

1 **Systematic genome-wide querying of coding and non-coding**
2 **functional elements in *E. coli* using CRISPRi**

4 **Harneet S. Rishi,^{1,2} Esteban Toro,³ Honglei Liu,⁴ Xiaowo Wang,^{4,5} Lei S. Qi,^{6,7,8} Adam P. Arkin^{3,9,*}**

6 ¹ Biophysics Graduate Group, University of California - Berkeley, Berkeley, CA, 94720, USA

7 ² Designated Emphasis Program in Computational and Genomic Biology, University of California -
8 Berkeley, Berkeley, CA, 94720, USA

9 ³ Department of Bioengineering, University of California - Berkeley, Berkeley, CA, 94720, USA

10 ⁴ Bioinformatics Division, Center for Synthetic and Systems Biology, Tsinghua National Laboratory for
11 Information Science and Technology

12 ⁵ Department of Automation, Tsinghua University

13 ⁶ Department of Bioengineering, Stanford University, Stanford, CA, 94305, USA

14 ⁷ Department of Chemical and Systems Biology, Stanford University, Stanford, CA, 94305, USA

15 ⁸ Stanford ChEM-H, Stanford University, Stanford, CA 94305, USA

16 ⁹ Environmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory,
17 Berkeley, CA, 94720, USA

19 *Correspondence: aparkin@lbl.gov

21 **ABSTRACT**

23 Genome-wide CRISPR interference (CRISPRi) screens have enabled the high-throughput identification of
24 essential genes in bacteria, yet they have been underutilized in systematically exploring both the coding
25 and non-coding genome. Here we perform CRISPRi screens in *Escherichia coli* targeting ~13,000 coding
26 and non-coding features using a ~33,000-member sgRNA library, which represents the most compact and
27 comprehensive genome-wide CRISPRi library in *E. coli* to date. We first leverage time-series
28 measurements to dramatically increase the resolution of CRISPRi screens by demonstrating a distinct
29 temporal ordering of essential cellular processes with genes involved in translation exhibiting the most
30 sensitivity to gene knockdown and genes involved in metabolism and cofactor biosynthesis exhibiting
31 among the least sensitivity, implying differences in expression buffering between essential processes. We
32 also query feature essentiality across biochemical conditions, reporting novel conditional phenotypes for
33 genes, small RNAs, and operons. Finally, we systematically target promoters and transcription factor
34 binding sites and (1) add phenotypic confidence to 141 known essential gene promoters and provide the
35 first phenotype-based experimental evidence for four computationally predicted essential gene promoters
36 and (2) refine CRISPRi targeting rules by demonstrating that promoter knockdown efficacy exhibits a strong
37 strand targeting orientation dependency. Together, this work reveals novel genomic feature phenotypes,
38 provides a characterized set of sgRNAs for future *E. coli* screens, and highlights key refinements to
39 CRISPRi screen design for microbial genome characterization.

41 **INTRODUCTION**

43 The nuclease deactivated dCas9 protein has been developed as a powerful tool for programmable gene
44 repression¹, and the ability to induce genetic perturbation at a user-defined time – a feature not available
45 in conventional gene disruption or deletion techniques – has enabled the CRISPRi-mediated
46 characterization of essential genes in a number of bacteria²⁻⁷. The programmability of CRISPRi targeting
47 also enables the interrogation of smaller non-coding DNA (ncDNA) features such as non-coding RNA
48 (ncRNA) genes, promoters, and transcription factor binding sites (TFBSs). ncDNA features, which

49 represent ~12 percent of the *E. coli* genome, play important roles in the regulation of gene expression in a
50 condition-dependent manner. For example, small RNAs (sRNAs) have been implicated in transient
51 regulatory processes involving membrane biogenesis, metabolism, and the synthesis of key transcription
52 factors⁸ while ncDNA regulatory elements drive key physiological decisions such as complex metabolism⁹,
53 pathogenicity¹⁰, and gene expression diversification¹¹. However, ncDNA features have been difficult to
54 perturb using traditional genome-scale methods (e.g. targeted modifications using λ-Red recombination¹²⁻
55¹⁶, random insertions using transposon elements¹⁷⁻²⁰) due to the random targeting of transposons and
56 disruption of local genomic context by insertions, making their interrogation via CRISPRi highly valuable.
57

58 Despite this potential value-add, previous bacterial CRISPRi screening studies have been limited in their
59 investigation of RNA genes beyond simple cases (e.g. tRNA, rRNA genes) and have rarely addressed non-
60 coding genomic features such as promoters and TFBSs. In comparison, CRISPRi screens in eukaryotic
61 systems have been routinely employed to find new regulatory sites in enhancer regions²¹⁻²³ and functionally
62 profile lncRNAs²⁴⁻²⁷, indicating the untapped potential of CRISPRi for the functional characterization of
63 bacterial genomes. In addition, existing CRISPRi screens measure phenotypes using end-point fitness
64 measurements by calculating the change in strain abundance between the beginning and end of a screen,
65 which ignores dynamic outcomes that may occur over the course of an experiment. However, the
66 physiological response resulting from CRISPRi-mediated gene repression could vary between different
67 genes, arising from differences in protein and mRNA decay rates, feedback regulation, interaction network
68 structure, and the physiological relevance of the targeted gene itself. Resolving such end-point
69 measurements by tracking transient responses upon CRISPRi knockdown can yield rich insights into the
70 primacy of processes for fitness and highlight the time it takes to demonstrate a physiological effect in
71 different environments.
72

73 Here we leverage the programmable nature of CRISPRi to target approximately 13,000 *E. coli* MG1655 K-
74 12 genomic features (protein-coding genes, non-coding RNAs (ncRNAs), promoters, and TFBSs) using a
75 compact, designed oligo array library of 32,992 sgRNAs. We first validated our technology by showing that
76 we could knock down 90% of essential genes (as annotated by the Profiling of *E. coli* Chromosome - PEC
77 - database^{28,29}) in a pooled screen with the entire library. Through this process, we showed that a designed,
78 compact library with ~4 guides/gene is sufficient for probing gene essentiality, which represents a
79 considerable reduction in comparison to a previous designed *E. coli* screening study using 15 guides/gene⁴.
80 Given that gene essentiality is context-dependent, we expected that querying essentiality under a variety
81 of biochemical conditions would allow us to delineate between a core set of essential genes and an
82 accessory set of conditionally-essential genes. We thus leveraged the inducible nature of CRISPRi to
83 propagate strains targeting essential genomic features and assay the library in several conditions to find
84 condition-dependent phenotypes for essential genes and ncRNAs. Having validated the library, we next
85 investigated if endpoint phenotypes could be further resolved by investigating how different genomic
86 features respond upon CRISPRi induction. We used time-series measurements to track the dynamic
87 response of genes in our library to CRISPRi perturbation and showed that essential genes exhibited distinct
88 profiles that were correlated with their physiological function – a phenomenon not reported from previous
89 CRISPRi screens due to their use of only endpoint measurements of fitness.
90

91 Finally, we studied the physiological effects of perturbing DNA regulatory elements such as promoters and
92 TFBSs as these features have been understudied in previous bacterial CRISPRi screens. We showed that
93 targeting promoters of essential genes could knock down gene expression and used this phenotypic
94 outcome to add annotation strength to RegulonDB promoters. We also showed that perturbing gene
95 expression was more successful when inhibiting transcription elongation (gene targeting CRISPRi) as
96 opposed to inhibiting transcription initiation (promoter targeting CRISPRi) in our library through a
97 comparison of guides targeting the promoter and gene sequences of known essential genes. By analyzing

98 differences in sgRNA design features and the genomic context of targeted promoters, we found targeting
99 the non-template strand of the promoter closest to the target gene was more effective in knocking down
100 gene expression than other promoter targeting orientations, indicating a new design consideration for
101 promoter CRISPRi. Finally, we looked at the effect of dCas9 targeting to TFBSs to see if TFBS-targeted
102 CRISPRi could perturb gene expression. We analyzed TFBSs regulating promoters of essential genes;
103 however, due to the proximity or overlap of targeted TFBSs with promoters we were unable to associate
104 phenotypes to specific TFBS features in most cases – finding only one case of a condition-dependent
105 phenotype for a TFBS cluster regulating the expression of a conditionally-essential aerobic respiration
106 gene. Together, this work represents an extension and characterization of bacterial CRISPRi screens as
107 well as a framework for the design, construction, pooled screening, and analysis of CRISPRi libraries for
108 the high-throughput functional annotation of bacterial genomic features.
109

110 RESULTS

112 Design and Construction of CRISPRi Library

113 We designed a CRISPRi library consisting of 32,992 unique sgRNAs to target 4,457 genes (including 130
114 small RNAs; sRNAs) and gene-like elements (e.g. insertion elements / prophages), 7,442 promoters and
115 transcription start sites (TSSs), and 1,060 transcription factor binding sites (TFBS) across the *E. coli* K-12
116 MG1655 genome using bioinformatic and biophysical design constraints (Fig. 1a, see [Supplementary Note](#)
117 1 for design details, [Supplementary Table 1a](#) for sequences). In brief, guides were designed to target
118 proximal to a protospacer adjacent motif (PAM) site (NGG for *S. pyogenes* dCas9 used in this work), target
119 a unique genomic sequence, maintain the secondary structure of the sgRNA, and avoid extreme GC
120 content (GC < 20%, GC > 80%), where possible. Gene-targeting guides were designed to target the non-
121 template strand and target close to the beginning of the gene, following previous observations^{1,30}. When
122 possible, multiple guides were designed for each feature. The designed sgRNAs were synthesized as an
123 oligo pool ([Supplementary Table 1b](#)). To allow for the screening of smaller, more focused libraries the
124 terminal 3' end of each oligo was designed with a category code that allows for the amplification of subsets
125 from the oligo library (Fig. 1b, [Supplementary Table 1c](#)). To construct the genome-wide library, sgRNAs
126 were PCR amplified from the oligo pool and then cloned into an expression vector using a golden gate
127 assembly strategy ([Methods](#)). This expression vector (ColE1 origin) maintains the guides under arabinose-
128 inducible control using a pBad promoter. The sgRNA library assembly was transformed into a strain
129 harboring a genetically-encoded dCas9 under aTc-inducible control using a pTet promoter.
130

131 The identity of each knockdown strain in the library is determined solely by the sgRNA plasmid it harbors,
132 specifically the 20-base pair variable region of the sgRNA that directs dCas9 targeting and encodes a DNA
133 barcode for the strain. The relative abundance of every sgRNA, and by extension every strain, can be
134 measured by amplicon sequencing of the variable sgRNA region from a plasmid DNA extraction of the
135 sgRNA library. To perform a pooled functional screen the library is induced to express dCas9 and the
136 sgRNAs and grown under selection for a short period of time (e.g. 24 population doublings) in a user-
137 defined experimental condition (Fig. 1c, [Methods](#)). During this competition, strains that carry an sgRNA
138 targeting a feature important for growth will decrease in abundance in the pool. This phenotypic outcome
139 can be quantified by measuring the starting and ending frequency of each strain and calculating a fitness
140 score, which is defined as the normalized log2 ratio of the relative abundance of the guide-strain after the
141 experiment to before the experiment ([Methods](#)). For gene targeting guides, we also define a composite
142 gene fitness score as the median of fitness scores for all guides targeting a gene.
143

