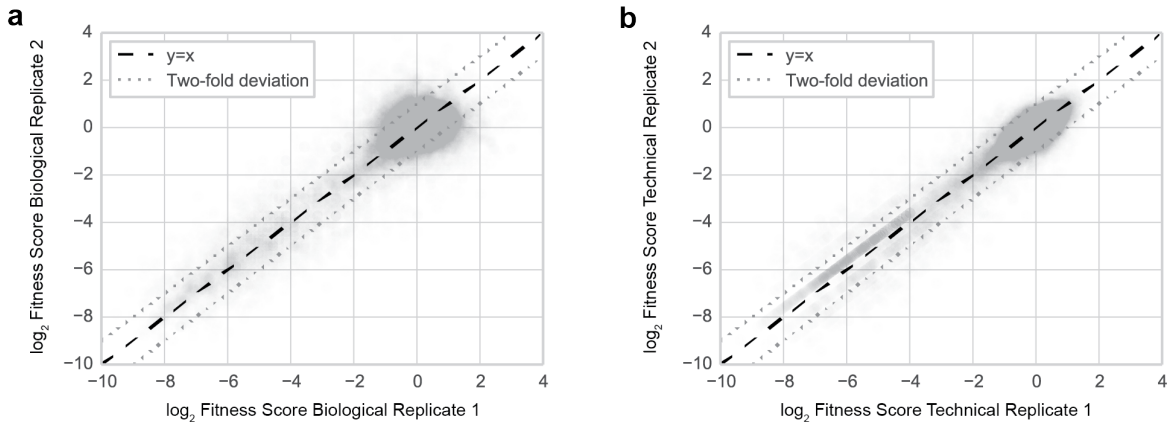
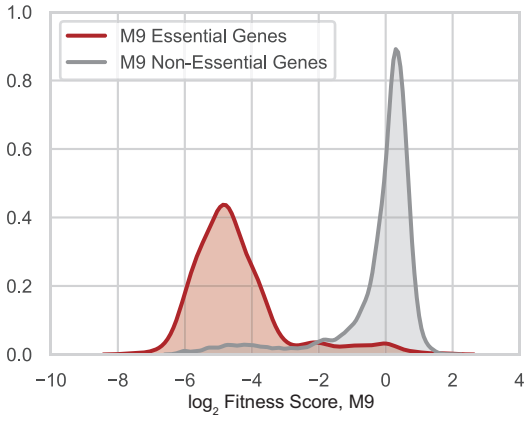
Supplementary Information

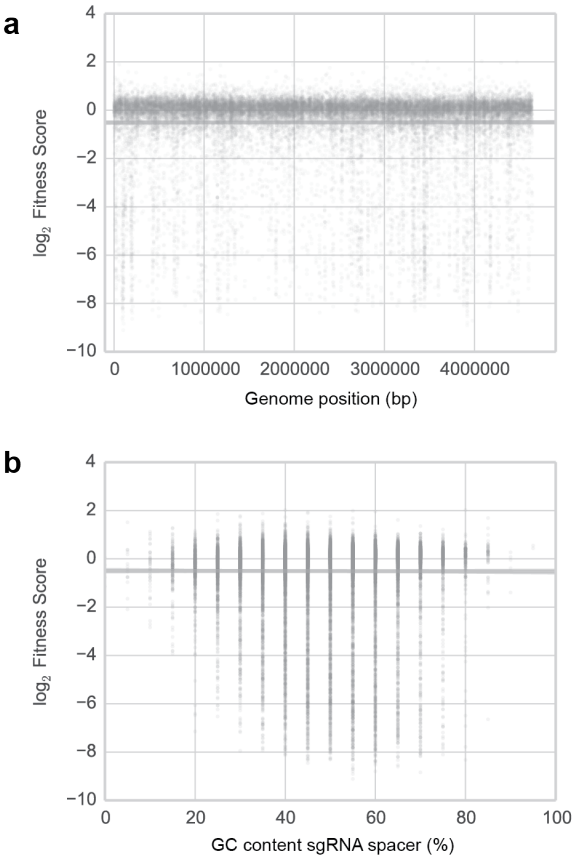
Table of Contents



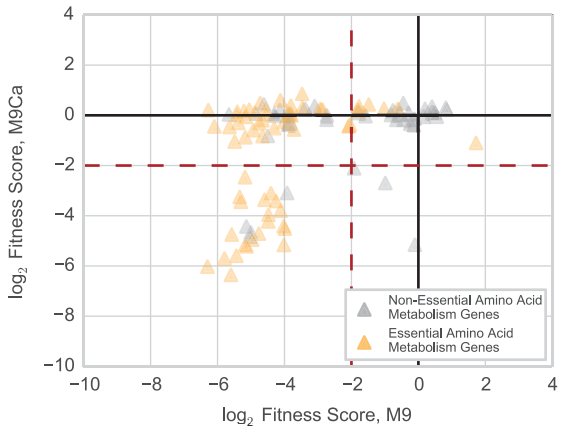
Supplementary Fig. 1 | CRISPRi library replicability. a Two biological replicates of a CRISPRi experiment where the library was grown in LB rich media. Each dot represents an sgRNA. A biological replicate represents a distinct library aliquot. **b** Two technical replicates of a CRISPRi experiment where the library was grown in LB rich media. Each dot represents an sgRNA. A technical replicate represents an aliquot of the library that was split prior to the start of the experiment.



