Title
Exploring the functional constraints on bacterial essential gene expression with CRISPRi

Author
Hawkins, John Sumblin

Publication Date
2019

Exploring the functional constraints on bacterial essential gene expression with CRISPRi

by
John Sumblin Hawkins

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

in

Biophysics

in the

GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:
Carol Gross
7AFBE9046E0744C...

Carol Gross
Chair

DocuSigned by:
Jonathan S. Weissman, Ph.D.
[illegible]C9...

Jonathan S. Weissman, Ph.D.

DocuSigned by:
Martin Kampmann, Ph.D.
0A1E23B9758644E...

Martin Kampmann, Ph.D.

Committee Members

Chapter 2 is reprinted largely as it appears in:

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.

Conceptualization, J.M.P., A.C., H.S., L.S.Q., K.C.H., and C.A.G.; Methodology, J.M.P., A.C., H.S., T.L.C., M.H.L., B.M.K., J.S.W., E.D.B., L.S.Q., K.C.H., and C.A.G.; Investigation, J.M.P., A.C., H.S., T.L.C., M.H.L., S.W., C.H.S.L., E.M., and E.H.W.; Formal Analysis, J.M.P., A.C., H.S., T.L.C., M.H.L., J.S.H., A.L.S., J.S.W., E.D.B., L.S.Q., K.C.H., and C.A.G.; Writing-Original Draft, J.M.P., A.C., H.S., L.S.Q., K.C.H., and C.A.G.; Writing-Review & Editing, J.M.P., A.C., H.S., L.S.Q., K.C.H., and C.A.G.; Funding Acquisition, J.M.P., J.S.W., E.D.B., L.S.Q., K.C.H., and C.A.G.; Supervision, J.M.P., J.S.W., E.D.B., L.S.Q., K.C.H. and C.A.G.

Chapter 3 is reprinted as a manuscript in preparation:

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*. https://doi.org/10.1101/805333.

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.

Exploring the functional constraints on bacterial essential gene expression with CRISPRi

John Sumblin Hawkins

Abstract

Essential genes are critical for cell viability. In bacterial cells these genes are unusually connected across their genetic network, frequently direct targets of antibiotics, and disproportionately expressed to the transcriptome and proteome. Despite their centrality to bacterial biology, essential genes have been poorly studied in vivo for want of the facile and high-throughput methods needed to systematically compromise activity essential genes while still maintaining the population levels necessary for measurements and assays of their functional impact. We applied CRISPR interference (CRISPRi) to create knockdowns of every essential gene in Bacillus subtilis, then assayed those knockdowns using chemical genomics, high-throughput microscopy, and growth experiments, which collectively provided significant insights into the organization and functional relevance of the bacterial essentialome. With this work as a springboard we developed a modified CRISPRi system leveraging the predictable reduction in efficacy of imperfectly matched sgRNAs to generate specific levels of CRISPRi activity and demonstrate its broad applicability in bacteria. Using libraries of such mismatched sgRNAs, we characterized the expression fitness relationships of essential genes in Escherichia coli and Bacillus subtilis. These organisms, though separated by ~2 billion years of evolution, conserve not only the essential genes themselves but in the majority of cases the relationships connecting their individual expression to overall fitness, suggesting that the tradeoffs underlying bacterial homeostasis are deeply fundamental.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

As recently as fifteen years ago few people would have recognized the acronym CRISPR, even within the field of Biology. Francisco Mojica(Mojica et al. 2005), Christine Pourcel(Pourcel, Salvignol, and Vergnaud 2005), and Alexander Bolotin(Bolotin et al. 2005), working independently, realized in 2005 that the pattern of **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (CRISPR) that they had been studying in bacterial genomes was connected to phage genomes, and correctly hypothesized that this might be a keystone of adaptive immune systems in bacteria. A year later, this hypothesis was elaborated by Eugene Koonin(Makarova et al. 2006). The chain of events which followed call to mind the rolling boulder chasing Indiana Jones through the prologue in "Raiders of the Lost Ark".

In an effort to arm themselves against cripplingly expensive outbreaks of phage infection in yoghurt production vats, researchers at Danisco France tested and confirmed this adaptive immune system hypothesis in 2007(Barrangou et al. 2007). They additionally identified the now-famous **C**RISPR-**as**sociated protein Cas9 as sufficient, in partnership with the CRISPR spacers, to effect genetic interference(Sapranauskas et al. 2011). Marraffini and Sontheimer(Marraffini and Sontheimer 2008) demonstrated the following year that, unlike RNAi silencing mechanisms, these effectors were targeting DNA. By 2012, researchers on both coasts of the United States and in Europe had taken the powerful promise of this simple system from abstract concept to practical laboratory tool(Barrangou 2012; Jinek et al.

2012; Zhang 2012). The web of use cases grew explosively from there, including functional genomics via genome-scale knockout screens in human cells(Tim Wang et al. 2014; Zhou et al. 2014), improved construction of mouse cancer models(Choi and Meyerson 2014; Chen et al. 2014; Torres et al. 2014), correction of disease-causing genetic mutations(Wu et al. 2013; Schwank et al. 2013), and acceleration of research on agricultural organisms such as rice and wheat(Jiang et al. 2013).

Cutting DNA at precisely controlled locations to inactivate or alter a gene remains an exciting and central part of the tree of tools rising from this fertile seed. But that second phrase, "precisely targeted locations", has deep implications. CRISPR/Cas affords the ability to target virtually *any* protein effector of interest to those same precisely controlled DNA locations. Jinek et. al.(Jinek et al. 2012; Gasiunas et al. 2012) and Gasiunas et. al.(Jinek et al. 2012; Gasiunas et al. 2012) showed that Cas9 could be used to cut DNA *in vitro*, and that this capability was tied to two domains -- a RuvC domain and an HNH domain, with each of the two cutting one strand of the target locus. However, these domains could be mutationally inactivated *without impeding Cas9's RNA-targeted binding of DNA*. Simply substitute in the catalytically inactive protein dCas9, directly or indirectly tethered to a transcriptional regulator, fluorophore, or any other effector domain, and "program" the partnered 20-base guide to point to a DNA target of interest. Your effector and DNA target will efficiently and robustly be held in close proximity by the dCas9.

In 2013, Stanley Qi demonstrated(Qi et al. 2013) the elegant potential of this idea's simplest possible implementation by fusing dCas9 to: absolutely nothing. Even though catalytically inactivated dCas9 no longer "does" anything beyond finding its target, the

tenacity with which it holds on to said target is sufficient to disrupt transcriptional expression of RNA molecules from the targeted locus. Dubbed "CRISPR interference", or simply "CRISPRi", this idea has since been successfully applied in both eukaryotes(Larson et al. 2013; Gilbert et al. 2013) and prokaryotes(Peters et al. 2016, 2019). In eukaryotes CRISPRi is a bit more complicated, generally depending on a fused effector domain such as KRAB to exact the desired degree of control over expression levels. The simplicity of CRISPRi is therefore especially attractive and exciting in the domain from which CRISPR was originally drawn: Bacteria.

When I joined UCSF as a graduate student, I was very excited to study the genetic networks underpinning the rich and complex life of bacterial cells. I was also, having been fortunate enough to spend the prior year volunteering with Jonathan Weissman's lab, deeply impressed by the power of genome-scale techniques like ribosome profiling(Ingolia et al. 2012) and NET-seq(Churchman and Weissman 2012). Opportunity to pursue both aims arose from a budding collaboration between Stanley Qi and Carol Gross. I joined both labs, and began working to develop CRISPRi technologies for high-throughput functional genomics in bacteria. When Stanley later left for Stanford, I remained at UCSF with the Gross lab. But Stanley's influence had been impactful, and this project had legs.

From the beginning of my thesis work I was excited about the power these new techniques could offer researchers exploring bacterial biology, yet frustrated to think how out-of-reach these new techniques might seem for labs *not* sitting scant dozens of feet from UCSF's Center for Advanced Technology, nor from the aforementioned labs at the vanguard of CRISPRi development. I wanted to expand the CRISPRi toolset, with a

particular eye towards the democratization of CRISPRi technology. I want other labs to benefit from powerful CRISPRi tools without expensive overhead and multi-year learning curves.

I started by developing a software pipeline to identify usable targets for sgRNA design. My publicly available suite of Python scripts automates the process of identifying all possible guide targets in a genome, and estimates the dangers of any off-target matches for said targets. It also annotates guides with important information about their position relative to interesting genomic features, particularly any overlap with the coding sequence of specific genes.

We soon found an important use for this technology. Complete (i.e. uncomplemented) removal of essential genes makes further study of an organism largely impossible -- it no longer grows(Koo et al. 2017). This feature made it very difficult to probe the precise functional relevance of essential genes in bacteria, and existing technologies such as introducing degradation tags(Cameron and Collins 2014) or targeting them with sRNAs(Storz, Vogel, and Wassarman 2011) are both labor intensive and imprecise. Yet essential genes -- often important targets of antibiotics and other chemicals, heavily enriched for genetic interactions with other genes, and certain to directly impact fitness in the neutral laboratory setting -- are critically important for understanding the overall genetic networks of bacteria. We used the guides designed by this pipeline to knock down essential genes in *B. subtilis* in order to examine their functions and network topology. (Chapter 1 -- Peters, et. al.)

Recently, our collaborators suggested(Gilbert et al. 2014) that insights gleaned from clarifying the *off-target* effects of an sgRNA/Cas9 pair could be used to modulate guide

efficacy. I ran with this idea. By introducing *mismatches into the guide* relative to the primary target, I might be able to reduce the frequency of *on-target* events, thereby creating guides that would impart some reduced level of knockdown in the presence of saturating dCas9 expression. Building a library of such mismatched guides could then provide a range of knockdowns for any targeted gene. It struck me that such a range of knockdowns presented a major opportunity. Having a range of phenotypes for each gene would make it much more likely that a given experiment would capture the phenotypic "sweet spot" in which a gene was depleted enough to impair fitness, but not so depleted as to exceed our limits of fitness detection or to occlude the interaction of that impairment with other sources of stress. This would hold true for any gene rendered contextually critical by a secondary stress such as a chemical antagonist or missing nutrient, but it was especially true for constitutively essential genes. Studies in yeast(Li et al. 2011) indicated that essential genes were interaction hubs for networks, highlighting the importance of including them in our analyses.

I realized that building and evaluating the mismatched guide library on the essential genes of *B. subtilis* was a major project on its own, and this became the cornerstone of my thesis. This work, reported in Chapter 2, expanded to include an analysis of a comprehensive knockdown library of gfp, as well as an analysis of a mismatched guide library targeted to the *E. coli* essential genes. We were able to develop a predictive model for mismatch knockdown efficacy from the GFP library and used that to characterize knockdown – fitness relationships of each essential gene in *E. coli* and *B. subtilis*. Additional aspects of the work are reported in several appendices. As the lab had never before performed large scale growth experiments, I validated the growth

defects of mismatched guides inferred from pooled growth with individual growth experiments (Appendix 1). We took a great deal of care in considering the diversity of representation within a pooled experiment, evaluating the importance of various bottlenecks and other sources of non-uniformity and noise in the constructed libraries of strains as well as the implications regarding the reproducibility and resolution of the fitness measurements for those strains (Appendix 2). I also used the library to explore how variably knocked down *dfrA* responds to low doses of Trimethoprim, to explore the idea that this method might refine our quantitative understanding of drug-target interactions. (Appendix 3). I developed small libraries of 10 guides per gene, to demonstrate the practical versatility of this approach (Appendix 4). Our original libraries included “double-mismatch” guides -- guides with a second mismatch. Single mismatch guides prove sufficient to our ends, but we show that our predictive model largely extends to the double-mismatch guides. (Appendix 5). In addition to the sgRNA design code discussed above (Appendix 6), I have provided software tools for the design of mismatch guide libraries and the analysis of pooled growth experiments on strains containing those guides (Appendix 7).

Before this thesis project, CRISPRi was a new idea, readily accessible for only the most fortunate of bacterial research labs. It is my sincere hope that the resources and analyses described in this thesis will allow other investigators to easily adapt these methods for their organism of choice, as well as to use the libraries described herein to examine conditions beyond the narrow set explored by this work.

Chapter 2

A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in Bacteria

SUMMARY

Essential gene products underpin the core reactions required for cell viability, but their contributions and relationships are poorly studied *in vivo*. Using CRISPR interference, we created knockdowns of every essential gene in *Bacillus subtilis* and probed their phenotypes. Our high-confidence essential gene network, established using chemical genomics, showed extensive interconnections among distantly related processes and identified modes of action for uncharacterized antibiotics. Importantly, mild knockdown of essential gene functions significantly reduced stationary phase survival without affecting maximal growth rate, suggesting that essential protein levels are set to maximize outgrowth from stationary phase. Finally, high-throughput microscopy indicated that cell morphology is relatively insensitive to mild knockdown but profoundly affected by depletion of gene function, revealing intimate connections between cell growth and shape. Our results provide a framework for systemati c investigation of essential gene functions *in vivo* that is broadly applicable to diverse microorganisms and amenable to comparative analysis.

INTRODUCTION

Essential gene functions underpin core cellular processes. Interrogating the relationships among essential gene functions is critical for understanding how bacterial growth is controlled and for facilitating drug development. Yet, few approaches can assess essential gene function *in vivo* to elucidate their connections. Neither gene-deletion libraries (Baba et al. 2006; Winzeler et al. 1999) nor saturating transposon mutagenesis (Goodman et al. 2009; van Opijnen, Bodi, and Camilli 2009) can be used to study essential genes, as cells cannot survive without their functions (Christen et al. 2011). Several high-throughput approaches have been used to identify or perturb essential genes in eukaryotes, including destabilizing the 3' UTR of mRNAs (DaMP alleles)(Breslow et al. 2008), CRISPR (*c*lustered *r*egularly *i*nterspaced *s*hort *p*alindromic *r*epeats)/Cas9 gene editing (Blomen et al. 2015; Tim Wang et al. 2015), and CRISPR/dCas9 transcriptional regulation technologies (Gilbert et al. 2014). The only study screening essential genes in bacteria used antisense RNA knockdowns to screen for antibiotic sensitivities (H. H. Xu et al. 2010), a method of limited utility due to variable efficacy (Forsyth et al. 2002).

Here we establish a CRISPR interference (CRISPRi) framework for systematic phenotypic analysis of essential genes in bacteria. CRISPRi uses a nuclease-deactivated variant of *Streptococcus pyogenes* Cas9 (dCas9) paired with a single guide RNA (sgRNA) to sterically hinder transcription at the sgRNA-base pairing genomic locus (Qi et al. 2013) and is a specific and efficient approach for knockdown, with demonstrated applicability in bacteria. We generated a comprehensive essential gene-knockdown library in the Gram-positive model bacterium *Bacillus subtilis* and used the library to enable drug target discovery, establish a functional network of essential gene processes, characterize how cell morphology and growth rate respond to reductions in essential gene expression, and dissect essentiality in a highly redundant genetic pathway. Our study provides a framework for comprehensive, high-throughput analysis of essential gene functions applicable to diverse bacteria.

RESULTS

CRISPRi is effective, specific, and titratable in B. subtilis

We established a CRISPRi system in *B. subtilis,* consisting of *S. pyogenes dcas9* driven by a xylose-inducible promoter (P_{xyl}) and sgRNAs expressed from a strong, constitutive promoter (P_{veg}), both transferred to the chromosome via integrating plasmids (**Figure 2.1 A**). This system is very efficient, exhibiting 3-fold repression of red fluorescent protein (RFP) without induction and 150-fold repression with full *dcas9* induction (**Figure 2.1 B, 2.8 A**). Although we also report a system with no basal repression based on a weak IPTG-inducible promoter (**Figure 2.8 B**), we used P_{xyl} exclusively throughout to consistently sensitize the library with slight knockdown and avoid inconsistencies resulting from slight variations in IPTG induction.

Our P_{xyl}-based CRISPRi system is titratable, with unimodal *rfp* repression at the single-cell level across sub-saturating inducer concentrations (**Figure 2.1 B, 2.8 C**), and high specificity, repressing only *rfp* expression at saturating inducer concentrations (**Figure 2.1 C**). We used NET-seq to identify the genomic positions of transcribing RNA polymerase (Larson et al. 2014), showing that CRISPRi sterically blocks transcription in *B. subtilis* (**Figure 2.1 D, 2.8 D**), as in *Escherichia coli* (Qi et al. 2013). CRISPRi is polar (Peters et al. 2015), with all downstream genes in an operon showing equivalent knockdown (**Figure 2.1 E**), and also reduces expression of upstream genes in the operon (**Figure 2.8 E**). Thus, CRISPRi technology is suitable for examining gene function at the operon level.

A CRISPRi knockdown library of essential genes

We constructed an arrayed library of *B. subtilis* strains expressing computationally optimized sgRNAs (Extended Experimental Procedures) targeting the 289 known or proposed essential genes (**Table 2.1**). The sgRNAs targeted unique DNA sequences at the 5´ ends of genes, where CRISPRi is most effective (Qi et al. 2013). Nearly all sgRNAs (~94%) targeting bona fide essential genes (258 genes total, Extended Experimental Procedures) (Koo et al. 2017) decreased colony size on agar plates with xylose (≥25% reduction in area compared to the

control; **Table 2.1**; Extended Experimental Procedures). Control cells expressing only *dcas9*, or *dcas9* and an sgRNA targeting an innocuous gene (*rfp*), had no growth defects (**Figure 2.8 F**). We conclude that a single sgRNA is sufficient for effective knockdown of essential genes, simplifying CRISPRi library design in bacteria.

CRISPRi-based essential gene phenotyping and drug target discovery

The ~3-fold repression of our knockdown library without induction (basal repression; **Figure 2.1 B, 2.8 A**) sensitized strains to various chemicals. This enabled us to define essential gene phenotypes via chemical-genomic analysis by measuring colony size against 35 unique compounds (Extended Experimental Procedures). We achieved high reproducibility, as measured by correlated colony sizes (R = 0.89, **Figure 2.9 A**). We converted colony sizes to chemical-gene scores (**Figure 2.9 A-D; Table 2.2**) (Nichols et al. 2011), and identified significant chemical-gene phenotypes (false discovery rate ≤5%) (Nichols et al. 2011). Most knockdowns of antibiotic targets were hypersensitized to their cognate drug (e.g. *dfrA*/folate biosynthesis to trimethoprim (Myoda et al. 1984), *fabF*/fatty acid metabolism to cerulenin (Moche et al. 1999); **Table 2.2**). Although *fabI*, the target of triclosan (Schujman et al. 2001), is not essential due to a gene duplication (Thomaides et al. 2007), another knockdown in the pathway (*fabG*) was sensitized (**Table 2.2**). We conclude that our CRISPRi platform effectively identifies known drug-gene interactions.

We tested our essential knockdown library as a platform for drug target discovery by screening against MAC-0170636, an antibiotic that upregulates the cell wall-damage responsive promoter P_{ywaC} (Czarny et al. 2014) by an unknown mechanism. Undecaprenyl pyrophosphate synthetase (*uppS*) was the most sensitized knockdown (**Figure 2.2 A; Table 2.2**), and we confirmed its sensitivity in liquid (**Figure 2.2 B**). Conversely, *uppS* overexpression increased MAC-0170636 resistance relative to wild-type (WT) cells (**Figure 2.2 B**). Purified *B. subtilis* UppS activity was inhibited by MAC-0170636 with an IC_{50} of 0.79 μM (**Figure 2.2 C**), indicating that UppS is the

direct target of MAC-0170636. UppS purified from another Firmicute, *Staphylococcus aureus*, was completely resistant to MAC-017063 (**Figure 2.9 E**), as was *S. aureus* itself. Likewise, *B. subtilis* expressing only *S. aureus uppS* was resistant (**Figure 2.2 D**). These results highlight the utility of our knockdown library for identifying direct targets of uncharacterized compounds; CRISPRi portability suggests its utility in future organism-specific drug-discovery efforts.

Functional analysis of the essential gene network

Highly correlated responses of gene knockdowns across chemical conditions (phenotypic signatures) indicate functional connections (Nichols et al. 2011). We established a network of gene-gene connections using statistically significant correlations among the phenotypic signatures of our essential gene knockdowns, based on both direct or indirect effects of drug-gene interactions (**Figure 2.3 A, 2.9 B-D, 2.10; Table 2.3**; Extended Experimental Procedures). The network was rich in known biological connections among genes in related processes such as cell-wall biosynthesis and cell division (**Figure 2.3 B**).

We quantitatively validated the network with a ROC curve. Because CRISPRi exhibits polarity, this analysis was based on operons containing essential genes ("essential operons", n = 203), rather than on the individual genes themselves (Extended Experimental Procedures, **Table 2.3**). We compared high-confidence gene-gene connections in the STRING database (http://string-db.org/) (Szklarczyk et al. 2015) with correlations between essential operons in our chemical-genomics dataset that met our correlation threshold (R=0.572; **Figure 2.3 C**; Extended Experimental Procedures). Our network showed excellent agreement with STRING, with a ~20-fold ratio of true to false positive rates (**Figure 2.3 C**, inset).

We investigated the relationship between network connections and gene function by assigning essential operons to functional groups using the SubtiWiki online resource (http://subtiwiki.uni-goettingen.de/) (Michna et al. 2014). Of the 61 non-overlapping functional groups, the 35 with two or more essential operons were assessed for connectivity within the group (intra-

connectivity). Mean intra-connectivity was ~2.4, compared with a background mean connectivity of ~0.3 ($p < 10^{-5}$), with 10 groups showing connectivity among ≥3 essential operons (**Table 2.3**).

Indeed, the ROC analysis showed strong specificity and sensitivity for recapitulating existing intra-process connections in STRING (**Figure 2.11 A**), with excellent recovery of connections within functional groups (**Figure 2.11 B**). Connections within the peptidoglycan (PG) cell wall biosynthesis (18) and DNA replication (9) functional groups were most dense, likely reflecting convergence of PG precursor biosynthesis and localized cell wall assembly (Turner, Vollmer, and Foster 2014) and replisome protein-protein interactions (Sanders, Dallmann, and McHenry 2010), respectively.

We next compared connectivity between functional groups (inter-connectivity) in our network with those in STRING, finding common connections and important distinctions (**Figure 2.10 D, 2.10 E**). First, the balance of intra- and inter-process connections differed; 46.2% of connections in STRING were inter-process, versus 59.0% in our network ($p < 10^{-4}$). Second, STRING inter-process connections were biased toward extensively studied processes, e.g. 39% of STRING but only 4.3% of our inter-process connections were between ribosomal proteins and translation factors. Moreover, 84% (113/134) of connections unique to our network were between processes (**Table 2.3**), highlighting the ability of our open-ended approach to detect such connections. Finally, our network revealed many connections between operons in distant functional groups. Using the hierarchical annotation levels from SubtiWiki (Michna et al. 2014) as an approximation of "annotation distance", we found that novel connections in our network were skewed toward processes furthest apart in annotation, whereas those in STRING were predominantly between related processes (**Figure 2.3 D**). Some distant connections were intuitive, such as DNA replication (*holB*) and folate biosynthesis (*folC* and the *sul-folB-folK* operon), likely reflecting the folate requirement for dTTP production (Hardy et al. 1987). Other connections were unexpected, such as those between peptidoglycan (PG) biosynthesis/cell

division and DNA replication/modification (e.g., *ftsL* and *dnaX*, *murC* and *gyrB*, *ddl-murF* and *ydiO*); these connections may be involved in failsafe mechanisms that link division and DNA replication (Arjes et al. 2014).

We explored the unexpected connection between transcription (*rpoB*) and cell division (*ftsL*), which is based on shared sensitivities to DNA intercalators and cell-wall antibiotics (**Figure 2.9 D**). Using RNAseq, we found that basal knockdown of *rpoB* reduced the *rpoB-rpoC* mRNA level by two-fold. Among the few other substantial expression changes (**Figure 2.11 D**), we found several down-regulated envelope genes including *manA* (4-fold), which is involved in cell-wall integrity (Elbaz and Ben-Yehuda 2010), and *sigW* (1.5-fold to 2-fold), a master regulator of envelope stress (Cao et al. 2002) (**Table 2.3**). Selective reduction of several envelope functions due to *rpoB* knockdown could result in cell-wall defects that mimic those caused by knockdown of late-acting cell-division genes.

We also identified novel connections to essential genes of unknown function (**Figure 2.3 A, 2.10, 2.11 E**). For example, resistance to cell wall-targeting antibiotics drove strong correlations among the largely uncharacterized gene *ylaN* (**Figure 2.9 F**) (Hunt et al. 2006; L. Xu et al. 2007), iron-sulfur cluster biogenesis, and isoprenoid biosynthesis, the latter of which depends on the iron-sulfur cluster enzyme IspH (Gräwert et al. 2004; Wolff et al. 2003). These results suggest that defects in cellular iron homeostasis underlie the connections to *ylaN*. Indeed, in follow up experiments, we determined that *ylaN* is non-essential with added iron(III) (**Figure 2.11 F**), further highlighting the ability of our unbiased approach to identify novel connections among essential processes.

Growth characteristics of the essential gene knockdown library

We measured apparent lag, maximum growth rate, and saturating density of growth curves for all basal knockdowns in LB medium (**Figure 2.4 A-B**; Extended Experimental Procedures). Almost all knockdowns (~80%) had a maximal growth rate equivalent to the control (**Figure 2.4**

B and inset); ribosomal proteins were the only functional category with slower growth rates relative to other knockdowns (p = 0.0008, t-test). However, most knockdowns (95%) had longer apparent lag times (**Figure 2.4 A**, inset), and cell wall-synthesis genes were enriched for the longest lags (p = 0.0013, Mann–Whitney U test, **Table 2.4**). CRISPRi repression was maintained in stationary phase (**Figure 2.11 G**).

We determined whether longer apparent lag times reflected strain-specific differences or decreased viability. Both the fraction of growing cells on agarose pads (Figure 4C) and plate viability measurements (**Figure 2.4 D**) of strains spanning the apparent lag times negatively correlated with apparent lag. Apparent lag times also tracked with expectations from the number of live cells present. For example, using the 45-min doubling time in batch culture (**2.4 B** inset), the ~8-fold reduction in *mraY* knockdown viable cells (**Figure 2.4 C-D**) requires ~3 additional divisions, accounting for the 2.1-h increase in apparent lag. Moreover, viable *mraY* knockdown cells and the control had equivalent growth rates after transfer to an agarose pad with fresh LB (**Figure 2.4 C**, inset**).** Together, these surprising results suggest that slight reductions in essential gene products affect outgrowth from stationary phase, creating a mixed population in which some cells exhibit WT outgrowth and others are either dead or non-growing.

Basal knockdown of essential genes results in changes in cellular dimensions that reveal shape actuators and modulators

Our library offers an unprecedented opportunity to examine whether basal knockdown of essential genes affects cell morphology. To provide baseline values, we quantified the dimensions of WT cells. Cells were imaged 3.5 h after different degrees of dilution into fresh medium (Extended Experimental Procedures), and therefore had different culture optical densities (OD). We found that median cell length and width varied systematically with extent of dilution (**Figure 2.5 A**), likely reflecting growth phase differences as cells cycle through nutrients in rich medium (LB). Cells diluted less were smaller, reflecting their slower growth rate at the time of imaging, similar to previous reports linking cell size and steady-state growth rate

(Schaechter, Maaloe, and Kjeldgaard 1958) and consistent with known shortening in stationary phase (Overkamp et al. 2015).

We next determined morphological parameters of the essential gene basal knockdowns to pinpoint proteins for which a small expression change results in altered shape, using high-throughput imaging at 3.5 h after dilution (Extended Experimental Methods). Because of variable stationary-phase outgrowth (**Figure 2.4 C-D**), culture OD varied widely at the time of imaging. We found a general relationship between growth rate and OD for virtually all knockdowns (**Figure 2.11 H**), indicating a large range of instantaneous growth rates at imaging. Nonetheless, the length and width of the knockdowns were highly correlated (**Figure 2.12 B**), similar to WT cells across dilutions (**Figure 2.12 A**). The fact that cells reach their maximum growth rate at a fixed OD (**Figure 2.11 H**), rather than at a fixed number of doublings after growth resumption, underscores the importance of cell density and the extracellular milieu in growth rate control.

Basal knockdowns that deviated from the length/width trendline (**Figure 2.5 B; Table 2.5**) could represent proteins involved in either actuating or regulating growth of the shape-determining cell wall. Only cell-envelope genes exhibited significant enrichment of outliers (average median length deviation = 0.99 μm, $p < 0.0005$ by bootstrapping, Extended Experimental Procedures; **Figure 2.5 B**), as expected because they are largely PG synthesis-related and are therefore actuators of cell shape. To validate our identification of outliers and ensure that cell chaining was rare, we used the membrane stain, FM4-64, on a subset of strains with different median cell lengths, finding that length measurements from phase images and peripheral fluorescence were highly correlated (**Figure 2.12 A**).

Several outliers in other functional groups were intriguing, as they could identify cell shape regulators. Basal knockdown of *tufA*, encoding the translation elongation factor Tu, resulted in cells that were substantially shorter than expected (**Figure 2.5 B, Table 2.5**); Tu interacts with the bacterial cytoskeleton protein and rod-shape determinant MreB (Defeu Soufo et al. 2010).

Knockdowns of several DNA replication genes (*dnaX*, *dnaA*, *nrdE/F*) resulted in longer cells (**Figure 2.5 B, Table 2.5**). Moreover, *dnaX* and *ftsL*, both of which are large mophological outliers, were also significantly correlated in our chemical screen, suggesting a functional connection between replication and division corroborated by multiple independent data types. As cell size is dependent on the degree of dilution prior to regrowth (**Figure 2.5 A**), direct comparisons of cellular dimensions among different strains require cells to be at the same OD. Large variation in lag times prevented us from measuring the entire library by this method. Instead, we examined two actin homologs (*mreB* and *mbl*) involved in coordinating PG synthesis (Scheffers and Pinho 2005) and the *mur* genes responsible for PG precursor synthesis at OD 0.3±0.05. Both actin homologs were wider than WT on average, with *mreB* exhibiting larger standard deviation, suggesting that cell width is particularly sensitive to MreB levels (**Figure 2.5 C**). The median cell lengths of *mur* strains were similar to WT (**Figure 2.5 D**), but widths (population median between 1.04-1.14 μm) were larger than no-sgRNA cells (1.02 μm), validating their classification as outliers (**Figure 2.5 D**). The *murB* cell width increased monotonically with *dcas9* induction (**Figure 2.5 E**), indicating that cell width is responsive to the cellular levels of PG precursors.

In summary, our systematic screen for the essential gene products most intimately tied to cell shape identified genes with known ties to morphology (e.g. *mreB*/*mbl*) and cell wall synthesis, revealed a quantitative relationship between PG precursor gene expression and cell width, and uncovered potential new regulators of cell morphology.

Single-cell characterization of terminal phenotypes provides a novel view of essential gene function

We examined whether substantial depletion results in more extreme morphological changes than basal knockdown by imaging the entire library after prolonged (~24 h) induction of *dcas9* (“terminal phenotypes”; Extended Experimental Procedures). Remarkably, >60% of the strains (174/289, or 166/258 bona-fide essentials) exhibited morphological phenotypes, even excluding

strains with growth defects only (over-division or growth halting; **Figure 2.6 Ai**), a number far exceeding the 48 known cell envelope-related strains. Knockdowns displayed several predominant morphologies (**Figure 2.6 Aii-viii, 2.12 B**), whereas the control formed a uniform lawn of normal, rod-shaped cells (**Figure 2.6 Aii**).

We quantified terminal morphologies by manually classifying phenotypes with an eight-dimensional phenotype vector including lysis, bulging, uniform shape loss, bending, filamentation, over-division (shorter cells), and extent of growth. We assessed whether functionally related genes mapped to particular phenotypes. Many cell-envelope genes bulged (e.g. *tagA*/*B*/*F*/*G*/*O* (teichoic acid), *mreC* (cell shape/PG biosynthesis); **Figure 2.6 Aiii**), while many ribosomal genes displayed severe cell bending (**Figure 2.6 Aiv**). However, in some cases genes in different processes had similar terminal phenotypes, such as growth halting (**Figure 2.6 Avi**) and bending (**Figure 2.6 Aiv**). Several metabolic genes (e.g. *metK*, *dxs*) exhibited bulging (**Table 2.5**), consistent with network connections between these processes and cell envelope synthesis/division (**Figure 2.3 A**). Finally, some strains displayed both bulging and lysis (**Table 2.5**), indicating that depletions in one process can lead to more than one terminal phenotype.

Phenotypic variability within functional groups and common phenotypes of different groups made it challenging to map phenotype to process with simple visual comparison. Instead, we evaluated the similarities between terminal phenotype vectors for all SubtiWiki annotation pairs (**Figure 2.6 B**, Extended Experimental Procedures), a process conceptually similar to the one we used to identify inter-process connections in the chemical-genomics dataset. Our similarity matrix recapitulated known connections (e.g. PG synthesis/cell division) while providing support for novel network connections (e.g. fatty acid metabolism with both isoprenoid biosynthesis and several central-dogma annotations; **Figure 2.6 B, 2.3 A, and 2.3 B**). Terminal-phenotype links are consistent with and complement the high level of inter-process connectivity of the essential gene network (**Figure 2.3 E**).

Variable terminal phenotypes within a group might reflect the rate of gene-product depletion. We tested this hypothesis by inhibiting tRNA charging (**Figure 2.6 C**) either by *valS* (ValS aminoacyl-tRNA synthetase) knockdown, or by addition of serine hydroxymate, ($tRNA^{ser}$ aminoacylation inhibitor). We observed dose-dependent phenotypes (**Figure 2.6 C**): shorter cells with slight depletion/inhibition (**Figure 2.12 C-D**), filamentation and bending with intermediate levels, and growth halting at high levels; these phenotypes were also exhibited by other aminoacyl-tRNA synthetase knockdowns (**Figure 2.6 Aiv, vi, viii**). Thus, the rate of inhibition of essential processes can qualitatively affect their terminal phenotypes.

In summary, the majority of essential genes knockdowns display an altered terminal morphological phenotype. Phenotypes group both within and across functional processes, demonstrating the utility of morphology for revealing interactions, and can vary with the kinetics and extent of gene knockdown.

Cellular behavior during the depletion of essential proteins

To determine how rapidly terminal phenotypes are established, we performed timelapse imaging on nine representative knockdown strains in the presence of inducer. We placed exponentially growing cells on agarose pads of LB+xylose to initiate depletion, and imaged every 5 min to determine microcolony growth rates (**Figure 2.6 D, 2.12 E**). During the first 60 min post-induction, WT cells maintained a constant elongation rate. The *mbl* knockdown maintained the WT growth rate, *murD* had a ~50% reduction, and the rest had a 20-30% reduction. Bulging (*tagD*) was the first prominent morphological phenotype observed (~40 min; 1-2 cell doublings on agarose pads) with filamentation (*pbpB*), bending (*accD*), and lysis (*murD*) occurring after ~60 min (2-3 doublings; **Figure 2.6 E**). As depletion is primarily by dilution, these data indicate that reduction of at least 50-75% below basal knockdown is required to substantially affect growth rate and morphology. Many factors prevent precise quantification, including differential protein stabilities and polar effects on transcription.

We examined growth of the entire library during *dcas9* induction by diluting stationary-phase cultures into liquid LB+xylose. Approximately 1/3 of the strains never emerged from stationary phase (OD_{600}<0.06 at 7 h; **Figure 2.13 A**). These strains covered the entire range of lags, hence lack of growth could not be explained by the variable stationary-phase outgrowth exhibited under basal knockdown conditions (**Figure 2.13 B**). Instead, these strains identify the subset of essential proteins present in limiting amounts such that additional depletion beyond basal knockdown results in growth cessation or lysis (**Table 2.4**).

Among strains that grew, we identified two distinct phenotypes (**Figure 2.13 C, Table 2.4**, Extended Experimental Procedures). Thirteen strains showed nearly linear growth; 9/13 affected cofactor biosynthesis (e.g. heme biosynthesis) or electron transport (**Figure 2.13 D**), and may reflect cofactor availability. Fourteen strains enriched in either cell-envelope synthesis (5/14) or DNA replication (7/14) (**Table 2.4**) stalled after initial growth and then exhibited a marked decrease in OD, indicative of lysis (**Figure 2.13 E**). This common phenotype may reflect the close network and morphological connections observed between DNA replication and envelope synthesis (**Figure 2.3 A**).

Dissecting a highly redundant gene network with multiplexed CRISPRi

Genetic redundancy, prevalent in complex processes (e.g. construction of PG) (Meeske et al. 2015) can mask the contributions of individual genes to essential processes. CRISPRi can simultaneously knockdown two genes (Qi et al. 2013); we show that CRISPRi can simultaneously knockdown eight unrelated, nonessential genes effectively (**Figure 2.7 A and 2.7 B**). Using this multiplexing capability, we examined redundancy for the 16 penicillin binding protein (PBP)-encoding genes involved in PG synthesis. Simultaneous knockdown of pairwise *pbp* genes were viable at full *dcas9* induction except *pbpA*/*pbpH*, a known synthetic lethal pair, and those including *pbpB*, the only essential PBP (**Figure 2.7 C, 2.14 A-C; Table 2.6**) (Wei et al. 2003). The observed fitness of double knockdowns was largely predicted by multiplying the

fitness of single mutants, a formalism developed for null mutations in parallel pathways (**Figure 2.14 A**, Extended Experimental Procedures).

We identified double knockdowns that were hypersensitive to PBP inhibitors (mecillinam, cefoxitin, and aztreonam; **Figure 2.14 D-E**), as these suggest possible triple synthetic gene combinations. Fitness of the *pbpA*/*ponA*/*pbpD* triple knockdown was lower than predicted (**Figure 2.7 D**) and we could not construct a Δ*pbpA*/Δ*pon*Δ/Δ*pbpD* triple deletion, although we could introduce an innocuous deletion (Δ*trpC*) or an unrelated *pbp* deletion (Δ*pbpC*) into Δ*pbpA*/Δ*pon*A. Thus, the triple deletion is either extremely sick or synthetic lethal.

PBP4 (*pbpD*), PBP2a (*pbpA*), and PBP1a/b (*ponA*) have transpeptidase activity and are septally localized (Scheffers, Jones, and Errington 2004). Reduced septal PG transpeptidation during cell division may underlie the severe growth defect of the triple knockdown. Control and single- and double-knockdown cells exhibited the expected rod-like shape, albeit with reduced cell length (**Figure 2.7 E-F**), potentially implicating these three genes in the regulation of cell division. By contrast, the triple mutant had distinct phenotypes including filamentation, cell lysis (**Figure 2.7 E**, white arrows), and bending along the cell contour (**Figure 2.7 G**, $p<10^{-6}$), whether constructed with complete knockdown of all three genes or *ponA* knockdown in the Δ*pbpA*/Δ*pbpD* double. This phenotype is consistent with lethality caused by reduced septal transpeptidation, and showcases the ease of dissecting redundant pathways with CRISPRi.

DISCUSSION

Bacteria typically have several hundred essential genes that encode the core reactions central to viability, together constituting ~10% of their total genetic complement. Lacking a facile way to reduce their expression, we had little understanding of *in vivo* relationships among essential gene processes, or how subtle imbalances in essential pathways impact cellular homeostasis. This work is a major advance in the study of bacterial essential genes, providing a systematic, unbiased study of their phenotypes *in vivo* using CRISPRi (**Figure 2.1**) to obtain facile and precise down-regulation. Using chemical genomics profiling and high-throughput microscopy, we identified complex phenotypes including chemical vulnerabilities (**Figure 2.3**), growth and shape phenotypes (**Figure 2.4, 2.5**), and terminal death phenotypes (**Figure 2.6**). Together, these revealed a complex web of connections among essential processes. Given that CRISPR systems are broadly active in bacteria, our approach can be readily extended to other bacterial species including pathogenic and non-culturable species.

Our essential gene network (**Figure 2.3**) reveals numerous inter-process connections not previously annotated, and is highly enriched in novel connections among distant processes. Process inter-connectivity may provide the cell with mechanisms for restoring cellular homeostasis in response to transient imbalances. Some connections make intuitive sense. For example, folate is likely linked to replication via its necessity for dTTP synthesis. Such correlations readily suggest specific hypotheses that can be directly tested experimentally; e.g. a particular DNA polymerase subunit senses dTTP levels. Other connections have no facile explanation, and underscore the fact that connections between core processes are abundant and understudied. Very likely, our network significantly underestimates connections as it is based only on the highest confidence interactions to avoid false positives. Comparable datasets in other organisms will provide the basis for evolutionary studies to investigate the extent to which the logic of these functional circuits is conserved across organisms.

Our studies suggest that the levels of essential *B. subtilis* proteins are higher than necessary to maintain optimal growth, as the vast majority of basal knockdown strains grew indistinguishably from WT in exponential phase (**Figure 2.4 B**). Thus, either protein levels are set high enough to be robust to small (1.5-3-fold) decreases in expression or their levels are maintained by an elaborate posttranscriptional regulatory system. Moreover, there is a sufficient excess of essential proteins that most strains (70%) emerge from stationary phase even when their gene products are depleted during regrowth. The 30% of strains unable to exit stationary phase (**Figure 2.13 A**) are those whose protein levels are closest to the levels necessary for normal growth.

In stark contrast, almost all basal knockdowns strains exhibited increased cell inviability during the exit from stationary phase. Thus, even a small decrease in protein product increases the vulnerability of cell regrowth (**Figure 2.4 A**). Importantly, there is little overlap between strains ceasing growth soon after depletion and those most vulnerable to inhibition of outgrowth from stationary phase (**Figure 2.13 B**). This discordance suggests that the set of essential genes whose expression level is close to that necessary for rapid cell growth in ideal conditions is distinct from the set necessary for survival during stationary phase. Speculatively, protein levels may reflect the ecological niche of *B. subtilis* in soil, where cells spend much time in a non-growing state and must survive long enough to enter the sporulation program with high efficiency.

Classically, genetic perturbations resulting in morphological variation, such as homeotic mutants and variegated maize, provided critical insights into fundamental cellular circuits. However, the essentiality or redundancy of most cell wall-synthesis proteins has made it difficult to uncover the molecular mechanisms underlying bacterial cell shape and size determination. Our CRISPRi knockdown approach allowed us to probe this relationship both under conditions of partial knockdown (1.5 to 3-fold) and during complete depletion. Partial knockdown identifies outliers exceptionally sensitive to depletion. As a consequence, we identified the critical envelope gene

actuators of the response, showed that cell width as well as length is controlled (**Figure 2.5**), and determined that cell width varies montonically with extent of depletion of *mur* genes (**Figure 2.5 D-E**). The few outliers involved in other processes identify putative regulators of cell wall synthesis. The translation and DNA replication-related outliers may tie the rate of protein synthesis and DNA replication to cell-wall growth, possibly using previously unrecognized moonlighting functions of these proteins. Moreover, since maximal growth rate is not substantially affected across the basal knockdown strains, our findings indicate that cell size and growth rate can be at least partially decoupled. Additionally, a graded morphological response to the levels of essential proteins may drive physiological heterogeneity enabling population adaptation to dynamic environments.

In contrast, complete depletion probes the intimate relationship between growth and morphology. Here, cells exhibit a wide array of terminal phenotypes with alterations in both growth and morphology (**Figure 2.6 A**). By comparison, only a single nonessential *E. coli* gene deletion (*rodZ*) has a morphological phenotype (Shiomi, Sakai, and Niki 2008). Some phenotypes result from gradual depletion of function (**Figure 2.6 C**), which may mimic imbalances that occur transiently due to stochasticity in gene expression. Tracing changes at different rates of depletion may provide clues to the cellular program for morphology. Moreover, the dynamics of removing an essential function may be as important to growth and viability as the presence or absence of such a protein.

The evolution of essential processes has remained largely mysterious. Long-term evolution experiments have detected many mutational events in the RNA polymerase complex, cell-wall synthesis, and cell-shape determination (Tenaillon et al. 2012), suggesting a great diversity of molecular adaptations, including to essential processes. The apparent excess abundance of most essential proteins (**Figure 2.6 D-E**) suggests that new functions could rapidly develop from the existing repertoire of proteins without compromising growth, possibly explaining the prevalence of moonlighting proteins (Huberts and van der Klei 2010). For example, actin and

tubulin homologs have diverse roles in bacteria and eukaryotes (Busiek and Margolin 2015). Phenotypic heterogeneity and variable survival during stationary phase, as observed here (**Figure 2.4**), may indicate that stressful environments have shaped the evolutionary history of essential genes. The complexity of the network (**Figure 2.3 A**) suggests context dependence that may be critical for the evolution of essential genes and their interactions; perhaps certain essential genes can be rendered non-essential in environments such as biofilms. Expansion of our study to varied environments may provide a more nuanced view of the instances when a particular process is limiting for growth.

EXPERIMENTAL PROCEDURES

CRISPRi Library Design, Cloning, Chemical Screens, and Growth Analysis

sgRNAs targeted all putative essential genes (subtiwiki.uni-goettingen.de) and those recently identified in our *B. subtilis* gene knockout library (Koo et al. 2017). sgRNAs were designed to target within the gene body near the 5´ end of the gene on the non-template strand. sgRNA libraries were cloned via inverse PCR as previously described (Larson et al. 2014) and strains were constructed using natural competence transformation. Chemical screening was performed and chemical-gene scores were calculated as previously described (Nichols et al. 2011). *B. subtilis* and *S. aureus* UppS proteins were purified using nickel-affinity chromatography, and assayed using a Kinetic EnzCheck pyrophosphate assay (Life Technologies). The *B. subtilis* essential gene network was constructed by calculating all pairwise Pearson correlations between sgRNA knockdown strains, randomly permuting gene identity relative to chemical-gene scores to generate a background distribution, then applying a significance cutoff to the correlations based on the 95% confidence interval of the background distribution. Population growth curves were obtained from a microplate reader, and growth information was extracted by fitting the curves to a Gompertz equation (Zwietering et al. 1990).

High-throughput Microscopy

Images of single cells were acquired after transferring cells from a 96-well plate onto a large-format agarose pad. Analysis of cellular morphologies was performed using custom MATLAB code. A minimum of 100 cells were analyzed for quantitative descriptions of single-cell phenotypes. Terminal phenotypes were examined by spotting cultures outgrown from stationary phase for two hours in liquid LB medium, transferred to agarose pads containing LB + 1% xylose, then imaged after overnight incubation. Time-lapse images were taken in an active-control environmental chamber at 37 °C (HaisonTech). Further details of methods are in the Extended Experimental Procedures.