144 Technology Validation of Genome-Wide CRISPRi Gene Knockdowns

145 To assess the ability of the library to yield biologically meaningful results, we profiled the phenotypic effect
146 of knockdown for all genes in the library via a fitness experiment in LB Lennox rich media (LB). We found

147 that CRISPRi was highly reproducible (Pearson $r_{biological} = 0.90$, $p < 0.05$, permutation test; Pearson $r_{technical} = 0.96$, $p < 0.05$, permutation test) ([Supplementary Fig. 1](#)). Furthermore, we observed that sgRNAs targeting known essential genes were severely depleted (i.e. strains harboring these guides exhibited a strong growth defect) over the course of an experiment when compared with sgRNAs targeting non-essential genes ([Fig. 2a](#)). We compared the fitness results with the Profiling of E. coli Chromosome (PEC) database, which reports 304 E. coli K-12 MG1655 genes for which a knockout could not be generated, implying that these genes were essential for growth in LB rich medium under aerobic conditions (i.e. the condition of library construction)^{28,29}. sgRNAs targeting 274 of 303 (~90%) essential genes were severely depleted (composite gene fitness ≤ -2) over the course of CRISPRi fitness experiments in the same condition, yielding 90 percent agreement with the PEC database. This also included proper depletion of all essential E. coli ncRNAs assayed in the experiment as well. Of the remaining 29 essential genes, 15 had at least one sgRNA with fitness ≤ -2 and an additional six had at least one sgRNA with fitness ≤ -1 ([Supplementary Table 2](#)). Overall, we found that 289 of 303 essential genes (~95%) could be knocked down by at least one designed sgRNA with fitness ≤ -2 , indicating high activity of the CRISPRi library. We also tested the library in M9 minimal medium (M9) under aerobic conditions and found that 385 out of 415 (93%) minimal media essential genes had a gene fitness score ≤ -2 when knocked down ([Supplementary Fig. 2](#)).

164

165 We also measured the tightness of inducible control for the CRISPRi library by growing it with no inducer
166 (i.e. no aTc or arabinose added to turn on expression of dCas9 and sgRNA) for the same period of time as
167 a regular fitness experiment (24 population doublings). Strains with essential gene-targeting sgRNAs
168 exhibited a negligible growth defect in this uninduced condition (with gene fitness scores near 0), and the
169 fitness defect of essential gene strains was significantly different between this uninduced case and an
170 induced case ($p < 0.001$, Mann-Whitney U-test; Cohen's d effect size = 3.7) ([Fig. 2b](#)). This suggested that
171 library strains with sgRNAs targeting essential genomic features can be maintained when the library is
172 propagated in an uninduced state.

173

174 We also checked if fitness was biased by factors such as position of targeting relative to chromosomal
175 origin, GC content of the sgRNA, or chromosomal strand of the targeted gene and found no significant
176 correlation ([Supplementary Fig. 3](#)). In agreement with prior reports of CRISPRi in bacteria^{2,4-6}, we found
177 CRISPRi-mediated polar operon effects where knockdown of an upstream nonessential gene in an
178 essential gene containing operon produced a growth defect similar to the essential gene itself, indicating
179 that CRISPRi can knockdown entire operons ([Fig. 2c](#)). Out of 160 operons containing at least one essential
180 gene targeted in our library, we focused on 47 operons where the essential gene was not the first gene in
181 the operon to assess the prevalence of polar operon effects. We found operon effects to be highly prevalent,
182 with every non-essential gene (based on PEC database) upstream of the essential gene in 38 out of the 47
183 operons exhibiting a growth defect when targeted with dCas9 ([Fig. 2d](#)).

184

185

186 **CRISPRi Screening of Essential Genes Under Various Environmental Conditions**

187 To evaluate whether CRISPRi could assess feature fitness in a condition-specific manner, we compared
188 feature enrichment in the library by varying two physiologically relevant parameters – nutrient availability
189 and oxygen availability. In the case of nutrient availability, we profiled the CRISPRi library in M9 media, M9
190 media supplemented with casamino acids (M9Ca), and LB media under aerobic growth conditions. In the
191 case of oxygen availability, we profiled the CRISPRi library in LB media under aerobic and anaerobic growth
192 conditions.

193

194 We first compared enrichment between varied nutrient availability conditions (LB, M9Ca, M9). As previously
195 discussed, we saw a strong depletion of sgRNAs targeting known essential genes (based on knockout

196 studies) in LB and M9 media. We next analyzed non-essential genes that should exhibit condition-
197 dependent phenotypes between these conditions by comparing the enrichment of known amino acid
198 metabolism genes for expected auxotrophic phenotypes. We found a strong depletion of guides targeting
199 genes involved in amino acid biosynthesis in the amino acid-deficient medium (M9) but not the
200 supplemented medium (M9Ca), indicating that CRISPRi can enrich for conditionally essential genes
201 ([Supplementary Fig. 4](#)).

202 Finally, we looked beyond phenotypes for protein-coding genes and analyzed sRNA feature enrichment.
203 Out of the 130 sRNAs with designed guides in the library, we had fitness data for 114 in each condition
204 (some sRNAs did not have data due to low read depth in one or more conditions). Of these 114 sRNAs,
205 we found novel phenotypes for the *hok/sok* Type I toxin-antitoxin (TA) system, which has been implicated
206 in bacterial persistence through the stringent response^{31,32}. Specifically, under stress or amino acid
207 starvation, (p)pGpp and Obg induce (via an unknown mechanism) expression of the *hokB* toxin gene, which
208 leads to membrane depolarization and persistence³³. In our CRISPRi screens, a knockdown of the *sokB*
209 antitoxin sRNA gene resulted in a successively stronger growth defect in LB, M9Ca, and M9 media
210 ([Supplementary Fig. 5](#)), likely due to its inability to inactivate the *hokB* toxin gene product under conditions
211 where it is expressed. The related *hokC-sokC* system exhibited a similar, yet even stronger, response to
212 the knockdown of antitoxin *sokC*. Previous literature has suggested that *hokC* is likely inactive due to an
213 insertion element located 22 bp downstream of the *hokC* reading frame³⁴. However, the *sokC* antitoxin
214 sRNA exhibits a strong deleterious phenotype when knocked down, implying that *hokC* may still be
215 functional. We hypothesize that this phenotype was not seen earlier because the *hokC-sokC* system had
216 only been investigated in nutrient-rich conditions (e.g. LB); however, here we are able to uncover this
217 phenotype by combining the programmability of CRISPRi targeting to investigate this small 55 bp feature
218 with the ability to assess feature fitness across conditions.

219 We next compared enrichment between the aerobically varied conditions, expecting to find condition-
220 specific phenotypes for genes involved in aerobic or anaerobic growth processes. Many strains with guides
221 targeting genes involved in aerobic respiration (eg pyruvate conversion genes, heme biosynthetic genes,
222 ubiquinol biosynthetic genes, cytochrome *bd*-I terminal oxidase subunits, ATP synthase F₁ synthase
223 complex subunits) were depleted in the aerobic condition but dispensable under anaerobic growth ([Fig. 3a](#)). NADH:quinone oxidoreductase I (*nuoABCEFGHIJKLMN*; NDH-1) and NADH:quinone oxidoreductase
224 II (*ndh*; NDH-2) showed a previously unreported phenotype ([Supplementary Fig. 6](#)). NDH-1 only exhibited
225 a defect in aerobic minimal media conditions (M9Ca, M9) while NDH-2 only exhibited a defect in the aerobic
226 rich media condition (LB), implying that NDH-1 may be the dominant oxidoreductase in nutrient-limited
227 conditions and NDH-2 may be dominant in nutrient-rich conditions. We noted that seven genes (*hemB*,
228 *hemC*, *hemD*, *hemH*, *ispB*, *nrdA*, *nrdB*) previously characterized as essential according to the Keio
229 database of essential genes in *E. coli* K-12 BW25113¹² and the PEC database of essential genes in *E. coli*
230 K-12 MG1655 were dispensable for growth under anaerobic conditions ([Fig. 3a](#)). These genes are involved
231 in heme biosynthesis (*hemB*, *hemC*, *hemD*, *hemH*) and ubiquinol biosynthesis (*ispB*), which play critical
232 roles in the aerobic electron transport chain. The essential genes *nrdA* and *nrdB*, which are involved in
233 aerobic nucleotide metabolism^{35,36}, were also dispensable under anaerobic growth. We clonally verified the
234 conditional essentiality of *nrdA* and *hemB* by showing that we could generate viable strains with deletions
235 of these genes under anaerobic conditions and that these deletion strains were not viable under aerobic
236 conditions ([Fig. 3b-c, Supplementary Table 3](#)). By demonstrating that these “essential” genes are only
237 conditionally essential, we show that they are not part of the core, essential genome but instead part of the
238 growth-supporting, conditionally-essential genome. We also noted that of the genes with conditional
239 phenotypes in [Fig. 3a](#), 20 were genes (genes with double asterisks in [Fig. 3a](#)) for which a gene disruption
240 mutant was not generated during a high-throughput transposon insertion screen using Rb-TnSeq due to
241 the attempted construction of the mutants under a condition where the underlying genes were essential.
242

245 We clonally verified one of these genes, *ubiD*, by showing that we could generate a viable deletion strain
246 under the condition determined as permissive via the CRISPRi screen (Fig. 3b-c, [Supplementary Table 3](#)).
247 This analysis presents a proof of concept for the use of two intertwined capabilities of CRISPRi screening
248 – the ability to induce CRISPRi to interrogate features traditionally regarded as essential and the ability to
249 probe feature essentiality across conditions – to delineate between the core, essential and accessory,
250 conditionally-essential genome.

251

252 **Time-series Measurements Elucidate Dynamic Knockdown Response to CRISPRi**

253 We next leveraged the ability to induce CRISPRi perturbations on-demand to probe the dynamic response
254 to knockdown for the library, focusing first on essential genes, and determine if time-series data could yield
255 further measurement resolution into essential gene phenotypes. Specifically, we grew the induced library
256 and sequenced samples at regular intervals over a period of 18 population doublings in LB rich media
257 ([Supplementary Fig. 7](#)). We examined the fitness of strains harboring guides targeting essential genes
258 across the timepoints and found that these strains exhibited successively stronger growth defects over
259 progressive time points (Fig. 4a). We next clustered the essential gene time-series data ([Supplementary](#)
260 [Note 2](#)) and found that essential genes could be classified into one of three groups (Early, Mid, Late) based
261 on their temporal growth trajectory (see [Figure 4b](#) for examples and [Figure 4c](#) for groupings). For example,
262 some essential genes showed a fitness defect soon after the first few population doublings while other
263 genes did not show a defect until several population doublings had occurred. Of the 287 essential genes
264 analyzed, 78 were in the Early group, 114 in the Mid group, and 95 in the Late group ([Supplementary Table](#)
265 [4a](#)).

