

Identification and genetic dissection of convergent persister cell states

<https://doi.org/10.1038/s41586-024-08124-2>

Received: 23 May 2023

Accepted: 26 September 2024

Published online: 6 November 2024

Open access

 Check for updates

Sydney B. Blattman¹, Wenyan Jiang^{1,4}, E. Riley McGarrigle¹, Menghan Liu¹, Panos Oikonomou¹ & Saeed Tavazoie^{1,2,3} 

Persister cells, rare phenotypic variants that survive normally lethal levels of antibiotics, present a major barrier to clearing bacterial infections¹. However, understanding the precise physiological state and genetic basis of persister formation has been a longstanding challenge. Here we generated a high-resolution single-cell RNA atlas of *Escherichia coli* growth transitions, which revealed that persisters from diverse genetic and physiological models converge to transcriptional states that are distinct from standard growth phases and instead exhibit a dominant signature of translational deficiency. We then used ultra-dense CRISPR interference³ to determine how every *E. coli* gene contributes to persister formation across genetic models. Among critical genes with large effects, we found *lon*, which encodes a highly conserved protease⁴, and *yqgE*, a poorly characterized gene whose product strongly modulates the duration of post-starvation dormancy and persistence. Our work reveals key physiologic and genetic factors that underlie starvation-triggered persistence, a critical step towards targeting persisters in recalcitrant bacterial infections.

Antibiotic persistence is the ability of rare bacterial cells to survive high concentrations of antibiotics, often by assuming a dormant or non-growing state⁵. One mode of persistence occurs following stress; persisters are cells that exhibit a long lag phase before resuming fast growth⁶. Selection after lethal antibiotic exposure yields mutants, including *metG** and *hipA7*, with high levels of persistence triggered by starvation^{7–11}. The *hipA7* mutation was also found in 5% of a small sample of clinical *E. coli* isolates¹².

Characterizing the cell state of rare persisters has been a longstanding challenge^{8,13,14} that is now tractable given the recent development of prokaryotic single-cell RNA sequencing^{2,15} (scRNA-seq). A recent study used scRNA-seq to identify an enriched population of *Klebsiella pneumoniae* persisters following antibiotic treatment¹⁶. However, it is also critical to agnostically define such persister cell states in antibiotic-naïve populations and determine the genetic basis of their emergence.

We thus set out to precisely characterize the emergence of antibiotic persisters, capture their transcriptional states, and contextualize those states relative to others. To that end, we generated a single-cell atlas of *E. coli* growth transitions. Within this atlas, we identified a unique persister cluster that is common across five genetic and physiological models of persistence. This persister cluster is distinct from all other growth phases and is primarily defined by translational deficiency. To understand which of the transcriptional markers of persisters may be causal to their formation, we carried out a comprehensive CRISPR-interference (CRISPRi) screen^{3,17,18} in three genetic models. We identified several gene products that contribute to persister formation, including *Lon* protease and *YqgE*, an uncharacterized protein that has not previously been implicated in these phenomena.

A distinct persister state

We focused first on *metG**, a hyper-persistent mutant of MG1655 with a 12-bp deletion in the methionine-tRNA ligase gene⁸. We hypothesized that *metG** hyper-persistence would be lag-dependent. Thus, we grew cells in a chemostat for more than 12 h (Fig. 1a) to establish a uniform exponential population. During exponential phase, *metG** and wild-type (MG1655) cells treated with ampicillin or ciprofloxacin had survival rates below 0.01% both when assayed directly from the bioreactor (Extended Data Fig. 1a, time <2 h) or after dilution into fresh medium (Fig. 1b, time <2 h). As cells reached stationary phase, *metG** cells showed a marked increase in lag phase persistence, increasing 100-fold in minutes (Fig. 1b, left, first versus second red arrow) before plateauing at 30–60% survival (Extended Data Fig. 1b, left). By contrast, survival of wild-type cells only reached 0.01–0.02% (Fig. 1b and Extended Data Fig. 1b, right). Concurrent with increased antibiotic survival, colony appearance times—a proxy for lag time¹⁹—became longer for *metG** cells but not for wild-type cells (Fig. 1c).

We sampled cells at critical timepoints (arrows in Fig. 1b) for high-throughput scRNA-seq by prokaryotic expression profiling by tagging RNA in situ and sequencing² (PETRI-seq). We used an updated PETRI-seq protocol, which includes Cas9-driven depletion of ribosomal RNA and multiple stopping points for flexible use (Supplementary Fig. 1 and Methods). At the indicated timepoints, we sequenced cells taken directly from the bioreactor ('undiluted') and 20 min or 2 h after dilution into fresh medium. We used uniform manifold approximation and projection (UMAP) to visualize single cells in two dimensions. Before *metG** cells became hyper-persistent, each population occupied a single area of the UMAP, which overlapped with matched wild-type samples (Fig. 1d, left). As *metG** survival increased to 0.3% (Fig. 1d,

¹Department of Biological Sciences, Columbia University, New York, NY, USA. ²Department of Systems Biology, Columbia University, New York, NY, USA. ³Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY, USA. ⁴Present address: Icahn School of Medicine at Mount Sinai, New York, NY, USA.  e-mail: st2744@columbia.edu

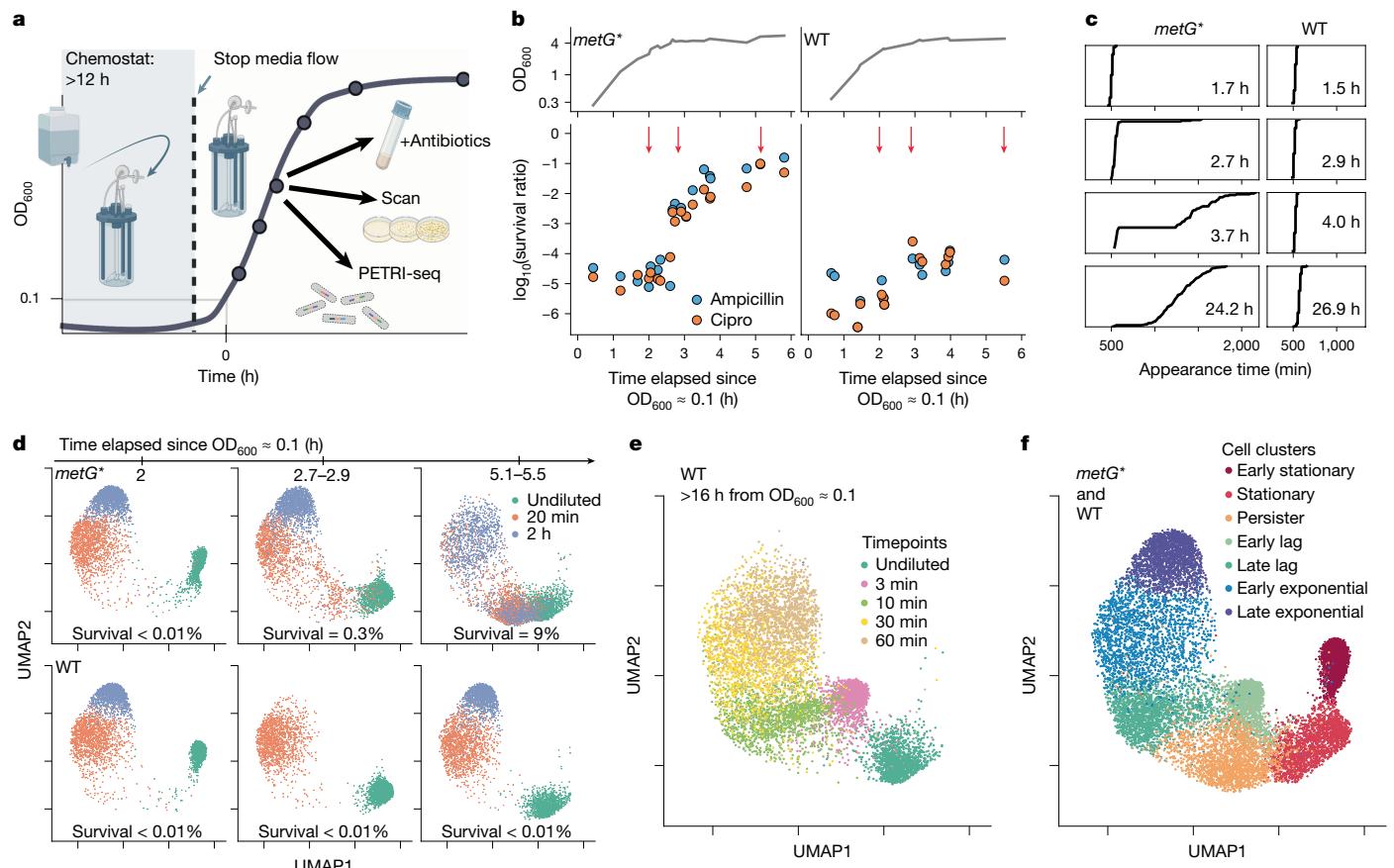


Fig. 1 | A single-cell RNA atlas of growth phases in *E. coli* reveals a distinct persister cell state. **a**, Schematic of bioreactor experiment. **b**, Left, *metG** cells exhibit lag-dependent hyper-persistence upon dilution from stationary phase. $n = 19$ measurements from 4 biological replicates. Top, optical density at 600 nm (OD_{600}) measurements over time. Bottom, survival after antibiotic treatment. Red arrows indicate timepoints used for PETRI-seq. Right, OD_{600} and survival measurements for wild-type *E. coli*. $n = 14$ from 3 biological replicates. Cipro, ciprofloxacin; WT, wild type. **c**, Cumulative distributions of colony appearance times. *metG** cells (left) exhibit extended and heterogeneous lag times relative to wild-type cells (right). **d**, Top, PETRI-seq of *metG** cells before and after critical

shift to hyper-persistence. Bottom, wild-type cells at matched timepoints. Cells were either sampled directly from the bioreactor (undiluted) or diluted into fresh medium and grown for 20 min or 2 h ($n = 1,500$ representative cells per population). In the bottom middle panel, the wild-type population was not sampled at 2 h. **e**, PETRI-seq of wild-type *E. coli* with higher resolution of growth states, including cells undiluted or after an indicated amount of time in fresh medium. $n = 1,500$ representative cells per population. **f**, Unsupervised clustering of combined *metG** and wild-type populations; 2,000 representative cells are shown for each cluster. Cluster names indicate corresponding phenotype and growth condition.

top middle) and then 9% (Fig. 1d, top right), single-cell transcriptomes assayed after dilution became bimodal for the *metG** cells, with one population resembling wild type and the other occupying a distinct space (best seen in the orange population in Fig. 1d, top right). We hypothesized that this distinct population corresponds to *metG** persisters, as it emerges during lag phase precisely at the transition to hyper-persistence. This state was highly reproducible in replicate scRNA-seq of *metG** cells (Extended Data Fig. 1d,e). Comparing these candidate persisters with co-existing non-persisters, we observed a marked reduction in absolute transcripts per cell (Extended Data Fig. 1f, purple violin and pie charts), which is similar to low mRNA abundance in dormant stationary cells^{2,20} (Extended Data Fig. 1f–h, undiluted). Given the large range in mRNA counts per cell, we downsampled every cell to a maximum of 38 mRNA counts for all analyses.

To better contextualize the emergent persister state, we next performed a high-resolution PETRI-seq experiment using wild-type cells in stationary, lag and exponential phases (Fig. 1e). Consistent with previous work, transcriptional changes occurred less than 3 min into lag phase²¹. We brought all cells together to generate a single-cell atlas of *E. coli* growth phases (Fig. 1f). Unsupervised clustering found seven transcriptional states, which we labelled on the basis of the samples in which they appeared. Sequencing of *metG** cells after ampicillin confirmed

annotation of the persister cluster (Fig. 2a). As shown in Fig. 1f, the persister state is distinct from stationary phase, meaning that persisters transcriptionally respond upon dilution into fresh medium, which occurs even in ampicillin-containing media (Extended Data Fig. 1i–k). Persisters are also distinct from wild-type cells in early lag (wild-type cells 3 min after dilution) or late lag (wild-type cells 10 min after dilution) (Fig. 1f and Extended Data Fig. 1l,m). However, both early lag and persister cells are in a transitional state between the stationary and exponential clusters based on principal component 1 (PC1) (Extended Data Fig. 1n). We used the principal component loadings to broadly define the dominant expression patterns in the cell atlas (Extended Data Fig. 1o). Translation and cold shock response genes²² increase from stationary to persister to exponential cells, whereas amino acid biosynthetic genes follow the reverse trend. Cells from exponential to early stationary phase appear the most aerobic; persisters and early lag cells are the least aerobic.

To specifically find markers of the *metG** persister state, we compared gene expression of the persister cluster to the early exponential cluster (Extended Data Fig. 2a(i)), the predominant type of co-occurring non-persister (orange and blue in Fig. 2a, left). Genes that are most upregulated in *metG** persisters include *rmf*, *cysK* and *mdtK* (Extended Data Fig. 2a(i)), expression of which we validated using transcriptional fusions (Extended Data Fig. 3a–e). However, many persister markers

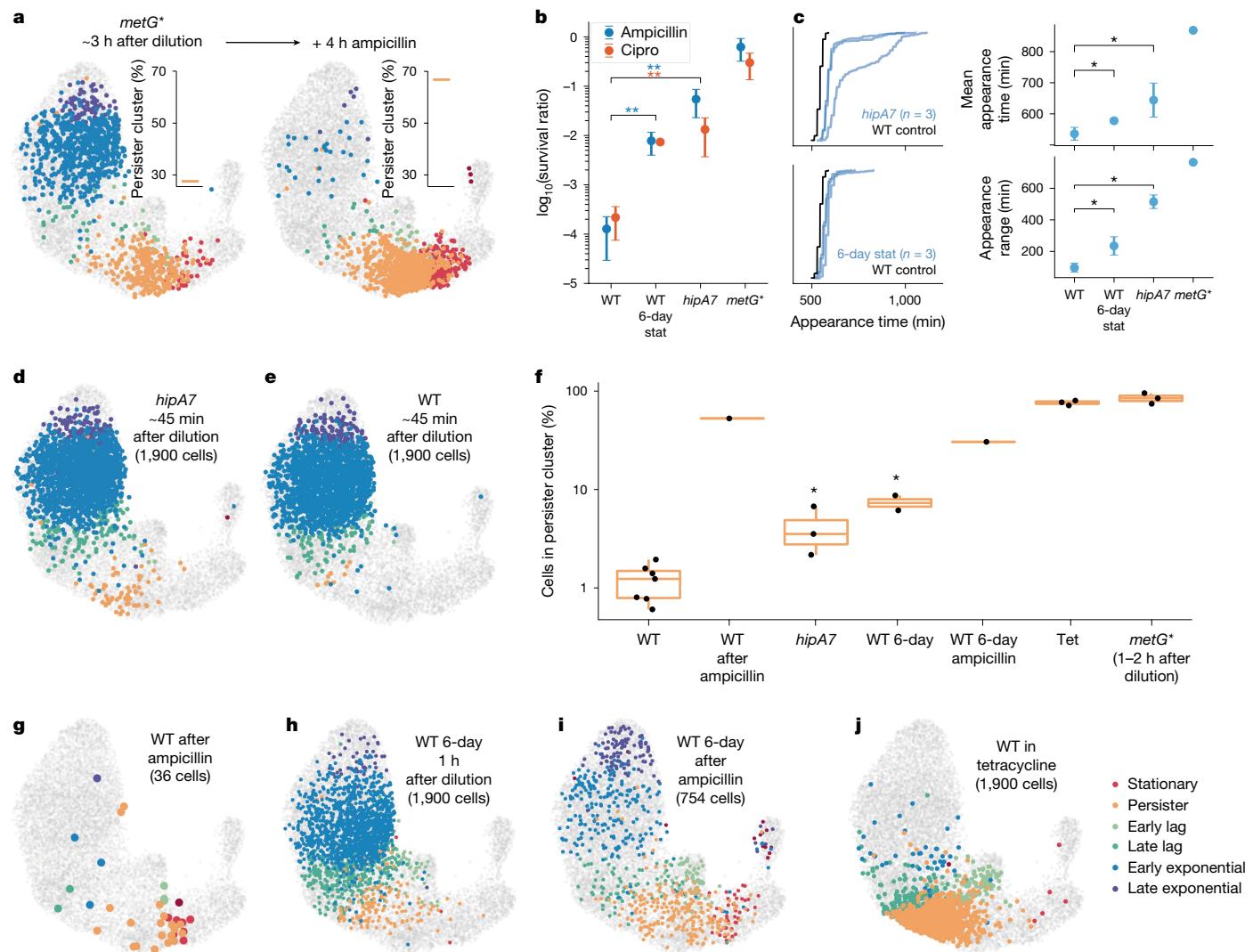


Fig. 2 | Convergence of persister states implicates translational deficiency as a defining feature of persister transcriptomes. **a**, *metG** cells before and after ampicillin treatment ($n = 1,100$ representative cells for each UMAP). Insets show the percentage of cells in the persister cluster. **b**, Lag phase antibiotic survival for wild-type cells after 6 days in stationary phase (WT 6-day stat) or for *hipA7* cells after standard overnight culture. Survival of both populations is significantly higher than for wild-type cells after standard overnight culture (WT) (one-sided Mann–Whitney U test; WT: $n = 14$; WT 6-day stat: $n = 3$ (ampicillin) or $n = 2$ (cipro); *hipA7*: $n = 4$; *metG**: $n = 10$). Data are mean \pm s.d. **c**, 6-day WT and *hipA7* cells have extended lag times. Left, cumulative distribution of colony appearance times. Right, mean and range of appearance times for *hipA7* and 6-day wild-type populations are higher than for wild-type control cells (one-sided Mann–Whitney U test; $P = 0.006$; WT: $n = 11$; WT 6-day: $n = 3$; *hipA7*: $n = 3$; *metG**: $n = 1$). Data are mean (of mean or range) \pm s.d. of replicate

populations. **d**, UMAP of *hipA7* cells ($n = 1,900$ representative cells; 3 biological replicates). **e**, UMAP of wild-type cells ($n = 1,900$ representative cells; 1 replicate). **f**, Percentage of cells in the persister cluster (one-sided Mann–Whitney U test versus WT; WT: $n = 7$; WT ampicillin: $n = 1$; *hipA7*: $n = 3$; WT 6-day: $n = 2$; WT 6-day ampicillin: $n = 1$; tetracycline: $n = 3$; *metG**: $n = 3$). In box plots, the centre line is the median, boxes delineate the 25th and 75th centiles and whiskers extend to minimum and maximum values. Tet, tetracycline. **g**, UMAP of wild-type cells in ampicillin ($n = 36$ cells in the entire dataset; 1 replicate). **h**, UMAP of 6-day wild-type cells ($n = 1,900$ representative cells; 2 biological replicates). **i**, UMAP of 6-day wild-type cells in ampicillin ($n = 754$ cells in the entire dataset; 1 replicate). **j**, UMAP of wild-type cells in tetracycline ($n = 1,900$ representative cells; 3 replicates). In all UMAPs, grey cells show all cells in the study for context. Points within the population of interest are coloured by cell cluster. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.0005$; NS, not significant ($P > 0.05$).

