences.

Motivated by this potential, here we systematically profile the chemical and morphological phenotypes of an *E. coli* essential gene knockdown library. By strategically aligning our assays with previously published datasets from a *B. subtilis* essential gene knockdown library (Peters et al. 2016), we are able to leverage this comparison to identify shared and Gram-negative specific mechanisms of intrinsic antibiotic resistance, as well as the primary regulators of cell morphology. Finally, we explore a novel phenomenon during sub-saturating CRISPRi targeting in bacteria which interacts non-productively with native feedback regulatory circuits on gene expression. We show that this feedback does not successfully restore homeostasis, and that dysregulation can take the form of increased heterogeneity in expression levels and other cellular phenotypes.

RESULTS AND DISCUSSION

Chromosomal CRISPRi enables the systematic functional study of essential genes in *E. coli*

To systematically analyze the functions of essential genes in the model Gram-negative bacterium *Escherichia coli*, we designed an inducible, chromosomally integrated CRISPRi system that is calibrated to achieve 50% knockdown in the absence of inducer (Figure 2.1A). Addition of IPTG reduces expression of an *rfp* target from 50% to ~7% in a uniform and

concentration-dependent manner (Figure 2.1B). Our arrayed library of 479 CRISPRi strains (Table 2.1), consists of 266 high-confidence essential genes, 81 genes of uncertain essentiality, and 132 non-essential or conditionally essential genes (Methods) (Koo et al. 2017; Baba et al. 2006; Yamamoto et al. 2009; Goodall et al. 2017; Patrick et al. 2007). Each gene is targeted by one computationally optimized sgRNA ("GitHub - Traeki/Sgrna_design" n.d.), with most (86-95%) either stronger than or within one standard deviation of the mean fitness impact of reported sgRNAs targeting that gene (Hawkins et al. 2019; Wang et al. 2018; Rousset et al. 2018) (Fig. 2.4B). The arrayed nature of our library enables many types of assays, including chemical-genetic screening and single-cell microscopy.

We reproducibly measured (Fig. 2.4C, r=0.897-0.974) relative strain fitness (RF = the number of doublings relative to that of non-targeted control strains (Methods)) using a pooled competition assay with or without induction (Figure 2.1C). Under induced conditions (1mM IPTG) few non-essential genes showed reduced fitness (median RF=1.00, Table 2.2), whereas most high-confidence essentials had reduced fitness (~75%, RF≤0.9, median RF=0.73), and many had fitness defects even without induction. The bottom 10th percentile of fitness defects without induction (RF<0.831, n=23) was enriched for ribosomal proteins and translation factors. A similar pooled competition experiment of *B. subtilis* CRISPRi strains (Peters et al. 2016) without induction (knockdown ~33-50%; RF<0.73, n=16) also showed enrichment for ribosomal proteins and translation factors with both shared and unique sets of translation factors among the most sensitized in these species (Fig. 2.4D). Thus, CRISPRi is effective at sensitizing cells specifically to essential gene knockdowns.

Chemical profiling of the essential gene knockdown library generates a robust dataset sensitive to operon structure for gene-phenotype assignment

We performed a chemical-genetic screen of the arrayed *E. coli* essential gene knockdown library, along with 91 representative non-essential deletion strains (Nichols et al.

2011) using sub-MIC concentrations of ~100 unique small molecule inhibitors, with some at multiple concentrations (150 conditions total, Table 2.3). The small molecules selected either elicited strong phenotypes from distinct sets of non-essential deletion strains (Nichols et al. 2011) or lacked known mechanisms of action (see Methods). We used endpoint growth to estimate condition-specific growth rates (quantified by S-scores (Collins et al. 2006)), as in our previous chemical-genetic screens (*B. subtilis* essential CRISPRi library (Peters et al. 2016); *E. coli* non-essential deletion library (Nichols et al. 2011)). We observed high correlation between knockdown growth measurements for condition replicates (≥4 replicates per condition, median Pearson r=0.711, Fig. 2.4E), and between strain replicates (2-4 replicates per strain, median Pearson r=0.680, Fig. 2.4F). Importantly, S-scores of the 91 deletion strains were well correlated to those in (Nichols et al. 2011) for the 70 overlapping conditions (Pearson r=0.565; Fig. 2.4G). Thus, the data from our screen was highly reproducible.

We defined significant chemical-gene phenotypes (FDR≤5%; (Nichols et al. 2011; Peters et al. 2016)), resulting in 1886 high-confidence phenotypes, with 64.5% of strains (309) having at least one phenotype (Figure 2.1D). We compared our phenotypic signatures to known functional interactions from several databases using ROC (receiver operating characteristic) curves to calculate a true positive rate (TPR) when the false positive rate (FPR) was set at 5% (Figure 2.1E). Highly correlated phenotypic signatures were better predictors of more specific metrics of functional interactions (e.g. members of the same complex, >4 shared GO biological process terms) than higher-level metrics (e.g. members of the same pathway, >2 shared GO biological process terms). As expected, the predictive power of all metrics increased by excluding strains with few significant phenotypes. This pattern also held in a reanalysis of the *B. subtilis* CRISPRi data in (Peters et al. 2016) for comparable metrics, although differences in annotation depth in some cases prevented a direct comparison. Therefore, phenotypic signatures capture specific functional interactions between essential genes.

In bacterial genomes, genes are often organized into co-transcribed operons. CRISPRi targeting reduces the transcript level of downstream operon members by blocking RNA polymerase transit and of upstream operon members by an unknown mechanism termed "reverse polarity" (Peters et al. 2015, 2016). Using a simple bioinformatic definition of operons (Methods), we found that phenotypic signatures were predictive of operons in both *E. coli* and *B. subtilis*, whether operons had consistent functions (>=50% of members sharing one GO biological process term) or were of mixed function (Figure 2.1E). However, using ROC analysis exclusively on genes from mixed function operons was less predictive of functional interactions than using only single genes (genes not predicted to be in an operon), suggesting that targeting mixed-function operons causes multiple perturbations that can obscure the phenotypic signature of the targeted gene.

Chemical profiling of the essential knockdown library provides a novel lens for cross-species comparisons of essential pathways

The direct targets of most antibiotics are conserved across even distantly related species. Thus, antibiotics are species-agnostic probes of bacterial cell biology, enabling comparison of *E. coli* and *B. subtilis* enrichment patterns (data from (Peters et al. 2016)). To broadly examine the chemical-genetic interactions, we grouped drugs by their targeted process (Figure 2.2A) and looked for patterns of drug sensitization (S-scores<0) that were significantly enriched within functional groups of genes (hypergeometric test, Bonferroni corrected $p<0.05$; Figure 2.2B). We observed enriched sensitivities in both species when drug target and gene function are directly related: e.g. DNA related stresses and DNA processes (e.g. DNA replication), PMF disruptors and quinone biosynthesis and recycling (e.g. riboflavin biosynthesis (*E. coli*) or menaquinone biosynthesis (*B. subtilis*)), cell wall synthesis inhibitors and cell wall biosynthesis genes (e.g. peptidoglycan biosynthesis), tetrahydrofolate inhibitors and genes in tetrahydrofolate biosynthesis, and protein synthesis inhibitors and translation factors (e.g.

aminoacyl-tRNA metabolism). Intriguingly, one shared enrichment pattern—knockdown of cell wall and cell division genes causing sensitivities to DNA damaging stresses—is not readily explained by direct synergistic effects. This observation suggests an evolutionarily conserved connection between the cell wall and DNA damage. Additionally, we observe differing enrichment patterns for some drug categories (e.g. protein synthesis) that are also not explained by direct effects. We next investigate these two cases of commonality and difference, and specifically how the latter may be attributed to the central distinction between *E. coli* and *B. subtilis*: the presence of the Gram-negative specific outer membrane (OM).

The knockdowns of essential cell wall and cell division genes independently contribute to DNA damage sensitivity

Knockdowns of cell wall or cell division genes increase sensitivity to DNA damaging stresses in both *E. coli* and *B. subtilis*, responding to ciprofloxacin in both species (Fig. 2.5A,B). As these knockdowns do not induce the SOS response in either species (Fig. 2.5C), they are unlikely to impact genome integrity in a way that is sensed by RecA/LexA. Two findings suggested that cell wall and cell division knockdowns independently contribute to DNA damage sensitization in *E. coli*. First, cell wall biosynthesis genes (e.g. *murB*) and cell division genes (e.g. *minE*) not co-transcribed with the other gene class are sensitized (Fig. 2.5A). Second, analysis of transcript levels in single knockdown strains of adjacent cell wall and cell division genes (Conway et al. 2014; Lalanne et al. 2018) (Figure 2.2C) reveal four transcriptionally independent units comprised exclusively or predominantly of either cell wall or cell division genes (Figure 2.2D), again suggesting independent effects on the DNA damage phenotype.

Using a liquid growth assay to monitor growth with or without sub-inhibitory concentrations of the chain-terminating nucleoside analogue azidothymidine (Elwell,' et al. 1987), we first demonstrated that both cell division (*ftsA, ftsL*) and cell wall biosynthesis (*murE*) gene knockdowns were sensitized to DNA damage compared to a control strain (Figure 2.2E,

Fig. 2.5D). We then asked whether the observed DNA damage sensitivity results from enhancing SulA-mediated division inhibition during SOS, where SulA prevents FtsZ polymerization during stress, and is then degraded by Lon during recovery (Mukherjee, Cao, and Lutkenhaus 1998; Dajkovic, Mukherjee, and Lutkenhaus 2008; Chen, Milam, and Erickson 2012). Consistent with this hypothesis, deleting *sulA* from the *ftsL* and *ftsA* cell division knockdown strains partially alleviated their sensitivity to azidothymidine (Figure 2.2E, Fig. 2.5D). Our finding that cell division knockdowns synergize with SulA is also consistent with the finding that *lon* is essential in strains with slight division defects (e.g. reduced-expression *ftsZ* mutants (Nazir and Harinarayanan 2016)) in a *sulA*-dependent manner, and that preventing SulA degradation (Δ*lon* strain) (Nichols et al. 2011) sensitizes cells to azidothymidine.

Interestingly, deleting *sulA* had no effect on DNA damage sensitization in the *murE* knockdown strain (Figure 2.2E, Fig. 2.5D). *murE* is less sensitized to DNA damage than *ftsL/A*, making it unlikely that the lack of effect of Δ*sulA* simply reflects a more severe division defect of *murE*. Instead, *murE* is likely to have a distinct mechanism for sensitization. Additionally, partial rescue of *ftsL* and *ftsA* knockdowns by Δ*sulA* raises the possibility that additional DNA damage-induced division inhibitors contribute to their sensitization (Maguin et al. 1986; Jaffe, D'Ari, and Norris 1986; Modell, Hopkins, and Laub 2011). In summary, our results are in agreement with complex coordination between peptidoglycan biosynthesis, cell division and genome integrity and/or replication, mediated in part by SulA and in part by as yet unknown mechanism(s).

Contributions to the barrier function of the outer membrane

Although the contributions of individual genes to barrier function has been documented (Vuorio and Vaara 1992; Galloway and Raetz 1990; Normark 1970; Young and Silver 1991; Grundstrom, Normark, and Magnusson 1980; Vaara and Nurminen 1999; Sperandeo et al. 2008; Sampson, Misra, and Benson 1989; Ruiz et al. 2005; Malinverni et al. 2006; Wu et al. 2005; Doerrler and Raetz 2005), our dataset is the first systematic interrogation of the

contributions of essential and many non-essential genes to the intrinsic resistance mechanisms provided by the Gram-negative outer membrane. There is currently great interest in reengineering Gram-positive-restricted antibiotics to meet the challenge of growing antibiotic resistance in Gram-negative pathogens, and design rules must contend with the unique permeability barrier imposed by the Gram-negative outer membrane (Richter and Hergenrother 2018). To advance the development of such rules, we comprehensively explored the signatures of outer membrane permeability in our dataset.

We first determined whether the processes inhibited by antibiotics or properties of the antibiotics themselves (irrespective of targeted process) were responsible for the enriched sensitivity of the 17 OM gene depletions to inhibitors of protein synthesis, transcription, and membrane stability (Figure 2.1A). Using tSNE to cluster all conditions based on OM gene knockdown phenotypes, we found no clustering based on targeted pathway (Figure 2.2F), but strong clustering of the 9 antibiotics in our screen unable to accumulate in *E. coli* (intracellular accumulation <300 nmol/10^{12} CFU (Richter et al. 2017)). We broadened this analysis by identifying additional strains specifically sensitized to non-accumulating antibiotics, and additional small molecule inhibitors highly correlated with at least one non-accumulating antibiotic (Figure 2.2H). We identified 24 additional compounds (33 total) whose phenotypic signatures among OM genes are strongly correlated with those of the non-accumulating drugs. These compounds were highly correlated as a group (median Pearson r=0.549 within outer membrane gene phenotypes; Figure 2.2H), and include molecules acting at the inner membrane either by disrupting the proton motive force (phenazine and triclosan (Domenech et al. 2019)) or acting as detergents (bile salts, benzalkonium, and chlorpromazine); as well as others with cytoplasmic targets, which may be OM-limited based on their large molecular weights (clarithromycin, spiramycin, bleomycin, actinomycin D, holomycin/phleomycin, ceftazidime, oxacillin) (Table 2.3).

Genes specifically sensitized to non-accumulating antibiotics (hypergeometric test, $p<0.05$) were predominantly from the pathways of LPS biosynthesis (*lpxABD, ΔrfaC, ΔrfaE, ΔlpcA*), LPS transport (*lptACEG*), outer-membrane protein chaperones and insertion machineries (*bamA, ΔbamB, ΔsurA*) and non-essential efflux transporters (*ΔacrB, ΔtolC*); each group exhibited distinct sensitization patterns. LPS biosynthesis genes were broadly sensitized to all 33 inhibitors, as were the major efflux machinery genes (*acrB, tolC*), except for retaining resistance to vancomycin and rifampicin, which may be poor efflux substrates because they have the highest molecular weights. The broad sensitization of *acrB* and *tolC* may result from reduced capacity to specifically efflux antibiotics, and/or from increased general permeability caused by membrane stress (Mateus et al. 2018). By contrast, LPS transport machinery genes (the Lpt complex) were predominantly sensitized to non-accumulating antibiotics (except *lptG*). These restricted sensitivities represent either distinct cellular outcomes from perturbing biosynthesis vs. transport or indicate that LPS biosynthesis is rate-limiting for populating the outer leaflet with LPS. Genes involved in the folding or insertion of outer membrane proteins (the Bam complex) were sensitized only to larger non-accumulating antibiotics, possibly indicating a defect in permeability but not efflux, the reverse of the phenotypic signatures of the efflux deletions. These data indicate that the processes of permeability and efflux can be decoupled. The sensitivity of LPS biosynthesis knockdown strains to both effluxed and (potentially) non-effluxed antibiotics raises the possibility that LPS content both limits permeability and is required for efficient efflux.

We were particularly interested in understanding the genes most strongly sensitized to non-accumulating antibiotics, most of which function in LPS biosynthesis. Because CRISPRi knockdown affects operons, and some of these genes are adjacent (Figure 2.2I), we determined whether each gene independently contributed to sensitization. Comparison of the total phenotypic signatures upon knockdown suggested a highly similar phenotypic outcome from

targeting *bamA*, *lpxD*, *fabZ*, *lpxA*, and *lpxB*, but not from targeting upstream of *bamA* (*uppS*, *cdsA*, *rseP*) (Figure 2.2I). *bamA* and *lpxD* had relatively isolated effects on transcription when assayed by mRNA-seq (Figure 2.2J), possibly because knockdown induced σ^E (Fig. 2.5E), whose multiple promoters might alleviate CRISPRi knockdown (Figure 2.2J). Expression of non-targeted variants of *bamA* or *lpxD* each restored wild-type levels of vancomycin resistance to their respective knockdown strains, suggesting independent contributions to OM integrity (Figure 2.2K). In contrast, targeting *lpxB* affected the expression of many genes (*fabZ, lpxA, lpxB, rnhB, dnaE* and *accA*) (Figure 2.2J). We discounted *dnaE* and *accA* since their phenotypes when targeted are inconsistent with an outer membrane integrity role (Figure 2.2I), and focused on *fabZ* and *lpxA,* the genes most strongly sensitized to vancomycin. Their sensitivity upon knockdown was alleviated both by complementation with non-targeted variants (Figure 2.2K) and by expression of the other gene (Fig. 2.5F), suggesting that each may contribute to vancomycin resistance. On its face this result presents an intriguing paradox, as LpxA and FabZ compete for substrates at the branchpoint between Lipid A biosynthesis and fatty acid biosynthesis (Figure 2.2L).

Our screen also identified novel sensitized CRISPRi strains whose targets had not previously been connected to barrier function. We discounted candidates reflecting secondary effects due to a) operon level knockdown (e.g. *leuS* and *holA* are in the *lptE* operon); b) upstream effects on LPS biosynthesis (e.g. *glmS*); c) imperfect overlap with non-accumulating drug set sensitivity (e.g. *zipA, parE, rpsK*). However, *alaS*, an alanine-tRNA ligase, did not have any of these confounding factors and hence may have a novel role in barrier function. Its knockdown phenotypes are strongly correlated with those of outer membrane genes (median r=0.476 with OM genes). We verified that knockdown in uninduced conditions (the conditions of the screen) affects transcript level, and showed that *alaS* expression was restored by full knockdown (Fig. 2.5G), as expected from the transcriptional feedback of AlaS (Putney and Schimmel 1981). This known feedback could explain lack of a phenotype in our vancomycin

sensitivity assay (Figure 2.2K), as these experiments are done under partial induction, where transcriptional compensation is expected. Importantly, *alaS* overexpression conferred slight resistance to vancomycin (Figure 2.2K), consistent with *alaS* having a homeostatic or stress-induced role in maintaining the OM barrier. The mechanism by which this occurs remains to be elucidated; aminoacyl-tRNA synthases (ARS) in other Gram-negative species often protect against cationic antimicrobial peptides (CAMPs) via Lipid A modifications (Fields and Roy 2018; Klein et al. 2009) although there are no known cases in *E. coli*.

In summary, we have systematically catalogued the contributions of essential and non-essential genes to OM integrity, with three important findings. First, our results reveal the profound effects of slight knockdown of LPS biosynthesis genes on drug entry, finding that a 2-fold knockdown of genes in Lipid A biosynthesis was as effective in allowing drug entry as deletion of later non-essential modifications of LPS core (*rfaA, rfaE*, *lpcA*: classical “deep rough” mutants). Our evidence is consistent with FabZ contributing to intrinsic vancomycin resistance via maintenance of the OM barrier, however it may also be the case that *fabZ* knockdown and complementation are primarily acting on flux through the LpxA Lipid A biosynthesis branch. We explore these possibilities further in the concluding perspective. Second, we find that LPS transport is more robust to knockdown than LPS synthesis in terms of increasing OM permeability. Finally, the broad and significant effects of *alaS* knockdown suggest that it may be an interesting drug target, as other aminoacyl-tRNA synthases have been (Hurdle, O’Neill, and Chopra 2005).

Predicted direct-target interactions reveal native feedback regulation of essential gene expression

Slight depletion of antibiotic targets should sensitize cells to the cognate antibiotic inhibitor, as has been observed in some cases in *B. subtilis* (Peters et al. 2016). Surprisingly, only 3/15 of the known antibiotic-target interactions in our screen were significantly sensitized

(FDR≤5%), although a majority of the remainder (7/12=58%) exhibited sensitization of "on-pathway" or related genes (Figure 2.3A). This suggests that these antibiotics and sgRNAs generally are active, but that compensatory regulation might obscure synergy. We selected 4 cases (*fabB*/cerulenin; *rho*/bicyclomycin; *fusA*/fusidic acid; *mreB/*A22) and one sensitized control (*fabI*/triclosan) for study. Quantifying transcript abundance without and with saturating induction, as compared to targeting *rfp*, revealed that our positive control, *fabI*/triclosan, and *fabB/*cerulenin, which was not sensitized, exhibited the expected knockdown efficacy (Figure 2.3B). Lack of a *fabB* phenotype may be due to the presence of FabF, a FabB homolog that is non-essential in *E. coli* but that is also inhibited by cerulenin (Price et al. 2001). Importantly, *B. subtilis* has only FabF (Koo et al. 2017) and the knockdown of *fabF* is sensitized to cerulenin (Figure 2.3A) (Peters et al. 2016). In contrast, *rho, fusA* and *mreB* transcripts were each present above wildtype levels when targeted with CRISPRi without induction. In each case, further induction of CRISPRi knockdown restored knockdown, suggesting that the cell has a limited capacity to compensate for reduced expression of these genes.

We developed two reporters to quantitatively measure knockdown and feedback at the single-cell level. The "knockdown reporter" drives *sfgfp* expression from an upstream region containing the native promoter and the 5' end of the gene (Zaslaver et al. 2006), and reports on expression at the endogenous locus: the net result of knockdown and compensatory regulation. The "feedback reporter" is identical to the knockdown reporter, except that its PAM sequence has been mutated so that expression reports only upregulation (Figure 2.3C). We explored two cases with well characterized negative feedback regulation (*rho, fusA)*, and a third where feedback is not completely understood (*gyrA*).

Rho prematurely terminates its own expression in its leader sequence (*rhoL*), allowing Rho to directly downregulate its own production (Matsumoto et al. 1986). Our reporter system confirms that non-saturating (uninduced conditions) CRISPRi slightly upregulates *rho* (Rousset et al. 2018). We further show that upregulation is uniform at the single-cell level, is overcome by

further induction of CRISPRi, and that bicyclomycin, a direct inhibitor of Rho, phenocopies the *rho* knockdown (Figure 2.3D). The bicyclomycin results confirm that both knockdown and feedback reporters are responsive to Rho activity levels. As the *rho* knockdown strain has no significant chemical phenotypes in the screen data, this overactive compensation may successfully approximate wildtype levels of expression without causing drastic cellular dysfunction.

Negative feedback of the *fusA* transcription unit is mediated by RpsG, which binds to an mRNA site upstream of its ORF and inhibits the translational coupling between *rpsL* and *rpsG-fusA-tufA* (Saito, Mattheakis, and Nomura 1994; Saito and Nomura 1994). A feedback reporter containing both the promoter and RpsG binding-site is upregulated in a *fusA* knockdown strain (Figure 2.3E), but mutations that disrupt RpsG binding (Saito and Nomura 1994) abrogate this upregulation. Thus, CRISPRi-induced feedback depends solely on RpsG binding, a finding we validated by showing that inhibiting FusA (EF-G) activity with fusidic acid does not upregulate expression (Figure 2.3E). Importantly, the *fusA* knockdown has significant chemical phenotypes, potentially owing to the dysregulation of multiple key translation factors (*fusA*/EF-G, *tufA*/EF-Tu) and ribosomal proteins (RpsL, RpsG) which are no longer responsive to (presumably) excess RpsG.

Finally, we probed the complex interactions that maintain genomic supercoiling. DNA gyrase (encoded by *gyrAB*) introduces negative supercoils and its promoter activity increases when supercoiling is decreased (Rolf Menzel and Gellert 1983; R. Menzel and Gellert 1987). The non-essential topoisomerase 1 (*topA*) antagonizes gyrase action by relaxing DNA supercoils (Gellert et al. 1982). Additionally, topoisomerase IV (*parCE*) has an essential role in decatenating chromosomes following replication (Zechiedrich, Khodursky, and Cozzarelli 1997). Novobiocin inhibits the activities of both *gyrB* and *parE in vivo* (Hardy and Cozzarelli 2003). Knockdown of *parE* but not *gyrB* was sensitized to novobiocin in both *E. coli* and *B. subtilis*, suggesting a compensatory mechanism specific to *gyrB* that may be shared.

Our chromosomally integrated feedback reporter revealed that increasing knockdown of either *gyrA* or *gyrB* upregulated expression of both *gyrA* and *gyrB* promoters (Figure 2.3F), likely preventing their knockdown (median feedback reporter expression=0.97 for *gyrA* and 1.4 for *gyrB* relative to non-targeting strain control). Notably, *topA* is also upregulated in the *gyrA* and *gyrB* knockdown strains even at basal knockdown, suggesting that compensation at the *gyrA* and *gyrB* promoters does not precisely restore native supercoiling, triggering overexpression of *topA* (Y. C. Tse-Dinh and Beran 1988; Y.-C. Tse-Dinh 1985). These results may suggest that cells oscillate between a hyper- and hypo-supercoiled state as DNA gyrase levels fluctuate with slight knockdown. Importantly, while the contributions of supercoiling-sensitive promoters of *gyrA* and *gyrB* in *E. coli* to homeostatic control of genomic supercoiling have been deeply explored, it is unclear whether the same paradigm maintains supercoiling in Gram-positives. Some Streptomyces have been shown to transcriptionally regulate *gyrA* and *gyrB* in response to short-term supercoiling stress (Szafran et al. 2016), while other evidence from *B. subtilis* suggests that mutations that increase Topo IV expression are a more facile way to rebalance supercoiling, perhaps indicating additional constraints on DNA gyrase levels (Reuß et al. 2019). The lack of sensitization of *gyrAB* knockdowns in *B. subtilis* may suggest that, like *E. coli* and Streptomyces, regulatory feedback is equipped to manage minor challenges to supercoiling homeostasis.

Taken together, these data show that feedback regulation actively tries to restore homeostasis when essential genes are depleted, and furthermore that even in conditions of sub-saturating CRISPRi targeting homeostasis may not be restored. Specifically, these studies indicate that sub-saturating targeting of *rho* leads to uniform upregulation (2-fold); that feedback on expression of the *fus* operon is regulated solely by RpsG with no contribution from *fusA* (EF-G); and that targeting of DNA gyrase subunits causes dysregulation of its own expression and that of other supercoiling-controlled genes. More broadly, this methodology will permit systematic analysis of the input-output relationships of regulatory circuits governing expression

at the single-cell level and open the door to systematic discovery and analysis of the regulatory mechanisms governing essential gene expression.

PERSPECTIVE

Here we describe the first systematic exploration of the phenotypes of partial essential gene depletion in the model Gram-negative bacterium *E. coli*, shedding new light on the roles of essential gene products *in vivo*. Using high-throughput quantitative chemical screens, we identified strengths and limitations of CRISPRi for the identification of gene phenotypes, namely that knockdown phenotypes reflect perturbations to the combined functions of all operon members. Nonetheless, we identify an underexplored mechanism of sensitization to DNA damage caused by slight knockdown of genes involved in cell wall biosynthesis and cell division; a connection that is shared between *E. coli* and *B. subtilis*. We identify one mechanism of interaction between these two processes which is the action of SOS-induced cell division inhibitors, and specifically demonstrate the importance of SulA in mediating sensitivity of cell division (partially) but not cell wall biosynthesis gene knockdowns. It is possible that a unifying connection can be found between the cell wall and cell division knockdowns that explains the remaining sensitivity of cell division and the totality of the cell wall biosynthesis gene knockdowns. Alternatively, perturbations to either cell wall synthesis or cell division may each connect to DNA damage by their own secondary mechanisms. For example, interfering with cell division has been shown in *B. subtilis* to irreversibly prevent DNA replication initiation (Arjes et al. 2014), though such a mechanism in *E. coli* has not been completely described (Sánchez-Gorostiaga et al. 2016). Among all general classes of drug targets, cell wall targeting and DNA targeting drugs demonstrate synergy when combined in gamma-proteobacteria, a relatively rare characteristic of antibiotic combinations (Brochado et al. 2018).

We pursue an interesting and medically important phenotype of outer membrane permeability, identified by its broad-spectrum antibiotic sensitization which was absent in the *B.*

subtilis CRISPRi dataset. We identify other small molecule stresses that synergize with outer membrane permeability, identifying compounds that disrupt the inner membrane and the proton motive force, suggesting that these two pathways are reliant on the protective barrier of the outer membrane for their optimal function. We also identify a large number of antibiotics from diverse classes that may be outer membrane limited; further analysis on this set may identify and/or clarify rules of compound accumulation in Gram-negative species, an important line of research.

The finding that FabZ may contribute to the OM permeability barrier is a novel one that raises intriguing possibilities about the regulation of LPS production. It is possible that the primary effect of *fabZ* targeting is on *lpxA* expression, suggesting that LpxA levels are a lever with which to tune LPS production. LpxA has been neglected as such a point of regulation because, although first in the Lipid A biosynthesis pathway, its reaction is thermodynamically unfavorable and requires the activity of LpxC downstream as the first committed step; however it is known that reduced activity *lpxA* mutants are sufficient to select for reduced activity *fabZ* mutants (Mohan et al. 1994), consistent with its activity being limiting in extreme cases. LpxC is sole known regulator of this branchpoint, and can become stabilized by FabZ overexpression or hyperactivity (Zeng et al. 2013; Ogura et al. 1999). Therefore, *fabZ* overexpression may restore the resistant phenotype to the *fabZ* or *lpxA* knockdowns by stabilizing LpxC and restoring flux towards LPS production. FabZ is the primary dehydratase of unsaturated fatty acid biosynthesis, and a model has been proposed in which LpxC stability is favored by increased acyl-ACP pools, the result of FabZ activity (Ogura et al. 1999). It is possible that compensation via LpxC is only equipped to rebalance flux towards LPS biosynthesis when FabZ activity is high, and cannot respond when FabZ activity is too low, or when LPS biosynthesis is depressed in the absence of FabZ activity. The synteny of these two genes—*fabZ* and *lpxA*—further suggests that other regulatory mechanisms are possible.

Finally, using a small set of predicted chemical-genetic interactions between antibiotics and their direct targets, we find that sub-saturating CRISPRi targeting of some essential genes does not cause the predicted sensitization because of regulatory compensation on the targeted gene itself. This regulatory compensation follows known features of feedback regulation, and some presented evidence suggests that during sub-saturating CRISPRi targeting, cells experience fluctuations in the concentration of essential gene products. In other cases, essential genes are readily knocked down by CRISPRi, suggesting that cells have flexible requirements for some essential gene products and strictly regulate others. The methodology we present here for identifying and characterizing feedback regulation caused by CRISPRi perturbation will be productively applied to broader screens for this behavior, not limited to essential genes.

MATERIALS AND METHODS

Experimental model and subject details:

Microbes

Escherichia coli strains were cultured in LB medium at 37C, or in MOPS complete with glucose (MOPS EZ Rich Defined, Teknova #M2105) at 37C, as indicated.

Method details:

General strain manipulations and procedures

CRISPRi strain construction:

The lambda-att integrating plasmid pCAH63 (Haldimann and Wanner 2001) was modified to contain an sgRNA expression cassette to generate pCs-550r in the following steps: the sgRNA constant region was cloned from pgRNA-bacteria (Addgene #44251 (Qi et al. 2013)), the terminators L3S3P22 and L3S2P21 were cloned up and downstream, respectively, to flank the sgRNA cassette, and the sgRNA promoter was changed from BBa_J23119 to PlLac-O1 (Lutz and Bujard 1997). New 20nt spacers were cloned into pCs-550r by inverse PCR (Larson et al.

2013), Sanger sequenced, and transformed into *E. coli* BW25113 harboring pINT-ts to promote integration at lambda att (Haldimann and Wanner 2001) using $CaCl_2$-competence and selecting for chloramphenicol resistance.