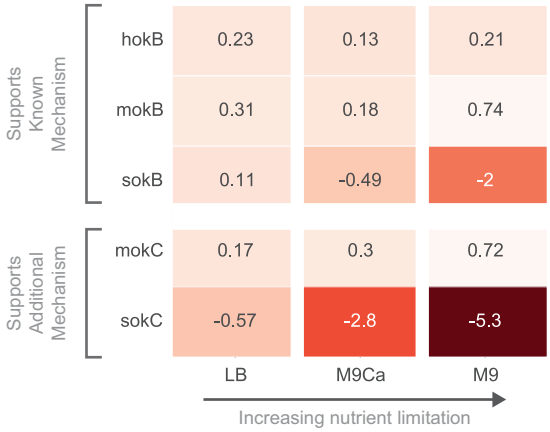
Supplementary Fig. 2 | CRISPRi library minimal media experiment. Depletion of minimal media (M9) essential gene targeting sgRNAs compared to non-essential gene targeting sgRNAs over the course of a pooled fitness experiment with the HT-CRISPRi library in M9 minimal media (with CRISPRi system induced) under aerobic growth conditions for 24 population doublings.



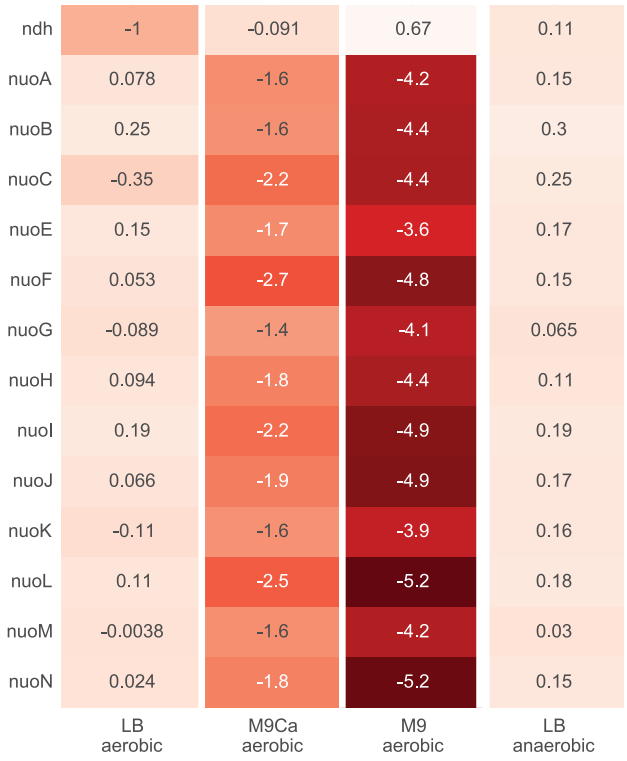
Supplementary Fig. 3 | Investigation of bias in CRISPRi library. a Genome position of library sgRNAs plotted against fitness of respective sgRNAs from a pooled experiment in LB media under aerobic conditions. Gray line represents linear relationship between fitness and genome position with 95% confidence interval. **b** GC content of sgRNA variable region for library sgRNAs plotted against fitness of respective sgRNAs from a pooled experiment in LB media under aerobic conditions. Gray line represents linear relationship between fitness and GC content of sgRNA spacer with 95% confidence interval. **c** Distribution of fitness scores for sgRNAs targeting features on the + or - strand of the genome.



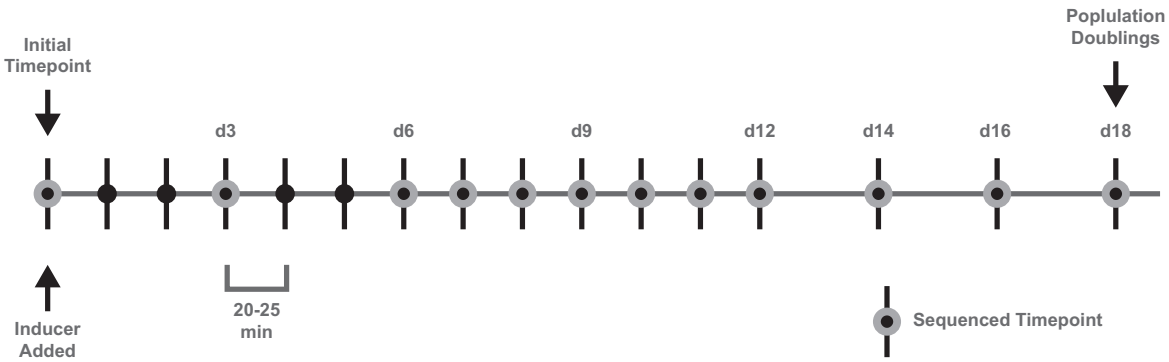
Supplementary Fig. 4 | CRISPRi library amino acid auxotrophy experiment. Depletion of amino acid biosynthetic gene targeting sgRNAs over the course of a pooled fitness experiment in either M9 minimal media (x-axis - M9) or M9 minimal media supplemented with casamino acids (y-axis - M9Ca) under aerobic growth conditions for 24 population doublings. Essential amino acid metabolism genes (yellow triangles) refer to genes classified as essential in Joyce et al *J Bacteriol* 2006 via screening of the Keio essential gene deletion collection on glycerol minimal medium.