found by this analysis are also expressed in stationary cells (Extended Data Figs. 2a(ii),(iv) and 3a,b). To find markers that are uniquely upregulated in persisters, we compared persisters to multiple neighbouring clusters (Extended Data Fig. 1p). Of these unique markers, the most upregulated genes include *yhaM*, which is involved in cysteine detoxification²³, and again *mdtK*, a putative drug efflux pump²⁴. We looked for known pathways that were enriched in persisters relative to all surrounding clusters (stationary, early/late lag and early exponential). No known pathways met these stringent criteria, but a single gene set—genes upregulated by PspF—was significantly enriched in persisters in all comparisons except versus early lag (Extended Data Fig. 1q). Upregulation by PspF typically occurs during cell envelope stress²⁵

and has been seen in *E. coli* persisters²⁶ and biofilms²⁷. By the same analysis, early lag cells exhibit much more marked upregulation of genes relative to proximal clusters, reflecting a distinct and coordinated expression programme (Extended Data Fig. 1r–u and Supplementary Table 1). Oxidative stress genes upregulated by OxyR and iron uptake genes have previously been discovered as lag phase markers²¹, whereas to our knowledge, upregulation of oligopeptide transport genes is a novel finding (Extended Data Fig. 1s–u). We conclude that *metG** persisters do not turn on lag-specific processes, which may contribute to their failure to resume growth. Knockout of *oxyR* has previously been shown to be sufficient to extend *E. coli* lag times²⁸, whereas non-oxidizing anaerobic conditions decrease lag times²⁹.

Convergence of persister transcriptomes

Having defined a distinct persister state in *metG** cells (Fig. 2a), we tested whether this transcriptional state resembles other models of persistence. HipA is a kinase that targets glutamyl tRNA synthetase (*gltX*)³⁰, and the *hipA7* mutation confers lag-dependent hyper-persistence^{6,31} (Fig. 2b,c). We performed PETRI-seq on wild-type and *hipA7* cells 45 min after dilution from a standard overnight (Fig. 2d,e and Extended Data Fig. 4a,b). For *hipA7* cells, we observed significantly increased occupancy in the same persister cluster as *metG** cells (Fig. 2f). We also analysed the same *hipA7* cells independently and found two cell clusters, one of which clearly mapped to the *metG** persister cluster (Extended Data Fig. 5a–d).

We next explored whether our persister cluster more broadly represents a state of persistence that may occur independently of mutations related to translation. In the wild-type populations sampled around 45 min after dilution from overnight growth, 1.2% of cells are in the persister cluster (Fig. 2f), which matches the persister fraction in ampicillin after the same amount of time (Extended Data Fig. 4b, kill curve). To determine whether wild-type cells in the persister cluster survive antibiotics, we treated them with ampicillin. Ampicillin increased the proportion of wild-type cells in the persister cluster by more than 40 times (Fig. 2f,g) and made persisters the most abundant cell cluster. Together, these data support the existence of shared transcriptional programmes in wild-type and *metG** persisters, although resolution at this low persistence rate is limited.

To increase the persistence rate without mutation, we starved wild-type *E. coli* for 6 days, which increased the persistence rate after dilution and 4 h of antibiotics to nearly 1% with a concurrent increase in lag times (Fig. 2b,c). We applied PETRI-seq to these ‘6-day’ wild-type cells after dilution into fresh medium (Fig. 2h and Extended Data Fig. 4c). Remarkably, 7.4% of these cells were in the persister cluster (Fig. 2f), significantly more than after a standard overnight culture (Fig. 2e). We analysed these cells independently, but they did not clearly group into two clusters (Extended Data Fig. 5e), possibly owing to a more continuous spectrum of states rather than bimodality (also evident in Fig. 2d versus Fig. 2h and in lag times (Fig. 2c)). To confirm that co-clustering was not simply because of the high complexity of the atlas, we analysed a subset of samples and again found enrichment in the persister cluster (Extended Data Fig. 5f–h). We also confirmed that convergence of models to the same persister cluster was not an artefact of low mRNA capture by downsampling cells to exactly 30 mRNAs and discarding cells with fewer than 30 counts (Extended Data Fig. 5i,j). As validation that 6-day wild-type cells in the persister cluster survive antibiotics, ampicillin treatment increased persister cluster occupancy fourfold and made persisters the most abundant cell type (Fig. 2f,i).

Finally, we turned to the clinical uropathogenic *E. coli* (UPEC) isolate strain CFT073 (ref. 32) which exhibits modestly increased antibiotic survival and concurrent lag time increase (Extended Data Fig. 6a,b). We sequenced UPEC cells 1 h after dilution from overnight and 4 h after dilution into ampicillin (Extended Data Fig. 6c). The 1-h population clustered into 4 cell states based on expression of fimbriae or flagellin genes or lack thereof (cluster 0) (Extended Data Fig. 6d,e). None of these clusters could be defined as a persister state, although ampicillin treatment enriched cells in cluster 0 (Extended Data Fig. 6f). We then integrated UPEC cells into our full atlas and found that UPEC cells that remained after ampicillin were highly enriched in the same cluster as MG1655 persister models (Extended Data Fig. 6g–j).

Co-clustering of different persister types is an important signal of convergence of persister states but does not mean that gene expression is identical across models. Although we find that marker genes and pathways significantly overlap, many genes and pathways also differ between persister models (Extended Data Figs. 2, 5k,l and 6k). Similar to *metG** persisters, *hipA7* persisters do not upregulate lag-specific pathways. 6-day wild-type persisters upregulate one of three lag-specific pathways—oligopeptide transport genes (Extended Data Fig. 5l).

Low translation underlies the persister state

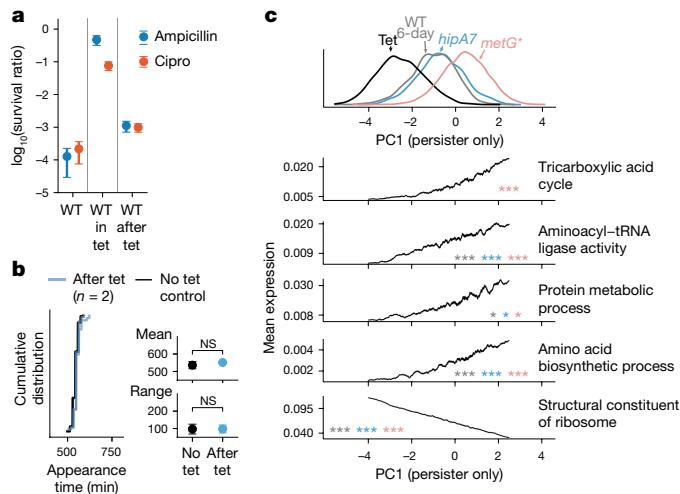
Given a lack of strongly upregulated known pathways to define the persister cluster, we considered an alternative approach to make sense of this transcriptional signature. Specifically, *metG** and *hipA7* mutations are translation-related, so we treated wild-type cells in lag phase with tetracycline, a bacteriostatic antibiotic that inhibits translation³³. Tetracycline-treated cells cluster with persister cells (Fig. 2f,j and Extended Data Fig. 4d), implicating translational deficiency as a defining feature of persister transcriptomes across models. Bulk proteomics of *metG** cells after dilution supports that these persister proteomes are minimally changed from stationary phase (Extended Data Fig. 7a–d), despite the transcriptome responding to fresh medium (Fig. 1f and Extended Data Fig. 7f). As further confirmation of translational deficiency in *metG** persisters, we analysed *metG** cells that expressed both *p*cysK-GFP, a reporter for persisters (Extended Data Fig. 3b) and *p*LlacO1-RFP, a proxy for protein expression. Assayed persister cells exhibited reduced translation relative to growing cells (Extended Data Fig. 3f,g). Previous studies have implicated low protein expression in persistence, specifically finding slow fluorescent protein production in *hipA7* persisters³⁴ and showing that wild-type cells with low protein expression are more likely to survive antibiotics^{13,35}.

We next assayed the role of transcriptional deficiency in persistence. Inhibition of transcription by rifampicin, a bacteriostatic antibiotic that targets RNA polymerase³⁶ (Extended Data Fig. 8a–c), revealed that cells without active transcription retain ribosomal RNA (rRNA) and thus can be captured by PETRI-seq (Extended Data Fig. 8d) but have substantially lower mRNA counts than non-rifampicin-treated cells (Extended Data Fig. 8e). Comparing rifampicin-treated cells to persister models shows active transcription in persister cells (Extended Data Fig. 8f). Although PETRI-seq captures fewer mRNAs per cell for persisters than for non-persisters, comparison of *metG** persisters to tetracycline treatment indicates that translational deficiency is sufficient to explain lower transcript counts (Extended Data Fig. 8g).

Tetracycline pharmacologically recapitulates a persister-like state but does not lead to sustained cell dormancy (Fig. 3a,b). We thus reasoned that genes that were differentially expressed between self-maintained persister cells and tetracycline-induced tolerant cells could be critical for persister formation and maintenance. Genes upregulated in *metG**, *hipA7* and 6-day wild-type persister cells versus tetracycline-treated cells significantly overlap but are not identical (Extended Data Fig. 9a). We computed PC1 for the persister cluster alone, which weakly separated sample types (Fig. 3c). Multiple pathways were significantly correlated with PC1 and differentially expressed in persisters relative to tetracycline-treated cells (Fig. 3c and Extended Data Fig. 9b). Upregulation of aminoacyl-tRNA ligase activity across persister types suggests a compensatory response to disrupted tRNA aminoacylation in the mutants. Surprisingly, *metG* itself was not one of the genes that was upregulated in *metG** persisters (Extended Data Fig. 9c), indicating failure to compensate for a hypomorphic mutation (Extended Data Fig. 9d). Other pathways that were upregulated in persister cells include protein metabolism, amino acid biosynthesis and tricarboxylic acid (TCA) cycle (*metG** only) (Fig. 3c). In the other direction, cells treated with tetracycline expressed more ribosomal components, which is consistent with downstream inhibition of translation relative to persister cells.

CRISPRi to identify persistence drivers

Having identified pathways with expression patterns suggestive of a role in persistence, we then carried out a systematic genetic interrogation to agnostically identify factors that are causal to persister formation and maintenance. We used CRISPR adaptation-mediated library manufacturing (CALM) to generate a comprehensive set of CRISPR RNAs (crRNAs) covering all *E. coli* genes³. We then used CRISPRi



to probe the contribution of every gene to lag phase duration and antibiotic survival in *metG**, *hipA7* and wild-type MG1655 cells (Fig. 4a). For reference, we measured the contribution of every gene to exponential growth and confirmed that essential genes were substantially depleted (Extended Data Fig. 9f,g). Fitness effects were highly correlated for *metG** and wild-type cells in exponential phase, when *metG** has no phenotype (Extended Data Fig. 9h).

To identify genes involved in lag-dependent persistence, we focused primarily on gene perturbations that shortened lag times. For *metG**, knockdown of two genes, *lon* and *yggE*, markedly shortened lag times (Fig. 4b), though these genes do not affect wild-type or *hipA7* lag times (Fig. 4b–d). *Lon* is a heat shock-induced protease with a disputed role in persistence^{37,38}, and *YggE* is an uncharacterized protein that has not previously been implicated in persistence or lag phase. In addition to these genes, CRISPRi perturbation of five other genes also significantly shortened lag phase in all *metG** replicates and reduced *metG** antibiotic survival (Fig. 4d and Extended Data Fig. 10a). Notably, these other driver genes are shared across genotypes, indicating that mechanisms of persister formation are simultaneously distinct and overlapping (Extended Data Fig. 9i). Specifically, crRNA targeting of *rpoH*³⁹, which encodes the heat shock-induced σ^{32} , and *sucA*, which encodes TCA cycle enzyme 2-oxoglutarate decarboxylase⁴⁰, shortened lag times across genotypes (Fig. 4d), as did perturbation of three genes encoding primosomal proteins (Extended Data Fig. 10a). The primosome is essential for propagating plasmids with the ColE1 origin⁴¹, including our crRNA plasmid, which may confound implication of these genes in persistence.

Notably, the only highly significant gene perturbation that specifically shortened *hipA7* lag time was *hipA* itself (Fig. 4c). Conversely, repression of *gltX*, which encodes the target of the HipA kinase,

specifically lengthened *hipA7* lag times (Fig. 4c). Repression of *metG* similarly extended lag times more in *metG** cells than in wild-type cells (Fig. 4b).

We tested the hypothesis that genes found by scRNA-seq to be upregulated in persister cells relative to tetracycline-treated cells would be candidate drivers of persistence. Indeed, *lon*, *yggE*, *rpoH* and *sucA* are significantly upregulated in *metG** persister cells, whereas *rpoH*, *sucA* and *hipA* are upregulated in *hipA7* persisters (Fig. 4e and Extended Data Fig. 9e). To consider a larger set of genes, we relaxed our criteria to include all 51 genes that were significantly enriched in *metG** lag in at least two replicates. Of these, 21 were significantly upregulated in *metG** persister cells (Extended Data Fig. 9j). This highly significant concordance ($P < 10^{-4}$) cross-validates the two independent approaches and supports our driver gene hypothesis. Considering these two datasets together gives key biological insight, as shared hits are likely to function in persister cells and/or proximal to their formation. Additional gene set comparisons substantiated that transcriptional markers significantly overlap with persistence drivers across genotypes (Extended Data Fig. 9k–n). Notably, many top marker genes found by PETRI-seq did not prove functionally important by CRISPRi (Extended Data Fig. 10b), demonstrating the importance of our dual approach and supporting a model in which causal genes are enriched but still a minority of those expressed.

On the pathway level, we similarly saw concordance between CRISPRi and expression data. crRNA targeting of TCA cycle-related genes shortened lag times in all cell types (Extended Data Fig. 10c). TCA cycle activity has previously been implicated in lag-dependent persistence in wild-type *E. coli*^{29,42}. We also found genotype-specific causal pathways. Targeting protein metabolism genes significantly reduced *metG** lag time (Extended Data Fig. 10d). This gene set is expressed in *metG**, *hipA7* and 6-day wild-type persisters (Fig. 3c); it includes proteases but not top driver gene *lon*. The targeting of some pathways, including amino acid biosynthesis, specifically reduced wild-type lag and antibiotic survival (Extended Data Fig. 10e). Amino acid biosynthetic genes are also expressed across persister types (Fig. 3c) and have previously been found to affect *E. coli* tolerance during stationary phase⁴³. Other expressed pathways (Fig. 3c and Extended Data Fig. 9b) may be important for persister resuscitation, as their repression lengthened lag times, although this may be due to slower growth (Extended Data Fig. 10f).

lon and *yggE* in *metG** hyper-persistence

Of the 7 top drivers of *metG** persistence (Fig. 4d and Extended Data Fig. 10a), we selected *lon*, *yggE* and *priA* for validation by genomic deletion. Of the genes not selected, *sucA* is part of the 2-oxoglutarate dehydrogenase complex, loss of which has already been shown to reduce lag-dependent persistence in wild-type cells²⁹. *dnaC*, *dnaT* and *rpoH* are essential genes⁴⁴. Excluded genes are also functionally related to selected genes (*dnaC/T* is part of the primosome with *priA*, and *lon* is induced by *rpoH*). As detailed below, deletion of *lon* or *yggE* significantly reduced *metG** lag-dependent persistence. However, deletion of *priA* did not affect *metG** lag and was not pursued further (Extended Data Fig. 11a).