266

267 We performed a gene ontology enrichment analysis to see if these classes were enriched for specific
268 biological functions ([Supplementary Table 4b](#), [Supplementary Note 2](#)). An analysis with TIGR Role
269 ontologies³⁷ revealed that essential genes in the Early group were significantly enriched for genes involved
270 in ribosomal protein synthesis and modification ($p < 0.001$, p-value from Hypergeometric test followed by
271 FDR correction) with 32 out of 41 essential genes with this TIGR Role present in the Early group. Resource
272 allocation studies in *E. coli* have shown that in rapidly dividing cells ribosomes are most abundant and
273 important for growth³⁸ and haploinsufficiency studies in yeast have shown that ribosomal genes exhibit
274 strong dose responses to gene expression perturbation in rich media³⁹. This would support our finding of
275 ribosomal protein synthesis and modification genes exhibiting a faster physiological response to expression
276 knockdown (via growth defect) relative to other essential genes queried. An analysis of the Mid group
277 revealed a strong enrichment in genes involved in tRNA aminoacylation ($p < 0.001$, Hypergeometric test
278 with FDR correction) with 19 out of 22 essential genes with this TIGR Role present in the Mid group. The
279 presence of tRNA aminoacylation genes in the Mid class also agrees with previous resource allocation
280 studies, which report that the dosage effects observed under exponential growth are present, but less
281 strong, for tRNA genes^{40,41}. Finally, an analysis of the Late group revealed an enrichment of all eight
282 essential genes involved in the 2-C-methyl-D-erythritol 4-phosphate/1-deoxy-D-xylulose 5-phosphate
283 (MEP/DOXP) pathway ($p < 0.05$, Hypergeometric test with FDR correction). The MEP/DOXP pathway⁴²
284 represents the mevalonate-independent pathway for producing the isoprenoid precursors isopentenyl
285 pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP), and its presence in a later, albeit still
286 essential class, in comparison to translation-related genes indicates that the abundance of certain pathway
287 metabolites may not be as rate-limiting to growth in rich media as genes related to translation.

288

289 Because the transcriptional and translational products of genes expressed prior to essential gene
290 knockdown are still present in cells upon CRISPRi induction, we hypothesized that differences in the initial
291 abundance and decay rate of these products may affect the time it takes to observe a measurable fitness
292 defect. We found that genes in the Early group had higher mRNA and protein abundance and were closer
293 to the chromosomal origin than genes in other groups ([Supplementary Fig. 8](#), [Supplementary Table 4c](#));

294 however, these trends were largely driven by the presence of protein synthesis related genes in the Early
295 group and likely represent correlative effects as opposed to causative drivers of group classification.
296

297 We next analyzed all genes targeted in the library to see whether genes classified as non-essential also
298 exhibited varied responses ([Supplementary Note 2](#)). We observed three categories after clustering, two of
299 which contained genes exhibiting a growth defect via the knockdown of both essential and non-essential
300 genes and a third category of genes that did not exhibit a growth defect ([Supplementary Fig. 9](#),
301 [Supplementary Table 4d](#)). Across the two categories of genes exhibiting a defect we saw an enrichment of
302 a number of processes including translation, transcription, aerobic respiration, and fatty acid metabolism
303 ([Supplementary Table 4e](#)).

304
305 The composite nature of the analyzed growth curves meant that the apparent decline in abundance of a
306 given strain could be the result of the slower growth of that strain, the faster growth of another strain, or a
307 combination of the two cases. To distinguish between these cases and validate the trends among essential
308 gene classes, we chose a representative essential gene from each class (Early, Mid, Late), generated
309 individual strains with dCas9 and sgRNAs to separately target these essential genes ([Supplementary Table](#)
310 [5](#)), and used the eVOLVER⁴³, an automated cell culture system, to monitor the temporal knockdown
311 response. We also generated a strain expressing dCas9 along with an sgRNA that did not target any
312 genomic locus to serve as a reference control. We used the eVOLVER as a turbidostat by programming it
313 to keep cells between two optical density (OD) ranges, which allowed us to track changes in doubling time
314 in response to CRISPRi induction. The control sgRNA strain exhibited no change in doubling time after
315 CRISPRi induction ([Supplementary Fig. 10a](#)). In comparing the essential gene-targeting validation strains,
316 we found that *rpsK* (Early gene) was the first to show an increase in doubling time upon induction of
317 CRISPRi, followed by *msbA* (Mid gene) and *folC* (Late gene), thus confirming our observations from the
318 pooled screen ([Fig. 4d](#)). We also found that even within a gene class, different genes could have different
319 profiles. For example, *msbA* showed a progressive increase in doubling time while *ftsZ* (another Mid gene)
320 consistently showed a halt in cell growth after a set number of doublings ([Supplementary Fig. 10b](#)).
321

322 Having validated the trends among essential gene classes, we compared the time-series measurements
323 from the screen to the endpoint measurements (i.e. fitness scores calculated from the initial and final time
324 points in a screen) from an earlier screen. We found that the time-series measurements successfully
325 enabled the further resolution of fitness for 219 essential genes via their stratification into gene classes
326 ([Figure 4e](#)). Together, these results demonstrate (1) the increased measurement resolution provided by
327 time-series measurements in resolving essential gene phenotypes and (2) that while CRISPRi knockdown
328 of an essential gene eventually leads to a fitness defect, different genes can exhibit varied dynamic
329 responses to perturbation, potentially indicating the functional importance of the genes and their biological
330 roles as well as highlighting target considerations for CRISPRi applications where transient dynamics are
331 important (e.g. CRISPRi-based genetic circuits).
332

333 **CRISPRi Screen Uncovers Design Considerations for Non-genic Targeting**

334 *Promoter Interference:* The CRISPRi library contains 14,188 sgRNAs targeting 3,237 promoters and 4205
335 transcription start sites (TSSs) from RegulonDB ([Supplementary Table 6a](#)). To measure the efficacy of
336 CRISPRi targeting for promoters on a genome-scale we assessed whether knockdowns of promoters
337 regulating essential genes produced a growth defect ([Fig. 5a](#)). An analysis of 1,102 sgRNAs targeting 337
338 essential gene promoters across experiments in rich and minimal media ([Supplementary Table 6b](#))
339 revealed that (i) for 74% of essential gene promoters at least 1 sgRNA produced a mild knockdown
340 phenotype (eg Fitness ≤ -1), and (b) for 51% of essential gene promoters, all sgRNAs produced a mild
341 knockdown phenotype. Through this survey, we collected additional experimental phenotypes (i.e. a
342 collection of fitness scores) for 141 known promoter annotations from RegulonDB, which primarily uses

343 RNA-seq as the primary source of experimental characterization for promoters ([Supplementary Table 7](#)).
344 We also found, to the best of our knowledge, the first phenotype-based experimental evidence for four
345 computationally predicted promoters of essential genes ([Supplementary Table 7](#)), highlighting the utility of
346 CRISPRi to improve the annotation strength of non-genic genomic features. We compared the fitness effect
347 of targeting essential gene sequences to that of targeting promoter sequences of essential genes and found
348 that targeting promoters to knockdown gene expression was less efficient than targeting the gene sequence
349 itself ([Fig. 5b](#)). However, we did find cases where promoter-targeting produced a knockdown phenotype
350 similar to gene-targeting knockdowns and where promoter-targeting yielded better knockdown performance
351 than the gene knockdown ([Supplementary Fig. 11](#)), indicating the potential of promoter CRISPRi as an
352 alternative to gene CRISPRi for control over gene expression. We also revisited the time-series data to
353 analyze how promoter CRISPRi compared to gene CRISPRi following a perturbation. To avoid the
354 confounding effects of multiple genes within the same transcriptional unit (TU) and multiple promoters
355 driving the same TU, we focused on 27 monocistronic essential gene TUs regulated by a single promoter
356 ([Supplementary Note 2](#)). We found a strong overlap between the trajectories of the two knockdown
357 implementations ([Supplementary Figure 12](#)), which further indicated the potential of promoter CRISPRi in
358 the presence of well-designed sgRNAs. To elucidate factors contributing to better promoter guide design,
359 we analyzed cases where promoter CRISPRi failed. We noted that 91% of essential promoters targeted by
360 the 334 guides that did not produce a growth defect (Fitness > -1) either were part of a promoter array (i.e.
361 two or more promoters in tandem regulating the same TU) or displayed a strong strand-dependency with
362 respect to knockdown efficiency.

363
364 We hypothesized that for cases where effective promoter knockdown was strongly dependent on the
365 targeted strand, the sgRNAs could be targeting more effective positions within the promoter to interfere
366 with transcriptional initiation or that the local genetic context was influencing knockdown efficacy. In the
367 latter scenario, we hypothesized a model of “transcriptional coupling” where CRISPRi targeting of a
368 promoter on the template strand failed to produce a fitness defect (while targeting the non-template strand
369 could produce a defect) due to its inability to block RNAP readthrough from an upstream transcriptional
370 event. We systematically identified 11 high-confidence cases where targeting the non-template strand
371 produced a growth defect while targeting the template strand did not ([Fig. 5c-d](#), see [Supplementary Table](#)
372 for cases and scoring metrics). One explanation for this result could be the transcriptional overlap of intra-
373 operonic promoters in operons containing multiple TUs (e.g. one TU within a larger TU). Recent reports
374 have also suggested that the transcription boundaries of operons are not as static as previously thought
375 with one study using long-read sequencing (SMRT-Cappable-seq) to demonstrate that 34% of RegulonDB
376 operons can be extended by at least one gene and that 40% of transcription termination sites have read-
377 through that alters operon content⁴⁴. Indeed, of the 11 high-confidence cases, five were TUs contained
378 within larger operons and the remaining six TUs were a part of an extended RegulonDB operon in the
379 SMRT-Cappable seq study ([Supplementary Table 8](#)). We also found an additional 26 cases of medium-
380 confidence ([Supplementary Table 7](#)) that are candidates for this transcriptional coupling that we could not
381 fully confirm either due to an insufficient number of guides available in both targeting orientations to test
382 our strand hypothesis or due to cases where most, but not all, guides produced phenotypes matching the
383 strand hypothesis ([Fig. 5d](#), [Supplementary Table 7](#)). Of these 26 cases, 15 were TUs that were part of
384 larger operons and seven were part of extended RegulonDB operons ([Supplementary Table 8](#)). Overall,
385 our results suggest that targeting the non-template promoter strand can lead to a higher likelihood of
386 successful CRISPRi knockdown for promoters in certain operonic contexts.

387
388 We also found that targeting CRISPRi in promoter arrays can yield distinct phenotypic profiles. Out of 59
389 tandem promoter arrays analyzed in the essential gene promoter data set, we found 40 tandem promoter
390 arrays where we observed one of two distinct phenotypic profiles: (1) all promoters in the array produced a
391 knockdown phenotype or (2) only the downstream promoter produced a fitness defect ([Fig. 5e](#),

392 Supplementary Table 7). In the case where all promoters produced a deleterious knockdown phenotype,
393 we hypothesized that either the most upstream promoter was the primary driver of expression or that all
394 promoters in the array were required for appropriate expression. In the case where only the downstream
395 most promoter showed a deleterious knockdown phenotype, we hypothesized that either the downstream
396 most promoter was the primary expression driver or that all promoters in the array are required for
397 appropriate expression. The remaining 19 tandem promoter arrays analyzed either had an insufficient
398 number of guides to draw any conclusions or were inconsistent with the aforementioned phenotypic profiles
399 (Supplementary Table 7). Overall, our results showed that the promoter closest to the target gene is more
400 likely to yield a knockdown phenotype and thus should be targeted when attempting to knockdown
401 expression of a gene regulated by multiple promoters via promoter CRISPRi.