FIGURES

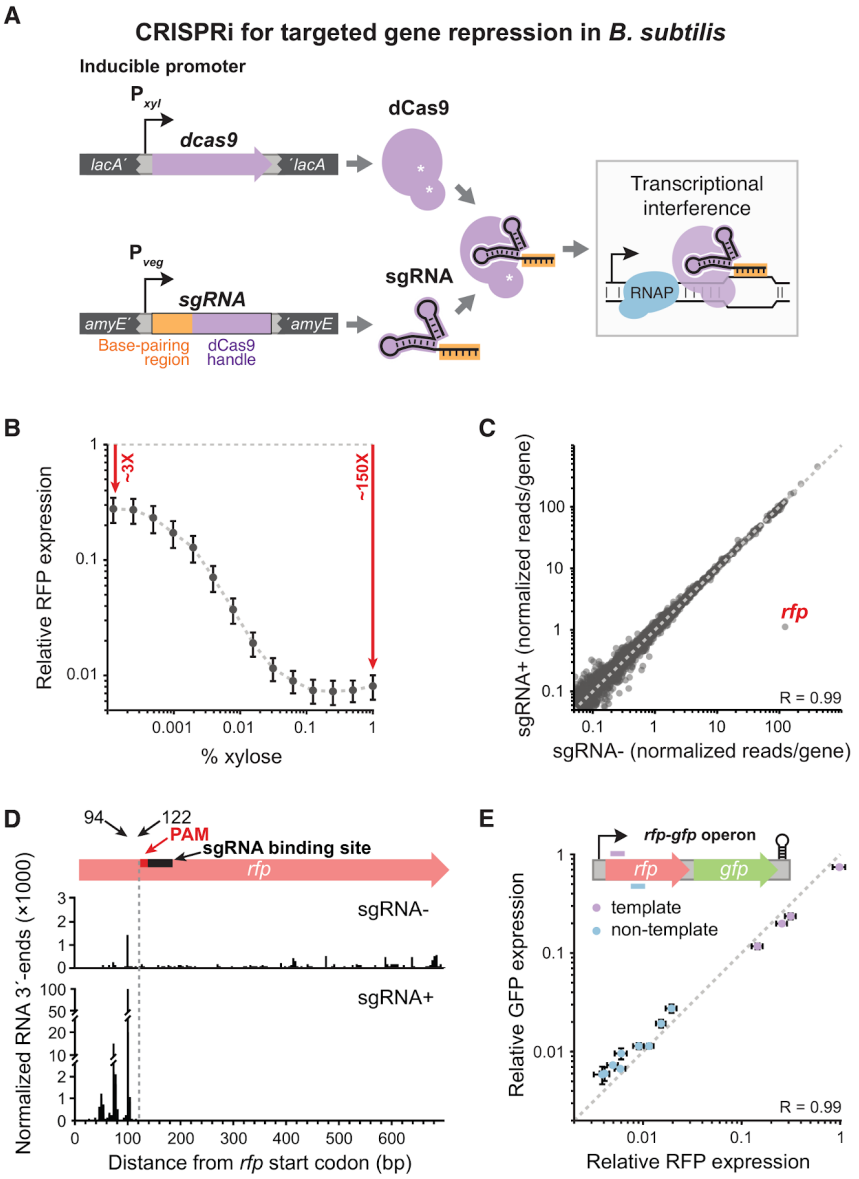


Figure 2.1. *B. subtilis* CRISPRi is Efficient, Titratable, and Specific.
A) *B. subtilis* xylose-inducible dCas9 is directed to specific DNA targets by constitutively expressed sgRNAs, where it represses transcription. *dcas9* was stably integrated into the *lacA* locus, and sgRNAs into *amyE* or *thrC*.
B) Flow cytometry of cells that constitutively express *rfp* and an *rfp*-targeted sgRNA, in which *dcas9* was induced by adding the specified concentration of xylose. Median RFP levels are relative to the no-sgRNA control (grey dashed line), and error bars are ±1 standard deviation.
C) RNA-seq of cells maximally induced for *dcas9* (1% xylose) and constitutively expressing *rfp* and an *rfp*-targeted sgRNA versus cells without an sgRNA. Reads per gene were normalized by reads per kilobase per million reads. The dashed line is $y = x$.
D) NET-seq of cells maximally induced for *dcas9* (1% xylose) and constitutively expressing *rfp* and an *rfp*-targeted sgRNA versus cells without an sgRNA. The *y*-axis is broken to

accommodate the wide range of reads. The dashed line corresponds to the upstream boundary of the PAM sequence. We suggest that RNA 3´ end peaks upstream of the dCas9 block result from RNA polymerase queuing.
E) Flow cytometry of cells maximally induced for *dcas9* (1% xylose) and constitutively expressing an *rfp-gfp* operon and various *rfp*-targeting sgRNAs. RFP and GFP levels are relative to the no-sgRNA control. The dashed line is $y = x$. Note that in this group of sgRNAs, only template strand-targeting sgRNAs exhibited low efficacy (violet).

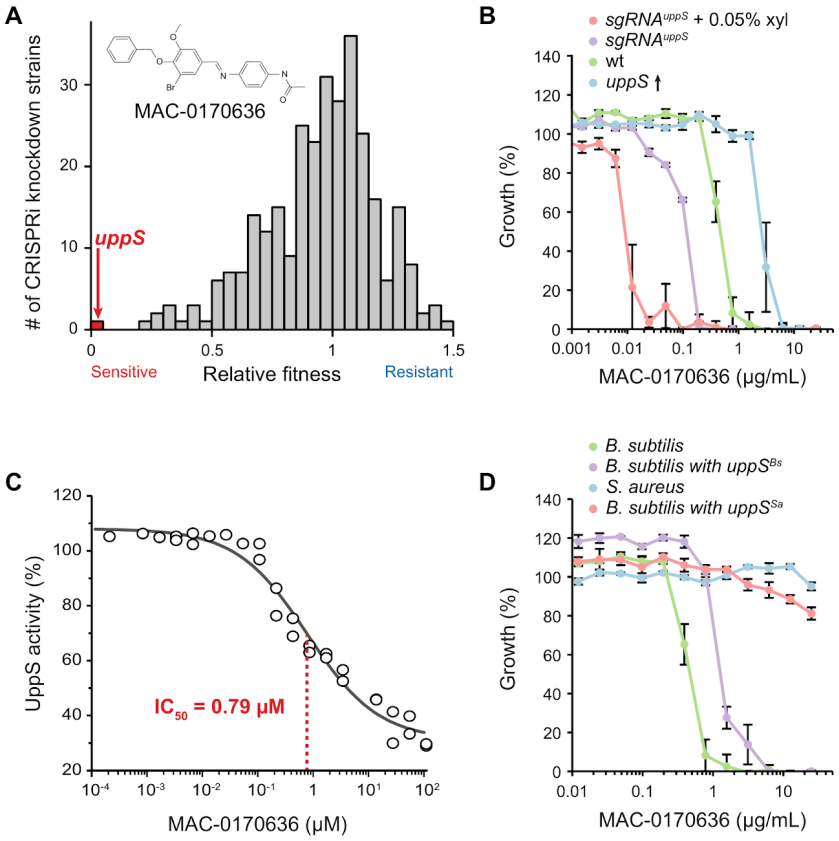


Figure 2.2. CRISPRi Knockdowns of Essential Genes Enable Discovery of Direct Antibiotic Targets.

A) Relative fitness of CRISPRi essential gene knockdown strains (*n* = 289) with basal *dcas9* expression (no xylose induction) grown on plates containing MAC-0170636, as determined by the ratio of normalized colony sizes on LB plates + DMSO versus LB + MAC-0170636.

B) Minimal inhibitory concentration (MIC) assay for strains over- or under-expressing *uppS* grown in liquid medium containing MAC-0170636. The MICs of these strains were (in μg/ml): *sgRNA*^{uppS} + 0.05% xyl (~0.012), *sgRNA*^{uppS} (~0.195), WT (~0.78), and *uppS* overexpression (~3.125). The values plotted are the means of at least three measurements, and error bars are ±1 standard deviation.

C) Concentration-dependent inhibition of purified *B. subtilis* UppS by MAC-0170636. Each point is an individual measurement.

D) MIC assay for strains expressing *B. subtilis uppS* (*BsΔuppS/amyE*::P$_{spank}$-*uppS*^{Bs}) or *S. aureus uppS* (*BsΔuppS/amyE*::P$_{spank}$-*uppS*^{Sa}) grown in liquid LB + MAC-0170636.

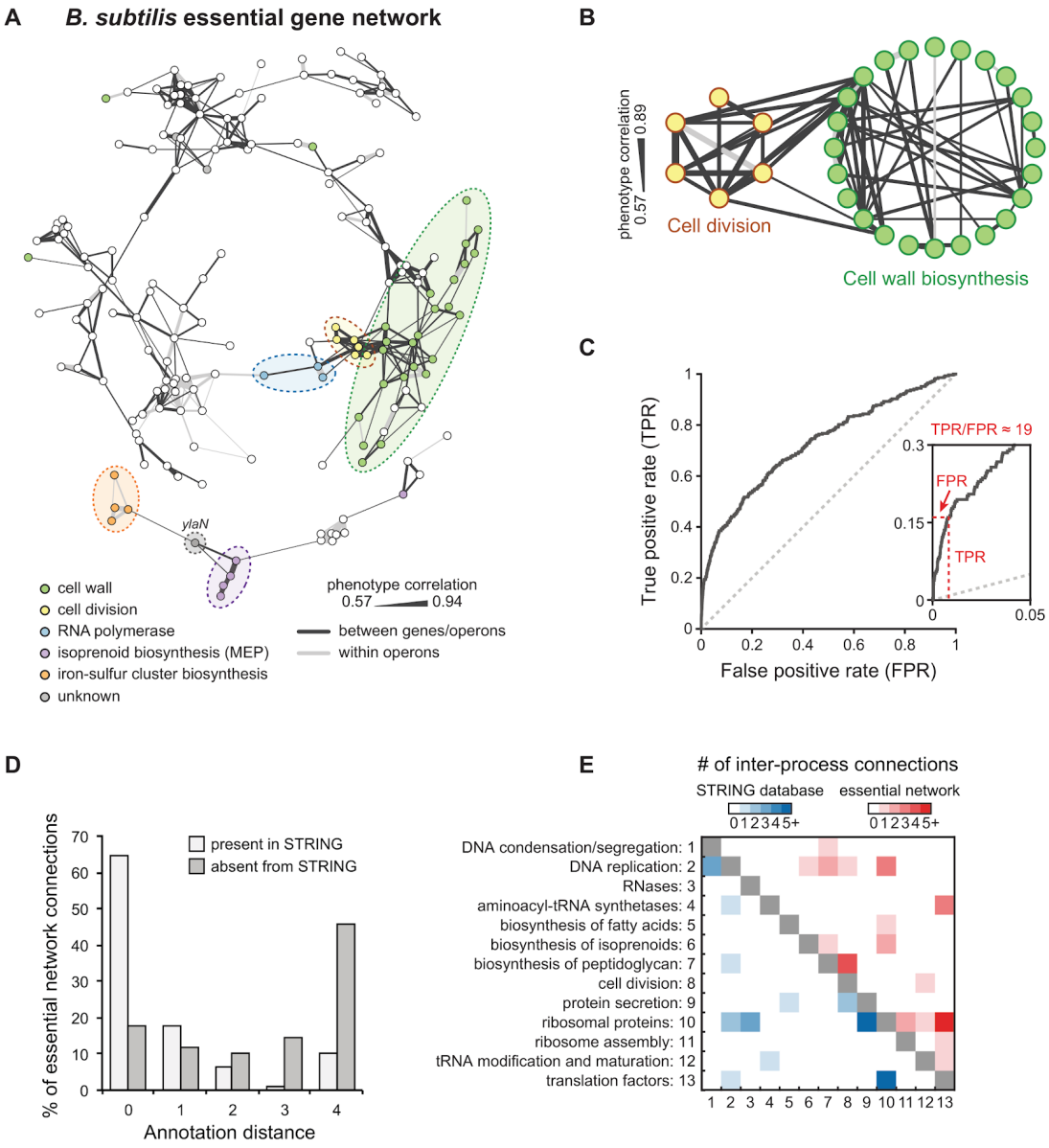


Figure 2.3. An Essential Gene Network Reveals Numerous Intra- and Inter-process Connections.

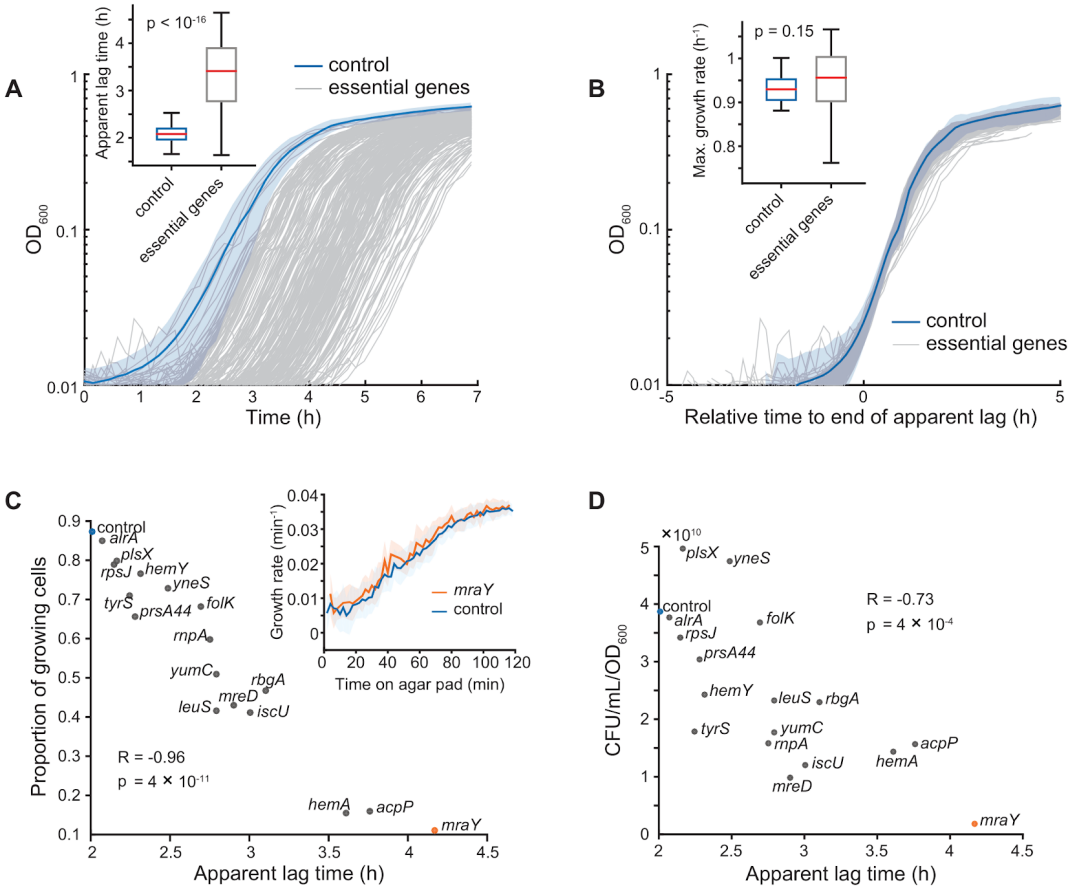
A) Essential gene network based on correlations between chemical-gene phenotypes. Edge thickness is proportional to the extent of correlation. See also **Figure 2.10** for network gene names.

B) Intra- and inter-process connections between cell division and cell wall-biosynthesis genes. Genes outside the main network or genes lacking intra-process connections were excluded.

C) ROC curve comparing connections between essential operons in our network to the STRING database. True-positive connections are those present in the high-confidence set of interactions from STRING, and false-positive connections are absent from STRING (these connections may be either truly false or novel).

D) Annotation distance for network connections between essential operons present or absent from the STRING database. Genes with an annotation distance of 0 are from the same functional group, while genes with an annotation distance of 4 are unconnected by annotation. E) Functional annotations for intra-process connections between essential operons present in the STRING database or present in our essential network.

Figure 2.4. High-resolution Growth Profiles of Essential Gene Knockdowns Reveal Widespread Defects in



Stationary-phase Survival.

A) Microplate reader growth curves of essential gene knockdown library strains. We grew cells for 18 h in LB before back-diluting into fresh LB (*t* = 0). Shaded area is mean ± standard deviation (S.D.) for the no-sgRNA control, *n* = 23.

B) Growth curves from A with the apparent lag time of each strain shifted to zero. Shaded area is mean ± S.D. for the no-sgRNA control, *n* = 23.

C) Data extracted from single-cell, time-lapse microscopy of selected essential gene knockdown strains grown on agarose pads with fresh LB after liquid growth in LB for 18 h. Although fewer *mraY* knockdown cells grow on LB pads after 18 h (11% versus 87% for the no-sgRNA control), the growth rate of elongating cells quantitatively matches that of the control (inset).

D) Viable cell plating of selected essential gene knockdown strains grown in LB for 18 h. CFU: colony-forming units.

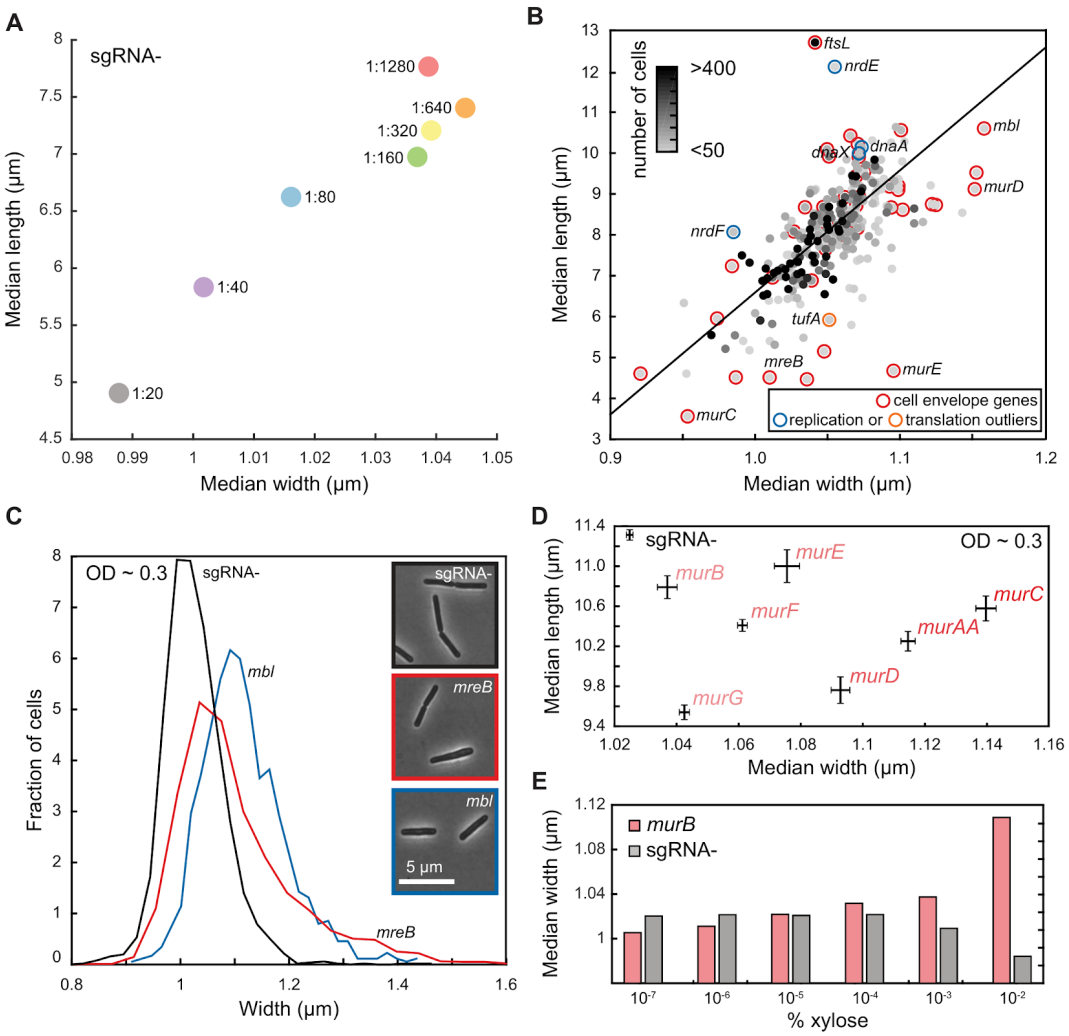


Figure 2.5. Partial Knockdown of Essential Genes Identifies Potential Morphological Regulators, with Envelope Gene Knockdown Leading to Changes in Cell Width.

A) Cell volume of the no-sgRNA control varies substantially after 3.5 h of growth across dilutions of the same overnight culture into fresh LB. Cell length and width were highly correlated (Pearson's $R = 0.96$, $p<0.001$).

B) Median cell length and width of essential knockdown strains are highly correlated after 3.5 h of growth (Pearson's $R = 0.66$, $p<10^{-38}$). "Cell envelope" is the only functional category that is enriched in the outliers to the best-fit line (black) between length and width. Selected non-cell envelope outliers in DNA replication (blue) and translation (orange) are also shown.

C) Distribution of cell widths of *mbl* and *mreB* knockdown strains at OD_{600}~0.3 reveals subtle distinctions between the two actin homologs, with *mbl* cells wider than no-sgRNA control cells and *mreB* cells adopting a broader range of widths. $n > 2000$ cells for each histogram.

D) Knockdown strains in the *mur* pathway have similar median cell length at OD_{600}~0.3, while cell width varies over a wide range. Bars represent standard error of the mean.

E) Cell width of the *murB* knockdown strain increases monotonically with the degree of *dcas9* induction, in contrast to the no-sgRNA control.

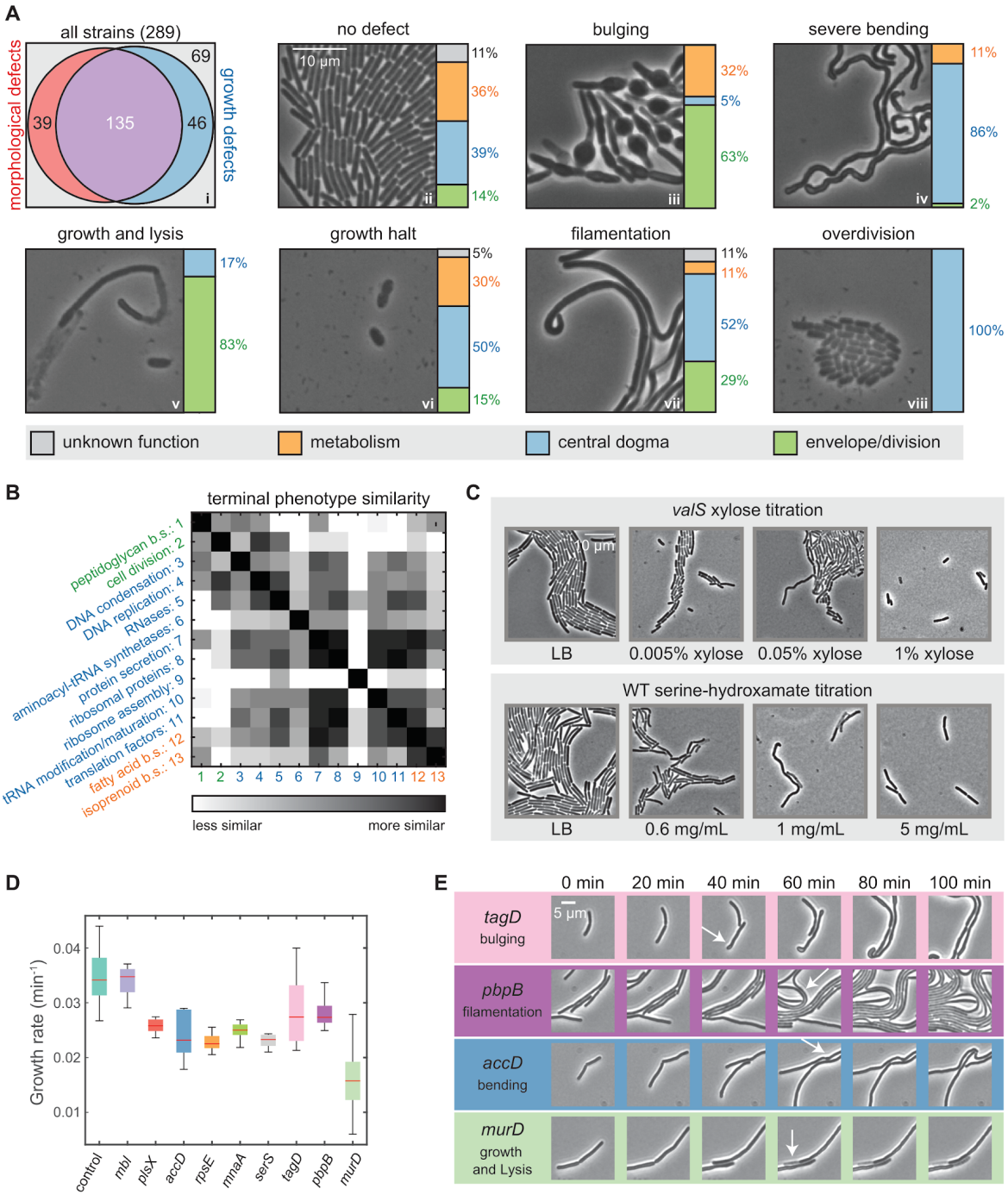


Figure 2.6. Essential Gene Depletion Reveals a Diversity of Terminal Phenotypes.
A) (i) Area-proportional graph of the fraction of essential gene knockdowns that give rise to morphological and growth terminal phenotypes. (ii-viii) Single-cell imaging of common terminal phenotypes of essential-gene knockdowns, with bar graphs depicting the broad functional categories underlying each terminal phenotype.
B) Matrix of the similarity of terminal phenotypes achieved by genes belonging to the most common essential-gene functional groups. b.s., biosynthesis.

C) Titrated depletion of the *valS* knockdown and inhibition of WT by serine hydroxamate led to similar trends in terminal phenotypes, from over-division to filamenting and bending, and finally growth halting.
D) Post-induction growth rates of selected strains with different terminal phenotypes. Most strains reduced growth rates by 20-30% in the first hour, except for *murD* and *mbl*. *murD* had a larger decrease in growth rate, whereas *mbl* had a similar growth rate as the no-sgRNA control.
E) Morphological phenotypes during depletion post-CRISPRi induction. Phenotypes were observable by 40-60 min post-induction.

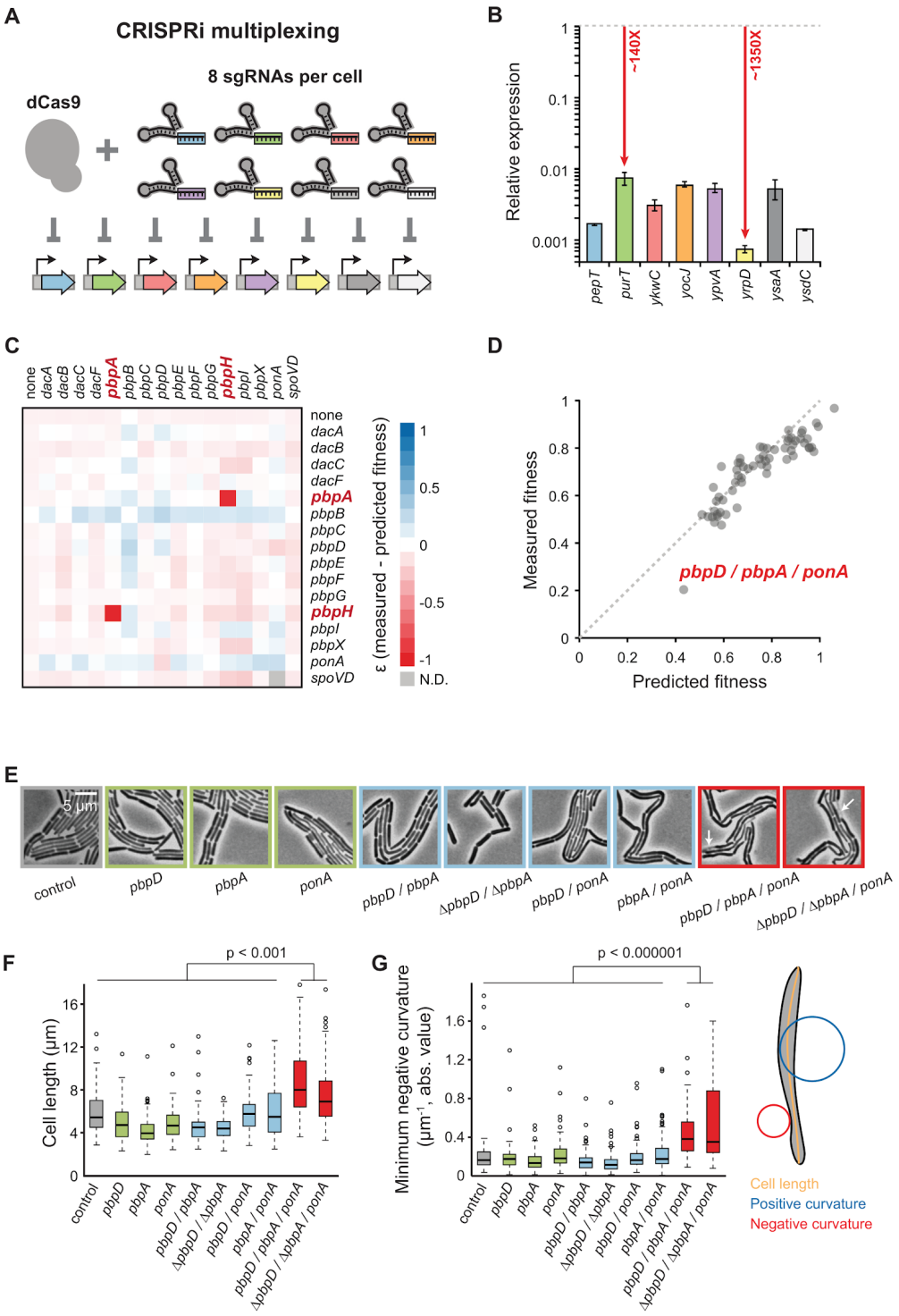


Figure 2.7. Multiplexed CRISPRi Knockdowns Facilitate Genetic Analysis of Complex Pathways.

A) Schematic of knockdown of eight genes in a single cell. Knockdowns correspond to the genes in (B).

B) Quantitative PCR of RNA levels in the eight-gene knockdown strain at maximal *dcas9* induction (1% xylose). Expression is relative to the no-sgRNA control. Points are the means of at least three measurements, and error bars are ±1 standard deviation.
C) The difference between measured and predicted fitness (ε, Extended Experimental Procedures) for each *pbp* double knockdown based on normalized colony size at maximal *dcas9* induction (1% xylose). Negative and positive ε represent reduced or improved fitness, respectively.
D) Measured versus predicted fitness of *pbp* triple knockdowns based on normalized colony size at maximal *dcas9* induction (1% xylose).
E) Terminal phenotypes of *pbp* knockdown strains. Arrows indicate lysed cells.
F) Box plots of cell length from *pbp* knockdowns in E. $n > 100$ cells for each strain.
G) Box plots of cell curvature from *pbp* knockdowns in E. $n > 100$ cells for each strain.

EXTENDED EXPERIMENTAL PROCEDURES

Plasmid Construction

Plasmids are available from the Bacillus Genetic Stock Center (http://www.bgsc.org/). *dcas9* was PCR-amplified from pdCas9-bacteria (Addgene #44249) using primers containing BamHI-compatible BsaI sites, digested with BsaI, then ligated into plasmid pAX01 digested with BamHI to generate pJMP1 (P_{xyl}-*dcas9*, Erm^R). The *rfp*-targeting $sgRNA^{RR1}$ was PCR-amplified from pgRNA-bacteria (Addgene #44251) using primers containing the *veg* promoter and EcoRI sites, digested with EcoRI, then ligated into either pDG1662 digested with EcoRI to generate pJMP2 (P_{veg}-$sgRNA^{RR1}$, Cm^R), or pDG1731 digested with EcoRI to generate pJMP3 (P_{veg}-$sgRNA^{NT1}$, Spc^R). Essential gene and *pbp* library plasmids were generated from either pJMP2 or pJMP3 using inverse PCR as previously described (Hawkins et al. 2015; Larson et al. 2013). $uppS^{Bs}$ was PCR-amplified from *B. subtilis* 168 genomic DNA using primers containing SalI and NheI sites, digested with SalI and NheI, then ligated into pDR110 digested with BamHI to generate pDR-$uppS^{Bs}$ (P_{spank}-$uppS^{Bs}$, Spc^R). $uppS^{Bs}$ was PCR-amplified from *S. aureus* N315 genomic DNA using primers containing HindIII and SphI sites, digested with HindIII and SphI, then ligated into pDR110 digested with HindIII and SphI to generate pJMP11 (P_{spank}-$uppS^{Sa}$, Spc^R). Multiple sgRNA plasmids were constructed using BsaI-mediated cloning as previously described (Hawkins et al. 2015), except for double *pbp* sgRNA plasmids, which were constructed by Gibson assembly (Gibson et al. 2009) using the Gibson Assembly master mix kit (New England Biolabs).

Strain Construction

Strains are available from the Bacillus Genetic Stock Center (http://www.bgsc.org/). All *B. subtilis* strains were constructed using natural competence via either a standard or high-throughput method (Koo et al. 2017).

Standard Method

3 ml of MC medium (10.7 g/L potassium phosphate dibasic, 5.2 g/L potassium phosphate monobasic, 20 g/L glucose, 0.88 g/L trisodium citrate dihydrate, 0.022 g/L ferric ammonium citrate, 1 g/L casein hydrolysate, 2.2 g/L potassium glutamate monohydrate, 20 mM magnesium sulfate, 150 nM manganese chloride, 20 mg/L tryptophan) were inoculated with a single colony of *B. subtilis* and incubated at 37 °C overnight (≥10 h). The overnight culture was diluted to an OD_{600} of 0.1 in 10 mL BMK medium (10.7 g/L potassium phosphate dibasic, 5.2 g/L potassium phosphate monobasic, 20 g/L glucose, 0.88 g/L sodium citrate dihydrate, 0.022 g/L ferric ammonium citrate, 2.5 g/L potassium aspartate, 10 mM magnesium sulfate, 150 nM manganese chloride, 40 mg/L tryptophan, 0.05% yeast extract), then grown in a 125 mL flask at 37 °C with shaking (250 rpm) until cells reached OD_{600}~1.5. 150 µL of culture were then mixed with ≥100 ng of plasmid DNA in a deep 96-well plate, covered with a breathable film, and incubated at 37 °C without shaking for 10 min, then incubated at 37 °C with shaking (900 rpm) for 2 h. After 2 h, cells were plated on LB agar containing selective antibiotics (7.5 µg/mL kanamycin, a combination of 1 µg/mL erythromycin and 15 µg/mL lincomycin, 6 µg/mL chloramphenicol or 100 µg/mL spectinomycin [by activity]).

High-throughput Method

Double and triple *pbp* knockdown libraries were constructed using high-throughput transformation. Individual wells of a deep 96-well plate containing 300 µL of MC medium were inoculated with single colonies of strains containing P_{xyl}-*dcas9* and an sgRNA targeting one of the 16 *pbp* genes; this plate was incubated at 37 °C with shaking (900 rpm) for at least 10 h. Cells were then diluted to OD_{600}~0.1 in BMK medium, and 25 µL of diluted cells were added to ≥100 ng of plasmid DNA in a shallow, v-bottom 96-well plate, covered with a breathable film, and incubated at 37 °C in a humidified incubator without shaking for 16 h. 50 µL of LB were added to each well and the well contents were plated on LB agar containing selective antibiotics.

Flow Cytometry

Strains were grown overnight in LB in deep 96-well plates, and then back-diluted 1:300 into fresh LB containing the appropriate concentration of xylose to induce expression of *dcas9*. After ~5 h of growth, cells were diluted 1:300 into phosphate buffered saline, and red fluorescence levels (B-A laser) were determined using an LSRII flow cytometer (BD Biosciences). Data for at least 10,000 cells were collected. Median red fluorescence signals were extracted from FCS files using FlowJo (FlowJo, LLC); error bars are from three biological replicates.

RNA-seq

Cells were grown in LB or LB + 1% xylose to OD_{600}~0.3. RNA extraction and RNA-seq was performed as previously described (Larson et al. 2013). RNA-seq data was deposited at the Gene Expression Omnibus (GSE74926).

NET-seq

Cells were grown in LB + 1% xylose to OD_{600}~0.3. NET-seq was performed as previously described (Larson et al. 2014). NET-seq data was deposited at the Gene Expression Omnibus (GSE74926).

sgRNA Design

The 20-nucleotide guide sequences for our sgRNA library were designed to be effective and specific. We first found all GG dinucleotides in the target genome, and extracted the 20 bases ending one base 5' of that GG as the potential target candidates for the organism. We then scored the specificity of each 20-nt guide by passing all the guides through Bowtie with the genome as the reference, then passing each 19-nt suffix through Bowtie, then each 18-nt suffix, etc. The score for a guide was the shortest suffix for which only a single alignment was found. Lower specificity score numbers were therefore better. If we never found a unique alignment at any length, the score was listed as -1, and the guide was avoided. Next, we annotated each

target with any coding regions overlapping that target. We noted the position of the target relative to the first coding base of the gene. For each gene we chose a target that was as close as possible to the 5' end of the coding sequence for that gene and that had a low (good) specificity score.

sgRNA Efficacy Analysis

To test the efficacy of our essential gene knockdown library, we pinned the library onto rectangular LB agar plates containing 1% xylose to fully induce expression of *dcas9*. Of the 299 strains in the library, 258 targeted bona fide essential genes as determined by a gold-standard gene deletion analysis that attempted to replace every open reading frame in the *B. subtilis* genome with an antibiotic resistance marker (Koo et al. 2017). 244 of the 258 (95%) strains showed a colony size defect of at least 25% (**Table 2.2**). We then tested the 14 strains that failed to show a colony size defect by measuring growth in liquid LB + 1% xylose using a microplate reader. We found that 5 of the 14 strains showed 50% or less growth compared to the growth of the no-sgRNA control at mid-log phase (OD_{600}~0.2; **Table 2.2**), bringing the percentage of effective sgRNAs to 97%. Of the nine strains that failed to show growth defects, five target phage genes; these genes may not be truly essential, as the antibiotic resistance marker used in the deletion analysis contains a strong promoter that may upregulate adjacent, toxic phage genes that are normally transcriptionally silent. Alternatively, dCas9 may be acting to repress phage transcription that is normally toxic (this appears likely for sgRNAs targeting the SKIN element repressor, *sknR*/*yqaE*). In the cases of *yezG* (a putative antitoxin) and *yhdL* (anti-SigM), CRISPRi repression of upstream genes on the same transcript appears to also silence expression of toxic proteins normally controlled by these gene products. We conclude that nearly all sgRNAs targeting essential genes showed repression activity.

Quantitative PCR

Overnight LB cultures of *B. subtilis* were diluted 1:1000 in 2 mL LB + 1% xylose. These liquid cultures were grown for 6 h at 37 °C. For RNA extraction, 600 µL of culture were mixed with 600 µL of -20 °C methanol and spun down at full speed for 5 min. Once decanted, the RNA was extracted with the Qiagen RNeasy RNA isolation kit protocol #7. Following RNA isolation, genomic DNA was eliminated with the Ambion DNA-Free kit per manufacturer's instructions. cDNA was synthesized with the Invitrogen SuperScript III First-Strand Synthesis System. Random hexamers were annealed, and then cDNA was synthesized via SuperScript III RT. cDNA was diluted 10-fold to be used in conjunction with Agilent Brilliant II SYBR Green QPCR Master Mix and 0.6 µM final concentration primers. Primers were designed with Primer3 with the following parameters: 60 °C melting temperature, 200 base pair-amplified region, and amplified region within the middle of the gene. Control genes employed consisted of *sigA*, *gyrA*, and *rpoB*. The qPCR included RNA samples that had DNase treatment, as well as RNase treatment, to verify that no genomic DNA remained. qPCR experiments were conducted in a Stratagene Mx3005P qPCR System. Three biological replicates were used in each qPCR.

CRISPRi *pbp* Double and Triple Knockdown Screening

Double and triple knockdown strains containing two sgRNA plasmids that integrate at either *amyE* or *thrC* were constructed using the high-throughput transformation method described above. Single colony isolates were stored as glycerol stocks in 96-well format. To screen the library, cells were robotically pinned from glycerol stocks onto rectangular LB agar plates in 384-colony format using a ROTOR robot (Singer Instruments), then pinned once more to 1536-colony format. Cells in 1536-colony format were then pinned to LB + 1% xylose plates to fully induce *dcas9* expression. LB plates with PBP inhibitors also contained 1% xylose.. Plates were imaged using a Powershot G10 camera (Canon) after ~7-9 h of growth at 37 °C, and colony size was extracted using the "opacity" setting in the Iris software package (Paradis-Bleau et al.

2014). Colony size was normalized within plates using internal controls that expressed *dcas9*, but did not contain sgRNAs.

The fitness (*W*) of double (or higher-order) deletion strains can be predicted by multiplying the fitness values of each of the single deletions (deletions of gene x and gene y) that comprise the double (predicted $W_{xy} = W_x \times W_y$); this is known as the multiplicative model (St Onge et al., 2007). Large values of the deviation between the measured fitness and the predicted fitness of double deletions (ε = measured W_{xy} – predicted W_{xy}) indicate a genetic interaction. Because our PBP genetic interaction screen used CRISPRi knockdowns rather than gene deletions, it was unclear if the multiplicative rule would apply to our analysis. To test for agreement between the multiplicative model and our CRISPRi screen data, we plotted predicted fitness versus measured fitness for all double knockdown strains (**Figure 2.14 A**). We found a strong, linear correlation (Pearson's $R = 0.95$) between predicted and measured fitness, suggesting that the multiplicative model can reliably be applied to our CRISPRi knockdown data. The small deviation we observed from the ideal multiplicative model (ideal fit: $y = x$; our fit: $y = 0.81x+0.13$) may be due to the relatively small sample size of our screen ($n = 240$ double knockdowns).

MAC-0170636 Screening and UppS Activity Assay

Cloning, Overexpression, and Purification of B. subtilis UppS

The gene *uppS* (GeneBank sequence NC_000964.3) was cloned into the pET-19b vector (Novagene) modified to encode an engineered Tobacco Etch Virus (TEV) protease cleavage site using primers *uppS*_FWD: CTAGCATATG CTCAACATACTCAAAAATTG and *uppS*_REV: CTAGCTACGAG CTAAATTCCGCCAAA. Primers used to generate the construct are listed below. UppS was over-expressed in *E. coli* BL21 (Rosetta) using auto-induction (Studier 2005). Cells were harvested by centrifugation at 6000 × *g* for 25 min at 4 °C. Cells were suspended in lysis buffer containing 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, 10 mM imidazole, 250 KU rLysozyme (Novagen) and an EDTA-free protease inhibitor tablet (Roche Diagnostics). Cells

were lysed using a cell disrupter (Constant Systems Limited, Daventry, UK). Lysates were cleared by centrifugation at 30,000 × *g* for 30 min at 4 °C. Nickel-affinity chromatography was performed on the lysates using a 30 mL free-flow gravity column and 5 mL of Ni-NTA agarose (Sigma). Once loaded, the column was washed with buffer (10 column volumes) containing 50 mM NaH_2PO_4 (PH 8.0), 300 mM NaCl, 20 mM imidazole. His-tagged UppS was then eluted in buffer containing 50 mM NaH_2PO_4 (PH 8.0), 300 mM NaCl, 250 mM imidazole. Elution fractions were dialyzed overnight against 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, and 1 mM DTT. Once dialyzed, the elution fraction was concentrated using Amicon Ultra centrifugation filters (10kDa cut-off) and quantified using a NanoDrop (Thermo Scientific).

Assay for UppS Inhibition by MAC-0170636

A Kinetic EnzCheck pyrophosphate assay (Life Technologies) was used to assess inhibition of MAC-0170636 *in vitro* in accordance with the manufacturers guidelines. The IC_{50} value was determined in 100 µL reaction volumes using a flat bottom 96-well plate (Costar 3370) in duplicate. Reactions with varying concentrations of MAC-0170636 were conducted with a final concentration of 1% DMSO in the presence of 0.2 mM 2-amino-6-mercapto-7-methyl-purine ribonucleoside (MESG), 0.625 U of purine ribonucleoside phosphorylase (PNP), 0.2 U of inorganic pyrophosphatase (PyroP), 0.125 µg of purified $UppS^{Bs}$ enzyme, 0.82 µM farensyl pyrophosphate (FPP) (1xKM) and 65 µM isopentenyl pyrophosphate (IPP) (5x KM). For $UppS^{Sa}$, 0.5 µg of purified UppS enzyme, 0.7 µM farensyl pyrophosphate (FPP) (1xKM), and 80 µM isopentenyl pyrophosphate (IPP) (5x KM) were used in the reaction. Data were fit using GraFit V5 (Erithacus Software).

CRISPRi Essential Gene Knockdown Screening and Network Construction

Essential Gene Knockdown Screening

sgRNA plasmids were transformed into a strain containing *dcas9* (CAG74209) using the standard method described above. Single colony isolates of each transformation were grown

overnight in LB + 6 μg/mL chloramphenicol, and stored as glycerol stocks at -80 °C in 96-well plates. Prior to screening, cells were robotically pinned onto rectangular LB agar plates in 384-colony format, and then pinned once more to 1536-colony format using a ROTOR robot (Singer Instruments). To screen the essential library, we pinned from rectangular LB agar plates in 1536-colony format to plates containing the indicated concentration of antibiotic or chemical stress. Each plate contained four technical replicates for each sgRNA strain and each chemical concentration was replicated at least twice; most concentrations had four replicate plates. Chemical concentrations were empirically determined by streaking wild-type *B. subtilis* 168 onto round agar plates containing chemicals and then visually inspecting the plates for colony size defects. We used concentrations in our screen that inhibited growth by 50% or less; chemical concentrations that inhibited growth such that the median colony size after 14 h was less than the median colony size on LB after 7 h were discarded. Importantly, the effective concentration of chemical in our screening plates is likely to be lower than the concentration added to the plate because of chemical breakdown during the two-day period in which plates were dried at room temperature to reduce problems with colony smearing due to wet plates. Pinned cells were grown on rectangular plates for 7-14 h at 37 °C, and then imaged using a custom light box with a Canon Powershot G10 (Canon).

Data Analysis and Network Construction

Colony sizes were extracted from plate images using the Iris software package (Paradis-Bleau et al. 2014). Spatial effects were normalized using a quadratic function, median and variance of colony sizes were normalized between plates, and S-scores were computed using previously developed software (Collins et al. 2006). False discovery rates were computed from S-scores as previously described (Nichols et al. 2011).

To construct an essential gene network from gene-chemical scores (i.e., S-scores), these scores were correlated (Pearson correlation) in a pairwise manner for all sgRNA knockdown strains, resulting in ~90,000 gene-gene correlations. To establish a cutoff for correlation

significance, we estimated the background distribution of correlations by randomly permuting gene identity relative to the chemical-gene scores 5,000 times, and then used the 95% confidence interval of this background distribution (after discarding self correlations that equaled 1) to establish a correlation significance threshold of 0.572 (method adapted from Nichols et al., 2011); this procedure resulted in 412 significant connections between distinct sgRNA knockdown strains. Network data was visualized using Cytoscape v3.2.1 (Cytoscape Consortium) (Shannon et al. 2003).