High-efficiency conjugation was used to transfer *dcas9* from the chromosome of a donor strain to the chromosome of a sgRNA-encoding recipient strain. A "pseudo-Hfr" strain isogenic to BW25113 carries the transfer region from F and a spectinomycin marker integrated downstream of *rhaM* (4086kb) (Typas et al. 2008). The *dcas9* donor strain was constructed by integrating *dcas9* and a gentamicin resistance marker at the Tn7 att site (Choi and Schweizer 2006), adjacent to the origin of transfer, using the Mobile-CRISPRi triparental mating strategy (Peters et al. 2019). To clone the Tn7 cassette plasmid, *dcas9* was amplified from pdCas9-bacteria (Addgene #44249) under control of the synthetic promoter BBa_J23105 (http://parts.igem.org/). Conjugation was performed on LB plates by mixing the *dcas9* donor and sgRNA recipient in equal ratios, incubating for 5hr at 37C, pinning onto double-selection plates (chloramphenicol + gentamicin), and growth overnight. Single colonies from each conjugation mix were isolated by streaking onto double-selection plates.

RFP and RFP-GFP reporter strains for knockdown quantification:

The *rfp* cassette including kan marker was PCR amplified from the entry vector used to construct the previously described RFP reporter strain (plasmid: pSLQ1232, strain: MG1655 *nfsA*::PlLac-O1-*mrfp*) (Qi et al. 2013), the *rfp* promoter changed from PLlac-O1 to a minimal synthetic promoter (BBa_J23119) {http://parts.igem.org/} to create pSLQ1232-P541-rfp, and integrated into BW25113 at *nfsA* by lambda red recombineering (Thomason et al. 2014) and selecting for kanamycin resistance. To construct the *rfp-gfp* synthetic operon reporter strain, *gfp* was cloned downstream of *rfp* in pSLQ1232-P541-rfp to create pSLQ1232-P541-rfp-gfp, and inserted into the chromosome at *nfsA* as described above. sgRNA plasmids targeting either *rfp* or *gfp* (pCs-550r or pCs-550-601, respectively) were integrated into the chromosome in the manner described for library plasmids above. Promoter variants were cloned along with *dcas9*

into the Tn7 cassette plasmid, and triparental mating was used to introduce *dcas9* cassette into the chromosome at Tn7att (as described above).

Transcriptional reporter plasmids:

We used transcriptional reporter plasmids selected from, or designed to mimic, the existing library from (Zaslaver et al. 2006). If the desired reporter was not a member of the library, the upstream region (150-400bp upstream of ORF and 50-100bp within ORF and containing the targeted protospacer) was amplified by PCR (see primers in Table 2.1) from the BW25113 genomic DNA with 25bp flanking sequence and assembled by HiFi (New England Biolabs #E2621L) with the PCR-amplified pUA66 vector. In the case of feedback reporters, PAM mutations were introduced by quick-change mutagenesis (see primers in Table 2.1). Plasmids were transformed into CRISPRi strains by electroporation, selecting for kanamycin resistance.

Complementation plasmids:

The ORFs of genes of interest were amplified from *E. coli* BW25113 genomic DNA (see primers in Table 2.1) and assembled into pBAD24 (Guzman et al. 1995) using Gibson assembly (NEB HiFi #E2621X). Plasmids were transformed into CRISPRi strains by electroporation, selecting for ampicillin resistance.

ΔsulA CRISPRi strains:

The *sulA::kan* allele was first moved into sgRNA recipient strains by P1 transduction (Thomason, Costantino, and Court 2007) and selecting for kanamycin resistance. Conjugation was used to move *dcas9* into the sgRNA *ΔsulA* recipients, as described above, double-selecting for kanamycin- and gentamicin-resistant colonies twice in succession, after which chloramphenicol resistance was confirmed by patching.

CRISPRi library design, construction, pooled growth experiment

Design:

sgRNAs were designed to target genes in *Escherichia coli* BW25113 having some evidence of essentiality in published datasets, as described in the main text and summarized in (see library description in Table 2.1). sgRNAs were designed to target within each gene's ORF near the 5' end, binding the non-template strand, and sgRNAs with multiple potential binding sites were avoided, as previously described (Peters et al. 2016). sgRNA design scripts are publicly available ("GitHub - Traeki/Sgrna_design" n.d.).

Arrayed library construction:

sgRNA plasmids were cloned, verified, and integrated into *E. coli* BW25113 as described for individual strains above. One isolate of each sgRNA recipient was stored by inoculating into 250ul LB with chloramphenicol in 96 deep-well plates, grown for 6.5hrs, mixed with glycerol, and stored at -80C.

Arrayed sgRNA recipient libraries and arrayed *dcas9* donor strain were pinned from glycerol stocks to separate LB agar plates using a ROTOR robot (Singer Instruments) and grown overnight. The arrayed recipient library was then mixed with the arrayed donor strain by pinning onto a new LB agar plate, and then grown for 8 hours to allow conjugation. Patches were mixed and transferred to a double-selection agar plate (gentamicin and chloramphenicol) using the ROTOR robot and grown overnight. Patches were each individually struck out on double selection plates for single colony isolation. To store the CRISPRi library, 2 isolates of each strain were inoculated in 250ul LB with chloramphenicol and gentamicin, supplemented with 0.2% glucose, in 96 deep-well plates, grown for 6.5hrs, mixed with glycerol, and stored at -80C in 96 well plates.

Pooled library construction:

To enable the use of deep sequencing to quantify relative fitness (see below), an additional ~50 non-targeting sgRNA plasmids were cloned and integrated into BW25113, as described above.

Control sgRNA spacers were selected as a random subset from previously characterized control sgRNAs (Hawkins et al. 2019). Construction of the pooled library (all library sgRNAs plus control sgRNAs) was identical to that of the arrayed library except that after the second double selection of the arrayed library, all patches were scraped from the agar plate, thoroughly mixed, and stored as glycerol stocks at -80C.

Pooled growth experiment:

To quantify the relative fitness of each CRISPRi strain, we enumerated the relative proportion of each sgRNA spacer in the mixed population by deep sequencing, before and after 15 doublings in saturating IPTG. Briefly, a single glycerol stock of the pooled library was fully thawed, inoculated into 10ml LB at 0.01 OD600, and grown for 2.5hr (final ~0.3 OD600). This culture was collected (10ml, t0) and used to inoculate replicate 4ml LB cultures (+/- 1mM IPTG) at 0.01 OD600, which were then repeatedly grown 130min to 0.3 OD600 (5x doublings) and back-diluted to 0.01 a total of 3 times (15x doublings). At the endpoint cultures were collected (4ml, t15) by pelleting (9000xg 2min) and stored at -80C. The following day genomic DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen #69506) with the recommended Gram-negative pre-treatment and RNAse A treatment. sgRNA spacer sequences were amplified from gDNA using Q5 polymerase (New England Biolabs #M0493S) for 14x cycles using custom primers containing TruSeq adapters and indices, followed by gel-purification from 8% TBE gels.

Spacer sequences were extracted from FASTQ files, counted by exact matching to expected library spacers, and their counts normalized within each sample to control for read depth. We calculate the fitness as Relative Fitness (Hawkins et al. 2019), where the log2 fold change is normalized by the median log2 fold change of the control sgRNAs, and adjusted by the number of doublings. All RF values are reported in Table 2.2.

The RF values of *B. subtilis* CRISPRi strains were recalculated from a previous dataset (Hawkins et al. 2019), using the same set of control sgRNAs as was used for the *E. coli*

CRISPRi pooled growth experiment. The measurements reflect the activities of the same sgRNAs used in the *B. subtilis* chemical genetic screen and original library (Peters et al. 2016).

Chemical screen of the arrayed CRISPRi library

Screen procedure

Chemical screening was performed and chemical-gene scores were calculated as previously described (Nichols et al. 2011; Peters et al. 2016; Shiver et al. 2016). Briefly, before screening, glycerol stock plates were thawed and robotically pinned onto LB agar plates in 1536 format using a ROTOR robot (Singer Instruments) to make the “source” plates. Following overnight growth, strains were transferred from source plates to plates containing sub-MIC chemical inhibitors/stresses using the ROTOR robot, and grown for 8-18hrs at 37C until the plate as a whole had measurable-sized patches, stored at 4C overnight, and imaged the following day using the spImager-M plate imaging system (S&P Robotics).

The strain array contains 2-4 total biological replicates of each CRISPRi strain, and a total of 91 deletion strains from the Keio collection (Baba et al. 2006) screened without replicates, in randomized positions within the array. Chemical concentrations were determined empirically by streaking the background strain (BW25113) on plates with 2-fold concentration ranges below the MIC. Concentrations were chosen such that growth was inhibited roughly 50% or less, and roughly 50% of small molecules were screened at multiple concentrations. The screen contained 4-5 replicates plates of each concentration. Chemical plates for screening were poured manually, dried 2 days at room temperature, and inspected for defects before screening. All conditions are described in Table 2.3.

Calculation of condition-specific phenotypes

Colony sizes were extracted from plate images using the Iris software package (Kritikos et al. 2017). Spatial effects were normalized using a quadratic function, median and variance of colony opacities were normalized between plates, and S-scores were computed using

previously described scripts in MATLAB (Collins et al. 2006; Shiver et al. 2016). False discovery rates for each condition-specific phenotype were computed from S-scores, on a condition-specific basis, as previously described (Nichols et al. 2011).

Further analyses of phenotypes

Gene-gene correlation validation

Genes were correlated (Pearson *r*) based on their phenotypic signatures, and the absolute value of gene-gene correlations (|*r*|) was compared to databases of functional connections in ROC analyses using the metrics package in sklearn. A simple definition of operons was used and applied to both *E. coli* and *B. subtilis* CRISPRi targets: co-directional genes where the ORF start is within 50bp of the upstream ORF stop were considered to be in the same operon. For comparison to the STRING database of functional interactions, several criteria were used to remove lower quality predictions: interactions were considered high quality only if their "experimental evidence" score was greater than 699.

Flow cytometry to quantify knockdown and reporter activities

Growth and flow cytometry analysis of the RFP reporter strains was done as described in (Rauch et al. 2017) with minor modifications. Strains were initially inoculated from single colonies and grown for ~5hr before dilution instead of overnight. Data was collected on a LSRII flow cytometer (BD Biosciences) using the yellow/green laser (561 nm) and the PE-Texas Red® detector (610/20 nm). Data for at least 20,000 cells were collected, and median fluorescence values were extracted using FlowJo (FlowJo, LLC). Error bars represent the standard deviation from 3 or more biological replicates. Data from representative samples were plotted as histograms using FlowJo.

For analysis of the transcriptional reporters, cultures were maintained in kanamycin selection throughout, and back-diluted in the presence of IPTG and/or drug treatment for 2-3hr,

before data was collected on the same instrument using the blue laser (488 nm) and the FITC detector (530/30 nm).

RT-qPCR to quantify knockdown

Growth and RT-qPCR

E. coli CRISPRi strains were grown in triplicate from single colonies in pre-warmed 4ml LB for 2.5hrs before back-dilution (1:80) in pre-warmed 4ml LB +/- 1mM IPTG and growth for 3hr prior to collection (OD600 ~ 0.2). The control strains express *rfp* with or without an sgRNAs targeting *rfp* ("non-targeting") and were treated identically. Samples were collected (300ul) in 900ul TRIzol-LS (Thermo Fisher #10296010) and stored at -20C overnight. The following day RNA was extracted according to the TRIzol protocol. RNA was quantified using a NanoDrop 2000c Spectrophotometer (Thermo Scientific) to normalize input (500ng input / 20ul reaction). For each RT-qPCR probe set and each sample replicate, reactions were performed in triplicate.

All RT-qPCR assays were done using the Luna Universal One-Step RT-qPCR kit (New England Biolabs #E3005S) according to its RT and cycling protocols, in 96 well PCR plates (Neptune #3732.X) and measured on a CFX Connect Real-Time System (Bio-Rad).

RT-qPCR analysis

Standard curves for each primer pair were first assessed on serially diluted RNA (extracted from the CRISPRi control strain) to confirm single melting peaks, strong correlations of technical replicates, and to calculate their efficiencies (in accordance with (Sinton, Finlay, and Lynch 1999)). The relative expression (or Normalized Relative Quantity (NRQ)) of each gene of interest in each experimental sample was calculated according to (Hellemans et al. 2008), which uses the geometric mean of two reference genes (here: *atpB* and *recA*) to normalize the probe of interest within each sample, and further calculates the fold-change in relative expression compared to a wildtype strain. The "non-targeting" *rfp*+ strain was considered the wildtype for the normalization of all other strains.

RNA-seq and 5'-end mapping

Cultures were grown and RNA extracted as described for RT-qPCR above.

RNA-seq library prep

1 ug of purified RNA was fragmented at 95°C for 7 minutes in 1x T4 RNA Ligase buffer (NEB) with an equal volume of 2X alkaline fragmentation buffer (0.6 volumes of 100 mM Na2CO3 plus 4.4 volumes of 100 mM NaHCO3. After 3'end healing with PNK (NEB) in T4 RNA ligase buffer for one hour, 3' ligation to a pre-adenylated, barcoded TruSeq R1 adapter with 5 random bases at its 5' end was performed overnight. The barcoded samples were then pooled and run onto a 6% TBE-Urea gel for size selection (>15nt insert size), eluted and ethanol precipitated before performing ribosomal RNA subtraction (RiboZero). Reverse transcription with SuperScript IV (Invitrogen) was performed using a TruSeq R1 RT primer, and followed by ligation of the TruSeq R2 adapter to the 3'end of the cDNA overnight, prior to another gel size selection as described above. A final PCR of the library was performed with indexed TruSeq PCR primers to add the index and P5/P7 flowcell adapters, followed by gel extraction, precipitation and a BioAnalyzer (Agilent) run for quality control before sequencing on a HiSeq4000 platform.

RPKM calculations

The indexed raw sequencing data was demultiplexed according to their R1 barcodes and the degenerate linker sequence was clipped using a custom script. Mapping of individual reads to the genome of *E. coli* (GenBank ID U00096.3) was performed with STAR (STAR: ultrafast universal RNA-seq aligner, (Dobin et al. 2013)), followed by read counting and calculation of RPKM for individual genome regions according to gene annotations from assembly ASM584v2.

Complementation of CRISPRi strains and quantification of chemical phenotype

CRISPRi strains were constructed that were complemented with either *gfp* or an essential gene of interest under control of an arabinose-inducible promoter in pBAD24 (Guzman et al. 1995). For each complementing essential gene, the toxicity of over-expression was first assessed by growing in 10-fold dilutions of arabinose from 0.2% to 0.0002% and monitoring OD600, and the maximum tolerated (without causing growth inhibition) concentration was used in subsequent experiments. For strains in which the CRISPRi-targeted gene was complemented, the complementation allele was mutated to preserve protein sequence but disrupt CRISPRi recognition by mutating either the PAM sequence or 2nt in the seed region (first 7nt adjacent to PAM). These mutations were made using quick-change mutagenesis.

CRISPRi stains complemented with either an essential gene or with *gfp* were grown in competition with an RFP-expressing strain (*nfsA::rfp-kan*) complemented with *gfp*. Competitions were done in the presence of ampicillin (to maintain the complementation plasmid), 0.01mM IPTG to induce CRISPRi knockdown, the previously determined maximum tolerated arabinose concentration (0.0002% or 0.2% arabinose), and with or without sub-inhibitory concentrations of vancomycin (40ug/ml).

Strains were first grown to mid-log phase in the presence of ampicillin and arabinose, and then diluted to 0.00025 OD with the competitor strain in the specified competition conditions, in 300ul in deep 96 well plates, and grown for a total of 4.5hrs (37C, 900RPM). Each competition well was then diluted and plates for CFU selecting for gentamicin resistance. 2-3 competition replicates were used for each experimental strain.

Growth curves to validate azidothymidine sensitivity

CRISPRi strains and those combined with the *sulA* deletion were grown in 4ml LB from single colonies, and diluted back to OD600=0.005 in 150ul LB with or without 0.025ng/ml

azidothymidine. The volumes were grown in microplates (Corning #3631) for 10hr in a Biotek Synergy H4 Microplate reader at 37C with fast shaking and measuring OD600 every 6min.

FIGURES

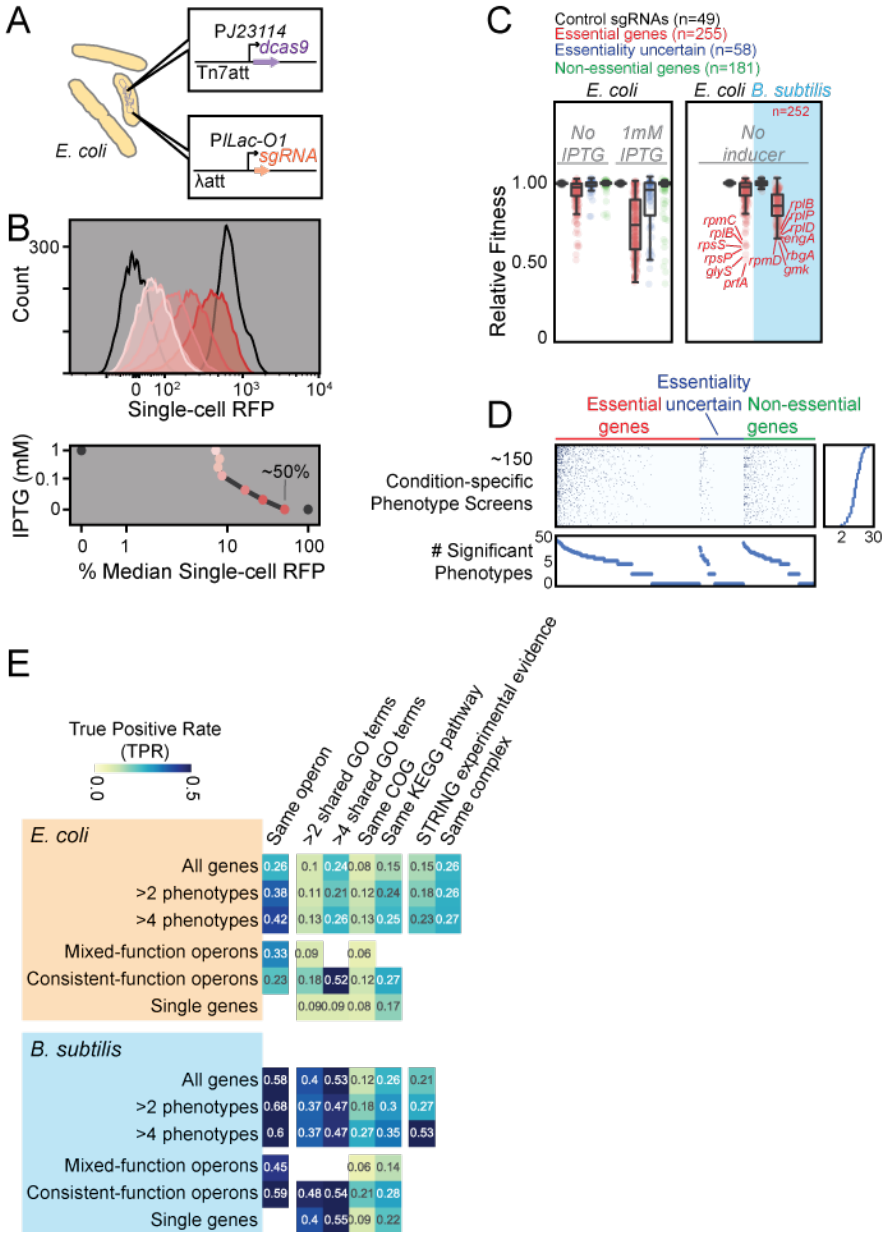


Figure 2.1. Chromosomal CRISPRi in *E. coli* enables the knockdown of essential genes and chemical phenotyping. (A) Diagram of *E. coli* chromosomal integration sites for *dcas9* from *Streptococcus pyogenes* and the chimeric sgRNA and their respective promoters. (B) Distribution of single-cell RFP fluorescence values for a chromosomal CRISPRi system as in (A) targeting chromosomal *rfp* and grown with titrating IPTG, compared to cells with no RFP (black line, left) and cells with no sgRNA (black line, right). Lower panel depicts the median single-cell RFP values as a percent of no-sgRNA control, compared to the concentration of IPTG. (C) Relative fitness of each CRISPRi strain grown in a pooled experiment, grown with or without 1mM IPTG for 15 generations. Strains are separated into categories based on the essentiality of their target gene: essential, non-essential, or essentiality uncertain. Right panel depicts the uninduced ("No IPTG") data for targeting essential genes in *E. coli* compared to the data from

(Hawkins et al. 2019) in which essential *B. subtilis* genes are targeted by slight knockdown (as in (Peters et al. 2016)) without induction. (D) Distribution of significant phenotypes (sensitive and resistant, FDR<0.05) from the chemical screen for the ~150 conditions tested. CRISPRi strains are separated into categories as in (C). Lower panel shows the sums of significant phenotypes for each CRISPRi strain, and right panel shows the sums of significant phenotypes for each condition. (E) True positive rates (TPR) at FPR=0.05 for all ROC analyses using the absolute value of correlation (|Pearson *r*|) between strains' phenotypic signatures and evaluated on the datasets of functional interaction described at the top. CRISPRi strains without at least one significant phenotype were excluded. Top panel shows the results for the chemical screen of the *E. coli* CRISPRi library. Bottom panel shows the results for the chemical screen of the *B. subtilis* CRISPRi library (Peters et al. 2016).

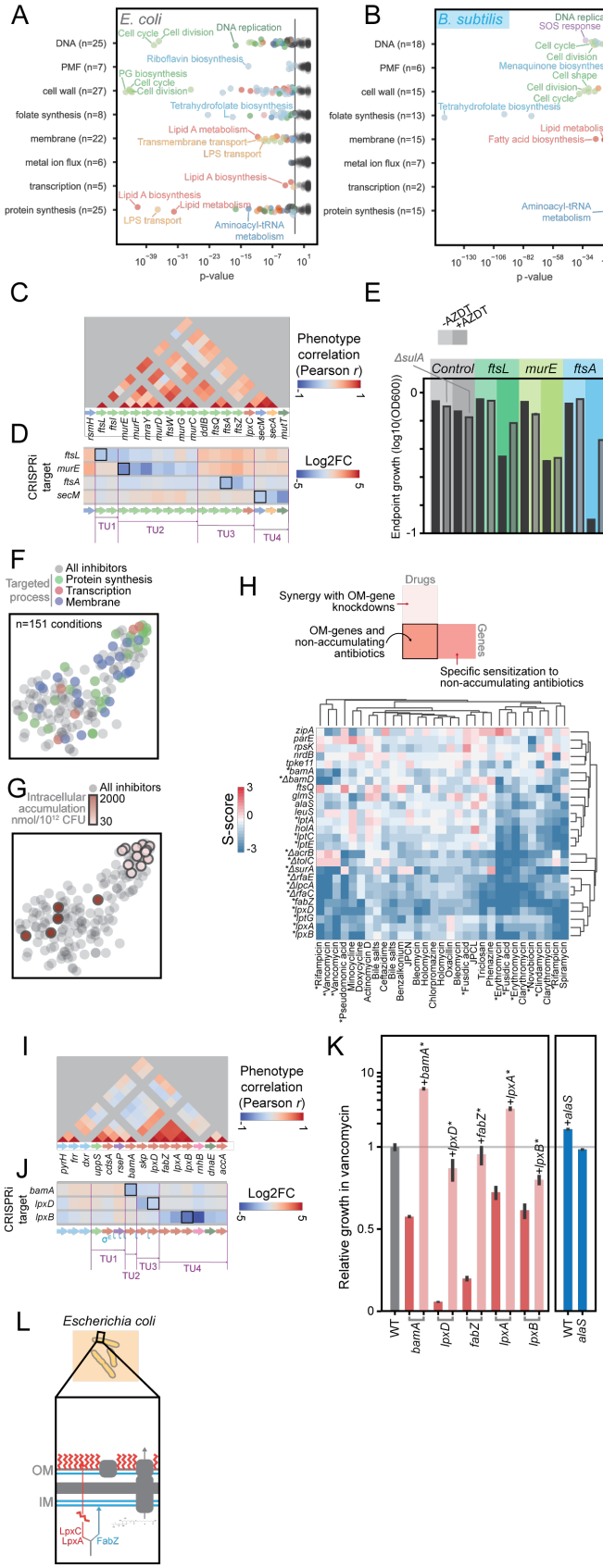
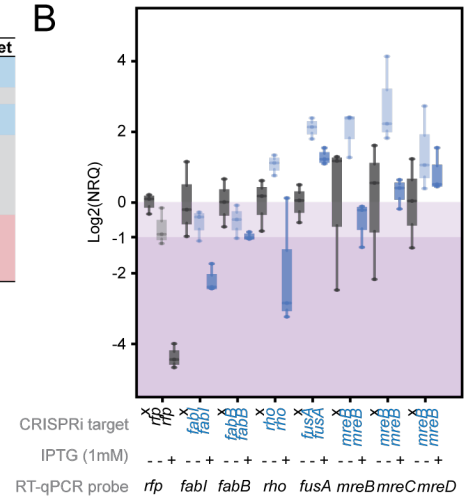

A
E. coli
DNA (n=25)
PMF (n=7)
cell wall (n=27)
folate synthesis (n=8)
membrane (n=22)
metal ion flux (n=6)
transcription (n=5)
protein synthesis (n=25)
p-value
B
B. subtilis
DNA (n=18)
PMF (n=6)
cell wall (n=15)
folate synthesis (n=13)
membrane (n=15)
metal ion flux (n=7)
transcription (n=2)
protein synthesis (n=15)
p-value
C
Phenotype correlation (Pearson r)
D
CRISPRi target
Log2FC
TU1
TU2
TU3
TU4
E
+AZDT
+PZDT
ΔsulA
Control
ftsL
murE
ftsA
Endpoint growth (log10(OD600))
F
Targeted process
All inhibitors
Protein synthesis
Transcription
Membrane
n=151 conditions
G
Intracellular accumulation nmol/10^12 CFU
All inhibitors
2000
30
H
Drugs
Synergy with OM-gene knockdowns
OM-genes and non-accumulating antibiotics
Genes
Specific sensitization to non-accumulating antibiotics
S-score
I
Phenotype correlation (Pearson r)
J
CRISPRi target
Log2FC
TU1
TU2
TU3
TU4
K
Relative growth in vancomycin
WT
bamA
lpxD
fabZ
lpxA
lpxB
alaS
L
Escherichia coli
OM
IM
LpxC
LpxA
FabZ

Figure 2.2. Cross-species comparison of sensitivity patterns reveals shared and Gram-negative-specific signatures. (A) Enrichment p-value (hypergeometric test, Bonferroni corrected) of GO biological process terms for which *E. coli* CRISPRi strains targeting genes annotated with that term had an enrichment of significant sensitivity phenotypes, for each group of stresses as categorized on the y-axis. X-axis shows the p-value of each enrichment. Each significantly enriched GO term is colored by a general functional categorization. (B) Enrichments as in (A) for significant sensitivity phenotypes in *B. subtilis* CRISPRi strains. (C) Correlation (Pearson *r*) between chemical phenotype signatures of genes in close genomic proximity with functions related to cell wall biosynthesis and cell division. (D) Change in expression (Log2FC) for genes of interest caused by CRISPRi targeting of genes selected from the cluster in (C). Independently repressible units are demarcated by purple lines. (E) Growth in the presence of sub-MIC azidothymidine of *E. coli* CRISPRi strains and their *ΔsulA* variants. Growth is measured by the median OD600 measurement following 6-7hrs of growth and averaged across two biological replicates. (F) t-SNE clustering of all conditions using the phenotypes of *E. coli* CRISPRi strains targeting genes with OM-related functions. Specific categories of conditions are colored as indicated. (G) t-SNE clustering as in (F) with conditions colored by their intracellular concentration as measured in (Richter et al. 2017). (H) Expanded analysis of outer membrane permeable mutants and their phenotypes. Heatmap showing S-scores of strains sensitized to non-accumulating antibiotics. Genes with OM-related functions are indicated with asterisks and are as used in (F-G). Non-accumulating antibiotics (<300nmol/10^{12} CFU) are indicated with asterisks. Additional genes are those with significantly enriched sensitivities to non-accumulating antibiotics (hypergeometric test, $p<0.05$). Additional conditions are those that were highly correlated with at least one non-accumulating antibiotic condition. (I) Correlation (Pearson *r*) between chemical phenotype signatures of genes in close genomic proximity with functions related to outer membrane biogenesis. (J) Change in expression (Log2FC) for genes of interest caused by CRISPRi targeting of genes selected from the cluster in (I). Independently repressible units are demarcated by purple lines, and annotated σ^E promoters are indicated in blue. (K) Relative growth in sub-MIC vancomycin of *E. coli* CRISPRi strains and their complemented variants, in competition against wildtype. Relative growth is assessed by CFU/ml and reflects the fraction of growth in vancomycin compared to untreated, relative to a wildtype control. (L) Schematic of the metabolic branch point between fatty acid biosynthesis and Lipid A biosynthesis, governed by the activities of FabZ and LpxA/LpxC, respectively.

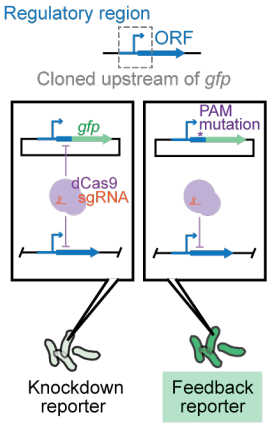
A

Drug	*E. coli* target	*B. subtilis* target
Novobiocin	*parE gyrAB parC*	*parCE gyrAB*
Pseudomonic acid	*ileS*	*ileS*
Triclosan	*fabI*	*n.t.*
Cerulenin	*fabB*	*fabF*
Sulfamonomethoxine	*folP*	*sul*
Sulfamethizole	*folP*	*n.t.*
5-fluorouridine	*thyA*	*n.t.*
Cinoxacin	*gyrAB parCE*	*n.t.*
D,L-serine hydroxamate	*serS*	*n.t.*
Bicyclomycin	*rho*	*n.t.*
Ciprofloxacin	*gyrAB parCE*	*gyrAB parCE*
Fusidic acid	*fusA*	*fusA*
A22	*mreB*	*mreB*
Rifampicin	*rpoBC*	*rpoBC*

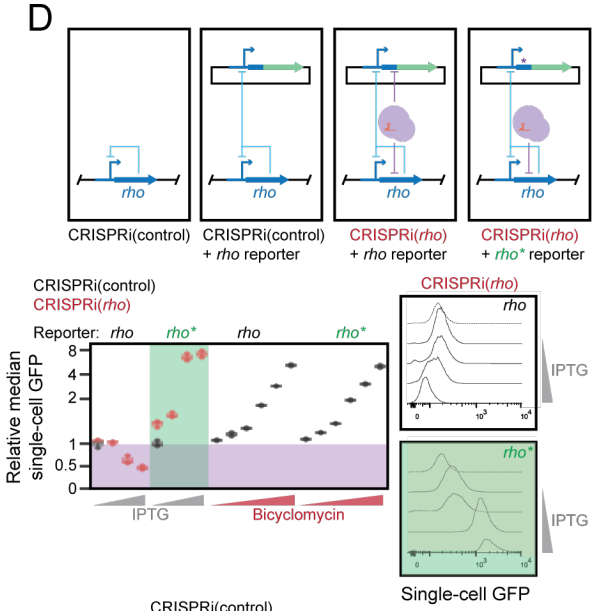
B



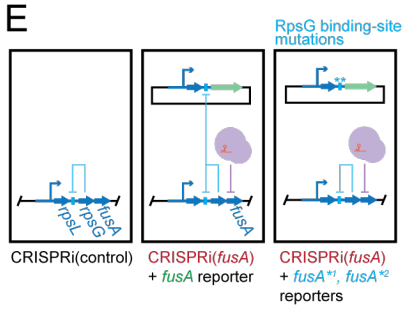
C



D



E



F

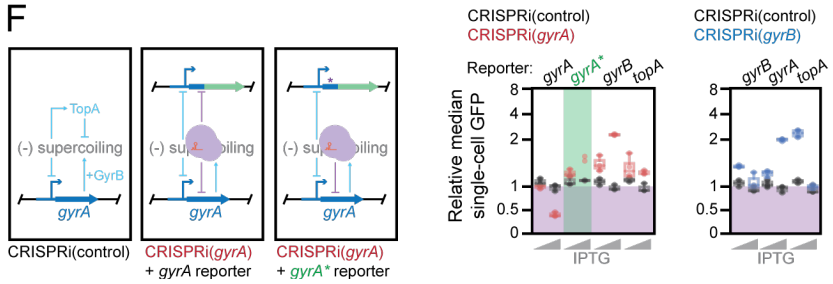


Figure 2.3. Predicted direct target interactions reveal pervasive feedback controlling essential gene expression and triggered by CRISPRi targeting. (A) Predicted direct target interactions in the *E. coli* and *B. subtilis* CRISPRi chemical genetic screens. All direct target pairs are listed and are shown in blue if the direct target knockdown was significantly sensitized (FDR<0.05) and in red if not. n.t. not tested. (B) Transcript levels of CRISPRi targeted genes measured by RT-qPCR. Wildtype levels of expression, uninduced CRISPRi levels (no IPTG) and maximally induced (1mM IPTG) are shown for each targeted gene. Transcript levels are shows as Normalized Relative Quantities (NRQ) and the distribution of 2-3 biological replicates are shown as boxplots. (C) Schematic of the dual-reporter system to quantify CRISPRi knockdown and feedback at the single-cell level. (D) Schematic describing negative autoregulation of *rho* and the knockdown/feedback reporter experiments. Lower panel shows the relative median single-cell GFP fluorescence for the strains and reporters as indicated. Induction of CRISPRi by IPTG or induction of feedback by drug treatment are each shown along the lower x-axis. Lower right panels show the distribution of single-cell GFP values for the *rho* knockdown and feedback reporters. (E) as in (D) for evaluation of *fusA* feedback regulation. (F) as in (D-E) for evaluation of *gyrA* and *gyrB* feedback regulation. All reporters in (F) are chromosomally integrated at *xylA*.