402 **TFBS Interference:**

403 Finally, we analyzed a set of 1810 sgRNAs in the library that were designed to target 1060 TFBSs on the
404 chromosome (Supplementary Table 9). We first focused on a subset of 175 sgRNAs that targeted 102
405 TFBSs regulating an individual promoter controlling the expression of at least one rich media (based on
406 PEC database) or minimal media (based on Joyce et al *J Bacteriol* 2006) essential gene. We found that
407 most TFBS knockdowns that yielded a deleterious knockdown phenotype were present within the RNAP
408 footprint for promoter binding, which we conservatively defined as between -60 to +20 nt relative to the
409 transcription start site (TSS) associated with the promoter (Supplementary Fig. 13). Due to this overlap, we
410 were unable to specifically associate such phenotypic outcomes to the TFBS alone as they could also be
411 (and likely were) a result of promoter knockdown. Ultimately, we found that it was challenging to parse the
412 phenotypic contribution of TFBSs due to their presence in promoters or binding site arrays with multiple
413 diverse transcription factors.

414 We next looked at all TFBSs that exhibited a growth defect when targeted across all conditions in which
415 the library was assayed. The activating NarL TFBS regulating the *cydDC* promoter, cydDp, exhibited a mild
416 condition-dependent phenotype between aerobic and anaerobic conditions in LB (Supplementary Fig. 14).
417 sgRNAs targeting *cydD*, which plays a role in respiration, and cydDp exhibited a growth defect in an aerobic
418 fitness assay in LB medium but displayed no such defect under anaerobic conditions where no terminal
419 electron acceptor was added and thus no respiration was active. Similarly, an sgRNA targeting the NarL
420 TFBS, which has a positive effect on gene expression for *cydDC* and is situated -126 nt from the cydDp
421 TSS, exhibited a mild growth defect as well (Fitness ~ -1.5) in the aerobic condition and a negligible growth
422 defect (Fitness ~ 0) in the anaerobic condition.

423 **DISCUSSION**

424 In this work we (1) explored key refinements to screen design and profiling using a genome-wide CRISPRi
425 library in *E. coli* and (2) used the CRISPRi platform to phenotypically interrogate the *E. coli* genome. During
426 the preparation of this manuscript, two other studies reported the use of genome-wide CRISPRi libraries to
427 identify essential genes and genes involved in phage-host interactions in *E. coli*^{4,45}. Our work here presents
428 a complementary and extended demonstration of the power of CRISPRi-based approaches to interrogate
429 microbial genomes with the discovery of novel phenotypes for essential genes using a more compact
430 library, application of time-series measurements to track and elucidate phenotypic changes arising after
431 CRISPRi induction, presentation of refined rules for CRISPRi targeting of promoters, and investigation of
432 CRISPRi targeting of TFBSs.

433 We leveraged the inducible nature of CRISPRi to propagate strains with sgRNAs targeting essential
434 genomic features and query them in a number of biochemical contexts, a task unfeasible using conventional
435 gene disruption or knockout approaches. This enabled us to generate 100s of essential gene strains not

441 covered by conventional knockout or Tn-Seq approaches in *E. coli*. Furthermore, we showed that a number
442 of genes classified as essential genes according to classical aerobically generated *E. coli* knockout
443 collections or unable to be assayed using Tn-Seq approaches were actually dispensable under anaerobic
444 conditions, representing a more comprehensive annotation of these genes. We validated the dispensability
445 of three of these genes by showing that we could generate strains with deletions of these genes under the
446 condition they were predicted to be dispensable from the CRISPRi screen.

447
448 We also utilized the inducible nature of CRISPRi to track the effect of knockdown on essential genes post-
449 induction of the CRISPRi machinery. Using time series measurements, we found that different essential
450 gene strains displayed growth defects at distinctly different times, and our results enabled us to classify
451 essential genes into specific categories based on how quickly a given gene's knockdown yielded a
452 measurable fitness defect. The genes in the most essential category had a remarkable overlap with genes
453 discovered to be most essential in other resource allocation studies of *E. coli* in the same condition and
454 also matched gene dosage studies in yeast. During preparation of this manuscript, a separate study
455 leveraged target-mismatched sgRNAs to induce a range of knockdown levels and interrogate the
456 expression-fitness landscape of essential genes in *E. coli* BW25113 (vs MG1655 in this work)⁴⁶. The results
457 of the study were largely complementary to this work: during exponential growth in rich (LB) media,
458 perturbation of genes involved in (1) translation (e.g. ribosomal proteins) produced a strong, linear dose-
459 response, (2) peptidoglycan biosynthesis produced a bimodal response (e.g. *mur* genes involved in Lipid II
460 synthesis exhibited a strong response while *mre* genes involved in peptidoglycan incorporation into the cell
461 wall exhibited a weak response), and (3) cofactor (e.g. heme, coenzyme A, folic acid, riboflavin)
462 biosynthesis produced among the weakest response. While both methods explore the expression-fitness
463 landscape of essential genes, they have distinct value propositions. The mismatch CRISPRi approach
464 maps gene expression levels to fitness, which can yield a dose response curve along with the expression
465 threshold below which an essential gene (and perhaps biological process) can impact cellular physiology.
466 The time-series CRISPRi approach records the time it takes for a physiologically deleterious knockdown of
467 an essential gene to impact cellular growth, which can yield the temporal order in which essential genes
468 (and perhaps biological processes) can impact cellular physiology. Both methods also have their limitations.
469 The time-series CRISPRi approach cannot generate a quantitative gene dosage profile since it only probes
470 one knockdown level. The mismatch CRISPRi approach is limited in that it only measures fitness at one
471 timepoint after induction of the library, and it is difficult to predict a priori if this timepoint is enough to observe
472 the dynamic range of fitness effects across all queried knockdown levels. For example, the majority of
473 cofactor synthesis genes highlighted as exhibiting a weak dose response (i.e. showed negligible change in
474 relative fitness across knockdown levels) in the mismatch CRISPRi work (e.g. *coaADE*, *folADEK*,
475 *ribABCDEF*, *hemABCDGHL*, *fabBGIZ*, *nadDEK*) after 10 population doublings (single endpoint of the work)
476 actually show fitness defects when observed at multiple timepoints up to 18 population doublings in the
477 time-series CRISPRi approach. By leveraging multiple timepoints in this work we were able to temporally
478 resolve these strong fitness effects associated with essential gene perturbation. Despite the limitations of
479 each method, together these approaches demonstrate the importance of CRISPRi dosage and duration as
480 important foundational handles for interrogating the expression-fitness landscape, and given their distinct
481 capabilities we believe end users can find utility in both.

482
483 The programmable nature of CRISPRi targeting also allowed us to interrogate promoters and TFBSs.
484 Specifically, we were able to compare gene-targeted CRISPRi (inhibit transcription elongation) to promoter-
485 targeted CRISPRi (inhibit transcription initiation), finding that gene-targeting CRISPRi largely outperformed
486 promoter-targeting CRISPRi. We also attributed phenotypic evidence to 141 known RegulonDB-annotated
487 promoters and associated, to our knowledge, the first experimental evidence to four predicted promoters
488 from RegulonDB. Finally, we explored phenotypic profiles associated with tandem promoter arrays and
489 promoters that displayed strand-dependent knockdown success to conclude that targeting the NT-strand

490 of the promoter closest to the target gene can yield more successful CRISPRi knockdowns in comparison
491 to other promoter-mediated orientations for certain genomic contexts.

492
493 While we demonstrated a high utility for microbial genome interrogation via CRISPRi-based screens in this
494 work, CRISPRi still has a number of limitations. First, targeting in operons yields polar effects, thus limiting
495 the analysis of essentiality to transcriptional units and assigning specific phenotypic confidence to only the
496 last gene in the transcriptional unit. As such, CRISPRi should serve as a complementary method to
497 transposon insertion and recombineering-based approaches, which are less prone to polar operon effects.
498 Second, the compact organization of bacterial genomes yields architectures with overlapping or tightly
499 spaced TFBS and promoter features. This makes it especially challenging to precisely attribute phenotypes
500 to a specific TFBS (due to its proximity or overlap with other TFBSs and promoters). Precise genome editing
501 methods such as MAGE and CREATE are likely more suitable for such cases. Regardless, the
502 programmability of CRISPRi targeting can be used to uncover intergenic regions of phenotypic importance
503 through tiled screens, which can be combined with TFBS and promoter predictions along with high-
504 throughput measurements (eg protein-DNA interactions, RNA-seq) to add annotation confidence for newly-
505 sequenced microbes. Overall, the CRISPRi library developed here presents a resource of curated and
506 phenotype-linked sgRNAs for use in *E. coli*, and the workflow developed here for interrogating genic and
507 non-genic chromosomal features provides the basis for high-throughput CRISPRi studies in other bacteria.
508

509 METHODS

510
511 **Chemicals, reagents, and media:** LB Lennox Medium (EZMix™ powder microbial growth medium, Sigma
512 Aldrich) was used to culture strains for experiments in rich media. M9 Minimal Medium (1X M9 salts, 2 mM
513 MgSO₄, 0.1 mM CaCl₂, 0.4% glycerol) was used to culture strains for experiments in minimal media.
514 Anhydrotetracycline (aTc; CAS 13803-65-1, Sigma-Aldrich) was used at 200 ng/mL to induce dCas9
515 expression. Arabinose was used at 0.1% to induce sgRNAs. Antibiotic concentrations used were 100 µg/mL
516 for carbenicillin and 30 µg/mL for kanamycin. Glucose was used at 0.2% in media for outgrowth of the library
517 from a freezer aliquot. Casamino acids (0.2%) were also used in M9 Minimal Medium for select assays.
518

519 **CRISPRi library design:** See [Supplementary Note 1](#) for details regarding sgRNA library design along with
520 [Supplementary Table 1](#) for sgRNA feature annotations, sequence-level details, and a summary of
521 category codes.