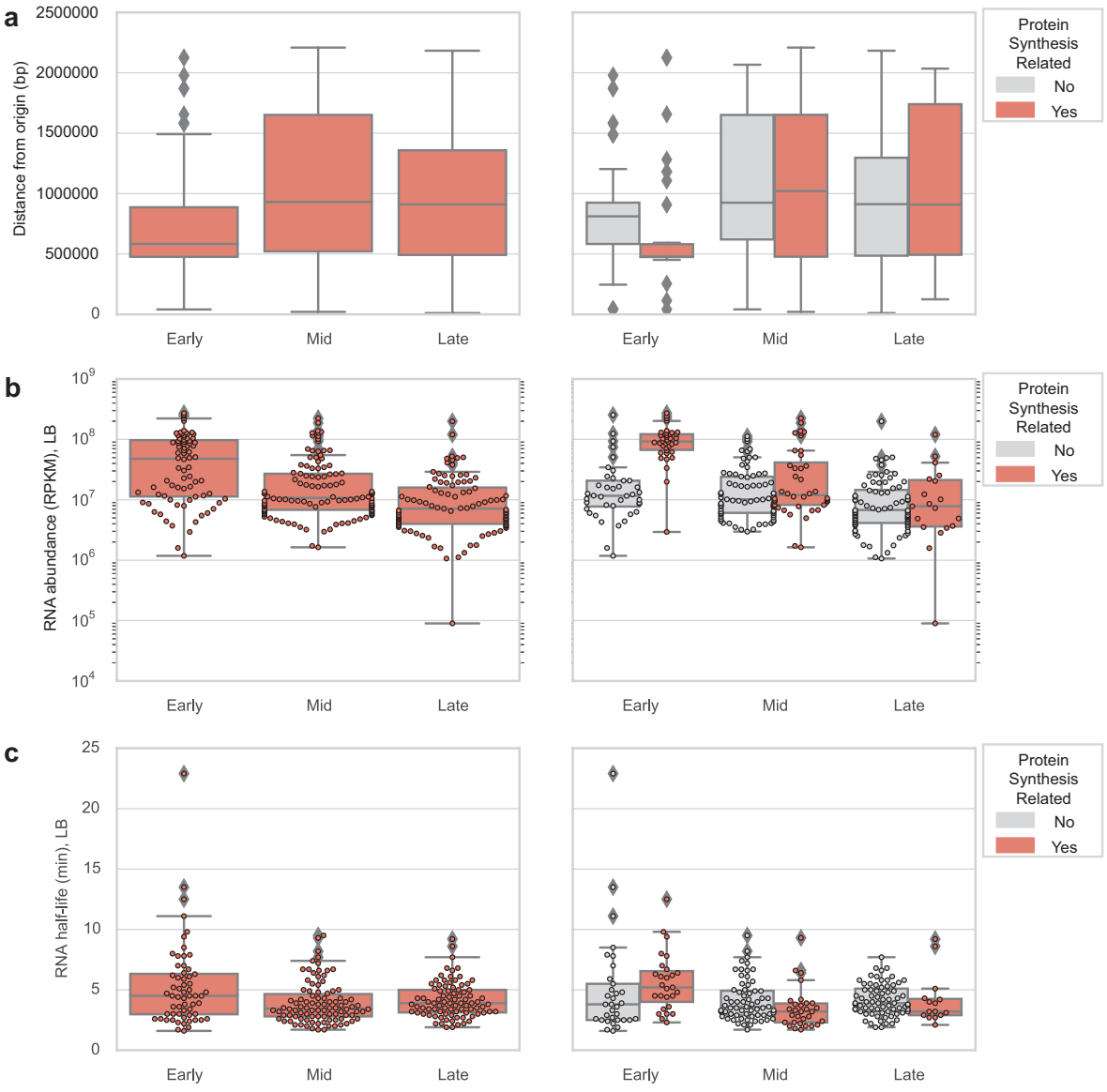
Supplementary Fig. 5 | Conditional phenotypes for *hok-sok* toxin-antitoxin system. Gene fitness scores for genes in the *hok-sok* toxin-antitoxin systems (B & C) showing increasing defect as a result of *sokB* and *sokC* knockdown under conditions of increasing nutrient limitation with *sokC* depicting a stronger phenotypic response than *sokB*. Mechanism for hokB-sokB is reported in Verstraeten et al *Molecular Cell* 2015. Nutrient conditions: LB (rich media), M9Ca (M9 minimal media supplemented with casamino acids), M9 (M9 minimal media). Gene fitness scores are averaged from a minimum of three replicates. Data from pooled fitness experiment with library grown for 24 population doublings under induction in stated condition.