Network Comparison to the STRING Database

Essential operons were defined based on manual curation of global expression data (Nicolas et al. 2012) and functional annotations from the SubtWiki database (http://subtiwiki.uni-goettingen.de/). For any essential operon with genes in different annotations, the operon was split into subsets of genes with the same annotation; correlations between genes in the same operon were eliminated from the analysis. Pairs of essential operons were defined as interacting if any pair of genes spanning both operons had a correlation above the threshold value 0.572. The connectivity between two functional groups was defined as the number of essential operon interactions between the two groups. If two genes within the same essential operon fell into distinct functional groups, the interaction between them did not contribute to the functional group connectivity. Similarly, the interaction between operons that have been classified as single operons in other studies did not contribute to the functional group connectivity.

Comparison to the STRING database was achieved by downloading the *Bacillus subtilis* 168-specific interactions scores from http://string-db.org/ (Szklarczyk et al. 2015). Scores were culled to the “experimental” or “database” categories. Genetic interactions in the STRING database were combined into essential-operon interactions in an equivalent manner to our chemical genomics data.

Annotation distance analysis between essential gene operons was performed using annotations from the SubtiWiki online database (http://subtiwiki.uni-goettingen.de/). SubtiWiki annotations

are broken down into four levels: the first level is very general (e.g., metabolism) and the fourth level is specific (e.g., RNA polymerase). In cases in which there were only three levels of annotation, the annotation from the third level was also used for the fourth level (**Table 2.3**). Operons with the same annotation at the fourth level (most specific) were given an annotation distance of 0, while operons with no overlapping annotations were assigned a 4. As the annotation distance was subject to bias based on the extent of annotation (e.g., whether or not strains were annotated at the fourth level), we suggest that the distance be interpreted as a qualitative metric for annotation differences.

High-resolution Liquid Growth Curve Acquisition and Analysis

All strains were grown in deep 96-well plates for 18 h, then back-diluted 1:200 into 200 µL fresh LB (for induced growth curves, LB was supplemented with 1% xylose, and the overnight cultures were grown for 26 h) and grown with shaking at 37 °C in an Infinite M200 plate reader (Tecan) for 7 h. The absorbance at wavelength 600 nm (OD_{600}) was measured at 7.5-min intervals.

For uninduced growth curves, the natural logarithm of the optical density was fit to the Gompertz equation (Zwietering et al. 1990) to determine apparent lag time, maximum specific growth rate, and from that, the doubling time. Each growth curve was fit individually.

For induced growth curves, several criteria were used to classify the curves into different categories: 1) no growth: OD_{600}<0.06 throughout the measurement; 2) linear growth: final OD_{600}>0.2, Pearson correlation coefficient >0.98 between time and OD_{600} after removal of the lag and saturation regions of the curve, and the slope at the end of the growth phase is >90% of the slope at the beginning of growth; 3) growth and death: maximum OD_{600} is >1.2 × the final OD_{600}, with final OD_{600}>0.06; 4) others: strains that do not fall into any of the previous three categories.

Viable Counts

Strains were grown in deep 96-well plates for 18 h. The OD of each overnight culture was measured using a Genesys 20 spectrophotometer (Thermal Scientific). Overnight cultures were serially diluted in LB, and at each dilution, 100 µL of culture were plated on LB. Plates were incubated overnight at 37 °C, and the dilutions with ~100 colonies were used to count the number of colony forming units (CFUs).

Each overnight culture used for CFU counting was also diluted 1:100 in fresh LB, and 1 µL of the diluted culture was spotted onto a pad of 1.5% agarose in fresh LB. The pads were incubated at 37 °C for 2 h and then imaged. Dead or non-growing cells remained small and isolated from other cells, and live cells grew into collections of larger cells. The number of dead/non-growing cells and live cells were manually counted for each strain; a total of at least 200 cells were counted for each strain.

High-throughput Imaging and Analysis

For phase-contrast and fluorescence imaging, cells were imaged on a Nikon Eclipse TE inverted fluorescence microscope with a 100X (NA 1.40) oil-immersion objective. Images were collected using an Andor DC152Q sCMOS camera (Andor Technology, South Windsor, CT, USA) and µManager v1.3 software (Edelstein et al. 2010). For time-lapse imaging, cells were maintained at 37 °C during imaging with an active-control environmental chamber (HaisonTech).

Custom MATLAB (MathWorks) image processing code was used to segment cells and identify active cell contours from phase microscopy (Monds et al. 2014; Ursell et al. 2014). Cell widths and lengths were calculated using the MicrobeTracker meshing algorithm (Sliusarenko et al. 2011). For all single-cell quantification, at least 100 cells were analyzed for each strain. Enrichment of morphological defects in each functional group was identified by conducting a two-sample KS test between the average length trend line deviations of strains corresponding to a functional group, and the entire collection background. For *pbp* strains, cell lengths were

approximated by half of the contour lengths, and cell bending was defined as the minimum negative curvature of the cell contour.

Terminal Phenotypes

Microscopy

Cells were back-diluted 1:200 from overnight culture into fresh LB and grown at 37 °C for 2 h in a plate shaker, and then 1 µL of cells was spotted onto a pad of 1.5% agarose in fresh LB supplemented with 1% xylose. Cells were incubated on the pads overnight at room temperature, and then imaged. For imaging under antibiotic treatment and xylose titration, the appropriate concentrations of antibiotics or xylose were added to the agarose pads. For staining, 1 µg/ML DAPI (Invitrogen) and/or 2 µg/mL FM4-64 (Invitrogen) were added to the pads.

Analysis

Terminal phenotype for each essential gene knockdown was manually classified using an eight-dimensional phenotype vector encompassing features of lysis, bulging, uniform shape loss, bending, filamentation, over-division (shorter cells), and level of post-induction growth. Severe bulging, filamentation, or lysing phenotypes were classified as dominant such that secondary phenotypes were ignored. A given functional group of essential operons was described by its fractional composition of terminal phenotypes from essential operons in that group. The cosine similarity of terminal phenotypes between functional groups was calculated between the normalized fractional composition vectors.

SUPPLEMENTAL FIGURES

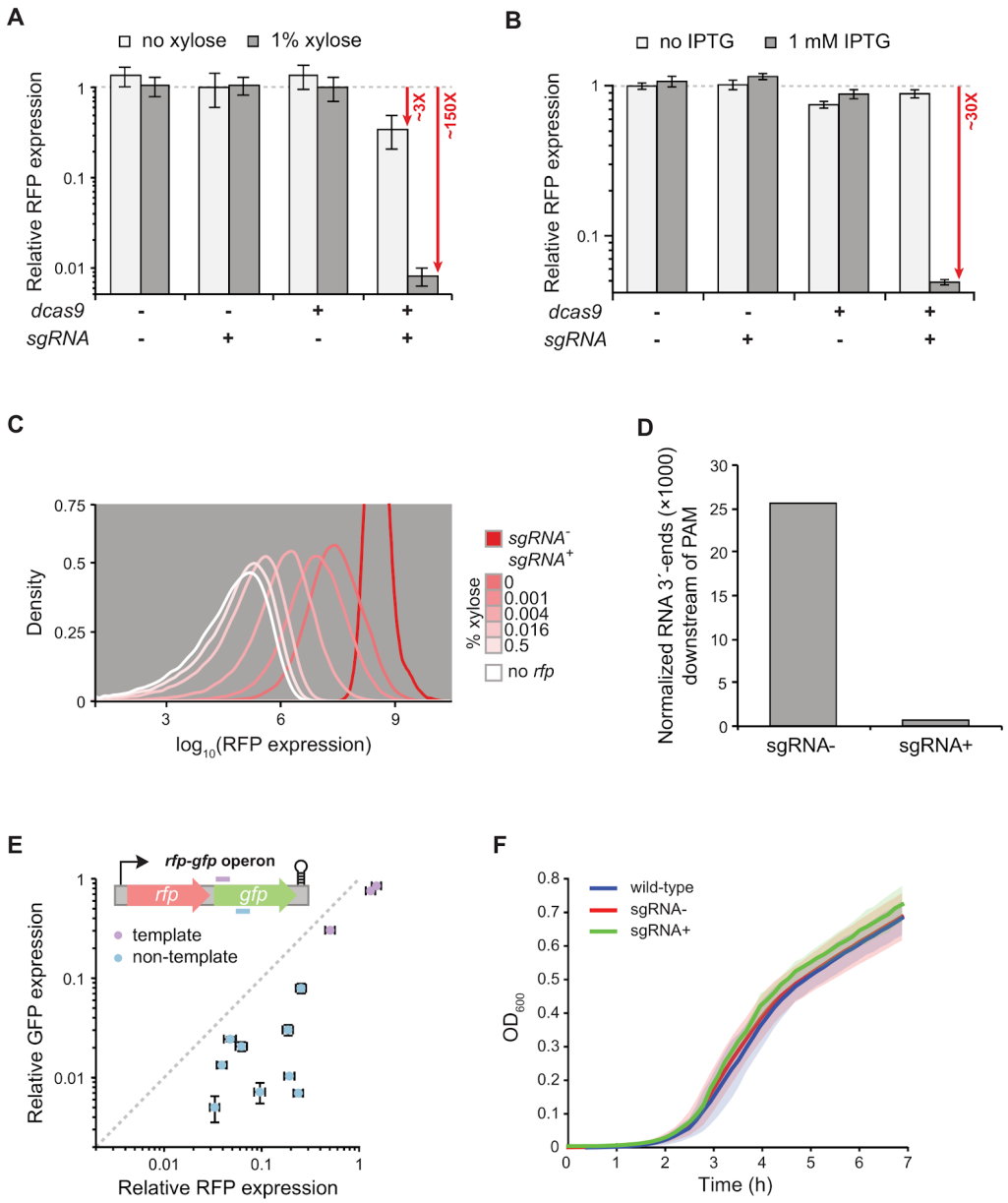


Figure 2.8. Further Characterization of the *B. subtilis* CRISPRi System, Related to Figure 2.1.

A) Flow cytometry of cells constitutively expressing *rfp*, with or without P_{xyl}-*dcas9* and an *rfp*-targeting sgRNA. RFP levels are relative to the sgRNA- control.

B) Flow cytometry of cells constitutively expressing *rfp*, with or without P_{IPTG}-*dcas9* and an *rfp*-targeting sgRNA. RFP levels are relative to the sgRNA- control. *rfp* knockdown without IPTG induction is <15%.

C) Flow cytometry of cells in which *dcas9* was induced by adding the specified concentration of xylose that constitutively express *rfp* and an *rfp*-targeting sgRNA shown as distributions of RFP

fluorescence for selected xylose concentrations. RFP levels per cell are relative to the sgRNA- control.
D) NET-seq of cells maximally induced for *dcas9* (1% xylose) and constitutively expressing *rfp* and an *rfp*-targeted sgRNA versus cells without an sgRNA. Bars show the total number of normalized RNA 3´-ends downstream of the sgRNA-dCas9 binding site.
E) Flow cytometry of cells constitutively expressing an *rfp-gfp* operon and various sgRNAs targeting *gfp* at maximal induction of *dcas9* (1% xylose). RFP and GFP levels are relative to the sgRNA- control. The dashed line is $y = x$. Note that the sgRNAs with low efficacy target the template strand (violet).
F) Growth curves of P_{xyl}-*dcas9* sgRNA- (red), P_{xyl}-*dcas9* sgRNA+ (green), and *B. subtilis* 168 cells (blue), with 1% xylose supplemented. Shaded areas are mean ± std for each strain.

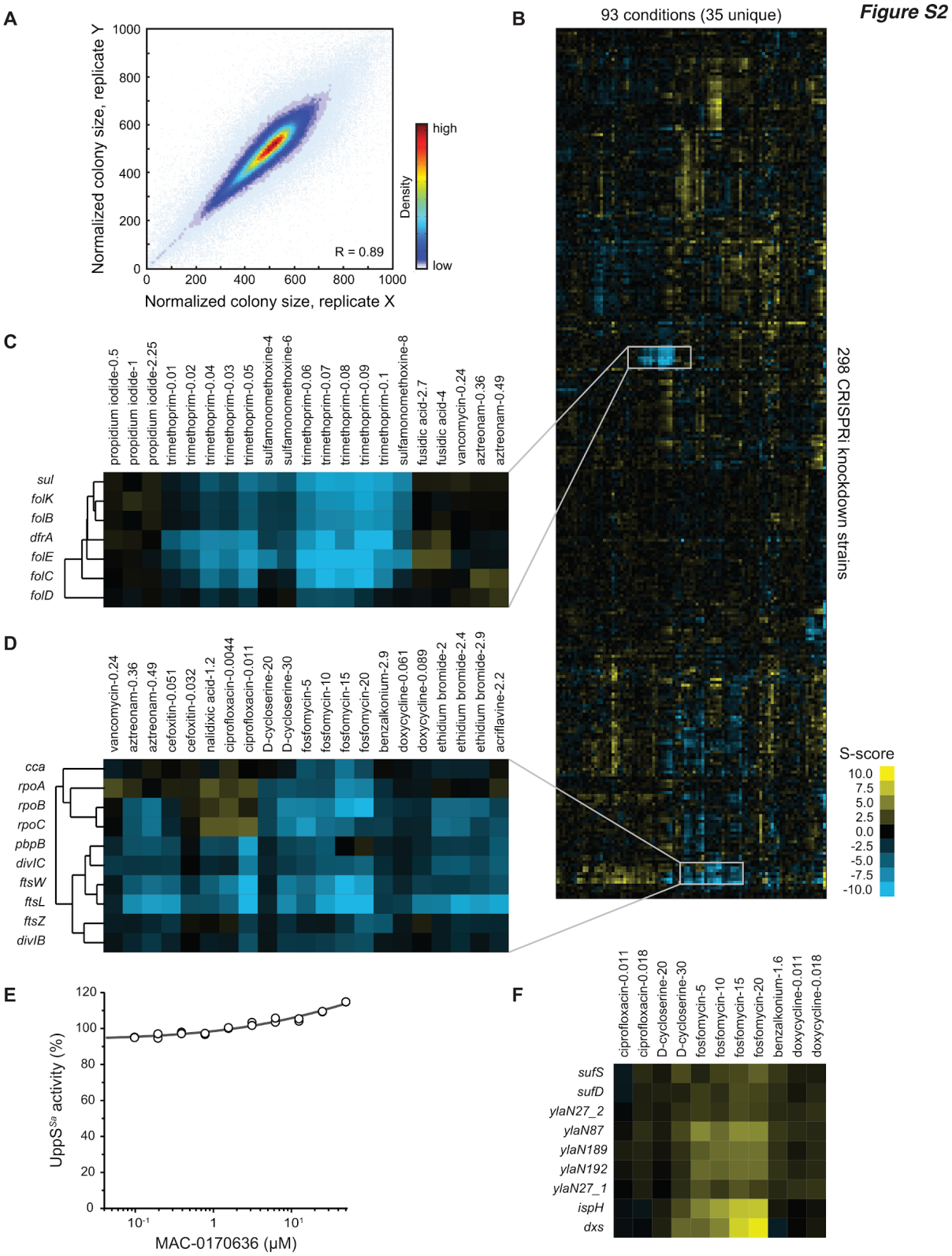


Figure 2.9. Essential Gene Knockdown Screen Reproducibility and Chemical-gene Phenotypes, Related to Figures 2.2 and 2.3.

A) Reproducibility between replicate colonies for all chemical conditions (*R*: Pearson's correlation).
B) Hierarchical clustering of chemical-gene scores (S-scores; (Collins et al., 2006)) for all chemical conditions.
C) A tetrahydrofolate biosynthesis and utilization cluster defined by knockdown strain sensitivity to the DfrA inhibitor trimethoprim and the Sul inhibitor sulfamonomethoxine.
D) A cell division/RNA polymerase cluster defined primarily by sensitivities to cell wall-acting antibiotics (aztreonam, cefoxitin, D-cycloserine, and fosfomycin) and DNA intercalators (ethidium bromide, and acriflavine).
E) Concentration-dependent inhibition of purified *S. aureus* UppS by MAC-0170636. Each point is an individual measurement.
F) A cluster containing *ylaN*, isoprenoid biosynthesis, and iron-sulfur cluster biosynthesis genes defined by resistance to cell wall-targeting antibiotics, especially fosfomycin.

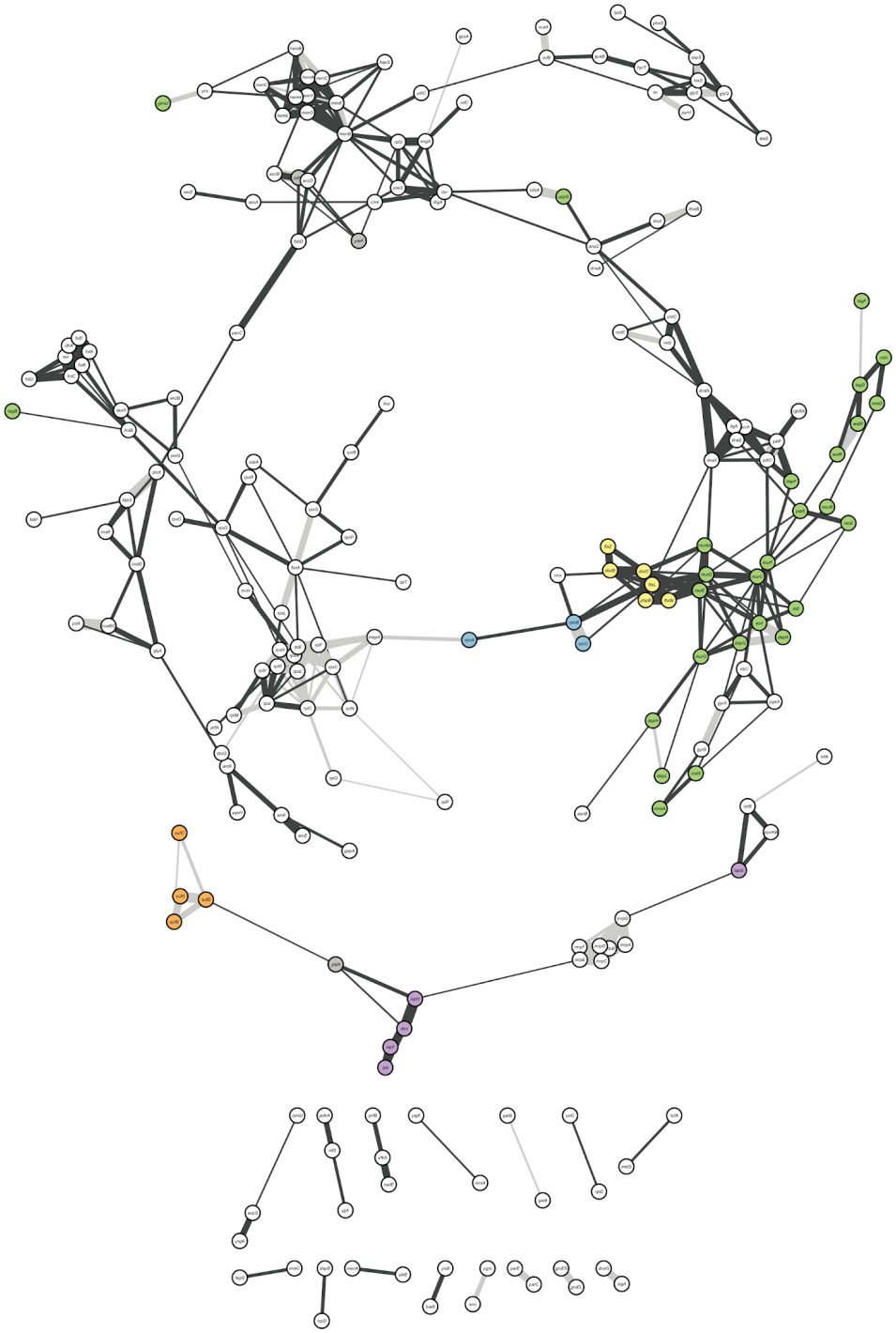


Figure 2.10. An Essential Gene Network Reveals Numerous Intra- and Inter-process Connections, Related to Figure 2.3.
The *B. subtilis* essential gene network as depicted in **Figure 2.3 A**, except showing sub-networks with only one or two connections and gene labels.

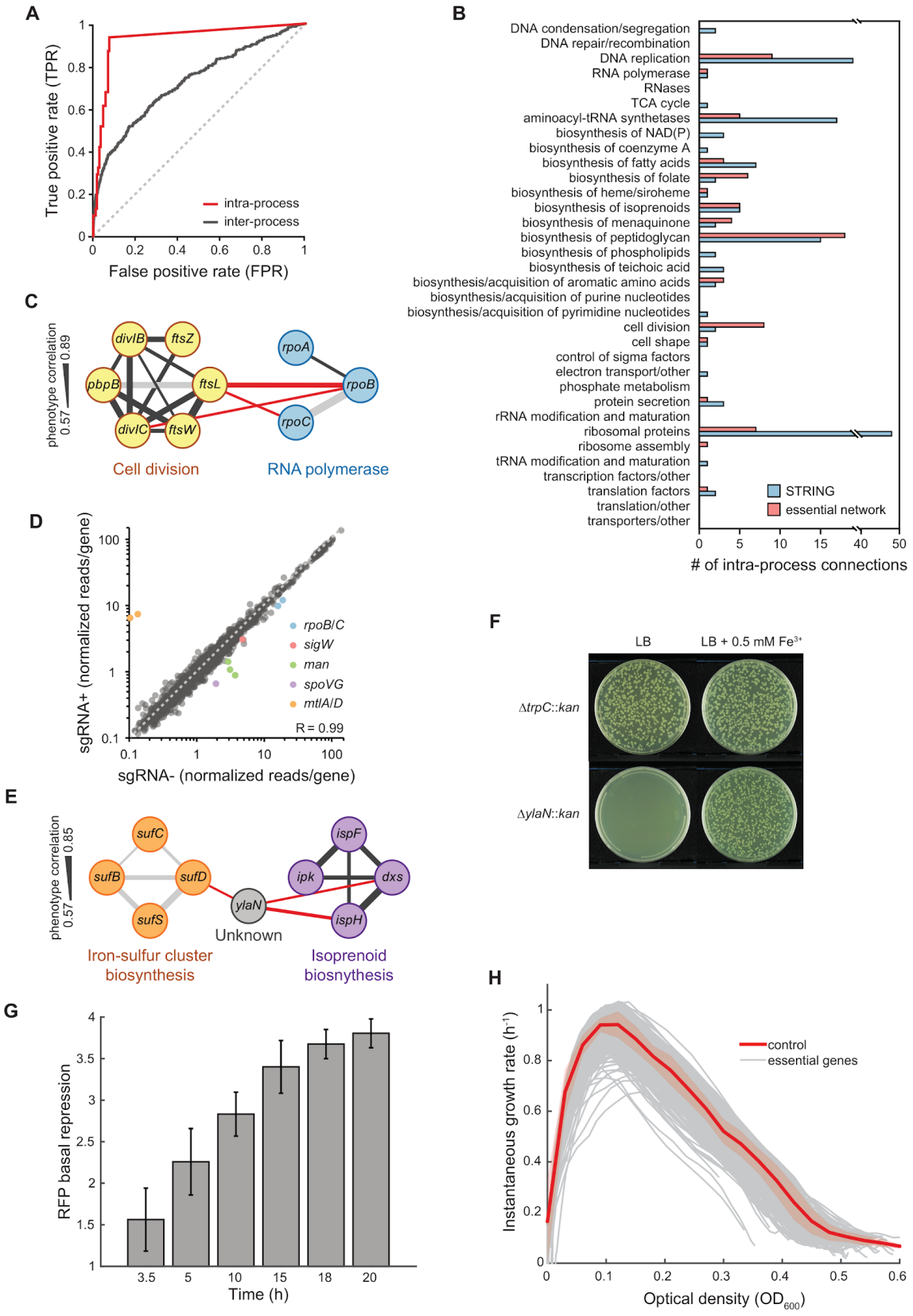


Figure 2.11. Further Characterization of the Essential Gene Network and Basal

Knockdown Levels, Related to Figures 2.3 and 2.4.
A) ROC curves depicting true positive rates and false positive rates for either intra- or inter-process connections between essential operons.
B) Absolute number of connections between essential operons in the STRING database (blue) or our essential network (red).
C) Previously uncharacterized connections between cell division and transcription (red lines); these connections are largely due to shared sensitivities of *ftsL*/*divIC* and *rpoB*/*C* knockdowns to cell wall-acting antibiotics and DNA intercalators (see also **Figure 2.9 D**). Black lines are connections between genes in different operons and grey lines are connections within operons.
D) RNA-seq of cells with basal expression of *dcas9* and an sgRNA targeting *rpoB* versus cells without an sgRNA. Reads per gene were normalized by RPKM. The dashed line is $y = x$. The *rpoBC* transcript is reduced ~40% by the *rpoB* sgRNA. Outliers and genes of interest are colored as indicated. The operon containing *manP*/*A-yjdF* is abbreviated as "*man*."
E) Previously uncharacterized connections among iron-sulfur cluster biosynthesis, isoprenoid biosynthesis, and the gene of unknown function *ylaN* (red lines). Black lines are connections between genes in different operons and grey lines are connections within operons.
F) *ylaN* is no longer essential in the presence of additional iron. *B. subtilis* 168 cells were transformed with PCR products that replaced either *trpC* (*B. subtilis* 168 is already mutant for *trpC*, so the replacement is neutral), or *ylaN* with a kanamycin-resistance marker.
G) Basal CRISPRi knockdown of *rfp* modestly increases during stationary phase.
H) Culture optical densities (ODs) determine the instantaneous growth rates for all strains. Instantaneous growth rates were calculated from OD measurements in **Figure 2.11 A**. Shaded area is mean ± std for sgRNA- control.

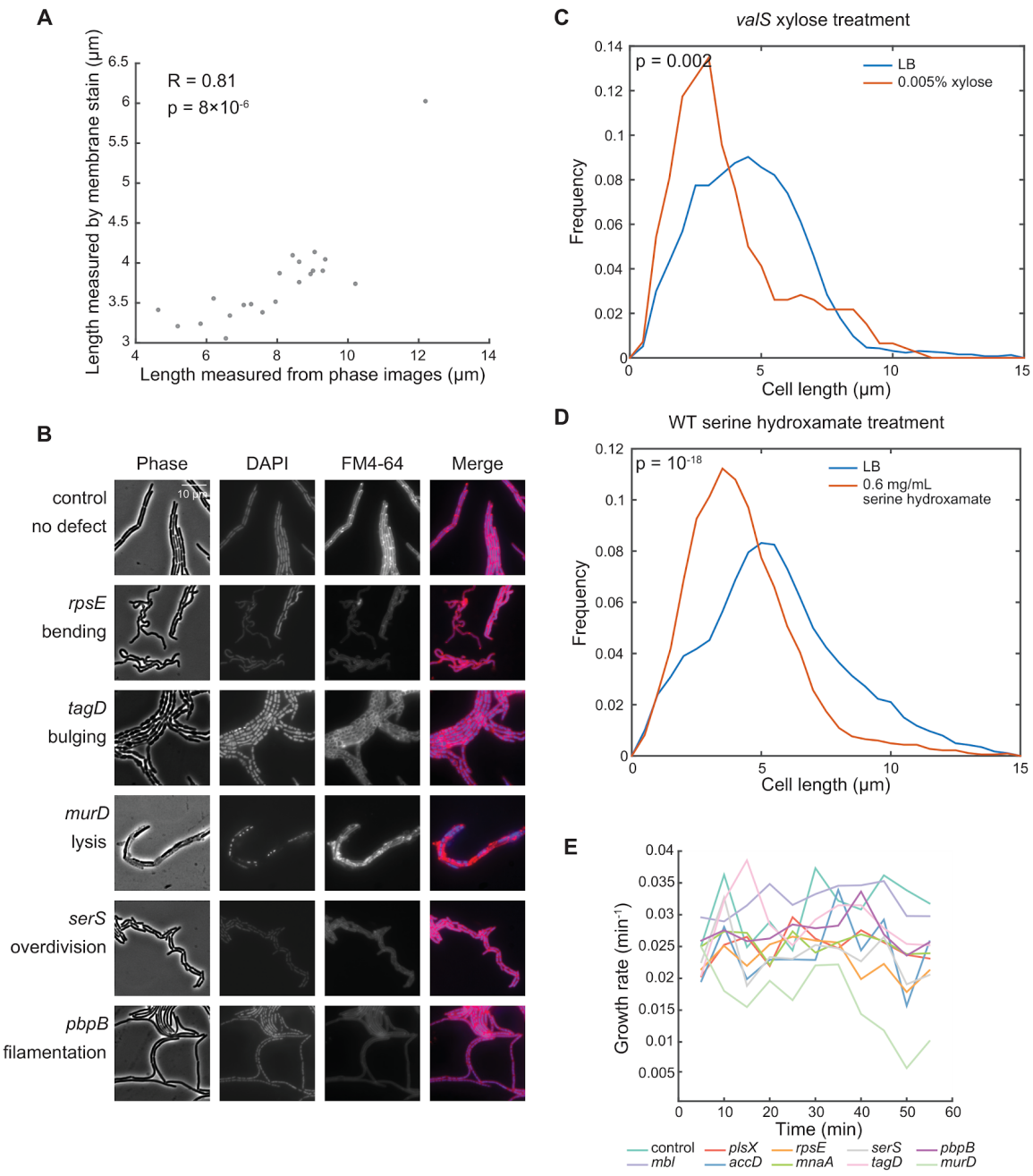


Figure 2.12. Further Characterization of Cellular Dimensions and Terminal Depletion Phenotypes, Related to Figures 2.5 and 2.6.

A) Median cell lengths for a subset of strains measured from both phase contrast images and FM4-64 staining. The two measurements are highly correlated. We note that our strains did not show as strong a propensity for chaining as other *B. subtilis* strains, and that it was possible to identify the positions of septa from phase contrast images.

B) Cells from a subset of strains with representative terminal phenotypes stained with DAPI and FM4-64 to reveal nucleoid structure and the cell membrane, respectively.

C) Distribution of cell lengths of the *valS* knockdown strain in its terminal phenotype in LB vs. LB + 0.005% xylose. Xylose-treated cells were significantly shorter and exhibited the overdivision phenotype.
D) Distribution of cell lengths of wild-type cells in its terminal phenotype in LB vs. LB + 0.6 mg/mL serine hydroxamate. Cells treated with serine hydroxamate were significantly shorter and exhibited the overdivision phenotype.
E) Post-induction instantaneous growth rates for selected strains with different terminal phenotypes. Most strains maintained constant growth rates in the first hour, except for *murD*, which decreased in growth rate after 40 min.

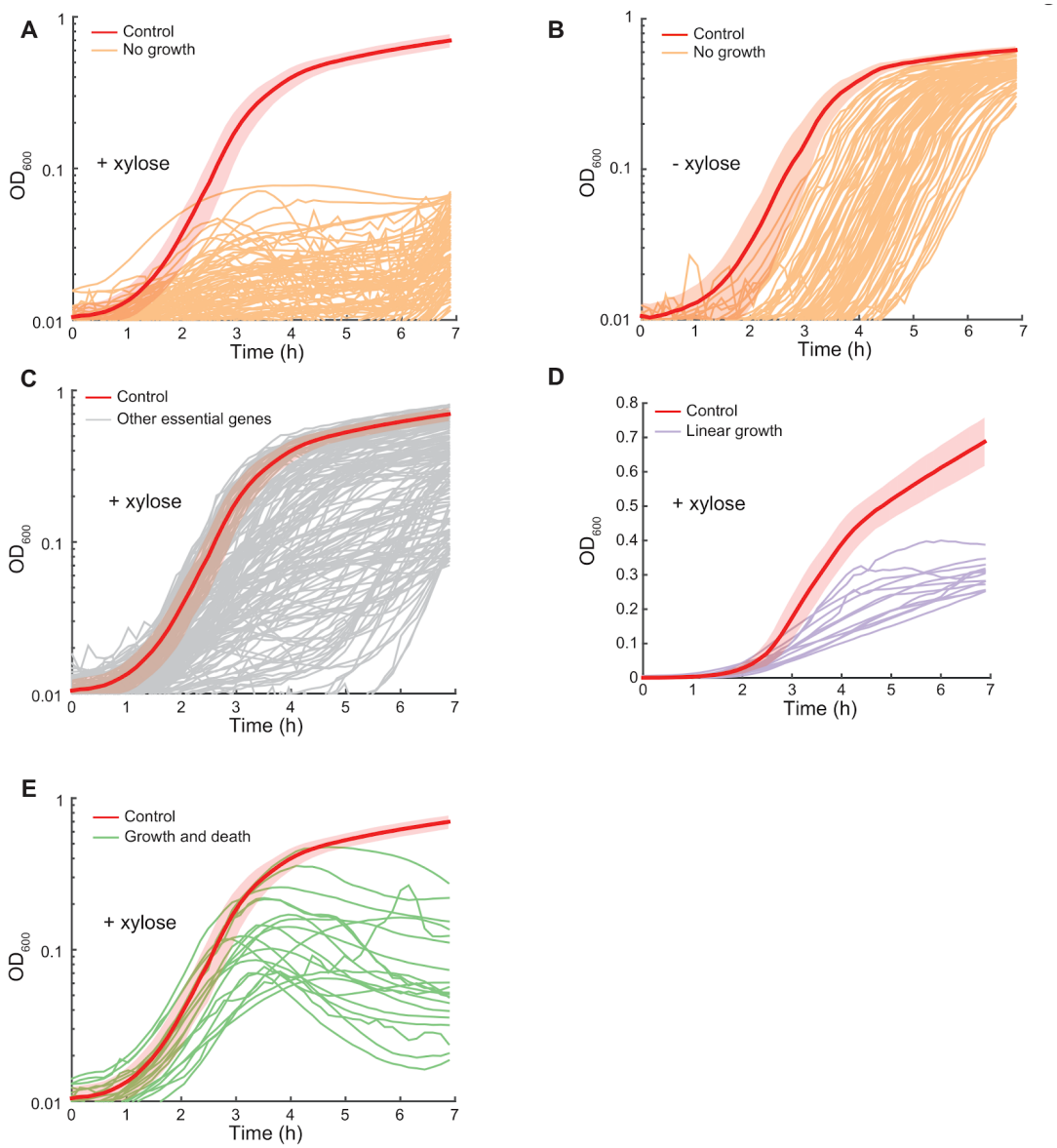


Figure 2.13. Essential Gene Depletion Growth Curves, Related to Figure 2.6.
Microplate reader growth curves of essential gene knockdown library strains at full induction of *dcas9*. Set of strains that displayed noteworthy growth patterns are plotted in each subfigure.
A) The set of strains that never emerged from stationary phase with *dcas9* induction.
B) Strains in A) displayed a range of apparent lags without *dcas9* induction, indicating that cell death during depletion is not correlated to vulnerability in stationary phase.
C) Strains that continued growing after *dcas9* induction.
D) Strains that exhibited slower, almost linear growth curves.
E) Strains that exhibited biphasic behavior, involving initial growth followed by stalling and a marked decrease in OD.

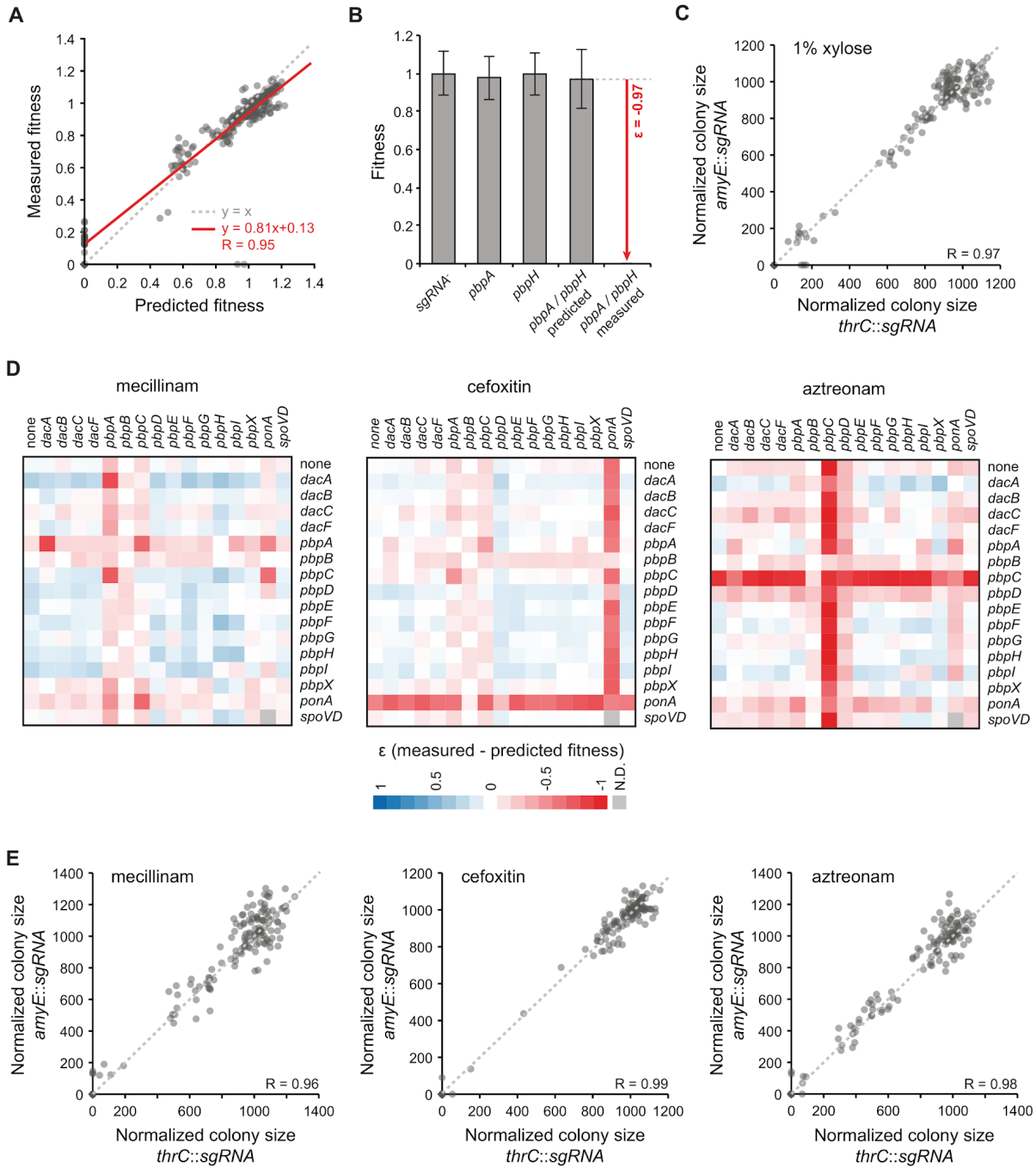


Figure 2.14. *pbp* Knockdown Specific Examples, Chemical Screens, and Reproducibility, Related to Figure 2.7.

A) Relationship between double knockdown fitness predicted by the multiplicative model and fitness measured by colony size. The grey, dashed $y = x$ line represents the ideal relationship between predicted and measured fitness according the multiplicative model; the red line is a linear fit to the data. Strains with two sgRNAs targeting the same gene were removed for clarity. B) Relative fitness defect of the *pbpB* single knockdown strain based on normalized colony size.

C) Agreement between normalized colony sizes for reciprocal double knockdown sgRNA pairs (e.g., *amyE*::*sgRNA*^{pbpA} and *thrC*::*sgRNA*^{pbpH} versus *amyE*::*sgRNA*^{pbpH} and *thrC*::*sgRNA*^{pbpA}) at maximal induction of *dcas9* (1% xylose).
D) The difference between measured and predicted fitness (ε values) for each *pbp* double knockdown based on normalized colony size at maximal induction of *dcas9* (1% xylose) and in the presence of the indicated PBP inhibitor.
E) Agreement between normalized colony sizes for reciprocal double knockdown sgRNA pairs at maximal induction of *dcas9* (1% xylose) and in the presence of the indicated PBP inhibitor.

Table 2.1. Strains, Plasmids, and Essential sgRNAs, Related to Figures 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, and 2.7.

[INCLUDED AS SEPARATE FILE]

Strains (Sheet 1) and plasmids (Sheet 2) used in this study, and sequences, annotations, and validation data for sgRNAs targeting essential genes (Sheet 3).

Table 2.2. CRISPRi Essential Gene Knockdown Chemical Screens, Related to Figures 2.2 and 2.3.

[INCLUDED AS SEPARATE FILE]

Essential knockdown screen chemical conditions (Sheet 1), chemical-gene scores (calculated from at least eight replicates; Sheet 2), sensitivities of essential knockdown strains for their cognate antibiotics (Sheet 3), and colony size data (median of at least eight replicates) from the MAC-0170636 screen.

Table 2.3. CRISPRi Essential Gene Knockdown Chemical Screen Correlations and Network, Related to Figure 2.3.

[INCLUDED AS SEPARATE FILE]

All gene-gene correlations from the essential knockdown screen (Sheet 1), significant gene-gene correlations (0.572 correlation cutoff; Sheet 2), essential operon annotations (Sheet 3), and *rpoB* knockdown RNA-seq data (Sheet 4).

Table 2.4. High-resolution Growth Profiles, Related to Figure 2.4.

[INCLUDED AS SEPARATE FILE]

Exponential growth rates and apparent lag times in basal knockdown conditions (Sheet 1), and categories of knockdown behavior during induced growth (Sheet 2).

Table 2.5. CRISPRi Essential Gene Knockdown Cell Morphologies, Related to Figures 2.5 and 2.6.
[INCLUDED AS SEPARATE FILE]

Uninduced cellular dimensions and deviation from expected cell shape based on correlation between length and width in **Figure 2.5 B** (Sheet 1), and terminal phenotypes of induced knockdown strains (Sheet 2).

Table 2.6. Multiplexed CRISPRi *pbp* Knockdowns and Chemical Screen, Related to Figure 2.7
[INCLUDED AS SEPARATE FILE]

Normalized colony sizes for *pbp* double knockdowns on xylose, xylose + mecillinam (3 μg/ml), xylose + cefoxitin (0.2 μg/ml), and xylose + aztreonam (1μg/ml) (colony sizes are normalized to the plate control and are the median of at least eight replicates; Sheet 1), measured minus predicted fitness for all *pbp* double knockdown screens (Sheet 2), and normalized colony sizes for the *pbp* triple knockdown screen (Sheet 3).

Chapter 3

Modulated efficacy CRISPRi reveals evolutionary conservation of essential gene expression-fitness relationships in bacteria

SUMMARY

Essential genes are the central hubs of cellular networks. Despite their importance, the lack of high-throughput methods for titrating their expression has limited our understanding of the fitness landscapes against which essential gene expression levels are optimized. We developed a modified CRISPRi system leveraging the predictable reduction in efficacy of imperfectly matched sgRNAs to generate specific levels of CRISPRi activity and demonstrate its broad applicability in bacteria. Using libraries of mismatched sgRNAs, we characterized the expression-fitness relationships of essential genes in Escherichia coli and Bacillus subtilis. Remarkably, these relationships co-vary by pathway and are predominantly conserved between E. coli and B. subtilis despite ~ 2 billion years of evolutionary separation, suggesting that deeply conserved tradeoffs underlie bacterial homeostasis.

Main Text

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 promoter replacement revealed gene-, environment-, and antibiotic-specific fitness effects of altering essential gene expression (Bauer, Li, and Siegal 2015; Keren et al. 2016; Dekel and Alon 2005; Eames and Kortemme 2012; Nichols et al. 2011; E. O. Johnson et al. 2019), 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. CRISPR interference (CRISPRi), which blocks bacterial transcription by targeting a catalytically dead Cas9 (dCas9) to a gene using a complementary sgRNA, has been used to perturb essential gene expression in its native context. However, tuning transcriptional repression by adjusting dCas9 or sgRNA abundance (Peters et al. 2016; Liu et al. 2017) is noisy and precludes the interrogation of multiple knockdown levels in a single experiment (Vigouroux et al. 2018). Building on previous work (Gilbert et al. 2014; Vigouroux et al. 2018), we reasoned that we could instead modulate transcriptional repression by programming a highly expressed CRISPRi system with sgRNAs imperfectly matched to their target. This would allow us to explore the fitness landscape of essential gene expression by enabling massively parallel interrogation of the fitness effects of multiple levels of CRISPRi activity across genes in a single pooled growth experiment.

We first explored how mismatches affect sgRNA activity by generating 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) (**Fig. 3.5 A**). Using FACS-seq (**Fig 3.1 A**, 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* (**Fig. 3.5 A-C, Table 3.1**). We found that sgRNAs with either single (**Fig. 3.1 B**) or double (**Fig. 3.7 A**) 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 (**Fig. 3.6 E-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, **Fig. 3.1 B. 3.7 A, and Table 3.1**), despite an evolutionary distance of several billion years and differences in experimental setup (*E. coli:* plasmid-encoded sgRNAs, *B. subtilis:* chromosomally integrated sgRNAs). This suggests that the primary determinant of CRISPRi efficacy in bacteria is the interaction between the dCas9-sgRNA complex and DNA, rather than organism-specific factors such as the host's transcriptional machinery.

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. Informed by previous work on CRISPRi off-target effects (Qi et al. 2013; Gilbert et al. 2014) and concurrent work on mismatched sgRNAs in a mammalian

context (Jost et al. 2019), we constructed a linear model using the position and base substitution of the mismatch and the GC% of the fully complementary spacer as features. We trained this model on the *E. coli, B. subtilis* (**Fig. 3.8, 3.9**), or species-averaged relative efficacy of our *gfp*-targeting singly mismatched sgRNAs (**Fig. 3.1 C**) and found that the effects of single mismatches could be robustly predicted in all cases (species-averaged R^2 = 0.56, 11-fold CV-MSE = 0.10 $^{+}/_{-}$ 0.08). Assuming that mismatches have independent effects on sgRNA efficacy, the model also accurately predicted double mismatch efficacy (R^2 = 0.53, **Fig. 3.7 B**). To further validate this model, we compared our predicted sgRNA activity to the previously measured association rates (k_{on}) of a dCas9-sgRNA complex to 60 singly mismatched and 1130 doubly mismatched DNA sequences (Boyle et al. 2017) (**Fig. 3.1 D, 3.3 C**). Our predicted efficacy 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. Consistent with this idea, our model recapitulates many biophysical properties of RNA-DNA interactions such as the relative stability of rG:dT basepairs implied by its high coefficient for A to G transitions (Sugimoto, Yasumatsu, and Fujimoto 1997) (**Table 3.2 and Fig. 3.6 D**). 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.

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 3.3**) and predicted to have a range of activity. 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, **Fig. 3.5 C**) and compact libraries in which each essential gene is targeted by 11 sgRNAs (*See* ***Note 3.1***). 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, **Fig. 3.5 B**). The libraries were grown for 10 doublings, maintaining exponential phase through back-dilution (**Fig. 3.2 A**). We calculated the relative fitness (Rest et al. 2013; Kampmann, Bassik, and Weissman 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 3.3**). Relative fitness is defined as the fraction of doublings a strain undergoes compared to wild-type over the 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 ($R^2 > 0.9$, **Fig. 3.10 A-B**), was validated by orthogonal measurements of individual strain fitness (**Fig. 3.10 C**), and was consistent within fully complementary sgRNAs targeting the same gene (Sup. Text 1). Our relative fitness values were correlated with previously reported measurements (Tianmin Wang et al. 2018; Rousset et al. 2018) but had greatly expanded dynamic

range due to greater sequencing depth and a shorter growth period (**Fig. 3.10 D-E**, Sup. Text 2). This expanded dynamic range enabled 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) caused negative relative fitness (**Table 3.4**). Consistent with an interpretation of negative relative fitness as lysis, a majority (15/24) of these *B. subtilis* genes caused lysis when targeted with a fully complementary sgRNA (Peters et al. 2016) (**Table 3.4**, Methods).