522
523 **CRISPRi library construction:** To clone the sgRNA library, sgRNAs were amplified from the OLS oligo pool
524 using primers 282 (5' CACATCCAGGTCTCTCCAT 3') and 284 (5'
525 cacatccaggctctCGGACTAGCCTTATTAACTTG 3') using Phusion II HS and the following protocol: 98°C
526 for 10 sec and 15 cycles of 98°C for 10 sec, 60°C for 30 sec, and 72°C for 20 sec followed by a final
527 extension of 72°C for 5 min. The PCR reaction was purified using a Zymo DNA Clean & Concentrator kit
528 and eluted in water. The purified library was cloned into the library receiver plasmid, pT154
529 (<https://benchling.com/s/seq-YGEVpcmWzQjGfRP8oDc>), via a golden gate reaction using Bsal and T7
530 DNA ligase. The golden gate reaction product was purified using a Zymo DNA Clean & Concentrator kit,
531 following the kit parameters for a plasmid cleanup. A derivative of *Escherichia coli* K-12 MG1655 (ET163:
532 MG1655 FRT-kanR-FRT tetR-pTet-dCas9; <https://benchling.com/s/seq-Gxu6IV96FF6y8jcpTrU>) was
533 used as the recipient strain for the sgRNA library. The purified library was electroporated into a competent
534 cell preparation of ET163 and maintained under carbenicillin (plasmid marker) and kanamycin (strain
535 marker) selection. Aliquots of the resulting library were stored at -80°C.
536

537 **CRISPRi fitness experiments:** An aliquot of the library was taken from storage at -80°C and thawed at room
538 temperature. The aliquot was used to inoculate a 5 mL culture of LB Lennox media (LB) with carbenicillin,

539 kanamycin, and glucose (multiple aliquots were used to inoculate distinct cultures for experiments with
540 biological replicates). The culture was grown at 37°C until it reached OD600 0.5. A 4 mL aliquot was taken
541 as an initial time point for the library (t0 sample); this sample was centrifuged (Eppendorf 5810R) at 4000
542 RPM (3202xg) and stored at -80°C. The remaining 1 mL of culture was centrifuged (Eppendorf 5417R) at
543 8000xg and washed twice with 1 mL of LB media. 156 uL of this washed sample was added to 10 mL of
544 LB media (~1:64 dilution) with arabinose (0.1%), aTc (200 ng/mL), carbenicillin (100µg/mL), and kanamycin
545 (30µg/mL). Technical replicates were generated by dividing this initial culture into 5 mL cultures. Cultures
546 were grown at 37°C until they reached OD600 ~0.5, indicating 6 population doublings of the library. The
547 library was again diluted 1:64 into 5 mL of LB media with arabinose, aTc, carbenicillin, and kanamycin and
548 grown at 37°C until the culture reached OD600 ~0.5. This process was repeated until the library had
549 undergone a total of 24 population doublings under induction. After 24 population doublings, the sample
550 was centrifuged (Eppendorf 5810R) at 4000 RPM (3202xg) and stored at -80°C.
551

552 For experiments in minimal media, the original freezer aliquot of the library was inoculated in M9 media
553 with glycerol (0.4%), glucose (0.2%), carbenicillin, and kanamycin. For induction of the CRISPRi system,
554 the library was cultured in M9 media with glycerol, arabinose, aTc, carbenicillin, and kanamycin. Casamino
555 acids (0.2%) were added depending on the assay condition.
556

557 For time-series experiments, samples were collected every doubling after the t0 sample was taken for the
558 first 12 doublings, after which samples were collected every two doublings until the library had undergone
559 a total of 18 doublings. During the experiment, the library was maintained between OD600 ~0.25 and ~0.50.
560

561 **CRISPRi sequencing library preparation:** Frozen, centrifuged samples from fitness experiments were taken
562 from storage at -80°C and thawed at room temperature. The CRISPRi sgRNA library was isolated using a
563 QIAprep® Spin Miniprep Kit. 10-20 ng of DNA from each sample was used for a PCR reaction to generate
564 NGS-ready sequencing samples in a 50 uL reaction using Phusion polymerase and two primers to add one
565 of two sets of indexed Illumina adaptors. The first set contained a constant reverse primer and a variable
566 forward primer with sample-specific 8 nucleotide barcodes that were sequenced “in-line” during an Illumina
567 sequencing read. The second primer set contained a constant forward primer and a variable reverse primer
568 with sample-specific indices that could be sequenced during an indexing read ([Supplementary Table 1d](#)).
569 Both primer sets yielded comparable sequencing results; however, we eventually shifted to using the
570 second primer set as the data could be readily demultiplexed using Illumina software.
571

572 Each reaction was performed using the following protocol: 98°C for 30 sec and 21 cycles of 98°C for 10
573 sec, 67°C for 15 sec, and 72°C for 10 sec followed by a final extension of 72°C for 5 min. 5 uL of each PCR
574 sample was pooled and purified using a Zymo DNA Clean & Concentrator kit. The purified sample was
575 quantified using the Qubit dsDNA HS assay kit and product size was confirmed using a Bioanalyzer 2100
576 automated electrophoresis system (DNA 1000 Kit). Final samples were run on either an Illumina MiSeq or
577 HiSeq instrument (2000/2500; Vincent J. Coates Genomics Sequencing Laboratory, UC Berkeley). All
578 relevant sequencing data have been deposited in the National Institutes of Health (NIH) Sequencing Read
579 Archive (SRA) at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA559958> under Accession code
580 PRJNA559958.
581

582 **CRISPRi sequencing data analysis:** Sequencing runs were demultiplexed using standard Illumina software
583 for samples using the second primer set or a custom python script (demultiplex_fastq.py) for samples using
584 the first primer set. Demultiplexed reads were processed using the following set of custom python scripts:
585 trim_sgRNA_reads.py to trim and filter reads according to quality thresholds; bwa.samtools.py to map the
586 trimmed sgRNA reads to a BWA index of the sgRNA library; parse_bam.py to convert mapped reads to a

587 table of counts that represent the abundance of each sgRNA in the sample. Custom scripts for analysis are
588 available at <https://github.com/hsrishi/HT-CRISPRi>.

589
590 **CRISPRi fitness score calculation:** A small constant (i.e. pseudocount of 1) was added to the raw read
591 counts to avoid errors in calculating fold-change in subsequent fitness calculations due to division by 0.
592 These adjusted read counts for each sample were normalized by the median abundance for that sample,
593 thus generating relative abundance (RA) values for each sgRNA library member and enabling comparisons
594 between different samples. The fitness score was calculated as the log₂ ratio of the RA of a guide strain in
595 a test condition relative to its RA in a control condition. In this framework, the test condition was a sample
596 of the library after being subjected to growth over the course of an experiment, and the control condition
597 was the t0 sample. The fitness scores from each sample were normalized such that the median fitness
598 score for the sample was 0. In practice, library members with t0 raw read counts < 10 were filtered out to
599 limit variability due to low read depth. Significance values for each sgRNA fitness score were calculated via
600 the edgeR package using raw read counts as the input^{47,48}.

601
602 We also created a gene fitness score, which we calculated as the median of fitness values for all sgRNAs
603 targeting a given gene. This provided a more stringent metric for quantifying strong fitness scores. For
604 example, for a given gene with four sgRNAs, at least two guides would have to yield a strong fitness score
605 in order for the median to be lower than -2. Fitness scores for all relevant experimental samples are listed
606 in [Supplementary Table 10](#).

607
608 **AUTHOR CONTRIBUTIONS**
609 H.S.R. led the experimental work, computational analyses, and manuscript preparation. H.S.R., E.T., and
610 A.P.A. designed experiments. E.T. cloned the CRISPRi library and performed initial experiments. H.L. and
611 X.W. designed the sgRNA library. A.P.A. supervised the research. All authors contributed to manuscript
612 preparation. L.S.Q. and A.P.A. conceived of the research.

613
614 **ACKNOWLEDGEMENTS**
615 The authors would like to thank (1) Agilent Technologies for providing the sgRNA library, (2) members of
616 the Arkin lab, especially Vivek Mutalik and Morgan Price, for insightful discussions over the course of the
617 work and during manuscript preparation, and (3) Guillaume Cambray and David Chen for initial help with
618 Illumina NGS read processing. This work used the Vincent J. Coates Genomics Sequencing Laboratory at
619 UC Berkeley. The authors would also like to acknowledge funding sources: Department of Energy Genome
620 Science program - Office of Biological and Environmental Research [DE-SC008812, Funding Opportunity
621 Announcement DE-FOA-0000640]; National Science Foundation (NSF) Graduate Research Fellowship (to
622 H.S.R.); National Institutes of Health (NIH) Genomics and Computational Biology Training Program
623 [5T32HG000047-18] (to H.S.R.). Funding for open access charge: DOE BER [DE-SC008812].

624
625 **COMPETING INTERESTS**
626 The authors declare no competing interests.

627
628 **DATA AVAILABILITY**
629 All relevant sequencing data have been deposited in the National Institutes of Health (NIH) Sequencing
630 Read Archive (SRA) at <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA559958> under Accession code
631 PRJNA559958. The source data underlying all figures (main text and Supplementary Information) are
632 provided as a Source Data file. Figures, Supplementary Information, and Source Data will also be available
633 at <https://github.com/hsrishi/HT-CRISPRi>. All other data available from the authors upon reasonable
634 request.

635

636 **CODE AVAILABILITY**637 Code for processing and analyzing CRISPRi data is available at <https://github.com/hsrishi/HT-CRISPRi>.

638

639 **REFERENCES**

640

- 641 1. Qi, L. S. *et al.* Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of
642 Gene Expression. *Cell* **152**, 1173–1183 (2013).
- 643 2. Peters, J. M. *et al.* A Comprehensive, CRISPR-based Functional Analysis of Essential Genes in
644 Bacteria. *Cell* **165**, 1–39 (2016).
- 645 3. Liu, X. *et al.* High-throughput CRISPRi phenotyping identifies new essential genes in
646 *Streptococcus pneumoniae*. *Molecular Systems Biology* **13**, 931–18 (2017).
- 647 4. Wang, T. *et al.* Pooled CRISPR interference screening enables genome-scale functional genomics
648 study in bacteria with superior performance. *Nature Communications* **9**, 1–15 (2018).
- 649 5. Rousset, F., Cui, L., Siouve, E., Depardieu, F. & Bikard, D. Genome-wide CRISPR-dCas9 screens
650 in *E. coli* identify essential genes and phage host factors. 1–31 (2018). doi:10.1101/308916
- 651 6. de Wet, T. J., Gobe, I., Mhlanga, M. M. & Warner, D. F. CRISPRi-Seq for the Identification and
652 Characterisation of Essential Mycobacterial Genes and Transcriptional Units. 1–24 (2018).
653 doi:10.1101/358275
- 654 7. Lee, H. H. *et al.* Functional genomics of the rapidly replicating bacterium *Vibrio natriegens* by
655 CRISPRi. *Nature Microbiology* **4**, 1105–1113 (2019).
- 656 8. Gottesman, S. & Storz, G. Bacterial small RNA regulators: versatile roles and rapidly evolving
657 variations. *Cold Spring Harbor Perspectives in Biology* **3**, a003798–a003798 (2011).
- 658 9. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & van Oudenaarden, A. Multistability in the
659 lactose utilization network of *Escherichia coli*. *Nature* **427**, 737–740 (2004).
- 660 10. Somvanshi, V. S. *et al.* A single promoter inversion switches *Photobacterium* between pathogenic
661 and mutualistic states. *Science* **337**, 88–93 (2012).
- 662 11. Oren, Y. *et al.* Transfer of noncoding DNA drives regulatory rewiring in bacteria. *Proc. Natl. Acad.
663 Sci. U.S.A.* **111**, 16112–16117 (2014).
- 664 12. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the
665 Keio collection. *Molecular Systems Biology* **2**, 1–11 (2006).
- 666 13. Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution.
667 *Nature* **460**, 894–898 (2009).
- 668 14. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A. & Gill, R. T. Rapid profiling of
669 a microbial genome using mixtures of barcoded oligonucleotides. *Nat Biotechnol* **28**, 856–862
670 (2010).
- 671 15. Freed, E. F. *et al.* Genome-Wide Tuning of Protein Expression Levels to Rapidly Engineer
672 Microbial Traits. *ACS Synth. Biol.* **4**, 1244–1253 (2015).
- 673 16. Garst, A. D. *et al.* Genome-wide mapping of mutations at single-nucleotide resolution for protein,
674 metabolic and genome engineering. *Nat Biotechnol* **35**, 1–12 (2016).
- 675 17. van Opijnen, T., Bodi, K. L. & Camilli, A. Tn-seq: high-throughput parallel sequencing for fitness
676 and genetic interaction studies in microorganisms. *Nat Meth* **6**, 767–772 (2009).
- 677 18. Langridge, G. C. *et al.* Simultaneous assay of every *Salmonella Typhi* gene using one million
678 transposon mutants. *Genome Research* **19**, 2308–2316 (2009).
- 679 19. Wetmore, K. M. *et al.* Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing
680 Randomly Bar-Coded Transposons. *mBio* **6**, e00306–15–15 (2015).
- 681 20. Price, M. N. *et al.* Mutant phenotypes for thousands of bacterial genes of unknown function.
682 *Nature* **557**, 503–509 (2018).
- 683 21. Fulco, C. P. *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR
684 interference. *Science* **354**, 769–773 (2016).
- 685 22. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed Engineering and Analysis of
686 Combinatorial Enhancer Activity in Single Cells. *Molecular Cell* **66**, 285–299.e5 (2017).
- 687 23. Simeonov, D. R. *et al.* Discovery of stimulation-responsive immune enhancers with CRISPR
688 activation. *Nature* **549**, 111–115 (2017).
- 689 24. Zhu, S. *et al.* Genome-scale deletion screening of human long non-coding RNAs using a paired-
690 guide RNA CRISPR-Cas9 library. *Nat Biotechnol* **34**, 1279–1286 (2016).