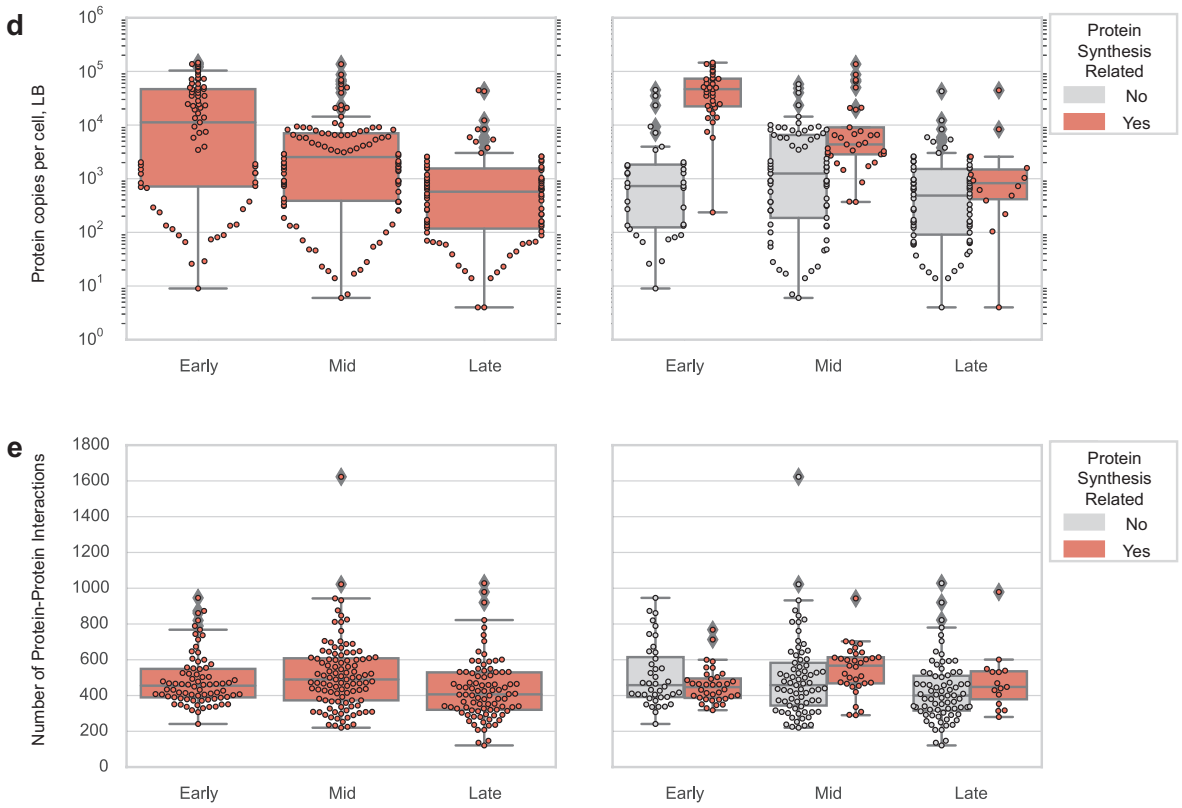


Supplementary Fig. 6 | Conditional phenotypes for NADH:quinone oxidoreductases. Comparison of CRISPRi phenotypes (gene fitness scores) between aerobic conditions in LB, M9Ca, and M9 media against anaerobic condition in LB for NADH:quinone oxidoreductase I (NDH-1; *nuo* genes) and NADH:quinone oxidoreductase 2 (NDH-II; *ndh*). Gene fitness scores are averaged from a minimum of three replicates. Data from pooled fitness experiment with library grown for 24 population doublings under induction in stated condition.

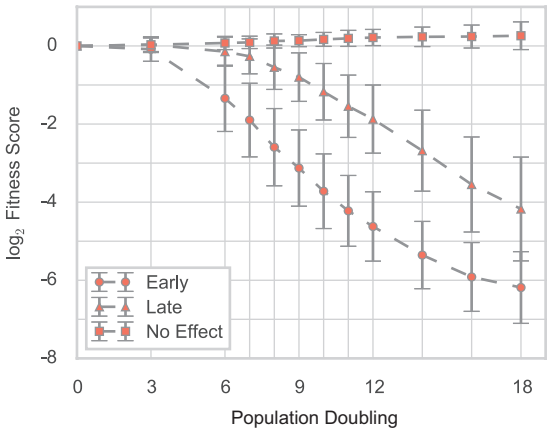


Supplementary Fig. 7 | Workflow of CRISPRi time-series experiment. The library was induced and an initial timepoint was taken. Samples of the library were taken every population doubling for the first 12 doublings and then every other doubling until population doubling 18. Timepoints with gray circles were sequenced.

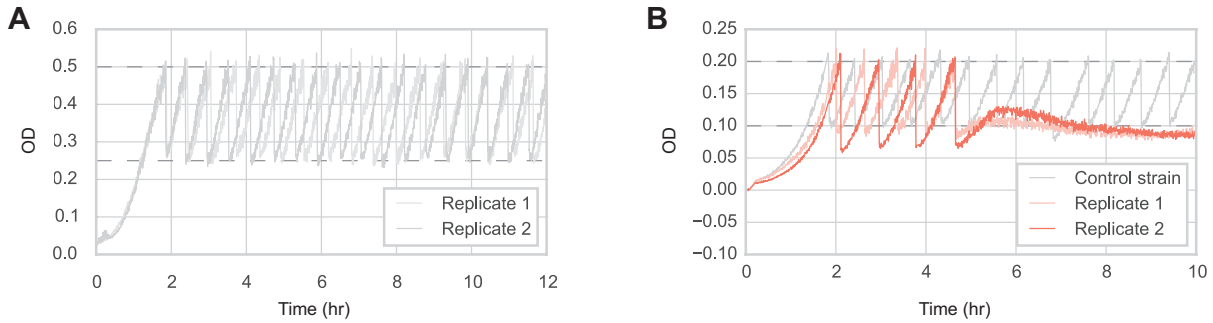

a
Distance from origin (bp)
2500000
2000000
1500000
1000000
500000
0
Early
Mid
Late
Protein Synthesis Related
No
Yes
b
RNA abundance (RPKM), LB
10^9
10^8
10^7
10^6
10^5
10^4
c
RNA half-life (min), LB
25
20
15
10
5
0



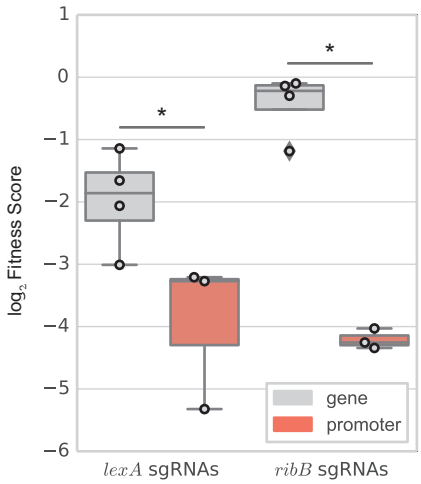
Supplementary Fig. 8 | Analysis of gene product features on time-series gene classification. Comparison of effect of (**a**) distance of gene start position from origin of replication, (**b**) RNA abundance in LB media, (**c**) RNA half-life in LB media, (**d**) protein abundance, and (**e**) number of protein-protein interactions for genes in each essential gene class from time-series data. Essential gene class comparisons (left) are further decomposed into subgroups of genes related (or not) to protein synthesis (right; protein synthesis related defined as corresponding TIGR Roles with leading descriptor "Protein synthesis" - e.g. "Protein synthesis:tRNA aminoacylation") to show if these genes are drivers of class-level trends. mRNA abundance data sourced from Venturelli et al *Nat. Comm.* 2017 GEO accession GSE94998. mRNA half-life data sourced from Bernstein et al *PNAS* 2002 Table 5. Protein abundance data sourced from Schmidt et al *Nat Biotechnol* 2016 Table S6. Protein-protein interaction data sourced from STRING database (string-db.org v10.5).