Lon protease, the top hit identified by CRISPRi (Fig. 4b,d), has been implicated in persistence previously, through a specific SulA-dependent mechanism. SulA is a cell division inhibitor that accumulates during antibiotic treatment and prevents growth resumption unless cleared by *Lon*^{38,45}. SulA accumulation can explain why *lon* perturbation reduces antibiotic survival across genotypes but not why it reduces lag time in *metG** cells (Fig. 4d). As expected, single deletion of *lon* strongly reduced antibiotic survival in both wild-type and *metG** cells (Extended Data Fig. 11b). Furthermore, lag phase antibiotic survival of *metG*-Δlon-ΔsulA* cells decreased more than 100,000-fold relative to the *metG** parental strain (Fig. 5a), indicating a major *sulA*-independent

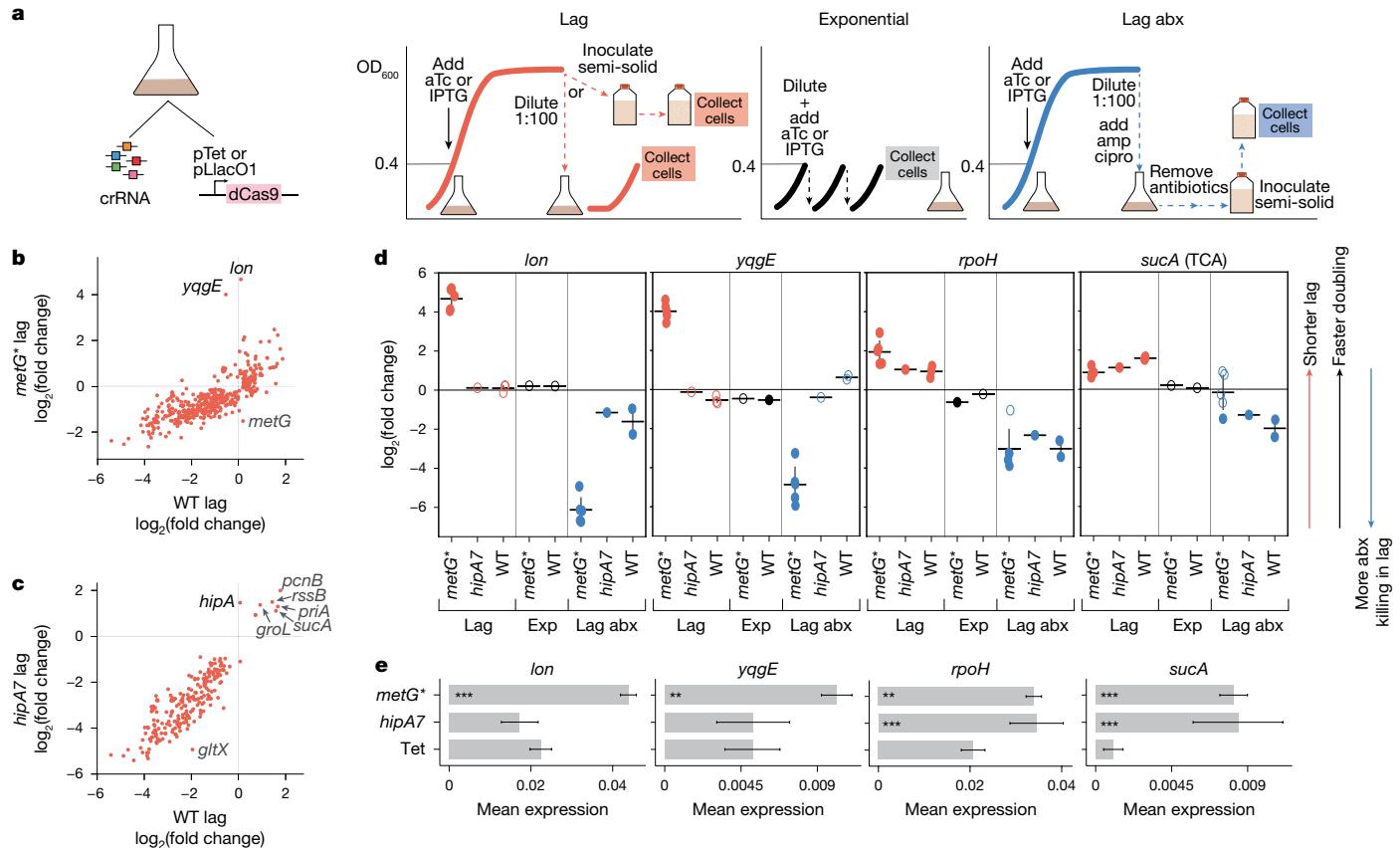


Fig. 4 | Genome-wide CRISPRi screen reveals genes and pathways that are causal to lag-dependent persistence. **a**, Schematic of CRISPRi screen (Methods and Supplementary Table 5). Abx, antibiotics; amp, ampicillin; aTc, anhydrotetracycline. **b**, Correlation between gene perturbation effects on *metG** lag versus wild-type lag ($\rho = 0.84$; $n = 413$ genes with false discovery rate (FDR) < 0.1 in at least 2 replicates of either comparison). **c**, Correlation between gene perturbation effects on *hipA7* lag versus wild-type lag ($\rho = 0.87$; $n = 215$ genes with Bonferroni-corrected P value < 0.05 (see Methods; more stringent because there is only 1 replicate)). **d**, Gene perturbations enriched after lag

phase outgrowth of *metG** cells in all replicates. Solid circles indicate significant enrichment or depletion (Methods; FDR = 0.1). Data are mean \pm s.d. Exp, exponential phase. **e**, Mean expression of genes in **d** for *metG** persisters, *hipA7* persisters or tetracycline-treated cells in the persister cluster. Genes that are significantly overexpressed in persisters versus tetracycline-treated cells are indicated (two-sided Mann–Whitney U test; $n = 4,576$ tet cells, 19,611 *metG** cells and 975 *hipA7* cells). Data are mean \pm s.e.m. Expression across clusters is shown in Extended Data Fig. 9r.

role for Lon. Similar to wild-type cells, *metG**- Δ *lon*- Δ *sulA* cells exhibit short, uniform lag times (Extended Data Fig. 11c). In the wild-type background, *lon* and *sulA* deletion decreased antibiotic survival 5.2-fold (Fig. 5a, grey versus black).

Lon can be pharmacologically inhibited by the small molecule bortezomib⁴⁶. Bortezomib has been shown to reduce exponential phase *E. coli* survival in response to ciprofloxacin⁴⁷ because of SulA accumulation. We found that without *sulA*, bortezomib also reduces antibiotic survival of *metG** cells, further validating a *sulA*-independent mechanism and supporting protease inhibitors as potentially useful antibiotic adjuvants (Extended Data Fig. 11d,e).

We next used bortezomib to temporally place the role of Lon in persister formation and maintenance. Using *metG**- Δ *sulA* cells, we added bortezomib either before cells reached stationary phase, during stationary phase, or at the start of lag phase (Fig. 5b, left). As a proxy for lag-dependent persistence, we measured population lag times. Bortezomib had the strongest effect when added before stationary phase but still had a significant effect during stationary phase (Fig. 5b, right, black and pink lines). However, inhibiting Lon at the start of lag phase had no effect (green line), indicating that Lon is key to persister formation during stationary phase. We then used proteomics to find candidate targets stabilized by deletion of Lon protease (Extended Data Fig. 12a,b and Supplementary Table 2). Proteomics implicates loss of iron–sulfur cluster assembly proteins IscS and IscU as well as reduction

in iron–sulfur cluster binding proteins overall as potentially important during *metG** stationary phase. Notably, Lon deletion also decreased expression of TCA cycle proteins, which could be an additional contributor to reduced persistence (Extended Data Figs. 10c and 12c,d).

yqqE, the other top hit in our screen, is uncharacterized and to our knowledge has not been previously implicated in persistence. We introduced *yqqE* deletions into *metG** and wild-type cells. *metG**- Δ *yqqE* survival to antibiotics was between 2,000- and 5-fold lower than the parental *metG** strain (Fig. 5c) with a substantial dependence on the length of the preceding overnight culture (Fig. 5c and Extended Data Fig. 11f). This strong correlation implicates YqqE activity as most important early in stationary phase, although after growth for over 24 h, *metG**- Δ *yqqE* cells still exhibit significantly reduced lag phase antibiotic survival (Fig. 5c, right and Extended Data Fig. 11g). *yqqE* deletion shortened lag times in the *metG** background (Fig. 5d, solid orange versus solid blue). We also expressed the endogenous *yqqE* locus on a high-copy plasmid (pYqqE+). In both *metG** and *metG**- Δ *yqqE* backgrounds, pYqqE+ increased the right tail of colony appearance times relative to each parental strain (Fig. 5d, dotted versus solid lines). Likely related, these overexpression strains give rise to fewer colony-forming units (CFUs) after overnight growth (Extended Data Fig. 11h). The missing CFUs may be viable but nonculturable cells, a more extreme version of the persister state⁴⁸. As expected, pYqqE+ increased antibiotic survival in *metG**- Δ *yqqE* cells (Extended Data Fig. 11i).

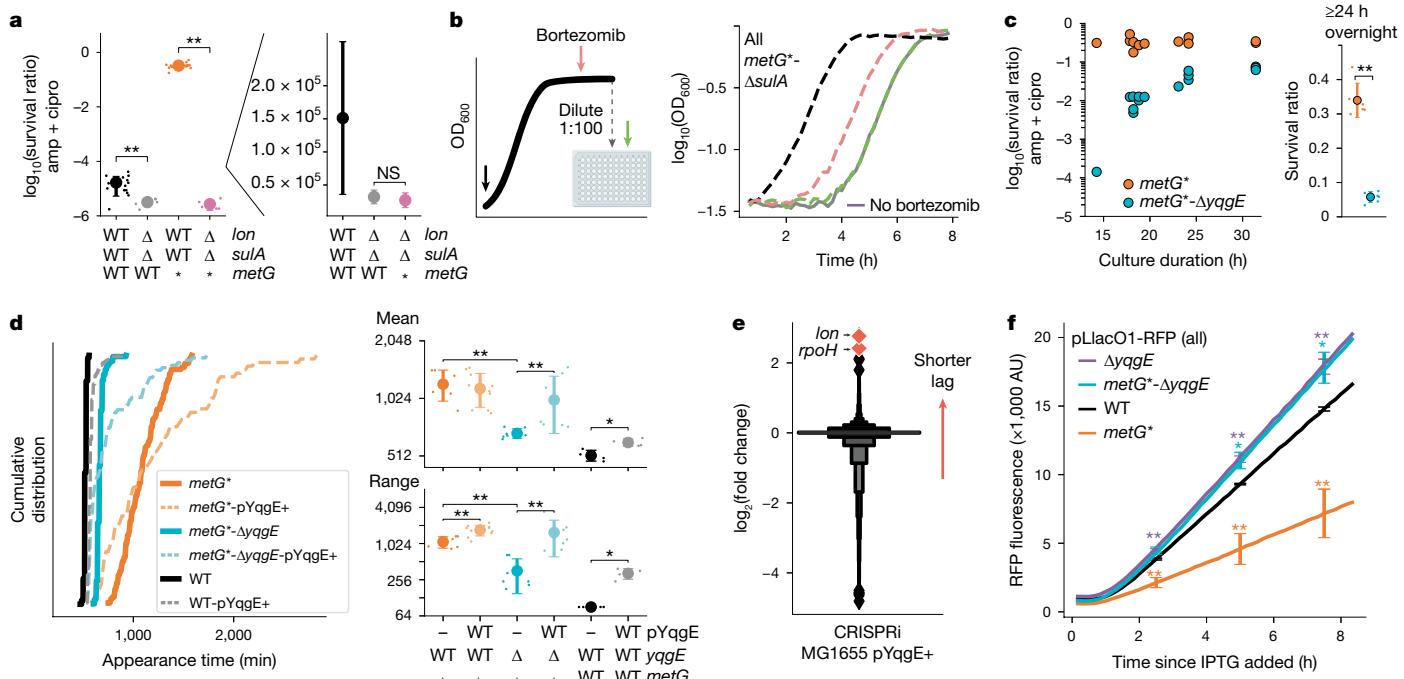


Fig. 5 | Lon protease and YqgE have key roles in lag-dependent hyper-persistence. **a**, *lon* deletion decreases lag phase persistence of wild-type and *metG** cells (one-sided Mann–Whitney U test; $n = 14$ (wild type), $n = 5$ (*Δlon-ΔsulA*, *metG**-*Δlon-ΔsulA*), $n = 15$ (*metG**)) independent of *sulA*. Data are mean \pm s.d. **b**, The Lon inhibitor bortezomib reduces *metG** lag phase when added before or during stationary phase. Left, cartoon of the experiment. Bortezomib was added either when colonies were picked (black arrow), after overnight growth but 3 h before dilution (pink arrow), or at dilution (green arrow). Right, growth curves after dilution and overnight culture. Line colours correspond to the arrows in the left panel. Adding bortezomib before or during the overnight culture significantly reduced lag times ($n = 3$; $P = 0.03$; one-sided Mann–Whitney U test) but did not affect doubling time. **c**, *yqgE* deletion decreases lag phase persistence of *metG** cells. Left, survival of *metG**-*ΔyqgE* cells is strongly correlated with the length of the preceding culture (two-sided Spearman's

$P = 0.9$; $P = 10^{-6}$), whereas survival of *metG** cells is not ($P = 0.1$; $P = 0.7$). Right, after growth for more than 24 h, lag phase survival is lower for *metG**-*ΔyqgE* cells than for *metG** cells (one-sided Mann–Whitney U test; $P = 0.002$, $n = 6$ biological replicates; fold change = 6). Data are mean \pm s.d. **d**, Left, representative cumulative distributions of colony appearance times. Right, mean (top) and range (bottom) of appearance times for replicate populations (one-sided Mann–Whitney U test; $n = 6$ (*metG**-pRFP+), $n = 10$ (*metG**-pYqgE+), $n = 7$ (*metG**-*ΔyqgE*-pRFP+; *metG**-*ΔyqgE*-pYqgE+), $n = 4$ (WT-pRFP+ and WT-pYqgE+); pRFP+ is omitted from labels for brevity). Data are mean (of mean or range) \pm s.d. of replicate populations. **e**, Distribution of CRISPRi effects on lag times of wild-type cells overexpressing YqgE ($n = 4,486$ genes). **f**, Protein expression in stationary phase. Significance is shown for each cell type relative to wild type at the indicated timepoint (one-sided Mann–Whitney U test; $n = 5$ (*metG*) or 6 (all others) biological replicates). Data are mean \pm s.e.m. AU, absorbance units.

In contrast to *metG**, but matching our screen (Fig. 4d), deleting *yqgE* in wild-type cells did not affect lag phase antibiotic survival or lag times beyond a minor fitness decrease (Extended Data Fig. 11j–l). Similarly, *yqgE* deletion had no effect on persistence in 6-day wild-type cells (Extended Data Fig. 11m). However, *yqgE* overexpression in wild-type cells did increase lag times (Fig. 5d; dotted grey versus black) and lag phase antibiotic survival (Extended Data Fig. 11n), indicating that YqgE can promote dormancy and lag-dependent persistence beyond the *metG** context. To find other factors critical for promotion of dormancy by YqgE, we carried out a CRISPRi screen in wild-type cells overexpressing *yqgE*. Resembling the *metG** context, *lon* and *rpoH* were top hits for reversing YqgE-promoted dormancy (Fig. 5e). A deletion strain confirmed that *lon* is epistatic to YqgE-induced dormancy (Extended Data Fig. 11o,p). Although the enzymatic function of YqgE has yet to be shown experimentally, in silico prediction suggests that it has protein disulfide isomerase activity⁴⁹ (Supplementary Fig. 2).

In investigating the timing of translational deficiency in *metG** persisters, we found that *metG** cells have reduced protein expression during stationary phase (Fig. 5f, orange versus black) prior to exhibiting lag-dependent hyper-persistence. Specifically, we tested this by transforming an inducible RFP and then growing cultures to stationary phase without inducer. Once cells reached stationary phase, we added isopropyl-β-D-thiogalactopyranoside (IPTG) to induce RFP expression. Given the apparent role of *yqgE* in promoting cell dormancy, we interrogated whether it may contribute to downregulation of translation.

By carrying out the same experiment with *ΔyqgE* and *metG**-*ΔyqgE* cells, we found that, indeed, the reduced protein expression in *metG** depends on *yqgE* (Fig. 5f, blue versus orange) and that stationary wild-type cells exhibit higher protein expression when *yqgE* is knocked out (Fig. 5f, purple versus black). We reason that under certain conditions, YqgE represses protein expression, which may be critical to extending post-starvation lag time. To determine whether *yqgE* could have such a role across multiple species, we searched for homologues in 2,421 microbial genomes from diverse taxa. We detected *yqgE* homologues in 35% of genomes, which was more frequent than 70% of all MG1655 genes (Extended Data Fig. 11q). Previous work has implicated *algH*, the homologue of *yqgE* in *Pseudomonas aeruginosa*, in regulation of virulence factors in that species⁵⁰, supporting a global regulatory role for YqgE-like proteins across Proteobacteria.

Discussion

Here we used unbiased systems approaches to characterize *E. coli* persister states and genetic factors that promote persister formation and maintenance. We applied PETRI-seq² to wild-type and hyper-persistent *E. coli* to generate an atlas of growth states and discover convergent persister states defined primarily by translational deficiency. To discover driver genes underlying persistence, we carried out comprehensive CRISPRi screening³. In wild-type cells, our screen supported previously discovered driver pathways, including the TCA cycle²⁹. In the

*metG** mutant, we found Lon protease and YqgE to be major drivers of hyper-persistence.

scRNA-seq facilitates a previously unattainable view of persister cell states and reveals that lag-dependent persister cells across multiple genotypes are in a transitional state between stationary and exponential phases (Fig. 1f and Extended Data Fig. 1n,o). Rather than markedly upregulating specific known pathways, persisters seem to be in a dysregulated state in which transcriptional patterns primarily reflect collateral responses to low translation (Fig. 2j). This convergent signature implicates translational deficiency as a core feature of persistence. Our work was conducted under laboratory conditions and primarily using non-pathogenic MG1655, but convergence of UPEC strain CFT073 persisters to the same persister cluster (Extended Data Fig. 6i,j) suggests broad relevance of these findings. Beyond basic phenomenology, scRNA-seq provides insight into causal genes and could be used to tailor drug adjuvants (Fig. 4d,e). As MG1655 and CFT073 strains have both been passaged for decades⁵¹, it will be critical for future research to extend scRNA-seq to fresh clinical isolates.