- 691 25. Liu, S. J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci
692 in human cells. *Science* **355**, eaah7111–16 (2017).
- 693 26. Joung, J. *et al.* Genome-scale activation screen identifies a lncRNA locus regulating a gene
694 neighbourhood. *Nature* **548**, 343–346 (2017).
- 695 27. Gasperini, M. *et al.* A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic
696 Screens. *Cell* **176**, 377–390.e19 (2019).
- 697 28. Kato, J.-I. & Hashimoto, M. Construction of consecutive deletions of the Escherichia coli
698 chromosome. *Molecular Systems Biology* **3**, 966–7 (2007).
- 699 29. Osterman, A. L. & Gerdes, S. Y. *Microbial Gene Essentiality: Protocols and Bioinformatics*. **416**,
700 (Humana Press, 2008).
- 701 30. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene
702 expression. *Nature Protocols* **8**, 2180–2196 (2013).
- 703 31. Gerdes, K. & Maisonneuve, E. Bacterial Persistence and Toxin-Antitoxin Loci. *Annu. Rev.
704 Microbiol.* **66**, 103–123 (2012).
- 705 32. Kint, C. I., Verstraeten, N., Fauvert, M. & Michiels, J. New-found fundamentals of bacterial
706 persistence. *Trends in Microbiology* **20**, 577–585 (2012).
- 707 33. Verstraeten, N. *et al.* Obg and Membrane Depolarization Are Part of a Microbial Bet-Hedging
708 Strategy that Leads to Antibiotic Tolerance. *Molecular Cell* **59**, 9–21 (2015).
- 709 34. Pedersen, K. & Gerdes, K. Multiple hok genes on the chromosome of Escherichia coli. *Mol.
710 Microbiol.* **32**, 1090–1102 (1999).
- 711 35. Fuchs, J. A. & Karlström, H. O. Mapping of nrdA and nrdB in Escherichia coli K-12. *J. Bacteriol.*
712 **128**, 810–814 (1976).
- 713 36. Garriga, X. *et al.* nrdD and nrdG genes are essential for strict anaerobic growth of Escherichia coli.
714 *Biochem. Biophys. Res. Commun.* **229**, 189–192 (1996).
- 715 37. Haft, D. H. *et al.* TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res* **41**, D387–D395
716 (2012).
- 717 38. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in
718 Bacteria. *Cell* **139**, 1366–1375 (2009).
- 719 39. Deutschbauer, A. M. Mechanisms of Haploinsufficiency Revealed by Genome-Wide Profiling in
720 Yeast. *Genetics* **169**, 1915–1925 (2005).
- 721 40. Ardell, D. H. & Kirsebom, L. A. The Genomic Pattern of tDNA Operon Expression in E. coli. *PLoS
722 Comput Biol* **1**, e12–14 (2005).
- 723 41. Couturier, E. & Rocha, E. P. C. Replication-associated gene dosage effects shape the genomes of
724 fast-growing bacteria but only for transcription and translation genes. *Mol. Microbiol.* **59**, 1506–
725 1518 (2006).
- 726 42. Hunter, W. N. The non-mevalonate pathway of isoprenoid precursor biosynthesis. *Journal of
727 Biological Chemistry* **282**, 21573–21577 (2007).
- 728 43. Wong, B. G., Mancuso, C. P., Kiriakov, S., Bashor, C. J. & Khalil, A. S. Precise, automated control
729 of conditions for high-throughput growth of yeast and bacteria with eVOLVer. *Nature Publishing
730 Group* **67**, 1–15 (2018).
- 731 44. Yan, B., Boitano, M., Clark, T. A. & Ettwiller, L. SMRT-Cappable-seq reveals complex operon
732 variants in bacteria. *Nature Communications* **9**, 318–11 (2018).
- 733 45. Rousset, F. *et al.* Genome-wide CRISPR-dCas9 screens in E. coli identify essential genes and
734 phage host factors. *PLoS Genet* **14**, e1007749–28 (2018).
- 735 46. Hawkins, J. S. *et al.* Mismatch-CRISPRi Reveals the Co-varying Expression-Fitness Relationships
736 of Essential Genes in Escherichia coli and Bacillus subtilis. *Cell Systems* (2020).
737 doi:10.1016/j.cels.2020.09.009
- 738 47. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential
739 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2009).
- 740 48. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq
741 experiments with respect to biological variation. *Nucleic Acids Res* **40**, 4288–4297 (2012).
- 742

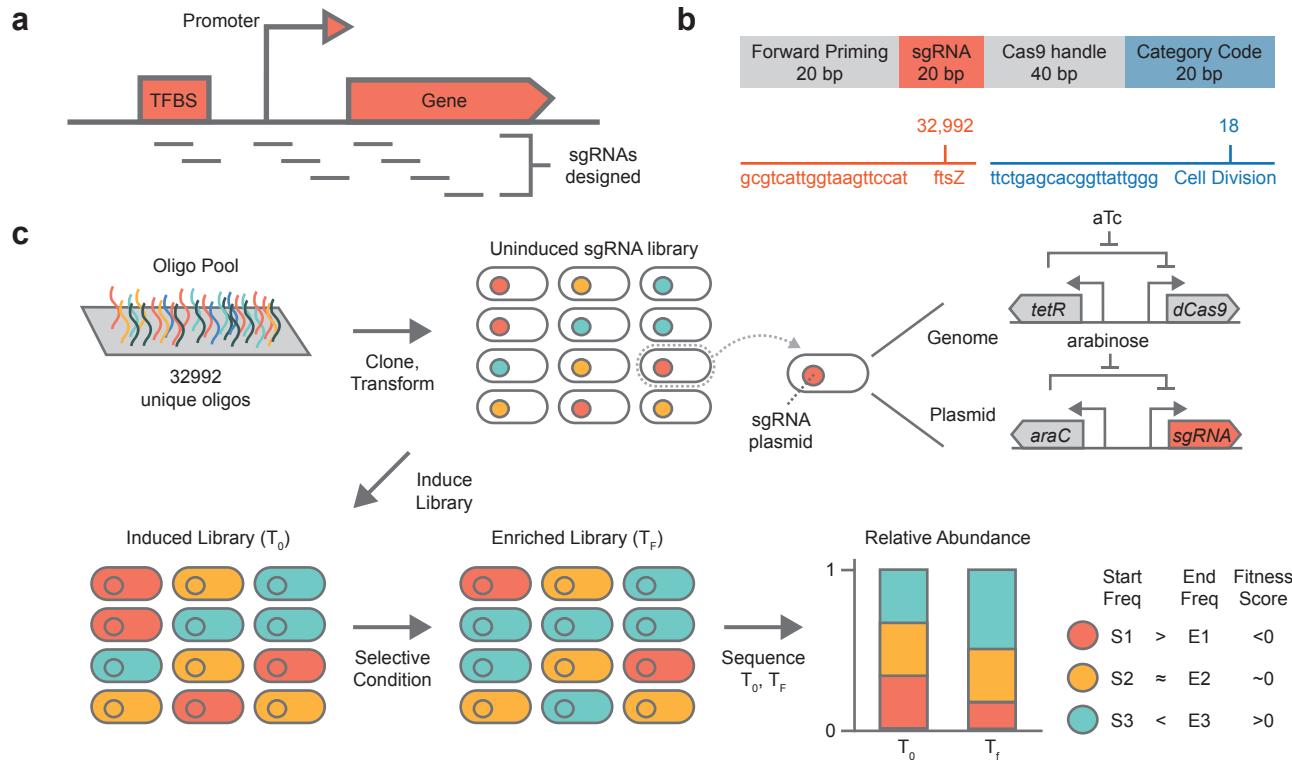


Fig. 1 | Overview of CRISPRi screening platform. a Guide sequences were designed to target three feature types on the *E. coli* genome: (i) gene sequences (ii) promoters (iii) transcription factor binding sites (TFBSs). Multiple guides were designed for each feature where possible (Methods). b Guide sequences were synthesized as oligos and ordered via Agilent Technologies as a pool. Category codes (short DNA barcodes) were included in designed oligos to enable amplification of subpools from the library. c Guides were first cloned into a receiver vector and transformed into a strain containing chromosomally integrated *dCas9*. At the beginning of an experiment the library is induced and an initial time-point (T_0) is taken. After growth in a selective condition for a period of time a final time-point (T_F) is taken. The initial and final samples are sequenced and the fitness of each library member is calculated.