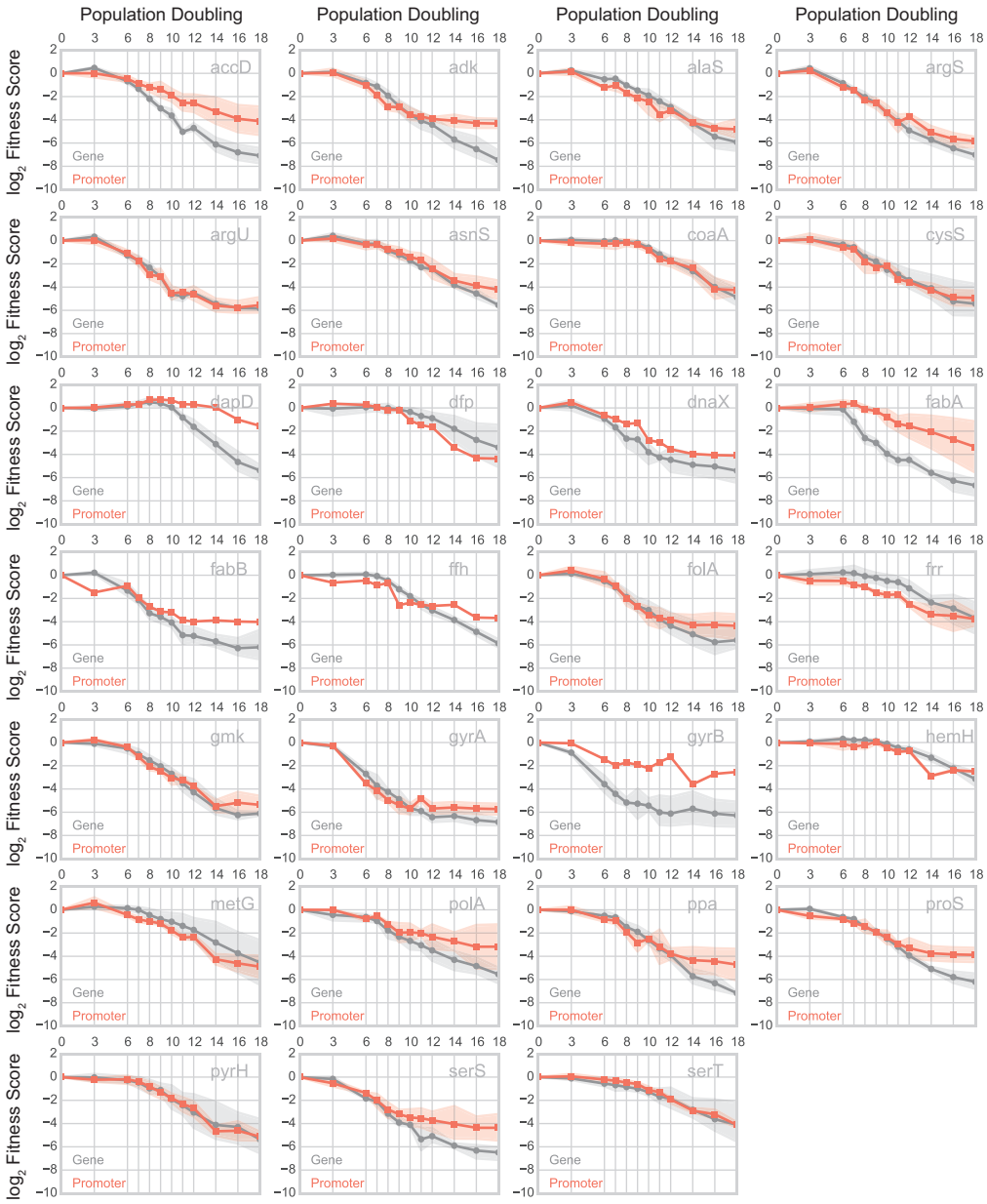
Supplementary Fig. 9 | Time-series classification of all genes in CRISPRi library. Grouping of all genes targeted in CRISPRi library into classes (Early, Late, No Effect) from K-means clustering and depiction of resulting composite growth curves. Each curve represents a gene class with each solid marker (circle, triangle, square) denoting the mean fitness score of genes (averaged across two replicates) with that gene class at a given population doubling (n_{Early} = 188, n_{Late} = 218, $n_{No\ Effect}$ = 4046; error bars represent ±1 standard deviation).