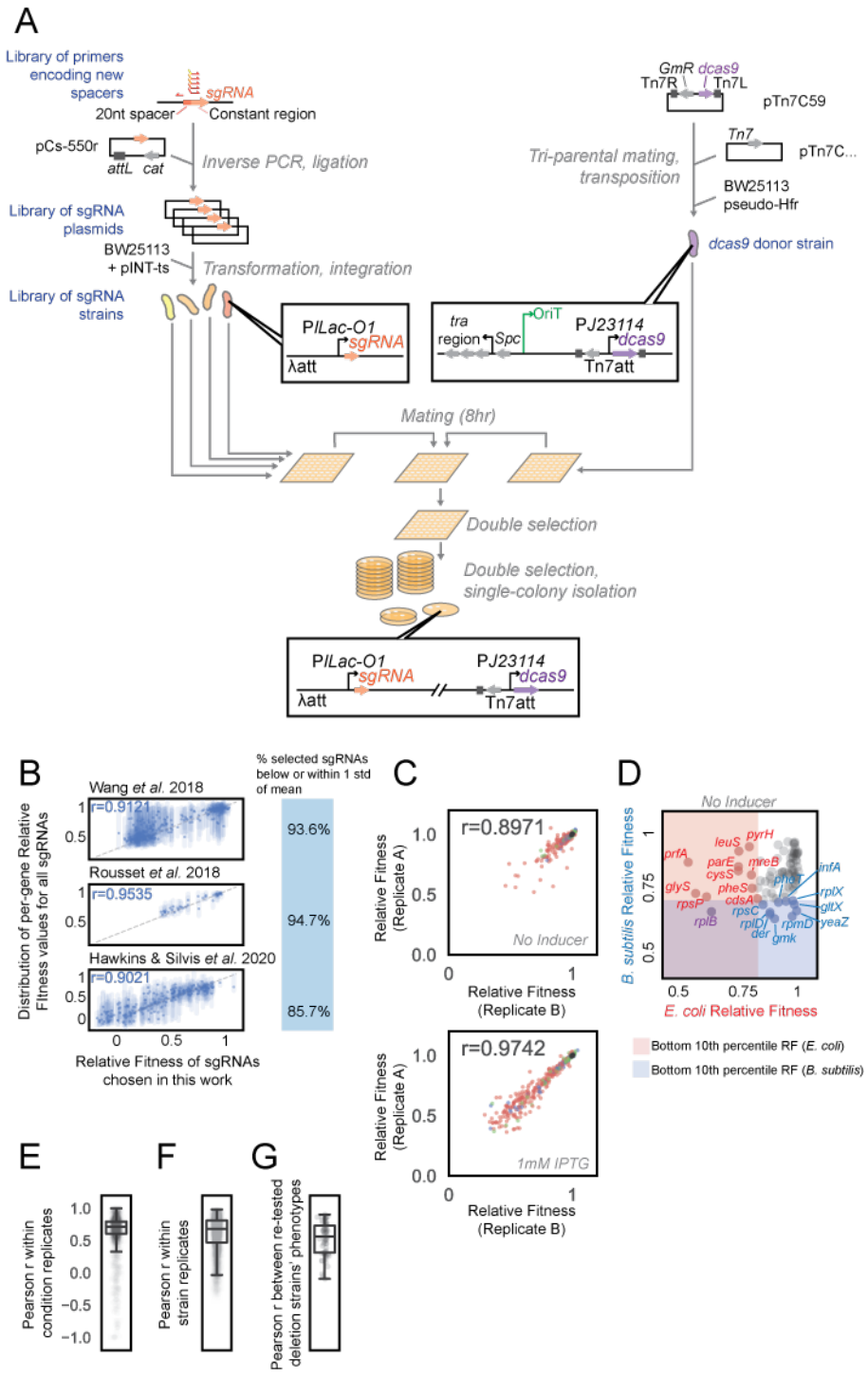


Figure 2.4. Construction and characterization of the *E. coli* essential gene CRISPRi library. (A) Workflow of the construction of the *E. coli* CRISPRi library. New spacers are cloned in parallel into the plasmid pCs-550r by iPCR, followed by ligation. These plasmids are competent to express the full length sgRNA under control of the IPTG-inducible promoter P*lLac-O1*. In parallel, each plasmid is integrated into the λatt site in the chromosome of a recipient wildtype (BW25113) cell using the helper plasmid pINT-ts, and selecting for chloramphenicol resistance. These strains are then arrayed in 96-well plates, and comprise the "sgRNA library".

The right branch of the workflow describes the construction of a *dcas9* donor strain which is competent to transfer *dcas9* in single copy to a recipient cell using conjugation. *dcas9* was cloned into the plasmid pTn7C59 inside a cassette recognized by the site-specific transposase Tn7. The plasmid pTn7C59 and the plasmid pTn7C1 (expressing Tn7) are each transferred to a recipient pseudo-Hfr strain (Typas et al. 2008) in a tri-parental mating strategy (Peters et al. 2019), resulting in integration of the *dcas9* cassette at the Tn7att in the recipient chromosome. The Tn7att site is adjacent to the origin of transfer (OriT) in the *dcas9* donor strain, allowing it to transfer the *dcas9* cassette but not the conjugative machinery (*tra* region) to a recipient cell. Below outlines the mating protocol, the mating and first selection steps of which take place on high-density array agar pads, and the final selection step using single colony isolation. (B) The relative fitness (RF) of sgRNAs chosen for the *E. coli* CRISPRi library and their activities compared to other sgRNAs targeting the same essential gene in three large-scale screens (Wang et al. 2018; Rousset et al. 2018; Hawkins et al. 2019). All fitness values are shown as relative fitness, with the mean fitness among all sgRNAs targeting the same gene shown as point in the line representing the full distribution. Right of the plots details the percentage of sgRNAs selected for this library whose RF values fall within or below 1 standard deviation of the mean RF for all sgRNAs targeting that gene. (C) Relative fitness measurements are reproducible. Relative fitness is shown for two biological replicates in uninduced “No inducer” and induced “1mM IPTG” conditions. Points are colored according to their essentiality designation as in Fig. 2.1C. (D) Relative fitness of high-confidence essential homologs in uninduced conditions in *E. coli* and *B. subtilis*. Essential homologs falling in the bottom 10^{th} percentile of essential homologs in either species are labeled. (E) Reproducibility of the chemical-genetic screen in *E. coli*. The distribution of correlations among all condition replicate pairs (≥4 replicates per condition, 151 conditions); median Pearson r=0.711. (F) The distribution of correlations among all strain replicate pairs (2-4 replicates per strain, 479 strains); median Pearson r=0.680. (G) The distribution of correlations for the deletion strains screened among all matched conditions with (Nichols et al. 2011) (91 strains, 70 overlapping conditions); Pearson r=0.565.

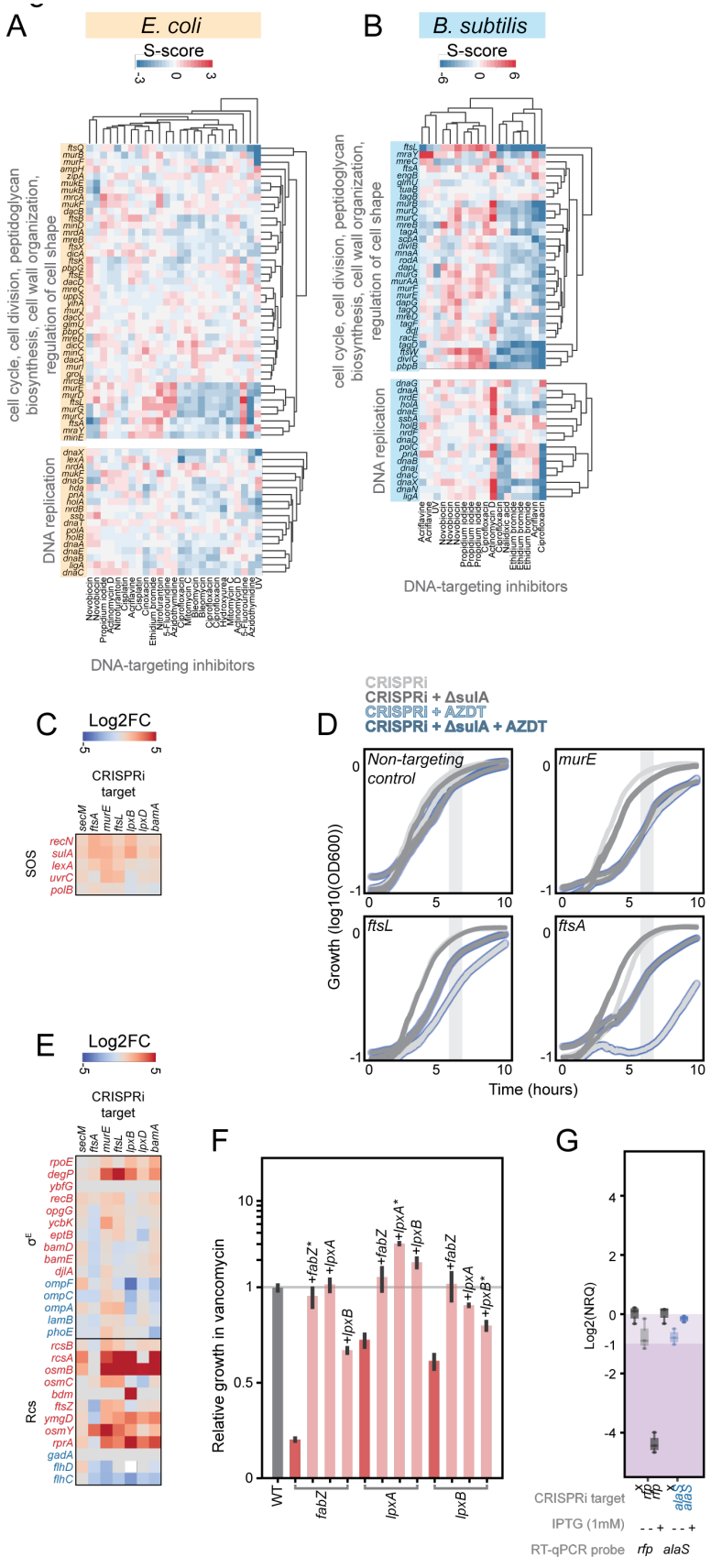

A
E. coli
S-score
cell cycle, cell division, peptidoglycan biosynthesis, cell wall organization, regulation of cell shape
DNA replication
DNA-targeting inhibitors
B
B. subtilis
S-score
cell cycle, cell division, peptidoglycan biosynthesis, cell wall organization, regulation of cell shape
DNA replication
DNA-targeting inhibitors
C
Log2FC
CRISPRi target
SOS
recN
sulA
lexA
uvrC
polB
D
CRISPRi
CRISPRi + ΔsulA
CRISPRi + AZDT
CRISPRi + ΔsulA + AZDT
Non-targeting control
murE
ftsL
ftsA
Growth (log10(OD600))
Time (hours)
E
Log2FC
CRISPRi target
σE
Rcs
F
Relative growth in vancomycin
WT
fabZ
lpxA
lpxB
G
Log2(NRQ)
CRISPRi target
IPTG (1mM)
RT-qPCR probe
rfp
alaS

Figure 2.5. Shared and Gram-negative specific sensitivity patterns from the chemical-genetic screens. (A) S-scores of all cell division and cell wall biosynthesis gene knockdowns (upper) and DNA replication genes (lower) in *E. coli* for conditions causing DNA damage. Upper shows strains and conditions clustered hierarchically using Euclidean distance. Lower shows conditions in the same order as above, and strains clustered. (B) as in (A) for S-scores of *B. subtilis* strains in which cell division or cell wall biosynthesis genes (upper) or DNA replication genes (lower) are targeted. Data as in (Peters et al. 2016). (C) RNA-seq of outer membrane, cell division, and cell wall biosynthesis genes demonstrates no strong activation of SOS response. Log2 fold change (Log2FC) in expression compared to a CRISPRi control strain is shown for SOS-induced genes. (D) Growth curves in the presence (blue outline) or absence (no outline) of sub-inhibitory azidothymidine. Each panel shows one CRISPRi target, alone or in combination with a sulA deletion (Δ*sulA*). Greyed section highlights the timepoints used to estimate end-point growth in Fig. 2.2E. (E) as in (C) for genes regulated by σ^E and Rcs pathways. (F) Complementation of vancomycin sensitivity as in Fig. 2.2K. (G) RT-qPCR quantification of target transcript levels in an *rfp* targeting and *alaS* targeting strain.

TABLES

Table 2.1. Strains and primers used in this study (including CRISPRi libraries). Supplemental file.

Table 2.2. Growth phenotypes from pooled screen. Supplemental file.

Table 2.3. Chemical screen details and phenotypes. Supplemental file.

REFERENCES

Arjes, Heidi A, Allison Kriel, Nohemy A Sorto, Jared T Shaw, Jue D Wang, and Petra Anne Levin. 2014. “Failsafe Mechanisms Couple Division and DNA Replication in Bacteria.” *Current Biology* 24 (18): 2149–55. https://doi.org/10.1016/j.cub.2014.07.055.

Baba, Tomoya, Takeshi Ara, Miki Hasegawa, Yuki Takai, Yoshiko Okumura, Miki Baba, Kirill A Datsenko, Masaru Tomita, Barry L Wanner, and Hirotada Mori. 2006. “Construction of Escherichia Coli K-12 in-Frame, Single-Gene Knockout Mutants: The Keio Collection.” *Molecular Systems Biology* 2. https://doi.org/10.1038/msb4100050.

Bikard, David, Wenyan Jiang, Poulami Samai, Ann Hochschild, Feng Zhang, and Luciano A Marraffini. 2013. “Programmable Repression and Activation of Bacterial Gene Expression Using an Engineered CRISPR-Cas System.” *Nucleic Acids Research* 41 (15). https://doi.org/10.1093/nar/gkt520.

Brochado, Ana Rita, Anja Telzerow, Jacob Bobonis, Manuel Banzhaf, André Mateus, Joel Selkrig, Emily Huth, et al. 2018. “Species-Specific Activity of Antibacterial Drug Combinations.” *Nature* 559 (7713): 259–63. https://doi.org/10.1038/s41586-018-0278-9.

Campos, Manuel, Sander K Govers, Irnov Irnov, Genevieve S Dobihal, François Cornet, and Christine Jacobs-Wagner. 2018. “Genomewide Phenotypic Analysis of Growth, Cell Morphogenesis, and Cell Cycle Events in Escherichia Coli.” *Molecular Systems Biology* 14 (6): e7573. https://doi.org/10.15252/msb.20177573.

Chen, Yaodong, Sara L Milam, and Harold P Erickson. 2012. “SulA Inhibits Assembly of FtsZ by a Simple Sequestration Mechanism.” *Biochemistry* 51 (14): 3100–3109. https://doi.org/10.1021/bi201669d.

Choi, Kyoung-Hee, and Herbert P Schweizer. 2006. “Mini-Tn7 Insertion in Bacteria with Single AttTn7 Sites: Example Pseudomonas Aeruginosa.” https://doi.org/10.1038/nprot.2006.24.

Collins, Sean R, Maya Schuldiner, Nevan J Krogan, and Jonathan S Weissman. 2006. “A Strategy for Extracting and Analyzing Large-Scale Quantitative Epistatic Interaction Data.”

Genome Biology 7 (7). https://doi.org/10.1186/gb-2006-7-7-r63.

Conway, Tyrrell, James P Creecy, Scott M Maddox, Joe E Grissom, Trevor L Conkle, Tyler M Shadid, Jun Teramoto, et al. 2014. “Unprecedented High-Resolution View of Bacterial Operon Architecture Revealed by RNA Sequencing.” *MBio* 5 (4). https://doi.org/10.1128/mBio.01442-14.

Dajkovic, Alex, Amit Mukherjee, and Joe Lutkenhaus. 2008. “Investigation of Regulation of FtsZ Assembly by SulA and Development of a Model for FtsZ Polymerization.” *Journal of Bacteriology* 190 (7): 2513–26. https://doi.org/10.1128/JB.01612-07.

Dobin, Alexander, Carrie A Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Doerrler, William T., and Christian R.H. Raetz. 2005. “Loss of Outer Membrane Proteins without Inhibition of Lipid Export in an Escherichia Coli YaeT Mutant.” *Journal of Biological Chemistry* 280 (30): 27679–87. https://doi.org/10.1074/jbc.M504796200.

Domenech, Arnau, Ana Rita Brochado, Vicky Sender, Karina Hentrich, Birgitta Henriques-Normark, Athanasios Typas, and Jan-Willem Veening. 2019. “Fighting the Spread of Antibiotic Resistance with Bacterial Competence Inhibitors.” *BioRxiv*, 683920. https://doi.org/10.1101/683920.

Elwell,’, Lynn P, Robert Ferone, G Andrew Freeman, James A Fyfe, John A Hill, Paul H Ray, Cynthia A Richards,’, et al. 1987. “Antibacterial Activity and Mechanism of Action of 3’-Azido-3’-Deoxythymidine (BW A509U).” *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*. Vol. 31. http://aac.asm.org/.

Fields, Rachel N., and Hervé Roy. 2018. “Deciphering the TRNA-Dependent Lipid Aminoacylation Systems in Bacteria: Novel Components and Structural Advances.” *RNA Biology* 15 (4–5): 480–91. https://doi.org/10.1080/15476286.2017.1356980.

Galloway, S. M., and C. R.H. Raetz. 1990. “A Mutant of Escherichia Coli Defective in the First Step of Endotoxin Biosynthesis.” *Journal of Biological Chemistry* 265 (11): 6394–6402.

Gellert, M., R. Menzel, K. Mizuuchi, M. H. O’Dea, and D. I. Friedman. 1982. “Regulation of DNA Supercoiling in Escherichia Coli.” *Cold Spring Harbor Symposia on Quantitative Biology* 47 (2): 763–67.

“GitHub - Traeki/Sgrna_design.” n.d. Accessed February 10, 2020. https://github.com/traeki/sgrna_design.

Goodall, Emily, Ashley Robinson, Iain Johnston, Sara Jabbari, Keith Turner, Adam Cunningham, Peter Lund, Jeffrey Cole, and Ian Henderson. 2017. “The Essential Genome of Escherichia Coli K-12.” *Mbio* 9 (1): 237842. https://doi.org/10.1101/237842.

Grundstrom, T., S. Normark, and K. E. Magnusson. 1980. “Overproduction of Outer Membrane Protein Suppresses EnvA-Induced Hyperpermeability.” *Journal of Bacteriology* 144 (3): 884–90.

Guzman, Luz-Maria, Dominique Belin, Michael J Carson, and Jon Beckwith. 1995. “Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P BAD Promoter.” *JOURNAL OF BACTERIOLOGY* 177 (14): 4121–30. https://doi.org/0021-9193/95/$04.00+0.

Haldimann, Andreas, and Barry L Wanner. 2001. “Conditional-Replication , Integration , Excision , and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria.” *JOURNAL OF BACTERIOLOGY* 183 (21): 6384–93. https://doi.org/10.1128/JB.183.21.6384.

Hardy, Christine D, and Nicholas R Cozzarelli. 2003. “Alteration of Escherichia Coli Topoisomerase IV to Novobiocin Resistance.” *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY* 47 (3): 941–47. https://doi.org/10.1128/AAC.47.3.941-947.2003.

Hawkins, John S, Melanie R Silvis, Byoung-Mo Koo, Jason M Peters, Marco Jost, Cameron C Hearne, Jonathan S Weissman, Horia Todor, and Carol A Gross. 2019. “Modulated

Efficacy CRISPRi Reveals Evolutionary Conservation of Essential Gene Expression-Fitness Relationships in Bacteria." *BioRxiv*, 805333. https://doi.org/10.1101/805333.

Hellemans, Jan, Geert Mortier, Anne De Paepe, Frank Speleman, and Jo Vandesompele. 2008. "QBase Relative Quantification Framework and Software for Management and Automated Analysis of Real-Time Quantitative PCR Data." *Genome Biology* 8 (2). https://doi.org/10.1186/gb-2007-8-2-r19.

Hurdle, Julian Gregston, Alexander John O'Neill, and Ian Chopra. 2005. "Prospects for Aminoacyl-TRNA Synthetase Inhibitors as New Antimicrobial Agents." *Antimicrobial Agents and Chemotherapy*. https://doi.org/10.1128/AAC.49.12.4821-4833.2005.

Jaffe, Aline, R. D'Ari, and Victor Norris. 1986. "SOS-Independent Coupling between DNA Replication and Cell Division in Escherichia Coli." *Journal of Bacteriology* 165 (1): 66–71. https://doi.org/10.1128/jb.165.1.66-71.1986.

Klein, Stefanie, Carlos Lorenzo, Sonja Hoffmann, Johannes M. Walther, Sonja Storbeck, Tanja Piekarski, Bryan J. Tindall, Victor Wray, Manfred Nimtz, and Jürgen Moser. 2009. "Adaptation of Pseudomonas Aeruginosa to Various Conditions Includes TRNA-Dependent Formation of Alanyl-Phosphatidylglycerol." *Molecular Microbiology* 71 (3): 551–65. https://doi.org/10.1111/j.1365-2958.2008.06562.x.

Koo, Byoung-Mo, George Kritikos, Jeremiah D. Farelli, Horia Todor, Kenneth Tong, Harvey Kimsey, Ilan Wapinski, et al. 2017. "Construction and Analysis of Two Genome-Wide Deletion Libraries for Bacillus Subtilis." *Cell Systems*. http://www.cell.com/cell-systems/pdf/S2405-4712(16)30447-1.pdf.

Kritikos, George, Manuel Banzhaf, Lucia Herrera-Dominguez, Alexandra Koumoutsi, Morgane Wartel, Matylda Zietek, and Athanasios Typas. 2017. "A Tool Named Iris for Versatile High-Throughput Phenotyping in Microorganisms." *Nature Microbiology* 2. https://doi.org/10.1038/nmicrobiol.2017.14.

Lalanne, Jean-benoît Benoît, James C. Taggart, Monica S. Guo, Lydia Herzel, Ariel Schieler,

and Gene-wei Wei Li. 2018. “Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry.” *Cell* 173 (3): 1–28. https://doi.org/10.1016/j.cell.2018.03.007.

Larson, Matthew H, Luke A Gilbert, Xiaowo Wang, Wendell A Lim, Jonathan S Weissman, and Lei S Qi. 2013. “CRISPR Interference (CRISPRi) for Sequence-Specific Control of Gene Expression.” *Nature Protocols* 8. https://doi.org/10.1038/nprot.2013.132.

Lee, Henry H, Nili Ostrov, Brandon G. Wong, Michaela A. Gold, Ahmad S. Khalil, and George M. Church. 2019. “Functional Genomics of the Rapidly Replicating Bacterium Vibrio Natriegens by CRISPRi.” *Nature Microbiology*. https://doi.org/10.1038/s41564-019-0423-8.

Lutz, Rolf, and Hermann Bujard. 1997. “Independent and Tight Regulation of Transcriptional Units in Escherichia Coli via the LacR/O, the TetR/O and AraC/I 1 -I 2 Regulatory Elements.” *Nucleic Acids Research* 25 (6): 1203–10. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC146584/pdf/251203.pdf.

Maguin, Emmanuelle, Howard Brody, Charles W Hill, and R. D’Ari. 1986. “SOS-Associated Division Inhibition Gene SfiC Is Part of Excisable Element E14 in Escherichia Coli.” *Journal of Bacteriology* 168 (1): 464–66. https://doi.org/10.1128/jb.168.1.464-466.1986.

Malinverni, Juliana C, John Werner, Seokhee Kim, Joseph G Sklar, Daniel Kahne, Rajeev Misra, and Thomas J Silhavy. 2006. “YfiO Stabilizes the YaeT Complex and Is Essential for Outer Membrane Protein Assembly in Escherichia Coli.” *Molecular Microbiology* 61 (1): 151–64. https://doi.org/10.1111/j.1365-2958.2006.05211.x.

Mateus, André, Jacob Bobonis, Nils Kurzawa, Frank Stein, Dominic Helm, Johannes Hevler, Athanasios Typas, and Mikhail M Savitski. 2018. “Thermal Proteome Profiling in Bacteria: Probing Protein State in Vivo.” *Molecular Systems Biology* 14 (7): e8242. https://doi.org/10.15252/msb.20188242.

Matsumoto, Yoshihiro, Katsuya Shigesada, Masanori Hirano, Mutsuo Imai, S Brown, B Albrechtsen, S Pedersen, et al. 1986. “Autogenous Regulation of the Gene for Transcription Termination Factor Rho in Escherichia Coli: Localization and Function of Its

Attenuators.” *JOURNAL OF BACTERIOLOGY*. Vol. 166.

Menzel, R., and Martin Gellert. 1987. “Fusions of the Escherichia Coli GyrA and GyrB Control Regions to the Galactokinase Gene Are Inducible by Coumermycin Treatment.” *Journal of Bacteriology* 169 (3): 1272–78. https://doi.org/10.1128/jb.169.3.1272-1278.1987.

Menzel, Rolf, and Martin Gellert. 1983. “Regulation of the Genes for Escherichia Coli DNA Gyrase: Homostatic Control of the DNA Supercoiling.” *Cell* 34: 105–13. https://pdf.sciencedirectassets.com/272196/1-s2.0-S0092867400X05272/1-s2.0-009286748390140X/main.pdf?x-amz-security-token=AgoJb3JpZ2luX2VjEFcaCXVzLWVhc3QtMSJHMEUCIQC7gH5Hph6qdDV4O%2FTD18HnT1g5kIHyIpgiav5gUOzAFAIgQIH0ZIQxS5UO78UPytO8pBz07agdUXRtV7QIc%2Fr3s.

Modell, Joshua W, Alexander C Hopkins, and Michael T Laub. 2011. “A DNA Damage Checkpoint in Caulobacter Crescentus Inhibits Cell Division through a Direct Interaction with FtsW.” *Genes & Development* 25 (12): 1328–43. https://doi.org/10.1101/gad.2038911.

Mohan, S., T. M. Kelly, S. S. Eveland, C. R.H. Raetz, and M. S. Anderson. 1994. “An Escherichia Coli Gene (FabZ) Encoding (3R)-Hydroxymyristoyl Acyl Carrier Protein Dehydrase. Relation to FabA and Suppression of Mutations in Lipid A Biosynthesis.” *Journal of Biological Chemistry* 269 (52): 32896–903.

Mukherjee, Amit, Chune Cao, and Joe Lutkenhaus. 1998. “Inhibition of FtsZ Polymerization by SulA, an Inhibitor of Septation in Escherichia Coli (CytokinesisGTPasetubulinpolymerization).” *Biochemistry*. Vol. 95. www.pnas.org.

Nazir, Aanisa, and Rajendran Harinarayanan. 2016. “Inactivation of Cell Division Protein FtsZ by SulA Makes Lon Indispensable for the Viability of a PpGpp 0 Strain of Escherichia Coli.” https://doi.org/10.1128/JB.00693-15.

Nichols, Robert J, Saunak Sen, Yoe Jin Choo, Pedro Beltrao, Matylda Zietek, Rachna Chaba, Sueyoung Lee, et al. 2011. “Phenotypic Landscape of a Bacterial Cell.” *Cell* 144 (1): 143–

56. https://doi.org/10.1016/j.cell.2010.11.052.

Nonejuie, Poochit, Michael Burkart, K. Pogliano, and Joe Pogliano. 2013. “Bacterial Cytological Profiling Rapidly Identifies the Cellular Pathways Targeted by Antibacterial Molecules.” *Proceedings of the National Academy of Sciences* 110 (40): 16169–74. https://doi.org/10.1073/pnas.1311066110.

Normark, S. 1970. “Genetics of a Chain-Forming Mutant of Escherichia Coli.” *Genet Res Camb* 16 (1970): 63–78.

Ogura, Teru, Koichi Inoue, Takashi Tatsuta, Toshinobu Suzaki, Kiyonobu Karata, Katherine Young, Lin Hui Su, et al. 1999. “Balanced Biosynthesis of Major Membrane Components through Regulated Degradation of the Committed Enzyme of Lipid A Biosynthesis by the AAA Protease FtsH (HflB) in Escherichia Coli.” *Molecular Microbiology* 31 (3): 833–44. https://doi.org/10.1046/j.1365-2958.1999.01221.x.

Patrick, Wayne M, Erik M Quandt, Dan B Swartzlander, and Ichiro Matsumura. 2007. “Multicopy Suppression Underpins Metabolic Evolvability.” *Molecular Biology and Evolution* 24 (12): 2716–22. https://doi.org/10.1093/molbev/msm204.

Peters, Jason M., Alexandre Colavin, Handuo Shi, Tomasz L. Czarny, Matthew H. Larson, Spencer Wong, John S. Hawkins, et al. 2016. “A Comprehensive, CRISPR-Based Functional Analysis of Essential Genes in Bacteria.” *Cell* 165 (6): 1493–1506. https://doi.org/10.1016/j.cell.2016.05.003.

Peters, Jason M., Melanie R. Silvis, Dehua Zhao, John S. Hawkins, Carol A. Gross, and Lei S. Qi. 2015. “Bacterial CRISPR: Accomplishments and Prospects.” *Current Opinion in Microbiology*. https://doi.org/10.1016/j.mib.2015.08.007.