It is also important to note that PETRI-seq captures only a fraction of transcripts per cell, and limited resolution may affect our ability to detect upregulated genes and pathways. This is particularly true for the wild-type MG1655 and CFT073 persisters, which could only be isolated in small numbers after antibiotic enrichment (Fig. 2g and Extended Data Fig. 6i). Total transcript counts for each cell type are included in Supplementary Table 4 and show a range of transcriptome coverage. Notably, we detect more transcripts in *metG** persisters than in early lag cells, which express multiple unique pathways found by PETRI-seq (Extended Data Fig. 1s–u). Thus, our ability to detect upregulated pathways in early lag cells but not in *metG** persisters supports our key conclusion that the persister state is primarily defined by translational deficiency rather than specific upregulation of defined pathways.

Although we find that transcriptional states across persister models are convergent, CRISPRi reveals that major upstream mechanisms diverge (Fig. 4). For *metG**, *lon* and *yqgE* are key drivers (Fig. 4b). By contrast, we find *hipA* itself to be the only major *hipA7*-specific driver gene (Fig. 4c). Both *metG** and *hipA7* affect tRNA aminoacylation, but CRISPRi hits overlap as much between these two strains as with the wild type (Extended Data Fig. 9i). This surprising result suggests that lag-dependent persisters are generated by diverse upstream mechanisms that nevertheless converge to similar transcriptional states. Notably, the *metG** mutation does not seem to directly prevent translation, as *lon* deletion completely reverses hyper-persistence (Fig. 5a) and *yqgE* deletion recovers stationary phase translation (Fig. 5f).

Causal genes for *metG**, *hipA7* and wild-type persistence also overlap, as seen for the TCA cycle (Extended Data Figs. 9i and 10c). The effect of TCA cycle disruption on wild-type persistence is most profound. Previous work has implicated the TCA cycle in persistence and attributed this to self-digestion to generate reducing power²⁹. It has also been found that lag-dependent persisters undergo more divisions as a population enters stationary phase, which may lead to stochastic generation of cells with very low protein levels⁵². Together, the combination of self-digestion and reductive division could lead to rare cells with extremely low protein abundance and translation rates. Once fresh nutrients are available, positive feedback from translation can exacerbate small differences and establish bimodality in the population. Compounding this divergence, translationally deficient persisters do not appear to express important lag pathways, including redox defence and iron uptake (Extended Data Figs. 1s–u and 5l).

In *metG**, hyper-persistence requires Lon, a ubiquitous protease that has evolved to degrade a broad set of targets, including misfolded proteins⁵³, antitoxins⁵⁴, ribosomal proteins⁵⁵ and sulfur-assimilation proteins⁵⁶. Ribosomal protein degradation could explain the role of Lon in persistence but is not supported by our proteomics data (Extended Data Fig. 12e,f). Instead, degradation of other targets, such as iron–sulfur cluster assembly proteins (Extended Data Fig. 12a,b), may leave cells

lacking key protein products that are needed to resume translation upon addition of fresh nutrients.

Finally, we identified YqgE as a major driver of *metG** persistence that is sufficient to increase lag times and persistence in wild-type *E. coli* (Fig. 5d and Extended Data Fig. 11n). *Lon* is epistatic to YqgE activity (Fig. 5e and Extended Data Fig. 11o,p), suggesting that YqgE may modulate Lon-mediated proteolysis. Functional annotation of YqgE suggests that it has disulfide isomerase activity (Supplementary Fig. 2). Of note, a disulfide redox switch is known to tune Lon activity⁵⁷. We hypothesize that YqgE acts as a checkpoint that responds to starvation, as its function is most important early in stationary phase (Fig. 5c). Truncated *metG* is likely to have reduced enzymatic activity⁵⁸ that could become limiting upon nutrient depletion. Many possible signals could then trigger YqgE activity, including but not limited to activation of the stringent response⁵⁹ or loss of translation fidelity⁶⁰. YqgE could then promote Lon-mediated degradation of critical targets to reduce translation (Fig. 5f). Without YqgE, *metG** cells remain partially hyper-persistent (Fig. 5c). In the absence of checkpoint YqgE, baseline Lon activity alone could ultimately lead to depletion of proteins critical for ramping up translation after starvation. In both cases, when fresh nutrients are added, low translation rates across the population lead to high persistence rates. As individual cells produce more proteins, positive feedback can drive a switch-like return to exponential growth, establishing a bimodal distribution of cell states.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41586-024-08124-2>.

1. Bigger, J. W. Treatment of staphylococcal infections with penicillin—by intermittent sterilisation. *Lancet* **2**, 497–500 (1944).
2. Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat. Microbiol.* **5**, 1192–1201 (2020).
3. Jiang, W., Oikonomou, P. & Tavazoie, S. Comprehensive genome-wide perturbations via CRISPR adaptation reveal complex genetics of antibiotic sensitivity. *Cell* **180**, 1002–1017.e1031 (2020).
4. Olivares, A. O., Baker, T. A. & Sauer, R. T. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat. Rev. Microbiol.* **14**, 33–44 (2016).
5. Balaban, N. Q. et al. Definitions and guidelines for research on antibiotic persistence. *Nat. Rev. Microbiol.* **17**, 441–448 (2019).
6. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. Bacterial persistence as a phenotypic switch. *Science* **305**, 1622–1625 (2004).
7. Fridman, O., Goldberg, A., Ronin, I., Shores, N. & Balaban, N. Q. Optimization of lag time underlies antibiotic tolerance in evolved bacterial populations. *Nature* **513**, 418–421 (2014).
8. Khare, A. & Tavazoie, S. Extreme antibiotic persistence via heterogeneity-generating mutations targeting translation. *mSystems* **5**, e00847 (2020).
9. Girgis, H. S., Harris, K. & Tavazoie, S. Large mutational target size for rapid emergence of bacterial persistence. *Proc. Natl Acad. Sci. USA* **109**, 12740–12745 (2012).
10. Moyed, H. S. & Bertrand, K. P. *hipA*, a newly recognized gene of *Escherichia coli* K-12 that affects frequency of persistence after inhibition of murein synthesis. *J. Bacteriol.* **155**, 768–775 (1983).
11. Van den Berg, B. et al. Frequency of antibiotic application drives rapid evolutionary adaptation of *Escherichia coli* persistence. *Nat. Microbiol.* **1**, 16020 (2016).
12. Schumacher, M. A. et al. HipBA-promoter structures reveal the basis of heritable multidrug tolerance. *Nature* **524**, 59 (2015).
13. Shah, D. et al. Persisters: a distinct physiological state of *E. coli*. *BMC Microbiol.* **6**, 53 (2006).
14. Keren, I., Shah, D., Spoerling, A., Kaldalu, N. & Lewis, K. Specialized persister cells and the mechanism of multidrug tolerance in *Escherichia coli*. *J. Bacteriol.* **186**, 8172–8180 (2004).
15. Kuchina, A. et al. Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371**, eaba5257 (2021).
16. Ma, P. et al. Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell* **186**, 877–891.e814 (2023).
17. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
18. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR–Cas system. *Nucleic Acids Res.* **41**, 7429–7437 (2013).
19. Levin-Reisman, I. et al. Automated imaging with ScanLag reveals previously undetectable bacterial growth phenotypes. *Nat. Methods* **7**, 737–739 (2010).

20. Chen, H., Shiroguchi, K., Ge, H. & Xie, X. S. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.* **11**, 781 (2015).
21. Rolfe, M. D. et al. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *J. Bacteriol.* **194**, 686–701 (2012).
22. Brandi, A. & Pon, C. L. Expression of *Escherichia coli* cspA during early exponential growth at 37°C. *Gene* **492**, 382–388 (2012).
23. Shimada, T., Tanaka, K. & Ishihama, A. Transcription factor DecR (YbaO) controls detoxification of L-cysteine in *Escherichia coli*. *Microbiology* **162**, 1698–1707 (2016).
24. Morita, Y. et al. NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrob. Agents Chemother.* **42**, 1778–1782 (1998).
25. Darwin, A. J. Stress relief during host infection: the phage shock protein response supports bacterial virulence in various ways. *PLoS Pathog.* **9**, e1003388 (2013).
26. Vega, N. M., Allison, K. R., Khalil, A. S. & Collins, J. J. Signaling-mediated bacterial persister formation. *Nat. Chem. Biol.* **8**, 431 (2012).
27. Beloin, C. et al. Global impact of mature biofilm lifestyle on *Escherichia coli* K-12 gene expression. *Mol. Microbiol.* **51**, 659–674 (2004).
28. Johnson, J. R., Clabots, C. & Rosen, H. Effect of inactivation of the global oxidative stress regulator oxyR on the colonization ability of *Escherichia coli* O1:K1:H7 in a mouse model of ascending urinary tract infection. *Infect. Immun.* **74**, 461–468 (2006).
29. Orman, M. A. & Brynildsen, M. P. Inhibition of stationary phase respiration impairs persister formation in *E. coli*. *Nat. Commun.* **6**, 7983 (2015).
30. Germain, E., Castro-Roa, D., Zenkin, N. & Gerdes, K. Molecular mechanism of bacterial persistence by HipA. *Mol. Cell* **52**, 248–254 (2013).
31. Korch, S. B., Henderson, T. A. & Hill, T. M. Characterization of the *hipA7* allele of *Escherichia coli* and evidence that high persistence is governed by (p)ppGpp synthesis. *Mol. Microbiol.* **50**, 1199–1213 (2003).
32. Mobley, H. L. et al. Pyelonephritogenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. *Infect. Immun.* **58**, 1281–1289 (1990).
33. Chopra, I. & Roberts, M. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. *Microbiol. Mol. Biol. Rev.* **65**, 232–260 (2001).
34. Gefen, O., Gabay, C., Mumcuoglu, M., Engel, G. & Balaban, N. Q. Single-cell protein induction dynamics reveals a period of vulnerability to antibiotics in persister bacteria. *Proc. Natl Acad. Sci. USA* **105**, 6145–6149 (2008).
35. Kwan, B. W., Valenta, J. A., Benedik, M. J. & Wood, T. K. Arrested protein synthesis increases persister-like cell formation. *Antimicrob. Agents Chemother.* **57**, 1468–1473 (2013).
36. Wehrli, W. & Staehelin, M. Actions of the rifamycins. *Bacteriol. Rev.* **35**, 290–309 (1971).
37. Harms, A., Fino, C., Sorensen, M. A., Semsey, S. & Gerdes, K. Prophages and growth dynamics confound experimental results with antibiotic-tolerant persister cells. *mBio* **8**, e01964-17 (2017).
38. Theodore, A., Lewis, K. & Vulić, M. Tolerance of *Escherichia coli* to fluoroquinolone antibiotics depends on specific components of the SOS response pathway. *Genetics* **195**, 1265–1276 (2013).
39. Grossman, A. D., Erickson, J. W. & Gross, C. A. The *htpR* gene product of *E. coli* is a sigma factor for heat-shock promoters. *Cell* **38**, 383–390 (1984).
40. Darlison, M. G., Spencer, M. E. & Guest, J. R. Nucleotide sequence of the *sucA* gene encoding the 2-oxoglutarate dehydrogenase of *Escherichia coli* K12. *Eur. J. Biochem.* **141**, 351–359 (1984).
41. Lee, E. H. & Kornberg, A. Replication deficiencies in *priA* mutants of *Escherichia coli* lacking the primosomal replication n' protein. *Proc. Natl Acad. Sci. USA* **88**, 3029–3032 (1991).
42. Cesar, S., Willis, L. & Huang, K. C. Bacterial respiration during stationary phase induces intracellular damage that leads to delayed regrowth. *iScience* **25**, 103765 (2022).
43. Shan, Y., Lazinski, D., Rowe, S., Camilli, A. & Lewis, K. Genetic basis of persister tolerance to aminoglycosides in *Escherichia coli*. *mBio* **6**, e00078–00015 (2015).
44. Goodall, E. C. A. et al. The essential genome of *Escherichia coli* K-12. *mBio* **9**, e02096-17 (2018).
45. Mizusawa, S. & Gottesman, S. Protein degradation in *Escherichia coli*: the *lon* gene controls the stability of *sulA* protein. *Proc. Natl Acad. Sci. USA* **80**, 358–362 (1983).
46. Liao, J. H. et al. Structures of an ATP-independent Lon-like protease and its complexes with covalent inhibitors. *Acta Crystallogr. D* **69**, 1395–1402 (2013).
47. Babin, B. M. et al. Leveraging peptide substrate libraries to design inhibitors of bacterial Lon protease. *ACS Chem. Biol.* **14**, 2453–2462 (2019).
48. Kim, J. S., Chowdhury, N., Yamasaki, R. & Wood, T. K. Viable but non-culturable and persistence describe the same bacterial stress state. *Environ. Microbiol.* **20**, 2038–2048 (2018).
49. Zhang, C., Freddolino, P. L. & Zhang, Y. COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information. *Nucleic Acids Res.* **45**, W291–W299 (2017).
50. Schlichter, D., Kubo, M., Shankar, S. & Chakrabarty, A. M. Regulation of nucleoside diphosphate kinase and secretable virulence factors in *Pseudomonas aeruginosa*: roles of algR2 and algH. *J. Bacteriol.* **177**, 2469–2474 (1995).
51. Hogins, J., Xuan, Z., Zimmern, P. E. & Reitzler, L. The distinct transcriptome of virulence-associated phylogenetic group B2 *Escherichia coli*. *Microbiol. Spectrum* **11**, e0208523 (2023).
52. Bakshi, S. et al. Tracking bacterial lineages in complex and dynamic environments with applications for growth control and persistence. *Nat. Microbiol.* **6**, 783–791 (2021).
53. Tomoyasu, T., Mogk, A., Langen, H., Goloubinoff, P. & Bukau, B. Genetic dissection of the roles of chaperones and proteases in protein folding and degradation in the *Escherichia coli* cytosol. *Mol. Microbiol.* **40**, 397–413 (2001).
54. Christensen, S. K., Pedersen, K., Hansen, F. G. & Gerdes, K. Toxin–antitoxin loci as stress-response-elements: ChpAK/MazF and ChpBk cleave translated RNAs and are counteracted by tmRNA. *J. Mol. Biol.* **332**, 809–819 (2003).
55. Kuroda, A. et al. Role of inorganic polyphosphate in promoting ribosomal protein degradation by the Lon protease in *E. coli*. *Science* **293**, 705–708 (2001).
56. Arends, J. et al. An integrated proteomic approach uncovers novel substrates and functions of the Lon protease in *Escherichia coli*. *Proteomics* **18**, 1800080 (2018).
57. Nishii, W. et al. A redox switch shapes the Lon protease exit pore to facultatively regulate proteolysis. *Nat. Chem. Biol.* **11**, 46–51 (2015).
58. Crepin, T., Schmitt, E., Blanquet, S. & Mechulam, Y. Structure and function of the C-terminal domain of methionyl-tRNA synthetase. *Biochemistry* **41**, 13003–13011 (2002).
59. Boutte, C. C. & Crosson, S. Bacterial lifestyle shapes stringent response activation. *Trends Microbiol.* **21**, 174–180 (2013).
60. Netzer, N. et al. Innate immune and chemically triggered oxidative stress modifies translational fidelity. *Nature* **462**, 522–526 (2009).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

Methods

Bacterial strains and culture conditions

E. coli MG1655, UPEC CFT073, and derivative mutant strains (Supplementary Table 6) were routinely grown at 37 °C with shaking. One millilitre of culture was grown in 14-ml round-bottom culture tubes shaking at 300 rpm, or larger volumes were grown in flasks at lower shaking speed. For all liquid experiments, we used supplemented M9 medium (SM9)⁸ (1× M9 salts (DF0485-17, Fisher Scientific), 0.4% glucose, 2 mM MgSO₄, 0.1 mM CaCl₂, 2 μM ferric citrate, 3.1 g l⁻¹ Neidhardt Supplement Mixture (NSM01, ForMedium) and micronutrient supplement as previously described⁶¹). Neidhardt Supplement was autoclaved for 20 min, stirred for 10 min, and then the other medium components were added.

Semisolid medium was prepared as previously described⁶² but using SM9. To prepare semisolid SM9, Neidhardt Supplement Mixture was combined with SeaPrep agarose (3.5 g l⁻¹, Lonza) and autoclaved for 22 min. Remaining medium components were added after autoclaving, as done for SM9 broth. Semisolid medium was then cooled to 37 °C. Cells were inoculated into semisolid medium at 37 °C and then briefly stirred. The resulting culture was placed in an ice bath for 30 min to let the medium gel. The ice bath must reach higher than the liquid level of the medium to evenly chill the entire volume. The semisolid culture was carefully transferred to 37 °C.

For strain construction and plasmid preparation, *E. coli* were grown in LB Miller Broth (DF0446-07-5, Fisher Scientific). LB Miller plates were used for growth on solid medium unless otherwise noted. For plasmid maintenance, plates or broth were supplemented with 25 μg ml⁻¹ chloramphenicol, 50 μg ml⁻¹ kanamycin, or 50 μg ml⁻¹ carbenicillin. Cells were routinely pelleted by centrifugation at 5,000g for 5 min.

Plasmid construction

p_{rmf}-RFP and *p_{mdtK}-RFP* were assembled by NEB HiFi assembly (E2621L) using pBbA6C-RFP⁶³ as backbone (amplified with SB226 and SB227) and *p_{rmf}-GFP*⁶⁴ or *p_{mdtK}-GFP*⁶⁴ as insert (amplified with SB224 and SB228).

All plasmids are listed in Supplementary Table 6. pBbS6C-*dCas9*, pBbS6C-*metG*, pBbS6C-*metG**, and pBbS6C-*yqgE* were cloned by NEB HiFi assembly (E2621L) using a common vector backbone (pBbS6C-RFP⁶³ amplified with SB167 and SB168) and the following gene inserts: pWJ445 amplified with SB170, SB171 (*dCas9*); MG1655 genomic DNA (gDNA) amplified with SB150, SB169 (*metG*); MG1655-*metG** gDNA amplified with SB150, SB169 (*metG**); MG1655 gDNA amplified with SB202, SB203 (*yqgE*).