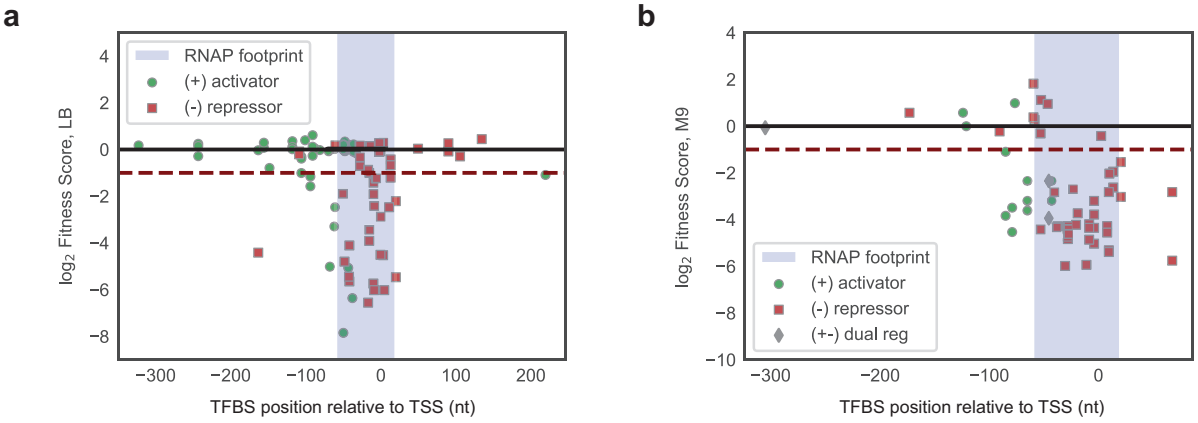
Supplementary Fig. 10| eVOLVER profiling of control and *ftsZ* CRISPRi strains. a eVOLVER growth curves of two replicates of a CRISPRi strain expressing dCas9 and a control sgRNA that does not target any locus on the chromosome. An uninduced culture of the strain was inoculated into the eVOLVER and grown until OD 0.50 in LB + antibiotics (carb/kan) media without inducers, after which each strain was diluted down to OD 0.25 with LB + antibiotics (carb/kan) + inducers (aTc, arabinose) media and then allowed to grow between OD 0.25 and 0.50. **b** eVOLVER growth curves of replicate *ftsZ*-targeting CRISPRi strains. An sgRNA targeting *ftsZ* was selected from the CRISPRi library and cloned into a strain expressing dCas9. An sgRNA designed to not target any locus in the *E. coli* genome was also cloned into a strain expressing dCas9 and used as a reference control strain. An uninduced culture of each strain was separately inoculated into the eVOLVER and grown until OD 0.20 in LB + antibiotics (carb/kan) media without inducers, after which each strain was diluted down to OD 0.10 with LB + antibiotics (carb/kan) + inducers (aTc, arabinose) media and then allowed to grow between OD 0.10 and 0.20 for multiple generations until ~10 hours.



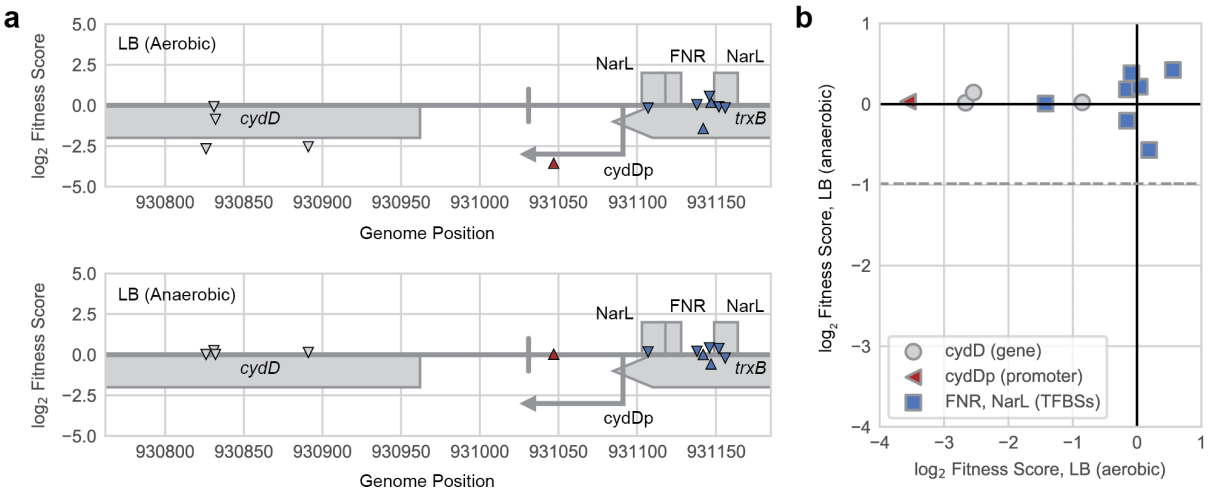
Supplementary Fig. 11 | Examples of promoter-tageting guides more effective than gene-targeting guides. Example case where promoter-targeting sgRNAs provide better knockdown of a known essential gene than functional gene-targeting sgRNAs (left - *lexA*) and gene-targeting sgRNAs that were unable to produce a fitness defect (right - *ribB*). *p < 0.05 (Mann-Whitney U-test); Cohen's *d* = 2.4 (left), 10.7 (right).



Supplementary Fig. 12 | Comparison of promoter- and gene-targeting CRISPRi time series. Composite fitness curves of promoter- and gene-targeting sgRNAs with Fitness ≤ -1 for monocistronic essential gene transcriptional units regulated by a single promoter (see Supplementary Note 2 for details). Each curve represents the mean fitness of gene- (gray; circle marker) or promoter- (red-orange; square marker) targeting sgRNAs (averaged across two replicates) for each measured time point with corresponding shaded regions repressenting 95% confidence intervals.



Supplementary Fig. 13 | CRISPRi knockdown of TFBSs regulating single essential gene promoters. a Fitness scores for sgRNAs targeting TFBSs regulating single promoters of transcription units containing at least one LB essential gene (as determined by PEC database). The RNAP footprint is defined as the window between -60 to +20 nt relative to the transcription start site (TSS) of the regulated promoter. Each object in the scatter plot represents the fitness of an sgRNA (y-axis) targeting a TFBS at a given distance from the TSS of the promoter it regulates (x-axis). A given TFBS can have a positive effect on gene expression (green circles), negative effect on gene expression (red squares), or dual effect on gene expression (gray diamonds) as determined by RegulonDB annotations. **b** Fitness scores for sgRNAs targeting TFBSs regulating single promoters of transcription units containing at least one M9 essential gene (as determined by Joyce et al *J Bacteriol* 2006).