Peters, Jason M, Byoung-Mo Koo, Ramiro Patino, Gary E Heussler, Cameron C Hearne, Jiuxin Qu, Yuki F Inclan, et al. 2019. “Enabling Genetic Analysis of Diverse Bacteria with Mobile-CRISPRi.” *Nature Microbiology*. https://doi.org/10.1038/s41564-018-0327-z.

Price, Allen C, Keum Hwa Choi, Richard J Heath, Zhenmei Li, Stephen W White, and Charles O

Rock. 2001. “Inhibition of β-Ketoacyl-Acyl Carrier Protein Synthases by Thiolactomycin and Cerulenin: Structure and Mechanism.” *Journal of Biological Chemistry* 276 (9): 6551–59. https://doi.org/10.1074/jbc.M007101200.

Putney, Scott D, and Paul Schimmel. 1981. “An Aminoacyl TRNA Synthetase Binds to a Specific DNA Sequence and Regulates Its Gene Transcription The Effect on Transcription of Proteins Other than Ala-TRNA l Beo.” *Nature* 291 (June): 25. https://www.nature.com/articles/291632a0.pdf.

Qi, Lei S, Matthew H Larson, Luke A Gilbert, Jennifer A Doudna, Jonathan S Weissman, Adam P Arkin, and Wendell A Lim. 2013. “Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression.” *Cell* 152 (5): 1173–83. https://doi.org/10.1016/j.cell.2013.02.022.

Rauch, Benjamin J, Melanie R Silvis, Judd F Hultquist, Christopher S Waters, Michael J. McGregor, Nevan J Krogan, and Joseph Bondy-Denomy. 2017. “Inhibition of CRISPR-Cas9 with Bacteriophage Proteins.” *Cell* 168 (1–2): 150-158.e10. https://doi.org/10.1016/j.cell.2016.12.009.

Reuß, Daniel R., Patrick Faßhauer, Philipp Joel Mroch, Inam Ul-Haq, Byoung Mo Koo, Anja Pöhlein, Carol A. Gross, Rolf Daniel, Sabine Brantl, and Jörg Stülke. 2019. “Topoisomerase IV Can Functionally Replace All Type 1A Topoisomerases in Bacillus Subtilis.” *Nucleic Acids Research* 47 (10): 5231–42. https://doi.org/10.1093/nar/gkz260.

Richter, Michelle F., and Paul J. Hergenrother. 2018. “The Challenge of Converting Gram-Positive-Only Compounds into Broad-Spectrum Antibiotics.” *Annals of the New York Academy of Sciences*, 1–21. https://doi.org/10.1111/nyas.13598.

Richter, Michelle F, Bryon S Drown, Andrew P Riley, Alfredo Garcia, Tomohiro Shirai, Riley L Svec, and Paul J Hergenrother. 2017. “Predictive Compound Accumulation Rules Yield a Broad-Spectrum Antibiotic.” *Nature*. https://doi.org/10.1038/nature22308.

Rock, Jeremy M., Forrest F. Hopkins, Alejandro Chavez, Marieme Diallo, Michael R. Chase,

Elias R. Gerrick, Justin R. Pritchard, et al. 2017. “Programmable Transcriptional Repression in Mycobacteria Using an Orthogonal CRISPR Interference Platform.” *Nature Microbiology* 2 (February): 1–9. https://doi.org/10.1038/nmicrobiol.2016.274.

Rojas, Enrique R., Gabriel Billings, Pascal D. Odermatt, George K. Auer, Lillian Zhu, Amanda Miguel, Fred Chang, Douglas B. Weibel, Julie A. Theriot, and Kerwyn Casey Huang. 2018. “The Outer Membrane Is an Essential Load-Bearing Element in Gram-Negative Bacteria.” *Nature* 559 (7715): 617–21. https://doi.org/10.1038/s41586-018-0344-3.

Rousset, Francois, Lun Cui, Elise Siouve, Christophe Becavin, Florence Depardieu, and David Bikard. 2018. “Genome-Wide CRISPR-DCas9 Screens in E. Coli Identify Essential Genes and Phage Host Factors.” *PLoS Genetics* 74 (4): 417–22. https://doi.org/10.1371/journal.pgen.1007749.

Ruiz, Natividad, Brian Falcone, Daniel Kahne, and Thomas J Silhavy. 2005. “Chemical Conditionality: A Genetic Strategy to Probe Organelle Assembly.” *Cell* 121 (2): 307–17. https://doi.org/10.1016/j.cell.2005.02.014.

Saito, Katsuyuki, Larry C Mattheakis, and Masayasu Nomura. 1994. “Post-Transcriptional Regulation of the Str Operon in Escherichia Coli Independent Translation.” *Biological Chemistry*, 111–24. https://doi.org/10.1016/S0022-2836(05)80020-8.

Saito, Katsuyuki, and Masayasu Nomura. 1994. “Post-Transcriptional Regulation of the Str Operon in Escherichia Coli: Structural and Mutational Analysis of the Target Site for Translational Repressor S7.” *Journal of Molecular Biology* 235 (1): 125–39. https://doi.org/10.1016/S0022-2836(05)80021-X.

Sampson, B. A., R. Misra, and S. A. Benson. 1989. “Identification and Characterization of a New Gene of Escherichia Coli K-12 Involved in Outer Membrane Permeability.” *Genetics* 122 (3): 491–501.

Sánchez-Gorostiaga, Citation, A Palacios, P Martínez-Arteaga, R Sánchez, M Casanova, and M Vicente. 2016. “Life without Division: Physiology of Escherichia Coli FtsZ-Deprived

Filaments.” *MBio* 7 (5): 1620–36. https://doi.org/10.1128/mBio.01620-16.

Shiver, Anthony L, Hendrik Osadnik, George Kritikos, Bo Li, Nevan Krogan, Athanasios Typas, and Carol A Gross. 2016. “A Chemical-Genomic Screen of Neglected Antibiotics Reveals Illicit Transport of Kasugamycin and Blasticidin S.” *PLoS Genetics* 12 (6). https://doi.org/10.1371/journal.pgen.1006124.

Sinton, Lester W., Rochelle K. Finlay, and Philippa A. Lynch. 1999. “The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments.” *Applied and Environmental Microbiology* 65 (8): 3605–13. https://doi.org/10.1373/clinchem.2008.112797.

Sperandeo, Paola, Fion K Lau, Andrea Carpentieri, Cristina De Castro, Antonio Molinaro, Gianni Dehò, Thomas J Silhavy, and Alessandra Polissi. 2008. “Functional Analysis of the Protein Machinery Required for Transport of Lipopolysaccharide to the Outer Membrane of Escherichia Coli †.” *JOURNAL OF BACTERIOLOGY* 190 (13): 4460–69. https://doi.org/10.1128/JB.00270-08.

Szafran, Marcin Jan, Martyna Gongerowska, Pawel Gutkowski, Jolanta Zakrzewska-Czerwińska, and Dagmara Jakimowicz. 2016. “The Coordinated Positive Regulation of Topoisomerase Genes Maintains Topological Homeostasis in Streptomyces Coelicolor.” *Journal of Bacteriology* 198 (21): 3016–28. https://doi.org/10.1128/JB.00530-16.

Thomason, Lynn C., Nina Costantino, and Donald L. Court. 2007. “ E. Coli Genome Manipulation by P1 Transduction .” *Current Protocols in Molecular Biology*, no. July: 1.17.1-1.17.8. https://doi.org/10.1002/0471142727.mb0117s79.

Thomason, Lynn C, James A Sawitzke, Xintian Li, Nina Costantino, and Donald L Court. 2014. “Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination.” *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.].* https://doi.org/10.1002/0471142727.mb0116s106.

Tse-Dinh, Yuk-Ching. 1985. “Regulation of the E. Coli DNA Top. I Gene by DNA Supercoiling.”

Nucleic Acids Research 44 (1): i–i. https://doi.org/10.1093/nar/gkv1387.

Tse-Dinh, Yuk Ching, and Rita K. Beran. 1988. “Multiple Promoters for Transcription of the Escherichia Coli DNA Topoisomerase I Gene and Their Regulation by DNA Supercoiling.” *Journal of Molecular Biology* 202 (4): 735–42. https://doi.org/10.1016/0022-2836(88)90554-2.

Typas, Athanasios, Robert J Nichols, Deborah A Siegele, Michael Shales, Sean R Collins, Bentley Lim, Hannes Braberg, et al. 2008. “High-Throughput, Quantitative Analyses of Genetic Interactions in E. Coli.” *Nature Methods* 5 (9): 781–87. https://doi.org/10.1038/NMETH.1240.

Vaara, Martti, and Marjatta Nurminen. 1999. “Outer Membrane Permeability Barrier in Escherichia Coli Mutants That Are Defective in the Late Acyltransferases of Lipid a Biosynthesis.” *Antimicrobial Agents and Chemotherapy* 43 (6): 1459–62. https://doi.org/10.1128/aac.43.6.1459.

Veening, Jan-Willem, and Xue Liu. 2017. “High-throughput CRISPRi Phenotyping Identifies New Essential Genes in Streptococcus Pneumoniae.” *Mol Syst Biol* 12 (888): 1–14. https://doi.org/10.15252/msb.

Vuorio, A, and Martti Vaara. 1992. “The Lipid A Biosynthesis Mutation LpxA2 of Escherichia Coli Results in Drastic Antibiotic Supersusceptibility.” *ANTIMICROBIAL AGENTS AND CHEMOTHERAPY*. http://aac.asm.org/.

Wang, Tianmin, Changge Guan, Jiahui Guo, Bing Liu, Yinan Wu, Zhen Xie, Chong Zhang, and Xin Hui Xing. 2018. “Pooled CRISPR Interference Screening Enables Genome-Scale Functional Genomics Study in Bacteria with Superior Performance-Net.” *Nature Communications* 9 (1). https://doi.org/10.1038/s41467-018-04899-x.

Wu, Tao, Juliana Malinverni, Natividad Ruiz, Seokhee Kim, Thomas J Silhavy, and Daniel Kahne. 2005. “Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in Escherichia Coli.” *Cell* 121 (2): 235–45.

https://doi.org/10.1016/j.cell.2005.02.015.

Yamamoto, Natsuko, Kenji Nakahigashi, Tomoko Nakamichi, Mihoko Yoshino, Yuki Takai, Yae Touda, Akemi Furubayashi, et al. 2009. “Update on the Keio Collection of Escherichia Coli Single-Gene Deletion Mutants.” *Molecular Systems Biology* 5. https://doi.org/10.1038/msb.2009.92.

Young, K., and L. L. Silver. 1991. “Leakage of Periplasmic Enzymes from EnvA1 Strains of Escherichia Coli.” *Journal of Bacteriology* 173 (12): 3609–14. https://doi.org/10.1128/jb.173.12.3609-3614.1991.

Zaslaver, Alon, Anat Bren, Michal Ronen, Shalev Itzkovitz, Ilya Kikoin, Seagull Shavit, Wolfram Liebermeister, Michael G Surette, and Uri Alon. 2006. “A Comprehensive Library of Fluorescent Transcriptional Reporters for Escherichia Coli.” *Nature Methods* 3 (8): 623–28. https://doi.org/10.1038/nmeth895.

Zechiedrich, E Lynn, Arkady B Khodursky, and Nicholas R Cozzarelli. 1997. “Topoisomerase IV, Not Gyrase, Decatenates Products of Site-Specific Recombination in Escherichia Coli.”

Zeng, Daina, Jinshi Zhao, Hak Suk Chung, Ziqiang Guan, Christian R.H. Raetz, and Pei Zhou. 2013. “Mutants Resistant to LpxC Inhibitors by Rebalancing Cellular Homeostasis.” *Journal of Biological Chemistry* 288 (8): 5475–86. https://doi.org/10.1074/jbc.M112.447607.

Chapter 3

Mismatch-CRISPRi reveals conserved expression-fitness relationships in bacteria

INTRODUCTION

Bacteria must optimize protein production to maximize survival and growth in constantly changing environments. Given the high energetic cost of protein synthesis, optimizing expression is particularly important for essential genes: although only ~5-10% of the genome, they constitute a disproportionate fraction (~50%) of the proteome (Lalanne et al., 2018) and insufficient expression is, by definition, fatal. Previous work using CRISPR interference (CRISPRi), hypomorphs, and promoter replacement revealed gene-, environment-, and antibiotic-specific fitness effects of altering essential gene expression (Bauer et al., 2015; Dekel and Alon, 2005; Eames and Kortemme, 2012; Johnson et al., 2019; Keren et al., 2016; Nichols et al., 2011), but the lack of a facile method for systematically perturbing bacterial gene expression has thus far prevented a comprehensive understanding of how bacteria optimize expression of their essential protein complement. CRISPRi, which represses bacterial transcription by targeting a catalytically dead Cas9 (dCas9) to an open reading frame using a complementary sgRNA, is an inducible and inherently barcoded system that has been used to perturb essential gene expression in its native context. However, tuning transcriptional repression by adjusting dCas9 or sgRNA abundance (Liu et al., 2017; Peters et al., 2016) is noisy and precludes the interrogation of multiple knockdown levels in a single experiment (Vigouroux et al., 2018).

Here we establish a species-independent approach for predictably titrating CRISPRi activity in bacteria using single mismatches in the base-pairing region of sgRNAs. Mismatched sgRNAs enable massively parallel interrogation of the fitness effects of many intermediate levels of CRISPRi efficacy across genes in a single pooled growth experiment. Building on previous studies of off-target and mismatched sgRNA activity (Gilbert et al., 2014; Jost et al., 2020; Vigouroux et al., 2018), we screened a comprehensive library of mismatched *gfp*-targeting sgRNAs in *Escherichia coli* and *Bacillus subtilis* and used the data to build a species-

independent model of mismatched sgRNA activity for bacterial CRISPRi. We used this model to explore the expression-fitness landscapes of all essential genes in *E. coli* and *B. subtilis* by comparing the fitness effects of ~90 mismatched sgRNAs to the predicted levels of CRISPRi activity for each essential gene in each species. Our analysis of per gene expression-fitness relationships revealed that CRISPRi targeting of different essential genes has different effects on cellular fitness, but that these effects are largely conserved within pathways, and between *E. coli* and *B. subtilis* homologs. The conservation of expression-fitness relationships for most genes over >2 billion years suggests that conserved homeostatic constraints underlie the optimization of essential gene expression and highlights processes with divergent evolutionary pressures. Our findings thus provide both new insights into bacterial physiology and an important new tool for exploring reduced-expression phenotypes in many bacterial species.

RESULTS AND DISCUSSION

CRISPRi efficacy is similarly titrated by sgRNA mismatches in *E. coli* and *B. subtilis*

Mismatched sgRNA efficacy has been sparsely tested in *E. coli* (Qi et al., 2013) using *rfp* targeted by variants of a single sgRNA. However, no comprehensive, multi-species measurements of repression by mismatched sgRNAs have been reported for bacterial CRISPRi. To directly quantify the impact of mismatches on the repression of mismatched sgRNAs in bacterial systems, we generated a comprehensive library of sgRNA spacers targeting *gfp* (3201 total), consisting of all spacers fully complementary to the non-template strand (33), a majority of their possible single mismatch variants (47/60), and a subset of their possible double mismatch variants (49/1710) (Figure 3.8A). Using FACS-seq (Figure 3.1A, Methods), we quantified the ability of these sgRNAs to repress transcription of a highly expressed chromosomal copy of *gfp* both in *E. coli* and *B. subtilis* (Figure 3.9A-C, Table S1). We found that sgRNAs with either single (Figure 3.1B) or double (Figure 3.10A) mismatches in

their base-pairing regions generated the full range of repression (no efficacy to full efficacy) in both species. Importantly, sgRNA activity was unimodal (Figure 3.9E-H) and highly correlated between *E. coli* and *B. subtilis* (R^2: singly mismatched sgRNAs = 0.65, doubly mismatched sgRNAs = 0.61, all sgRNAs = 0.71; Figure 3.1B, Figure 3.10A, and Table S1), despite an evolutionary distance of several billion years and differences in experimental setup (*E. coli:* plasmid-encoded sgRNAs, *B. subtilis:* chromosomally integrated sgRNAs).

Mismatched sgRNA efficacy has also been explored in mammalian CRISPRi systems (Gilbert et al., 2014; Jost et al., 2020), however, substantial differences exist between CRISPRi modalities in bacteria (blocking RNA-polymerase elongation) and mammalian systems (recruiting chromatin modifiers to promoters). To compare mismatched sgRNA efficacy between bacteria and mammalian systems, we calculated the mean relative activity of bacterial *gfp*-targeting and mammalian essential gene-targeting singly mismatched sgRNAs (Jost et al., 2020) for all combinations of mismatch position and base substitution. We found that although sgRNA activity is correlated between the two systems (R^2 = 0.61, Figure 3.1C and Figure 3.11), the activity of the mammalian system is more strongly impacted by mismatches in general, particularly in the PAM-proximal seed region (Figure 3.1C and Figure 3.11. Whereas almost all mismatches in the seed region completely abolish sgRNA activity in the mammalian system, sgRNAs with equivalent mismatches still retain measurable activity in the bacterial system, consistent with previous reports (Qi et al., 2013). These differences may be due to differences in how CRISPRi functions. In bacteria, dCas9 efficiently blocks transcriptional elongation when targeted within the ORF, while in mammalian systems, efficient repression requires targeting a dCas9-KRAB fusion to occlude the promoter region and recruit chromatic modifying proteins (Gilbert et al., 2013). Previous work compared the activity of mismatched sgRNAs in mammalian cells using either dCas9 or dCas9-KRAB (Gilbert et al., 2014) and found that mismatches in the seed region were better tolerated by a dCas9 repression system than by a dCas9-KRAB system. This indicates that KRAB function may be responsible for the sensitivity

of the mammalian system to mismatches. Taken together, these data suggest that although the primary determinant of mismatched sgRNA efficacy in both bacteria and mammalian systems is shared, differences in how CRISPRi functions in these two systems manifest as quantitative differences in mismatched sgRNA efficacy. Remarkably, mismatched sgRNAs function similarly in *E. coli* and *B. subtilis*, potentially allowing facile design of sgRNAs with specific degrees of knockdown across diverse bacterial systems.

A species-independent linear model robustly predicts mismatched-sgRNA activity

Given the species-independent performance of *gfp*-targeting mismatched sgRNAs, we next asked whether we could accurately predict the effects of single mismatches on sgRNA activity. Previous work on CRISPRi off-target effects (Gilbert et al., 2014; Qi et al., 2013) and concurrent work on mismatched sgRNAs in a mammalian context (Jost et al., 2020), identified mismatch position, base substitution, and the GC% of the fully complementary spacer as the strongest determinants of mismatched sgRNA efficacy. We therefore constructed a simple linear model that used one-hot encoded mismatch position (20 parameters), one-hot encoded base substitution (12 parameters), and spacer GC% (1 parameter) to predict the relative efficacy of mismatched *gfp*-targeting sgRNAs (Figure 3.2A).

We separately trained this linear 33-parameter model on the *E. coli, B. subtilis*, or species-averaged relative efficacy of our 1,551 *gfp*-targeting singly mismatched sgRNAs. Regardless of which data was used for training, model weights for mismatch location, base substitution, and spacer GC% were similar (Figure 3.12, Table S2), could be used for cross-prediction (Figure 3.13) and reflected known dCas9 behavior. Model weights for mismatch location indicated decreasing sgRNA efficacy for mismatches closer to the PAM (Figure 3.2B), as expected based on the mechanism of dCas9 binding and R-loop formation (Gong et al., 2018). Model weights for base substitution were correlated ($R^2 = 0.60$, $p < 0.005$) to the

changes in the free energy of sgRNA-DNA pairing (ΔΔG) caused by the base substitution (Figure 3.2C). Finally, the negative coefficient assigned to GC% suggests that sgRNAs with high GC% are more tolerant of mismatches, consistent with what has been found in a mammalian system (Jost et al., 2020). Despite the simplicity of this model, the effects of single mismatches were robustly predicted (Figure 3.2D, Figure 3.13, species-averaged R^2 = 0.56, 11-fold CV-MSE = 0.10 $^{+}/_{-}$ 0.08). Additionally, when applied to doubly mismatched sgRNAs by assuming that mismatches independently affect sgRNA efficacy (as suggested in Qi et al., 2013), our model accurately predicted the relative efficacy of these sgRNAs (species-averaged R^2 = 0.53, Figure 3.10B).

To validate a biophysical interpretation of our model, we took advantage of a previously published data set containing measured association rates (k_{on}) of a dCas9-sgRNA complex to 60 singly mismatched and 1130 doubly mismatched DNA sequences (Boyle et al., 2017). Re-interpreting this data as sgRNA mismatches, we compared the measured association rates to the relative sgRNA activity predicted by our model for these orthogonal sgRNAs (Figure 3.2E, Figure 3.10C). Our predicted sgRNA activity was highly correlated (R^2: single mismatches = 0.71, double mismatches = 0.45) to the k_{on} measured in this *in vitro* system, supporting the hypothesis that mismatched-CRISPRi functions by reducing the association rate of the dCas9-sgRNA complex for the target DNA, likely by slowing R-loop formation (Gong et al., 2018). Taken together, these data strongly suggest that a simple linear model trained on the relative efficacy of our *gfp*-targeting singly mismatched sgRNA library can be used to design mismatched sgRNAs with a specific activity level targeting any gene.

Measuring the fitness of libraries of mismatched sgRNAs in *E. coli* and *B. subtilis*

Using our model of mismatched sgRNA activity, we designed a set of sgRNAs targeted to the essential gene complement of *E. coli* and *B. subtilis* (~300 genes in each species, Table S3) and predicted to have a range of activities. We generated large pooled libraries of strains in

which each essential gene is targeted by 100 sgRNAs (10 fully matched guides each with 9 singly mismatched variants, Methods, Figure 3.8C) and compact libraries in which each essential gene is targeted by 11 sgRNAs. Additionally, for two well characterized essential genes encoding UDP-GlcNAc-1 carboxyvinyltransferase (*E. coli: murA*, *B. subtilis: murAA*), and dihydrofolate reductase (*E. coli: folA*, *B. subtilis: dfrA*), we generated comprehensive libraries (at least 47/60 single mismatch variants for each sgRNA within the gene, Methods, Figure 3.8B). The libraries were grown for 10 doublings, maintaining exponential phase through back-dilution (Figure 3.3A). We calculated the relative fitness (Kampmann et al., 2013; Rest et al., 2013) of each strain by comparing its relative abundance (quantified by next-generation sequencing of the sgRNA spacers) to the relative abundance of 1000 non-targeting sgRNAs at the start and end of each experiment (Methods, Table S3). Relative fitness is defined as the number of doublings of any strain relative to the number of wildtype doublings over the time course of the experiment. Strains with a relative fitness of 1 grow as well as wild-type; lower values imply slower growth. Relative fitness was highly reproducible in both species ($R^2 > 0.9$, Figure 3.14A-B) and was validated by orthogonal measurements of individual strain fitness (Figure 3.14C). Our relative fitness values for fully complementary guides were correlated with previously reported measurements (Rousset et al., 2018; Wang et al., 2018) but had greatly expanded dynamic range (Figure 3.14D-E) due to differences in experimental design. Whereas previous studies were optimized for determining essentiality by quantifying fitness over >15 generations, we optimized our experiments for an expanded dynamic range by quantifying fitness over 10 generations and sequencing to greater depth. This expanded dynamic range enabled the quantification of strong fitness defects, including measurement of negative relative fitness, which indicates active depletion from the pool. CRISPRi targeting of 23 *E. coli* genes and 24 *B. subtilis* genes reproducibly (>5 sgRNAs/gene) caused negative relative fitness (Table S4). Consistent with an interpretation of negative relative fitness as lysis, a majority (15/24) of these

B. subtilis genes caused lysis (as assayed by microscopy) when targeted with a fully complementary sgRNA (Peters et al., 2016) (Table S4, Methods).

Per gene expression-fitness relationships are robustly quantified using mismatched-sgRNAs

We next assessed whether comparing the activity of sgRNAs predicted from our model to their measured relative fitness would allow us to infer the expression-fitness relationships of essential genes. Inferring per gene expression-fitness relationships requires both that our *gfp*-trained model accurately predicts relative sgRNA efficacy and that fully complementary sgRNAs have similar efficacy at all loci within a gene.

We first tested the applicability of our *gfp*-trained model to sgRNAs targeting endogenous genes. Since repression of essential gene expression monotonically decreases cellular fitness, we reasoned that if our model is accurate, predicted sgRNA efficacy should be negatively correlated to relative fitness within a series of sgRNAs targeted to a specific locus. Consistent with this hypothesis, we found that predicted sgRNA activity within series was negatively correlated to the relative fitness of those strains in both *E. coli* (median r = -0.74, Figure 3.3B) and *B. subtilis* (median r = -0.86, Figure 3.3B), suggesting that relative sgRNA activity was correctly predicted. Weaker correlations in *E. coli* likely reflect variation in sgRNA plasmid copy number and/or *E. coli* specific effects (Cui et al., 2018). To further probe the generality of our model, we trained it on the relative fitness effects of our comprehensive mismatched sgRNA libraries (Figure 3.8B) targeting the essential endogenous dihydrofolate reductase genes: *E. coli folA* (1,525 sgRNAs) and *B. subtilis dfrA* (1,281 sgRNAs). Because dihydrofolate reductase abundance is linearly related to fitness above an initial threshold of activity (Bhattacharyya et al., 2016), we interpreted the fitness defects of strains containing mismatched sgRNAs targeting these genes as readouts of knockdown efficacy. We found that when trained on the *folA* or *dfrA* data, model weights (Figure 3.12) and performance (Figure

3.13) were similar to the *gfp*-trained model, suggesting that both our model of mismatched sgRNA efficacy and the parameters fit from the *gfp* data are broadly applicable.

We next tested whether knockdown efficacy was consistent across targeted loci. Because our model predicts knockdown efficacy with respect to the fully complementary sgRNA ("relative knockdown efficacy"), it can be applied across sgRNA families to determine expression-fitness curves only if fully complementary sgRNAs achieve similar knockdown efficacy at all loci within a gene. Fully complementary sgRNAs targeting *gfp* and *folA/dfrA* showed limited variability that did not correlate with the location of the sgRNA (Figure 3.15). To determine if this pattern held true for other endogenous essential genes, we reasoned that differences in knockdown at different loci within a gene would manifest as differences in the fitness effect of fully complementary sgRNAs targeting the same gene (Figure 3.3C-D). Comparing the variability of the fitness effect of fully complementary sgRNAs targeting the same gene to the overall variability in the fitness effect of fully complementary sgRNAs using sum of squares, we found that within gene variability accounted for only 26.7% of total variability in *E. coli* and 18.6% of total variability in *B. subtilis*. This suggests that fully complementary sgRNAs targeting the same gene are substantially more similar with regards to their fitness outcomes than fully complementary sgRNAs as a whole, and supports the assumption that fully complementary sgRNAs targeting the same gene have similar levels of activity.

Consistent with these outcomes, predicted sgRNA activity was negatively correlated with cellular fitness for all sgRNAs targeting the same gene in both *E. coli* (median r = -0.65, Figure 3.3B) and *B. subtilis* (median r = -0.75, Figure 3.3B). Taken together, these analyses strongly suggest that we can accurately and sensitively probe the expression-fitness relationships of essential genes in *E. coli* and *B. subtilis* by comparing the predicted activity of mismatched sgRNAs to their measured fitness using a pooled screening approach.

Expression-fitness relationships are conserved within biological processes and between essential homologs

Examining the essential gene expression-fitness relationships, we were struck by their diverse and gene-specific nature (Table S3). To quantitatively characterize these differences, we first binned the sgRNAs targeting each gene according to predicted sgRNA activity and calculated the median fitness within each bin (Methods, Figure 3.16A-B, and Table S5). Next, we used these simplified representations of per gene expression-fitness relationships to calculate pairwise distances between *E. coli* and *B. subtilis* essential genes. Within each organism, we found that the expression-fitness relationships of genes involved in the same biological process (whether defined by KEGG, GO biological process, or COG) were significantly more similar to each other than to those of genes involved in different biological processes, even when excluding gene pairs in the same operon to account for CRISPRi polarity (all $p < 10^{-16}$, Methods). Inversely, clustering genes by the shape of their expression-fitness curves produced functional enrichments (Table S6) in both *E. coli* and *B. subtilis*. Finally, in a cross-species comparison, the expression-fitness curves of essential genes were, as a group, more similar ($p < 10^{-10}$) to that of their homologs than to other genes in the opposing species.

To eliminate the possibility that these similarities were the result of systemic biases in the prediction of mismatched sgRNA activity, we also performed these analyses using the average per gene fitness effect of fully complementary sgRNAs, which do not depend on our model of mismatched sgRNA activity. We found that the per gene fitness effects of fully complementary sgRNAs were also significantly more similar within biological processes (all $p < 10^{-16}$, Methods) than between biological processes. However, clustering solely by the per gene fitness effects of fully complementary sgRNAs produced fewer functional enrichments, and a less statistically significant similarity between *E. coli* and *B. subtilis* homologs ($p < 10^{-6}$). Taken together, these data suggest that these expression-fitness curves are both biologically

meaningful and representative of deeply conserved homeostatic constraints on bacterial physiology.

Expression-fitness relationships of biological processes

To explore the conserved optimizations of bacterial essential gene expression, we examined three functional categories having similar expression-fitness relationships in *E. coli* and *B. subtilis*. CRISPRi targeting of essential cofactor biosynthesis genes (KEGG pathways under “Metabolism of cofactors and vitamins”) did not strongly affect fitness in either species after 10 generations (Figure 3.4A-B). This observation is consistent with the small-colony but non-culturable phenotype of essential cofactor biosynthesis gene deletions (Koo et al., 2017) and suggests that these cofactors and/or the enzymes producing them are present in excess of what is required for exponential growth. This buffer may be required to enable rapid shifts in metabolism in response to changing environmental conditions, similar to what has been proposed for the pentose-phosphate pathway (Christodoulou et al., 2018).

The robustness of both bacteria to CRISPRi targeting of essential cofactor synthesis genes contrasts with the strong, approximately linear effect of targeting genes involved in translation (KEGG pathways under “Translation”, Figure 3.4C-D). Previous work has established a linear relationship between growth rate and the number of ribosomes per cell during exponential growth in *E. coli, B. subtilis,* and other bacteria (Borkowski et al., 2016; Schaechter et al., 1958; Scott et al., 2010). By linearly inhibiting ribosomal protein expression, we likely decrease the number of functional ribosomes, leading to a corresponding linear decrease in growth rate. Moreover, feedback to restore ribosomal protein expression is unlikely because most ribosomal proteins are negatively regulated by their excess relative to rRNA (Nomura et al., 1980; Scott et al., 2014). Depletion of translation factors has a similarly linear effect on growth rate, likely due to slowed elongation rate (Dai et al., 2016) as has been shown for some antibiotics that inhibit translation elongation (Scott et al., 2010). The conserved linear

relationship between the expression of proteins involved in translation and growth rate reinforces the universal importance of translational capacity for determining growth rate.