To assemble pYqgE+ and pRFP+, intermediate plasmid pBbE6A-RFP was assembled by ligation of Zral- and Xhol-digested pBbS6C-RFP⁶³ and pBbE2A-RFP⁶³. pBbE6A-RFP was amplified by PCR with SB168 and SB180 to generate the vector fragment. For pYqgE+, *yqgE* was amplified from MG1655 genomic DNA with SB178 and SB179. For pRFP+, RFP was amplified from pBbE2A-RFP⁶³ with SB197 and SB198. Vector and inserts were assembled using the HiFi assembly kit (E2621L, New England Biolabs).

pBbS6A-*yqgE* (also called ‘i-pYqgE+’ for ‘inducible pYqgE+’) was assembled by NEB HiFi assembly (E2621L) using pBbS6A-*yqgE*⁶³ as backbone (amplified with SB212 and SB213) and pYqgE+ for the insert (amplified with SB180 and SB214).

Strain construction

hipA7 cells were constructed by transferring the *hipA7* mutation from TH1269³¹ to our MG1655 strain. All deletion strains were constructed using λ Red-mediated recombination⁶⁵. The following primers were used to amplify the template from pKD4 for recombination: SB165 and SB166 (*yqgE*), SB183 and SB184 (*lon*), SB185 and SB186 (*priA*), SB191 and SB192 (*sul4*). For all strains except Δ*priA*, the kanamycin resistance (kan') cassette was removed with pCP20, which was subsequently lost after non-selective overnight growth at 42 °C. Strains were confirmed to have no remaining antibiotic resistance.

Removal of kan' was not successful for MG1655-Δ*priA* or *metG**-Δ*priA*, as cultures did not grow at 42 °C. Instead, experiments in Extended Data Fig. 11a were done with the kanamycin marker still present. After outgrowths (Extended Data Fig. 11a), deletion of *priA* was confirmed again, and *dnaC* was checked for compensatory mutations⁶⁶. Δ*priA* strains were grown in M9 for cloning steps and then in SM9 for growth curves (Extended Data Fig. 11a).

Bioreactor growth

Ten litres of SM9 medium was prepared in a carboy. Around 160 ml were transferred into an autoclaved vessel (DASGIP) with 500 ml capacity. A silicone heater (GBH0250-1, BriskHeat) was used to bring the temperature of the medium to 37 °C, after which 100 μl of overnight culture was inoculated into the vessel, and medium flow into the vessel was turned on. An outflow pump maintained a constant level of medium. The culture was stirred with a stirrer bar at 500 rpm and air was flowed in at 0.1 l min⁻¹. Medium flow rate was manually tuned to be above the *E. coli* doubling time. After >12 h, medium flow was turned off. Using a sampling port and syringe, samples were taken for OD₆₀₀ measurement, antibiotic treatment, ScanLag, and/or PETRI-seq.

Antibiotic survival assays

To measure antibiotic tolerance, cells were incubated in SM9 containing 200 μg ml⁻¹ ampicillin and/or 5 μg ml⁻¹ ciprofloxacin for 4 h (unless otherwise noted) with shaking at 37 °C. For lag phase antibiotic survival, cells were taken either from the bioreactor or from an overnight culture and diluted 1:50 or 1:100 into SM9 plus antibiotics. When important, the length of the ‘overnight’ culture was noted (as in Fig. 5c or for 6-day stationary in Fig. 2b), but typical overnight cultures were grown for 16–24 h. For stationary phase antibiotic survival (‘undiluted’ in Extended Data Fig. 1a), antibiotics were added directly to the overnight culture. To count colonies after treatment, cells were pelleted, resuspended in PBS, and then plated on LB. CFUs were counted after 48 h and compared to CFUs before antibiotic treatment. Unless otherwise noted, replicates were biological replicates from distinct single colonies picked for overnight cultures.

To assay survival ‘in tet’ or ‘after tet’ (Fig. 3a,b), overnight cultures were diluted into fresh medium containing 54 μg ml⁻¹ tetracycline. Cultures were incubated in tetracycline for 30 min at 37 °C with shaking. Then, antibiotics were added (in tet), or cells were pelleted, washed twice in tetracycline-free fresh medium, then treated with antibiotics (after tet). Cells were kept in antibiotics for 4 h.

To assay antibiotic survival after rifampicin (Extended Data Fig. 8b), overnight cultures were diluted into fresh medium containing 200 μg ml⁻¹ rifampicin. Cultures were incubated for 30 min at 37 °C with shaking, then ampicillin or ciprofloxacin was added and incubated for 4 h (‘WT in rifampicin’). For comparison, an overnight culture was diluted into antibiotic-containing medium (without rifampicin, ‘WT’ on plot). To assay survival in rifampicin alone (Extended Data Fig. 8c), overnight cultures were diluted into fresh medium containing 200 μg ml⁻¹ rifampicin and incubated for 1 h at 37 °C with shaking. CFUs were counted before and after rifampicin.

Appearance time assays

Cells were taken either from the bioreactor or from a standard overnight culture and ~100 CFU were spread on an LB or SM9 agar plate. Unless otherwise noted, replicates were measured from distinct single colonies picked for inoculation. To maximize reproducibility, all plates contained 25 ml of medium. Colony appearance times were not different between LB and SM9 plates. As detailed previously¹⁹, plates were put on a scanner (Epson V500 Photo) in the 37 °C incubator and scanned at 15-min intervals for 24–48 h.

Scanners were controlled by ScanningManager software, and images were analysed using Matlab scripts previously published¹⁹. Appearance times were found using the appearance output of

Article

getAppearanceGrowth. Minimum colony size was set to 20 and maximum set to 100.

PETRI-seq library preparation

Growth conditions for all PETRI-seq samples are detailed in Supplementary Table 3.

PETRI-seq of *E. coli* cells was carried out as detailed previously². A stepwise protocol is available at: <https://tavazoielab.c2b2.columbia.edu/PETRI-seq/>. In brief, cells were pelleted and fixed overnight in 4% formaldehyde. The following day, cells were washed twice in PBS with RNase inhibitor (PBS-RI) and then resuspended in 50% ethanol in PBS-RI. In 50% ethanol, cells could be stored at -20 °C for at least 2 weeks. Cells were washed twice in PBS-RI to remove the ethanol and then permeabilized with lysozyme. Cells were washed twice again and then treated with DNase. After DNase inactivation, cells were washed twice in PBS-RI. As a stopping point, cells could then be resuspended in 50% ethanol in PBS-RI and saved at -20 °C for at least 2 weeks; then they were washed twice again in PBS-RI before resuming. To continue cell preparation, the cell pellet was resuspended in PBS-RI and counted using a haemocytometer. Split-pool barcoding, cell lysis, and second strand synthesis were performed as described, yielding 20 µl purified cDNA². For fragmentation, EZ-Tn5 (TNP92110, Biosearch Technologies) was loaded by annealing SB117 and SB118 (Supplementary Table 6), diluting the oligonucleotides to 5 µM each in 50% glycerol, and then adding 2 µl EZ-Tn5 to 8 µl of the oligonucleotides. EZ-Tn5 was incubated with the oligonucleotides for 30 min at room temperature; loaded EZ-Tn5 was stored at -20 °C. 0.125 µl of loaded EZ-Tn5, 24.875 µl TD buffer (FC-131-1096, Illumina), and 5 µl water were added to 20 µl purified cDNA and incubated at 55 °C for 5 min then brought to 10 °C. 12.5 µl NT (FC-131-1096, Illumina) was immediately added to stop the reaction. Fragmented cDNA was amplified in a 500 µl PCR with Q5 polymerase (M0491L, New England Biolabs); 100 µl 5× buffer, 10 µl 10 mM dNTPs (N0447L, New England Biolabs), 5 µl Q5 polymerase, 85 µl Q5 High GC Enhancer, 0.5 µM N70x (Nextera Index Kit v2 Set A, TG-131-2001, Illumina; or equivalent from Integrated DNA Technologies), 0.5 µM i50x (E7600S, New England Biolabs; or equivalent from Integrated DNA Technologies). Libraries were amplified until the early exponential phase (-16–18 cycles): 72 °C 3 min; 95 °C 30 s; cycle: 95 °C 10 s, 55 °C 30 s, 72 °C 30 s; 72 °C 5 min. PCR reactions were pooled (if the 500 µl reaction had been split into multiple PCR tubes), and 100 µl was taken, purified twice with AMPure XP beads (A63881, Beckman Coulter), and eluted in 30 µl water. The resulting libraries could be sequenced directly (non-depleted) or rRNA-depleted using Cas9.

rRNA depletion of PETRI-seq libraries by Cas9

PETRI-seq libraries were subjected to rRNA depletion by the canonical Cas9::crRNA::tracrRNA tripartite complex⁶⁷. To prepare tracrRNA, a dsDNA template (C2425) was made by PCR of pWJ402³ with Q5 polymerase and primers W2031 and W2032. Alternatively, C2425, which is 96 bases long, could be made by ordering and annealing complementary oligonucleotides. C2425 was used for T7 in vitro transcription with the TranscriptAid T7 High Yield Transcription Kit (K0441, Thermo Scientific) by combining the following in a 20 µl reaction: 4 µl 5× reaction buffer, 2 µl 100 mM ATP, 2 µl 100 mM CTP, 2 µl 100 mM GTP, 2 µl 100 mM UTP, 1 µl T7 RNAP enzyme, 700 ng C2425. The reaction was incubated at 37 °C for 4 h, during which a white precipitate became visible. 1 µl DNase I (AMPD1, Millipore Sigma) was added and incubated at 37 °C for an additional 45 min to digest the DNA template. RNA was purified using the Norgen Biotek Total RNA purification kit (37500, Norgen) to generate J703 (tracrRNA). To prepare crRNAs, 45 µM W2034 (T7 promoter) and 45 µM W2035-W2141 (separate reaction for each) were combined in annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA), heated to 95 °C for 5 min then cooled to room temperature. 1 µl of annealed product

was used for T7 in vitro transcription (K0441, Thermo Scientific) by adding the following: 4 µl 5× reaction buffer, 2 µl 100 mM ATP, 2 µl 100 mM CTP, 2 µl 100 mM GFP, 2 µl 100 mM UTP, 1 µl T7 RNAP enzyme, 6 µl water. The reaction was incubated at 37 °C for 4 h, during which a white precipitate became visible. One microlitre DNase I (AMPD1, Millipore Sigma) was added and incubated at 37 °C for an additional 45 min to digest the DNA template. Each resulting crRNA was purified using the Norgen Biotek Total RNA purification kit (37500, Norgen). To anneal tracrRNA to crRNA, 70 pmol tracrRNA (J703) and 70 pmol crRNA were combined in 10 µl of annealing buffer (10 mM Tris pH 7.5, 50 mM NaCl, 1 mM EDTA), heated to 95 °C for 5 min, then slowly cooled to room temperature to yield 7 pmol µl⁻¹ tracrRNA::crRNA. All 59 annealed tracrRNA::crRNA were pooled in an equimolar ratio. rRNA was depleted by combining the following in a 50 µl reaction: 5 µl 10× reaction buffer (Z03386, GenScript), 0.74 µl tracrRNA::crRNA (5.18 pmol total tracrRNA::crRNA; 0.088 pmol of each), 10 µl Cas9 (Z03386, GenScript), 49–80 ng PETRI-seq library. The reaction was incubated at 37 °C for 90 min then purified twice with 1× AMPure beads. The library concentration was measured using the Agilent Bioanalyzer High Sensitivity DNA Kit (5067-4626, Agilent). Libraries were sequenced for 75 cycles (58 R1, 17 R2) using the NextSeq 500/550 High Output Kit v2.5 (20024906, Illumina). rRNA-depleted libraries were loaded at -2× the recommended concentration to account for cleaved rRNA fragments without both Illumina adapters. Non-depleted libraries were loaded at -1.5× the recommended concentration.

Our rRNA depletion strategy is in theory very similar to DASH⁶⁸, which amplifies cDNA after Cas9 cleavage. Further optimization could include testing the differences between these techniques.

Fluorescence-activated cell sorting

*metG** cells were transformed with fluorescent transcriptional reporters for *rmf*, *cysK* and *mdtK* promoters⁶⁴ (Supplementary Table 6). Overnight cultures were diluted 1:100 (*rmf*, *cysK*, dual markers) or 1:50 (*mdtK*) into SM9 then grown for 3.5 (*rmf*), 3.17 (*cysK*), 2.5 (*mdtK*), 5.8 (dual *cysK/rmf*), or 5.25 (dual *cysK/mdtK*) hours at which point they reached OD₆₀₀ of 0.401 (*rmf*), 0.336 (*cysK*), 0.238 (*mdtK*), 0.35 (dual *cysK/rmf*), or 0.59 (dual *cysK/mdtK*). Cells were centrifuged at 5,000g for 5 min and then resuspended in PBS. Using an S3e Cell Sorter (12007058, Bio-Rad), cells were analysed, gated by forward scatter versus side scatter (Bio-Rad ProSort; Extended Data Fig. 3h,i), then sorted by GFP expression (high or low) into PBS. Sorted cells were counted (CFU), inoculated into antibiotic-containing SM9, and/or used for ScanLag. For the protein expression assay shown in Extended Data Fig. 3f,g, *metG*-p**cysK*-GFP cells were transformed with pBbA6C-RFP⁶³ (35290, Addgene), which expresses RFP under the LlacO1 promoter. Overnight cultures were diluted 1:50 into SM9 + 500 µM IPTG then grown for 4.6 h (OD₆₀₀ = 0.182). We noted that because of the stochasticity in *metG** lag times, time to reach a particular OD₆₀₀ after dilution from an overnight varied substantially by experiment. Cells were resuspended in PBS, analysed, gated by forward scatter versus side scatter (Extended Data Fig. 3h,i), then sorted by GFP and RFP expression into SM9 (Extended Data Fig. 3f). GFP- or RFP-only controls were used to compensate for overlapping emissions of GFP and RFP. Because the cells come out of the sorter in PBS, the final composition of medium was 71% SM9 in PBS. Cell density was too low to successfully pellet the cells and change medium. Cells were grown at 37 °C with shaking (300 rpm) and analysed at given timepoints over the next day. OD₆₀₀ stayed below 0.01 for the duration of the experiment, likely reflecting high purity of cells with long lag times and possibly reduced growth rate from 29% PBS. To see RFP expression in persister cells (Extended Data Fig. 3g), populations were gated on high GFP (*cysK*⁺). To make Extended Data Fig. 3, FlowJo 10.8.1 was used. Distributions were plotted using the layout editor. *metG** cells without a fluorescent protein expression vector were used to subtract background autofluorescence (for Extended Data Fig. 3g).

Generating crRNA library with CALM

E. coli crRNA libraries were generated using CALM, as previously described³ with one minor modification. C2185 (insert library) and C2184 (backbone) were assembled by Gibson reaction (E2621L, New England Biolabs) and transformed into MG1655 cells without pWJ445 (dCas9 plasmid). The library was grown in LB broth containing 50 µg ml⁻¹ kanamycin at 37 °C until OD₆₀₀ reached ~0.4 (about 4 h). The resulting library was pelleted and used to miniprep an assembled crRNA plasmid library, labelled M60. Sequencing of this library confirmed high coverage of the genome with each gene targeted on average by 56 unique crRNAs.

CRISPRi screen sample collection

Supplementary Table 5 includes details about each CRISPRi library. Generally, electrocompetent cells were prepared from the parental strain containing either pWJ445 (^pTet-dCas9) or pBbS6C-dCas9 (^pLlacO1-dCas9). Different inducers (IPTG or aTc) were used for replicates to avoid inducer-specific effects. ~200 ng of M60 (crRNA plasmid library) was electroporated with 50 µl of cells using the MicroPulser (Bio-Rad) set to the default *E. coli* program 1 (1 mm, 1.8 kV, 6.1 ms). Cells were recovered in 500 µl SOC medium for 1.5 h at 37 °C. A small volume was taken to count colonies on LB agar with or without selection antibiotics (kanamycin + chloramphenicol) in order to calculate transformation efficiency and ensure adequate library coverage. The crRNA library contains at most 500,000 crRNAs³, and transformations routinely yielded >100 million transformants. The remaining volume of recovered cells was transferred to a flask containing 225 ml SM9 with kanamycin + chloramphenicol. For YqgE/RFP overexpression screens (SBC308-SBC315; Supplementary Table 5), 500 µM IPTG was added 2 h later to induce *yqgE* or RFP. For all libraries, cells were grown at 37 °C until they reached an OD₆₀₀ of ~0.4 (3–4 h). The resulting library (L) was divided for either late exponential/stationary dCas9 induction before lag phase assays or exponential dCas9 induction before exponential assays. Samples were also taken from L for SBC95 and SBC125 (Supplementary Table 5). Late exponential/stationary induction allowed for minimal loss of essential crRNAs (area under the curve = 0.48–0.52 for wild-type/*metG** before and after induction), so all genes could be assayed in lag phase.

Lag phase assays. For lag phase assays, 20 ml of cell library (L) was pelleted and resuspended in 20 ml SM9 containing kanamycin, chloramphenicol, and either anhydrotetracycline (aTc; 20 nM) or isopropyl-β-D-thiogalactopyranoside (IPTG; 500 µM). Centrifugation likely was not necessary here but was done every time for consistency. The culture was grown overnight (14–22 h) to induce dCas9 and reach stationary phase. The next day, cells were pelleted and resuspended in the same volume of SM9 without inducer or antibiotics. Samples were taken for SBC96, SBC126, SBC191, SBC205, SBC316, SBC308, SBC310, SBC312 and SBC314 (Supplementary Table 5).