We next assessed whether comparing the predicted activity of sgRNAs to their relative fitness would allow us to infer per gene expression-fitness relationships. First, we asked whether predicted sgRNA activity was inversely correlated to relative fitness across our data set. As expected, we found strong negative correlations both within sgRNA families (sgRNAs targeting the same locus) and within genes (**Fig. 3.11**, Methods). Weaker correlations in *E. coli* likely reflect variation in sgRNA plasmid copy number and/or *E. coli* specific effects (Cui et al. 2018). Second, we examined the expression-fitness relationships of *murA/murAA* and *folA/dfrA* using comprehensive mismatched sgRNA libraries. Consistent with previous studies (22-24), we find that CRISPRi targeting of *murA/murAA* bimodally affects fitness (**Fig. 3.2 B-C**), while CRISPRi targeting of *folA/dfrA* linearly affects growth rate above an initial threshold of activity (**Fig. 3.2 D-E**). Highlighting the approximately linear effect of *folA/dfrA* repression on fitness, our model could accurately predict *gfp* knockdown after being trained on the fitness effects of mismatched sgRNAs targeting those genes (**Fig. 3.8, 3.9**). Third, we measured the ability of 18 mismatched sgRNAs to repress a *murAA-gfp* transcriptional fusion in *B.*

subtilis. To enable quantification of lethal levels of knockdown and to minimize transcriptional feedback these measurements were conducted in a *B. subtilis* strain complemented with non-targeted *murAA* (Methods). The predicted activity of sgRNAs targeting *murAA* in this experiment closely tracked actual knockdown (**Fig. 3.2 F**), suggesting that the non-linear expression-fitness relationship of *murAA* without complementation (**Fig. 3.2 B**) reflect non-linearly decreasing growth due to MurAA depletion, transcriptional feedback, cell lysis, or other host specific effects. Finally, we confirmed that low efficacy (relative fitness > 0.95) sgRNAs targeting *dfrA* were functional by measuring the fitness of the *B. subtilis dfrA* library in the presence of trimethoprim, a direct inhibitor of DfrA. Trimethoprim decreased the fitness of low efficacy sgRNAs targeting *dfrA* suggesting that these sgRNAs impact *dfrA* expression even in the absence of a measurable fitness defect (**Fig. 3.2 G**). Taken together, these validation experiments 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.

Examining the essential gene expression-fitness relationships, we were struck by their diverse and gene-specific nature (**Fig. 3.12, 3.13, Table 2.3**). 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, **Fig. 3.2 B-E, Table 3.5**). 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 3.6**) 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. Taken together, these data suggest that these expression-fitness curves are both biologically meaningful and representative of deeply conserved homeostatic constraints on bacterial physiology.

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 (**Fig. 3.3 A-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”, **Fig. 3.3 C-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 (Schaechter, Maaloe, and Kjeldgaard 1958; Scott et al. 2010; Borkowski et al. 2016). 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 (**Fig. 3.12, 3.13**), likely due to slowed elongation rate (Dai et al. 2016) as has been shown for some translation inhibitors (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 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, peptidoglycan synthesis genes exhibited bimodal fitness outcomes that depended on

predicted sgRNA activity (**Fig. 3.3 E-F**). Cells tolerated partial repression of these genes without exhibiting a fitness defect, perhaps due to transcriptional feedback and/or an excess of enzyme. If expression was sufficiently repressed, these strains lysed (**Table 3.4**) 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)*. However, the dearth of intermediate fitness outcomes upon repression of peptidoglycan precursor synthesis in both species is surprising. It suggests that neither species is able to slow growth rate in response to reduced flux through cytoplasmic peptidoglycan precursor synthesis 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 peptidoglycan precursor synthesis and may explain the sharp transition between growth and lysis.

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 illustrate biologically relevant differences between the two organisms. We identified 9 homologs as significantly different between the two organisms (**Table 3.7**, FDR < 0.2), most of which encoded enzymes involved in peptidoglycan synthesis and maturation. In contrast to the conserved bimodal expression-fitness relationships of genes involved in cytoplasmic peptidoglycan precursor synthesis just discussed (**Fig. 3.4**, group 3), CRISPRi targeting of genes required for producing either UDP-GlcNAc (**Fig. 3.4**, group 1) or *meso*-DAP (**Fig. 3.4**,

group 2) differentially affected fitness in the two species. *E. coli* was robust to CRISPRi targeting of these genes, while *B. subtilis* was sensitive, lysing when these genes were targeted with high activity sgRNAs (**Fig. 3.4, Table 3.4, 3.7**). The differential effect of CRISPRi targeting on these genes could be attributed to buffering of either expression or activity in *E. coli* but not in *B. subtilis*, perhaps mediated by divergent regulatory mechanisms (Rodionov et al. 2003; Barreteau et al. 2008). Differences in the fitness effect of DAP pathway knockdowns may also be accounted for by peptidoglycan stem-peptide recycling through the *E. coli* enzyme Mpl (J. W. Johnson, Fisher, and Mobashery 2013). *E. coli* was also significantly more tolerant of perturbation of *mreBCD* than *B. subtilis* (**Fig. 3.4**, group 4), exhibiting a minimal fitness defect after 10 generations (but lysis after 15 generations, **Table 3.3**). This observation is consistent with the small effect of CRISPRi targeting of *mrdA* (the PBP2 associated with MreBCD) on fitness in *E. coli* (**Fig. 3.8**), and with previous work which found that *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 may play a role. Alternatively, the substantially higher turgor pressure in *B. subtilis (Osawa and Erickson 2018)* may make it less tolerant of cell wall abnormalities.

Singly mismatched CRISPRi is a universal approach for systematically perturbing bacterial gene expression. Leveraging this technique, we explore the expression-fitness relationships of essential genes in *E. coli* and *B. subtilis* and reveal that the basic biological constraints driving essential gene fitness landscapes are conserved over >2

billion years of evolution. 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.

Funding

MRS was supported by the U.S. National Science Foundation through T32 GM007810a and a Graduate Research Fellowship. JSW is a Howard Hughes Medical Institute Investigator. This work was supported in part by the National Institutes of Health (F32 GM108222 and K22AI137122 to JMP; F32 GM116331 and K99 GM130964 to MJ; P50 GM102706, U01 CA168370, R01 DA036858, and RM1 HG009490 to JSW; R35 GM118061 to CAG) and the Innovative Genomics Institute, UC Berkeley (CAG).

Competing interests

JSW and MJ have filed patent applications related to CRISPRi/a screening and mismatched sgRNAs in eukaryotic systems. JSW consults for and holds equity in KSQ Therapeutics, Maze Therapeutics, and Tenaya Therapeutics. JSW is a venture partner at 5AM Ventures. MJ consults for Maze Therapeutics.

Data and materials availability

All sequencing data have been deposited in the Short Read Archive with BioProject accession PRJNA574461. All analyzed data is available in the main text or the supplementary materials. Unpublished code used in this manuscript is available on GitHub (https://github.com/traeki/mismatch_crispri and https://github.com/traeki/sgrna_design).

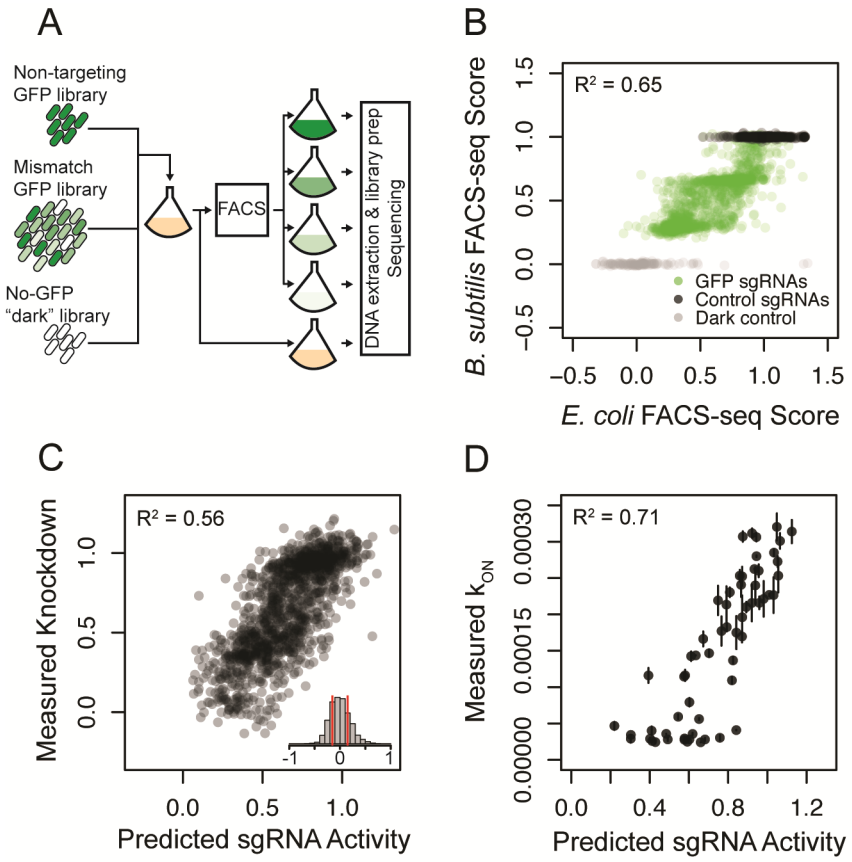


Figure 3.1

Singly mismatched sgRNAs reproducibly generate a range of knockdown efficacies in *B. subtilis* and *E. coli* and are accurately predicted by a simple linear model. (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) 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). (D) The predictions of the linear model compared to the measured singly mismatched sgRNA association rates (k_{ON}) in vitro (Boyle et al. 2017).

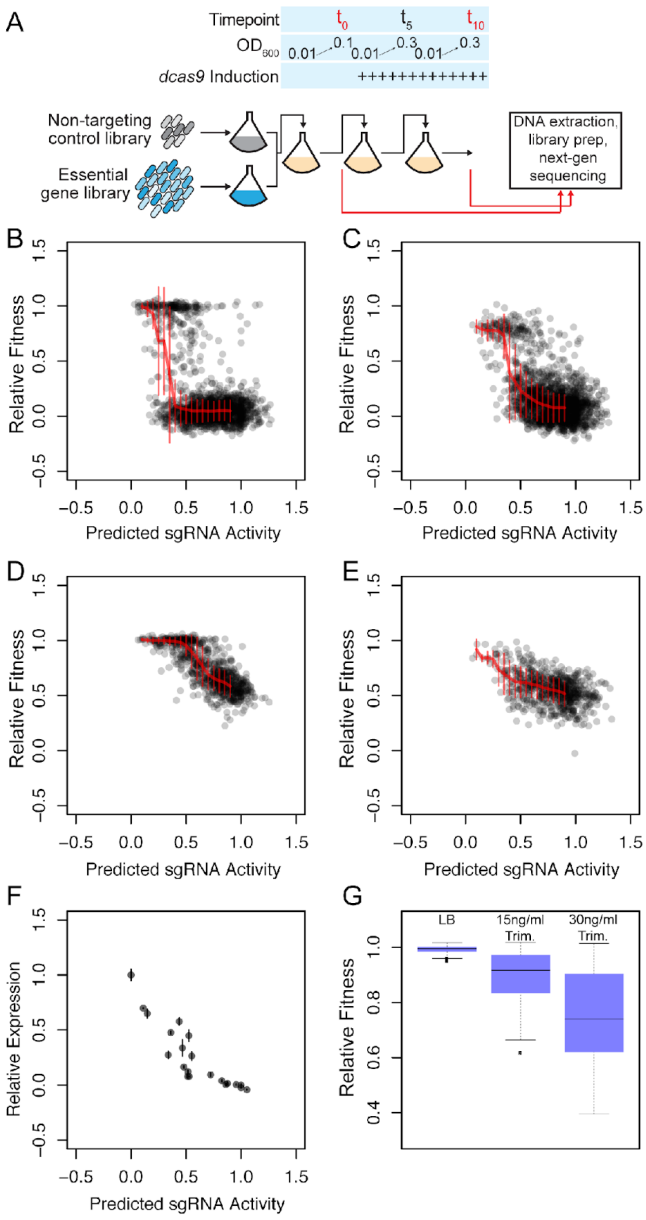


Figure 3.2

Singly mismatched sgRNAs targeting two essential genes in *E. coli* and *B. subtilis* illustrate the gene-specific nature of expression-fitness relationships. (A) Schematic of the fitness experiment design. (B-E) Predicted sgRNA activity and measured relative fitness of singly mismatched sgRNA targeting (B) *murAA* in *B. subtilis*. (C) *murA* in *E. coli*. (D) *dfrA* in *B. subtilis*. (E) *folA* in *E. coli*. Median and SD values of relative fitness for sgRNAs grouped into 17 bins based on predicted knockdown are shown in red. (F) 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). (G) Measured relative fitness of singly mismatched sgRNAs targeting *dfrA* in *B. subtilis* with relative fitness > 0.95 in LB

(n=252). Measured fitness values of the same sgRNAs in the presence of sub-inhibitory concentrations of trimethoprim.

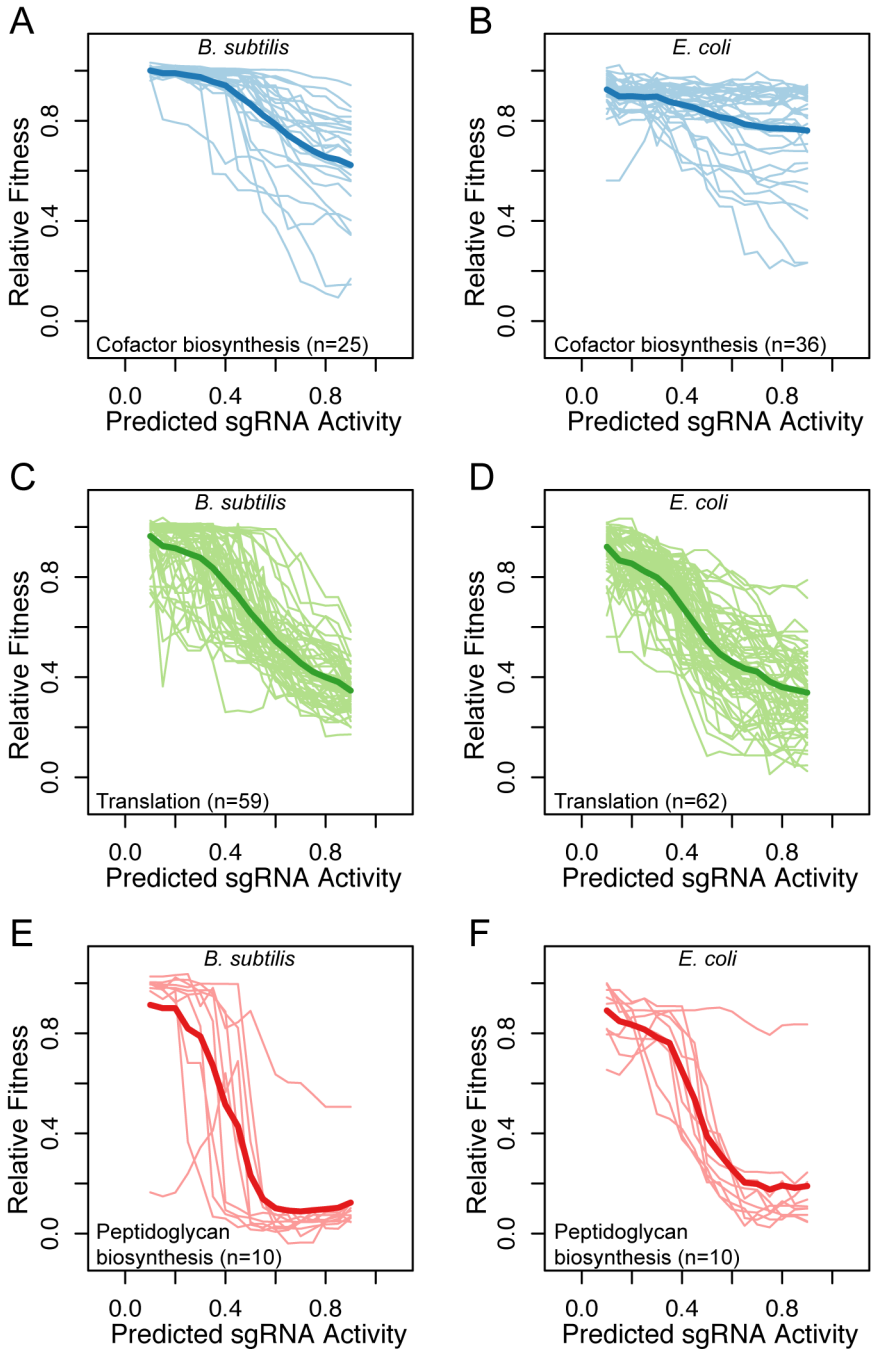


Figure 3.3
Expression-fitness relationships of essential genes are conserved within biological process and between *B. subtilis and E. coli.* Relative fitness compared to predicted knockdown for: essential cofactor biosynthesis genes (KEGG pathways under "Metabolism of cofactors and vitamins") in *B. 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).

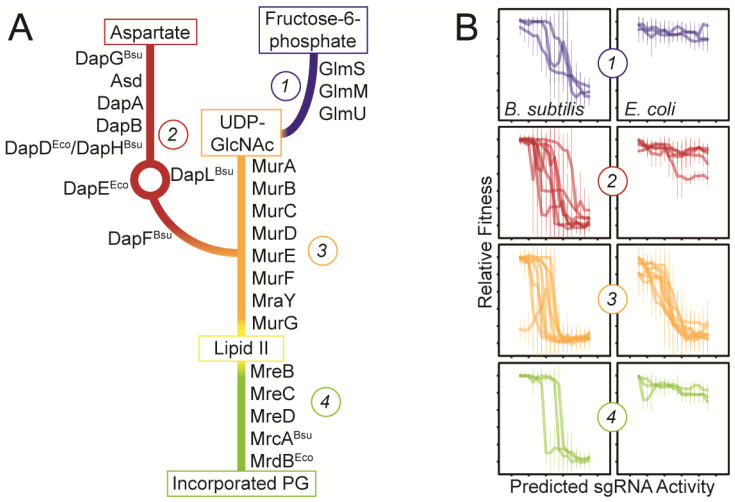


Figure 3.4
Similar and different expression-fitness relationships of cell wall biosynthesis genes in *B. subtilis* and *E. coli*. (A) Pathway of peptidoglycan synthesis and incorporation, color coded by portion of the pathway. (B) Predicted knockdown vs. relative fitness for the groups of essential genes from pathway sections indicated in (A), in *B. subtilis* and *E. coli*.

Note 3.1

To facilitate the exploration of essential gene requirements in diverse conditions, we also constructed smaller (11 sgRNA/gene) libraries that can be easily manipulated, screened, and multiplexed. These libraries generate a broad range of phenotypes for most genes (**Fig. 3.14**) in B. subtilis and E. coli.

Materials and Methods

Materials

Bacillus subtilis strain construction and growth conditions

All primers are described in **Table 3.8**, and all strains are described in **Table 3.9**.

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 3.8**) 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).

Escherichia coli strain construction and growth conditions

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 transduction (Thomason, Costantino, and Court 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 in this case the *dcas9* is expressed by the minimal synthetic promoter P*BBa_J23105* {https://parts.igem.org}, and transconjugates were selected using gentamicin and chloramphenicol.

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).

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 3.8**). The oligonucleotide pools were amplified by 15 cycles of PCR using Q5 polymerase (New England Biolabs #M0493S) and custom primers (**Table 3.8**). 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 3.3**).

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 3.3**). 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 3.3**) from V1 of the library. The V2 *E. coli* libraries included a majority of genes with evidence for essentiality (**Table 3.3**) (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 (**Fig. 3.5**). 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 {https://github.com/traeki/mismatch_crispri, 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 (**Fig. 3.5**). 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 (**Fig. 3.5**).

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.

Methods

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 (**Fig. 3.6 A**). The trimethoprim experiments were carried out in an identical manner, except that both 1% xylose and trimethoprim (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 3.8**), followed by gel-purification from 8% TBE gels (Invitrogen #EC62152BOX), and sequencing on HiSeq 4000 with single-end 50bp reads using a custom sequencing primer (UCSF Center for Advanced Technology).

Relative fitness analysis

Raw FASTQ files were aligned to the library oligos and counted using {https://github.com/traeki/mismatch_crispri, count_guides.py}, and relative fitness was calculated using {https://github.com/traeki/mismatch_crispri, 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 3.3**).

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 3.3**.

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") (**Fig. 3.5 A**). 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 5th 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 3.1**.

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 (**Fig. 3.6 E-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 (**Fig. 3.6 E-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 3.1** and plotted in **Fig. 3.6 A**.

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 a sgRNA, after first subtracting the background GFP fluorescence from a non-fluorescent *B. subtilis* strain. Relative GFP fluorescence measurements are provided in **Table 3.10**.

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:

$$knockdown_{sgRNA} = 1 - FACS.seq\ score$$

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

$$sgRNA\ activity_{singly\ mismatched\ sgRNA} = \frac{knockdown_{\ singly\ mismatched\ sgRNA}}{knockdown_{\ fully\ complementary\ sgRNA}}$$

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 3.2** and **Fig. 3.14**, the raw data in **Table 3.1**.

Expression-fitness relationship analysis

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 (**Fig. 3.15** and supplementary text 1), 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 a 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 sgRNA and per gene patterns are shown in **Fig. 3.12** (*E. coli*) and **Fig. 3.13** (*B. subtilis*). Per gene bin medians for essential genes can be found in **Table 3.5**.

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

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 3.6**.

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 = 155, 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 155 genes (i.e. when one gene from one species is compared to the 154 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.

Supplementary Text

1. 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. To do so, we compared the total sum of squares (totalSS) of fully complementary sgRNAs to within gene the sum of squares (withinSS). In *E. coli*, the withinSS accounted for 26.7% of the totalSS and in *B. subtilis* the withinSS accounted for 18.6% of the totals. 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.

2. 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 (**Fig. 3.10 D-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.

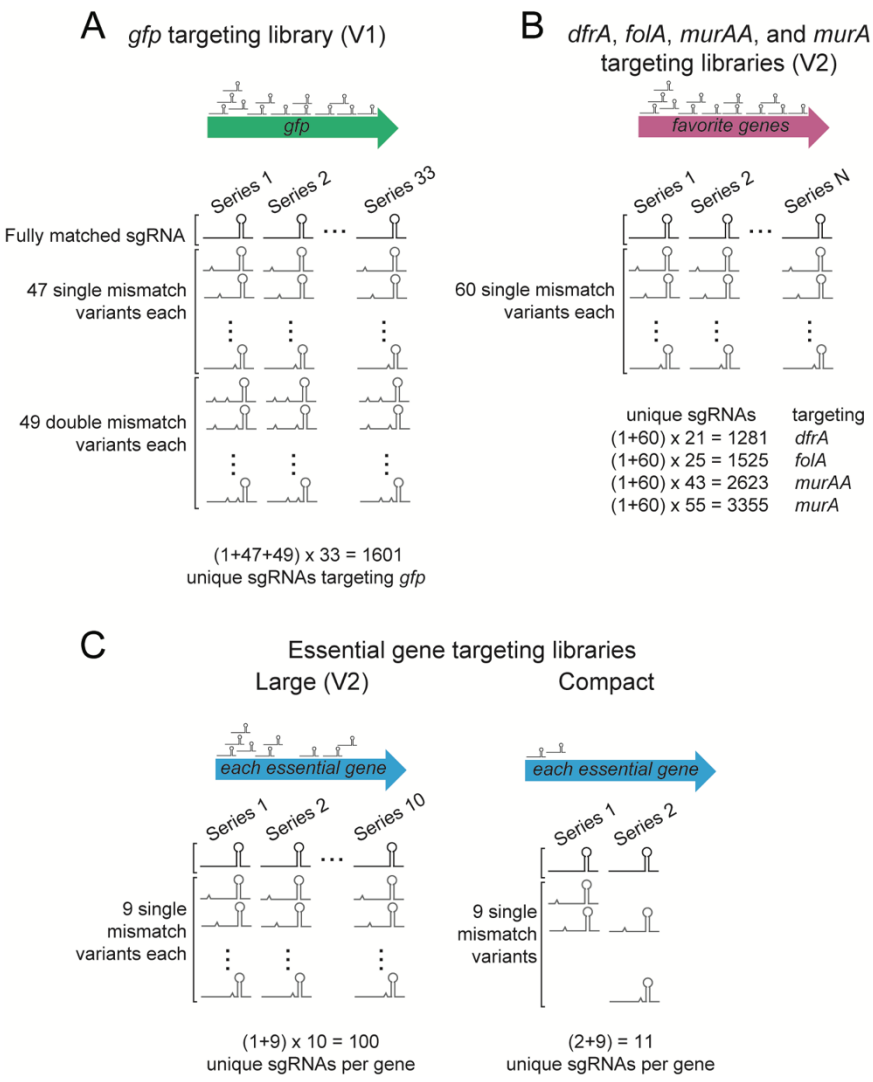


Figure 3.5
Design of mismatched sgRNA libraries targeting (A) *gfp*, (B) comprehensive libraries targeting *dfrA*, *folA*, *murAA*, and *murA*, and (C) each essential gene for the large libraries and the compact libraries. The breakdown of single mismatch variants per series and the total unique sgRNAs per gene are shown for each.

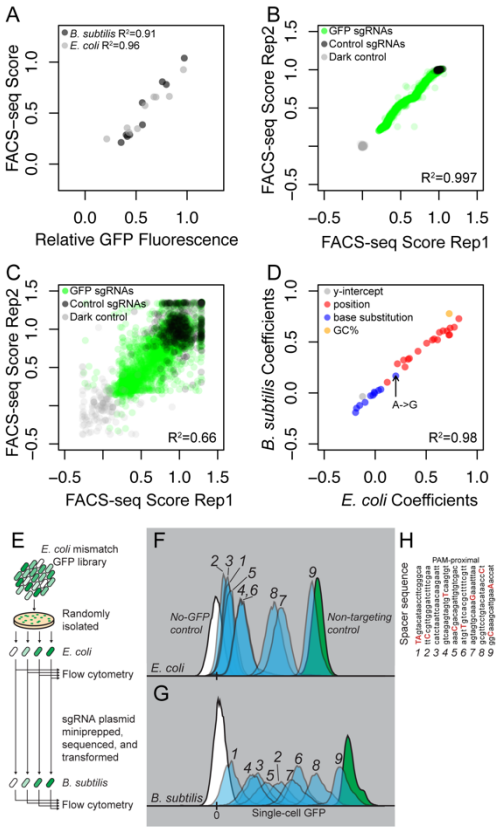


Figure 3.6
Details related to the linear model, FACS-seq data, and its validation. (A) Mismatched sgRNA efficacy measured individually (relative GFP fluorescence) in either *E. coli* or *B. subtilis* compared to their FACS-seq score from measurements done in the same species. Relative fluorescence is 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) Coefficients of each variable from linear models trained on FACS-seq data from either *E. coli* or *B. subtilis*. The A to G transition discussed in the main text is highlighted. y_int is the y-intercept; GC is the GC% of the fully complementary sgRNA spacer. (E) 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. (F-G) The distribution of single-cell GFP fluorescence values for strains of *E. coli* (G) or *B. subtilis* (F). (H) The sequences of the spacers indicated in (F) and (G), with mismatched bases highlighted in red.

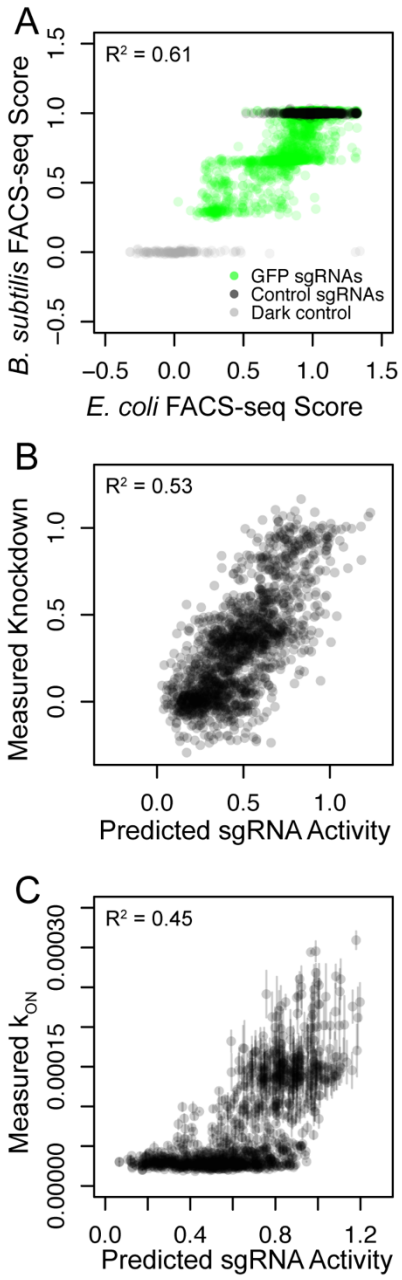


Figure 3.7
Doubly mismatched sgRNAs are accurately predicted as the combined independent effects of singly mismatched sgRNAs. (A) FACS-seq enrichment scores (average of 2 biological replicates) for each doubly mismatched sgRNA targeting *gfp* in *B. subtilis* and *E. coli*. (B) The predictions of the 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 (*14*).

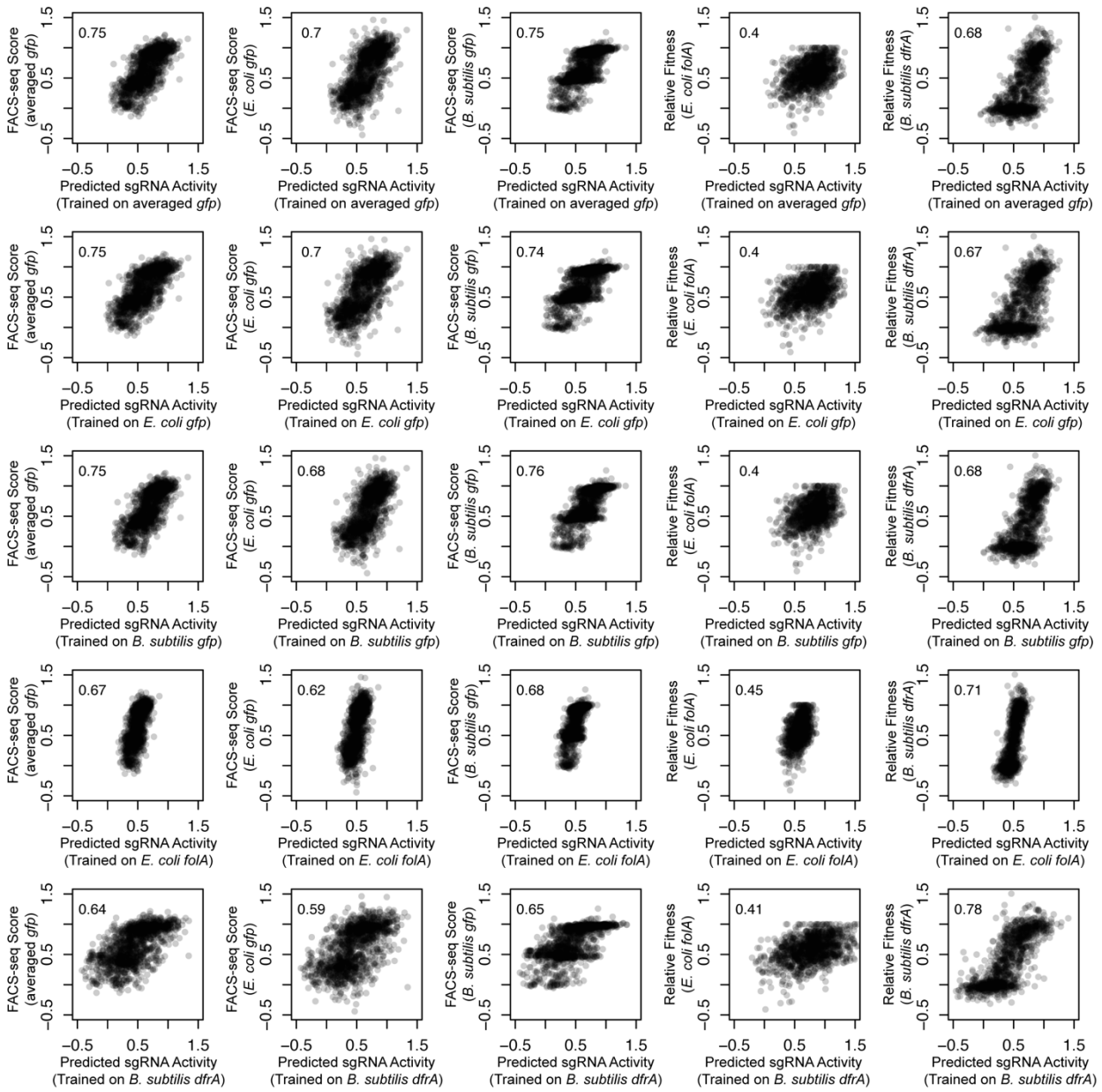


Figure 3.8
Linear models of singly mismatched sgRNA efficacy trained on FACS-seq or relative fitness data from *E. coli, B. subtilis,* or the average of both (averaged *gfp*) retain a majority of their predictive power on other singly mismatched sgRNA datasets. Each panel compares the predictions of 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*.

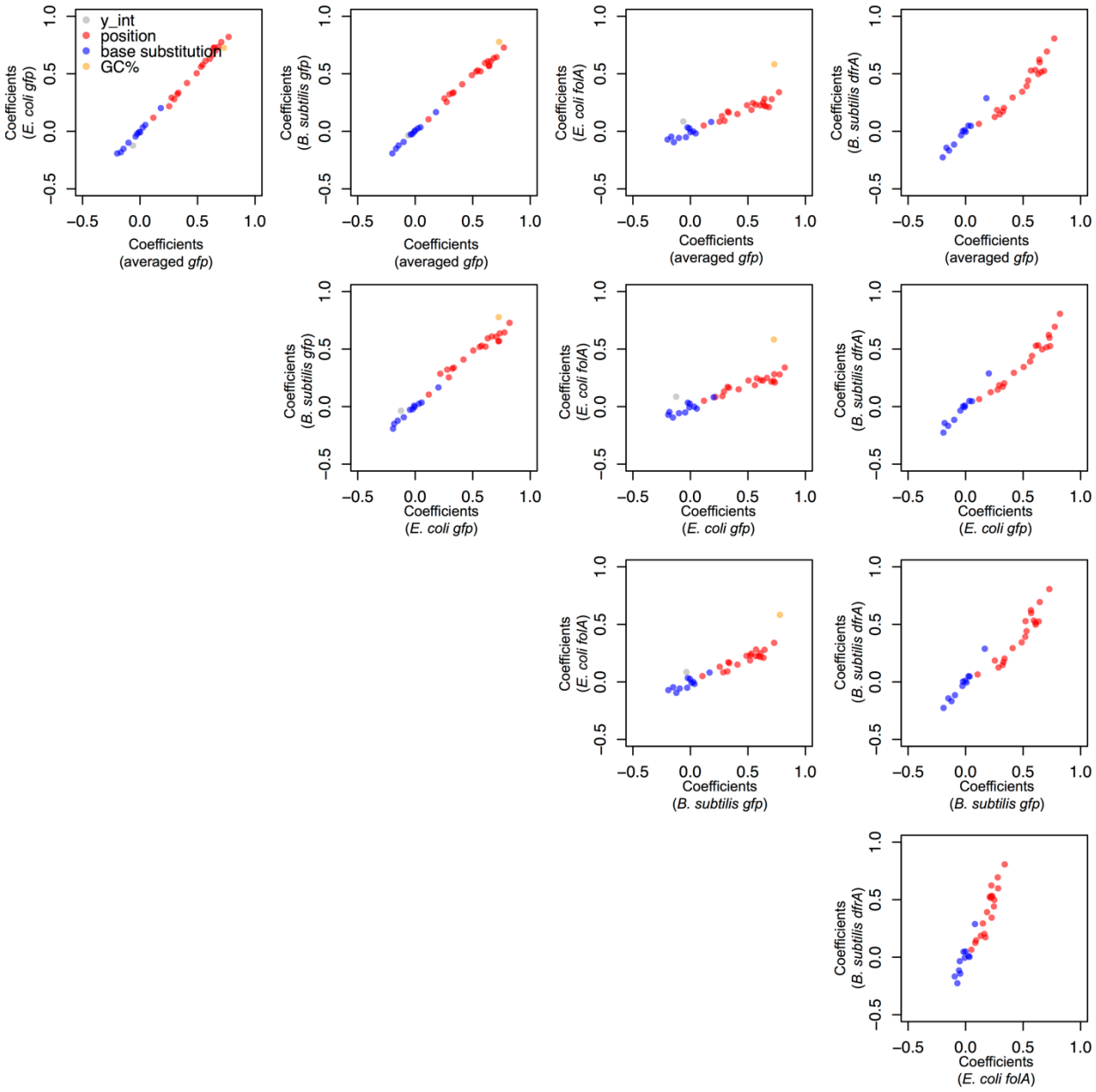


Figure 3.9
Parameters of linear models of singly mismatched sgRNA efficacy trained on FACS-seq or relative fitness data from either *E. coli* or *B. subtilis* have strongly correlated coefficient values. Each panel compares the coefficient values from the linear models trained on the two specified datasets.

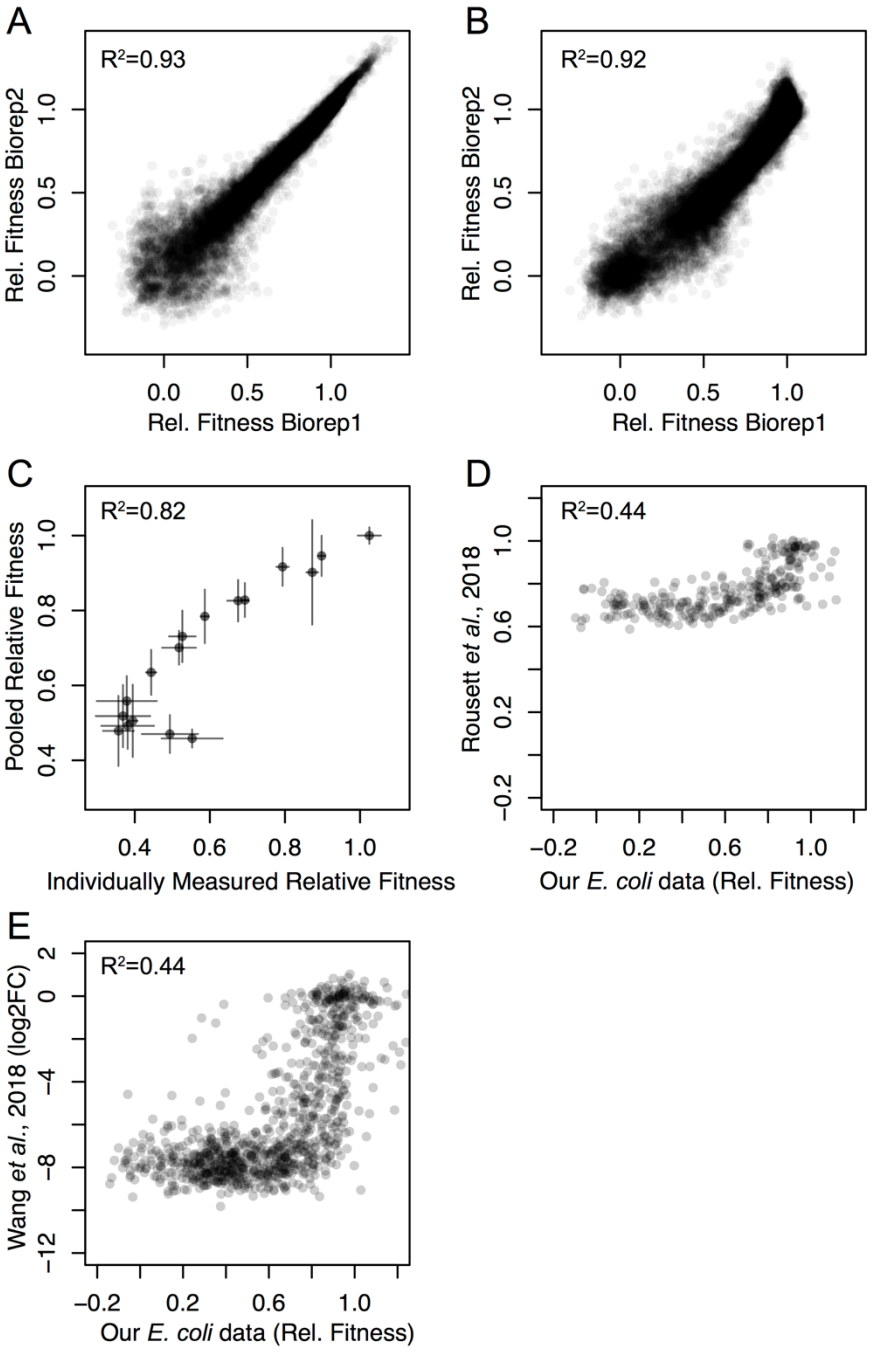


Figure 3.10
Relative fitness measurements are reproducible, orthogonally validated, and capture a large dynamic range. (A-B) Relative fitness measurements from two biological replicates in *E. coli* (A), and *B. subtilis* (B). (C) Relative fitness from pooled experiment compared to a relative fitness metric from competing individual *dfrA*-targetomg 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 (*19*) and (*18*) 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 (*18*), 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 (*19*), 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. See Supplementary Text 2 for further discussion.

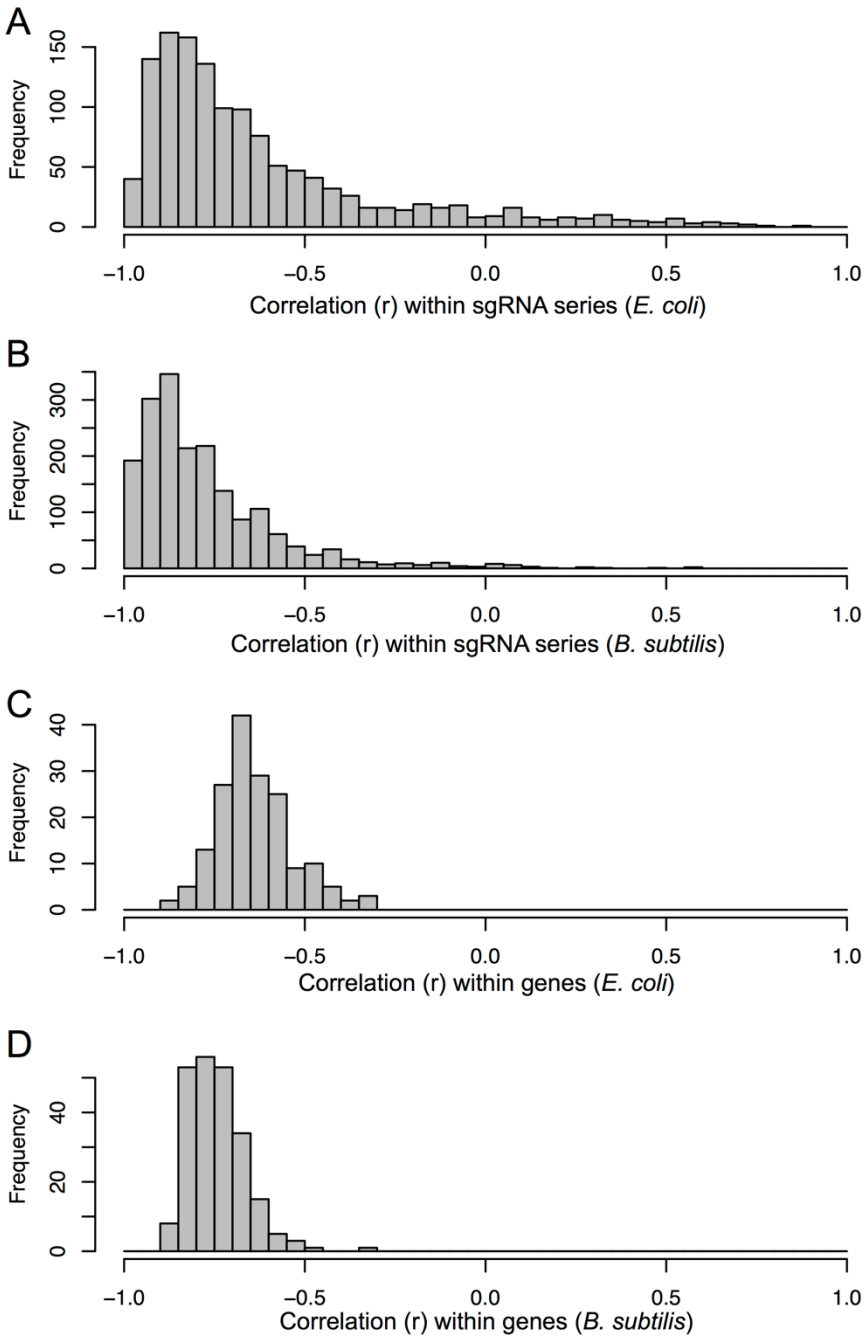


Figure 3.11
Singly mismatched sgRNAs targeting essential genes are strongly and negatively correlated within sgRNA series and within genes. (A-B) Distribution of per sgRNA series correlations (r) for sgRNAs targeting genes in *E. coli* (A) and *B. subtilis* (B). (C-D) Distribution of per gene correlations (r) for sgRNAs targeting genes in *E. coli* (C) and *B. subtilis* (D).

Figure 3.12. All *E. coli* per gene knockdown-fitness curves
[INCLUDED AS A SEPARATE FILE]

Figure 3.13. All *B. subtilis* per gene knockdown-fitness curves
[INCLUDED AS A SEPARATE FILE]

Table 3.1
[INCLUDED AS A SEPARATE FILE]

Table 3.2
[INCLUDED AS A SEPARATE FILE]

Table 3.3
[INCLUDED AS A SEPARATE FILE]

Table 3.4
[INCLUDED AS A SEPARATE FILE]

Table 3.5
[INCLUDED AS A SEPARATE FILE]

Table 3.6
[INCLUDED AS A SEPARATE FILE]

Table 3.7
[INCLUDED AS A SEPARATE FILE]

Table 3.8
[INCLUDED AS A SEPARATE FILE]

Table 3.9
[INCLUDED AS A SEPARATE FILE]

Table 3.10
[INCLUDED AS A SEPARATE FILE]

Appendix #1

Validation

To validate our pooled-growth fitness measurements, we assessed the growth rate of individual strains and compared it to the measured growth rate in the pooled experiment. We chose 20 of the *dfrA* sgRNAs exhibiting a diverse set of fitness outcomes for individual testing. Each experimental strain was constructed by transforming the appropriate *dfrA* sgRNA into a parental strain containing dCas9 and additionally an RFP or GFP marker (two replicates were used, one of each color where available, two technical replicates of a single color otherwise --see **Table 4.1**). Control strains contained dCas9 and a fluorescent marker but lacked a guide RNA.