CRISPRi targeting of genes involved in cytoplasmic peptidoglycan (PG) precursor synthesis (KEGG ko00550) also generated strong phenotypes in both species. However, in contrast to the linear expression-fitness relationship of genes involved in translation, PG synthesis genes exhibited bimodal fitness outcomes that depended on predicted sgRNA activity (Figure 3.4E-F). This bimodality is highlighted by the fitness outcomes of the comprehensive *murA* and *murAA*-targeting libraries, even when considered independently of predicted knockdown (Figure 3.16C-D). Cells tolerated partial repression of these genes without exhibiting a fitness defect. If expression was sufficiently repressed, these strains lysed (Table S4) as has been described for *murA*, *murG*, and *mraY* inhibition in *E. coli* (Fransen et al., 2017; Mengin-Lecreulx et al., 1991; Zheng et al., 2008) and for *murC*, *murD*, and *murG* depletion in *B. subtilis* (Peters et al., 2016). To determine whether the bimodality observed for these genes was due to bimodal CRISPRi activity, we measured the ability of 18 mismatched sgRNAs to repress a *murAA-gfp* transcriptional fusion in *B. subtilis*. These measurements were conducted in a *B. subtilis* strain complemented with non-targeted *murAA* (Methods) to enable quantification of lethal levels of knockdown and to avoid the potential for transcriptional feedback. Measured knockdown closely tracked the predicted activity of sgRNAs targeting *murAA* in this experiment (Figure 3.16E, Table S10), suggesting that the non-linear expression-fitness relationship of uncomplemented *murAA* reflects non-linearly decreasing growth due to MurAA depletion, transcriptional feedback, cell lysis, or other host specific effects. Previous work in *E. coli* (Mengin-Lecreulx and van Heijenoort, 1985) and *Salmonella typhimurium* (Kahan et al., 1974) demonstrated that the levels of enzymes involved in cytoplasmic peptidoglycan precursor synthesis are not affected by growth rate or fosfomycin inhibition. Additionally, recent work in *E. coli, Caulobacter crecentus*, and *Listeria monocytogenes* (Harris and Theriot, 2016) found that low doses of fosfomycin impact cell morphology, as expected if peptidoglycan synthesis is

inhibited, but did not impact growth rate. This suggests that the absence of homeostatic regulation in peptidoglycan precursor synthesis is conserved. Given this lack of regulation, the dearth of intermediate fitness outcomes in either species upon repression of PG precursor synthesis is surprising. It suggests that neither species is able to slow growth rate or up-regulate cytoplasmic PG precursor synthesis in response to reduced flux through this pathway to prevent lysis. It has been proposed that bacteria use peptidoglycan precursor concentration to sense and balance cellular metabolism and growth (Harris and Theriot, 2016). This would be incompatible with direct feedback regulation of cytoplasmic PG precursor synthesis and may explain the sharp transition between growth and lysis.

The expression-fitness relationships of many cell wall synthesis genes differ between *E. coli* and *B. subtilis*

Given the similarity between the expression-fitness curves of most essential genes in *E. coli* and *B. subtilis,* we reasoned that homologs with substantially different expression-fitness curves may reflect biologically meaningful differences between the two organisms. We identified 9 homologs as significantly different between the two organisms (Table S7, FDR < 0.2).

Remarkably, 7/9 of these genes encoded enzymes involved in peptidoglycan (PG) synthesis and maturation, highlighting that although these pathways are conserved between Gram-positive and -negative species, major distinctions have evolved in how they contribute to construction of viable cells. Whereas CRISPRi targeting of genes involved in cytoplasmic PG precursor synthesis (Figure 3.5, group 3) generated bimodal expression-fitness relationships in both *E. coli* and *B, subtilis*, CRISPRi targeting of genes encoding enzymes involved in UDP-GlcNAc synthesis, *meso*-DAP synthesis, and longitudinal cell wall synthesis differentially affected the two species (Figure 3.5).

E. coli was significantly more tolerant of *mreBCD* perturbation than *B. subtilis* (Figure 3.5, group 4). In contrast to *B. subtilis*, which lysed at intermediate levels of *mreBCD*

knockdown, *E. coli* exhibited a minimal fitness defect after 10 generations and only lysed after 15 generations (Table S3). This observation is consistent with the small effect of CRISPRi targeting of *mrdA* (the PBP2 associated with MreBCD) on fitness in *E. coli*, the lack of transcriptional knockdown when targeting *mreC* (Reis et al., 2019), and with previous work which found that the fitness of *Enterobacter cloacae* is also relatively unaffected by *mreBCD* CRISPRi targeting (Peters et al., 2019). It is unclear why *E. coli* and other Gram-negative bacteria are less affected by *mreBCD* CRISPRi targeting than *B. subtilis*, however transcriptional buffering through feedback appears to play a role.

E. coli was also more robust than *B. subtilis* to CRISPRi targeting of genes required for producing either UDP-GlcNAc (Figure 3.5, group 1) or meso-DAP (Figure 3.5, group 2). Whereas *B. subtilis* lysed when these genes were targeted with high activity sgRNAs, *E. coli* exhibited a minimal fitness effect after 10 generations (Figure 3.5, Table S4, and Table S7). To determine whether transcriptional feedback, mediated by divergent regulatory mechanisms (Barreteau et al., 2008; Rodionov et al., 2003) is responsible for the lack of observed phenotype in *E. coli,* we measured the ability of 2 fully complementary and 2 mismatched sgRNAs to reduce the expression of *E. coli* genes encoding enzymes involved in *meso*-DAP (*asd, dapD, dapE*), and UDP-GlcNAc (*glmS*) synthesis using RT-qPCR. We found that both fully complementary and mismatched sgRNAs were effective in significantly reducing the expression of these genes, with mismatched sgRNAs affecting gene expression less than fully matched sgRNAs (Figure 3.6A). All fully complementary sgRNAs targeting *dapD* and *dapE* generated similar knockdown (~20 fold) to a sgRNA targeting *rfp*, suggesting that if compensatory mechanisms exist, they must be post-transcriptional. In contrast, fully complementary sgRNAs targeting *asd* and *glmS* were less efficacious, generating ~5 fold knockdown, suggesting that the robustness of *E. coli* to CRISPRi targeting of *asd* and *glmS* may be due to transcriptional feedback. Consistent with this hypothesis, *asd,* which encodes an important branchpoint enzyme in amino acid synthesis is multivalently repressed by lysine, threonine, and methionine

(Boy and Patte, 1972) and negatively regulated by the small RNA SgrS (Bobrovskyy and Vanderpool, 2016). By de-repressing *asd* in response to low amino acid levels, these mechanisms may be responsible for the increased robustness of *E. coli* to *asd* knockdown. Similarly, the robustness of *E. coli* to CRISPRi targeting of *glmS* may also be due to feedback regulation by a positively acting sRNA, *glmZ*, which stabilizes *glmS* mRNA in response to low intracellular GlcN-6-P levels (Urban and Vogel, 2008), such as those caused by CRISPRi targeting of *glmS.* Taken together these results suggest that feedback, mediated by divergent regulatory mechanisms (Barreteau et al., 2008; Rodionov et al., 2003) may contribute to the reduced sensitivity of *E. coli* to CRISPRi targeting of these genes.

An additional difference between cell wall synthesis in *E. coli* and *B. subtilis* is PG recycling (Figure 3.6B). Whereas *E. coli* recycles cleaved PG in both exponential and stationary phase, *B. subtilis* does so only in stationary phase (Johnson et al., 2013). We reasoned that increased robustness of *E. coli* to knockdown of genes in *meso*-DAP and UDP-GlcNAc synthesis might be a consequence of the ability of *E. coli* to supplement *de novo* synthesis of these compounds with recycled PG. We therefore repeated our small library fitness experiments in *E. coli* strains harboring deletions of either *ampG* or *mpl*, two key enzymes involved in recycling, and compared the relative fitness of essential gene knockdowns in the different backgrounds (Methods, Figure 3.8C). AmpG is the sole permease involved in PG recycling, and its deletion abolishes PG recycling (Johnson et al., 2013). Deletion of *ampG* did not sensitize *E. coli* to knockdown of PG synthesis genes (Figure 3.6C), suggesting that recycled PG does not contribute robustness to knockdown. Surprisingly deletion of *mpl*, which acts downstream of AmpG to ligate salvaged tripeptides to UDP-GlcNAC, sensitized *E. coli* to knockdown of *murI* and *dapA* (Figure 3.6D). These two genes encode enzymes responsible for isomerizing L-Glu to D-Glu (*murI*) and catalyzing the rate limiting step in *meso*-DAP synthesis (*dapA,* Reverend et al., 1982). D-Glu and *meso*-DAP are two of the three amino acids in the tripeptide ligated by *mpl.* The sensitizing and specific effect of an *mpl* deletion suggests that the flux through PG

recycling is important for fitness, and the lack of sensitizing effects in the *ampG* deletion suggests that the absence of this flux can be overcome. Compensatory regulation in response to a cytoplasmic PG intermediate upstream of *mpl* ligation (e.g. UDP-GlcNAC) would be activated in response to *ampG* deletion, but not in response to *mpl* deletion, potentially explaining our observations. However, the mechanism of such regulation remains to be elucidated.

These experiments also identified a novel phenotype, highlighting the ability of these approaches to generate new biology. Deleting either *ampG* or *mpl* de-sensitized *E. coli* to the depletion of FtsH (Figure 3.6C-D), an essential protease responsible for balancing flux through lipopolysaccharide biosynthesis and phospholipid synthesis by regulating the stability of LpxC, an enzyme catalyzing the first committed step in lipopolysaccharide synthesis (Ogura et al., 1999). It remains to be determined whether PG recycling affects the balance between lipopolysaccharide biosynthesis and phospholipid synthesis (perhaps by depleting the pool of UDP-GlcNAc), or an ancillary FtsH function. Underscoring the importance of screening multiple levels of essential gene knockdown, no significant epistatic interactions were identified in either strain background when only the two fully complementary sgRNAs targeting each gene were considered. These data highlight the utility of combining mismatched CRISPRi libraries with other genetic backgrounds to identify novel modulators of essential gene requirements.

Expression-fitness curves are modulated by external perturbations

Bacteria have been previously shown to produce some enzymes at higher levels than needed for immediate survival in order to buffer against future environmental perturbations (Figure 3.4A-B, Christodoulou et al., 2018). To explore the ability of mismatched sgRNAs to capture shifts in the expression-fitness relationship of individual genes driven by external perturbations, we measured the relative fitness of the comprehensive sgRNA library targeting *drfA* in *B. subtilis* treated with sub-MIC doses of the antibiotic trimethoprim. Trimethoprim

directly inhibits DfrA (Figure 3.7A) and has been shown to act synergistically with partial knockdown of *dfrA* (Peters et al., 2016). However, the degree of synergy as a function of *dfrA* knockdown has not been investigated. We found that a low dose of trimethoprim abolished the initial buffering observed in the untreated strain (Figure 3.7B). A higher (but still sub-MIC) dose of trimethoprim further depressed the expression-fitness relationship, and caused a phenotype even at the lowest levels of knockdown. These data suggest that the DfrA concentration is buffered against external perturbations and highlights the ability of mismatched sgRNAs to enable exploration of these subtle shifts.

PERSPECTIVE

Bacterial essentialomes typically consist of several hundred genes encoding the core reactions central to viability. Apart from studies of the *lac* operon (Dekel and Alon, 2005; Eames and Kortemme, 2012), the lack of precise, high-throughput methods for titrating gene expression has precluded an understanding of how bacteria respond to a continuum of essential gene repression. Here we report a modified CRISPRi system that generates graduated, species-independent levels of repression by programming dCas9 with singly mismatched sgRNAs. Leveraging this system, we assessed the fitness effects of titrating essential gene expression for almost all essential genes in *E. coli* and *B. subtilis.* These data revealed striking differences between the expression-fitness landscapes of genes and pathways that highlighted conserved and divergent biological constraints driving the optimization of essential gene expression. Because CRISPRi tools have now been established for many bacteria (e.g. (Peters et al., 2019), our approach can be applied to pathogenic and microbiome strains to inform target selection for drug design and illuminate unique constraints for bacterial growth.

Our comprehensive characterization of mismatched sgRNAs targeting *gfp* identifies the organism-independent rules that determine mismatched sgRNA activity in bacteria (Figure 3.1 and Figure 3.2). By applying these rules, either independently or with the software developed

here (Methods), sgRNAs can be readily designed that generate a specific level of knockdown targeting any bacterial gene. These mismatched sgRNAs enable a range of high-throughput techniques to accelerate biological discovery. First, mismatched sgRNAs allow the rapid generation of reduced expression mutants. Reduced expression mutants of essential genes have been used for drug target discovery and lead optimization in *Mycobacterium tuberculosis*; however, extensive up-front strain optimization was required to identify gene-specific expression levels that facilitate identification of chemical-genetic interactions (Johnson et al., 2019). Our system enables facile generation and testing of a broad range of repression levels, potentially accelerating drug discovery and target identification. Second, our high-throughput pooled screening methodology to determine relative fitness is amenable to testing the effects of essential gene titration in varying environmental and genetic backgrounds. To simplify the exploration of essential gene requirements in diverse conditions, we constructed smaller (11 sgRNA/gene) libraries for *E. coli* and *B. subtilis* essential genes that can be easily screened in varying conditions or transferred into different genetic backgrounds. The reduced complexity of these libraries aids multiplexing while retaining a broad range of phenotypes for most genes in both species (Figure 3.17). Finally, our approach also allows the use of CRISPRi to measure epistatic interactions between essential and non-essential genes. This approach had previously been hampered by the need to fully repress the non-essential gene (so as to maximize the chance of a phenotype) while only partially repressing the essential gene (so as to enable cell survival). This hurdle can be overcome by targeting the essential gene with a mismatched sgRNA.

Although a mounting body of evidence supports the idea that gene essentiality is a quantitative trait (Rancati et al., 2018), systematically exploring this hypothesis has been challenging due to the lack of universally applicable methods of essential gene titration. Here we firmly establish quantitatively different fitness effects of essential gene depletion by targeting each gene at 10 separate loci using directly comparable CRISPRi methodologies and fitness

measurements in two species. This is the first dataset that allows meaningful comparisons of expression-fitness relationships across species, and we use it to compare homologous essential genes in *E. coli* and *B. subtilis*. The similarity of most expression-fitness relationships between these diverged species underscores conserved evolutionary constrains and also highlights the significant differences in DAP synthesis and in the Rod PG synthetic apparatus as new targets for study. In contrast to synthetic promoter-based methods of studying expression-fitness relationships (Keren et al., 2016) which inherently eliminate native transcriptional feedback loops, our method is sensitive to the effects of essential gene regulation. Whether a cell buffers the effects of gene repression by regulatory feedback or by producing an excess of gene product is secondary to the importance of having buffering capacity, which is readily observed for some genes and not others. These studies inform target selection for drug design, illuminate aspects of bacterial growth, and provide a starting point for investigating how bacteria program robustness into their essential gene network.

MATERIALS AND METHODS

Experimental model and subject details:

Microbes

Escherichia coli strains were cultured in LB medium at 37C. *Bacillus subtilis* strains were cultured in LB medium at 37C.

Methods details:

General strain manipulations and procedures

Bacillus subtilis strain construction and growth conditions

All *B. subtilis* strains were constructed in the wildtype 168 background using natural competence as previously described (Koo et al., 2017). For all individual CRISPRi strains and

libraries, a recipient strain encoding *dcas9* under control of the P*xyl* promoter at the *lacA* locus (strain CAG74209) (Peters et al., 2016), was transformed with an sgRNA plasmid (see "sgRNA plasmid construction") which recombines in single copy at the *amyE* locus, selecting for chloramphenicol resistance. In select cases, single- vs. double-crossover events from plasmid integration were distinguished by streaking on starch plates to assay disruption of *amyE*.

For the GFP knockdown FACS-seq experiments, two modified recipient strains expressing *dcas9* were constructed: one encoding *gfp* (strain CAG78920) and the other encoding *rfp* (strain CAG78921). To construct these, the *dcas9* strain (strain CAG74209) was transformed with pDG1731-gfp or pDG1731-rfp to integrate P*veg-gfp-spc* or P*veg-rfp-spc*, respectively, at the *thrC* locus, selecting for spectinomycin resistance. All subsequent transformations of the *gfp* and *rfp*-marked strains required threonine supplementation in the competence media (40µg/ml), as *thrC* is disrupted.

For flow cytometry-based competition experiments (see "Relative fitness validation"), the *dcas9* recipient strain was transformed with a modified sgRNA plasmid that also encodes either P*veg-gfp* or P*veg-rfp* (see "sgRNA plasmid construction").

A *murAA-gfp* transcriptional fusion knock-down reporter strain was constructed by transformation of the *dcas9* strain (above) with the DNA fragments containing constitutively expressed *rfp* with removable kan^R cassette, *murAA-gfp* transcriptional fusion with removable kan^R cassette, and constitutively expressed non-targeted *murAA* with spectinomycin resistant gene, of which fragments were integrated into *sacA*, *murAA* and *thrC* locus respectively, sequentially in that order. The *B. subtilis* P*veg* promoter was used for constitutive expression of *rfp* and non-targeted *murAA*. The DNA fragment containing constitutively expressed *rfp* with removable kan^R cassette was constructed by joining PCR of three fragments: P*veg-rfp-*kan^R fragment amplified from pACYC-*rfp-kanR*, and 1kb each 5′ and 3′ flanking sequences of *sacA*. The DNA fragment containing *murAA-gfp* transcriptional fusion with removable kan^R cassette was constructed by the joining of *gfp-kanR* amplified from pACYC-gfp-kanR, 1kb of the

3’ end of the *murAA* open reading frame, and 1kb downstream of the *murAA* open reading frame. Before the *gfp-kanR* fragment was integrated downstream of *murAA* to generate strain CAG78923, the *kanR* cassette was removed from *rfp* strain as described previously to generate strain CAG78922 (Koo et al., 2017). Non-targeted *murAA* was designed to remove PAM sequence or alter the sgRNA targeting sequence without substituting amino acid sequence of *murAA*. Non-targeted *murAA* DNA was generated by overlapping PCR with mutagenic primers (Table S8) and its *BsiWI/NruI* digested fragment were cloned into pJMP3 (Addgene #79875) digested with *BsrGI/PmeI*. The cloned plasmid was transformed into the *dcas9, rfp*, *murAA-gfp* strain, selecting for spectinomycin resistance, to generate strain CAG78924. Finally, this strain was transformed with sgRNA plasmids as described above.

Unless otherwise noted, all strain construction and growth assays for *B. subtilis* were done in LB medium and using antibiotics at the specified concentrations: erythromycin (1µg/ml), spectinomycin (100µg/ml), chloramphenicol (7.5µg/ml), kanamycin (7.5µg/ml).

<u>*Escherichia coli* strain construction and growth conditions</u>

All CRISPRi library strains were constructed in the wildtype BW25113 background by electroporating an sgRNA plasmid or plasmid pool (see “sgRNA plasmid construction”) into a recipient strain encoding *dcas9* (for essential gene knockdown libraries), or *dcas9* and *gfp* or *rfp* (for GFP knockdown libraries), selecting for ampicillin resistance.

For the essential gene knockdown library recipient strain (strain CAG78830), Tn7 transposition was used to integrate a *dcas9* expression cassette into the Tn7att site using triparental mating of DAP(diaminopimelic acid)-dependent donors and selecting for gentamicin resistance in the absence of DAP, as previously described (Peters et al., 2019). The *dcas9* expression cassette is modified from previously described versions (Peters et al., 2019), contains *dcas9* from *S. pyogenes* (Qi et al., 2013) with a 3X Myc C-terminal tag, and is

expressed from the IPTG-inducible promoter P*lLac-O1* (Lutz and Bujard, 1997) and regulated by *lacIq*.

For the GFP knockdown FACS-seq experiments, two recipient strains expressing *dcas9* were constructed: one encoding *gfp* (strain CAG78108) and the other encoding *rfp* (strain CAG78107). Each was generated by first cloning the constitutive *gfp* and *rfp* expression cassettes from pDG1731-gfp and pDG1731-rfp upstream of *frt-cat-frt* from pKD3 (Datsenko and Wanner, 2000), integrating them into the chromosome between *yjaA* and *yjaB* using recombineering (Thomason et al., 2014), and selecting for chloramphenicol resistance. P1 phage transduction (Thomason et al., 2007) was then used to move the *gfp-frt-cat-frt* or *rfp-frt-cat-frt* cassettes into BW25113, selecting for chloramphenicol resistance. Chromosomal *dcas9* was then introduced to these strains by conjugation using a pseudo-Hfr *dcas9* donor, as described previously (Rauch et al., 2017), where *dcas9* is expressed by the minimal synthetic promoter P*BBa_J23105* {https://parts.igem.org}, and transconjugates were selected using gentamicin and chloramphenicol.

For the RT-qPCR experiments, CRISPRi strains expressing sgRNAs targeting peptidoglycan biosynthesis genes were individually reconstructed by electroporating the *dcas9* strain (strain CAG78830) with sgRNA plasmids constructed as described below (see “sgRNA plasmid construction”).

For the experiments combining the compact sgRNA libraries with deletions of peptidoglycan recycling pathway genes, the desired deletion alleles (*ampG::kan* or *mpl::kan*) were isolated from the Keio collection. Briefly, the deletion mutants were isolated, confirmed by PCR of kanamycin cassette junctions, and P1 phage was made from verified strains. Transduction of the *dcas9* strain (strain CAG78830) was performed (Thomason et al., 2007) with each the phage, selecting for kanamycin resistance, and the resulting strains were transformed with sgRNA plasmid libraries as detailed below (see “*Escherichia coli* CRISPRi library construction)

Unless otherwise noted, all strain construction and growth assays for *E. coli* were done in LB medium and using antibiotic selection at the specified concentrations: ampicillin (100µg/ml), carbenicillin (50µg/ml), gentamicin (10µg/ml), chloramphenicol (25µg/ml), kanamycin (30ug/ml).

Bacillus subtilis CRISPRi library construction

As in the individual CRISPRi strain construction (above), CRISPRi libraries were constructed by transforming sgRNA plasmids into the *dcas9* strain. The protocol was modified in one of two ways in order to increase the scale; we found both methods were sufficient to maintain coverage of the pooled plasmids. In one method, cells were grown in *B. subtilis* competence medium to OD600=1.5, and then incubated with plasmid DNA (300µl cells + 300ng plasmid DNA) in 96-well deep-well plates. Incubations were performed for 2hr at 37C with shaking (900RPM), after which point plates were spun down at 5000g for 10 minutes and resuspended in 2mL LB medium before plating on plates (Falcon #351058) with chloramphenicol at a density ~0.4M CFU/plate and growth overnight at 37C. A second method incubated competent cells (grown in *B. subtilis* competence medium to OD600=1.5) with plasmid DNA in culture flasks, for 2hr at 37C with shaking (900RPM), after which point cells were spun down in 50ml tubes and resuspended in 2-6ml LB before plating on chloramphenicol plates as before.

To store the transformed CRISPRi library, plates were scraped, pelleted and resuspended in S7 salts (Koo et al., 2017) with 15% glycerol, and stored in 500uL aliquots at -80C.

Escherichia coli CRISPRi library construction

Strain library construction from plasmid libraries was achieved by electroporating plasmid DNA into the recipient strains, and plating on plates (Falcon #351058) with carbenicillin and 0.2% glucose (to repress uptake of residual lactose in LB that can induce the IPTG-controlled *dcas9* in the essential gene knockdown strains) at a density ~0.4M CFU/plate and growth overnight at

37C. To store the libraries, plates were scraped, pelleted, and resuspended in 15% glycerol to be stored at -80C.

sgRNA plasmid construction

The sgRNA plasmid pJSHA77 was modified from pDG1622 to increase transformation and double-crossover efficiency. 1.5kb of DNA upstream of *amyE* was PCR amplified from *B. subtilis* 168 genomic DNA and inserted into pDG1662 by HiFi Assembly (New England Biolabs #E2621L), replacing the shorter upstream fragment of *amyE* in pDG1662. Synthetic DNA containing a transcription terminator, an sgRNA driven by P*veg* with BsaI cut sites for spacer cloning, and downstream tandem transcription terminators was purchased from IDT and cloned into the previously described pDG1662 derivative by HiFi Assembly (New England Biolabs #E2621L), generating pJSHA77.

Oligonucleotide pools containing the desired elements with flanking restriction sites and library-specific PCR adapters were obtained from Agilent Technologies (Table S8). The oligonucleotide pools were amplified by 15 cycles of PCR using Q5 polymerase (New England Biolabs #M0493S) and custom primers (Table S8). The PCR product was digested with BsaI-HFv2 (New England Biolabs #R3733) and gel purified from 10% TBE gels (Invitrogen #EC6275BOX) to remove adapter ends. pJSHA77 vector was midi-prepped (Qiagen #12143), digested with BsaI-HFv2 for 1hr, and treated with Antarctic phosphatase (New England Biolabs # M0289S), and ligation was carried out at a 1:2 (vector:insert) molar ratio using T4 DNA Ligase (New England Biolabs #M0202L). Ligations were transformed into electrocompetent cells (New England Biolabs #C3020K), recovered for 1hr at 37C in LB, and then inoculated into 100ml with carbenicillin and grown overnight. Plasmid libraries were collected by midiprep (Qiagen #12143) and analyzed by deep sequencing (Illumina MiSeq #MS-103-1002) to assess cloning efficiency and library diversity.

For individual sgRNA strains, inserts were prepared by annealing two single-stranded DNA oligos together to create the 4-base overhangs, and then annealed inserts were ligated using T4 DNA Ligase (New England Biolabs #M0202L) individually into pJSHA77 digested with BsaI-HFv2 and treated with Antarctic phosphatase (New England Biolabs # M0289S).

For the single-strain competition validation strains, pJSHA77 was first modified to incorporate a constitutively expressed P*veg-gfp* or P*veg-rfp* using HiFi Assembly (New England Biolabs #E2621L). Strains were then constructed as described above, ligating annealed-pair inserts into the modified vector after digesting with BsaI-HFv2.

sgRNA plasmid library design

Code for designing (fully matched) sgRNA spacers targeting a list of genomic loci can be found at https://github.com/traeki/sgrna_design.

Non-targeting sgRNA controls were designed by creating random 20nt sequences with a distribution of GC content similar to *B. subtilis* (~45%), and then using bowtie (Langmead et al., 2009) to identify (and subsequently filter out) sgRNAs which aligned (allowing 3 or fewer mismatches) to other intragenic targets in the combined genomes of *E. coli* and *B. subtilis*, or any targets in *gfp* or *rfp*.

For the libraries targeting all essential genes in *B. subtilis*, multiple iterations of sgRNA library design (i.e. spacer design), construction, and analysis were used. For *B. subtilis* libraries, all presented data is from V2 library measurements, with the exception of the trimethoprim experiments which used measurements of the V1 libraries (all data in Table S3).

For the V1 libraries targeting *B. subtilis* genes we chose target genes to be all those previously identified as essential, putative essential, or low-fitness (Koo et al., 2017; Peters et al., 2016) (Table S3). For every gene in the V1 set, two non-overlapping fully complementary spacers were chosen, each targeting the non-template strand as close to the start of the ORF as possible. For each fully complementary spacer, a set of 25 spacer variants were designed

and ordered: 2x the fully complementary spacer, 5x randomly chosen single-mismatches within 7 bases of the PAM, 5x randomly chosen single-mismatches 8-12 bases from the PAM, 3x randomly chosen single-mismatches 13-19 bases from the PAM (to exclude the outermost base), 10x randomly chosen double mismatches 1-19 bases from the PAM. In addition, for every gene the first three non-overlapping template-strand spacers were included.

The V2 *B. subtilis* libraries included all essential *B. subtilis* genes as well as a subset of non-essential but fitness-impacting genes (Table S3) from V1 of the library. The V2 *E. coli* libraries included a majority of genes with evidence for essentiality (Table S3) (Koo et al., 2017). For every gene in this set, ten non-overlapping fully complementary spacers were chosen on the non-template strand, as close to the start of the ORF as possible. For each fully complementary spacer, a set of 10 spacer variants was designed and ordered (for a total of 100 sgRNAs per gene): 1x the original fully complementary spacer, 9x single-mismatches (Figure 3.8). Single-mismatches were chosen using the following criteria: all possible single-mismatch variants were evaluated by the trained linear model for a predicted sgRNA activity (*GitHub - traeki/sgrna_design*, no date) (train_linear_model.py and choose_guides.py). These predicted sgRNA activities were categorized into five bins: <10%, >90%, and three equally sized bins between 10% and 90% predicted sgRNA activity. Three sgRNAs were chosen from each of the middle three bins. For the design of all libraries using this strategy, a preliminary version of the linear model was used.

The compact libraries with 11 sgRNAs per gene were selected as above, with the following modifications for both species: for each gene 2x fully complementary sgRNAs were chosen, and 9x single-mismatch variants were selected from among all possible single-mismatch variants of each, using a binning strategy as described above (Figure 3.8). For *E. coli*, also as described above, the bins were generated using predicted sgRNA activity. For *B. subtilis* the bins were instead generated using the measured relative fitness values from the V1 experiment, and the selected sgRNAs were therefore a subset of those used in the V1 library.

The *dfrA, gfp,* and *rfp* V1 comprehensive libraries (used in the trimethoprim experiment and all FACS-seq experiments) were designed analogous to the V1 essential gene libraries, with 100 sgRNAs per target: 4x the original fully complementary spacer, 20x randomly chosen single-mismatches within 7 bases of the PAM, 15x randomly chosen single-mismatches 8-12 bases from the PAM, 12x randomly chosen single-mismatches 13-20 bases from the PAM, and 49x randomly chosen double mismatches 1-20 bases from the PAM (Figure 3.8).

For the V2 comprehensive libraries targeting *dfrA*, *murAA*, *folA*, or *murA*, we designed all possible non-template spacers, each with all possible single-mismatches, for a total of 60x mismatch variants per fully complementary sgRNA.

Relative fitness experiments

Relative fitness experimental details

Glycerol stocks of the *B. subtilis* essential-gene library (V1 or V2), the *dfrA* and *murAA* libraries (V1 or V2), and the library of non-targeting control sgRNAs were fully thawed, mixed, and inoculated into 150 mL cultures of LB at a combined OD600 of 0.01 (5% control, 75% essential-gene library, 10% *dfrA* library, 10% *murAA* library). This culture was allowed to grow to OD600 0.1, at which point the culture was back-diluted to OD600 0.01 in fresh 150 mL culture of LB + 1% xylose. This culture was then grown to OD600 0.3 (~5 doublings), back-diluted to OD600 0.01 in LB + 1% xylose, and grown to OD600 0.3 (total ~10 doublings). Samples were collected a) immediately before back dilution into xylose and b) after the final growth phase, ~10 doublings apart (Figure. 2A). The trimethoprim experiments were carried out in an identical manner, except that both 1% xylose and trimethoprim (Sigma-Aldrich #T7883-5G) (0 ng/mL, 15ng/mL, or 30ng/mL) were added from the first back-dilution and maintained throughout growth. Concentrations of trimethoprim were chosen such that wildtype growth rate was unaffected.

Fitness experiments for the *E. coli* V2 libraries were carried out in an identical manner to the *B. subtilis* fitness experiments with the following exceptions: all growth occurred in the presence of ampicillin, and induction was achieved with 1mM IPTG instead of 1% xylose.

For both *B. subtilis* and *E. coli*, compact library experiments were carried out in an identical manner as the larger scale fitness experiments above, save that the volume of cultures was 15mL, and only compact libraries and non-targeting control libraries were mixed together (90% compact library, 10% controls).