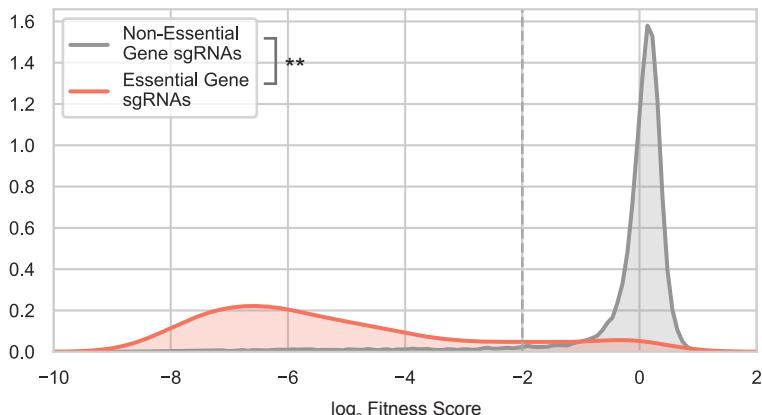
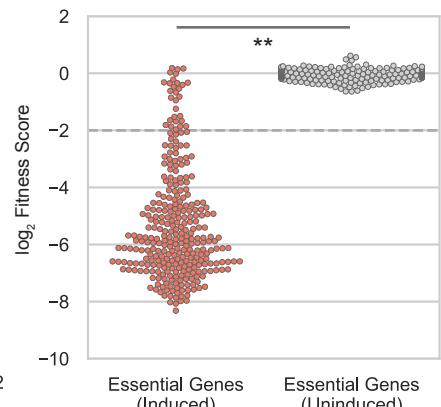
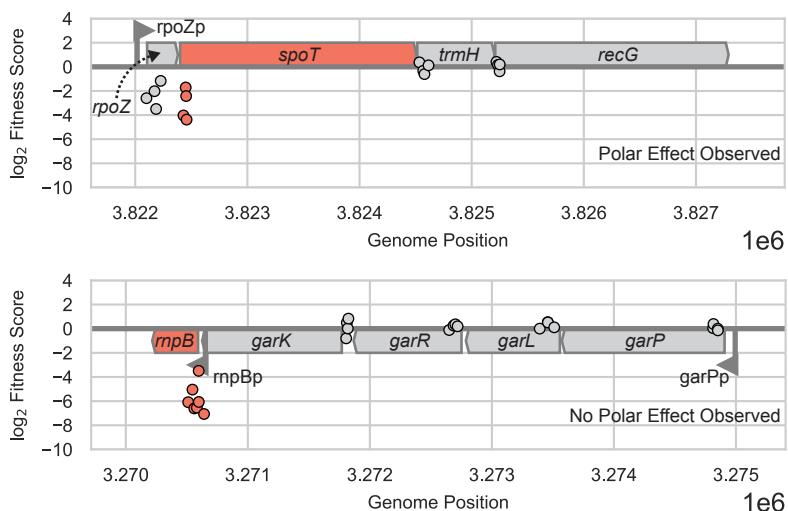
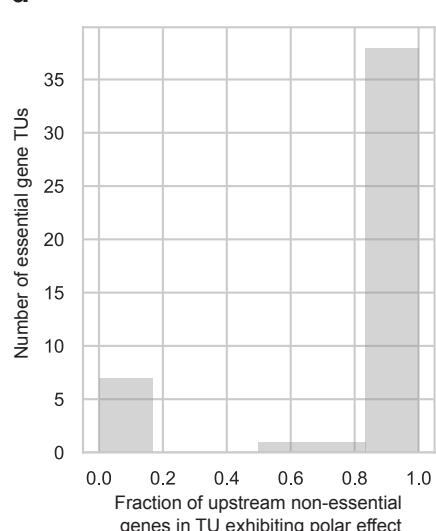
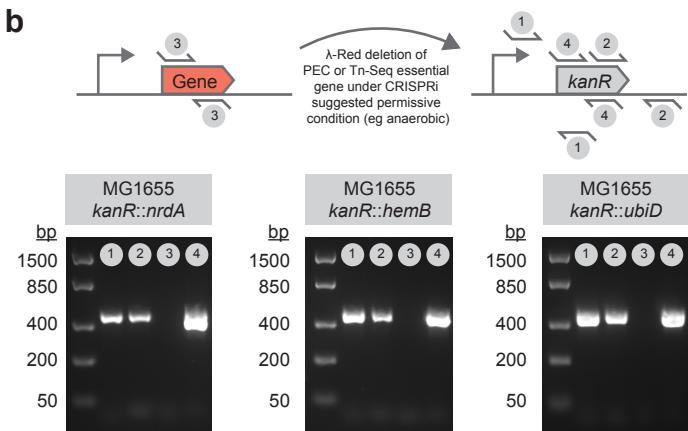
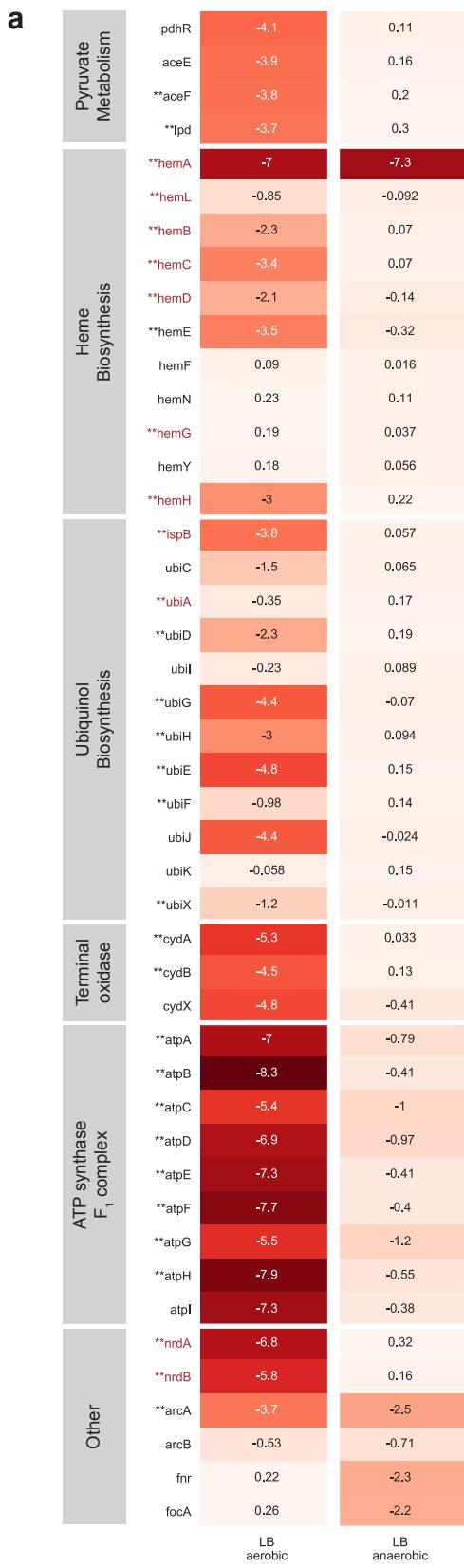
a**b****c****d**

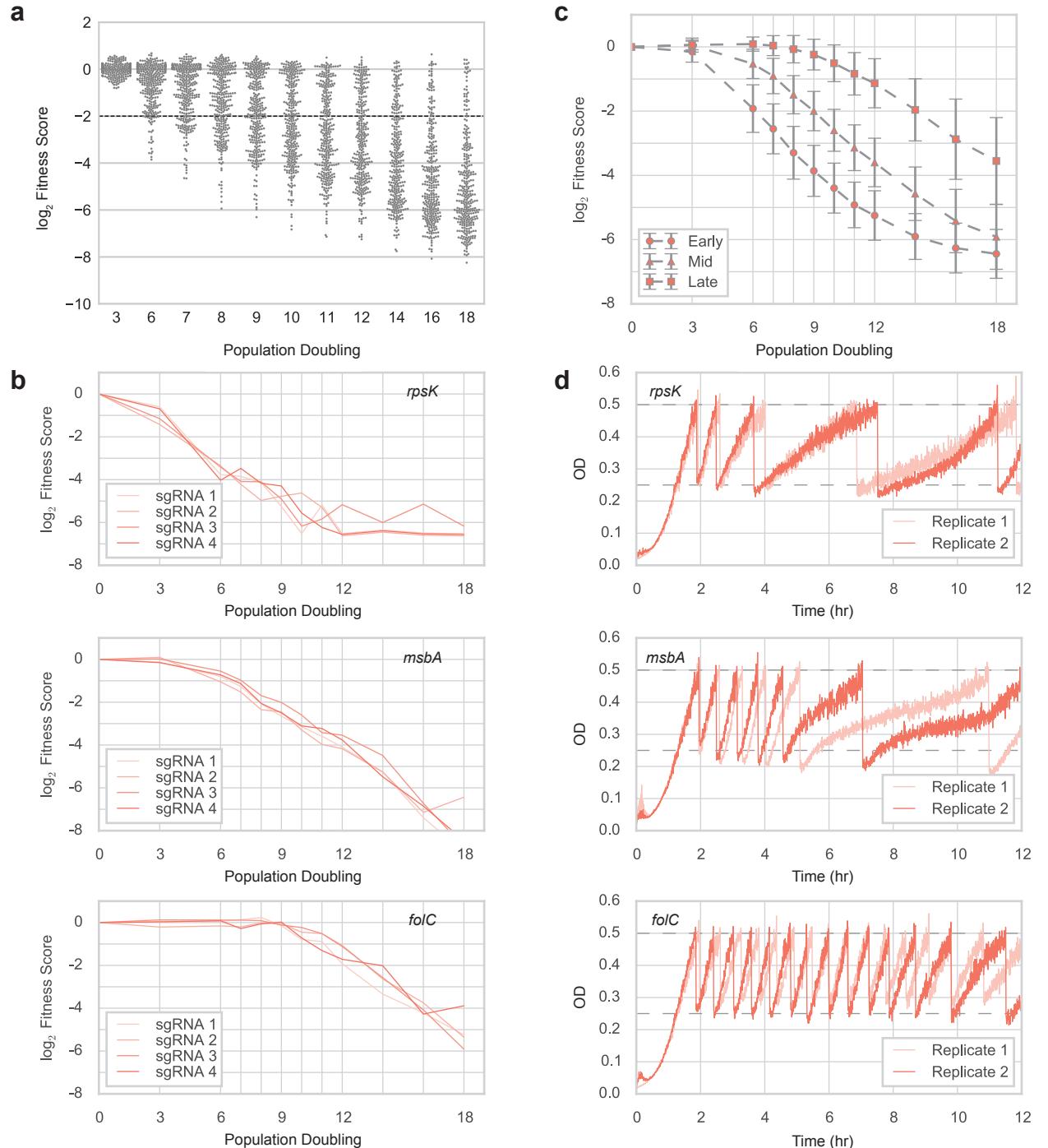
Fig. 2 | Technology validation of CRISPRi screening platform. **a** Depletion of essential gene targeting sgRNAs compared to non-essential gene targeting sgRNAs over the course of a pooled fitness experiment with the CRISPRi library in LB rich media (with CRISPRi system induced) under aerobic growth conditions for 24 population doublings. ** $p < 0.001$, Mann-Whitney U-test (two-tailed); Cohen's $d = 3.7$. **b** Demonstration of tight, inducible control of sgRNA library via comparison of essential gene fitness scores ($n=303$) from pooled fitness experiments where the CRISPRi library was either induced (left) or uninduced (right). In the induced condition, the library was induced with aTc and arabinose to express dCas9 and sgRNA and then grown in LB media for 24 doublings as in a regular fitness experiment (Methods). In the uninduced condition, the library was also grown in similar culturing conditions (e.g. LB media for 24 doublings); however, neither aTc nor arabinose were added. ** $p < 0.001$, Mann-Whitney U-test (two-tailed); Cohen's $d = 3.7$. **c** Example of CRISPRi-mediated polar operon effects where targeting a non-essential gene (*rpoZ*) upstream of an essential gene (*spoT*) in the same transcriptional unit (*rpoZ-spoT-trmH-recG*) produces a fitness defect (top panel). In the presence of an intra-operonic promoter (eg *rnpBp*), knockout of upstream non-essential genes (*garK*, *garR*, *garL*, *garP*) in the same transcriptional unit (*garP-garL-garR-garK-rnpB*) does not produce a fitness defect because essential gene expression can be rescued by the intra-operonic promoter (bottom panel). Targeting the intra-operonic promoter (*rnpBp*) or essential gene (*rnpB*) itself does produce a fitness defect. Each dot represents an sgRNA (centered at midpoint of chromosomal target) targeting either an essential (red-orange) or non-essential (grey) gene. **d** Fraction of non-essential genes upstream of an essential gene within the same transcriptional unit (TU) that also show a fitness defect when knocked down, likely indicating a CRISPRi-mediated polar operon effect. Source data are provided as Source Data file.