For lag outgrowth, resuspended cells were either inoculated into 500 ml semisolid SM9 (~80 million cells per litre for SBC101, SBC131, SBC210, SBC318; ~2 × 10⁹ cells per litre for SBC100, SBC130) or diluted 100× into SM9 broth (SBC102, SBC132, SBC192, SBC309, SBC311, SBC313, SBC315). Outgrowth times are shown in Supplementary Table 5.

For lag antibiotic treatment, resuspended cells were diluted 100× into SM9 containing 200 µg ml⁻¹ ampicillin and 5 µg ml⁻¹ ciprofloxacin and incubated for the indicated amount of time (Supplementary Table 5). After antibiotic treatment, cells were pelleted, washed in PBS, then resuspended in SM9 and inoculated into 500 ml semisolid SM9 medium at a density ~50 × 10⁶ cells per litre. After 2 days, cell samples were collected for SBC98, SBC99, SBC128, SBC129, SBC207, SBC208, SBC209 and SBC317. Semisolid medium was used to minimize interclone competition.

Exponential assays. For exponential assays, dCas9 was induced in exponential phase, and cells were not grown overnight to stationary phase. The cell library (L) was diluted 200x into SM9 containing kanamycin, chloramphenicol, and 20 nM aTc or 500 µM IPTG. Cells were grown for 3.5–4.5 h (OD₆₀₀ = ~0.2–0.4). Samples were taken for SBC133 and SBC211 (Supplementary Table 5) and also diluted 100× into SM9 containing kanamycin, chloramphenicol, and 20 nM aTc or 500 µM IPTG. These cells were grown for 3–4 h then sampled for SBC134 and SBC212.

CRISPRi library preparation

Collected cell samples (described above) were pelleted and mini-prepped (Qiagen). 400 ng of DNA were amplified in a 60 µl PCR with Q5 polymerase (M0491L, New England Biolabs), 0.5 µM of forward primer (equimolar mixture of W1397, W1398, W1399, W1400), and 0.5 µM of reverse primer (W1699). The reaction was thermocycled as follows: 98 °C 30 s; 10 × 98 °C 10 s, 55 °C 20 s, 72 °C 30 s; 72 °C 2 min. PCR products were purified by double-sided AMPure cleanup (left-side ratio = 0.8×; right-side ratio = 1.4×) then eluted in 40 µl H₂O. 2.5 µl of purified DNA was used for a second PCR in 100 µl using Q5 polymerase, 0.5 µM forward primer (CRISPRi_PCR_2_F; Supplementary Table 6), and 0.5 µM reverse primer (CRISPRi_PCR_2_R; Supplementary Table 6). The reaction was thermocycled as follows: 98 °C 30 s; 6 × 98 °C 10 s, 55 °C 20 s, 72 °C 30 s; 72 °C 2 min. PCR products were purified by two AMPure cleanups (first 0.9×, then 0.8×) and eluted in 30 µl. The library concentration was measured using the Agilent Bioanalyzer High Sensitivity DNA Kit (5067-4626, Agilent). Libraries were sequenced for 75 cycles using the NextSeq 500/550 High Output Kit v2.5 (20024906, Illumina). Single-end reads are ideal because only Read 1 is useful for mapping crRNAs. However, depending on the forward primer used for PCR 2, as few as 58 cycles can be allocated to read 1.

Antibiotic susceptibility with bortezomib

Bortezomib (5043140001, Millipore Sigma) stock was prepared by dissolving in DMSO. For the assays in Extended Data Fig. 11d,e, single colonies were picked into SM9 containing 1% DMSO and 100 µM bortezomib. Control (−bzmb) cultures were started in the same way but SM9 contained 1% DMSO and no bortezomib. After overnight culture, antibiotic survival was assayed as described in 'Antibiotic survival assays', but for lag phase assays, antibiotic-containing SM9 was supplemented with 1% DMSO ± 100 µM bortezomib.

Lag times after bortezomib treatment

For full growth with bortezomib (dotted black line in Fig. 5b), single colonies of *metG**-Δ *sulA* cells were picked into 1 ml SM9 containing 1% DMSO and 100 µM bortezomib (5043140001, Millipore Sigma). For bortezomib addition during stationary phase (dotted pink line in Fig. 5b), single colonies of *metG**-Δ *sulA* cells were picked into 1 ml SM9. After 20 h, bortezomib was added to 100 µM. For bortezomib treatment during lag phase (dotted green line in Fig. 5b) or not at all (control; grey line in Fig. 5b), single colonies of *metG**-Δ *sulA* cells were picked into 1 ml SM9 containing 1% DMSO. After 24 h of growth, all cultures were diluted 100× into SM9 containing either 100 µM bortezomib and DMSO (green line in Fig. 5b) or only 1% DMSO (all other samples). Cells were grown on a plate reader (37 °C with continuous shaking; PowerWave XS2, BioTek) and OD₆₀₀ measured at 10-min intervals.

Stationary phase translation assay

E. coli cells of the indicated genotype (Fig. 5f) containing pBbS6C-RFP were grown for 24 h in 1 ml SM9. Two-hundred microlitres of overnight culture were transferred to a 96-well plate and supplemented with 2.5 mM IPTG. Cells were grown on a plate reader (37 °C with continuous shaking; Synergy Neo2, BioTek). RFP (570 excitation, 620 emission) and OD₆₀₀ were measured at 10-min intervals.

***metG* complementation of *metG** mutation**

Wild-type or *metG** cells carrying either pBbS6C-*metG* or pBbS6C-*metG** (Supplementary Table 6) were grown overnight in 1 ml SM9 containing chloramphenicol. Overnight cultures were diluted 1,000× into SM9 containing chloramphenicol and 500 µM IPTG. These cultures were grown overnight again. The following day, lag phase antibiotic survival was assayed with ampicillin and ciprofloxacin (Extended Data Fig. 9d).

Quantitative proteomics

Overnight cultures of 3 colonies each of MG1655, *metG**, and *metG*-Δlon-ΔsulA* were grown in 1 ml SM9 for ~19 h at 37 °C. Stationary samples were taken directly from the overnight cultures. For wild-type (MG1655) exponential cells (Extended Data Fig. 7), MG1655 overnight cultures were diluted 200× into fresh SM9 and grown for 90 min (final OD₆₀₀ = 0.2–0.23). For *metG** lag/persister cells (Extended Data Fig. 7), *metG** overnight cultures were diluted 100× into fresh SM9 and grown for 30 min. OD₆₀₀ did not increase in that 30 min (replicate 1: OD_{initial} = 0.069, OD_{30min} = 0.065; replicate 2: OD_{initial} = 0.07, OD_{30min} = 0.065; replicate 3: OD_{initial} = 0.072, OD_{30min} = 0.066).

Cells were collected in Eppendorf tubes and washed twice with ice-cold PBS. Cells were then lysed in lysis buffer containing 8 M urea, 0.1 M ammonium bicarbonate, and protease inhibitors (1 mini-Complete EDTA-free tablet). The lysate was cleared by centrifugation at 14,000 rpm for 30 min at 4 °C. The supernatant was transferred to a new tube, and the protein concentration was determined using a BCA assay (Pierce). Subsequently, 10 µg of total protein was subjected to disulfide bond reduction with 10 mM DTT (at 56 °C for 30 min) followed by alkylation with 10 mM iodoacetamide (at room temperature for 30 min in the dark). Excess iodoacetamide was quenched with 5 mM DTT (at room temperature for 15 min in the dark). Samples were then diluted sixfold with 50 mM ammonium bicarbonate and digested overnight at 37 °C with a trypsin/Lys-C mix (1:100). The next day, digestion was stopped by the addition of 1% TFA (final v/v), followed by centrifugation at 14,000g for 10 min at room temperature to pellet precipitated lipids. Cleared digested peptides were desalted on an SDB-RPS Stage-Tip disk⁶⁹ and dried down in a speed-vac. Peptides were resuspended in 10 µL of 3% acetonitrile/0.1% formic acid and injected onto a Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer using a DIA method for peptide MS/MS analysis.

The UltiMate 3000 UHPLC system coupled with an EASY-Spray Pep-Map RSLC C18 column was used to separate fractionated peptides with a gradient of 5–30% acetonitrile in 0.1% formic acid over 90 min at a flow rate of 300 nL min⁻¹. After each gradient, the column was washed with 90% buffer B for 10 min and re-equilibrated with 98% buffer A (0.1% formic acid, 100% HPLC-grade water) for 30 min. Survey scans of peptide precursors were performed from 350–1,200 m/z at 120 K FWHM resolution with a 1×10^6 ion count target and a maximum injection time of 60 ms. The instrument was set to run in top speed mode with 3-s cycles for the survey and MS/MS scans. After a survey scan, 26 m/z DIA segments were acquired from 200–2,000 m/z at 60 K FWHM resolution with a 1×10^6 ion count target and a maximum injection time of 118 ms. HCD fragmentation was applied with 27% collision energy, and resulting fragments were detected using the rapid scan rate in the Orbitrap. The spectra were recorded in profile mode.

DIA data were analysed with the MaxDIA software platform within the MaxQuant software environment using a library-free approach⁷⁰. The search was set up with the reference *E. coli* proteome database downloaded from UniProt. The false discovery rate (FDR) was set to 1% at the peptide precursor level and 1% at the protein level. Results obtained from MaxQuant were further analysed using the standard pipeline for differential analysis with the DEP package⁷¹. Proteins were filtered for inclusion in 2 out of 3 replicates of at least one condition. Data was normalized by variance stabilizing transformation. Missing data was imputed using the MinProb method with $q = 0.01$. Significantly enriched

proteins were defined by alpha = 0.05 and Ifc = log₂(1.5) (Supplementary Table 2). For the principal components analysis (PCA) in Extended Data Fig. 7c, LFQ intensity (for included samples) was log-transformed and scaled with StandardScaler to centre each protein with mean of 0 and s.d. of 1. Principal components were calculated from all proteins using sklearn⁷². See next section for PCA and UMAP in Extended Data Fig. 7a,b.

PETRI-seq analysis

Barcode demultiplexing was carried out as previously described² with the following minor modification⁷³: before extracting the unique molecular identifier (UMI) sequence, PEAR⁷⁴ was used to merge reads 1 and 2 when they overlapped. Only non-overlapping reads were carried forward because read 2 should contain cDNA sequence, and the end of read 1 should contain barcode 1. Note that this may not apply when sequencing more than 75 cycles. Also, read 2 was trimmed if it matched the reverse complement of the end of read 1, an artefact we think occurs due to hairpin formation. The full pipeline uses trimmomatic⁷⁵ (v0.33) to filter reads, Cutadapt⁷⁶ (v1.18) to demultiplex, UMI-tools⁷⁷ (v0.5.5) to extract UMIs, bwa⁷⁸ (v0.7.17) to align, and featureCounts⁷⁹ (v1.6.3) to annotate features.

Seurat (version 4.1.1)⁸⁰ was used for normalization, dimensionality reduction, and clustering of PETRI-seq data. In brief, the matrices produced by demultiplexing and UMI collapsing were read into a Seurat object. All MG1655 samples in this study (Supplementary Table 3) were combined in the same Seurat object. For Extended Data Fig. 6g–j, a new Seurat object was made with all MG1655 cells plus CFT073 cells; accessory genes only in the CFT073 genome were omitted. For all analysis, rRNA counts were excluded except for Extended Data Fig. 8. Barcodes were filtered for more than 9 and fewer than 1,000 mRNA UMIs. All cells were then downsampled to 38 UMIs using the SampleUMI function (max.umi = 38, upsample = FALSE). UMI counts were log-normalized using the geometric mean of all cell UMI counts as a scale factor. Gene counts were scaled and centred to a mean of 0 and s.d. of 1 (Seurat ScaleData). Principal components were calculated with all genes. For the full cell atlas (Fig. 1f), principal components 1–10 were used to compute UMAP⁸¹ coordinates (default parameters) and to find neighbouring cells. Clusters were found using default parameters⁸² (Louvain algorithm) at resolution 0.32. For *hip47* cells alone (Extended Data Fig. 5a), principal components 1–5 were used to find neighbouring cells, and clusters were found at resolution 0.1. For extended stationary (6-day) wild-type cells with *metG** and standard wild-type cells (Extended Data Fig. 5f), principal components 1–6 were used to find neighbouring cells, and clusters were found at resolution 0.16. For the full atlas downsampled to ~30 mRNA UMIs (Extended Data Fig. 5i), cells were downsampled as described with max.umi = 30. Then, only cells with exactly 29 or 30 mRNA UMIs were kept in the Seurat object. Cells were processed and clustered as with the full atlas (10 principal components, resolution = 0.34). For CFT073 cells alone (Extended Data Fig. 6d,e), CFT073 accessory genes were included, principal components 1–10 were used to find neighbouring cells, and clusters were found at resolution 0.31. For CFT073 with MG1655 cells (Extended Data Fig. 6g), principal components 1–10 were used to find neighbouring cells, and clusters were found at resolution 0.38.

To project proteomics samples with scRNA-seq (Extended Data Fig. 7a,b), proteomics samples were log-normalized using the geometric mean of the scRNA-seq library. Proteomics samples were then merged into a single Seurat object with downsampled, log-normalized scRNA-seq data. This entire Seurat object was scaled and centred with ScaleData. For the PCA (Extended Data Fig. 7a), loadings were extracted from the scRNA-seq Seurat object and used to project all cells and proteomic samples. For UMAP (Extended Data Fig. 7b) and clustering (Extended Data Fig. 7a,b), principal components 1–6 were used with n.neighbors = 50 and k.param = 50. Clusters were found at resolution 0.32. If principal components 1–10 and default n.neighbors and k.param are used (as with scRNA-seq alone), then the stationary

and lag proteomes form their own cluster; exponential proteomes still cluster with early exponential transcriptomes.

When the full cell atlas is shown or analysed (Fig. 1 and Extended Data Figs. 1n–u and 5i), only cell samples relevant up to that point in the text are shown or included in expression analysis, but all cells (as listed in Supplementary Table 3) were used to compute principal components, UMAP coordinates, and cell clusters. See Supplementary Table 3 for details of which cell samples are included in each figure.

For Extended Data Fig. 8, which defines transcriptional deficiency, different thresholds were used to retain cells with very low mRNA counts. Specifically, in Extended Data Fig. 8f,g, all cells with total RNA above a library-specific threshold (between 16–64 total UMIs) were retained. By contrast, Fig. 2f includes only cells with at least 10 mRNAs, as these are the cells used for UMAP and clustering. rRNA depletion is also important to consider when defining transcriptional deficiency (Extended Data Fig. 8d–f). Extended Data Fig. 8d,e only shows libraries that were not subjected to rRNA depletion. In Extended Data Fig. 8f, all libraries are included with a slightly different threshold used for depleted or non-depleted libraries.

Differential expression analysis from scRNA-seq

To find genes differentially expressed between cell clusters or pre-defined populations, a custom pipeline combining edgeR⁸³ and Seurat's FindMarkers tool was used. EdgeR was used with TMM normalization to calculate log₂(fold change) from pseudobulk samples. Pseudobulk samples are calculated by summing all counts from all single cells of a given population; single-cell transcriptomes are taken before downsampling. For *P* values, limma's⁸⁴ rankSumTestWithCorrelation (the default for Seurat's FindMarkers; two-sided Wilcoxon–Mann–Whitney) was used with downsampled, log-transformed single-cell data as input. Using downsampled cells for significance testing gives the result most consistent with the centred edgeR data. Total UMI counts by sample (before and after downsampling) are provided in Supplementary Table 4.

CRISPRi analysis

CRISPRi sequencing reads were aligned to reference genomes for *E. coli* then to *S. aureus* (used for library manufacturing³). Functional spacers were identified as described³ based on presence of an “NGG” PAM sequence. Only functional *E. coli* spacers were used for downstream analysis.

For lag and exponential comparisons, spacer abundance post-outgrowth was compared to pre-outgrowth. For lag antibiotic treatment, spacer abundance post-antibiotics plus outgrowth was compared to after outgrowth only. For simplicity, consider post-antibiotics as a “post” condition relative to outgrowth only (“pre”) in the description below.

To compare CRISPRi libraries, spacers were filtered to remove any position with fewer than 10 reads in both post and pre libraries. Then, the frequency of each spacer was calculated by dividing the number of reads for that spacer by the total number of reads in the library. A pseudocount of 0.99 was added to spacers with 0 counts. Based on the assumption that spacers targeting intergenic regions outside of promoters would not affect phenotypes, we used these intergenic spacers to normalize spacer abundance in both pre and post libraries. All spacer frequencies were normalized as follows (for exemplified spacer labelled A):

$$\text{enrichment}_A = \log_2(\text{spacer}_A/\text{GM}_{\text{null}})_{\text{post}} - \log_2(\text{spacer}_A/\text{GM}_{\text{null}})_{\text{pre}}$$

where GM_{null} is the geometric mean of the frequencies of all null (intergenic) spacers.

To calculate gene enrichment scores, mean enrichment scores for spacers aligned within or directly upstream of each gene were calculated. The number of spacers mapping to each gene varied, which was

important for computing gene enrichment *P* values. The null distribution of enrichment scores for intergenic spacers was randomly sampled to generate pseudogenes with n spacers. This was repeated to generate 100,000 simulated replicates for every relevant *n*. To assign a *P* value, each gene enrichment score was compared to a simulated null distribution with the same number of spacers as included for that gene. Significantly enriched or depleted genes were found based on a FDR of 0.1 using the Benjamini–Hochberg method⁸⁵. In most cases, significant genes were further filtered by significance in multiple replicates. For *hipA7*, only one replicate of each screen was done. In Fig. 4b, we wanted to highlight top hits, so we used Bonferroni correction to threshold only the most significant hits. In other figures, we used FDR of 0.1 for the *hipA7* replicates.