At the desired time points, *B. subtilis* cultures were collected (1ml) by pelleting (9000xg 2min) and genomic DNA was extracted using the DNeasy Blood & Tissue kit (Qiagen #69506) with the recommended Gram-positive pre-treatment and RNAse A treatment. For the *E. coli* fitness experiments, *E. coli* cultures were collected (4ml) by pelleting (20000xg 2min) and plasmid DNA was extracted using the QIAprep Spin miniprep kit (Qiagen #27106). sgRNA spacer sequences were amplified from gDNA or plasmid DNA using Q5 polymerase (New England Biolabs #M0493S) for 14x cycles using custom primers containing TruSeq adapters and indices (Table S8), followed by gel-purification from 8% TBE gels (Invitrogen #EC62152BOX), and sequencing on HiSeq 4000 with single-end 50bp reads at the UCSF Center for Advanced Technology using a custom sequencing primer (Table S8).

Relative fitness analysis

Raw FASTQ files were aligned to the library oligos and counted using (*GitHub - traeki/sgrna_design*, no date) (count_guides.py), and relative fitness was calculated using (compute_gammas.py and gamma_to_relfit.py). For each strain (x) with at least 100 counts at t_0 we calculate the relative fitness $F(x)$ according to:

$$F(x) = \frac{log2\frac{r_{wt}(t_0) * r_x(t_{10})}{r_{wt}(t_{10}) * r_x(t_0)}}{g_{wt}} + 1$$

where $r_x(t_i)$ is the fraction of strain *X* in the population at time *i* and g_{wt} is the number of generations of wildtype growth in the experiment. A derivation of this equation can be found in (Keren et al., 2016) and (Rest et al., 2013). In our experiments, g_{wt} is calculated from the OD measurements of the culture, and $r_{wt}(t_i)$ is calculated as the median of 1000 non-targeting control sgRNAs from that sample. For strains with at least 100 counts at t_0 and 0 counts at t_{10}, we set:

$$log2\frac{r_x(t_{10})}{r_x(t_0)} = 0$$

Finally, the relative fitness measurements of each sgRNA were averaged across samples (*B. subtilis* experiments: 6 replicates, *E. coli* experiments: 4 replicates) to calculate the final relative fitness value and standard deviation (Table S3).

Detection limits of relative fitness measurements

Our relative fitness experiments seek to quantify the number of doublings each strain experiences during the course of the experiment, relative to the number of doublings a wild-type (or a non-targeting sgRNA control) strain experiences during this time. To do so, we measure the bulk growth of the population, and quantify the relative abundance of each strain at the start and end of each experiment *via* next-generation sequencing. Changes in the relative abundance of a strain are determined by the growth rate of the individual strain relative to the population as a hole. For example, there is a 2^{10} ~ 1,000-fold increase in the number of cells during a 10-doubling experiment. Therefore, cells that do not divide (but remain intact) will experience a 1,000-fold decrease in relative abundance.

Our ability to measure the relative abundance of strains is constrained by sequencing depth. Assuming an equal number of reads at the start and end of the experiment, measurement of a 1,000-fold decrease in relative abundance requires that a strain have at least 1,000 reads at the start of the experiment. A poorly represented strain (e.g. 50 read counts at the start of the experiment) cannot decrease 1,000-fold and be meaningfully measured.

Previously reported pooled fitness experiments of CRISPRi libraries in *E. coli* prioritized sensitivity to slight growth defects over quantifying the extent of a strong fitness defect (Figure 3.13D-E). To do so, these experiments were run for many generations (15+) and were sequenced with relatively less depth (median counts ~100). This limited their ability to quantify strong fitness effects. In contrast, this study prioritized quantification of the full range of possible fitness outcomes. As a result, our experiments were run for 10 generations and deeply sequenced (median counts > 1,000), allowing us to quantify a broad range of fitness defects.

Many strains were abundant enough at the start of the experiment to allow accurate quantification of decreases greater than 2^{10} ~ 1,000-fold. These events (relative fitness < 0) represent active depletion from the pool.

Relative fitness validation

To validate the practice of using pooled growth measurements as an approximation of relative fitness, we also measured the relative fitness of individual *dfrA* knockdown strains grown in the presence of a wildtype strain. For each *dfrA* sgRNA, the spacer was cloned separately into pJSHA77-gfp and pJSHA77-rfp, each transformed into the *dcas9* strain, and then competed against a wildtype constitutively expressing the opposite fluorophore (i.e. strains with a *dfrA* sgRNA and expressing *gfp* were competed against a wildtype expressing *rfp*). Strains were mixed at a starting OD600 of 0.01 in 300µL of LB in four replicate wells of a 96-well deep-well plate, covered with a breathable film, and grown shaking at 900 RPM at 37C. Cells were diluted to OD600 0.01 in fresh LB with 1% xylose and grown again (900 RPM, 37C) to OD600 0.3.

Immediately after each back-dilution (and at end of experiment) the previous plate was fixed with 50µL of 37% formaldehyde per well, incubated for 10min at room temperature, and quenched with 50µL of 2.5 M glycine. The quenched reaction was diluted 1:20 into 1X PBS before measurement by flow cytometry (LSRII, BD Biosciences) using the blue laser (488 nm) and the FITC detector (530/30 nm) for GFP detection, and the yellow/green laser (561 nm) and the PE-Texas Red detector (610/20 nm) for RFP detection. Data for at least 20,000 cells were collected, and thresholds based on control wells were used to define the GFP+ and RFP+ populations to determine the ratio of each population in each sample using FlowJo (FlowJo, LLC). All calculated relative fitness measurements from this validation experiment are provided in Table S3.

Relative expression measurements

Growth and RT-qPCR

Reconstructed *E. coli* CRISPRi strains targeting genes involved in peptidoglycan precursor biosynthesis were grown in triplicate from single colonies in pre-warmed 4ml LB + ampicillin for 2.5hrs before back-dilution (1:80) in pre-warmed 4ml LB + ampicillin + 1mM IPTG and growth for 3hr prior to collection (OD600 ~ 0.2). The control strains express *rfp* and harbor sgRNA plasmids expressing sgRNAs targeting either *rfp* or *gfp* (“non-targeting”) and were treated identically. Samples were collected (300ul) in 900ul TRIzol-LS (Thermo Fisher #10296010) and stored at -20C overnight. The following day RNA was extracted according to the TRIzol protocol. RNA was quantified using a NanoDrop 2000c Spectrophotometer (Thermo Scientific) to normalize input (500ng input / 20ul reaction). For each RT-qPCR probe set and each sample replicate, reactions were performed in triplicate.

All RT-qPCR assays were done using the Luna Universal One-Step RT-qPCR kit (New England Biolabs #E3005S) according to its RT and cycling protocols, in 96 well PCR plates (Neptune #3732.X) and measured on a CFX Connect Real-Time System (Bio-Rad).

RT-qPCR analysis

Standard curves for each primer pair were first assessed on serially diluted RNA (extracted from the CRISPRi control strain) to confirm single melting peaks, strong correlations of technical replicates, and to calculate their efficiencies (in accordance with (Bustin et al., 2009)). The relative expression (or Normalized Relative Quantity (NRQ)) of each gene of interest in each experimental sample was calculated according to (Hellemans et al., 2007), which uses the geometric mean of two reference genes (here: *atpB* and *recA*) to normalize the probe of interest within each sample, and further calculates the fold-change in relative expression compared to a wildtype strain. The "non-targeting" *rfp*+ strain was considered the wildtype for the normalization of all other strains.

FACS-seq experiments

FACS-seq experimental details

Three separate strain libraries were constructed and mixed together for use in the sorting experiments: a gfp^+ strain with the *gfp*-targeting sgRNA library (mismatch-GFP), a gfp^+ strain with the non-targeting sgRNA control library ("high-GFP" or "control sgRNA" in figure), and a gfp^- strain with the *rfp*-targeting sgRNA library ("no-GFP" or "dark control") (Figure 3.1A). Glycerol stocks of each library were fully thawed, inoculated into replicate 12.5ml cultures of LB (*B. subtilis*) or LB with ampicillin (*E. coli*) at 0.01 OD600, and allowed to grow for 2.5-3hr. Then cultures were back-diluted to 0.01 OD600 in LB with 1% xylose (*B. subtilis*) or LB with ampicillin (*E. coli*) and grown for 2.5hr. Immediately before sorting the cultures were mixed at a ratio reflecting the overall diversities of their libraries (40% mismatch-GFP, 40% low-GFP, 20% high-GFP), and then the mixture was diluted 1:10 in PBS at room temperature (*B. subtilis*) or on ice (*E. coli*).

Sorting was done on the mixed cultures using a BD FACSAria II (Laboratory for Cell Analysis in Helen Diller Family Comprehensive Cancer Center at UCSF), using the blue laser (488 nm) and the FITC detector (530/30 nm), and at a flow rate of 5 and collecting for 20min total. Post-sorting the collected bins were filtered using either cellulose nitrate membranes with 0.2um pore (Thermo Scientific #145-0020) or mixed cellulose esters 0.22um pore disc filters (MF-Millipore #GSWP02500) on a glass filtration apparatus. Filters were resuspended in 9ml LB (*B. subtilis*) or LB with ampicillin (*E. coli*) by vortexing at max speed for 30s, then split into two outgrowth cultures and grown overnight in 4ml LB (*B. subtilis*) or LB with ampicillin (*E. coli*). A portion of the input mixed sample (i.e. pre-sorting) was treated similarly and grown overnight. DNA was extracted from each outgrowth culture separately and analyzed by deep sequencing as described above.

FACS-seq analysis

For each species, two biological replicates (i.e. cultures starting from unique glycerol stocks) were sorted by FACS, and from each biological replicate's 4 bins (plus unsorted mixture) two technical replicates (i.e. two overnight outgrowth cultures from which DNA was extracted) were sequenced. Library spacers were counted in each sequenced sample, normalized to the sample's total number of spacers counted, and technical replicate normalized counts were added together. For each biological replicate, the sorted bins were further normalized with respect to the mixed (i.e. pre-sorting) sample in the following manner: a linear model was used to determine the appropriate weights for each bin in order to recapitulate the mixed sample, and those weights were applied as scaling factors for all read counts from the given bin. This normalization was essential to correct for sequencing depth and cell number differences between bins. Briefly, we used the sklearn package (sklearn.linear_model) in Python and applied it to the mixed sample after removing from it the top and bottom 5^{th} percentiles.

We sought to define a metric for enrichment in the GFP-high bins vs. the GFP-low bins that would be similar in scale to relative fitness. We define an enrichment ratio (ER) for each sgRNA as:

$$ER = \frac{3}{3} n.norm_{Bin4} + \frac{2}{3} * n.norm_{Bin3} \frac{1}{3} * n.norm_{Bin2} + \frac{0}{3} * n.norm_{Bin1}$$

where $n.norm_{Bin\,i}$ is the normalized counts in Bin *i*, and Bin1 has the lowest GFP fluorescence while Bin4 has the highest. By this metric, values close to 1 have the highest GFP fluorescence (or weakest sgRNA activity) and values <1 have lower GFP fluorescence (or stronger sgRNA activity). Enrichment scores were normalized on a per experiment basis by subtracting the mean enrichment score of the "dark controls" and dividing by the mean enrichment score of the "high-GFP" strains. The resulting scores for each sgRNA (called the "FACS-seq score" in the main text) are available in Table S1.

FACS-seq validation

To validate our sorting procedure and the relationship between the calculated FACS-seq score and the fluorescence of a single strain, we randomly isolated 9 strains from the *E. coli* GFP knockdown library and analyzed them by flow cytometry to quantify knockdown relative to a non-targeting sgRNA (Figure 3.9E-H). Strains were grown in deep 96-well plates in 300ul LB overnight, diluted back and grown to ~0.4 OD600 before measurement. Briefly, data was collected on a LSRII flow cytometer (BD Biosciences) using the blue laser (488 nm) and the FITC detector (530/30 nm). Data for at least 20,000 cells were collected, and median fluorescence values were extracted using FlowJo (FlowJo, LLC). Data from representative samples were plotted as histograms using FlowJo to confirm that single-cell fluorescence was unimodal within the population (Figure 3.9E-H). sgRNA plasmids were miniprepped (Qiagen

#27106) from each library isolate and Sanger sequenced to ascertain their identity in the library experiment. To assay the behavior of the same sgRNAs in *B. subtilis,* the miniprepped plasmid was transformed into *B. subtilis* as described above, double-crossover events were verified by streaking on starch plates, and the strains were analyzed by flow cytometry as described above. All relative fluorescence measurements are provided in Table S1 and plotted in Figure 3.9A.

Predicted sgRNA activity validation

To validate the linear model's ability to predict sgRNA activity based on sgRNA sequence, we measured the knockdown of a *murAA-gfp* transcriptional fusion in a *B. subtilis* strain that was complemented by a non-targeted copy of *murAA*. These strains also expressed a chromosomal *rfp* that allowed for calculation of the GFP/RFP ratio on a per cell basis. Strains were grown as described above (FACS-seq validation), with the exception that *dcas9* was induced using 1% xylose after dilution. Data was collected on a LSRII flow cytometer (BD Biosciences) using the blue laser (488 nm) and the FITC detector (530/30 nm) for GFP detection, and the yellow/green laser (561 nm) and the PE-Texas Red detector (610/20 nm) for RFP detection. Data for at least 20,000 cells were collected, and the per-cell GFP/RFP ratios as well as the population median GFP/RFP ratios were extracted using FlowJo (FlowJo, LLC). Relative knockdown was normalized to a *murAA-gfp* strain lacking an sgRNA, after first subtracting the background GFP fluorescence from a non-fluorescent *B. subtilis* strain. Relative GFP fluorescence measurements are provided in Table S9.

Linear model of singly mismatched sgRNA efficacy

Having measured the ability of ~1,600 singly mismatched sgRNAs to knockdown GFP expression, we sought to build a model to predict the effect of mismatches on sgRNA efficacy. Since an enrichment score of 1 represent maximal GFP fluorescence, and a score of 0 represents no GFP fluorescence, we define knockdown for each sgRNA as:

$$knockdown_{sgRNA} = 1 - FACS.seqscore$$

We then normalized the ability of each mismatched sgRNA to knockdown GFP compared to its equivalent fully complementary sgRNA using the equation below:

$$sgRNAactivity_{singlymismatchedsgRNA} = \frac{knockdown_{singlymismatchedsgRNA}}{knockdown_{fullycomplementarysgRNA}}$$

We next built a model that fit the activity of each sgRNA using the position of the mismatch (from 0 to 19, with 19 being PAM proximal, one hot encoded), the transition of the mismatch (from X to Y, one hot encoded), and the GC% of the fully complementary sgRNA. Mismatched sgRNAs were excluded from the analysis if they were variants of fully complementary sgRNAs with less than 0.5 knockdown (as described above). The parameters from this model trained on *E. coli, B. subtilis,* or species-averaged per sgRNA activity are presented in Table S2 and Figure 3.12, the raw data in Table S1.

Expression-fitness relationship analysis

Quantifying similarity between fully complementary guides targeting the same gene

Our *gfp* based model predicts the activity of singly mismatched sgRNAs relative to the activity of the fully complementary sgRNA from which they are derived. To use this relative sgRNA activity as a proxy for absolute activity, fully complementary sgRNAs targeting the same gene should have the same activity. Since we cannot easily measure the sgRNA activity directly when targeting endogenous essential genes, we reasoned that we could validate this assumption by comparing fitness effect of fully complementary sgRNAs targeting the same gene (plotted in Figure 3.14).

To determine whether fully complementary sgRNAs targeting the same essential gene had similar effects, we compared the total sum of squares (totalSS) to the within gene sum of squares (withinSS) for fully complementary sgRNAs targeting essential genes. In *E. coli*, the withinSS accounted for 26.7% of the totalSS and in *B. subtilis* the withinSS accounted for 18.6% of the totalSS. This suggests that fully complementary sgRNAs targeting the same gene are substantially more similar with regards to their fitness outcomes than fully complementary sgRNAs as a whole, and supports the assumption that fully complementary sgRNAs targeting the same gene have similar levels of activity.

Expression-fitness relationship analysis details

In order to quantitatively assess the expression-fitness relationship of genes targeted by the V2 *E. coli* and *B. subtilis* libraries, we developed a per gene pipeline, described below.

1. In general, fully complementary sgRNAs targeting the same gene had similar fitness effects (Figure 3.17), suggesting that all fully complementary sgRNAs induce a similar level of knockdown. We identified outlier sgRNAs that were significantly less effective at inducing a fitness defect (and therefore were likely to be ineffective at knocking down their target) by comparing the distribution of fitness values for each series (series = the fully complementary sgRNA and its 9 singly mismatched variants) to the fitness distribution of the remaining series targeting the same gene. Using a two-sided t-test, we assessed whether the distribution of their relative fitness values was significantly different ($p < 0.05$) from the relative fitness distribution of the remaining sgRNAs targeting the gene. If their distribution was significantly different and their mean relative fitness was higher than the other sgRNAs targeting the same gene, we surmised that the fully matched sgRNA was likely not functional and excluded its series from further analysis.

2. We next predicted the sgRNA activity of all sgRNAs using the model of sgRNA efficacy described above trained on the two species averaged GFP data also described above. Consistent with the definition of sgRNA activity above, fully complementary sgRNAs were assigned an sgRNA activity of 1.
3. We binned sgRNAs that passed our filter (Step 1) based on their predicted sgRNA activity (bin width = 0.2, bin spacing = 0.05, for a total of 17 bins), and within each bin we calculated the median relative fitness. A fully healthy (relative fitness = 1, predicted sgRNA activity = 0) pseudocount was included for each gene. Per gene bin medians for essential genes can be found in Tables S5.

Per gene bin medians were used in all analyses of gene expression-fitness relationship similarity.

Per gene and per sgRNA family correlation

For both *E. coli* and *B. subtilis,* per sgRNA family correlations were calculated for sgRNA families that passed the filter described above, had at least one relative fitness value less than 0.7, and had measurements for at least 6/10 possible sgRNAs. Similarly, per gene correlations for genes with a last bin fitness less than 0.7 and computed using sgRNA families that passed the filter described above.

Gene expression-fitness relationship clustering and enrichment analysis

To determine whether per gene expression-fitness curves were biologically meaningful, we clustered the bin medians (described above) for all essential gene in *E. coli* and *B. subtilis* into 9 clusters using k-means with 10,000 random restarts. Functional enrichment within clusters was calculated for COG categories, GO biological process terms, and KEGG terms using the hypergeometric test. Only p-values with Bonferroni corrected ($p < 0.05$) are shown in Table S6.

Gene similarity comparisons

To determine whether the expression-fitness relationships of genes within COG categories, GO biological process, or KEGG categories were more similar to each other than to those of other genes we first calculated pairwise Euclidean distances between the expression-fitness relationships of all essential genes within each species. We then used a two-sided t-test to compare the distances between genes within each category to the distances between those genes and genes in different categories. We accounted for CRISPRi polarity due to operon structure by excluding any distances between genes within the same operon (defined as two genes in the same direction <50bp apart) from both the "inside category" and the "outside category" set.

To determine whether the expression-fitness relationships of homologous genes were more similar to each other than to those of other genes in the opposing organism, we calculated the pairwise Euclidean distance between the expression-fitness relationships of all essential genes that have essential homologs in both *E. coli* and *B. subtilis* (n = 150, as defined in Koo et at., 2017). We next used a two-sided t-test to determine if the distance between homologs was, on average, different from the overall distribution of distances between these 150 genes (i.e. when one gene from one species is compared to the 149 genes in the opposing species). To determine which pairs of homologs were significantly dissimilar, for each gene pair (including homologs), we calculated how many cross-species comparisons involving either gene were more similar than the comparison in question. We compared this number in homologs and non-homologs to calculate a FDR.

Quantification and statistical analysis:

Statistical parameters—including r, R^2, SD—are reported in the Figures, Figure Legends, or Supplementary Tables, as indicated and described in Methods Details (above).

Data and code availability:

Data

All raw sequencing data is deposited in the Short Read Archive under accession PRJNA574461.

FIGURES

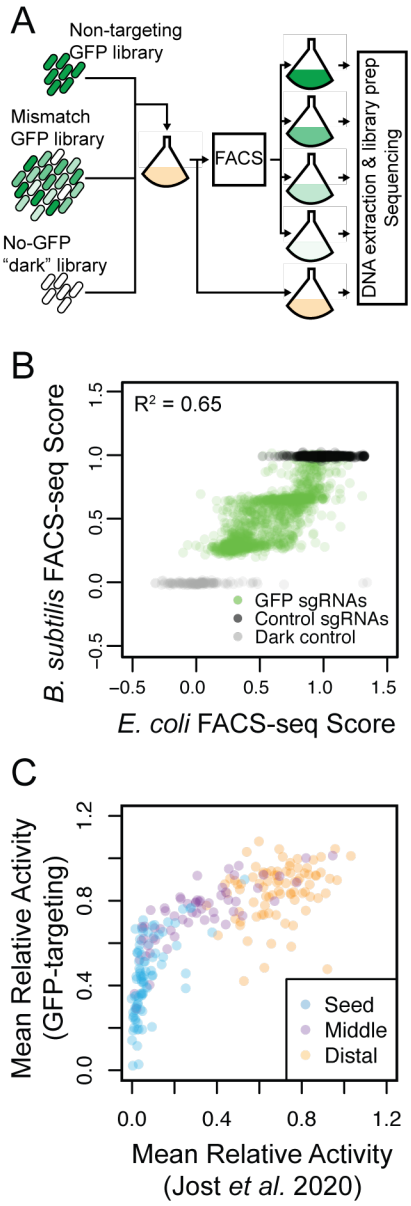


Figure 3.1. Singly mismatched sgRNAs reproducibly generate a range of knockdown efficacies in *B. subtilis* and *E. coli* but perform differently from a dCas9-KRAB system in mammalian cells. (A) Workflow of a FACS-seq experiment. (B) FACS-seq scores (average of 2 biological replicates) for each singly mismatched sgRNA targeting *gfp* in *B. subtilis* and *E. coli*. Additional noise in *E. coli* likely represents changes in plasmid copy number during outgrowth. (C) Mean relative activity of sgRNAs with all possible single base substitutions at every possible position in *E. coli* and *B. subtilis* targeting *gfp* data and in a mammalian CRISPRi system targeting essential genes (Jost et al., 2020), total of 26,248 mismatched sgRNAs)

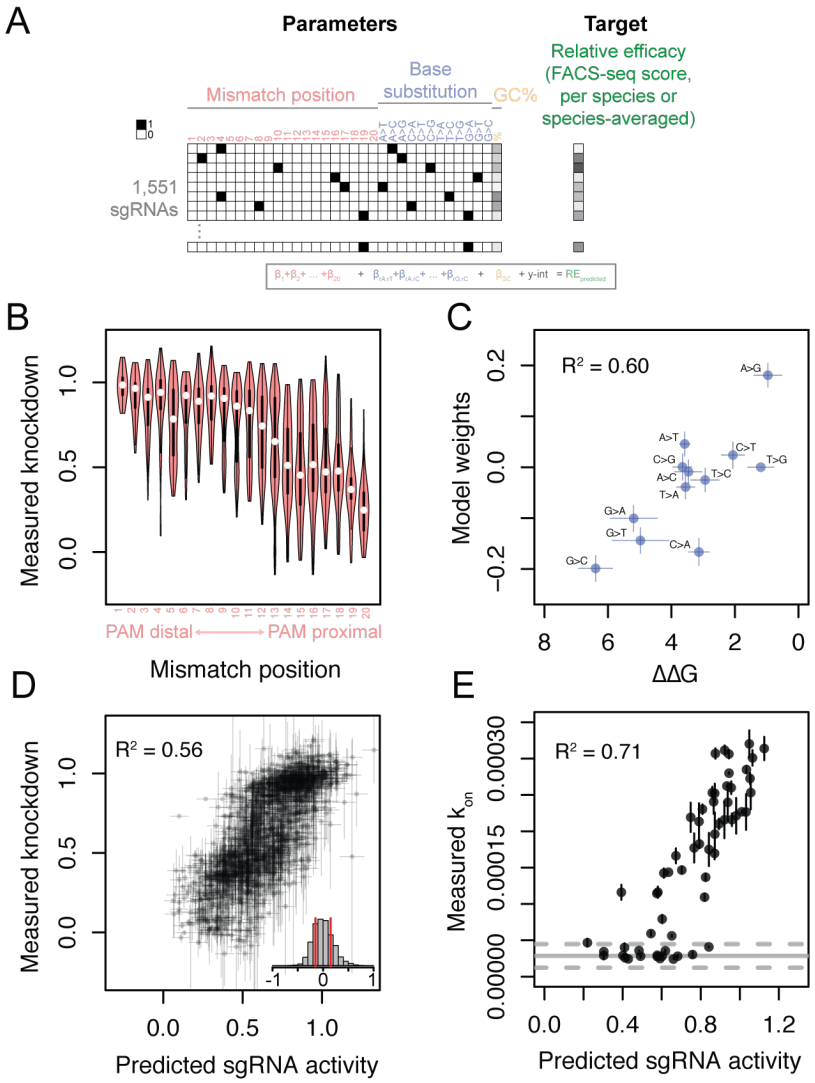


Figure 3.2. Mismatched sgRNA activity is accurately predicted by a simple linear model. (A) Schematic representation of a simple linear model for predicting the relative activity of mismatched sgRNAs. (B) Distributions of singly mismatched sgRNA relative activities by mismatch position. Each distribution represents 36-93 sgRNAs. (C) Comparison of model parameters for base substitution and the average ΔΔG of the mismatch calculated using a nearest neighbor approximation and the values from (Alkan et al., 2018). (D) The predictions of a linear model trained on GC%, mismatch position, and mismatch identity compared to the measured relative *gfp* knockdown efficacies of each sgRNA averaged over both species. Inset is a histogram of the differences between predicted and measured knockdown, reflecting both prediction and measurement error: 56% of sgRNAs measured within 0.15 of their predicted activity (red bars). (E) The predictions of the linear model compared to the measured singly mismatched sgRNA association rates (k_{ON}) *in vitro* (Boyle et al., 2017). Grey lines indicate the average (solid) and average +/- 1 SD (dashed) association rate of sgRNAs with mutated PAMs. Since such sgRNAs have no measurable association rate, this represents the detection limit of the assay in (Boyle et al., 2017).

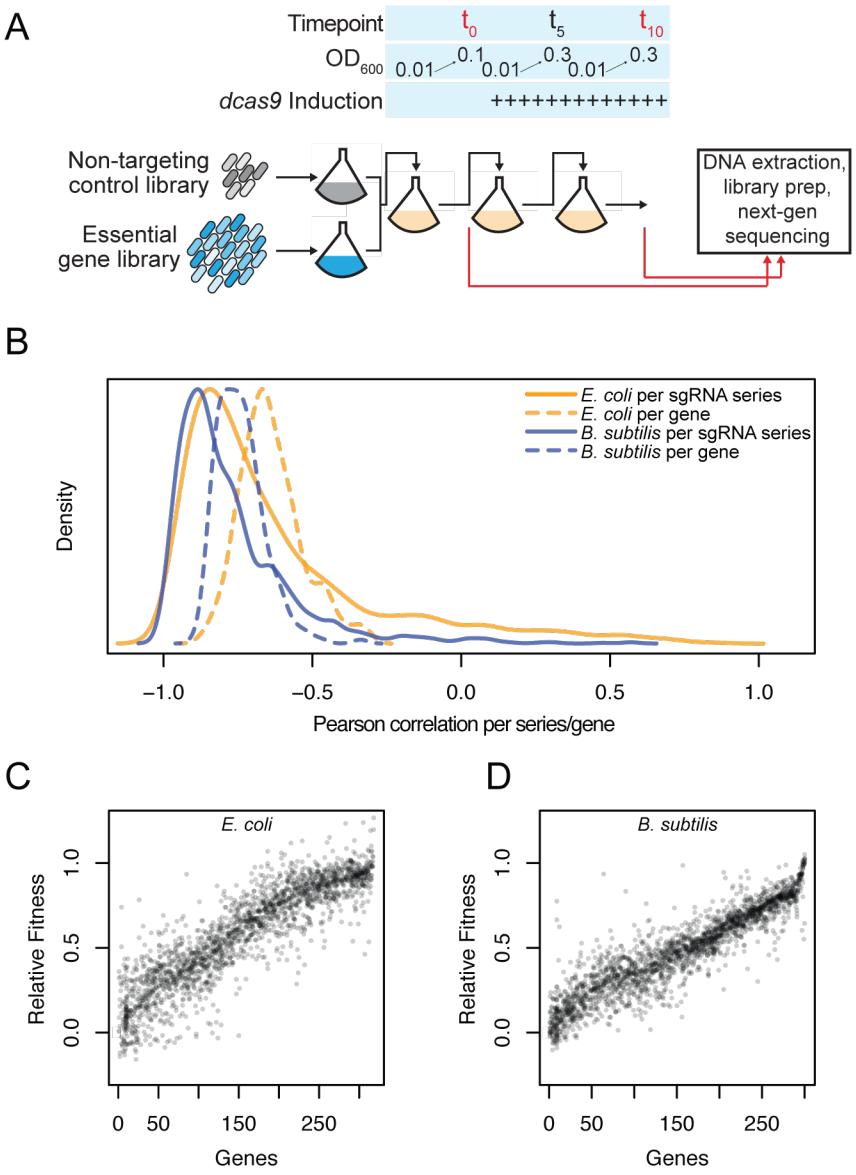


Figure 3.3. The expression-fitness curves of essential genes in *E. coli* and *B. subtilis* can be studied using singly mismatched sgRNAs. (A) Schematic of the fitness experiment design. (B) Distribution of per sgRNA locus (solid lines) and per gene (dashed lines) correlations (Pearson *r*) for sgRNAs targeting genes in *E. coli* (orange) and *B. subtilis* (blue). (C-D) The fitness effects of all fully complementary sgRNA targeting essential genes in *E. coli* (C) and *B. subtilis* (D) showing that the identity of the targeted gene is the driving factor in determining the fitness effect of an sgRNA. Genes are arranged in order of median fitness defect.