The growth protocol was similar to that used for pooled growth experiments. Each experimental strain and control strains were grown individually to exponential phase, at which point each experimental strain was mixed 1:1 with the control strain containing the opposite fluorescent marker. Each pair was back-diluted for growth to start the experiment and again at the 5-doubling midpoint as in the pooled-growth experiment. Formaldehyde-fixed samples taken at the start of the experiment and after 10 doublings were counted by flow cytometry to determine the ratio of the experimental to the control strain at each timepoint. We computed relative fitness by comparing the ratio at t=0 to that after 10 doublings of growth.

We saw strong agreement between our pooled-growth measurements and the gold-standard measurements (**Figure 4.2**). The pooled-growth measurements gave higher fitness values overall, which we attribute to differences in experimental conditions (e.g. growth in deep-well plates in the individual measurements vs large-volume flasks for the

pooled growth experiments). The pooled-growth measurements also had greater noise at lower relative fitness, which is an expected consequence of distributing resolution (i.e. read depth) over 30,000 strains instead of 2. See Appendix #2 for more discussion of this latter issue.

Table 4.1 -- fluorophore(s) used, by variant

variant sequence	flouorophore
AGATCATTGGGCAAATGCCA	both
AGATCATTGGGCTAATGCTA	both
GTCTTTGCCGATAAGCCTGT	both
GTCTTTGCCGATAAGCCTCT	both
AGATCAATGGGCAAATGCCA	both
AGAGCAATGGGCAAATGCCA	both
AGATCATTGGGCATATGCCA	both
AGACCATCGGGCAAATGCCA	both
GGATCATTGGGCAAATGCCA	both
AGATCATTGGGCAAAGGCCA	both
AAATCATTAGGCAAATGCCA	gfp
AGATCATCGGGCAAATGGCA	gfp
AGATCATTGGGCAAATTCCA	red
AGATCATTGGGCAAATGCCC	gfp
GTCTGTGCCGATAAGCCTGT	both
GTGTTTGCCGATAAGCCTGT	both
GTCTTTGCCGACAAGCCTGT	both
GTCTTTGCCGATAAGCTTGT	both
GTCTTGTCCGATAAGCCTGT	both
GTCTTTGCCGATCAGCCTGT	both
GTCTTTGCCGATAAGCATGT	red
GTCTTTGCCTATAAGCCTGT	red
GTCTTTGCCGATAAGCCTGG	red
GTCTTTGCCGATAAGGCTGG	red

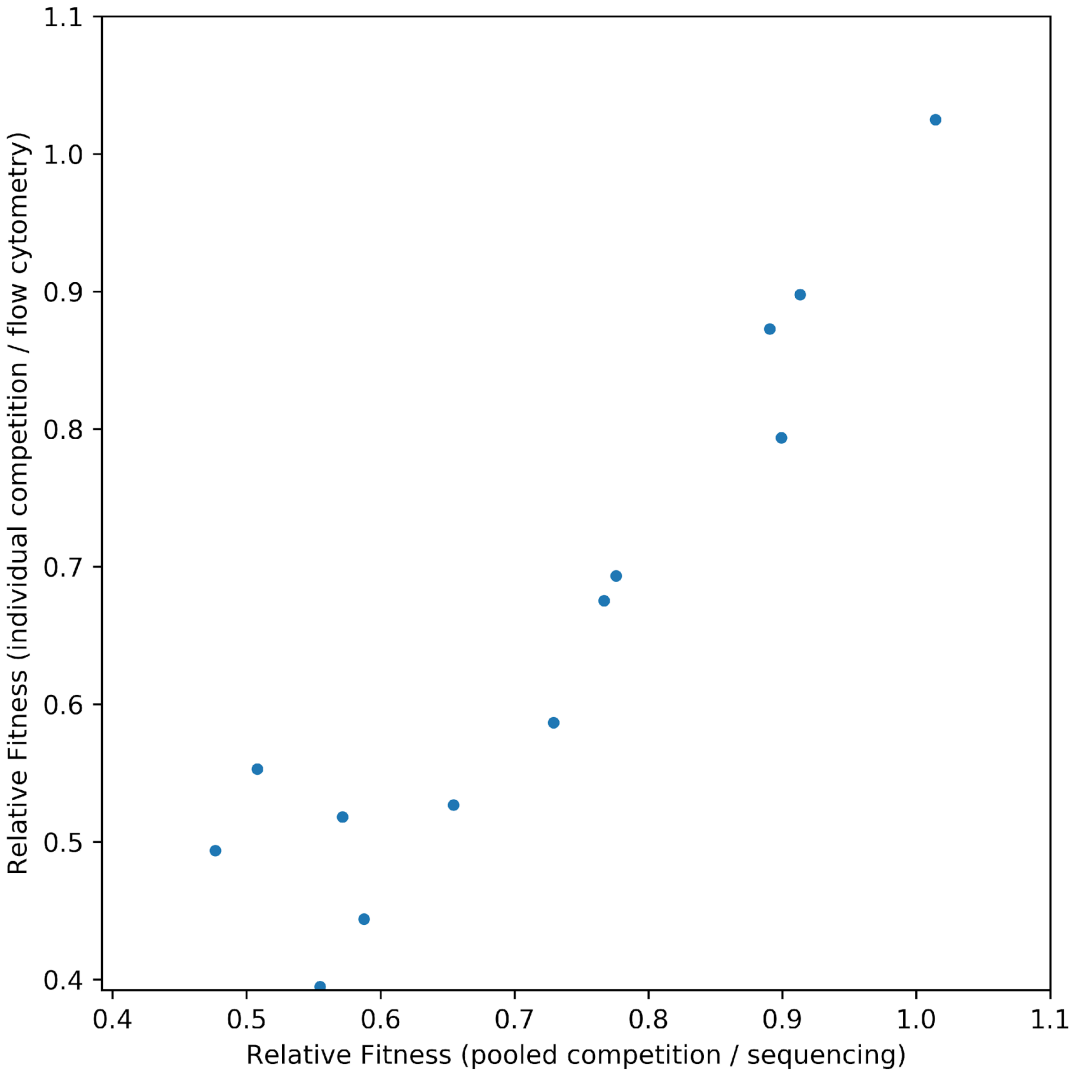


Figure 4.2
Comparison of relative fitness measurements from pooled growth experiments and individual competition experiments for mismatched dfrA guides.

Fitness of strains with mismatched, variable efficacy sgRNA guides targeting *dfrA* was computed by using deep sequencing at the beginning and end of the experiment to determine changes in their fractional representation in a large strain library containing mismatched guides against every B. subtilis essential genes at t=0 and t=10 doublings (x-axis). Individual strains containing these same sgRNAs against *dfrA* and a GFP/RFP fluorescent marker were grown in competition with a strain with the other fluorescent marker but lacking the guide. The fractional representation of each guide strain vs the control at 0 and 10 doublings was determined by flow-cytometry. Relative fitness values were computed from the change in fractional representation (y-axis).

Appendix #2

Evaluating Success

Experimental design is critical for extracting maximal information about strain phenotypes. We calculate growth rate of each strain relative to the wildtype growth rate by using deep sequencing to compare the fractional representation of that strain at the start of dcas9 induction (t=0) and after 10 wildtype doublings (t=10). Since there is a 2^10 = 1000-fold increase in wildtype cells during 10 doublings, cells that do not grow will be diluted out and therefore experience a 1000-fold decrease in abundance relative to wildtype. To observe a no-growth phenotype therefore requires the ability to detect a decrease of this magnitude. The feasibility of achieving this depends on the complexity of the library, and the depth of sequencing.

We can conservatively expect to get 3 x 10^7 sequencing reads per sample. Our main pooled-growth experiments each comprise around 3 x 10^4 distinct strains. If our library were perfectly uniform at the start of our growth measurement, each guide would have a starting read count of 10^3 -- just enough to detect a relative fitness of 0. But no library is perfectly uniform in actual practice. Some strains will be heavily over-represented in the starting pool, and some will be altogether absent. Therefore, the number of strains is not, by itself, adequate to specify library complexity.

Incorporating this non-uniformity into our reasoning, we therefore introduce a third factor directly tied to our desired outcome. We declare a minimum starting count of 100 for our strains, below which we cease to consider a strain part of the pool for relative fitness calculation in a given sample. This ensures that we will correctly measure relative fitness values at least as extreme as 0.35. We can then work backwards from this cutoff

to determine an *effective library complexity* -- i.e. the size of sample needed to capture the diversity *relevant to our final measurement*. 100 reads out of a 3 x 10^7-read depth represents a frequency of 1/(3 x 10^5). This tells us that our *effective library complexity* is 3 x 10^5 elements. Any step with *fewer* than 3 x 10^5 elements therefore represents a bottleneck, almost certainly causing loss of diversity even if the step is itself otherwise perfectly uniform. (For example, if we collect only 1 x 10^5 transformants during the transformation step of library construction, elements representing only 1/(3 x 10^5) will likely be lost from the pool entirely.)

To minimize information loss, we worked to maintain a coverage of strains at 10-fold over threshold at every step by ensuring a minimum of 3 x 10^6 events (e.g. transformants collected, cells in starting inoculum, ...). There are three ways to think about information retention and evaluate our success in this regard. First, is the distribution of sub-population frequencies consistent through the process of constructing and preparing our library? Second, what fraction of guides are empirically retained above the designated cutoff to the start of our experiments? And finally are the resulting computations replicable?

We explored these questions with our guide libraries for *B. subtilis*.

To address the first question, we sequenced guides to interrogate numerical diversity at three stages: after ligating synthesized oligos into a plasmid library, after transforming plasmids carrying genomically integrating sgRNA constructs into a strain library, and after recovering strains to exponential phase from frozen aliquots to begin an experiment. Comparing the first two (**Figure 5.1**) or the latter two (**Figure 5.2**) samples, we saw that the representation of most guides did not change. The greater

discrepancies seen in 5.1 (vs. 5.2) highlight the importance of understanding -- and where possible widening -- bottlenecks. In constructing our library, the most difficult place to maintain high numbers of entities/events was in the transition from plasmid library to strain library. Transformation and, especially, plating/recovery of individual colonies, is extremely resource intensive and difficult to scale. We were able, but just barely, to maintain our desired 10-fold coverage through these transitions.

The second question can be answered by simply observing which guides are above the cutoff. Overall, the vast majority of guides (83%) were retained above cutoff to the start of our experiment (i.e. to the right of the red-line demarcation in **Figure 5.3**). More importantly, on a per gene basis, we saw retention of a vast majority of guides (median of 84/100) with even the worst-case gene close to 50% retention (**Figure 5.4**). Finally, our cutoffs were sufficiently high, and our bottlenecks sufficiently wide, to allow high reproducibility in our pooled-growth fitness measurements. This was true for *B. subtilis* (representative replicates compared in **Figure 5.5**, R = .95) and for *E. coli* (representative replicates compared in **Figure 5.6**, R=0.96).

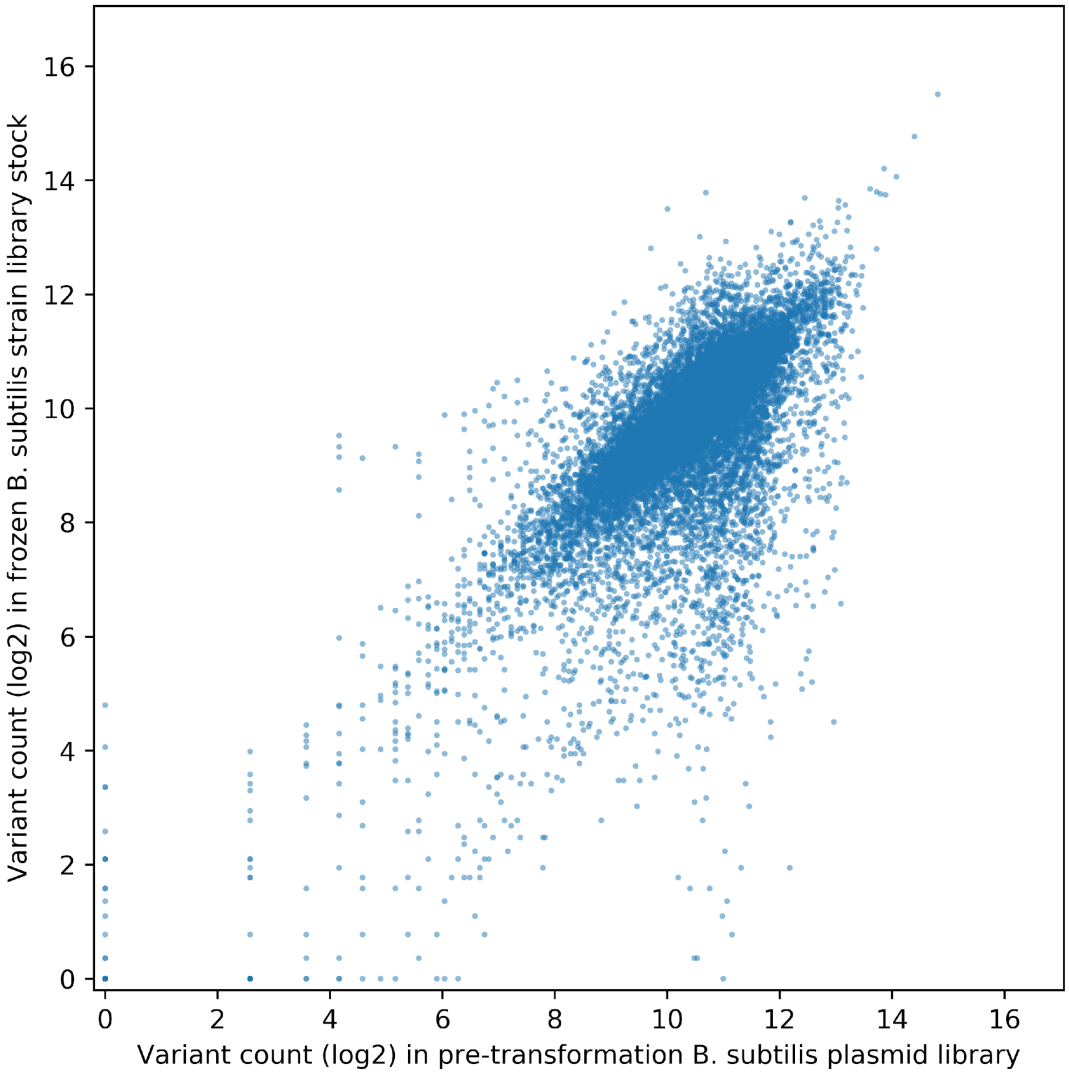


Figure 5.1
Information preservation between plasmid library and strain library

sgRNA guide loci were amplified from *B. subtilis* plasmid library DNA (x-axis) or genomic DNA of frozen *B. subtilis* strain library stocks (y-axis) and counted via deep sequencing.

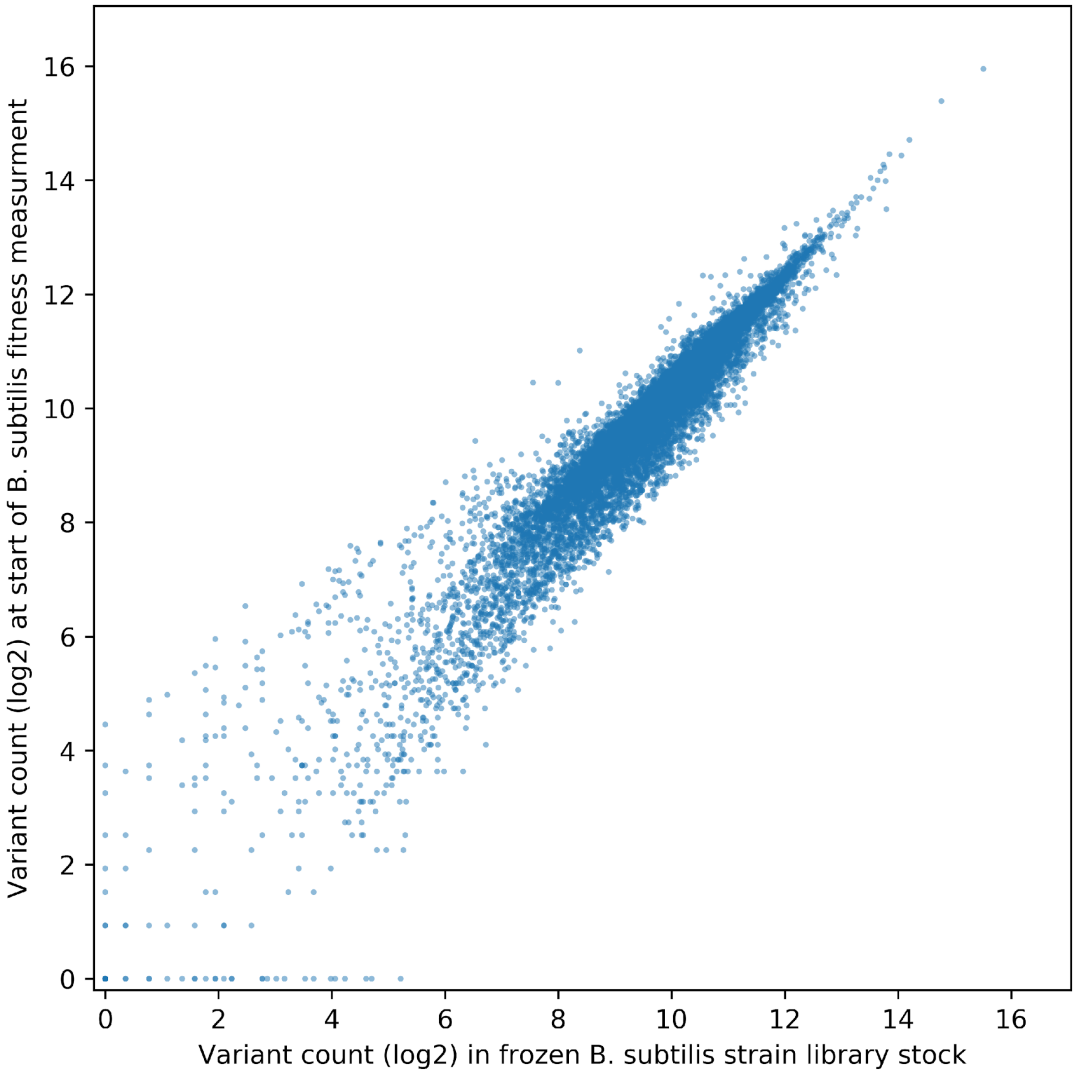


Figure 5.2
Information preservation between strain library construction and start of growth experiment

sgRNA guide loci were amplified from genomic DNA of frozen *B. subtilis* strain library stock (x-axis) or post-recovery samples at the outset of growth experiment (y-axis) and counted via deep sequencing.

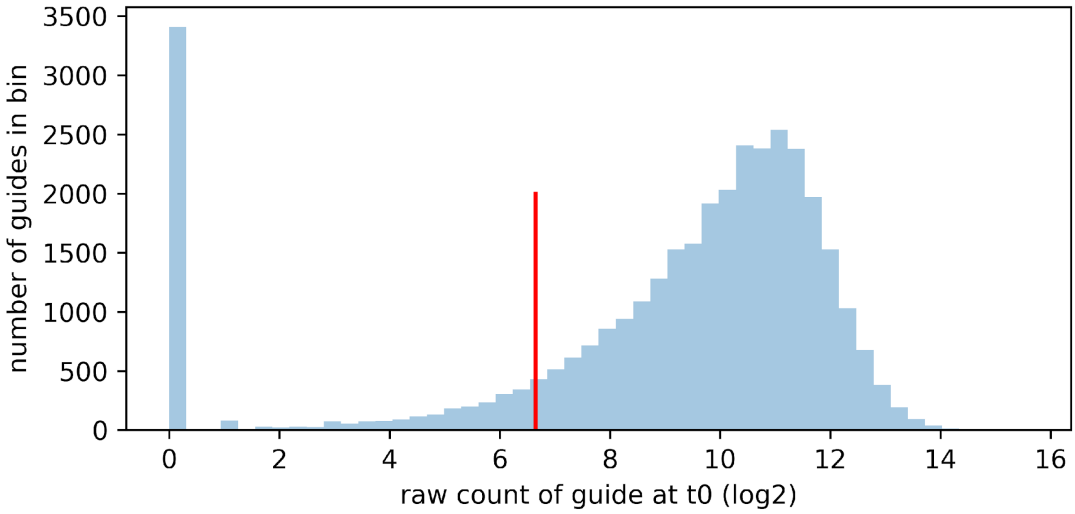


Figure 5.3
Overall guide retention above cutoff

Unnormalized guide counts from sequencing a single t0 sample of *B. subtilis* guides (~65M total reads). Red line highlights cutoff at log2(100).

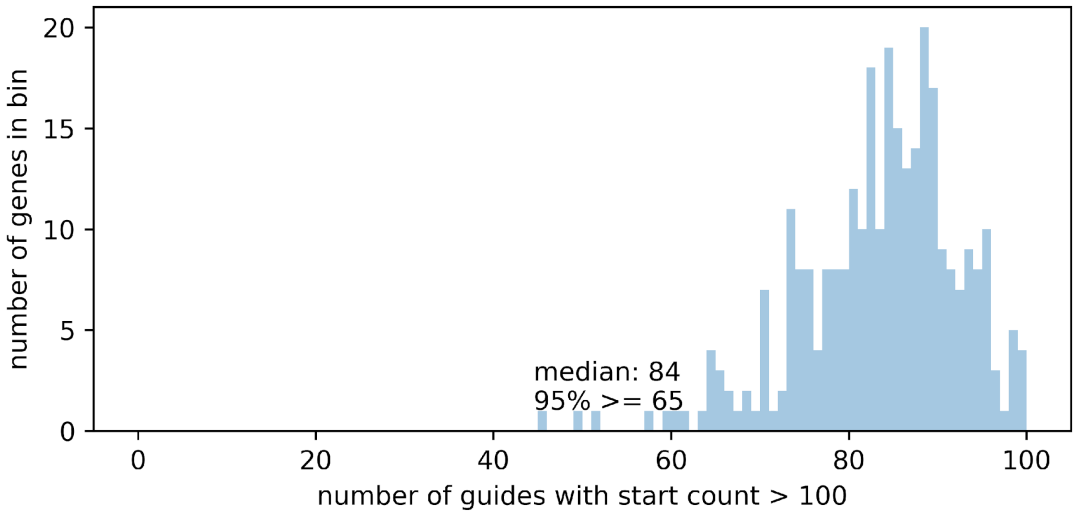


Figure 5.4
Per-gene guide retention above cutoff

Genes binned by number of guides remaining after filtering sub-cutoff guide counts as shown in **Figure 5.1**.

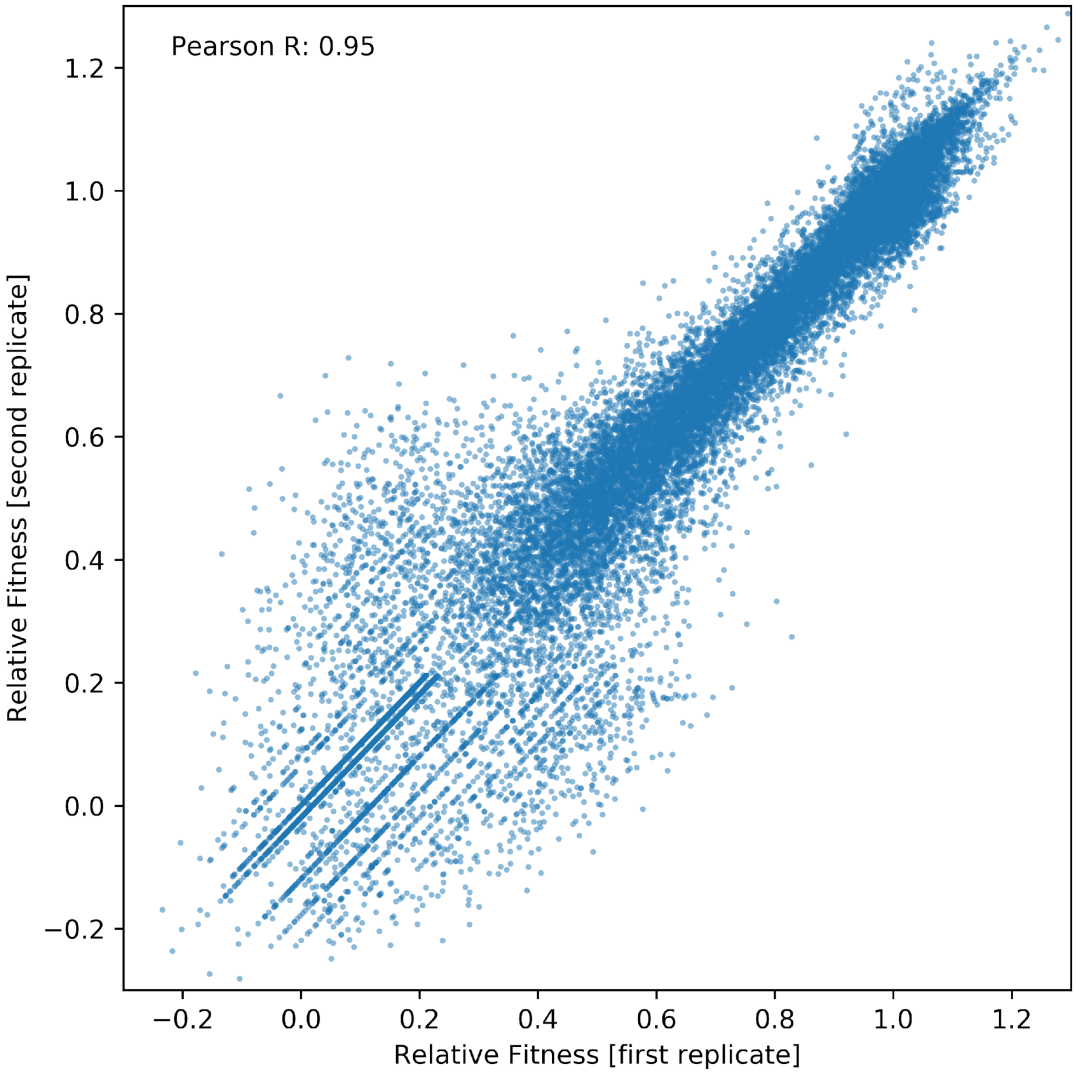


Figure 5.5
B. subtilis replicates

Representative comparison of fitness measurements for two growth replicates (*B. subtilis*).

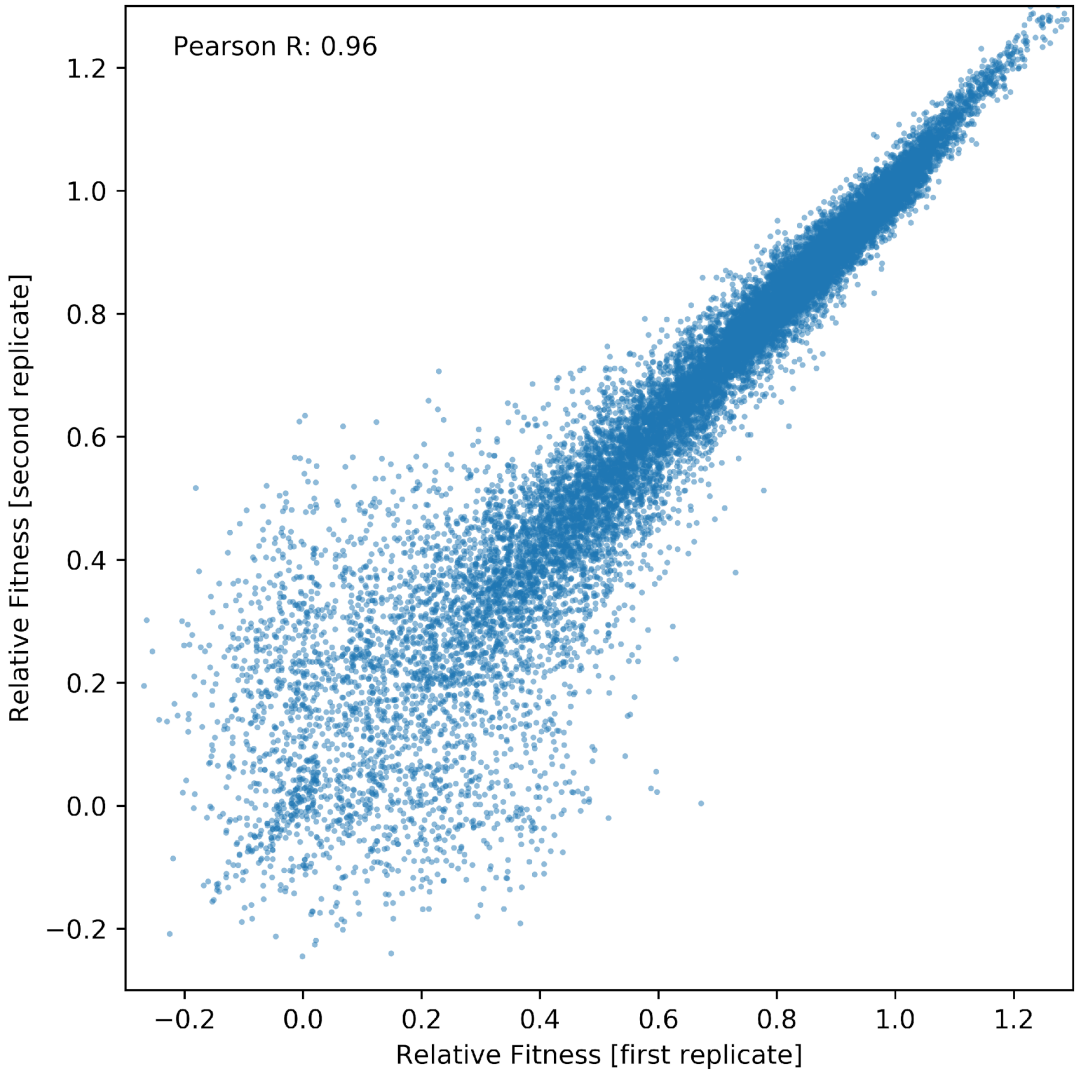


Figure 5.6
E. coli replicates

Representative comparison of fitness measurements for two growth replicates (*E. coli*).

Appendix #3

Trimethoprim Experiments

Measuring relative fitness during pooled competitive growth of singly mismatched CRISPRi guides can tell us a great deal about the importance of a gene. The meaning of this metric is quite clear so long as a strain is both healthy enough to remain entirely in the observed pool for the duration of an experiment and unhealthy enough to fall in fractional representation in time to distinguish it from the random walk of control replicates. Appendix 2 frames the considerations necessary to account for the former concern -- capturing extremely low fitness. Here we consider the latter -- guides/strains with fitness scores not readily distinguished from background.

Many guides in our singly mismatched CRISPRi experiments did not impart phenotypes outside the variance of the internal controls (strains with non-targeting sgRNAs). The guides in question may actually be non-functional, failing to alter transcription of the target gene. Alternatively, these guides may decrease expression of the targeted gene but not affect fitness either because the gene product in question is normally expressed at a slight excess or because the cell has a buffering mechanism to counteract downregulation.

To determine whether these guides actually decrease expression, we took advantage of the known synergy between the drug trimethoprim and its direct target DfrA, established in Chapter 2. We performed pooled competitive growth experiments on our full library, including our large collection of *dfrA* guides, in the presence of two sub-MIC doses of trimethoprim (15 and 30 ng/ml). The higher dose minimally, but detectably, impacts the overall growth of our library as measured by overall optical density (**Figure 6.1**).

Because 10 ng/mL was insufficient to perturb overall growth, we selected a slightly higher dose of 15 ng/mL for our "low" dosage. Our measurements of the effects of these concentrations of trimethoprim on our entire library, **excluding those encoding the folate pathway**, indicate that strain growth was not *differentially* affected at these concentrations of trimethoprim (**Figure 6.2**). In contrast, strains with *dfrA* guides (**Fig 6.3**) were severely affected, and growth of other genes in the pathway was affected less severely (**Fig 6.5-6.10**). This result strongly suggests that, without drug, the "no-phenotype" *dfrA* guides were knocking down gene expression, although they do not affect growth.

Exploring this issue further reveals that for the higher of the two sub-MIC doses (30 ng/mL trimethoprim), those *dfrA* strains with reduced growth in the absence of trimethoprim were predominantly pushed into a regime of such low fitness that it is difficult to distinguish one from another, while those *dfrA* strains whose growth was indistinguishable from the wt cover nearly the full range of *dfrA* phenotypes in the presence of trimethoprim (**Figure 6.4**). 15ng/ml trimethoprim had similar but less severe effects. Guides like these will be useful in other experiments, where we perturb the strains with a panel of secondary stressors to uncover phenotypes, and will provide a wealth of data about regulation and function of the essential genes.

Our trimethoprim experiment captures other expected phenotypes for important secondary targets in the folate pathway (**Figure 6.5-6.10**). The one member of the fol family in our set which is not essential, folD, shows a slightly different phenotype than the others, with the lower dose phenotypes closely resembling those for the no-drug case, but showing striking synergy at the higher dose. A small number of genes outside

the folate pathway also appear to have slight synergy phenotypes in the presence of trimethoprim, such as aroE -- known to be peripherally associated with folate pathway (Babitzke, Gollnick, and Yanofsky 1992) (**Figure 6.11**), sufB and fmt -- two genes related to iron metabolism (**Figure 6.12, 6.13**), and the as yet poorly characterized reductase ylbQ (**Figure 6.14**). These phenotypes were detectable for the parent guides, but a consistent shift across a series of knockdowns makes detection of subtle phenotypes an easier proposition. We believe these guide series libraries consequently stand to make screening for unknown interactions with secondary stressors much more sensitive.

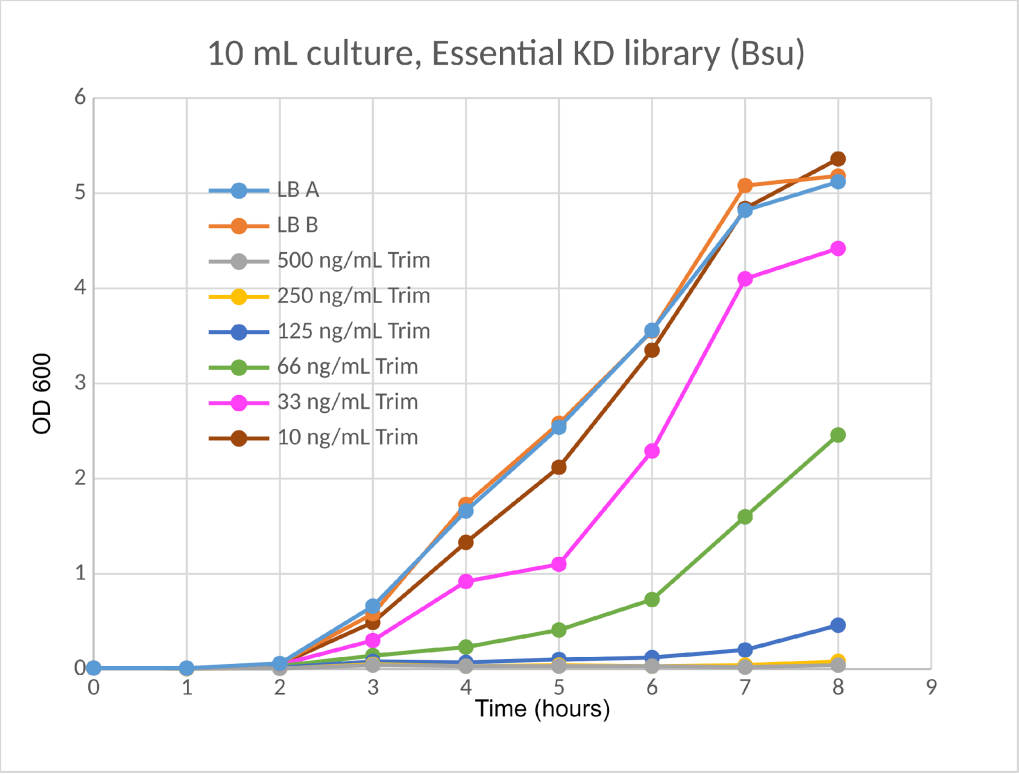


Figure 6.1
Selection of trimethoprim doses for growth experiments.

B. subtilis library was grown with a range of trimethoprim doses, tracking OD 600. 10 ng/mL trimethoprim did not substantially impact growth of library, whereas 33 ng/mL did.

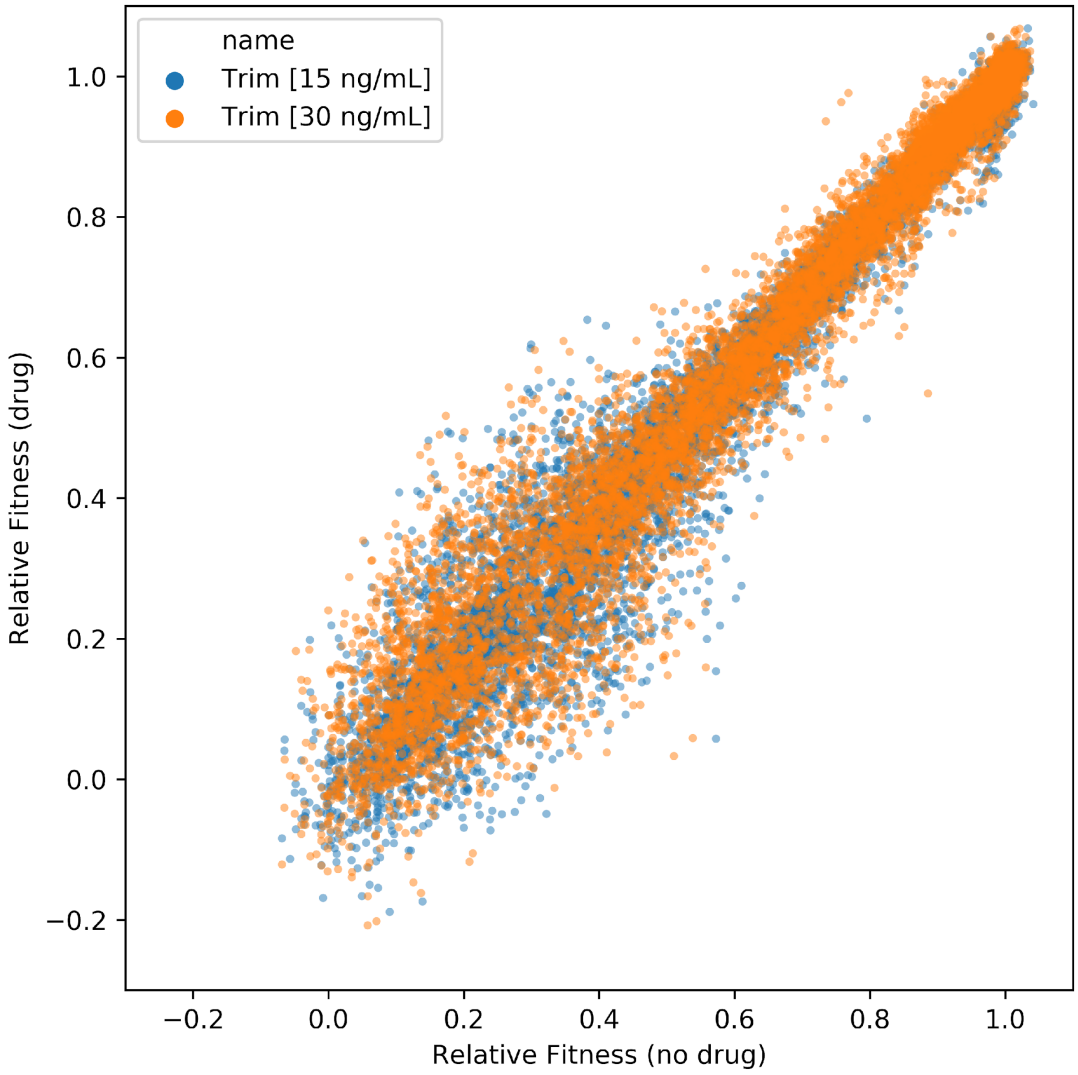


Figure 6.2
Effect of trimethoprim dose on B. subtilis essential genes [excluding the folate pathway]

Competitive pooled-growth experiments were performed on the initial *B. subtilis* ve-CRISPRi library in LB + 1% Xylose, with either 0, 15, or 30 ng/mL trimethoprim. Average relative fitness values plotted for each strain show the comparison between no-trimethoprim replicates (x-axis) and 15 ng/mL (blue) or 30 ng/mL (orange) replicates (y-axis). Most strains are not differentially impacted relative to the control strains. This analysis excludes strains with guides targeting folate pathway genes: *dfrA*, *folB*, *folC*, *folD*, *folE*, *folK*, *sul* (c.f. **Figures 6.3-6.10**).

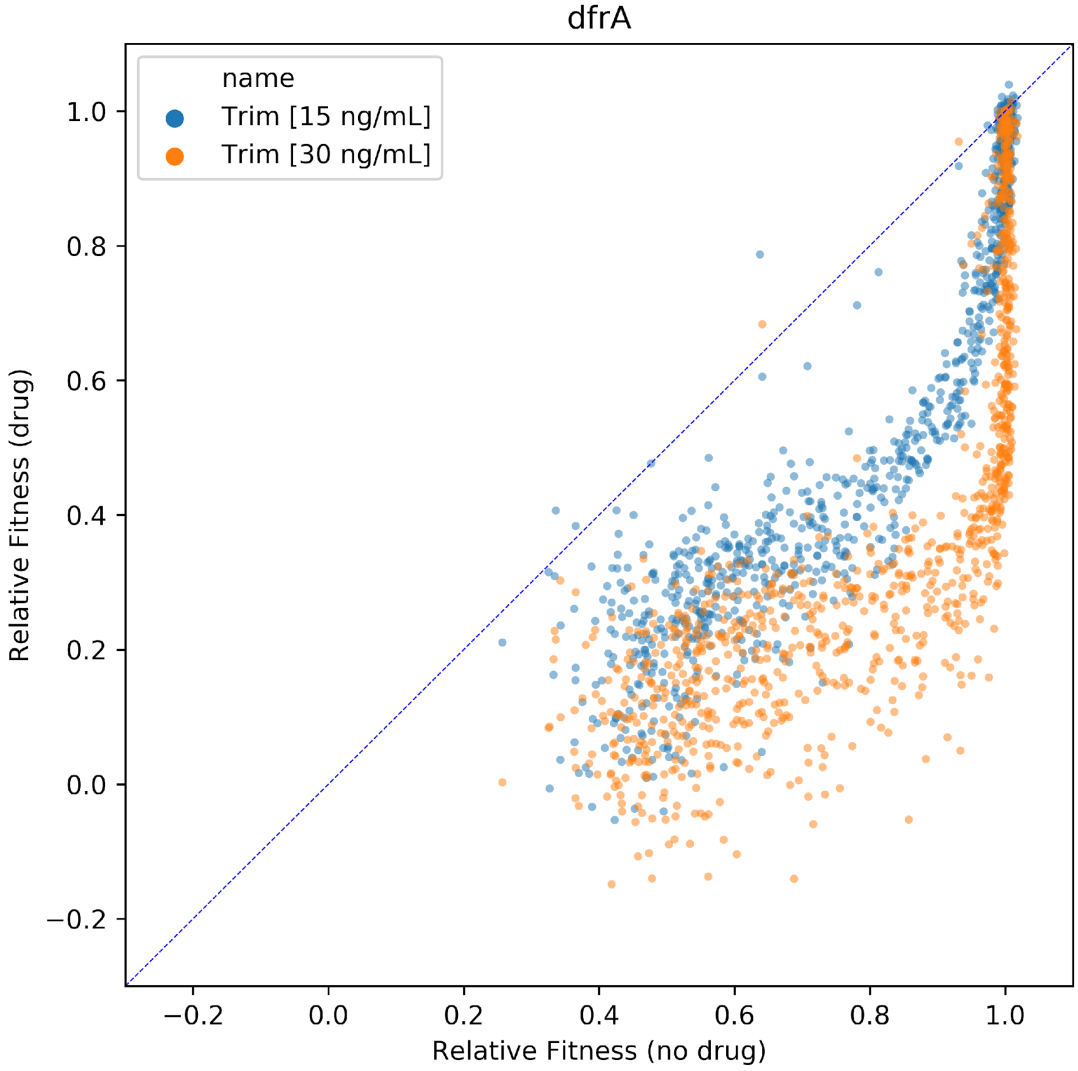


Figure 6.3
Effect of trimethoprim dose [dfrA]

Data as described in **Figure 6.2** but restricted to guides targeting *dfrA*

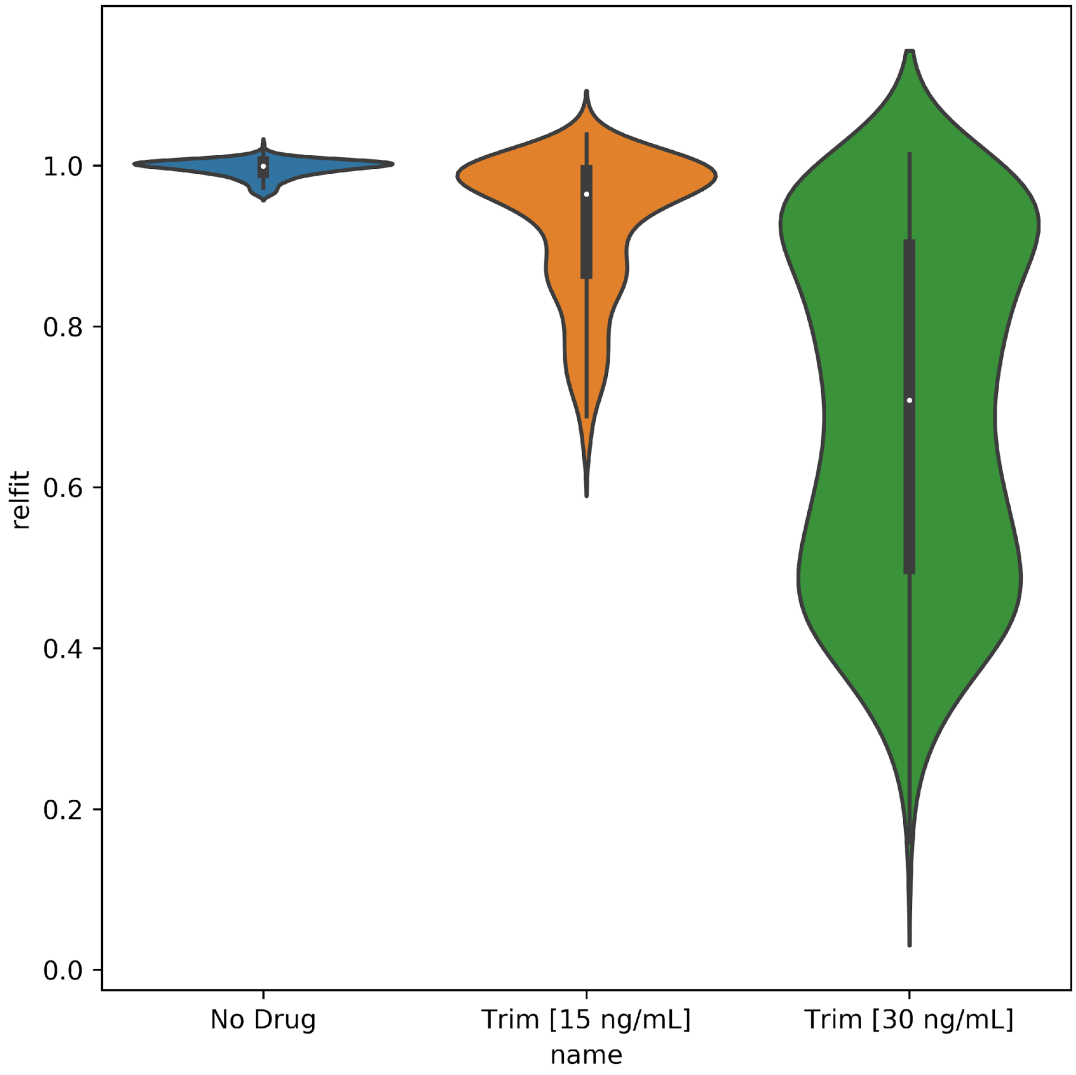


Figure 6.4
Trimethoprim dose exposes phenotypes for dfrA strains

Alternate representation of data from **Figure 6.3**. *dfrA* guides are further sub-filtered to those effecting no-drug phenotypes within one standard deviation of control median (left violin plot). Distribution of fitness scores for these same strains greatly expands in the presence of 15 ng/mL (center violin plot) or 30 ng/mL (right violin plot) trimethoprim.