For each gene in the CRISPRi screens, enrichment and significance were calculated independently for crRNAs targeting the antisense or sense strand; the strand with strongest effect (by significance then enrichment score) is determined and included in each relevant figure. For Fig. 4b,c, strand is noted in source data. In Fig. 4d and Extended Data Fig. 10, the strand shown is antisense unless otherwise noted. To assess depletion of essential genes (Extended Data Fig. 9f,g), we used a stringent set of genes found to be essential in all of four previous datasets⁸⁶.

Pathway enrichment with iPAGE

To find pathways significantly correlated with principal component 1 or 2 of the cell atlas (Extended Data Fig. 1o), we divided the principal component loadings into high (greater than 0.025) and low (less than −0.025) groups. We ran iPAGE⁸⁷ in discrete mode (up, down) with maximum *P* value of 0.001 and independence = 0. Redundant pathways were filtered manually, and representative ones are shown.

To find genes enriched in either the early lag or the persister cluster (Extended Data Figs. 1p–u and 5k,l), differential expression analysis was performed as described. Using the Benjamini–Hochberg method⁸⁵, an FDR of 0.01 was applied to select significantly over- and under-represented genes. These significance scores were used as input for iPAGE⁸⁷, which was run in discrete mode (up, down, neutral) with maximum *P* value of 0.05 and independence = 0. Pathways were then filtered further by *P* values indicated in figure legends. Redundant pathways and those indicating enrichment of a single operon were filtered manually; representative terms are shown. To compute mean expression, the AverageExpression function in Seurat was used, and mean gene expression values were averaged for all genes in a given set.

To find genes enriched in each persister type versus early exponential cells (Extended Data Figs. 2 and 6k) differential expression analysis was performed as described. Using the Benjamini–Hochberg method⁸⁵, an FDR of 0.01 was applied to select significantly over- and under-represented genes. These significance scores were used as input for iPAGE⁸⁷, which was run in discrete mode (up, down, neutral) with maximum *P* value of 0.1 and independence = 0. Pathways shown in Extended Data Fig. 2a–c(iv) are the top (*P*<0.0005) non-redundant gene sets overexpressed in each persister type; when these pathways are also significant for another persister type, they are also labelled in that panel (**P*<0.05, ***P*<0.005, ****P*<0.0005).

To find genes enriched in persister cell groups versus tetracycline-treated cells (Figs. 3c and 4e and Extended Data Figs. 9a,b,e,i–n and 10a,b), differential expression analysis was performed as described. An FDR of 0.05 was applied⁸⁵ to select over- and under-represented genes. These significance scores were used as input for iPAGE⁸⁷, which was run in discrete mode (up, down, neutral) with maximum *P* value of 0.05 and independence = 0. To select pathways to show in Fig. 3c and Extended Data Fig. 9b, loadings of principal component 1 were also used as input for iPAGE (continuous mode, 8 bins, max_p = 0.005, independence = 0). The intersection of gene sets enriched in the top PC1 bin (*P*<0.001) and gene sets over-represented in at least one persister type (*P*<0.05; based on marker genes versus tetracycline-treated cells) are shown in Fig. 3c and Extended Data Fig. 9b. Redundant pathways were manually filtered.

Article

To find genes enriched after CRISPRi perturbation (Extended Data Fig. 10c–e), an FDR of 0.1 was applied⁸⁵ using *P* values (computed as described above from null distribution) to select over- and under-represented genes. These significance scores were used as input for iPAGE⁸⁷, which was run in discrete mode (up, down, neutral) with maximum *P* value of 0.05 and independence = 0. Gene sets were subsequently filtered by *P* value < 0.01. For each gene, antisense- and/or sense-targeting crRNAs can be significant. For this analysis, only one strand was used for input; if one or both were significant in a single direction, the gene was assigned that direction, but if the antisense and sense crRNAs were significant in opposite directions, then the one with the higher enrichment score (absolute value) was used. Pathways significantly enriched in >3 of 5 *metG** replicates, >1 of 3 wild-type replicates, or in 1 *hipA7* replicate are shown in Extended Data Fig. 10c–e.

To find pathways enriched in proteomics data (Extended Data Fig. 12a–d), differential protein analysis was performed as described with DEP. Fold changes were used as input for iPAGE⁸⁷, which was run in continuous mode with 5 bins and maximum *P* value of 0.05.

Identification of *E. coli* gene homologues

The proteins of *E. coli* K12 (*n* = 4,136) were downloaded from Biocyc⁸⁸ version 25.1. To identify potential homologues, *E. coli* proteins were searched against genomes of diverse microbial organisms. A total of 2,421 genomes downloaded from JGI IMG were included in the search⁸⁹. These genomes were selected based on quality (High Quality = “Yes” in IMG portal) and optimization for biodiversity. They represent 39 phylum, 68 classes and 168 orders (Supplementary Table 7) based on GTDB taxonomic classification⁹⁰. The protein search was done using DIAMOND⁹¹ under specific parameters: “blastp -e 1e-10 -k 10000000 -query-cover 66 --subject-cover 50 -b8 -cl”. Protein hits with maximum *e*-value equal to E-10 were kept as potential homologues for downstream analysis. For each protein, the number of genomes with homologues was counted and converted to frequency, shown in Extended Data Fig. 11q.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Sequencing data have been uploaded to the Gene Expression Omnibus (GEO) under accession number GSE229435. Source data are provided for Figs. 1–5 and Extended Data Figs. 1, 2, 5, 6, 8, 9, 10, 11.

Code availability

Code to generate all figures is available at <https://github.com/tavalab/PETRI-seq-persistence> and <https://doi.org/10.5281/zenodo.13362784> (ref. 73).

61. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119**, 736–747 (1974).
62. Momen-Roknabadi, A., Oikonomou, P., Zegans, M. & Tavazoie, S. An inducible CRISPR interference library for genetic interrogation of *Saccharomyces cerevisiae* biology. *Commun. Biol.* **3**, 723 (2020).
63. Lee, T. S. et al. BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *J. Biol. Eng.* **5**, 12 (2011).
64. Zaslaver, A. et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods* **3**, 623–628 (2006).
65. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
66. Sandler, S. J. et al. *dnaC* mutations suppress defects in DNA replication- and recombination-associated functions in *priB* and *priC* double mutants in *Escherichia coli* K-12. *Mol. Microbiol.* **34**, 91–101 (1999).
67. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

68. Prezza, G. et al. Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. *RNA* **26**, 1069–1078 (2020).
69. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **11**, 319–324 (2014).
70. Sinitcyn, P. et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat. Biotechnol.* **39**, 1563–1573 (2021).
71. Zhang, X. et al. Proteome-wide identification of ubiquitin interactions using UbIa-MS. *Nat. Protoc.* **13**, 530–550 (2018).
72. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
73. tavalab, tavalab/PETRI-seq-persistence: Identification and genetic dissection of convergent persister cell states. Zenodo <https://doi.org/10.5281/zenodo.13362785> (2024).
74. Zhang, J., Kober, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics* **30**, 614–620 (2014).
75. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
76. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBNet* **17**, 10–12 (2011).
77. Smith, T. S., Heger, A. & Sudbery, I. Modelling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
78. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
79. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
80. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e3529 (2021).
81. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform manifold approximation and projection for dimension reduction. Preprint at <https://doi.org/10.48550/arXiv.1802.03426> (2018).
82. Waltman, L. & Van Eck, N. J. A smart local moving algorithm for large-scale modularity-based community detection. *Eur. Phys. J. B* **86**, 471 (2013).
83. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
84. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
85. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
86. Gerdes, S. Y. et al. Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *J. Bacteriol.* **185**, 5673–5684 (2003).
87. Goodarzi, H., Elemento, O. & Tavazoie, S. Revealing global regulatory perturbations across human cancers. *Mol. Cell* **36**, 900–911 (2009).
88. Karp, P. D. et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **20**, 1085–1093 (2017).
89. Chen, I. M. A. et al. The IMG/M data management and analysis system v.7: content updates and new features. *Nucleic Acids Res.* **51**, D723–D732 (2022).
90. Parks, D. H. et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat. Biotechnol.* **36**, 996–1004 (2018).
91. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat. Methods* **18**, 366–368 (2021).
92. Novick, A. & Weiner, M. Enzyme induction as an all-or-none phenomenon. *Proc. Natl. Acad. Sci. USA* **43**, 553–566 (1957).
93. VanBogelen, R. A. & Neidhardt, F. C. Ribosomes as sensors of heat and cold shock in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **87**, 5589–5593 (1990).
94. Cheng-Guang, H. & Gualerzi, C. O. The ribosome as a switchboard for bacterial stress response. *Front. Microbiol.* **11**, 619038 (2021).
95. Michel, B. & Sandler, S. J. Replication restart in bacteria. *J. Bacteriol.* **199**, e00102–e00117 (2017).
96. Ciesielski, S. J., Schilke, B., Marszalek, J. & Craig, E. A. Protection of scaffold protein Isu from degradation by the Lon protease Pim1 as a component of Fe-S cluster biogenesis regulation. *Mol. Biol. Cell* **27**, 1060–1068 (2016).

Acknowledgements The authors thank the members of the Tavazoie laboratory for many helpful discussions throughout this project; R. Soni for assistance with mass spectrometry proteomics; A. Khare for *E. coli* strains and initial advice on persistence experiments; and L. Freddolino for assistance with COFACTOR. This work was supported by National Institutes of Health grants 2R01AI077562 and R01GM139215 (S.T.), National Science Foundation Graduate Research Fellowship DGE 16-44869 (S.B.B.), and National Institutes of Health grant 1K99AI153530 (W.J.).

Author contributions S.B.B. and S.T. conceived the study and designed experiments. S.B.B. performed experiments and data analysis. W.J. assisted with PETRI-seq and CALM technology development. E.R.M. performed survival and lag time assays for UPEC cells. M.L. performed gene homology search. S.B.B. and ST wrote the paper. E.R.M., W.J., M.L. and P.O. reviewed and edited the paper.

Competing interests The authors declare no competing interests.

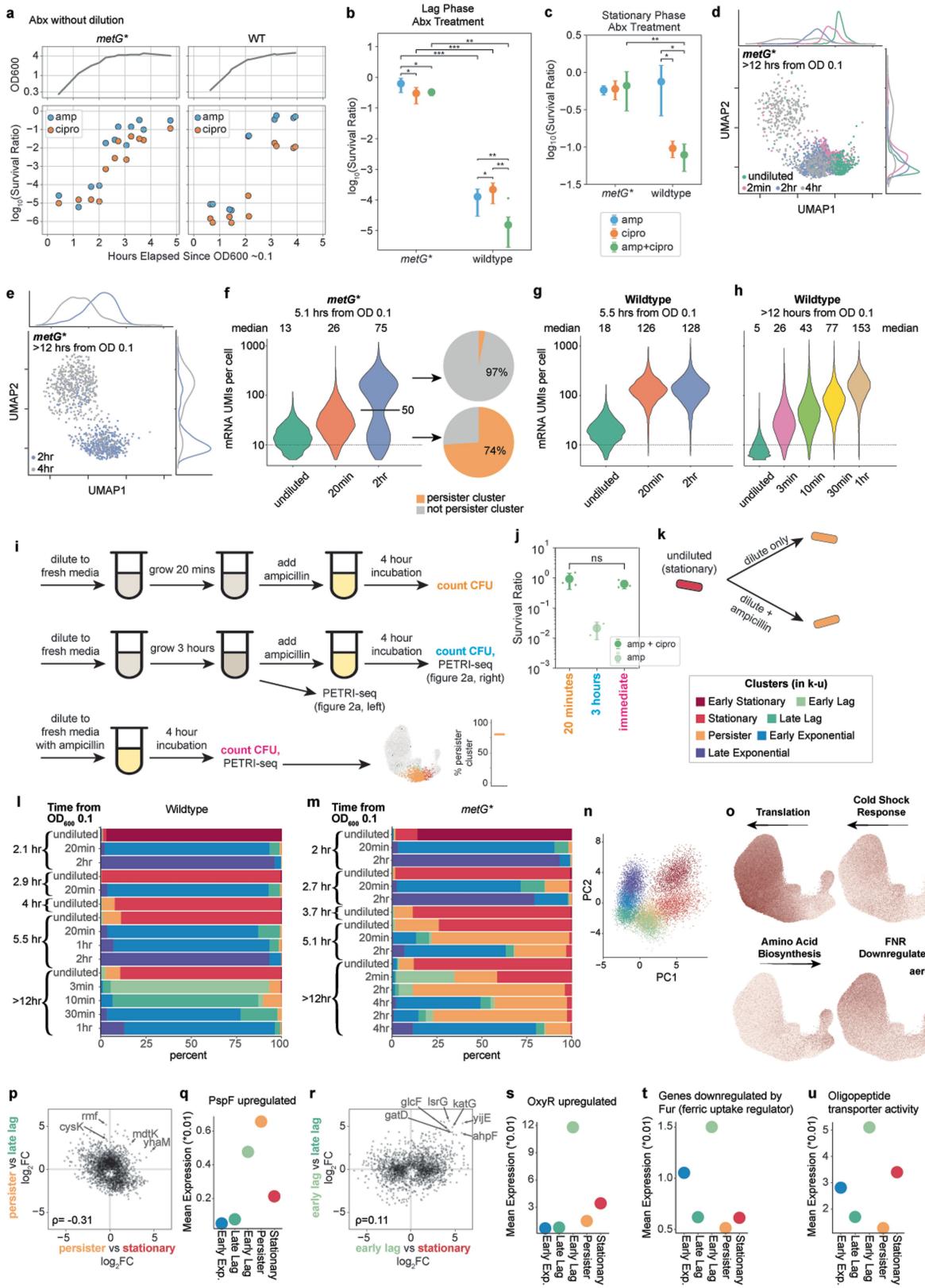
Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41586-024-08124-2>.

Correspondence and requests for materials should be addressed to Saeed Tavazoie.

Peer review information *Nature* thanks the anonymous reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permissions information is available at <http://www.nature.com/reprints>.



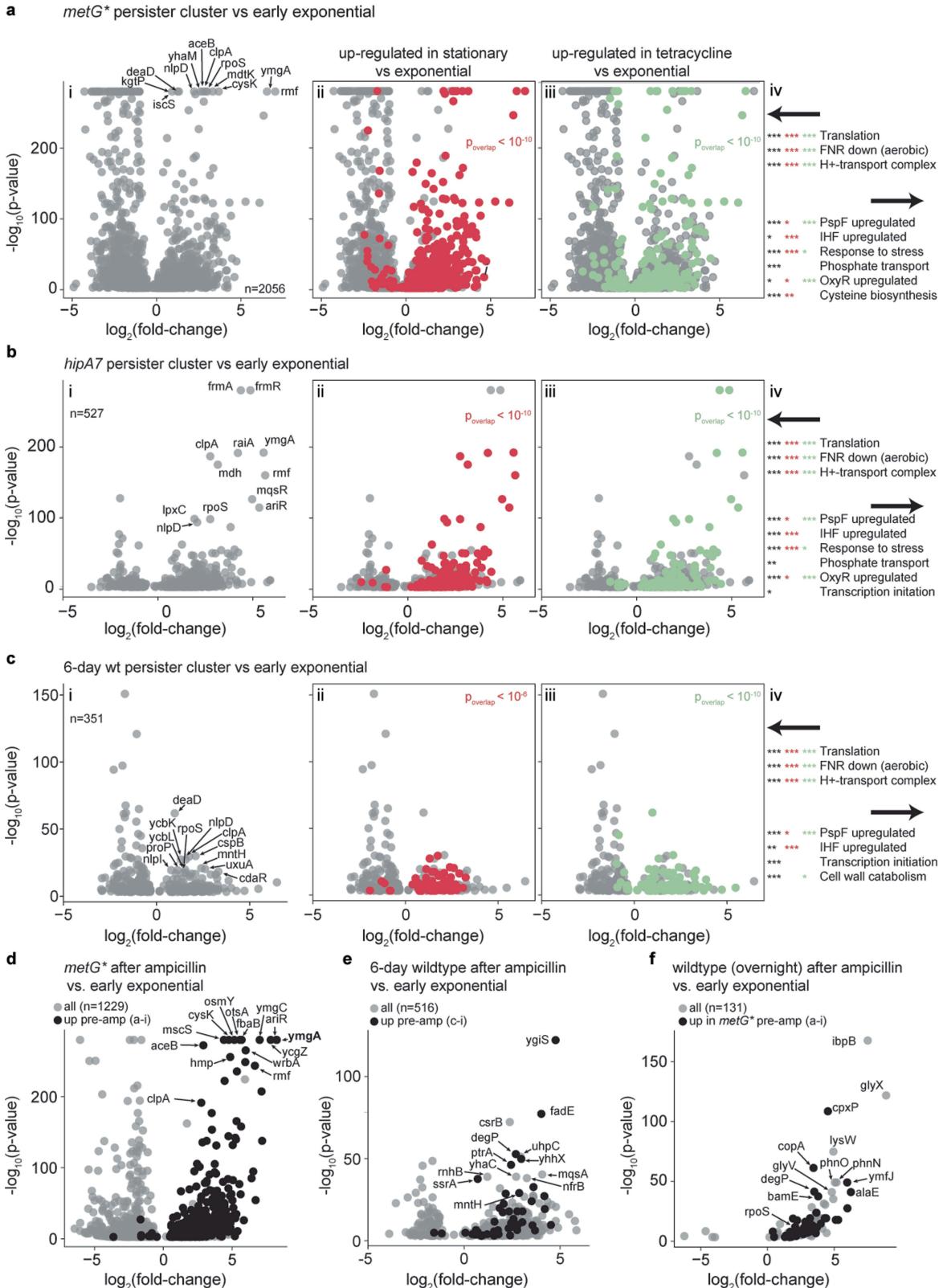
Extended Data Fig. 1 | See next page for caption.