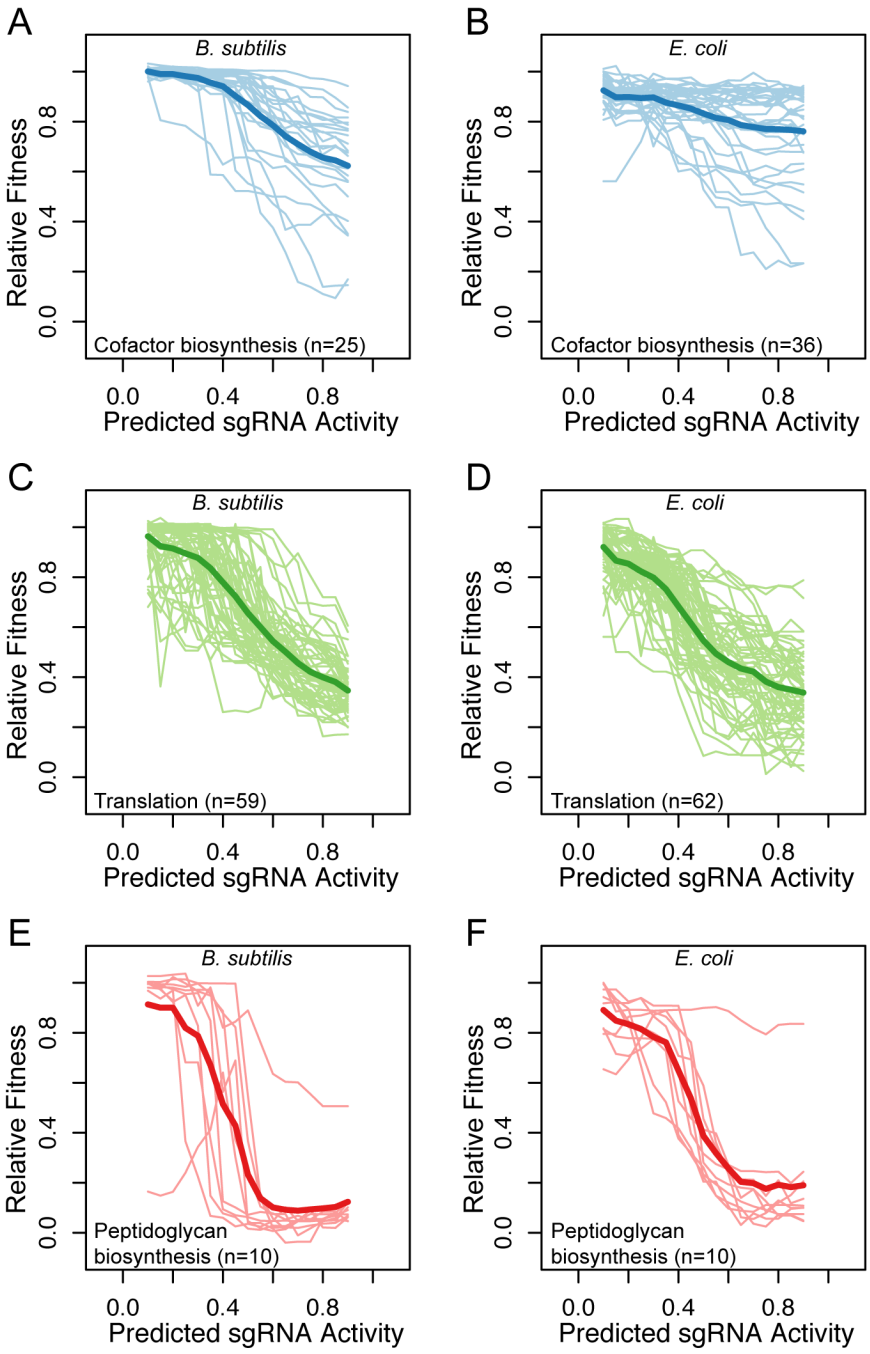


Figure 3.4. Expression-fitness relationships of essential genes are conserved within biological process and between *B. subtilis and E. coli.* Relative fitness compared to predicted knockdown for: essential cofactor biosynthesis genes (KEGG pathways under "Metabolism of cofactors and vitamins") in *B. subtilis* (A) or *E. coli* (B); KEGG pathways under "Translation" in *B. subtilis* (C) or *E. coli* (D); peptidoglycan biosynthesis (KEGG pathway ko00550) in *B. subtilis*. (E) or *E. coli* (F).

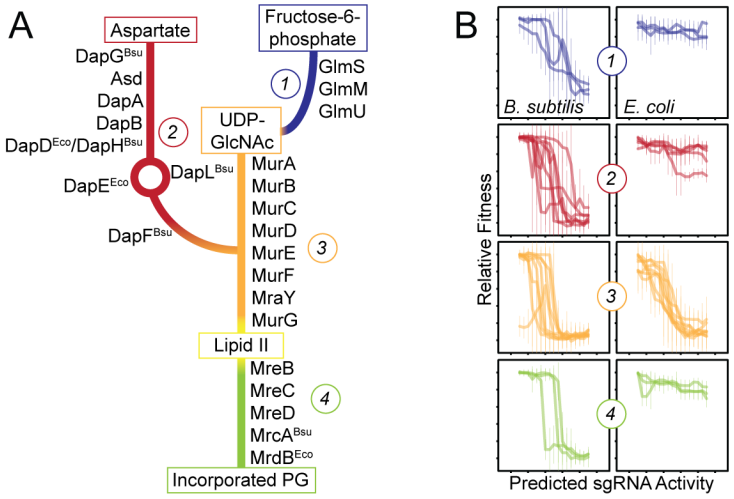


Figure 3.5. Similar and different expression-fitness relationships of cell wall biosynthesis genes in *B. subtilis* and *E. coli*. (A) Pathway of peptidoglycan synthesis and incorporation, color coded by portion of the pathway. (B) Predicted knockdown vs. relative fitness for the groups of essential genes from pathway sections indicated in (A), in *B. subtilis* and *E. coli*.

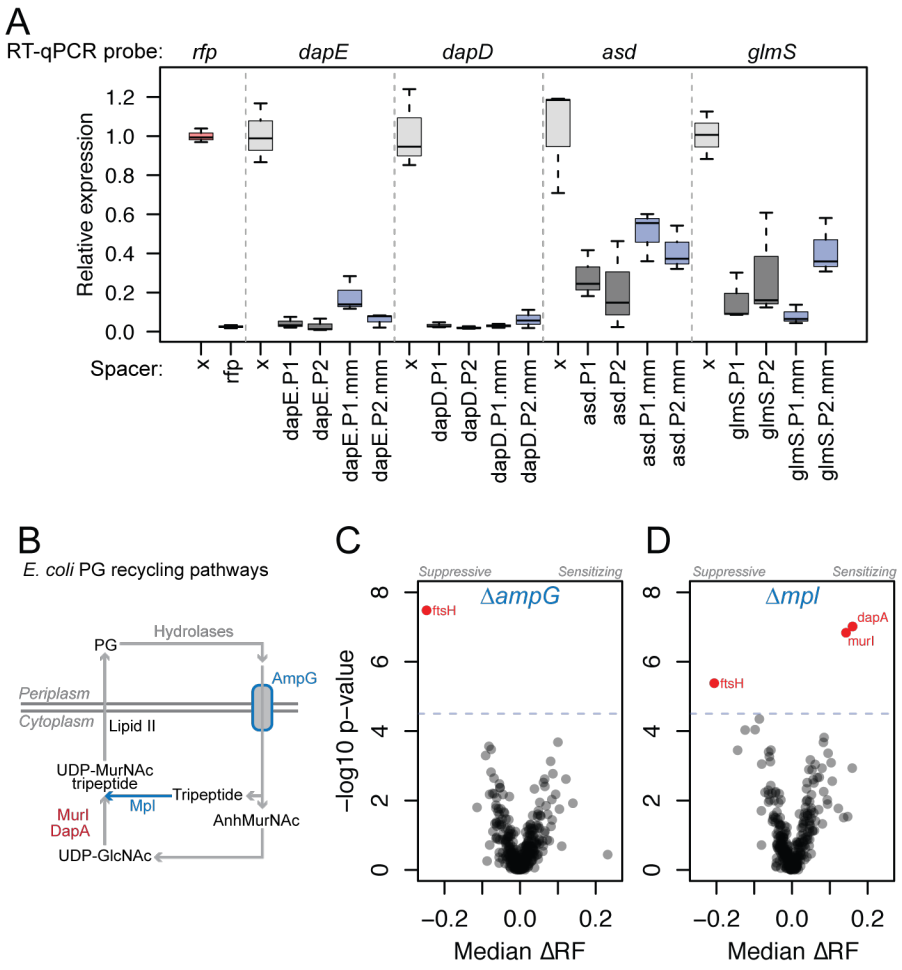


Figure 3.6. Screening mismatched sgRNA libraries in combination with genetic perturbations reveals modulators of essential gene requirements. (A) RT-qPCR measurements of repression by sgRNAs targeting 4 essential genes in *E. coli* shows that both mismatched and fully complementary sgRNAs are able to repress transcription. (B) Schematic of peptidoglycan recycling and synthesis pathway in *E. coli.* (C-D) Volcano plots comparing the median change in relative fitness for all sgRNAs targeting a gene to the statistical significance of those changes as quantified by a Wilcox test in the Δ*ampG* (C) and Δ*mpl* (D) genetic backgrounds. The dashed line represents a Bonferroni corrected *p*-value < 0.01

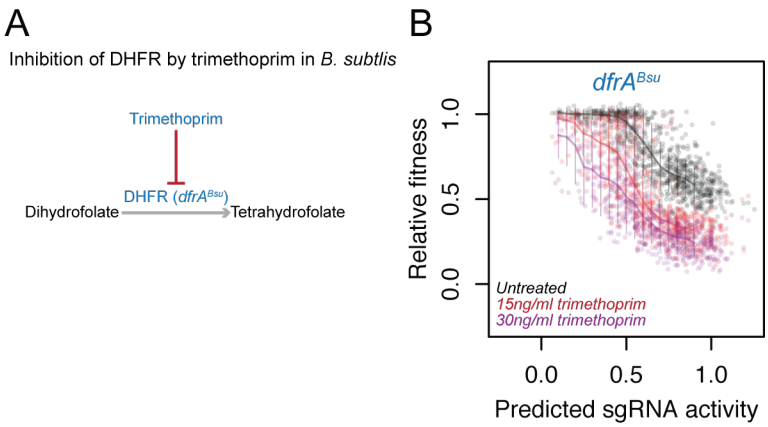


Figure 3.7. Environmental changes can modulate essential gene requirements. (A) Schematic of dihydrofolate reductase (*dfrA*^{Bsu}) inhibition by trimethoprim. (B) Expression fitness curve of *B. subtilis dfrA* in LB (black), LB+15ng/ml trimethoprim (red), and LB+30ng/ml trimethoprim (purple) showing the expression dependent synergy between DfrA depletion and trimethoprim.

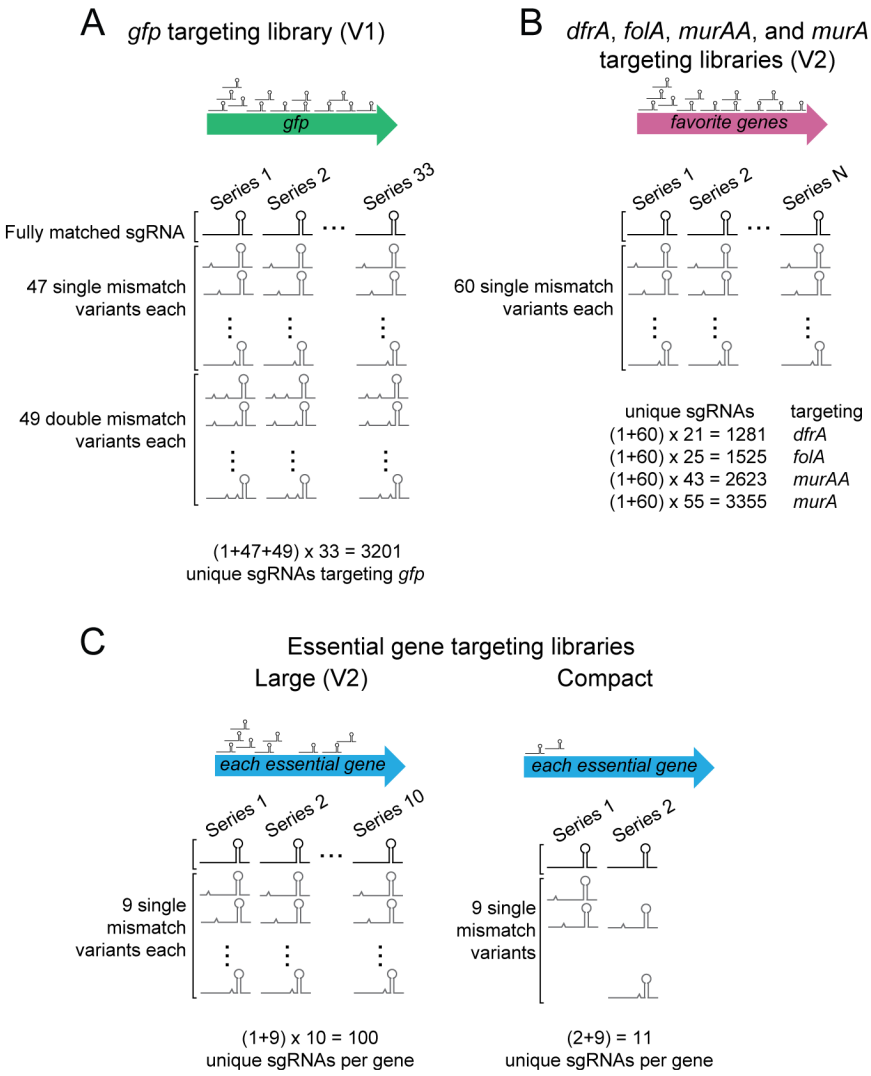


Figure 3.8. Design of mismatched sgRNA libraries targeting (A) *gfp*, (B) comprehensive libraries targeting *dfrA*, *folA*, *murAA*, and *murA*, and (C) each essential gene for the large libraries and the compact libraries. The breakdown of single mismatch variants per series and the total unique sgRNAs per gene are shown for each.

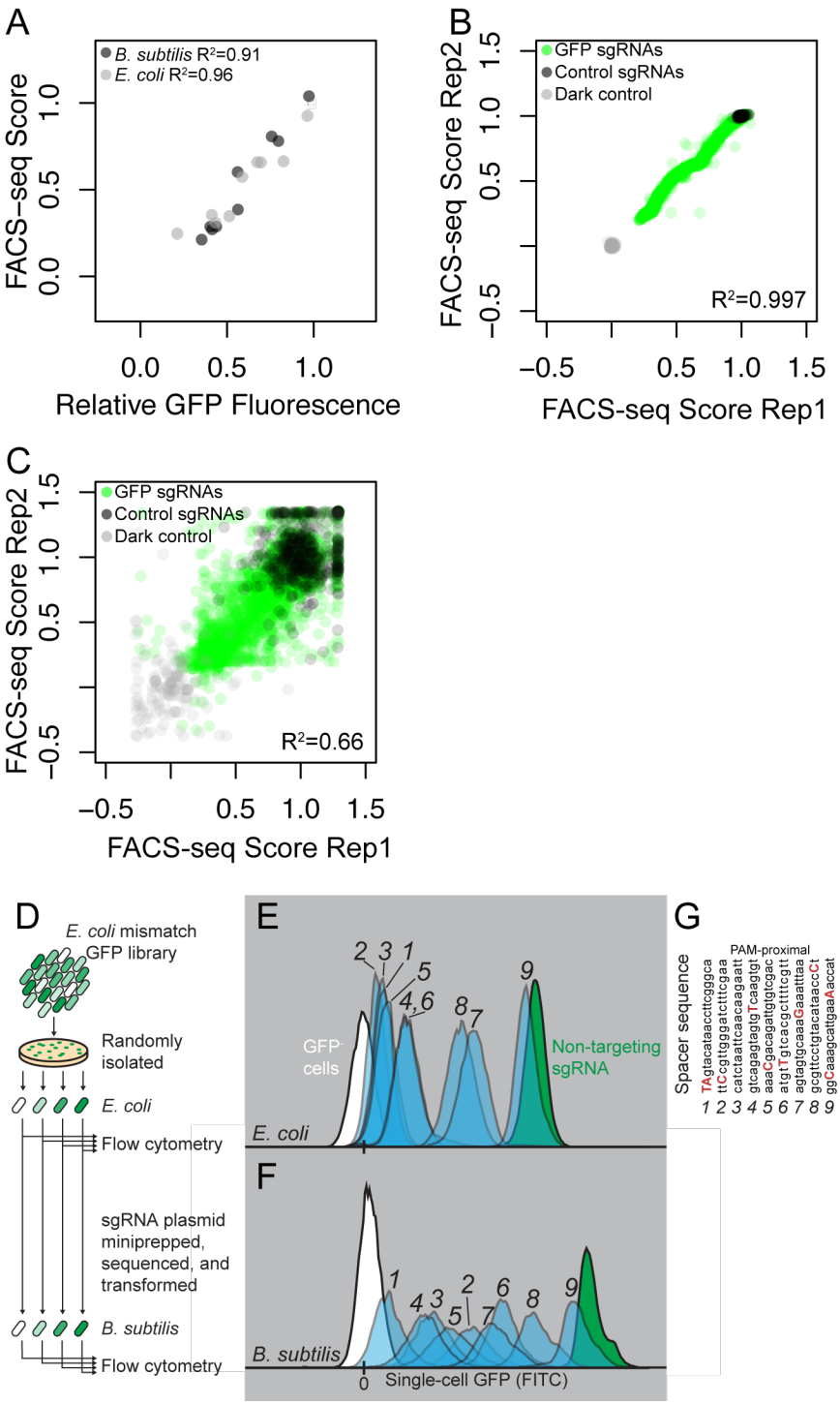


Figure 3.9. Details related to the linear model, FACS-seq data, and its validation. (A) Mismatched sgRNA efficacy measured individually (relative GFP fluorescence) in either *E. coli* or *B. subtilis* compared to their FACS-seq score from measurements done in the same species. Relative fluorescence is the median GFP single-cell fluorescence, normalized as a fraction of non-targeted control. (B) FACS-seq scores for all sgRNAs comparing two biological replicates in *B. subtilis*. (C) FACS-seq scores for all sgRNAs comparing two biological replicates in *E. coli*. Increased noise in *E. coli* likely reflects variation in sgRNA plasmid copy number at the time of

DNA extraction and/or sequence-based *E. coli* specific effects on sgRNA efficacy (Cui et al., 2018). (D) Schematic describing the isolation of random singly mismatched sgRNA strains from the *E. coli gfp* library, their analysis by flow cytometry, the introduction of the same sgRNA plasmids into *B. subtilis*, and their analysis by flow cytometry. (E-F) The distribution of single-cell GFP fluorescence values for strains of *E. coli* (E) or *B. subtilis* (F). (G) The sequences of the spacers indicated in (E) and (F), with mismatched bases highlighted in red.

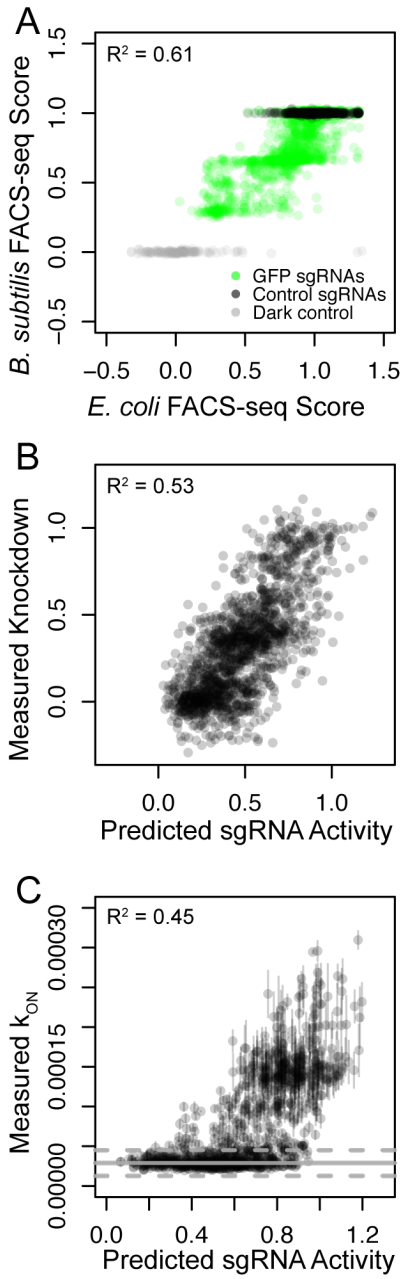


Figure 3.10. Doubly mismatched sgRNAs are accurately predicted as the combined independent effects of singly mismatched sgRNAs. (A) FACS-seq enrichment scores (average of 2 biological replicates) for each doubly mismatched sgRNA targeting *gfp* in *B. subtilis* and *E. coli*. (B) The predictions of the linear model for doubly mismatched sgRNA efficacy, treating each mismatch as independently affecting sgRNA efficacy, compared to the doubly mismatched sgRNAs' measured *gfp* knockdown efficacies (two species averages). (C) The predictions of the linear model for doubly mismatched sgRNAs compared to the measured doubly mismatched sgRNA association rates (k_{ON}) in vitro (Boyle et al., 2017). Grey lines indicate the average (solid) and average +/- 1 SD (dashed) association rate of sgRNAs with mutated PAMs. Since such sgRNAs have no measurable association rate, this represents the detection limit of the assay in (Boyle et al., 2017).

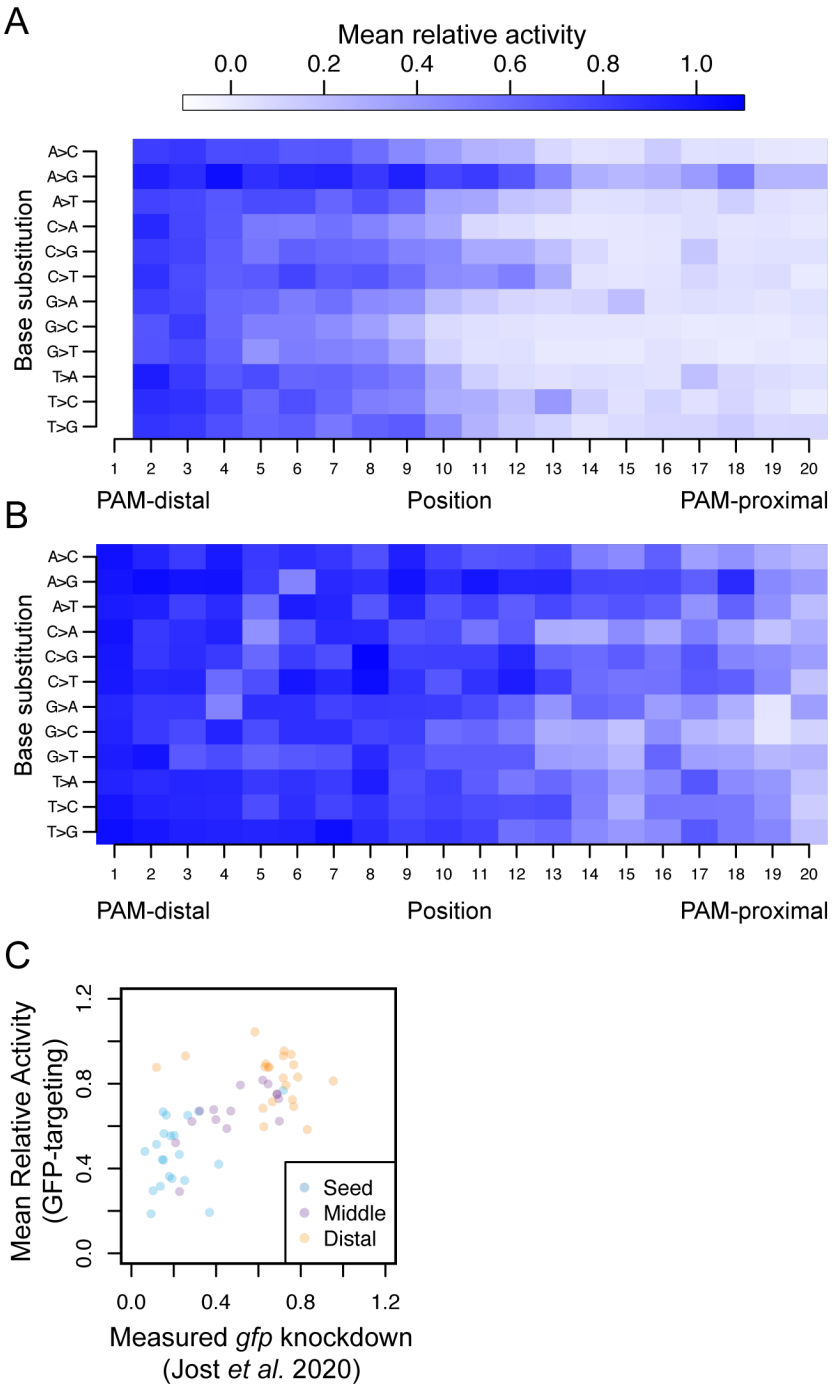


Figure 3.11. Mismatched sgRNAs affect CRISPRi efficacy similarly in mammalian systems and bacterial systems. (A) The mean relative activity of sgRNAs targeting essential mammalian genes with specific base substitutions at specific positions. Mismatches at position 20 are not shown because all sgRNAs contained a "G" at position 20 for compatibility with the U6 promoter. (B) The species-averaged mean relative activity of sgRNAs targeting *gfp* in *E. coli* and *B. subtilis* with specific base substitutions at specific positions. Darker color indicates stronger sgRNA activity. (C) Comparison of the species-averaged mean relative activity of sgRNAs targeting *gfp* in *E. coli* and *B. subtilis* with specific base substitutions at specific positions and the relative activity of singly mismatched sgRNAs targeting *gfp* in a mammalian system (Jost et al., 2020). Mismatch effect is correlated (R^2 = 0.44, $p < 10^{-7}$), but non-linear.

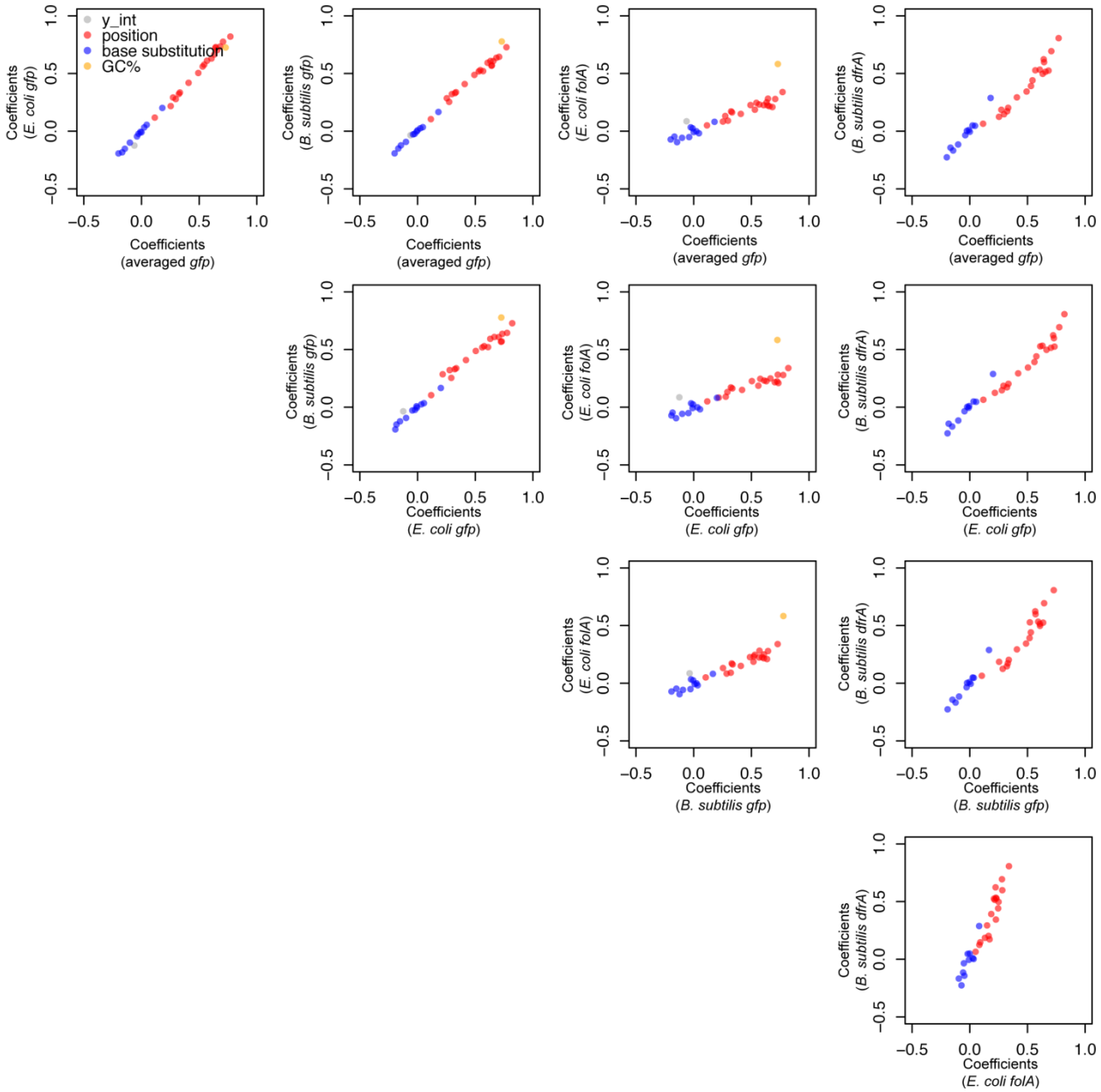


Figure 3.12. Parameters of linear models of singly mismatched sgRNA efficacy trained on FACS-seq or relative fitness data from either *E. coli* or *B. subtilis* have strongly correlated coefficient values. Each panel compares the coefficient values from the linear models trained on the two specified datasets.

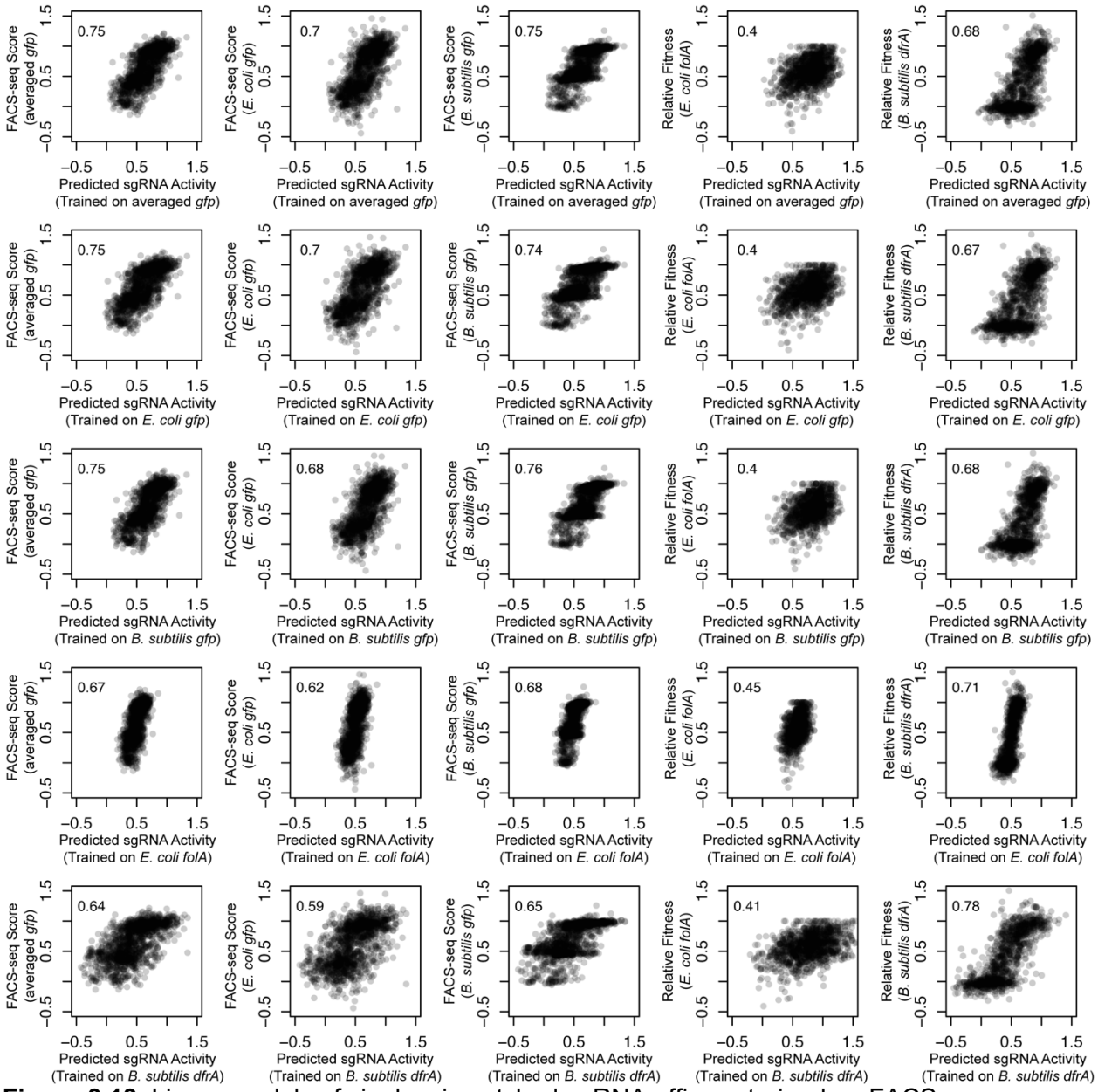


Figure 3.13. Linear models of singly mismatched sgRNA efficacy trained on FACS-seq or relative fitness data from *E. coli, B. subtilis,* or the average of both (averaged *gfp*) retain a majority of their predictive power on other singly mismatched sgRNA datasets. Each panel compares the predictions of a linear model trained on the specified dataset to the measured efficacy (relative fitness or FACS-seq score) of sgRNAs in the other specified dataset, with the pearson correlation coefficient shown in the inset. The datasets used and evaluated are, in order from top-bottom and left-to-right: species averaged *gfp, E. coli gfp*, *B. subtilis gfp*, *E. coli folA*, and *B. subtilis dfrA*.