Supplementary Fig. 14 | Feature cofitness of *cydD* gene, promoter, and TFBS-targeting sgRNAs. a Fitness data for *cydD* gene, its corresponding promoter (cydDp), and TFBSs (NarL - gene expression activator, FNR - gene expression activator) regulating its promoter from fitness assays in LB media between aerobic (top panel) and anaerobic (bottom panel) conditions. Each triangle represents an sgRNA (centered at midpoint of chromosomal target) targeting either the chromosomal strand corresponding to the non-template (downward facing triangle) or template (upward facing triangle) strand of the *cydD* gene. **b** Scatter plot comparing conditional phenotypes for sgRNAs targeting *cydD* (gray circles), *cydD* promoter (red triangle), and *cydD* TFBSs (blue squares) between aerobic and anaerobic conditions.

Supplementary Table 3. List of essential gene knockout validation strains

Strain Number	Plasmid Name	Description	Host	Resistance	Link to Modified Sequence
HR715	n/a	*kanR::nrdA*	MG1655 K-12	Kan	https://benchling.com/s/seq-LQ6uGkCQr09iU68FCnzh
HR716	n/a	*kanR::hemB*	MG1655 K-12	Kan	https://benchling.com/s/seq-Wu1pfiZX4M5JlaIFtora
HR717	n/a	*kanR::ubiD*	MG1655 K-12	Kan	https://benchling.com/s/seq-XOLhFwSV5CbGRdp5i1Yt

Supplementary Table 5. List of strains used for eVOLVER CRISPRi experiment

Strain Number	Plasmid Name	Description	Host	Resistance	Link to Plasmid Sequence
ET169	pT169	Pbad:control sgRNA	ET163	Amp, Kan	https://benchling.com/s/seq-aLVjhEiBggyDQgeltKfg
ET170	pT170	Pbad:*ftsZ* sgRNA	ET163	Amp, Kan	https://benchling.com/s/seq-Zq5ApvVfBGbslqFt2RLW
HR664	pHR664	Pbad:*rpsK* sgRNA	ET163	Amp, Kan	https://benchling.com/s/seq-s88pSK0iEeEyRJ7xwr6G
HR665	pHR665	Pbad:*msbA* sgRNA	ET163	Amp, Kan	https://benchling.com/s/seq-JZWG9whqXzBqjFkrt4iA
HR666	pHR666	Pbad:*folC* sgRNA	ET163	Amp, Kan	https://benchling.com/s/seq-SqywQzhOAWZCmbDTgYHT

Supplementary Note 1. sgRNA library design

We designed the sgRNA library following rules described in Larson et al *Nature Protocols* 2013:

Selection of sgRNAs for oligo pool:

1. We first identified all 5'- XXXXX XXXXX XXXXX XXXXX NGG-3' sequences by searching both the sense and anti-sense strand of the genome, to generate the original pool of the potential sgRNA binding sites.
2. To avoid potential off-target effects, we mapped all 5'-XX XXXXX XXXXX-NGG-3' from step #1 back to the genome using the short reads mapping program Seqmap (http://www-personal.umich.edu/~jianghui/seqmap/) with parameter setting "1 /output_all_matches", and filtered out the sequences with multiple mappings.
3. We required that the designed sgRNAs should be able to fold properly. To check that this was true, we linked the 42nt scaffold sequence 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG-3' to the 3' end of the 20 nt specific target binding sequence and checked the folding structure of this 62nt sequence by RNA secondary structure prediction using RNAfold (http://www.tbi.univie.ac.at/~ronny/RNA/) with default parameters. We only kept the ones that the scaffold region could fold to the hair-pin structure as reported in Jinek et al *Science* 2012.
4. Finally, we filtered out any sgRNA sequences containing the BsaI restriction site (GGTCTC), which we used for cloning purposes.

The sequences that passed these four steps composed our pool of potential sgRNAs. Next, we chose sgRNAs from the sgRNA pool to target all (1) annotated genes, (2) promoters and (3) TFBSs, according to RegulonDB.

1. sgRNAs that target coding sequences:

   We tried to collect 4 sgRNAs for each annotated gene in the *E. coli* genome. We implemented a recursive approach to select sgRNAs as close to the ATG as possible and on the non-template strand for each gene. We first looked at the first 50%. Next, we looked at the annotated 5' UTR regions, and the ones close to the start codon where selected with higher priority. Finally, we looked at the last half of the CDS sequence, and chose the sites closer to the start codon with higher priority. By using this approach, 4281 genes could be targeted with 4 sgRNAs, 193 additional genes could be targeted by 1-4 sgRNAs, and 158 genes could not be targeted by any sgRNA.

   We further looked at the 158 genes that could not be targeted by the previous pipeline. We noticed that 39 of them were located in operons where an upstream gene in that operon had properly selected sgRNAs.

   For the rest 109 genes, we found many of them had closely related homologs on the genome, which caused the sgRNAs targeting these region to be not unique on the genome and could target both of the homologs. So we compared the sequences of all the annotated genes, and defined a homolog gene set by performing a megablast search with parameter setting of "-F F -D 3 -e 1e-10". We searched for the potential sgRNA target sites that locate in both the homolog genes but not any other sites on the genome. 48 genes could be targeted in this way.

   Finally, there are 71 genes could not be targeted by our sgRNA design procedure. Most of them are small RNAs that don't have any PAM site.

   Finally, we designed 17622 sgRNAs, which could target 4561 genes (4522 directly, 39 indirectly) on the *E.coli* genome.

2. sgRNA target promoters:

   For the promoters that did not overlap with any annotated UTR or CDS regions, we selected the sgRNA from both the sense and anti-sense strand in the region from upstream 60 bp to downstream 10 bp relative to the transcription start site.