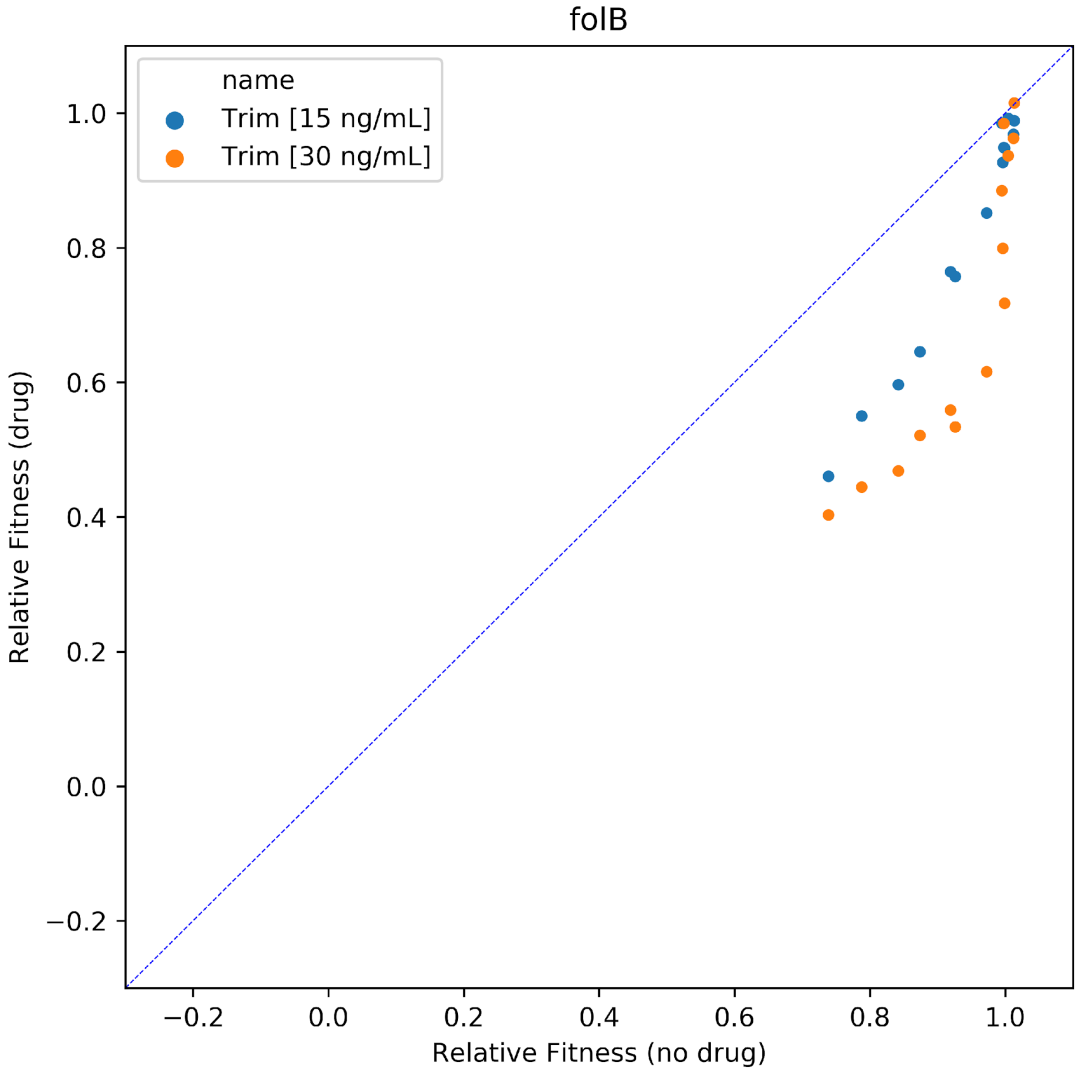


Figure 6.5
Effect of trimethoprim dose [folB]

Data as described in **Figure 6.2** but restricted to guides targeting folB

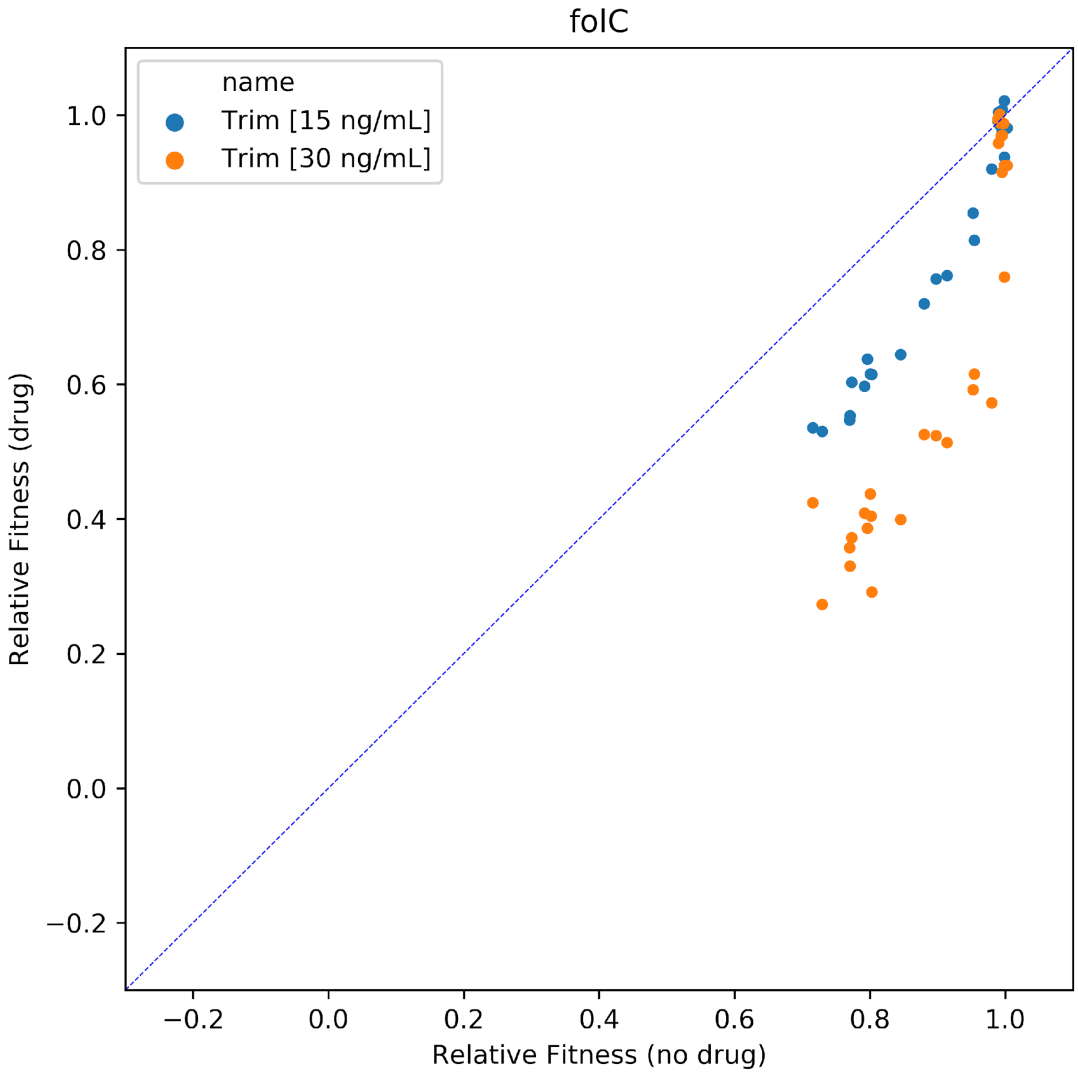


Figure 6.6
Effect of trimethoprim dose [folC]

Data as described in **Figure 6.2** but restricted to guides targeting folC

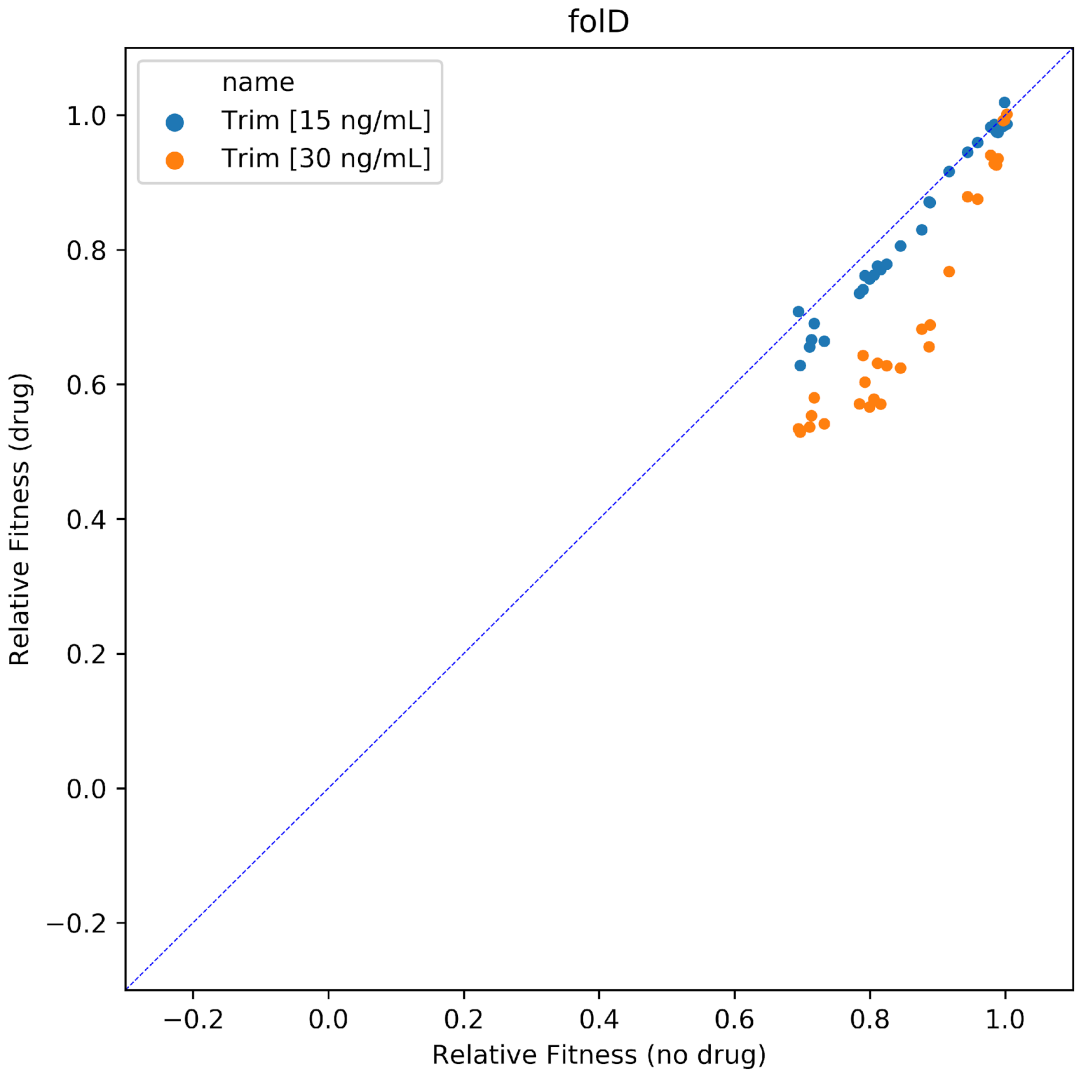


Figure 6.7
Effect of trimethoprim dose [folD]

Data as described in **Figure 6.2** but restricted to guides targeting folD

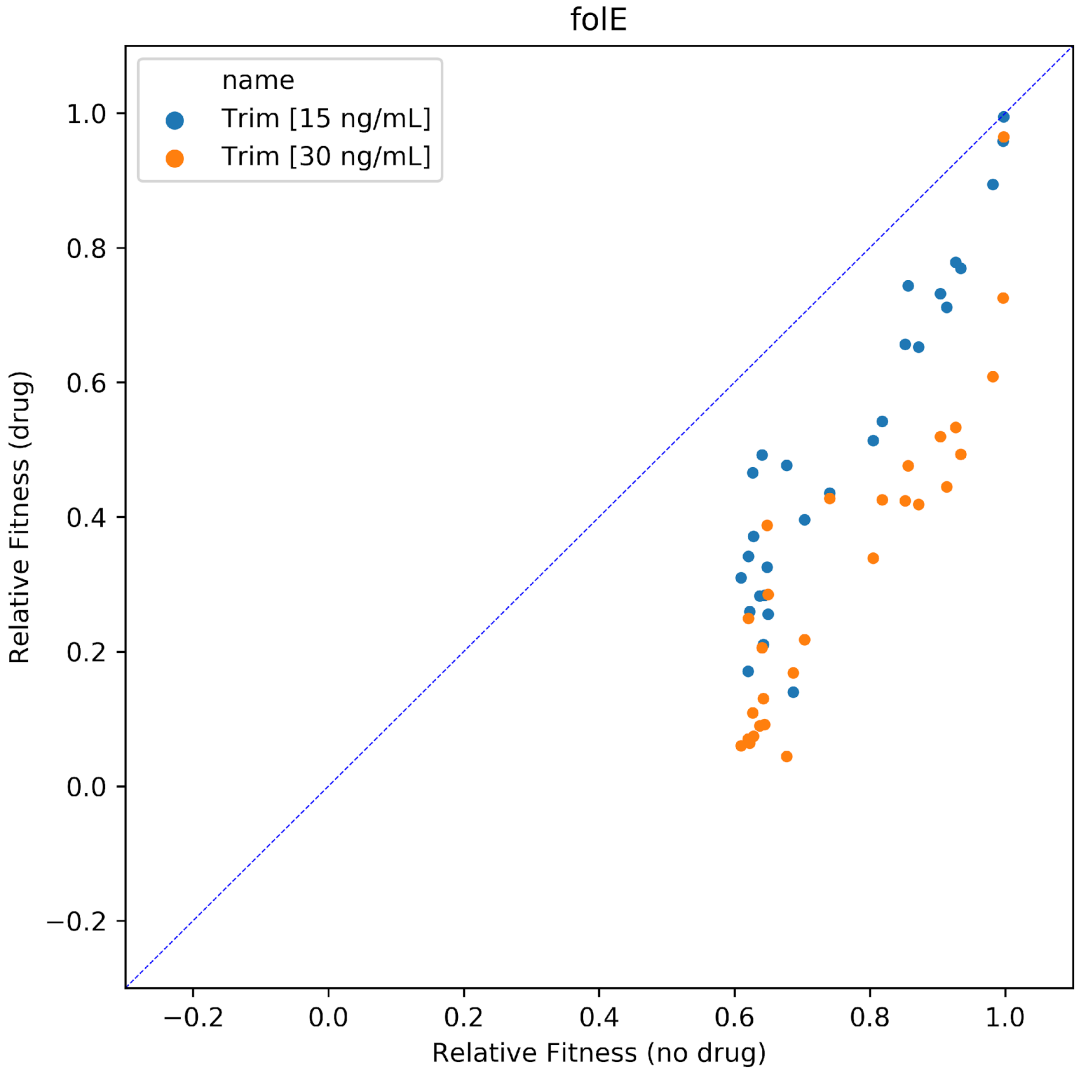


Figure 6.8
Effect of trimethoprim dose [folE]

Data as described in **Figure 6.2** but restricted to guides targeting folE

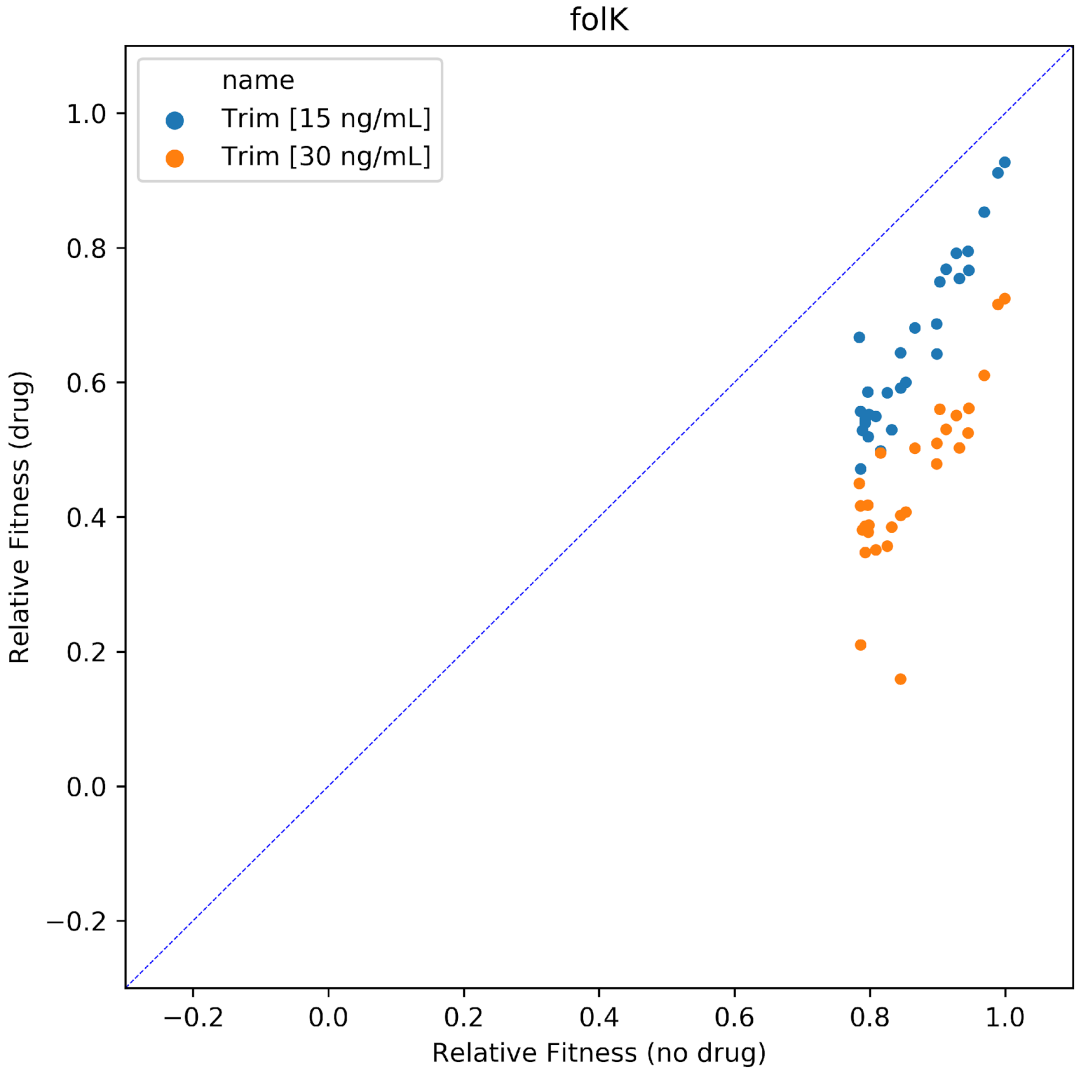


Figure 6.9
Effect of trimethoprim dose [folK]

Data as described in **Figure 6.2** but restricted to guides targeting folK

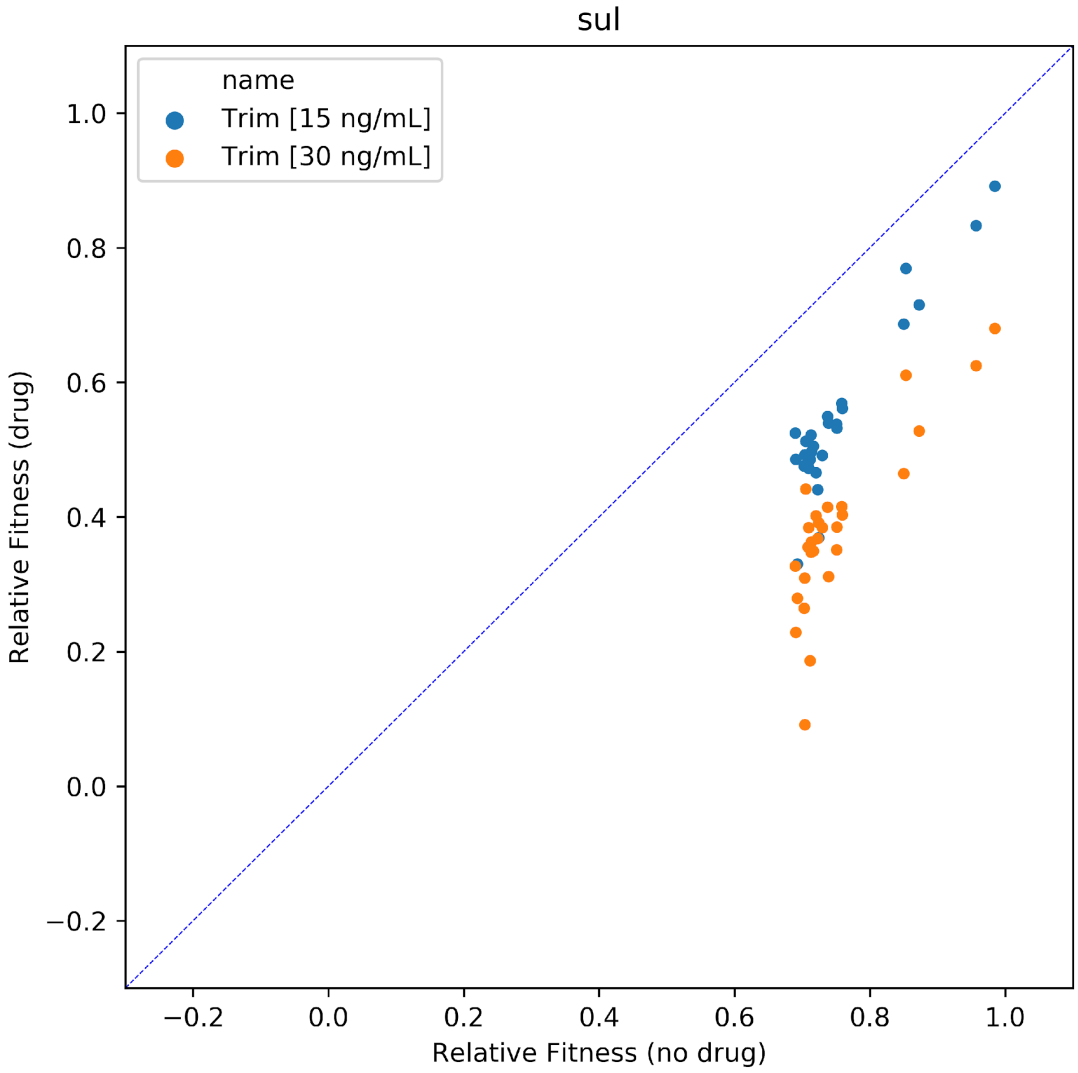


Figure 6.10
Effect of trimethoprim dose [sul]

Data as described in **Figure 6.2** but restricted to guides targeting sul

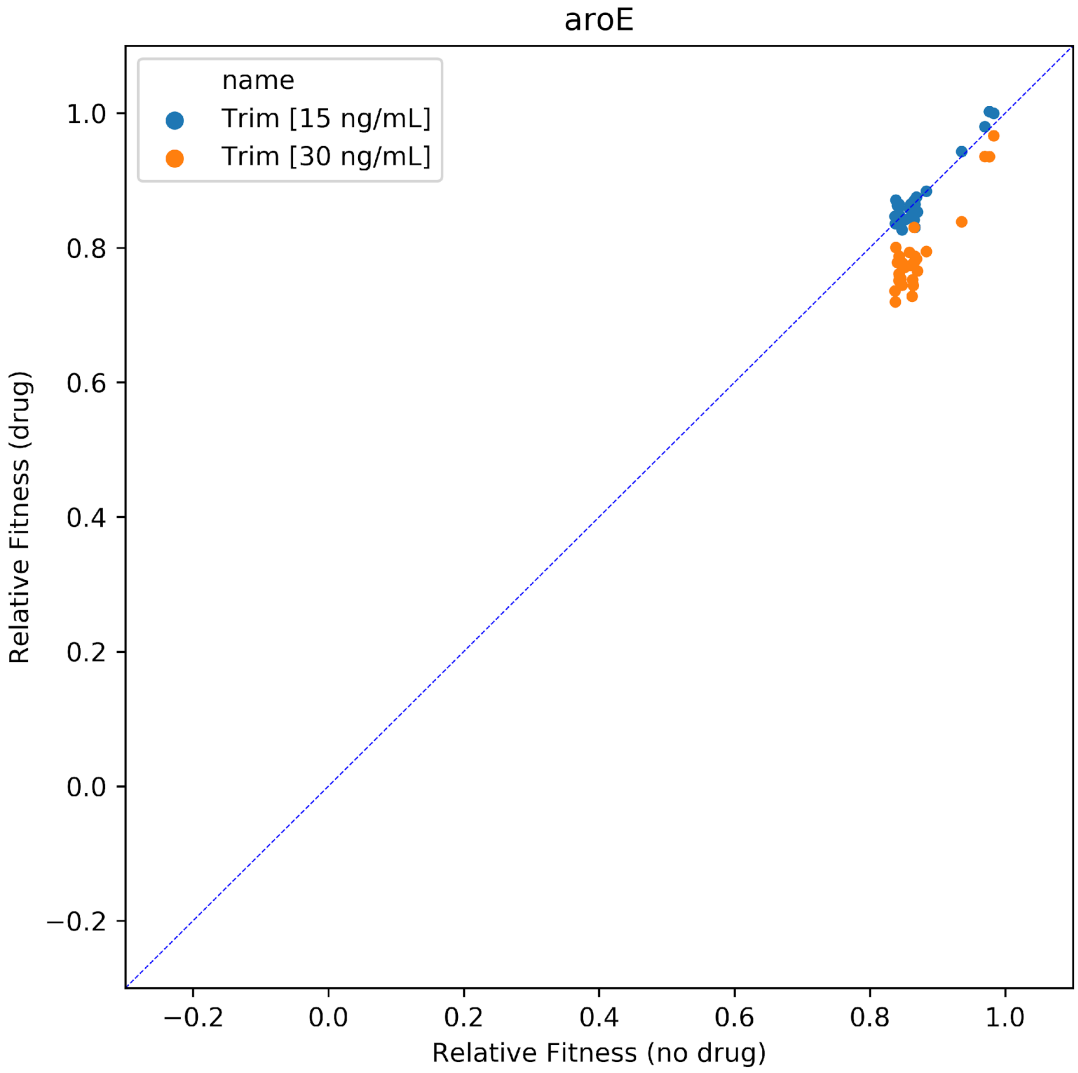


Figure 6.11
Effect of trimethoprim dose [aroE]

Data as described in **Figure 6.2** but restricted to guides targeting aroE

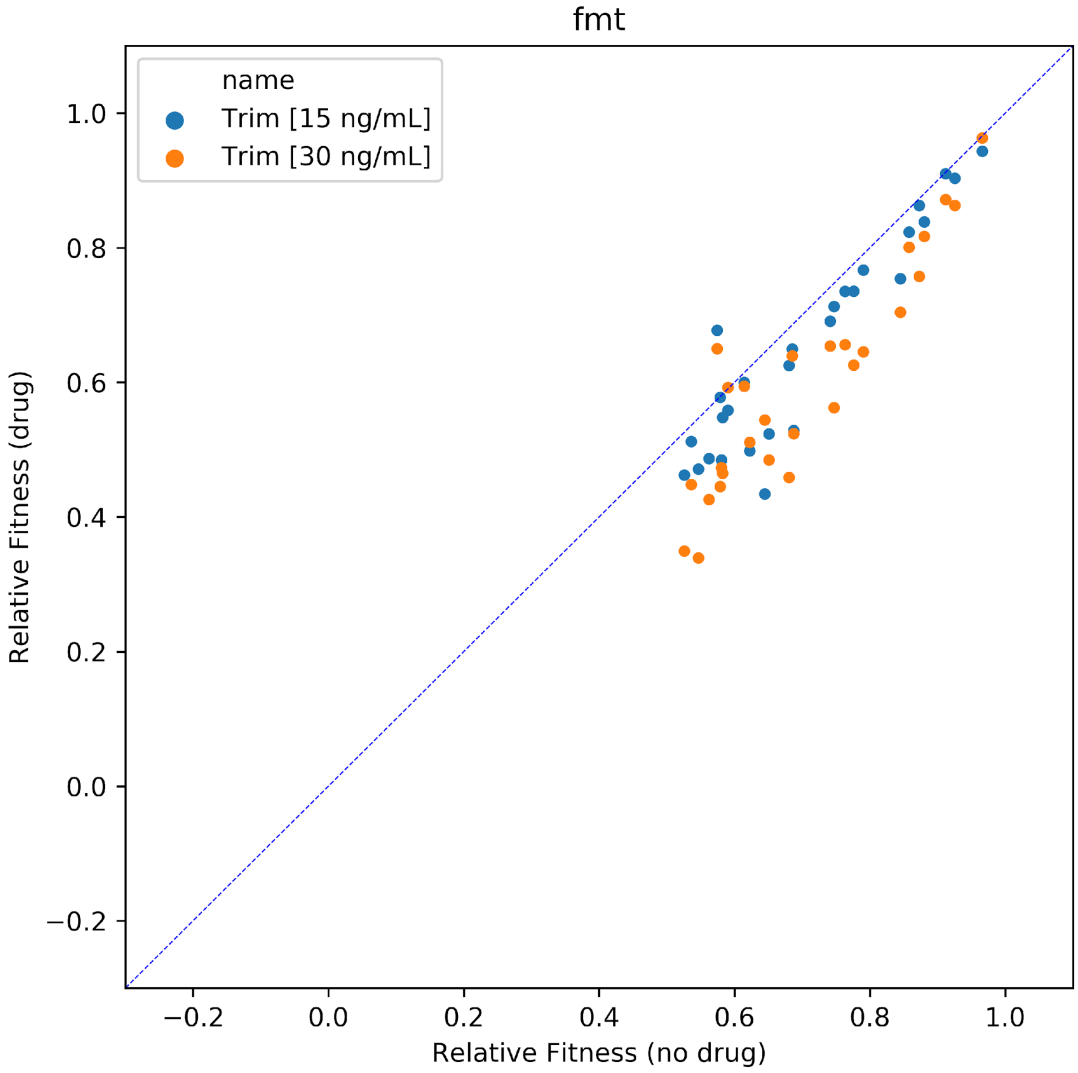


Figure 6.12
Effect of trimethoprim dose [fmt]

Data as described in **Figure 6.2** but restricted to guides targeting fmt

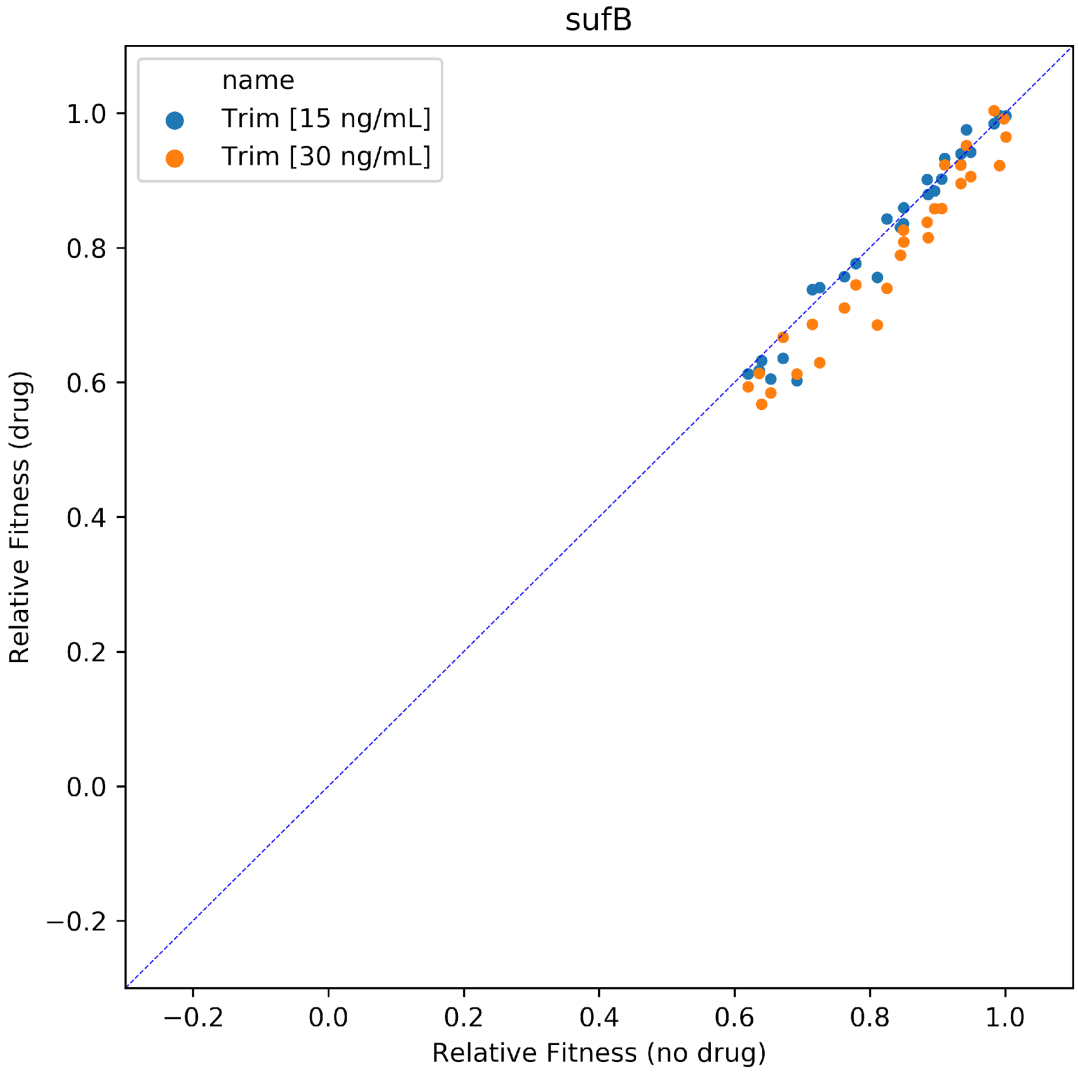


Figure 6.13
Effect of trimethoprim dose [sufB]

Data as described in **Figure 6.2** but restricted to guides targeting sufB

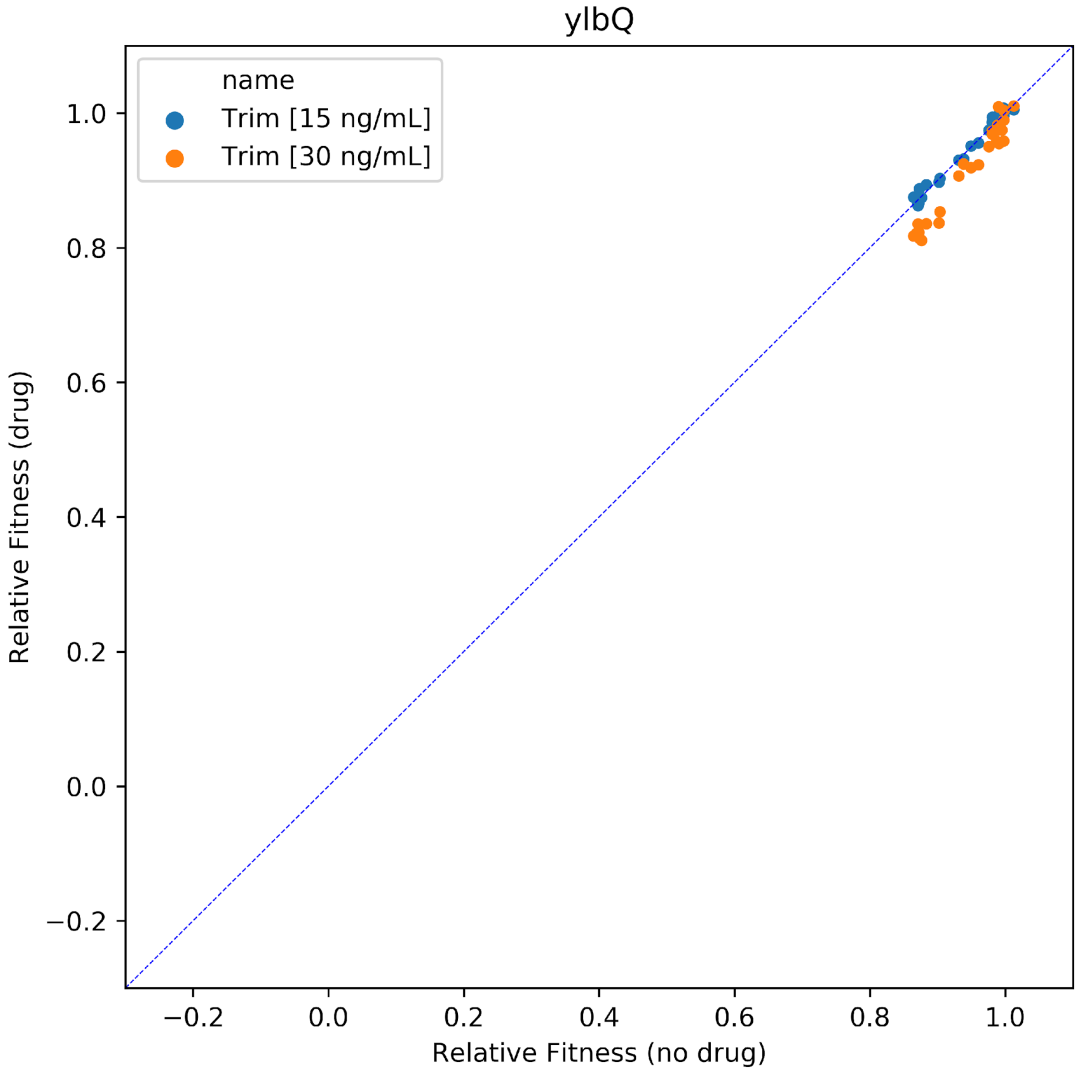


Figure 6.14
Effect of trimethoprim dose [ylbQ]

Data as described in **Figure 6.2** but restricted to guides targeting ylbQ

Appendix #4

Small Libraries

Our model's robust predictions of knockdown efficacy suggest that we can retrieve the most salient features of knockdown/phenotype relationships with a much smaller number of guides. To test this idea, we built small-scale libraries comprising ~11 guides per essential gene for both *B. subtilis* and *E. coli*, and then grew and analyzed these libraries as described for the larger libraries. Whereas the *B. subtilis* library was designed from phenotypes measured in an initial larger-scale experiment, the *E. coli* library was designed solely based on the predictive model.

Phenotypes in the minimal libraries correlate well with the model predictions for *E. coli* (**Figure 7.1**, R=0.63) and *B. subtilis* (**Figure 7.2**, R=0.59) and with measurements from overlapping experiments with large pools (**Figure 7.3, 7.4**). When knockdown predictions are considered on a per-gene basis they correspond even more closely with median correlations of 0.78 (**Figure 7.5**) for *B. subtilis* and 0.58 (**Figure 7.6**) for *E. coli*. This compares favorably to the relationships exhibited by the large libraries, with correlations of 0.70 (**Figure 7.7**) and 0.41 (**Figure 7.8**) respectively. Genes in *B. subtilis* generally show good correspondence for both libraries (**Figure 7.9**). Genes with poor correlation for the small *E. coli* library generally also have low correlations for the large library and low-magnitude phenotypes overall, pointing to the non-linearity of the true knockdown/phenotype curve of the gene as the primary source of the discrepancy (**Figure 7.10**).

We successfully obtained intermediate phenotypes for the vast majority of our targeted genes. 95% of genes had at least 2 guides with phenotypes in the targeted range of

10% to 90% of wildtype growth in *subtilis* (**Figure 7.11**), and for the strictly prediction-based *E. coli* library this rose to 3 guides (**Figure 7.12**). Half of genes in the latter library had at least 10 (i.e. nearly all) guides in the targeted range. This compares favorably to outcomes for the larger libraries (**G**), and demonstrates that even a small library of mismatch guides was sufficient to introduce a meaningful variety of knockdown phenotypes for each included gene.

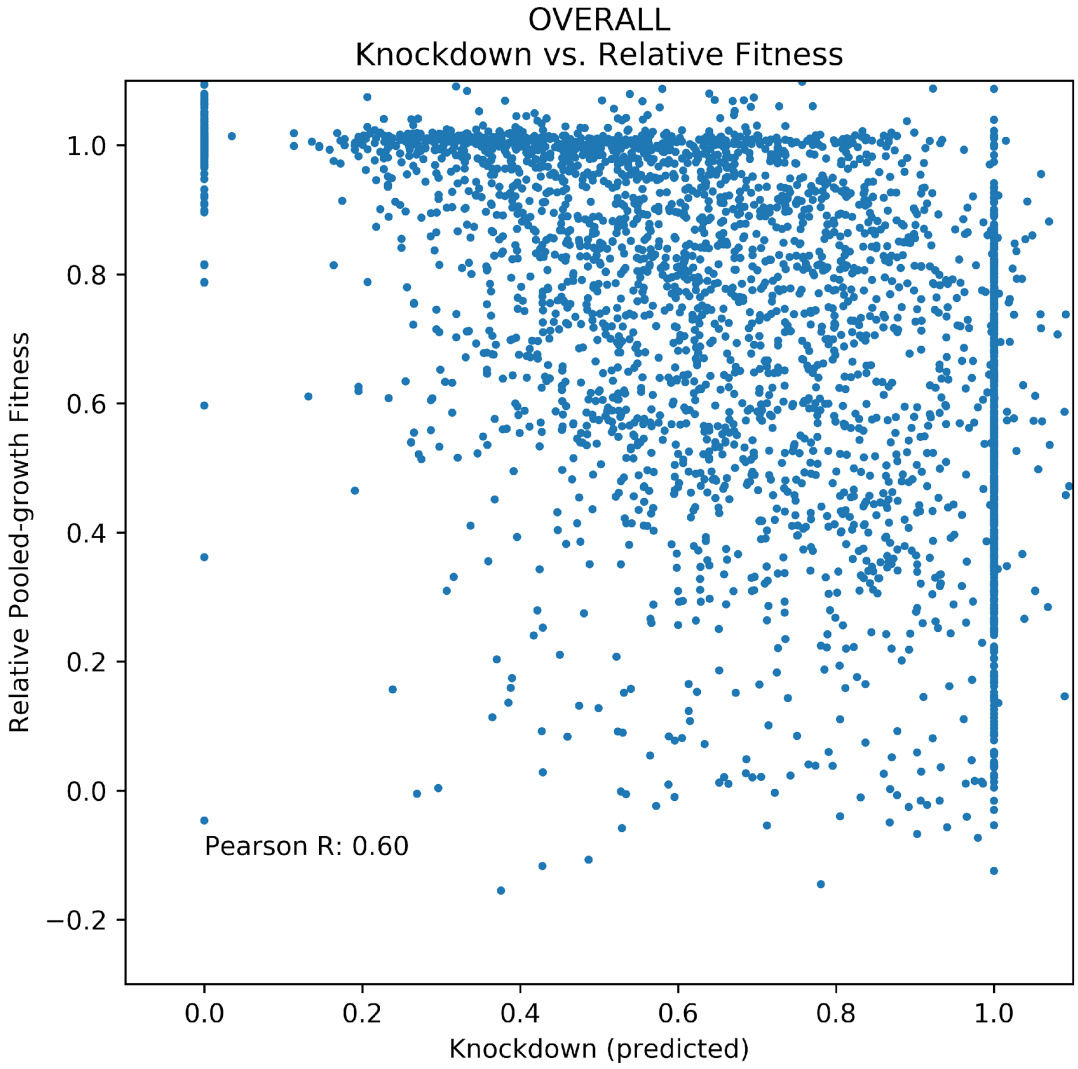


Figure 7.1
B. subtilis small library fitness vs model prediction.

Fitness measurements for strains in small *B. subtilis* library. Competitive pooled growth experiments were structurally identical to those performed on the large libraries save that a smaller physical volume was used for each replicate (15 mL instead of 150 mL). Guides with no mismatches were assigned knockdown "predictions" of 1.0 (targeting) or 0.0 (non-targeting control).