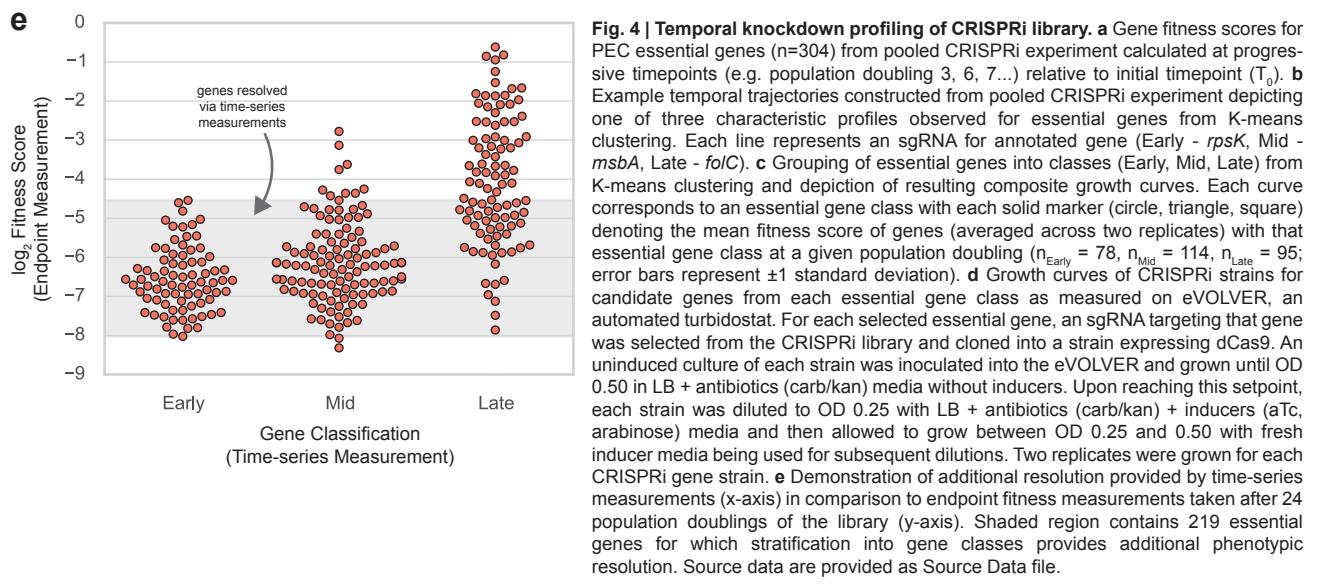


C

Strain Genotype	Plating Condition	Plating Dilution				
		10 ⁰	10 ⁻¹	10 ⁻²	10 ⁻⁴	10 ⁻⁵
<u>ET163</u> MG1655 pTet-dCas9 <i>kanR</i>	aerobic					
	anaerobic					
<u>HR715</u> MG1655 <i>kanR::nrdA</i>	aerobic					
	anaerobic					
<u>HR716</u> MG1655 <i>kanR::hemB</i>	aerobic					
	anaerobic					
<u>HR717</u> MG1655 <i>kanR::ubiD</i>	aerobic					
	anaerobic					

Fig. 3 | Conditional phenotypes from CRISPRi screening. **a** Comparison of CRISPRi phenotypes (gene fitness scores) between aerobic and anaerobic conditions in LB. Gene names in maroon represent genes classified as essential by the Keio collection (*E. coli* K-12 BW25113) and PEC database of essential genes in *E. coli* K-12 MG1655. Gene names with a preceding “**” superscript represent genes for which a mutant could not be generated using RbTnSeq during a high-throughput screen in *E. coli* K-12 BW25113. Gene fitness scores are averaged from a minimum of three replicates. **b** λ-Red recombining mediated deletion of select aerobic essential genes from Keio collection/PEC database (*nrdA*, *hemB*) or sick genes from Rb-TnSeq (*ubiD*) under permissive condition (anaerobic) as discovered via the CRISPRi screen. Gel images with reactions validating in-frame deletion of each essential gene via PCRs showing successful integration of *kanR* resistance cassette and removal of essential gene at native gene locus. **c** Confirmation that anaerobically generated knockouts of selected genes are non-viable under aerobic condition (non-permissive condition). An MG1655 strain with *kanR* cassette integrated on the chromosome is provided as a WT-like reference (ET163). Source data are provided as Source Data file.





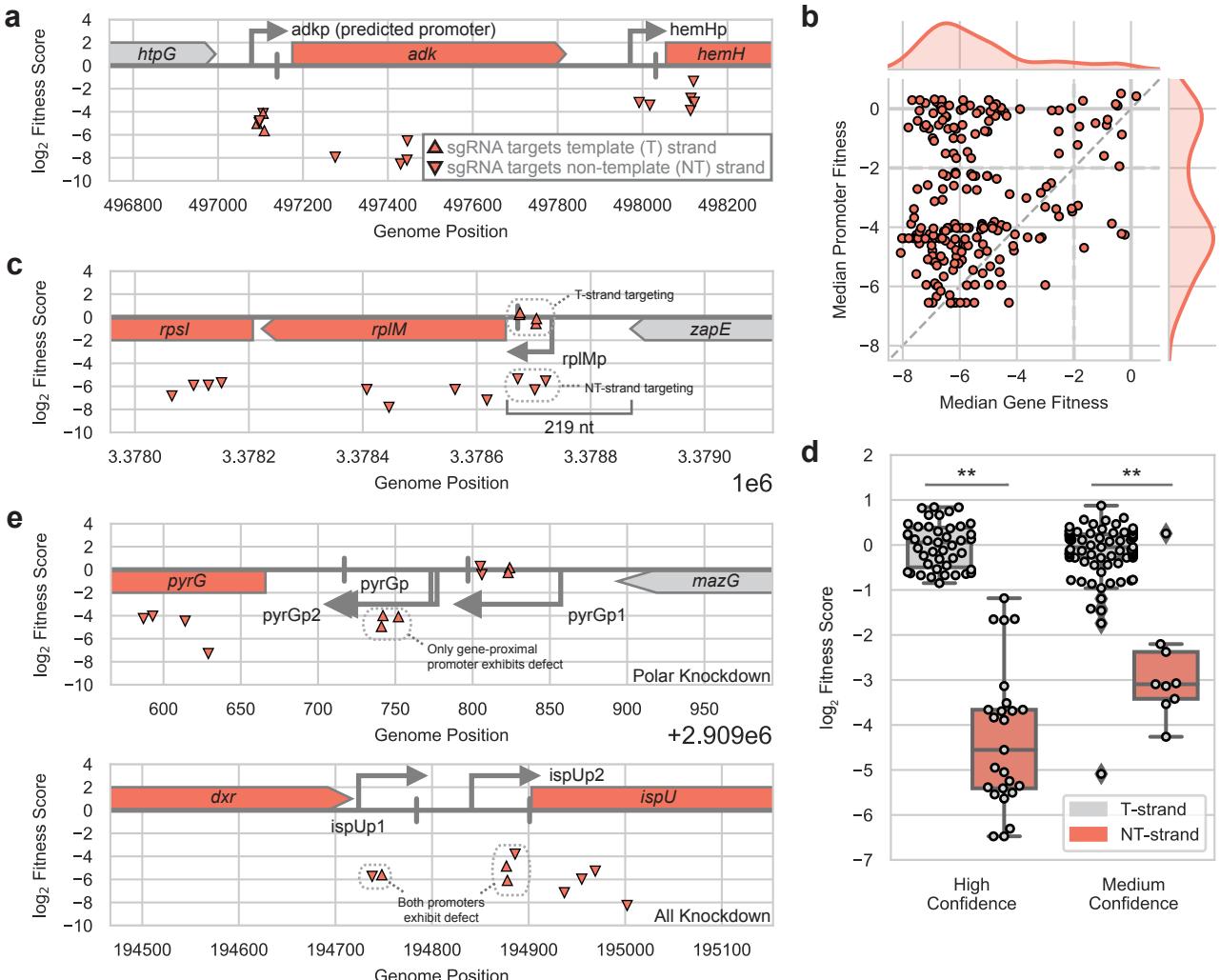


Fig. 5 | Non-genic phenotypes from CRISPRi library. **a** Demonstration of how CRISPRi fitness data for promoter knockdowns can add experimental confidence to predicted promoters (eg adkp) and known promoters (eg hemH) by confirming that targeting the promoter produces a similar phenotype (i.e. fitness outcome) in comparison to targeting its regulated gene (eg adkp - adk; hemH - hemH). **b** Comparison of efficacy of gene-targeting CRISPRi against promoter-targeting CRISPRi for gene expression knockdown. For all essential genes for which guides targeting both gene and promoter sequences were present in the library, the median of fitness scores for sgRNAs targeting the gene sequence (x-axis) is plotted against the median of fitness scores for sgRNAs targeting the promoter sequence (y-axis). Note that the thin diagonal dashed line represents $y = x$. **c** Depiction of strand-dependency of CRISPRi-mediated promoter knockdown for rplMp driving expression of the *rplM-rpsI* operon. Only sgRNAs targeting the NT-strand of the promoter (relative to the gene) produce a fitness defect, while T-strand targeting sgRNAs do not. **d** Boxplots (with data points overlaid) showing strand dependent promoter CRISPRi for 12 high-confidence cases and 26 medium-confidence cases. Each case represents a TU and all of the promoters regulating it (Methods). ** $p < 0.001$ (Mann-Whitney U-test, two-tailed); Cohen's $d = 4.3$ (left), 3.2 (right). **e** Phenotypic profiles of tandem promoter arrays where only knockdown of an essential gene proximal promoter yields a CRISPRi-mediated growth defect (top) or where a knockdown of any promoter regulating the essential gene can yield a growth defect (bottom). Source data are provided as Source Data file.