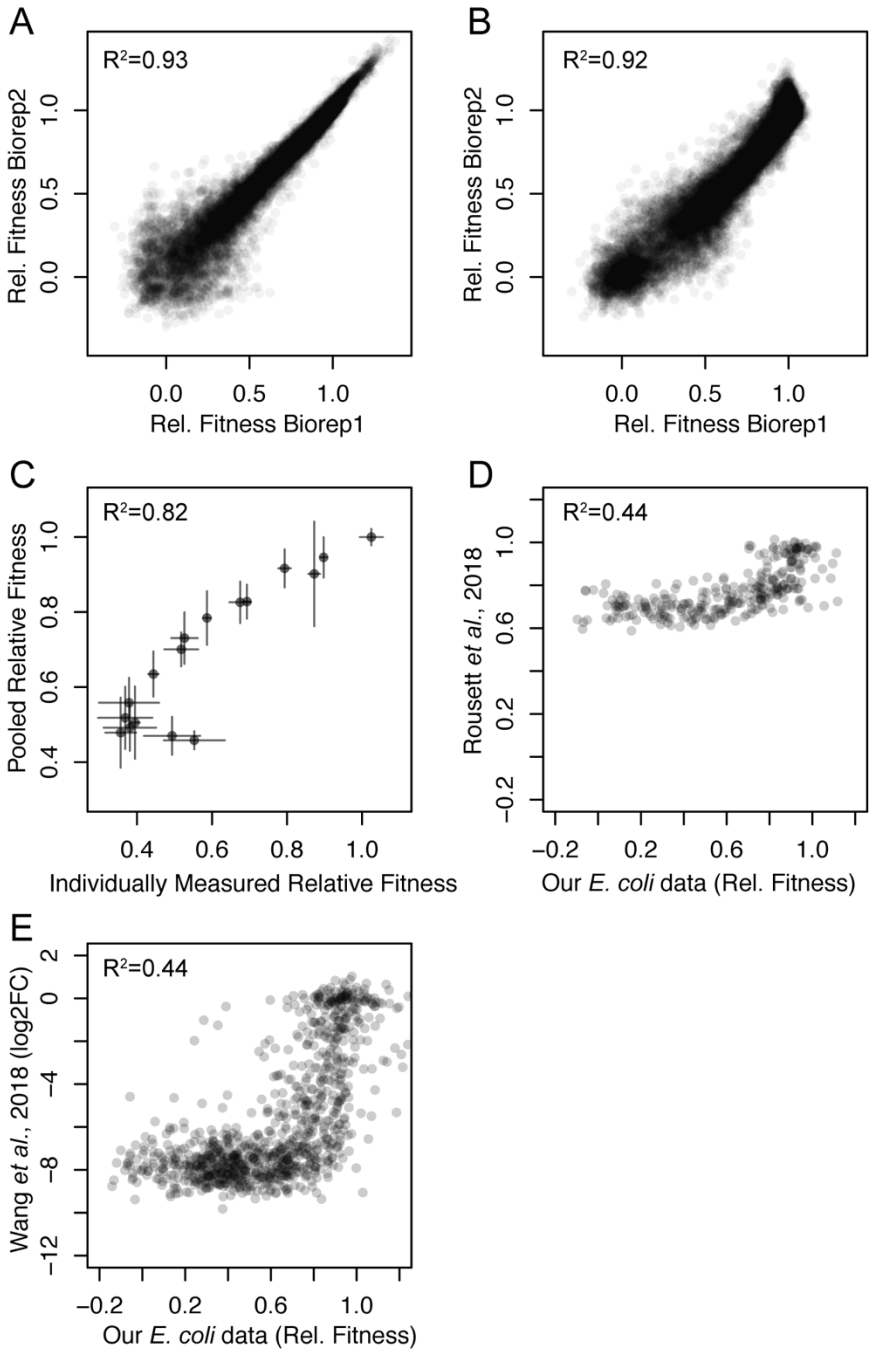


Figure 3.14. Relative fitness measurements are reproducible, orthogonally validated, and capture a large dynamic range. (A-B) Relative fitness measurements from two biological replicates in *E. coli* (A), and *B. subtilis* (B). (C) Relative fitness from pooled experiment compared to a relative fitness metric from competing individual *dfrA*-targeting CRISPRi strains against a fluorescently labeled wildtype and enumerating their relative abundance by flow cytometry before and after 10 doublings. (D & E) Per sgRNA relative fitness compared to previously reported fitness measurements (Rousset et al., 2018; Wang et al., 2018) showing the increased dynamic range of our measurements. The minimum quantifiable relative fitness can be approximated by the log2(median per sgRNA read count) divided by number of generations

of growth. In (Wang et al., 2018), median read count per sgRNA was ~100, and strains were grown for ~15 generations; therefore, relative fitness below ~0.6 is not resolvable. Similarly, in (Rousset et al., 2018), median read count per sgRNA was >200 (~17 million total counts, 92,919 elements), and strains were grown for ~17 generations; therefore, relative fitness below 0.6 is not resolvable.

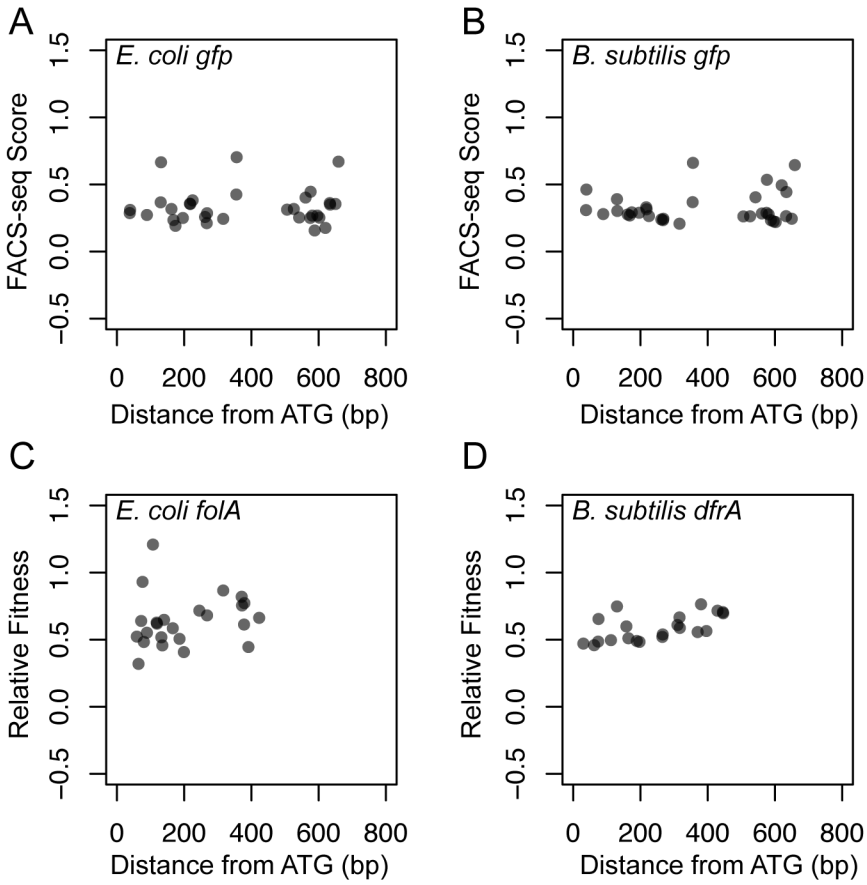


Figure 3.15. Fully complementary sgRNA efficacy is not significantly correlated with distance within the open reading frame in *E. coli* or in *B. subtilis*. (A-B) sgRNA distance from *gfp* ATG compared to its FACS-seq score in *E. coli* (A) and *B. subtilis* (B). (C) sgRNA distance from *folA* ATG compared to its impact on fitness in *E. coli*. (D) sgRNA distance from *dfrA* ATG compared to its impact on fitness in *B. subtilis*.

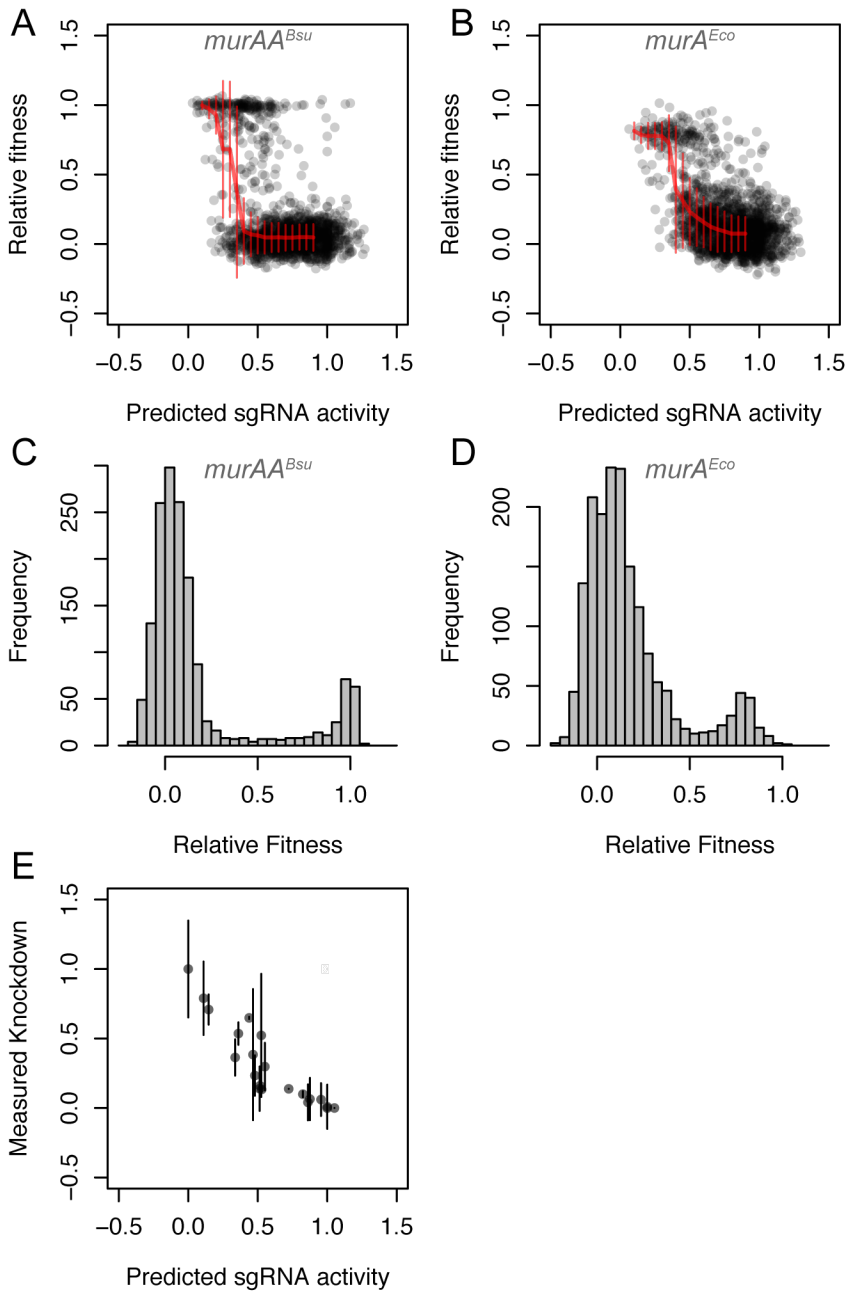


Figure 3.16. Singly mismatched sgRNAs targeting *E. coli murA* and *B. subtilis murAA* generate bimodal phenotypes that are not due to bimodal knockdown activity. (A-B) Predicted sgRNA activity and measured relative fitness of singly mismatched sgRNA targeting: (A) *murAA* in *B. subtilis*. (B) *murA* in *E. coli*. (C-D) Histogram of fitness outcomes for the same sgRNAs (E) Predicted sgRNA activity and relative expression for 18 singly mismatched sgRNA targeting a *murAA-gfp* transcriptional fusion in a *murAA*–complemented *B. subtilis* strain (Methods). Relative expression is shown as the median single-cell GFP fluorescence, normalized as a fraction of control (no sgRNA).

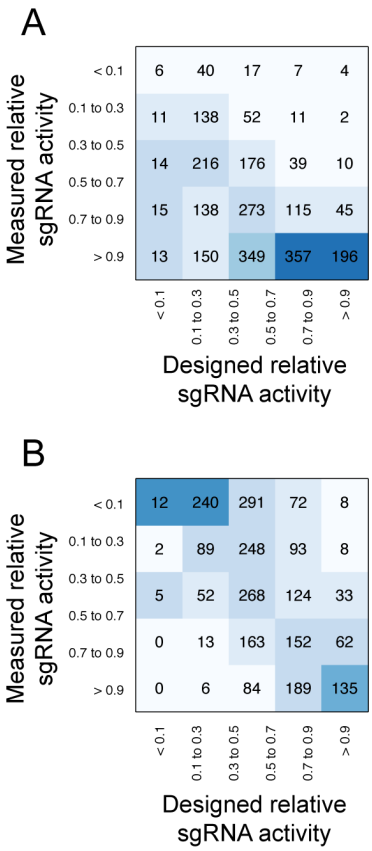


Figure 3.17. Design constraints and measured efficacy of singly mismatched sgRNAs in the compact libraries. (A) Confusion matrix showing the singly mismatched sgRNAs selected for the compact library based on their predicted knockdown vs. their measured relative fitness in *E. coli*. Designed relative fitness shows the predicted knockdown normalized by the strongest efficacy sgRNA's relative fitness. (B) Same as (A) for the compact library in *B. subtilis*.

TABLES

Table 3.1. Key resources

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Lysogeny broth (LB), Lennox	Fisher scientific	Cat# BP1427-2
Bacillus subtilis MC medium	Koo *et al.*, 2017	N/A
Bacillus subtilis competence medium	Koo *et al.*, 2017	N/A
IPTG	Denville scientific	Cat# C18280-13
Xylose		
Ampicillin sodium salt	Sigma-Aldrich	Cat# A9518
Kanamycin sulfate	Sigma-Aldrich	Cat# K1377
Erythromycin	Sigma-Aldrich	Cat# E5389
Spectinomycin dihydrochloride pentahydrate	Sigma-Aldrich	Cat# S9007
Chloramphenicol	Sigma-Aldrich	Cat# C0378
Carbenicillin	Millipore-Sigma	Cat# 205805
Gentamicin sodium salt	Fisher Scientific	Cat# AAJ1605103
Trimethoprim	Sigma-Aldrich	Cat# T7883-5G
Q5 High-Fidelity DNA polymerase	New England Biolabs	Cat# M0493S
HiFi Assembly	New England Biolabs	Cat# E2621L
BsaI-HFv2	New England Biolabs	Cat# R3733
T4 DNA Ligase	New England Biolabs	Cat# M0202L
Critical Commercial Assays		
DNeasy Blood & Tissue Kit	Qiagen	Cat# 69506
Midiprep Kit	Qiagen	Cat# 12143
QIAprep Spin miniprep kit	Qiagen	Cat# 27106
Deposited Data		
Raw sequencing data (FASTQs) for relative fitness experiments and FACS-seq experiments	This study	SRA: PRJNA574461

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Organisms/Strains		
Bacillus subtilis 168	BGSC	1A1
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm)*	Peters *et al.*, 2016	CAG74209
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), amyE::*P*veg-sgRNA(cat)* (CRISPRi libraries: sgRNA spacers listed in Table S3)	This study	N/A
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), thrC::*P*veg-gfp(Spc)*	This study	CAG78920
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), thrC::*P*veg-gfp(Spc)*, pJSHA77 (CRISPRi libraries: sgRNA spacers listed in Table S1)	This study	N/A
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), thrC::*P*veg-rfp(Spc)*	This study	CAG78921
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), thrC::*P*veg-rfp(Spc)*, pJSHA77 (CRISPRi libraries: sgRNA spacers listed in Table S1)	This study	N/A
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), sacA::*P*veg-rfp*	This study	CAG78922
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), sacA::*P*veg-rfp, murAA-gfp(Kan)*	This study	CAG78923
Bacillus subtilis 168 *lacA::*P*xyl-dcas9(Erm), sacA::*P*veg-rfp, murAA-gfp(Kan), thrC::*P*veg-murAA*(Spc)*	This study	CAG78924
Escherichia coli BW25113	Baba *et al.*, 2006	N/A
Escherichia coli BW25113 *Tn7att::*P*lLac-O1-dcas9(Gent)*	This study	CAG78830
Escherichia coli BW25113 *Tn7att::*P*lLac-O1-dcas9(Gent)*, pJSHA77 (CRISPRi libraries: sgRNA spacers listed in Table S3)	This study	N/A
Escherichia coli BW25113 *Tn7att::*P*BBa_J23105-dcas9(Gent), yjaA:*P*veg-gfp(Cat):yjaB*	This study	CAG78108
Escherichia coli BW25113 *Tn7att::*P*BBa_J23105-dcas9(Gent), yjaA:*P*veg-gfp(Cat):yjaB*, pJSHA77 (CRISPRi libraries: sgRNA spacers listed in Table S1)	This study	N/A
Escherichia coli BW25113 *Tn7att::*P*BBa_J23105-dcas9(Gent), yjaA:*P*veg-rfp(Cat):yjaB*	This study	CAG78107
Escherichia coli BW25113 *Tn7att::*P*BBa_J23105-dcas9(Gent), yjaA:*P*veg-rfp(Cat):yjaB*, pJSHA77 (CRISPRi libraries: sgRNA spacers listed in Table S1)	This study	N/A
10-beta Electrocompetent *Escherichia coli*	New England Biolabs	Cat# C3020K
Oligonucleotides		
Primers used in this study are listed in Table S8	This study	N/A
Recombinant DNA		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pDG1731	Radeck *et al.*, 2013	pBS4S (Addgene# 55170)
pDG1731-gfp	This study	N/A
pDG1731-rfp	This study	N/A
pDG1622	BGSC	ECE119
pJSHA77	This study	N/A
pJSHA77-rfp	This study	N/A
pJSHA77-gfp	This study	N/A
Software and Algorithms		
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
sgRNA design (fully matched sgRNAs)	This study	https://github.com/traeki/sgrna_design
Linear model training (train_linear_model.py)	This study	https://github.com/traeki/mismatch_crispri
Design a subset of mismatch sgRNA (choose_guides.py)	This study	https://github.com/traeki/mismatch_crispri
FASTQ analysis to calculate sgRNA abundance and relative fitness (count_guides.py, compute_gammas.py, gamma_to_relfit.py)	This study	https://github.com/traeki/mismatch_crispri
FlowJo v10	FlowJo, LLC	

REFERENCES

Alkan, F., Wenzel, A., Anthon, C., Havgaard, J.H., and Gorodkin, J. (2018). CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. Genome Biol. *19*, 177.

Barreteau, H., Kovac, A., Boniface, A., Sova, M., Gobec, S., and Blanot, D. (2008). Cytoplasmic steps of peptidoglycan biosynthesis. FEMS Microbiol. Rev. *32*, 168–207.

Bauer, C.R., Li, S., and Siegal, M.L. (2015). Essential gene disruptions reveal complex relationships between phenotypic robustness, pleiotropy, and fitness. Mol. Syst. Biol. *11*, 773.

Bhattacharyya, S., Bershtein, S., Yan, J., Argun, T., Gilson, A.I., Trauger, S.A., and Shakhnovich, E.I. (2016). Transient protein-protein interactions perturb E. coli metabolome and cause gene dosage toxicity. Elife *5*.

Bobrovskyy, M., and Vanderpool, C.K. (2016). Diverse mechanisms of post-transcriptional repression by the small RNA regulator of glucose-phosphate stress. Mol. Microbiol. *99*, 254–273.

Borkowski, O., Goelzer, A., Schaffer, M., Calabre, M., Mäder, U., Aymerich, S., Jules, M., and Fromion, V. (2016). Translation elicits a growth rate-dependent, genome-wide, differential protein production in Bacillus subtilis. Mol. Syst. Biol. *12*, 870.

Boy, E., and Patte, J.C. (1972). Multivalent repression of aspartic semialdehyde dehydrogenase in Escherichia coli K-12. J. Bacteriol. *112*, 84–92.

Boyle, E.A., Andreasson, J.O.L., Chircus, L.M., Sternberg, S.H., Wu, M.J., Guegler, C.K., Doudna, J.A., and Greenleaf, W.J. (2017). High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. Proc. Natl. Acad. Sci. U. S. A. *114*, 5461–5466.

Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. (2009). The MIQE guidelines: minimum

information for publication of quantitative real-time PCR experiments. Clin. Chem. *55*, 611–622.

Christodoulou, D., Link, H., Fuhrer, T., Kochanowski, K., Gerosa, L., and Sauer, U. (2018). Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia coli's Rapid Response to Oxidative Stress. Cell Syst *6*, 569–578.e7.

Cui, L., Vigouroux, A., Rousset, F., Varet, H., Khanna, V., and Bikard, D. (2018). A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9. Nat. Commun. *9*, 1912.

Dai, X., Zhu, M., Warren, M., Balakrishnan, R., Patsalo, V., Okano, H., Williamson, J.R., Fredrick, K., Wang, Y.-P., and Hwa, T. (2016). Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. Nat Microbiol *2*, 16231.

Datsenko, K.A., and Wanner, B.L. (2000). One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. *97*, 6640–6645.

Dekel, E., and Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. Nature *436*, 588–592.

Eames, M., and Kortemme, T. (2012). Cost-benefit tradeoffs in engineered lac operons. Science *336*, 911–915.

Fransen, F., Hermans, K., Melchers, M.J.B., Lagarde, C.C.M., Meletiadis, J., and Mouton, J.W. (2017). Pharmacodynamics of fosfomycin against ESBL- and/or carbapenemase-producing Enterobacteriaceae. J. Antimicrob. Chemother. *72*, 3374–3381.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-

Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Gong, S., Yu, H.H., Johnson, K.A., and Taylor, D.W. (2018). DNA Unwinding Is the Primary Determinant of CRISPR-Cas9 Activity. Cell Rep. *22*, 359–371.

Harris, L.K., and Theriot, J.A. (2016). Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size. Cell *165*, 1479–1492.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. *8*, R19.

Johnson, E.O., LaVerriere, E., Office, E., Stanley, M., Meyer, E., Kawate, T., Gomez, J.E., Audette, R.E., Bandyopadhyay, N., Betancourt, N., et al. (2019). Large-scale chemical-genetics yields new M. tuberculosis inhibitor classes. Nature *571*, 72–78.

Johnson, J.W., Fisher, J.F., and Mobashery, S. (2013). Bacterial cell-wall recycling. Ann. N. Y. Acad. Sci. *1277*, 54–75.

Jost, M., Santos, D.A., Saunders, R.A., Horlbeck, M.A., Hawkins, J.S., Scaria, S.M., Norman, T.M., Hussmann, J.A., Liem, C.R., Gross, C.A., et al. (2020). Titrating gene expression using libraries of systematically attenuated CRISPR guide RNAs. Nat. Biotechnol.

Kahan, F.M., Kahan, J.S., Cassidy, P.J., and Kropp, H. (1974). The mechanism of action of fosfomycin (phosphonomycin). Ann. N. Y. Acad. Sci. *235*, 364–386.

Kampmann, M., Bassik, M.C., and Weissman, J.S. (2013). Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. Proc. Natl. Acad. Sci. U. S. A. *110*, E2317–E2326.

Keren, L., Hausser, J., Lotan-Pompan, M., Vainberg Slutskin, I., Alisar, H., Kaminski, S., Weinberger, A., Alon, U., Milo, R., and Segal, E. (2016). Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness. Cell *166*, 1282–1294.e18.

Koo, B.-M., Kritikos, G., Farelli, J.D., Todor, H., Tong, K., Kimsey, H., Wapinski, I., Galardini, M., Cabal, A., Peters, J.M., et al. (2017). Construction and Analysis of Two Genome-Scale

Deletion Libraries for Bacillus subtilis. Cell Syst *4*, 291–305.e7.

Lalanne, J.-B., Taggart, J.C., Guo, M.S., Herzel, L., Schieler, A., and Li, G.-W. (2018). Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry. Cell *173*, 749–761.e38.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Liu, X., Gallay, C., Kjos, M., Domenech, A., Slager, J., van Kessel, S.P., Knoops, K., Sorg, R.A., Zhang, J.-R., and Veening, J.-W. (2017). High-throughput CRISPRi phenotyping identifies new essential genes in Streptococcus pneumoniae. Mol. Syst. Biol. *13*, 931.

Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res. *25*, 1203–1210.

Mengin-Lecreulx, D., and van Heijenoort, J. (1985). Effect of growth conditions on peptidoglycan content and cytoplasmic steps of its biosynthesis in Escherichia coli. J. Bacteriol. *163*, 208–212.

Mengin-Lecreulx, D., Texier, L., Rousseau, M., and van Heijenoort, J. (1991). The murG gene of Escherichia coli codes for the UDP-N-acetylglucosamine: N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis. J. Bacteriol. *173*, 4625–4636.

Nichols, R.J., Sen, S., Choo, Y.J., Beltrao, P., Zietek, M., Chaba, R., Lee, S., Kazmierczak, K.M., Lee, K.J., Wong, A., et al. (2011). Phenotypic landscape of a bacterial cell. Cell *144*, 143–156.

Nomura, M., Yates, J.L., Dean, D., and Post, L.E. (1980). Feedback regulation of ribosomal protein gene expression in Escherichia coli: structural homology of ribosomal RNA and ribosomal protein MRNA. Proc. Natl. Acad. Sci. U. S. A. *77*, 7084–7088.

Ogura, T., Inoue, K., Tatsuta, T., Suzaki, T., Karata, K., Young, K., Su, L.H., Fierke, C.A.,

Jackman, J.E., Raetz, C.R., et al. (1999). Balanced biosynthesis of major membrane components through regulated degradation of the committed enzyme of lipid A biosynthesis by the AAA protease FtsH (HflB) in Escherichia coli. Mol. Microbiol. *31*, 833–844.

Peters, J.M., Colavin, A., Shi, H., Czarny, T.L., Larson, M.H., Wong, S., Hawkins, J.S., Lu, C.H.S., Koo, B.-M., Marta, E., et al. (2016). A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria. Cell *165*, 1493–1506.

Peters, J.M., Koo, B.-M., Patino, R., Heussler, G.E., Hearne, C.C., Qu, J., Inclan, Y.F., Hawkins, J.S., Lu, C.H.S., Silvis, M.R., et al. (2019). Enabling genetic analysis of diverse bacteria with Mobile-CRISPRi. Nat Microbiol *4*, 244–250.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Rancati, G., Moffat, J., Typas, A., and Pavelka, N. (2018). Emerging and evolving concepts in gene essentiality. Nat. Rev. Genet. *19*, 34–49.

Rauch, B.J., Silvis, M.R., Hultquist, J.F., Waters, C.S., McGregor, M.J., Krogan, N.J., and Bondy-Denomy, J. (2017). Inhibition of CRISPR-Cas9 with Bacteriophage Proteins. Cell *168*, 150–158.e10.

Reis, A.C., Halper, S.M., Vezeau, G.E., Cetnar, D.P., Hossain, A., Clauer, P.R., and Salis, H.M. (2019). Simultaneous repression of multiple bacterial genes using nonrepetitive extra-long sgRNA arrays. Nat. Biotechnol.

Rest, J.S., Morales, C.M., Waldron, J.B., Opulente, D.A., Fisher, J., Moon, S., Bullaughey, K., Carey, L.B., and Dedousis, D. (2013). Nonlinear fitness consequences of variation in expression level of a eukaryotic gene. Mol. Biol. Evol. *30*, 448–456.

Reverend, B.D.-L., Boitel, M., Deschamps, A.M., Lebeault, J.-M., Sano, K., Takinami, K., and Patte, J.-C. (1982). Improvement of Escherichia coli strains overproducing lysine using

recombinant DNA techniques. European Journal of Applied Microbiology and Biotechnology *15*, 227–231.

Rodionov, D.A., Vitreschak, A.G., Mironov, A.A., and Gelfand, M.S. (2003). Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? Nucleic Acids Res. *31*, 6748–6757.

Rousset, F., Cui, L., Siouve, E., Becavin, C., Depardieu, F., and Bikard, D. (2018). Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and phage host factors. PLoS Genet. *14*, e1007749.

Schaechter, M., Maaloe, O., and Kjeldgaard, N.O. (1958). Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium. J. Gen. Microbiol. *19*, 592–606.

Scott, M., Gunderson, C.W., Mateescu, E.M., Zhang, Z., and Hwa, T. (2010). Interdependence of cell growth and gene expression: origins and consequences. Science *330*, 1099–1102.

Scott, M., Klumpp, S., Mateescu, E.M., and Hwa, T. (2014). Emergence of robust growth laws from optimal regulation of ribosome synthesis. Mol. Syst. Biol. *10*, 747.

Thomason, L.C., Costantino, N., and Court, D.L. (2007). E. coli genome manipulation by P1 transduction. Curr. Protoc. Mol. Biol. *Chapter 1*, Unit 1.17.

Thomason, L.C., Sawitzke, J.A., Li, X., Costantino, N., and Court, D.L. (2014). Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination. Curr. Protoc. Mol. Biol. *47*, 1.16.1–1.16.39.

Urban, J.H., and Vogel, J. (2008). Two seemingly homologous noncoding RNAs act hierarchically to activate glmS mRNA translation. PLoS Biol. *6*, e64.

Vigouroux, A., Oldewurtel, E., Cui, L., Bikard, D., and van Teeffelen, S. (2018). Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes. Mol. Syst. Biol. *14*, e7899.

Wang, T., Guan, C., Guo, J., Liu, B., Wu, Y., Xie, Z., Zhang, C., and Xing, X.-H. (2018). Pooled CRISPR interference screening enables genome-scale functional genomics study in bacteria with superior performance. Nat. Commun. *9*, 2475.

Zheng, Y., Struck, D.K., Bernhardt, T.G., and Young, R. (2008). Genetic analysis of MraY inhibition by the phiX174 protein E. Genetics *180*, 1459–1466.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:
Melanie Silvis
11249830E5F24D4...
Author Signature

2/19/2020
Date