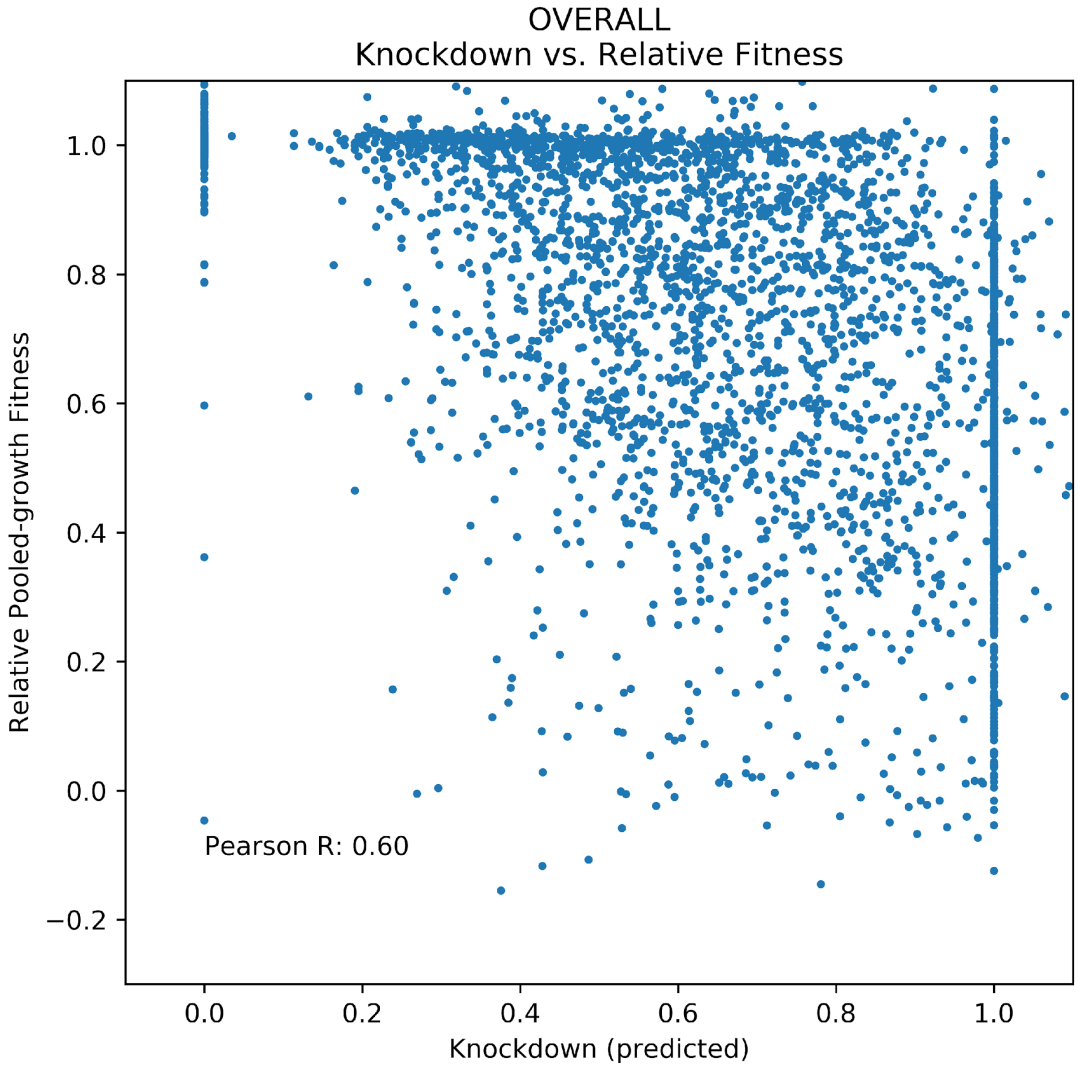


Figure 7.2
E. coli small library fitness vs model prediction.

Fitness measurements for strains in small *E. coli* library. Competitive pooled growth experiments were performed exactly as decribed for the large libraries save cultures were 15 mL rather than 150 mL. Guides with no mismatches were assigned knockdown "predictions" of 1.0 (targeting) or 0.0 (non-targeting control).

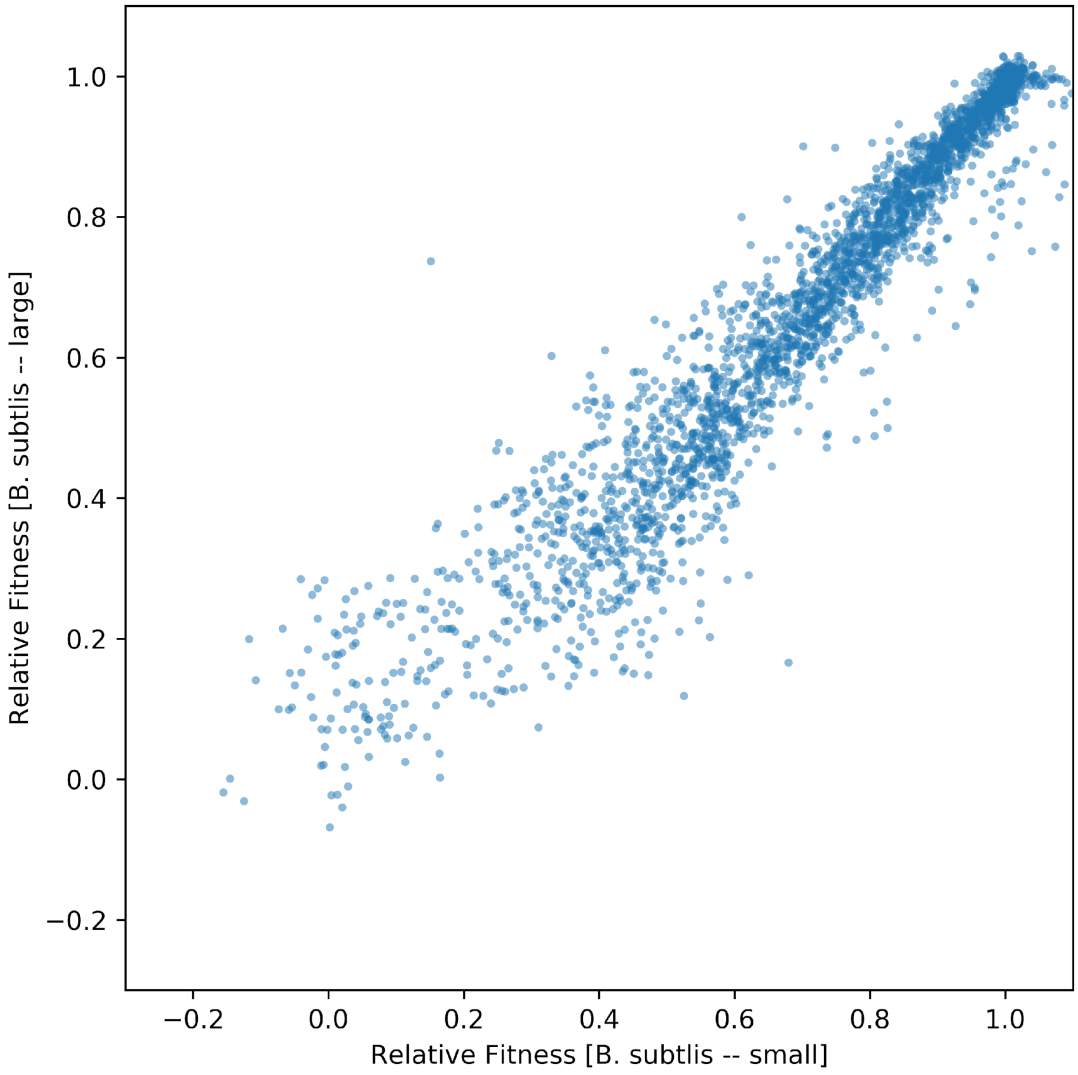


Figure 7.3
Comparison of B. subtilis fitness scores measured for small library vs. original pre-screen library

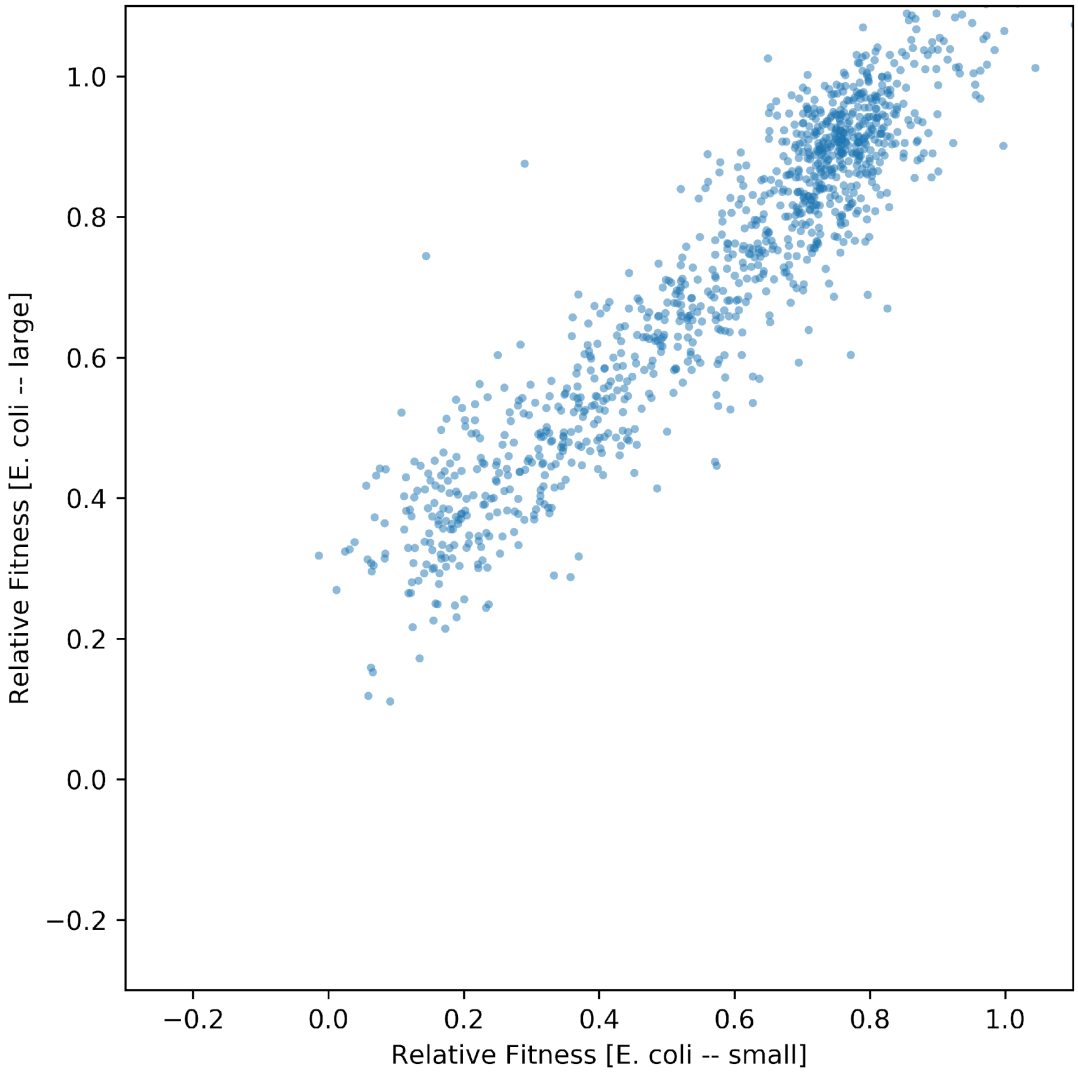


Figure 7.4
Comparison of E. coli fitness scores measured for small library vs. large library

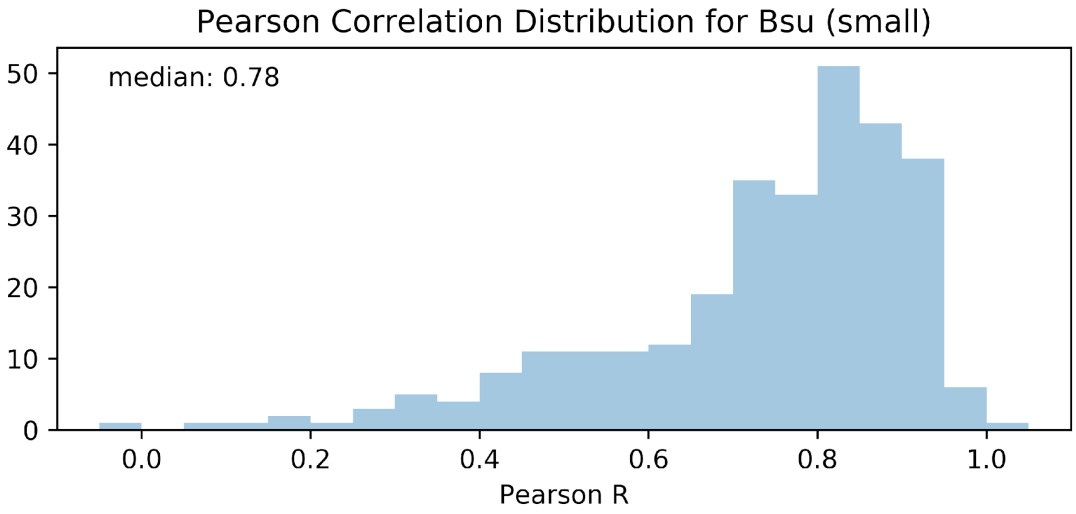


Figure 7.5
B. subtilis small library Pearson correlations by gene

Pearson R was computed between predicted knockdown and measured fitness for guides targeting each gene. Histogram shows distribution of per-gene Pearson R scores.

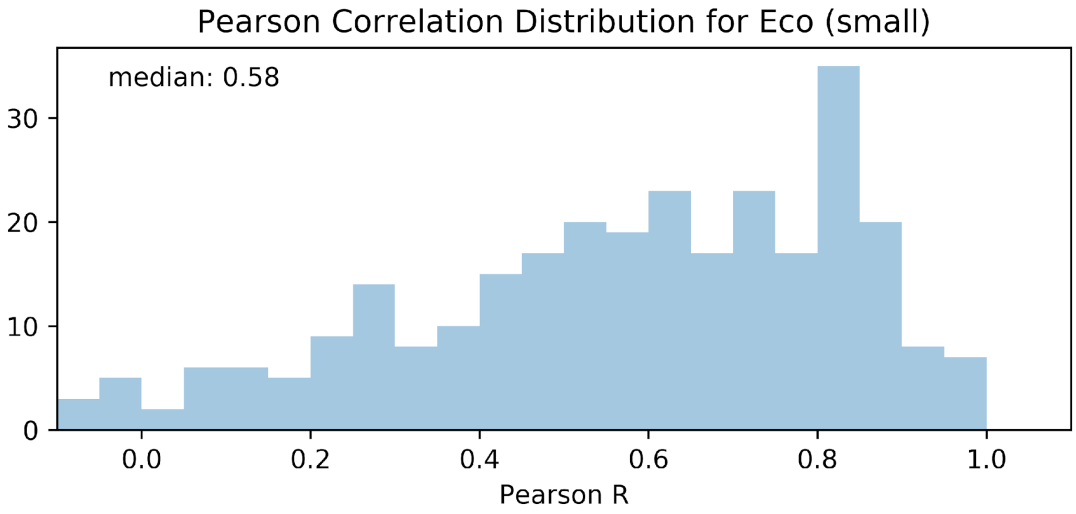


Figure 7.6
E. coli small library Pearson correlations by gene

[See **Figure 7.5**.]

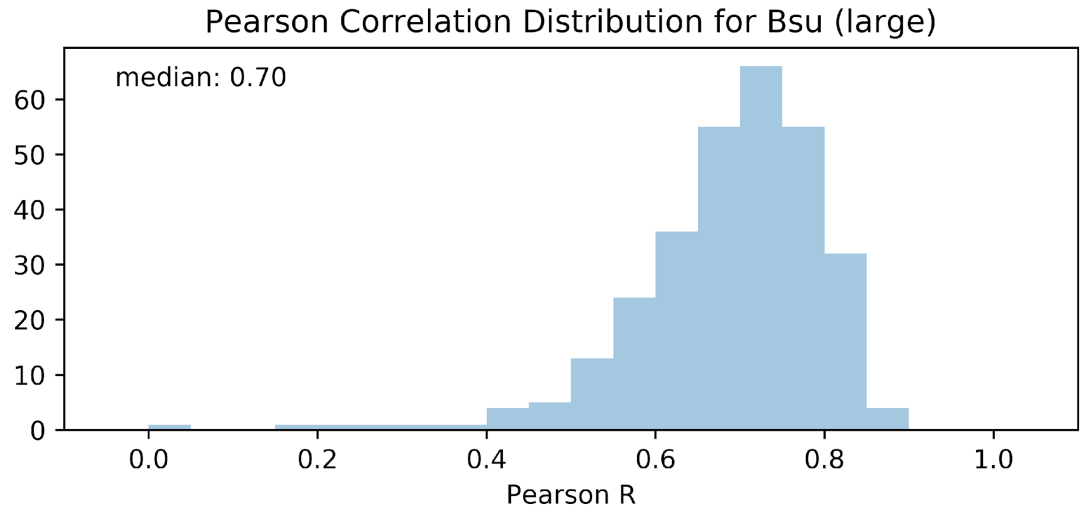


Figure 7.7
B. subtilis large library Pearson correlations by gene

[See **Figure 7.5**.]

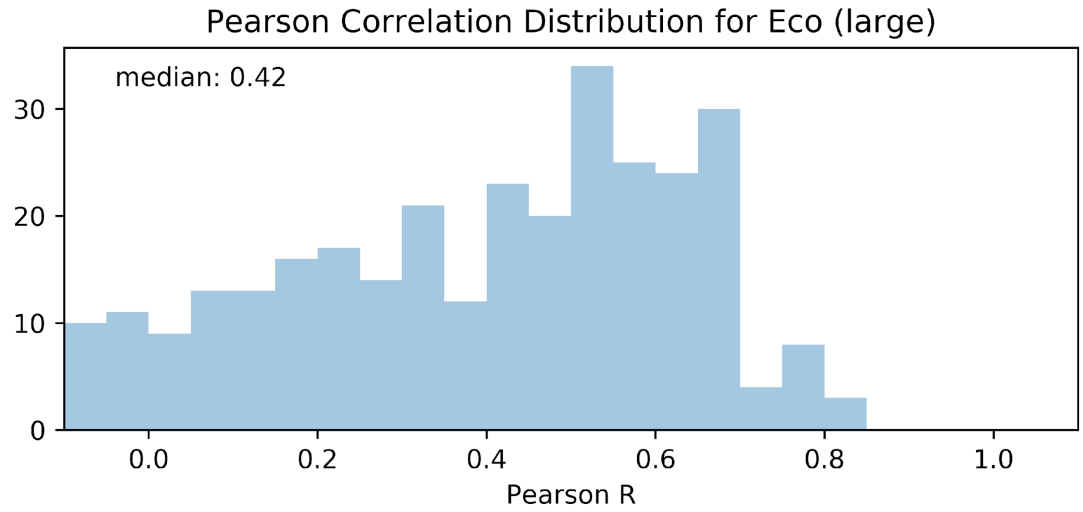


Figure 7.8
E. coli large library Pearson correlations by gene

[See **Figure 7.5**.]

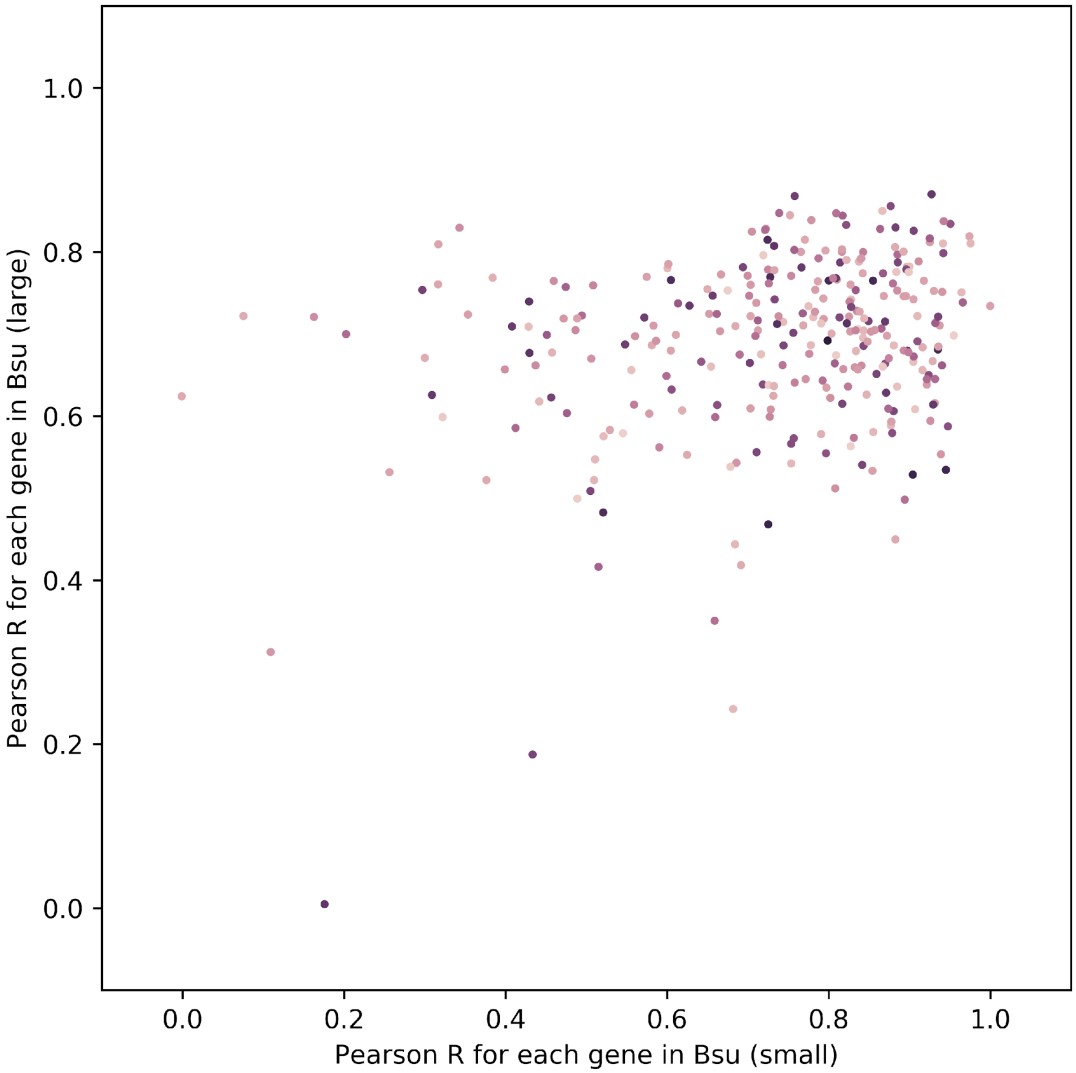


Figure 7.9
B. subtilis small v. large library comparison of Pearson correlations by gene.

Individual values as computed for 7.5/7.6, compared. Coloration of dots indicates extremity of phenotype; genes with darker colors have more extreme phenotypes. Pearson correlations were generally high for all genes in both libraries.

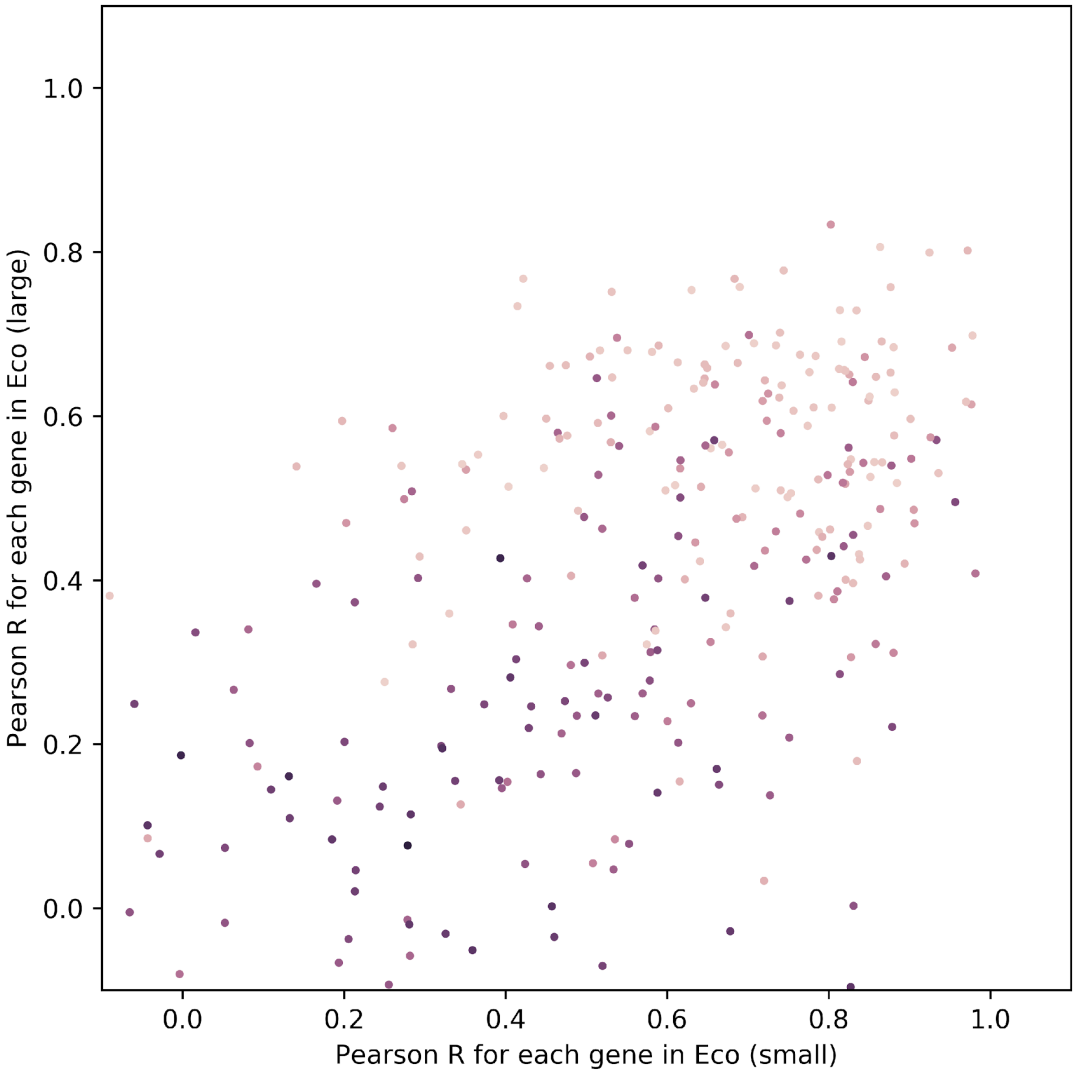


Figure 7.10
E. coli small v. large library comparison of Pearson correlations by gene.

Individual values as computed for 7.7/7.8, compared. Coloration of dots indicates extremity of phenotype; genes with darker colors have more extreme phenotypes. More genes show low correlation in *E. coli*, particularly genes with low-magnitude phenotypes. More phenotypes were captured by the small library experiment.

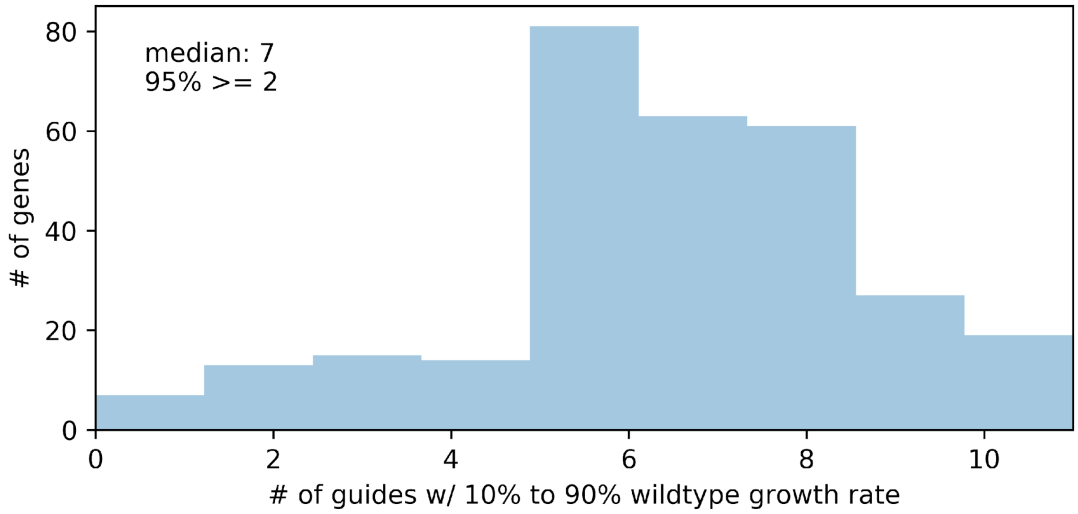


Figure 7.11
Guides in targeted range per gene for B. subtilis (small library)

Guides with phenotypes in the desired range (between 10% and 90% wildtype growth rate) were counted for each gene. Histogram shows distribution of per-gene coverages for small *B. subtilis* library.

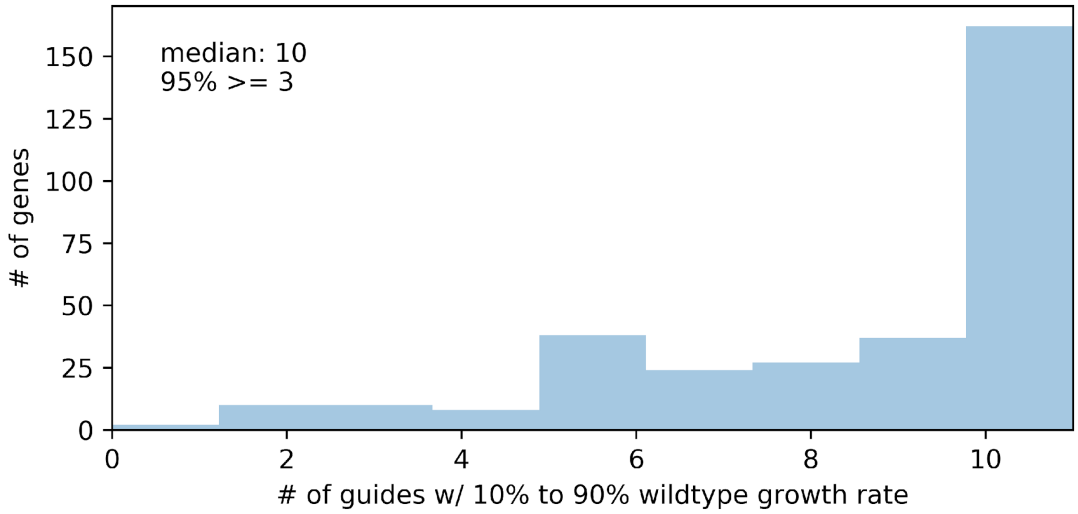


Figure 7.12
Guides in targeted range per gene for E. coli (small library)

Guides with phenotypes in the desired range (between 10% and 90% wildtype growth rate) were counted for each gene. Histogram shows distribution of per-gene coverages for small *E. coli* library.

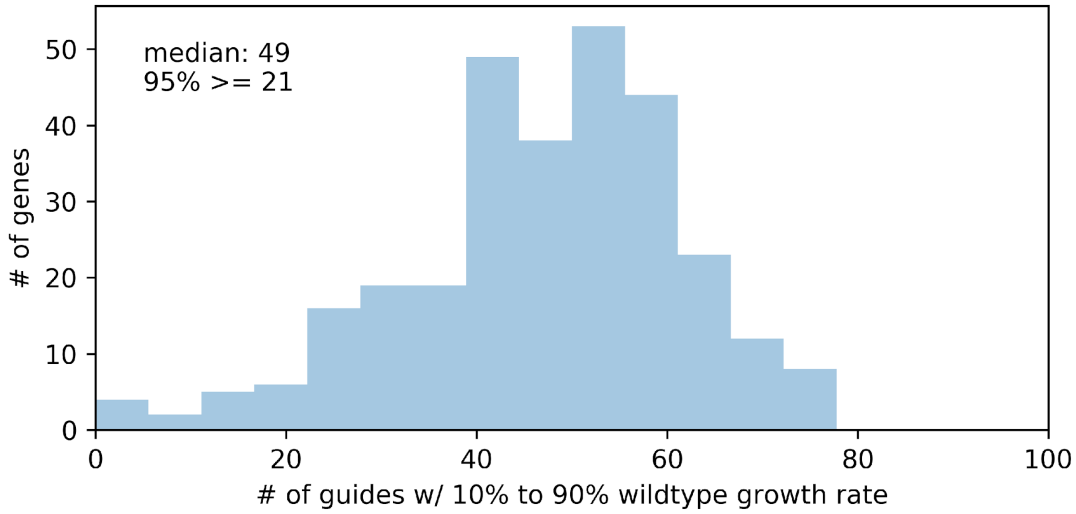


Figure 7.13
Guides in targeted range per gene for B. subtilis (large library)

Guides with phenotypes in the desired range (between 10% and 90% wildtype growth rate) were counted for each gene. Histogram shows distribution of per-gene coverages for large *B. subtilis* library.

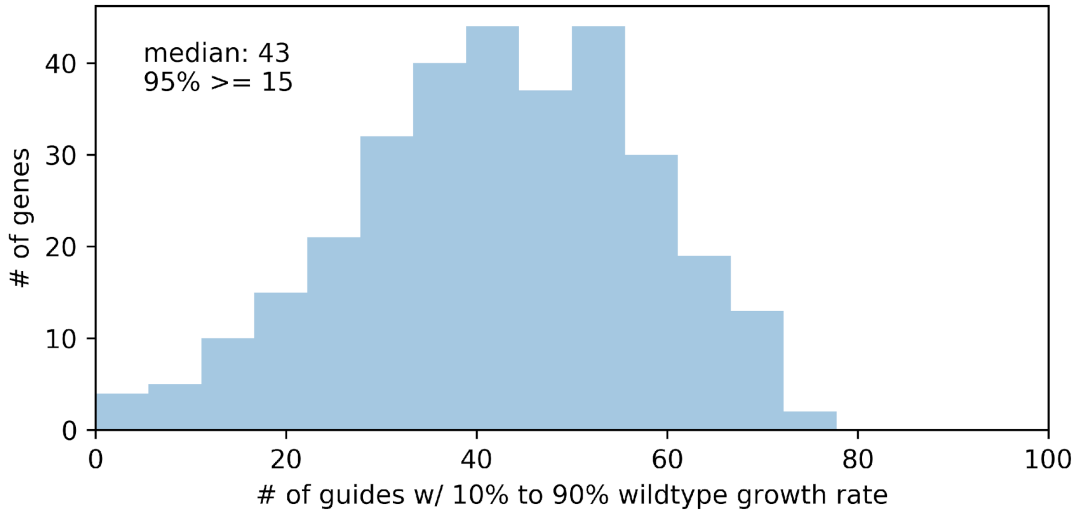


Figure 7.14
Guides in targeted range per gene for E. coli (large library)

Guides with phenotypes in the desired range (between 10% and 90% wildtype growth rate) were counted for each gene. Histogram shows distribution of per-gene coverages for large *E. coli* library.

Appendix #5

Double-Mismatch Guides

Our approach of titrating guide efficacy with a single mismatch captures a wide range of knockdowns for any gene target. In Appendix #3 we showed evidence that such guides can impart even the very low levels of knockdown necessary to capture a range of phenotypes in the presence of a synergistic secondary stress. Prior to measuring actual outcomes for these guides we were concerned that a single mismatch might be insufficient to our aims, and so our initial libraries (including our GFP library) incorporated a sizeable number of guides (49 per parent locus) with two mismatched bases. These guides did not prove necessary for our primary objectives. They do, however, provide a second lens through which to examine the performance of the predictive model (trained exclusively on single-mismatch guide outcomes).

Because the model was not built to consider multiple mismatches, we cannot directly apply it to these guides. However, because the model predicts knockdown as a *fraction* of parent knockdown, we can assume independence --i.e. assume that the impact of each mismatch is unchanged by the match at other positions in the guide -- and simply multiply the fractional predictions together, declaring that a mismatch imparting a 75% knockdown and a mismatch imparting an 80% knockdown would together impart a (75% * 80%) = 60% knockdown. Certainly the relationship is unlikely to be truly independent in practice. Nevertheless, we find the predictions from this conscious over-application of the single-mismatch model to be surprisingly accurate.

For our GFP data, predictions of the double-mismatch guides correspond very well to measured outcomes across both organisms (**Figure 8.1**, R = 0.73). Furthermore, when

we look at fitness outcomes for our broad libraries of *dfrA* or *murAA* guides in *B. subtilis* (for which our initial libraries included double-mismatch guides), we saw a meaningful correspondence even though many of the guides present no phenotype across 10 generations (*dfrA*: **Figure 8.2**, R = 0.50; *murAA*: **Figure 8.3**, R = 0.56). Furthermore, in the presence of trimethoprim (see Appendix #3) we find the prediction/fitness correspondence rises even further (15 ng/mL: **Figure 8.4,** R = 0.60; 30 ng/mL: **Figure 8.5**, R = 0.66).

We do not recommend the application of these double-mismatch guides in practice. They were more challenging to predict precisely and, in light of the broad titration coverage of single-mismatch variants, unnecessary. These data do however reinforce the predictive power of our single-mismatch linear model, and provide further evidence that the fundamental mechanism of the CRISPRi system is robust to many particulars of context.

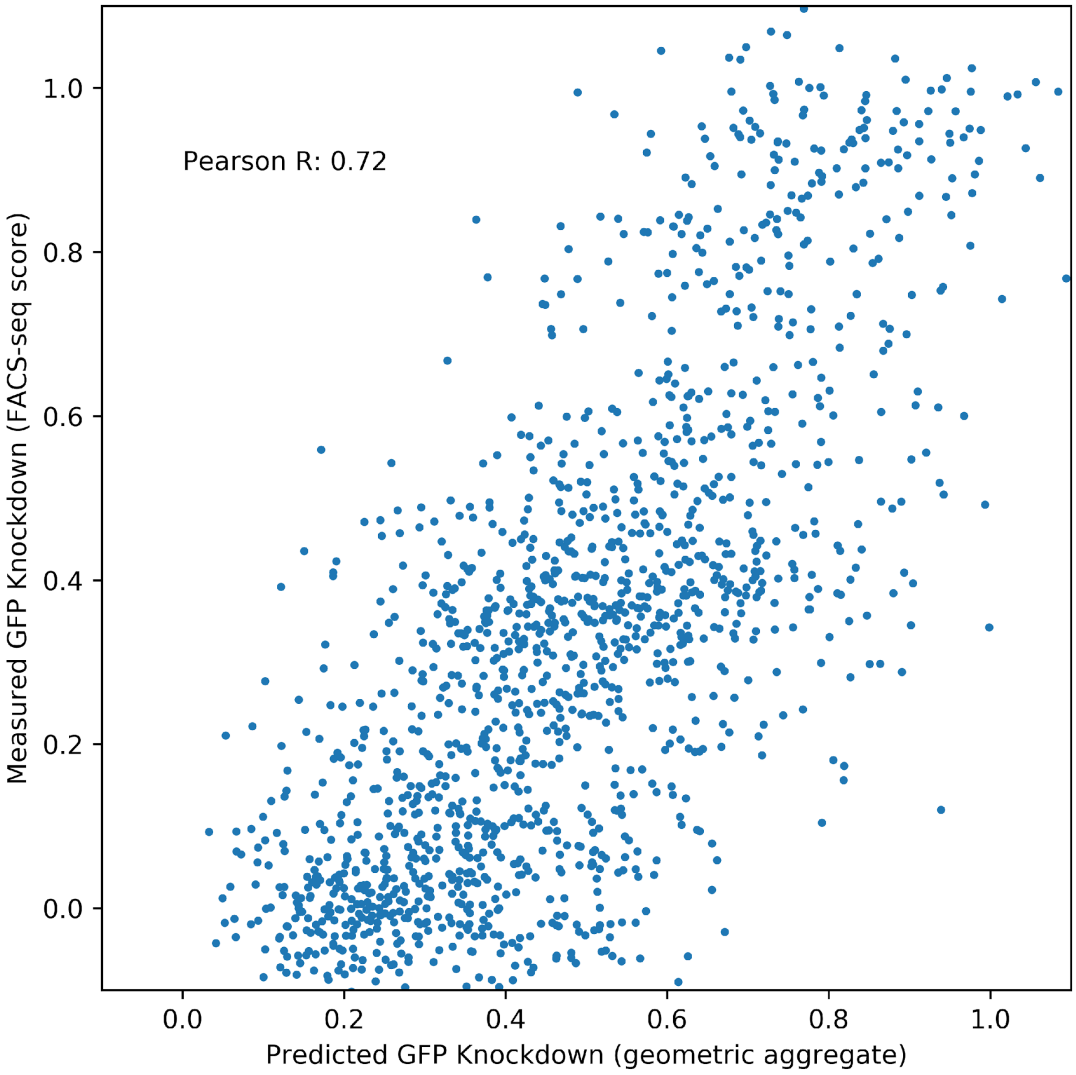


Figure 8.1
Predicted vs. Measurement (via FACS-seq) of GFP Knockdown

Comparison of predicted knockdown to average measured knockdown for GFP library. Predictions were made by separately applying single-mismatch model to each mismatch in isolation and multiplying the outputs.

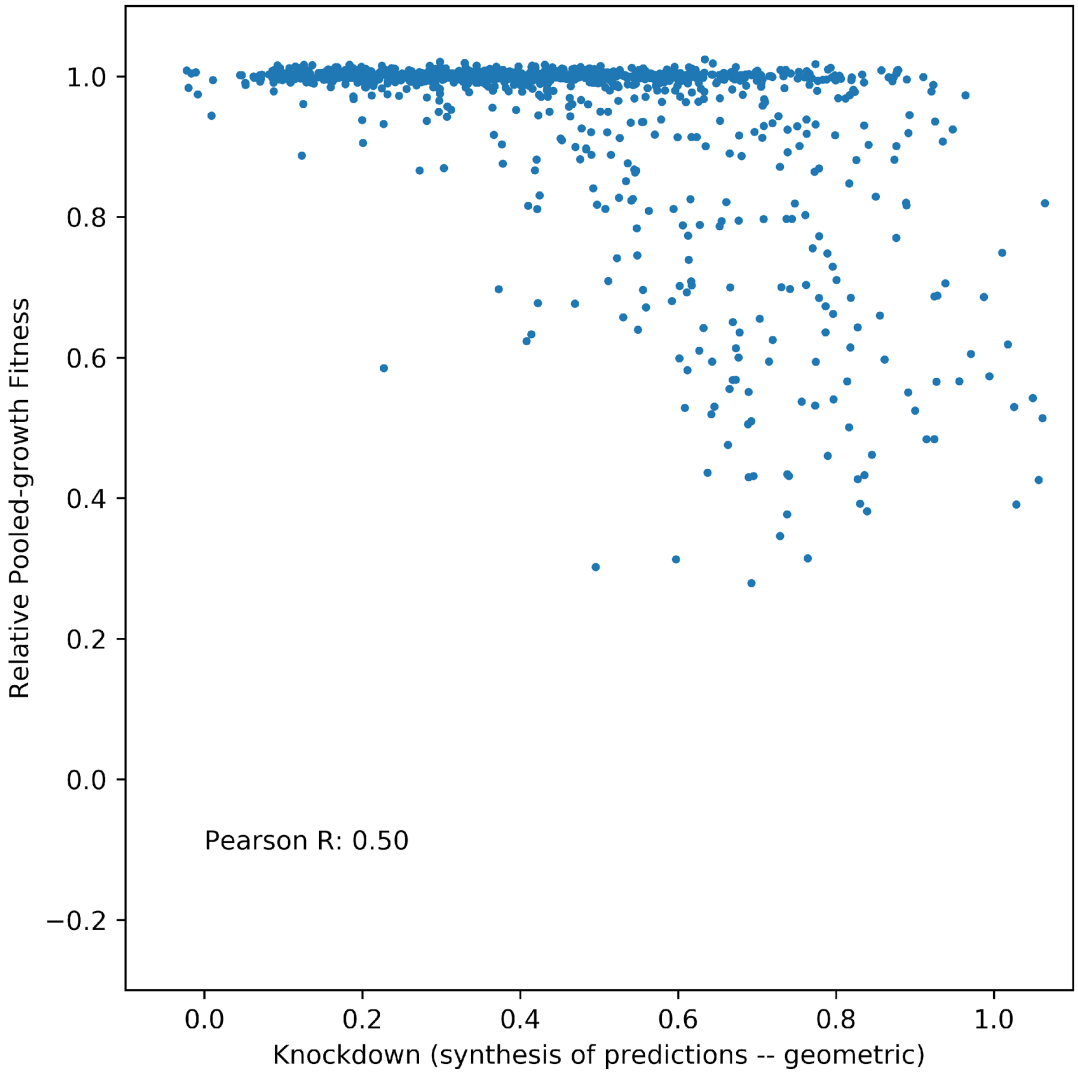


Figure 8.2
Predicted Knockdown vs Measured Relative Fitness (Bsu/dfrA)

Comparison of predicted knockdown to measured relative fitness for *dfrA* library in *B. subtilis*. Predictions were made by separately applying single-mismatch model to each mismatch in isolation and multiplying the outputs.

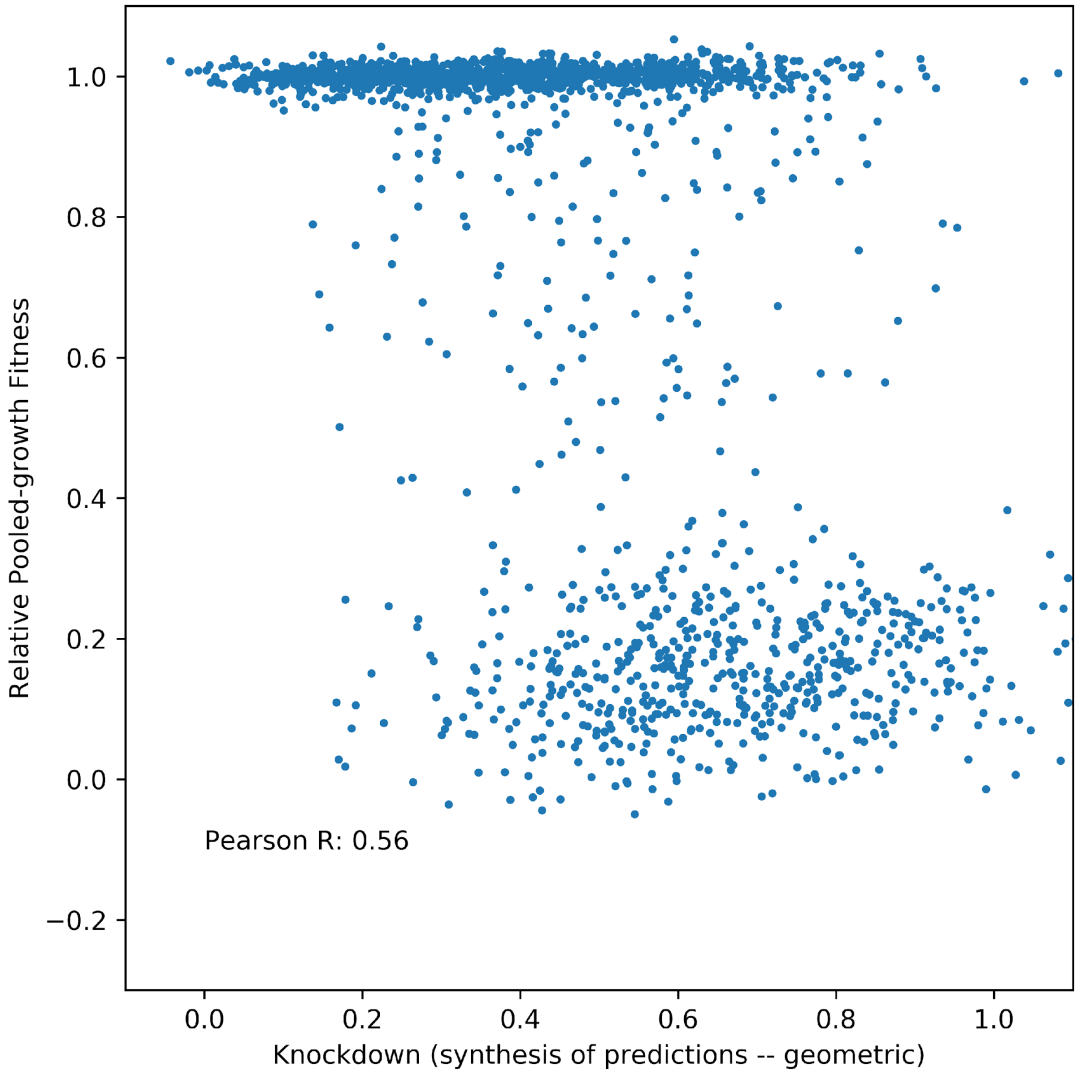


Figure 8.3
Predicted Knockdown vs Measured Relative Fitness (Bsu/murAA)

Comparison of predicted knockdown to measured relative fitness for *murAA* library in *B. subtilis*. Predictions were made by separately applying single-mismatch model to each mismatch in isolation and multiplying the outputs.