   For the promoters located within a gene body, we only designed sgRNAs that binds to the template strand of that region. 14257 sgRNAs were selected to target 7404 Promoters.

3. sgRNA target TFBS sites:

   We designed all the sgRNAs that could target the TFBSs annotated in the RegulonDB database. An sgRNA is selected if it could cover at least one-third of the annotated TFBS. If the TFBS is shorter than 15 bp, we required that the overlap should be at least 5bp. 1867 sgRNAs were selected to target 1264 TFBS sites.

4. sgRNA for subcategories:

   a. We designed sgRNAs for 21 genes subcategories (eg cell division, small RNAs, central intermediary metabolism). These sgRNAs are encoded with an additional category code in the 3' end of each library oligo to enable amplification of subpools of the library. Categories and their corresponding category codes for amplification can be found in Supplementary Table 1.

We used the following external files as annotations for our sgRNA design:

- Genome sequence: Escherichia coli str. K-12 substr. MG1655, complete genome, NCBI Reference Sequence NC_000913.2 (http://www.ncbi.nlm.nih.gov/nuccore/NC_000913.2)
- Genome annotations from RegulonDB v8.1: (http://regulondb.ccg.unam.mx/download/Data_Sets.jsp)
  - Gene coordinate: Gene_sequence.txt
  - Promoter annotation: PromoterSet.txt
  - UTR annotation: UTR_5_3_sequence.txt
  - Transcription factor binding sites: BindingSiteSet.txt

Note: To keep with genome annotation updates, sgRNAs were remapped to promoter and TFBS features using more recent RegulonDB annotations:

- Promoter annotation: PromoterSet.txt (RegulonDB v9.4; release date 05-08-2017)
- TFBS annotation: BindingSiteSet.txt (RegulonDB v10.5; release date 09-13-2018)

Supplementary Note 2. Analysis of time-series data

The fitness of each sgRNA strain was calculated at each sequenced time point relative to the initial timepoint of the experiment. This constructed a time-series fitness curve for each sgRNA in the library.

Time-series Analysis 1 – Clustering of Essential Genes:

1. Calculate gene fitness scores for each gene annotated as essential in the PEC database
2. Filter out any genes that did not have a gene fitness score $\leq$ -1 (i.e. keep only essential genes that showed a knockdown phenotype)
3. Keep only timepoints with a Pearson correlation $\geq$ 0.8 across two replicates
4. Average the remaining timepoints across replicates
5. Performed a min-max scaling of each timepoint (i.e. i.e. fitness values at each timepoint were scaled to between 0 and 1) from Step 4 to ensure that all timepoints were treated equally
6. Used the Elbow method to track the variation of the within-cluster-sum-of-squares (WCSS) with the number of clusters (k – ranging from 1 to 14) and found k = 3 to be the optimal number of clusters for K-means based on visual inspection.
7. Performed K-means clustering with selected k from Step 6 to classify essential gene curves
8. Visualize K-means clusters (Early / Mid / Late)

Time-series Analysis 2 – Clustering of All Genes:

1. Calculate gene fitness scores for each gene targeted in the CRISPRi library
2. Keep only timepoints with a Pearson correlation $\geq$ 0.8 across two replicates
3. Average the remaining timepoints across replicates
4. Performed a min-max scaling of each timepoint (i.e. i.e. fitness values at each timepoint were scaled to between 0 and 1) from Step 3 to ensure that all timepoints were treated equally
5. Used the Elbow method to track the variation of the within-cluster-sum-of-squares (WCSS) with the number of clusters (k – ranging from 1 to 14) and found k = 3 to be the optimal number of clusters for K-means based on visual inspection.
6. Performed K-means clustering with selected k from Step 5 to classify essential gene curves
7. Visualize K-means clusters (Early / Late / No Effect)

Gene Ontology Enrichment for Analysis 1 and 2:
For either time-series analysis, each gene was associated with its annotated TIGR Role. A hypergeometric test was carried out for each TIGR Role in each gene class (for analysis 1 – Early / Mid / Late; for analysis 2 – Early / Late / No Effect) with parameters: N = #total essential genes in data set, K = #total genes in class, n = #total genes with TIGR Role in data set, k = #genes with TIGR Role in class. The Benjamini-Hochberg correction was applied to the resulting p-values using the multitest function (parameter: “fdr_bh”) in the statsmodels python module (http://www.statsmodels.org/stable/index.html). The threshold of $p_{FDR\text{-}adjusted} \leq 0.05$ was used as the significance threshold.

Time-series Analysis 3 – Comparison of gene-targeting and promoter-targeting CRISPRi:

1. Select all essential genes for which guides targeting the corresponding promoter and the gene itself were designed in the library
2. Of these promoter-gene pairs, select all essential genes that are the first and only gene in their respective transcription unit (TU). This enables association of a specific promoter knockdown or gene phenotype to the specific gene itself.
3. Of the remaining promoter-gene pairs, select cases where the gene only has one promoter
4. Keep only sgRNAs that had t0 counts $\geq$ 10 and had a fitness score $\leq$ -1 by the final timepoint (i.e. timepoint 15)
5. Plot time-series using lineplot function from seaborn plotting library (v0.9.0) with the parameter setting “ci = 95” to generate 95% confidence intervals via bootstrapping.
    a. Lineplot function: https://seaborn.pydata.org/generated/seaborn.lineplot.html
6. For each gene, compare the overlap of the 95% confidence intervals between population doublings 6 and 12 (these timepoints were selected because they are both highly correlated across replicates and because after doubling 12 we start to see fitness scores leveling out due to limitations in sequencing read depth)

References

1. Larson,M.H., Gilbert,L.A., Wang,X., Lim,W.A., Weissman,J.S. and Qi,L.S. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols*, **8**, 2180–2196.
2. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*, **337**, 816–821.