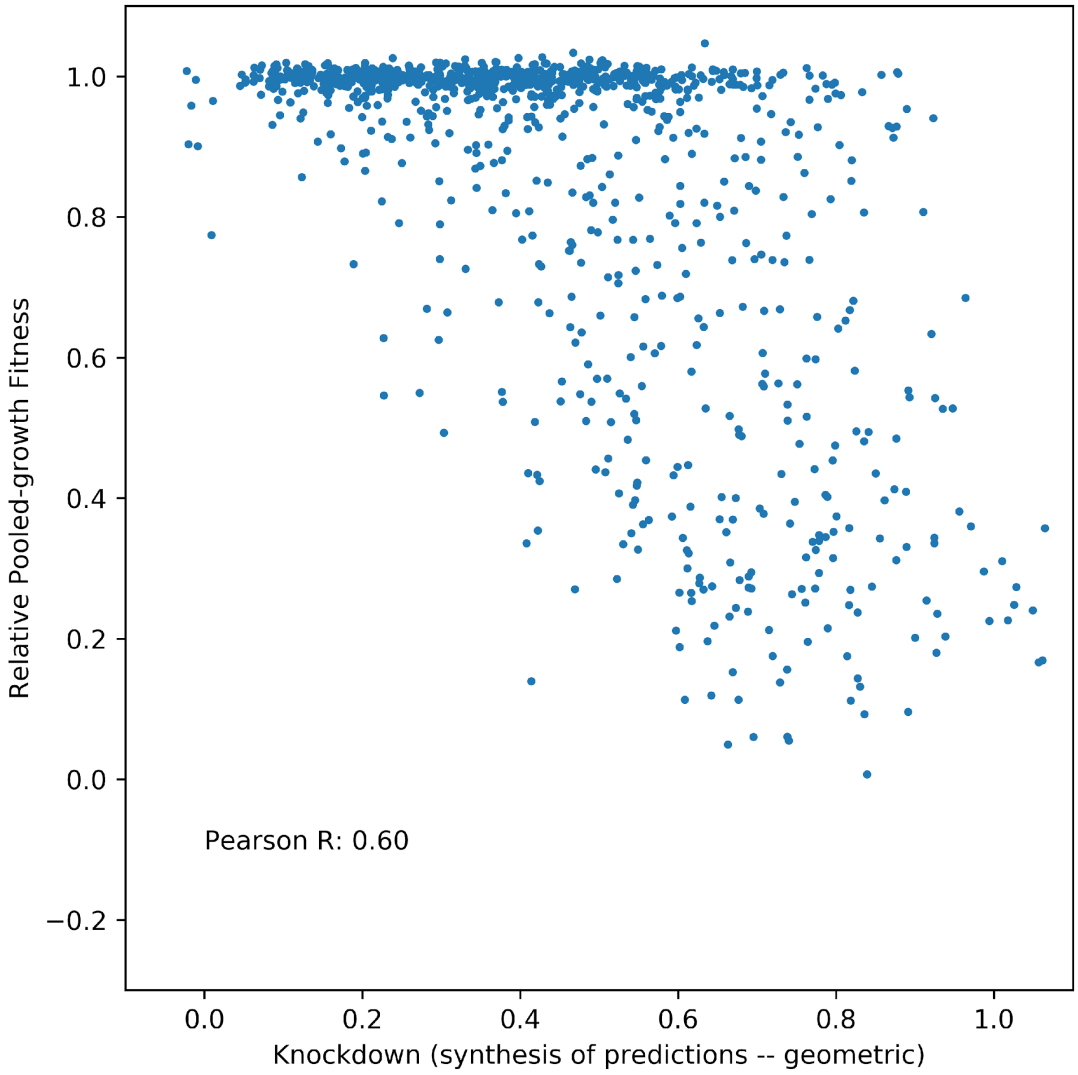


Figure 8.4
KD Prediction vs. Fitness (Bsu/dfrA) [15 ng/mL trimethoprim]

Data as for **Figure 8.2**, but using fitness measurements from samples with 15 ng/mL trimethoprim.

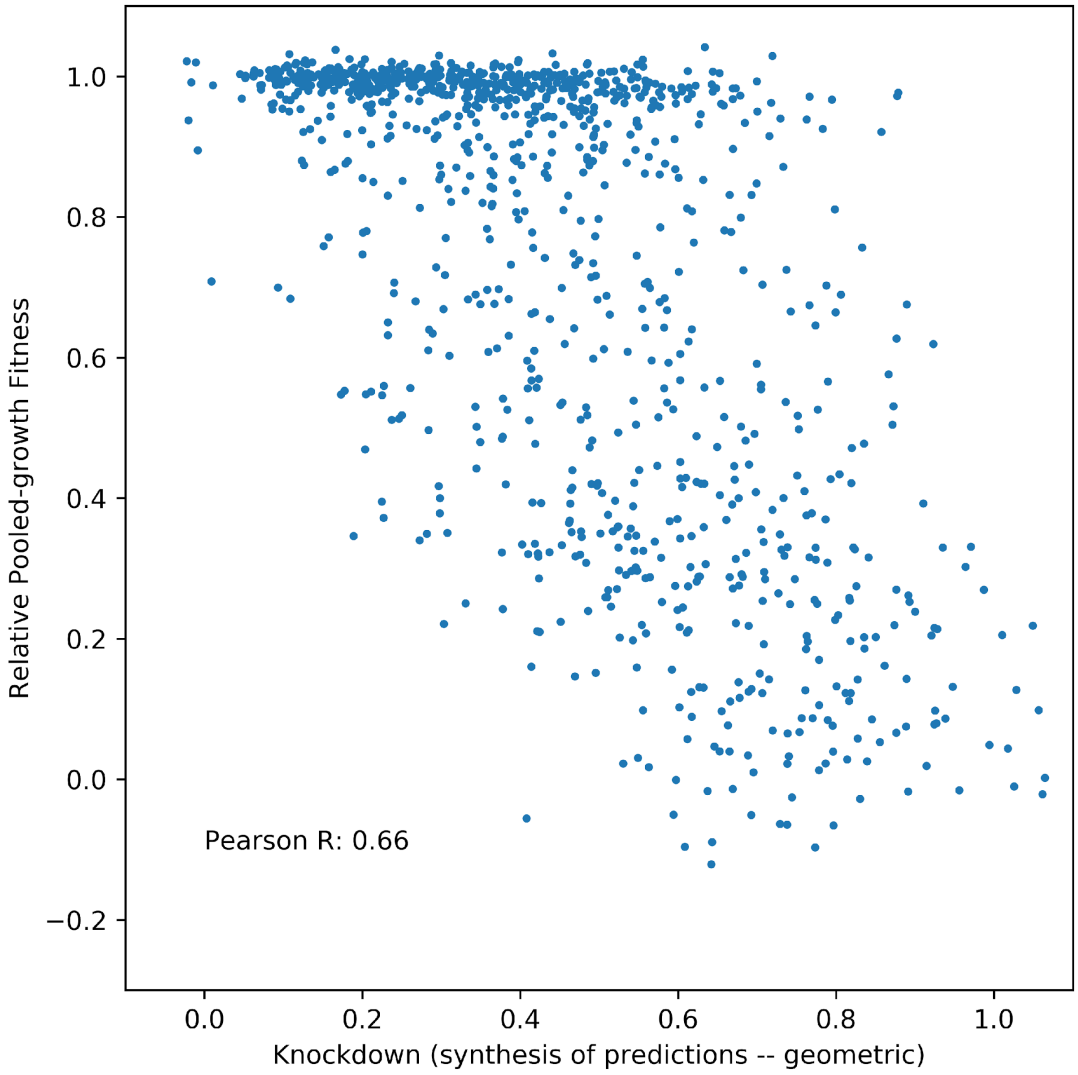


Figure 8.5
KD Prediction vs. Fitness (Bsu/dfrA) [30 ng/mL trimethoprim]

Data as for **Figure 8.2**, but using fitness measurements from samples with 30 ng/mL trimethoprim.

Appendix #6

sgRNA Design

Before any of the libraries described in this work could be created or used, we needed to design guides to target particular genomic loci. When we began our project, there were no broadly available tools for this purpose, and so I created scripts to process available information about a genome and identify the dCas9 targets available for each gene therein.

In hopes of enabling other researchers to perform this task for their bacterium, I have made this code available as a public-facing github repository at https://www.github.com/traeki/sgrna_design

The code used a clever idea originally conceived by Max Horlbeck to ascribe a Specificity score to each guide. In essence, we used the bowtie alignment tool to find alternate off-target matches for each designed guide. The bowtie tool uses the per-base "phred score" provided by Illumina sequencers to soften its search stringency lest mis-called nucleotides prevent alignments. By treating the list of available PAM-adjacent guide candidates as "reads" and providing a "phred-score" that loosely captures the "targeting importance" of each position in the 20-nucleotide guide, we can induce bowtie to give back off-target match candidates. By setting our tolerance threshold to be progressively more permissive in iterative sweeps, we can identify the candidate guides that become "non-specific" within each band of tolerance, and score the recorded targets accordingly.

Guides were then saved to a file annotated with information about their position, both in absolute terms, and in relation to any coding features (i.e. genes) recorded in the provided genome. Users can then subselect guides for specific genes according to their preferences and needs.

Appendix #7

Mismatch Guide Design

Working with mismatch-CRISPRi adds an obvious layer of complication to the process of designing guide libraries. Furthermore, pooled-growth competition libraries require additional layers of analysis to understand. I created scripts allowing us to model and predict knockdowns for mismatch-CRISPRi guides, choose guides for mismatch-CRISPRi libraries based on our predictive model, and interpret and analyze the sequencing output from pooled-growth experiments using those libraries.

I have made the code, along with our predictive model, available at https://www.github.com/traeki/mismatch_crispri. I hope this will serve not only to document our work, but also to facilitate adoption of these ideas and techniques in other labs.

Conclusions / Future Directions

The work described in the preceding chapters develops and demonstrates tools for systematically compromising and interrogating the function of essential genes in bacteria. Where previous efforts to explore the function of these genes required significant investment of time and effort just to investigate a single gene in vitro, we can now construct and study strain libraries spanning the entire essentialome of an organism. In chapter 2, we explored a variety of assays afforded by such libraries. In chapter 3 we showed that these techniques remain effective in extremely divergent bacterial models. These methods and assays, along with the software tools we provide to simplify their adoption, will permit the rapid interrogation of gene function across the entire genomes of many -- if not most -- bacteria of interest. This work can be carried forward in many directions, several of which present themselves as immediately accessible and interesting.

First and most straightforward is the unification of Chapters 2 & 3. In Chapter 2, we explored the power of secondary environmental stressors such as chemicals or nutrient deficiencies to expose phenotypes otherwise buffered in strains with very low (i.e. uninduced) essential knockdowns. In Chapter 3, we used systematically compromised mismatch libraries to quantify the expression breakpoint beneath which each essential gene first begins to significantly impair growth, and how rapidly growth deteriorates as expression is constrained even further. By performing analogous pooled-growth assays in the context of secondary environmental stressors, we can combine these insights. Not only will this allow us to further refine our genetic network, it will in many cases allow us to flesh out deeper questions about their relationships. In Chapter 3 and

Appendix 3 we discuss a small proof-of-concept using trimethoprim. Pooled-growth assays under very low doses of trimethoprim reveal the expected outcome that dfrA and other genes involved in folate metabolism are much more impaired at much lower levels of knockdown than other knockdowns. Our comparisons of outcomes with and without trimethoprim demonstrate that mismatched sgRNAs can induce expression phenotypes even milder than those achieved by the uninduced system (with perfectly matched sgRNAs) in Chapter 2. By using these more sensitive strains for our measurements of stress interactions, we will be able to make much more refined estimates of the precise degree to which cells are buffering/shielding against specific problems. The trimethoprim assays also demonstrate that the rate of impairment depends both on drug concentration and on degree of knockdown, and most importantly that this interaction differs by gene. We can use these differences to infer not just the similarity or interaction between two genes, but indeed to make and test hypotheses about the quantitative and directional nature of those interactions. Furthermore, as we have already constructed small versions of these libraries (discussed at length in Appendix 4), the necessary growth assays will have much a smaller footprint, and require little to no new effort beyond the execution of the assays themselves.

Another exciting opportunity afforded by this work is epistasis interrogation with double-knockdown strains. Though they code for a specific genetic target, sgRNAs act in trans relative to their own genomic location, and also are quite short compared to the genes they are. This allows installation of two sgRNAs to a single small locus, from which they can subsequently be amplified and sequenced (e.g. with paired-end techniques), identifying two-sgRNAsknockdown strains within a pooled growth context and permitting

experiments analogous to those performed in Chapter 3 using single sgRNA libraries. We scratched the surface of this possibility in Chapter 2, using multiple-knockdown strains to explore redundancies among the PBP family of genes. Working with perfectly matched sgRNAs, however, it is not possible to meaningfully measure knockdown pairs comprising both an essential and a non-essential gene. With induction sufficient to produce a phenotype for a non-essential gene, the phenotype of the essential gene (death) dominates the measurement. What's more, even if two essential genes are examined in the absence of induction -- or two non-essential genes, under full induction -- the actual situational importance of those genes is not binary. Essentiality is a quantitative trait, including "non"-essentiality! Such pairs might work to elucidate some epistatic interactions, in some contexts, but it would be very difficult to control for the possibility that one phenotype would simply overshadow the other. The possible domain of epistasis maps constructed with fully-matched sgRNAs is therefore importantly constrained. Furthermore, essential genes are the connection hubs of genetic networks, and their absence from any effort to explore genetic interactions dramatically diminishes the scope and granularity of inquiry. Mismatch-CRISPRi allows separately tuned modulation of multiple targets in a single cell. sgRNAs can be drawn from pre-measured sets (indeed for libraries in E. coli or B. subtilis they can be drawn directly from those presented in this work), or multiple sgRNAs can be used for each gene. The latter approach of course has significant scale implications, and pairs best with a curated subset of genes. Such a library would produce many measurements for each gene pair. This obviates the need for pre-screening to determine optimal knockdowns, but it also increases the robustness of detection for any but the most fine-grained distinctions.

Negative control sgRNAs of course can (and should) be included in each "axis" of such a two-dimensional knockdown library, both to provide control-control pairs for determination of "wildtype" phenotype and to provide statistically robust measurements of the phenotype for each individual sgRNA in isolation, internal to the experimental context. Using multiple guides per essential gene, where feasible, will afford similar statistical refinement to the measured fitness of paired experimental knockdowns. Double-sgRNA libraries will allow us to even more directly associate pairs of genes, and can even be assayed using competitive pooled growth in the context of another (now, "tertiary") stress condition -- drug, starvation, or otherwise.

Third, all of these techniques can be applied in organisms other than these two specific models. Mismatch-CRISPRi requires the ability to transform both a Cas effector and an sgRNA into a background strain. In some organisms (Mycobacteria smegmatis, for example), dCas9 does not seem to work effectively, and alternate Cas systems, or other tuning, might be required. We nonetheless expect that this system will be extremely effective in a great many bacterial hosts, given the evolutionary dissimilarity already spanned by the two discussed in this document. The software tools we have provided require only basic access to some publicly available software (primarily a python installation and access to the bowtie aligner) and a genbank file for the target organism, and can then provide as comprehensive a library as needed for partnership with S. pyogenes dCas9. (For other Cas systems, the scripts will need to be updated to allow for particulars such as rules governing PAM regions.) Researchers are free to design arbitrarily scaled libraries of sgRNAs, adjusting the genes covered and number of guides included to accommodate their needs and constraints. Tthese libraries must

be ordered, cloned, and transformed into background, and that could be more difficult in some species than in others. But once the library is installed, the actual competitive growth assays should be extremely similar to those describe in Chapter 3, and we think this will dramatically speed forays into functional genomics of organisms where such exploration was previously prohibitively difficult and slow.

While these are certainly not the only applications of mismatch-CRISPRi, we anticipate that they will be more than sufficient to generate an explosion of insights and hypotheses. And while we think such future directions demonstrate the exciting power of the techniques explored in this work, it is our sincere hope that these tools will also be of even more immediate use for comparitively quotidian applications. Researchers already focused on a single gene can separately construct libraries, whether 10 sgRNAs, 100 sgRNAs, or a comprehensive library like those used for dihydrofolate reductase or GFP, which they can in turn apply to focused exploration of narrowly targeted questions and hypotheses. In the timeless words of Emerson, "To know even one life has breathed easier because you lived -- this is to have succeeded." We can think of few outcomes more satisfying than having simplified, accelerated, or eased the work of our brilliant peers, even if only a little.

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: CB* 24 (18): 2149–55.

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 (February): 2006.0008.

Babitzke, P., P. Gollnick, and C. Yanofsky. 1992. “The mtrAB Operon of Bacillus Subtilis Encodes GTP Cyclohydrolase I (MtrA), an Enzyme Involved in Folic Acid Biosynthesis, and MtrB, a Regulator of Tryptophan Biosynthesis.” *Journal of Bacteriology* 174 (7): 2059–64.

Barrangou, Rodolphe. 2012. “RNA-Mediated Programmable DNA Cleavage.” *Nature Biotechnology*.

Barrangou, Rodolphe, Christophe Fremaux, Hélène Deveau, Melissa Richards, Patrick Boyaval, Sylvain Moineau, Dennis A. Romero, and Philippe Horvath. 2007. “CRISPR Provides Acquired Resistance against Viruses in Prokaryotes.” *Science* 315 (5819): 1709–12.

Barreteau, Hélène, Andreja Kovac, Audrey Boniface, Matej Sova, Stanislav Gobec, and Didier Blanot. 2008. “Cytoplasmic Steps of Peptidoglycan Biosynthesis.” *FEMS Microbiology Reviews* 32 (2): 168–207.

Bauer, Christopher R., Shuang Li, and Mark L. Siegal. 2015. “Essential Gene Disruptions Reveal Complex Relationships between Phenotypic Robustness,

Pleiotropy, and Fitness." *Molecular Systems Biology* 11 (1): 773.

Blomen, Vincent A., Peter Májek, Lucas T. Jae, Johannes W. Bigenzahn, Joppe Nieuwenhuis, Jacqueline Staring, Roberto Sacco, et al. 2015. "Gene Essentiality and Synthetic Lethality in Haploid Human Cells." *Science* 350 (6264): 1092–96.

Bolotin, A., B. Quinquis, A. Sorokin, and S. D. Ehrlich. 2005. "Clustered Regularly Interspaced Short Palindrome Repeats (CRISPRs) Have Spacers of Extrachromosomal Origin." *Microbiology* 151: 2551–61.

Borkowski, Olivier, Anne Goelzer, Marc Schaffer, Magali Calabre, Ulrike Mäder, Stéphane Aymerich, Matthieu Jules, and Vincent Fromion. 2016. "Translation Elicits a Growth Rate-Dependent, Genome-Wide, Differential Protein Production in Bacillus Subtilis." *Molecular Systems Biology* 12 (5): 870.

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.

Breslow, David K., Dale M. Cameron, Sean R. Collins, Maya Schuldiner, Jacob Stewart-Ornstein, Heather W. Newman, Sigurd Braun, Hiten D. Madhani, Nevan J. Krogan, and Jonathan S. Weissman. 2008. "A Comprehensive Strategy Enabling High-Resolution Functional Analysis of the Yeast Genome." *Nature Methods* 5 (8): 711–18.

Busiek, Kimberly K., and William Margolin. 2015. "Bacterial Actin and Tubulin Homologs in Cell Growth and Division." *Current Biology: CB* 25 (6): R243–54.

Cameron, D. Ewen, and James J. Collins. 2014. “Tunable Protein Degradation in Bacteria.” *Nature Biotechnology* 32 (12): 1276–81.

Cao, Min, Tao Wang, Rick Ye, and John D. Helmann. 2002. “Antibiotics That Inhibit Cell Wall Biosynthesis Induce Expression of the Bacillus Subtilis sigma(W) and sigma(M) Regulons.” *Molecular Microbiology* 45 (5): 1267–76.

Chen, Chong, Yu Liu, Amy R. Rappaport, Thomas Kitzing, Nikolaus Schultz, Zhen Zhao, Aditya S. Shroff, et al. 2014. “MLL3 Is a Haploinsufficient 7q Tumor Suppressor in Acute Myeloid Leukemia.” *Cancer Cell* 25 (5): 652–65.

Choi, Peter S., and Matthew Meyerson. 2014. “Targeted Genomic Rearrangements Using CRISPR/Cas Technology.” *Nature Communications* 5 (April): 3728.

Christen, Beat, Eduardo Abeliuk, John M. Collier, Virginia S. Kalogeraki, Ben Passarelli, John A. Coller, Michael J. Fero, Harley H. McAdams, and Lucy Shapiro. 2011. “The Essential Genome of a Bacterium.” *Molecular Systems Biology* 7 (August): 528.

Christodoulou, Dimitris, Hannes Link, Tobias Fuhrer, Karl Kochanowski, Luca Gerosa, and Uwe Sauer. 2018. “Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia Coli’s Rapid Response to Oxidative Stress.” *Cell Systems* 6 (5): 569–78.e7.

Churchman, L. Stirling, and Jonathan S. Weissman. 2012. “Native Elongating Transcript Sequencing (NET-Seq).” *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* Chapter 4 (April): Unit 4.14.1–17.

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): R63.

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): 1912.

Czarny, T. L., A. L. Perri, S. French, and E. D. Brown. 2014. “Discovery of Novel Cell Wall-Active Compounds Using P ywaC, a Sensitive Reporter of Cell Wall Stress, in the Model Gram-Positive Bacterium Bacillus Subtilis.” *Antimicrobial Agents and Chemotherapy* 58 (6): 3261–69.

Dai, Xiongfeng, Manlu Zhu, Mya Warren, Rohan Balakrishnan, Vadim Patsalo, Hiroyuki Okano, James R. Williamson, Kurt Fredrick, Yi-Ping Wang, and Terence Hwa. 2016. “Reduction of Translating Ribosomes Enables Escherichia Coli to Maintain Elongation Rates during Slow Growth.” *Nature Microbiology* 2 (December): 16231.

Datsenko, K. A., and B. L. Wanner. 2000. “One-Step Inactivation of Chromosomal Genes in Escherichia Coli K-12 Using PCR Products.” *Proceedings of the National Academy of Sciences of the United States of America* 97 (12): 6640–45.

Defeu Soufo, Hervé Joël, Christian Reimold, Uwe Linne, Tobias Knust, Johannes Gescher, and Peter L. Graumann. 2010. “Bacterial Translation Elongation Factor EF-Tu Interacts and Colocalizes with Actin-like MreB Protein.” *Proceedings of the National Academy of Sciences of the United States of America* 107 (7): 3163–68.

Dekel, Erez, and Uri Alon. 2005. “Optimality and Evolutionary Tuning of the Expression Level of a Protein.” *Nature* 436 (7050): 588–92.

Eames, Matt, and Tanja Kortemme. 2012. “Cost-Benefit Tradeoffs in Engineered Lac Operons.” *Science* 336 (6083): 911–15.

Edelstein, Arthur, Nenad Amodaj, Karl Hoover, Ron Vale, and Nico Stuurman. 2010.

"Computer Control of Microscopes Using μManager." *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* Chapter 14 (October): Unit14.20.

Elbaz, Maya, and Sigal Ben-Yehuda. 2010. "The Metabolic Enzyme ManA Reveals a Link between Cell Wall Integrity and Chromosome Morphology." *PLoS Genetics* 6 (9): e1001119.

Forsyth, R. Allyn, Robert J. Haselbeck, Kari L. Ohlsen, Robert T. Yamamoto, Howard Xu, John D. Trawick, Daniel Wall, et al. 2002. "A Genome-Wide Strategy for the Identification of Essential Genes in Staphylococcus Aureus." *Molecular Microbiology* 43 (6): 1387–1400.

Fransen, Fiona, Kelly Hermans, Maria J. B. Melchers, Claudia C. M. Lagarde, Joseph Meletiadis, and Johan W. Mouton. 2017. "Pharmacodynamics of Fosfomycin against ESBL- And/or Carbapenemase-Producing Enterobacteriaceae." *The Journal of Antimicrobial Chemotherapy* 72 (12): 3374–81.

Gasiunas, Giedrius, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. 2012. "Cas9-crRNA Ribonucleoprotein Complex Mediates Specific DNA Cleavage for Adaptive Immunity in Bacteria." *Proceedings of the National Academy of Sciences of the United States of America* 109 (39): E2579–86.

Gibson, Daniel G., Lei Young, Ray-Yuan Chuang, J. Craig Venter, Clyde A. Hutchison 3rd, and Hamilton O. Smith. 2009. "Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases." *Nature Methods* 6 (5): 343–45.

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 (3): 647–61.

Gilbert, Luke A., Matthew H. Larson, Leonardo Morsut, Zairan Liu, Gloria A. Brar, Sandra E. Torres, Noam Stern-Ginossar, et al. 2013. "CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes." *Cell* 154 (2): 442–51.

Goodman, Andrew L., Nathan P. McNulty, Yue Zhao, Douglas Leip, Robi D. Mitra, Catherine A. Lozupone, Rob Knight, and Jeffrey I. Gordon. 2009. "Identifying Genetic Determinants Needed to Establish a Human Gut Symbiont in Its Habitat." *Cell Host & Microbe* 6 (3): 279–89.

Gräwert, Tobias, Johannes Kaiser, Ferdinand Zepeck, Ralf Laupitz, Stefan Hecht, Sabine Amslinger, Nicholas Schramek, et al. 2004. "IspH Protein of Escherichia Coli: Studies on Iron-Sulfur Cluster Implementation and Catalysis." *Journal of the American Chemical Society* 126 (40): 12847–55.

Hardy, L. W., J. S. Finer-Moore, W. R. Montfort, M. O. Jones, D. V. Santi, and R. M. Stroud. 1987. "Atomic Structure of Thymidylate Synthase: Target for Rational Drug Design." *Science* 235 (4787): 448–55.

Harris, Leigh K., and Julie A. Theriot. 2016. "Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size." *Cell* 165 (6): 1479–92.

Hawkins, John S., Spencer Wong, Jason M. Peters, Ricardo Almeida, and Lei S. Qi. 2015. "Targeted Transcriptional Repression in Bacteria Using CRISPR Interference (CRISPRi)." *Methods in Molecular Biology* 1311: 349–62.

Huberts, Daphne H. E. W., and Ida J. van der Klei. 2010. "Moonlighting Proteins: An Intriguing Mode of Multitasking." *Biochimica et Biophysica Acta* 1803 (4): 520–25.

Hunt, Alison, Joy P. Rawlins, Helena B. Thomaides, and Jeff Errington. 2006.

"Functional Analysis of 11 Putative Essential Genes in Bacillus Subtilis." *Microbiology* 152 (Pt 10): 2895–2907.

Ingolia, Nicholas T., Gloria A. Brar, Silvia Rouskin, Anna M. McGeachy, and Jonathan S. Weissman. 2012. "The Ribosome Profiling Strategy for Monitoring Translation in Vivo by Deep Sequencing of Ribosome-Protected mRNA Fragments." *Nature Protocols* 7 (8): 1534–50.

Jiang, Wenzhi, Huanbin Zhou, Honghao Bi, Michael Fromm, Bing Yang, and Donald P. Weeks. 2013. "Demonstration of CRISPR/Cas9/sgRNA-Mediated Targeted Gene Modification in Arabidopsis, Tobacco, Sorghum and Rice." *Nucleic Acids Research* 41 (20): e188.

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.

Johnson, Eachan O., Emily LaVerriere, Emma Office, Mary Stanley, Elisabeth Meyer, Tomohiko Kawate, James E. Gomez, et al. 2019. "Large-Scale Chemical-Genetics Yields New M. Tuberculosis Inhibitor Classes." *Nature* 571 (7763): 72–78.

Johnson, Jarrod W., Jed F. Fisher, and Shahriar Mobashery. 2013. "Bacterial Cell-Wall Recycling." *Annals of the New York Academy of Sciences* 1277 (January): 54–75.

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*. 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 of the United States of America* 110 (25): E2317–26.

Keren, Leeat, Jean Hausser, Maya Lotan-Pompan, Ilya Vainberg Slutskin, Hadas Alisar, Sivan Kaminski, Adina Weinberger, Uri Alon, Ron Milo, and Eran Segal. 2016. "Massively Parallel Interrogation of the Effects of Gene Expression Levels on Fitness." *Cell* 166 (5): 1282–94.e18.

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-Scale Deletion Libraries for Bacillus Subtilis." *Cell Systems* 4 (3): 291–305.e7.

Lalanne, Jean-Benoît, James C. Taggart, Monica S. Guo, Lydia Herzel, Ariel Schieler, and Gene-Wei Li. 2018. "Evolutionary Convergence of Pathway-Specific Enzyme Expression Stoichiometry." *Cell* 173 (3): 749–61.e38.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and Memory-Efficient Alignment of Short DNA Sequences to the Human Genome." *Genome Biology* 10 (3): R25.

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 (11): 2180–96.

Larson, Matthew H., Rachel A. Mooney, Jason M. Peters, Tricia Windgassen, Dhananjaya Nayak, Carol A. Gross, Steven M. Block, William J. Greenleaf, Robert Landick, and Jonathan S. Weissman. 2014. "A Pause Sequence Enriched at

Translation Start Sites Drives Transcription Dynamics in Vivo." *Science* 344 (6187): 1042–47.

Liu, Xue, Clement Gallay, Morten Kjos, Arnau Domenech, Jelle Slager, Sebastiaan P. van Kessel, Kèvin Knoops, Robin A. Sorg, Jing-Ren Zhang, and Jan-Willem Veening. 2017. "High-Throughput CRISPRi Phenotyping Identifies New Essential Genes in Streptococcus Pneumoniae." *Molecular Systems Biology* 13 (5): 931.

Li, Zhijian, Franco J. Vizeacoumar, Sondra Bahr, Jingjing Li, Jonas Warringer, Frederick S. Vizeacoumar, Renqiang Min, et al. 2011. "Systematic Exploration of Essential Yeast Gene Function with Temperature-Sensitive Mutants." *Nature Biotechnology* 29 (4): 361–67.

Lutz, R., and H. Bujard. 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 Research* 25 (6): 1203–10.

Makarova, K. S., N. V. Grishin, S. A. Shabalina, Y. I. Wolf, and E. V. Koonin. 2006. "A Putative RNA-Interference-Based Immune System in Prokaryotes: Computational Analysis of the Predicted Enzymatic Machinery, Functional Analogies with Eukaryotic RNAi, and Hypothetical Mechanisms of Action." *Biology Direct* 1: 7.

Marraffini, Luciano A., and Erik J. Sontheimer. 2008. "CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA." *Science* 322 (5909): 1843–45.

Meeske, Alexander J., Lok-To Sham, Harvey Kimsey, Byoung-Mo Koo, Carol A. Gross, Thomas G. Bernhardt, and David Z. Rudner. 2015. "MurJ and a Novel Lipid II Flippase Are Required for Cell Wall Biogenesis in Bacillus Subtilis." *Proceedings of*

the National Academy of Sciences of the United States of America 112 (20): 6437–42.

Mengin-Lecreulx, D., L. Texier, M. Rousseau, and J. van Heijenoort. 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.” *Journal of Bacteriology* 173 (15): 4625–36.

Michna, Raphael H., Fabian M. Commichau, Dominik Tödter, Christopher P. Zschiedrich, and Jörg Stülke. 2014. “SubtiWiki-a Database for the Model Organism Bacillus Subtilis That Links Pathway, Interaction and Expression Information.” *Nucleic Acids Research* 42 (Database issue): D692–98.

Moche, M., G. Schneider, P. Edwards, K. Dehesh, and Y. Lindqvist. 1999. “Structure of the Complex between the Antibiotic Cerulenin and Its Target, Beta-Ketoacyl-Acyl Carrier Protein Synthase.” *The Journal of Biological Chemistry* 274 (10): 6031–34.

Mojica, F. J. M., C. Dıez-Villasenor, J. Garcıa-Martınez, and E. Soria. 2005. “Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements.” *Journal of Molecular Evolution* 60: 174–82.

Monds, Russell D., Timothy K. Lee, Alexandre Colavin, Tristan Ursell, Selwyn Quan, Tim F. Cooper, and Kerwyn Casey Huang. 2014. “Systematic Perturbation of Cytoskeletal Function Reveals a Linear Scaling Relationship between Cell Geometry and Fitness.” *Cell Reports* 9 (4): 1528–37.

Myoda, T. T., S. V. Lowther, V. L. Funanage, and F. E. Young. 1984. “Cloning and Mapping of the Dihydrofolate Reductase Gene of Bacillus Subtilis.” *Gene* 29 (1-2):

135–43.

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.

Nicolas, Pierre, Ulrike Mäder, Etienne Dervyn, Tatiana Rochat, Aurélie Leduc, Nathalie Pigeonneau, Elena Bidnenko, et al. 2012. “Condition-Dependent Transcriptome Reveals High-Level Regulatory Architecture in Bacillus Subtilis.” *Science* 335 (6072): 1103–6.

Nomura, M., J. L. Yates, D. Dean, and L. E. Post. 1980. “Feedback Regulation of Ribosomal Protein Gene Expression in Escherichia Coli: Structural Homology of Ribosomal RNA and Ribosomal Protein MRNA.” *Proceedings of the National Academy of Sciences of the United States of America* 77 (12): 7084–88.

Opijnen, Tim van, Kip L. Bodi, and Andrew Camilli. 2009. “Tn-Seq: High-Throughput Parallel Sequencing for Fitness and Genetic Interaction Studies in Microorganisms.” *Nature Methods* 6 (10): 767–72.

Osawa, Masaki, and Harold P. Erickson. 2018. “Turgor Pressure and Possible Constriction Mechanisms in Bacterial Division.” *Frontiers in Microbiology* 9 (January): 111.

Overkamp, Wout, Onur Ercan, Martijn Herber, Antonius J. A. van Maris, Michiel Kleerebezem, and Oscar P. Kuipers. 2015. “Physiological and Cell Morphology Adaptation of Bacillus Subtilis at near-Zero Specific Growth Rates: A Transcriptome Analysis.” *Environmental Microbiology* 17 (2): 346–63.

Paradis-Bleau, Catherine, George Kritikos, Katya Orlova, Athanasios Typas, and

Thomas G. Bernhardt. 2014. “A Genome-Wide Screen for Bacterial Envelope Biogenesis Mutants Identifies a Novel Factor Involved in Cell Wall Precursor Metabolism.” *PLoS Genetics* 10 (1): e1004056.

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.

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* 4 (2): 244–50.

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* 27 (October): 121–26.

Pourcel, C., G. Salvignol, and G. Vergnaud. 2005. “CRISPR Elements in Yersinia Pestis Acquire New Repeats by Preferential Uptake of Bacteriophage DNA, and Provide Additional Tools for Evolutionary Studies.” *Microbiology* 151: 653–63.

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.

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–

58.e10.

Rest, Joshua S., Christopher M. Morales, John B. Waldron, Dana A. Opulente, Julius Fisher, Seungjae Moon, Kevin Bullaughey, Lucas B. Carey, and Demitri Dedousis. 2013. “Nonlinear Fitness Consequences of Variation in Expression Level of a Eukaryotic Gene.” *Molecular Biology and Evolution* 30 (2): 448–56.

Rodionov, Dmitry A., Alexey G. Vitreschak, Andrey A. Mironov, and Mikhail S. Gelfand. 2003. “Regulation of Lysine Biosynthesis and Transport Genes in Bacteria: Yet Another RNA Riboswitch?” *Nucleic Acids Research* 31 (23): 6748–57.

Rousset, François, 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* 14 (11): e1007749.

Sanders, Glenn M., H. Garry Dallmann, and Charles S. McHenry. 2010. “Reconstitution of the B. Subtilis Replisome with 13 Proteins Including Two Distinct Replicases.” *Molecular Cell* 37 (2): 273–81.

Sapranauskas, Rimantas, Giedrius Gasiunas, Christophe Fremaux, Rodolphe Barrangou, Philippe Horvath, and Virginijus Siksnys. 2011. “The Streptococcus Thermophilus CRISPR/Cas System Provides Immunity in Escherichia Coli.” *Nucleic Acids Research* 39 (21): 9275–82.

Schaechter, M., O. Maaloe, and N. O. Kjeldgaard. 1958. “Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Grown of Salmonella Typhimurium.” *Journal of General Microbiology* 19 (3): 592–606.

Scheffers, Dirk-Jan, Laura J. F. Jones, and Jeffery Errington. 2004. “Several Distinct Localization Patterns for Penicillin-Binding Proteins in Bacillus Subtilis.” *Molecular*

Microbiology 51 (3): 749–64.

Scheffers, Dirk-Jan, and Mariana G. Pinho. 2005. “Bacterial Cell Wall Synthesis: New Insights from Localization Studies.” *Microbiology and Molecular Biology Reviews: MMBR* 69 (4): 585–607.

Schujman, G. E., K. H. Choi, S. Altabe, C. O. Rock, and D. de Mendoza. 2001. “Response of Bacillus Subtilis to Cerulenin and Acquisition of Resistance.” *Journal of Bacteriology* 183 (10): 3032–40.

Schwank, Gerald, Bon-Kyoung Koo, Valentina Sasselli, Johanna F. Dekkers, Inha Heo, Turan Demircan, Nobuo Sasaki, et al. 2013. “Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients.” *Cell Stem Cell* 13 (6): 653–58.

Scott, Matthew, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. 2010. “Interdependence of Cell Growth and Gene Expression: Origins and Consequences.” *Science* 330 (6007): 1099–1102.

Scott, Matthew, Stefan Klumpp, Eduard M. Mateescu, and Terence Hwa. 2014. “Emergence of Robust Growth Laws from Optimal Regulation of Ribosome Synthesis.” *Molecular Systems Biology* 10 (August): 747.

Shannon, Paul, Andrew Markiel, Owen Ozier, Nitin S. Baliga, Jonathan T. Wang, Daniel Ramage, Nada Amin, Benno Schwikowski, and Trey Ideker. 2003. “Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks.” *Genome Research* 13 (11): 2498–2504.

Shiomi, Daisuke, Masako Sakai, and Hironori Niki. 2008. “Determination of Bacterial Rod Shape by a Novel Cytoskeletal Membrane Protein.” *The EMBO Journal* 27

(23): 3081–91.

Sliusarenko, Oleksii, Jennifer Heinritz, Thierry Emonet, and Christine Jacobs-Wagner. 2011. “High-Throughput, Subpixel Precision Analysis of Bacterial Morphogenesis and Intracellular Spatio-Temporal Dynamics.” *Molecular Microbiology* 80 (3): 612–27.

Storz, Gisela, Jörg Vogel, and Karen M. Wassarman. 2011. “Regulation by Small RNAs in Bacteria: Expanding Frontiers.” *Molecular Cell* 43 (6): 880–91.

Studier, F. William. 2005. “Protein Production by Auto-Induction in High Density Shaking Cultures.” *Protein Expression and Purification* 41 (1): 207–34.

Sugimoto, N., I. Yasumatsu, and M. Fujimoto. 1997. “Stabilities of Internal rU-dG and rG-dT Pairs in RNA/DNA Hybrids.” *Nucleic Acids Symposium Series*, no. 37: 199–200.

Szklarczyk, Damian, Andrea Franceschini, Stefan Wyder, Kristoffer Forslund, Davide Heller, Jaime Huerta-Cepas, Milan Simonovic, et al. 2015. “STRING v10: Protein-Protein Interaction Networks, Integrated over the Tree of Life.” *Nucleic Acids Research* 43 (Database issue): D447–52.

Tenaillon, Olivier, Alejandra Rodríguez-Verdugo, Rebecca L. Gaut, Pamela McDonald, Albert F. Bennett, Anthony D. Long, and Brandon S. Gaut. 2012. “The Molecular Diversity of Adaptive Convergence.” *Science* 335 (6067): 457–61.

Thomaides, Helena B., Ella J. Davison, Lisa Burston, Hazel Johnson, David R. Brown, Alison C. Hunt, Jeffery Errington, and Lloyd Czaplewski. 2007. “Essential Bacterial Functions Encoded by Gene Pairs.” *Journal of Bacteriology* 189 (2): 591–602.

Thomason, Lynn C., Nina Costantino, and Donald L. Court. 2007. “E. Coli Genome

Manipulation by P1 Transduction." Edited by Lynn C. Thomason, Nina Costantino, and Donald L. Court. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* 2 (July): 1.17.1–1.17.8.

Thomason, Lynn C., James A. Sawitzke, Xintian Li, Nina Costantino, and Donald L. Court. 2014. "Recombineering: Genetic Engineering in Bacteria Using Homologous Recombination." Edited by Frederick M. Ausubel, Roger Brent, Robert E. Kingston, David D. Moore, J. G. Seidman, John A. Smith, and Kevin Struhl. *Current Protocols in Molecular Biology / Edited by Frederick M. Ausubel ... [et Al.]* 47 (April): 1.16.1–1.16.39.

Torres, R., M. C. Martin, A. Garcia, Juan C. Cigudosa, J. C. Ramirez, and S. Rodriguez-Perales. 2014. "Engineering Human Tumour-Associated Chromosomal Translocations with the RNA-Guided CRISPR-Cas9 System." *Nature Communications* 5 (June): 3964.

Turner, Robert D., Waldemar Vollmer, and Simon J. Foster. 2014. "Different Walls for Rods and Balls: The Diversity of Peptidoglycan." *Molecular Microbiology* 91 (5): 862–74.

Ursell, Tristan S., Jeffrey Nguyen, Russell D. Monds, Alexandre Colavin, Gabriel Billings, Nikolay Ouzounov, Zemer Gitai, Joshua W. Shaevitz, and Kerwyn Casey Huang. 2014. "Rod-like Bacterial Shape Is Maintained by Feedback between Cell Curvature and Cytoskeletal Localization." *Proceedings of the National Academy of Sciences of the United States of America* 111 (11): E1025–34.

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.

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." *Nature Communications* 9 (1): 2475.

Wang, Tim, Kıvanç Birsoy, Nicholas W. Hughes, Kevin M. Krupczak, Yorick Post, Jenny J. Wei, Eric S. Lander, and David M. Sabatini. 2015. "Identification and Characterization of Essential Genes in the Human Genome." *Science* 350 (6264): 1096–1101.

Wang, Tim, Jenny J. Wei, David M. Sabatini, and Eric S. Lander. 2014. "Genetic Screens in Human Cells Using the CRISPR-Cas9 System." *Science* 343 (6166): 80–84.

Wei, Yuping, Teresa Havasy, Derrell C. McPherson, and David L. Popham. 2003. "Rod Shape Determination by the Bacillus Subtilis Class B Penicillin-Binding Proteins Encoded by pbpA and pbpH." *Journal of Bacteriology* 185 (16): 4717–26.

Winzeler, E. A., D. D. Shoemaker, A. Astromoff, H. Liang, K. Anderson, B. Andre, R. Bangham, et al. 1999. "Functional Characterization of the S. Cerevisiae Genome by Gene Deletion and Parallel Analysis." *Science* 285 (5429): 901–6.

Wolff, Murielle, Myriam Seemann, Bernadette Tse Sum Bui, Yves Frapart, Denis Tritsch, Ana Garcia Estrabot, Manuel Rodríguez-Concepción, Albert Boronat, Andrée Marquet, and Michel Rohmer. 2003. "Isoprenoid Biosynthesis via the Methylerythritol Phosphate Pathway: The (E)-4-Hydroxy-3-Methylbut-2-Enyl Diphosphate Reductase (LytB/IspH) from Escherichia Coli Is a [4Fe-4S] Protein."

FEBS Letters 541 (1-3): 115–20.

Wu, Yuxuan, Dan Liang, Yinghua Wang, Meizhu Bai, Wei Tang, Shiming Bao, Zhiqiang Yan, Dangsheng Li, and Jinsong Li. 2013. “Correction of a Genetic Disease in Mouse via Use of CRISPR-Cas9.” *Cell Stem Cell* 13 (6): 659–62.

Xu, H. Howard, John D. Trawick, Robert J. Haselbeck, R. Allyn Forsyth, Robert T. Yamamoto, Rich Archer, Joe Patterson, et al. 2010. “Staphylococcus Aureus TargetArray: Comprehensive Differential Essential Gene Expression as a Mechanistic Tool to Profile Antibacterials.” *Antimicrobial Agents and Chemotherapy* 54 (9): 3659–70.

Xu, Ling, Svetlana E. Sedelnikova, Patrick J. Baker, Alison Hunt, Jeff Errington, and David W. Rice. 2007. “Crystal Structure of S. Aureus YlaN, an Essential Leucine Rich Protein Involved in the Control of Cell Shape.” *Proteins* 68 (2): 438–45.

Zhang, F. 2012. Systems Methods and Compositions for Sequence Manipulation. U. S. Provisional Patent Application 61/736,527, filed. *Patent*, issued December 12, 2012.

Zheng, Yi, Douglas K. Struck, Thomas G. Bernhardt, and Ry Young. 2008. “Genetic Analysis of MraY Inhibition by the phiX174 Protein E.” *Genetics* 180 (3): 1459–66.

Zhou, Yuexin, Shiyou Zhu, Changzu Cai, Pengfei Yuan, Chunmei Li, Yanyi Huang, and Wensheng Wei. 2014. “High-Throughput Screening of a CRISPR/Cas9 Library for Functional Genomics in Human Cells.” *Nature* 509 (7501): 487–91.

Zwietering, M. H., I. Jongenburger, F. M. Rombouts, and K. van ’t Riet. 1990. “Modeling of the Bacterial Growth Curve.” *Applied and Environmental Microbiology* 56 (6): 1875–81.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.

DocuSigned by:
John Sumblin Hawkins
0AB7C703C1CF4B9...
Author Signature

11/15/2019
Date