Article

Extended Data Fig. 1 | PETRI-seq discovers a distinct persister state, supplement to Fig. 1.

a, As in Fig. 1b but antibiotic survival was assayed without dilution into fresh media. **b,c**, Antibiotic survival after overnight growth and dilution (**b**) or no dilution (**c**) (one-sided Mann-Whitney U test; * $p < 0.05$, ** $p < 0.05$; *** $p < 0.05$). Mean and standard deviation are shown. **d,e**, UMAP of biological replicates of *metG** transcriptional states. Cells were grown overnight then sampled without dilution or diluted into rich media and sampled after indicated time. Proportions of persister cells are higher than Fig. 1d because cells were grown overnight. 500 representative cells are shown for each population. **f-h**, mRNA UMIs captured per cell in populations shown in Fig. 1 (1d top right in **f**, 1d bottom right in **g**, 1e in **h**). The dotted line at 10 mRNA UMIs indicates minimum threshold to include cells in PCA, UMAP, and clustering. In **f**, pie charts show that cells in the persister cluster are enriched below 50 mRNA UMI per cell threshold. **i-k**, *metG** persisters form with or without ampicillin, seen by comparing PETRI-seq and CFU counts for 3 different cell treatments (**j**). Experimental overview in **i** and conclusion illustrated in **k**. Diluting directly into ampicillin or adding ampicillin after 20 min in fresh media does not change *metG** survival (**j**, $p = 0.2$; one-sided Mann-Whitney U test; ‘20 min’: $n = 3$; ‘immediate’: $n = 3$; ‘3 h’: $n = 2$; mean and

standard deviation shown). **l,m**, Cluster assignments for cell populations shown on UMAP (panels d-e; Fig. 1d,e). **n**, Cell atlas as in Fig. 1f but cells are projected onto principal components 1 and 2. 2000 representative cells are shown for each cluster. **o**, Pathways most strongly correlated with PC1 or PC2. Normalized expression values are shown and each use a different scale (min=0; max=0.14, 0.30, 0.08, 0.06 [clockwise]; $n = 249,219$ cells). **p**, Differential expression analysis between persister vs. neighboring clusters. Only significant genes are shown (two-sided Mann-Whitney U; FDR = 0.01; $n = 2460$). **q**, Mean expression by cluster of genes upregulated by PspF, the only gene module enriched in the persister cluster relative to stationary and late lag cells ($p < 0.002$; iPAGE). The cells used here include all *metG** and wildtype cells in the persister cluster (Supplementary Tables 3-4), but restricting to *metG** cells after dilution into fresh media recapitulates the finding of only this upregulated pathway. **r**, Differential expression analysis between early lag vs. neighboring clusters. Only significant genes are shown (two-sided Mann-Whitney U; FDR = 0.01; $n = 2009$). **s-u**, Mean expression by cluster of gene modules enriched in the early lag cluster relative to all neighboring clusters ($p < 0.002$; iPAGE).

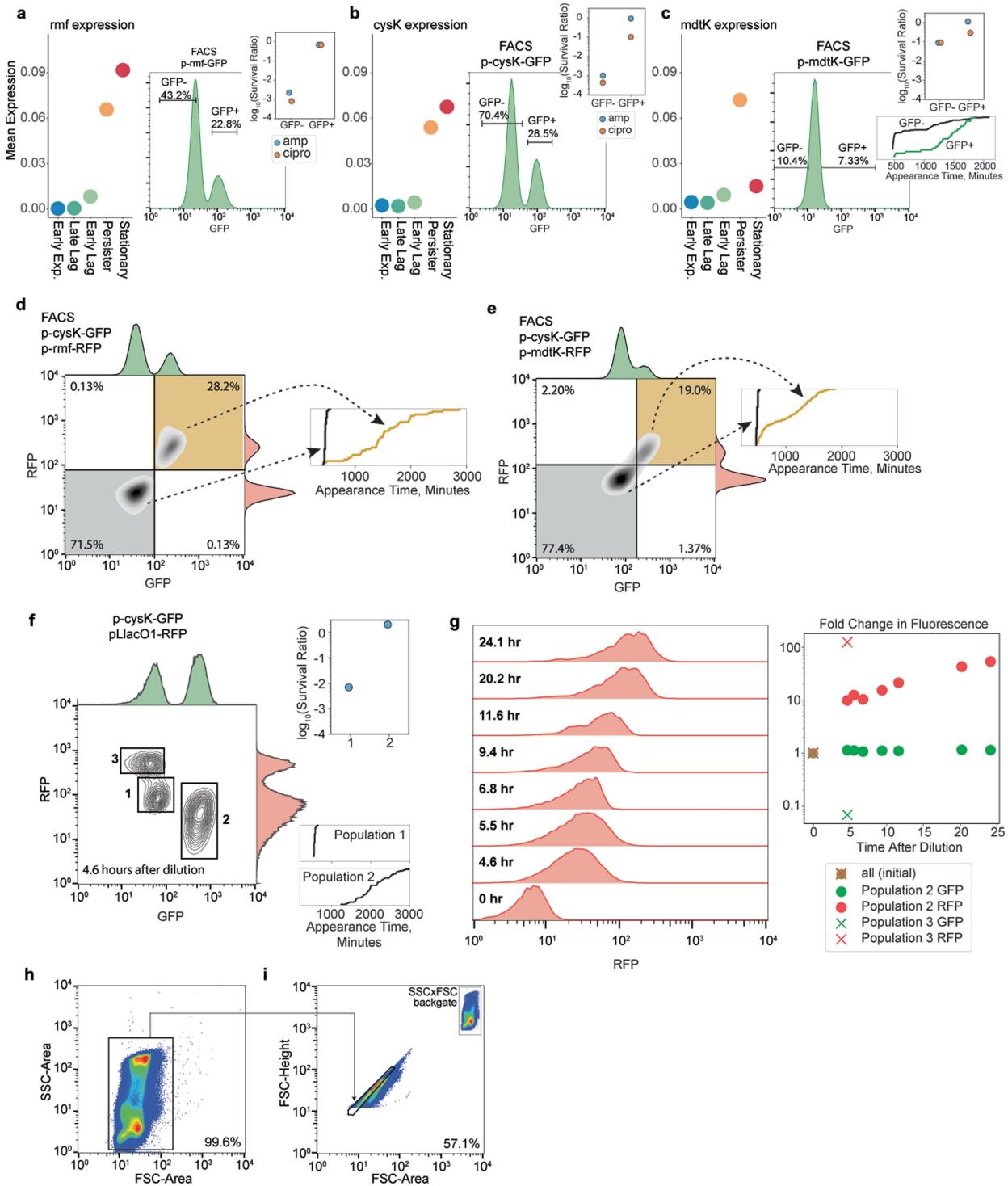


Extended Data Fig. 2 | See next page for caption.

Article

Extended Data Fig. 2 | Differential expression between persister and non-persister (exponential) cells. Volcano plots; in all panels, only significant genes are shown (two-sided Mann-Whitney U; FDR = 0.01), and the 12 most significantly upregulated are labeled. Select common hits (*rmf*, *clpA* in d; *rpoS* in f) are also labeled. Genes expressed by different persister types significantly overlap (hyper-geometric tests; source data/p-overlap). **a–c.** *i*. Genes differentially expressed between *metG** (a), *hipA7* (b), and 6-day wildtype (c) cells in the persister cluster vs. early exponential cells. Persister populations only include cells after dilution into fresh media and not treated with ampicillin. Early exponential cells of all strains were included. Note repeated top hits (all: *rpoS*, *clpA*, *nlpD*; *metG** + *hipA7*: *rmf*, *ymgA*; *metG** + 6-day wildtype: *deAD*). *ii*, *iii*. Colored dots indicate genes significantly overexpressed in stationary cluster or tetracycline-treated cells (all clusters) vs. early

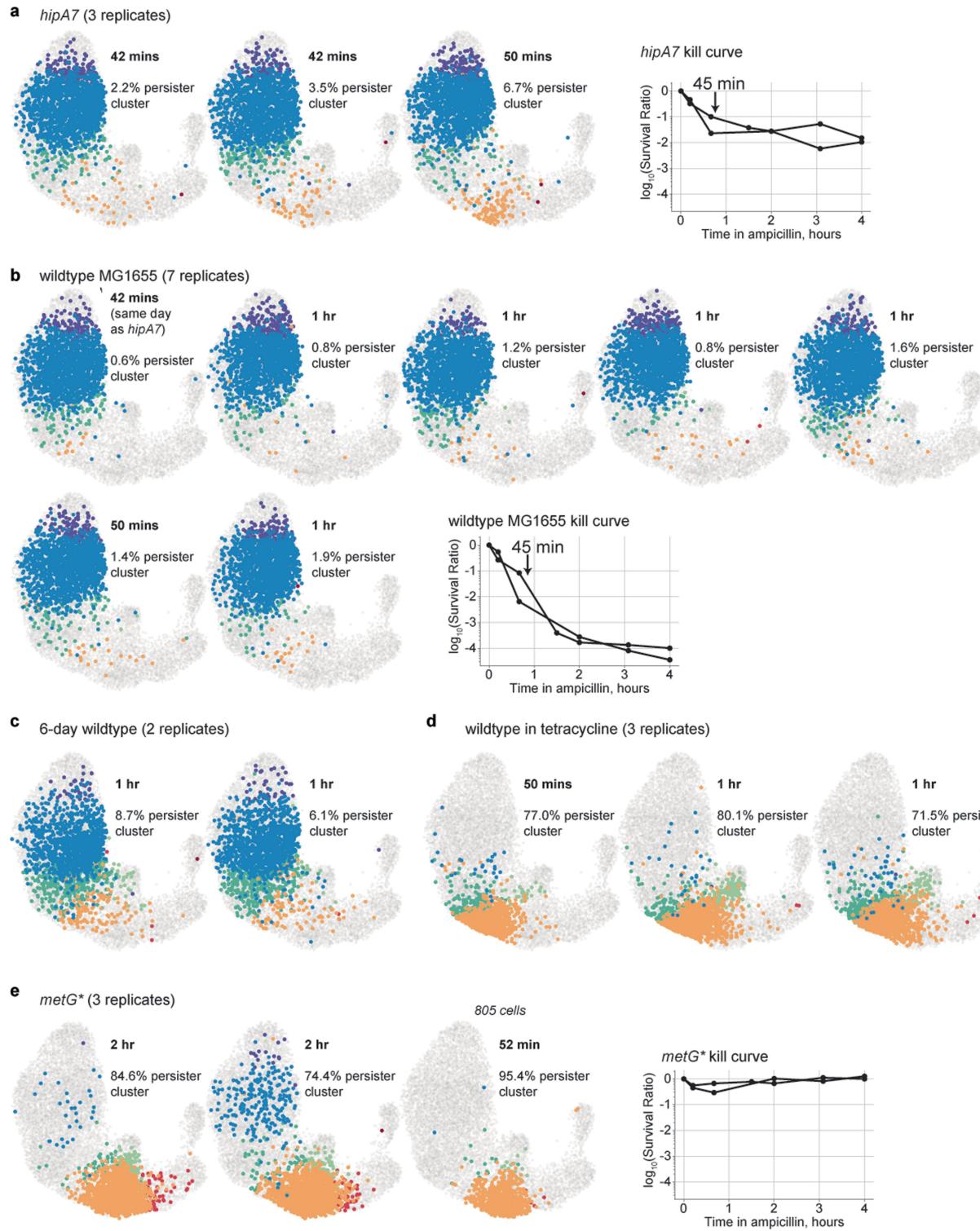
exponential cells. *iv*. Selected pathways significantly enriched in either early exponential ($p < 10^{-4}$; iPAGE) or persister ($p < 0.05$; iPAGE) cells. Asterisks indicate pathways also significant for stationary (red) or tetracycline-treated (green) cells (* $p < 0.05$, ** $p < 0.005$, *** $p < 0.0005$). **d**, Genes differentially expressed between *metG** cells in ampicillin (all clusters) vs. early exponential cluster. All pathways shown in panel a except phosphate transport are significant here ($p < 0.05$). **e**, Genes differentially expressed between 6-day wildtype cells in ampicillin (all clusters) vs. early exponential cluster. All pathways shown in panel c except cell wall catabolism are also significant here ($p < 0.05$). **f**, Genes differentially expressed between wildtype cells in ampicillin (all clusters) vs. early exponential cluster. Translation genes are significantly under-expressed, and IHF upregulated genes are significantly overexpressed ($p < 0.05$).



Extended Data Fig. 3 | Validation of persister markers and translational deficiency of *metG persisters.** **a–c**, FACS of transcriptional reporters validates *metG** persister markers predicted by PETRI-seq. *Left*: gene expression from PETRI-seq. *Middle*: Flow cytometry analysis of *metG** lag populations containing reporter plasmid. Populations were sorted (gates shown), then antibiotic survival was measured (*top right*). For *mdtK*, colony appearance times were also assayed (*bottom right*). *rmf* has previously been validated as a persister marker⁸. *cysK* and *mdtK* have not. **d,e**, FACS of dual reporter *metG** strains validates co-occurrence of persister marker expression. Indicated populations (gold, grey) were sorted and colony appearance assayed (*right*). **f**, Protein expression assay in *metG** persister cells. *metG**+_pcysK-GFP+_pLacO1-RFP (pBbs6C-RFP) overnights were diluted into fresh media + IPTG then analyzed (*left*) and sorted according to the gates shown. Survival to

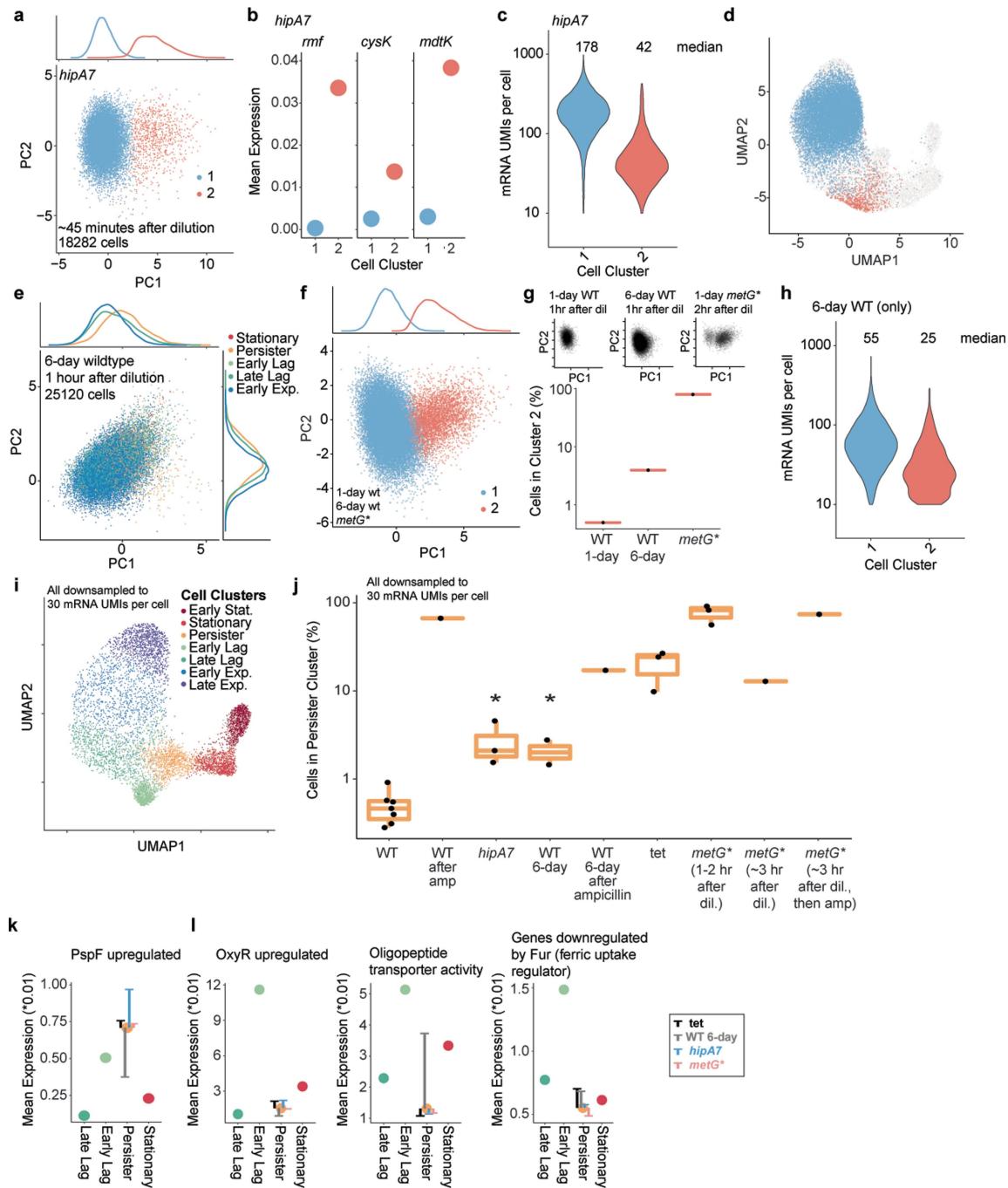
ampicillin (*top right*) and appearance times (*bottom right*) were assayed to confirm *cysK* persister expression. Persisters (high GFP) have reduced protein expression (low RFP). The different RFP expression between populations 1 and 3 is likely due to bimodal induction of the lac operon⁹². **g**, After sorting into fresh media + IPTG, population 2 in panel f was analyzed by flow cytometry at the indicated timepoints. *Left*, distributions of RFP expression across population 2. *Right*, mean fold-change in GFP or RFP expression. After 25 h, RFP expression in population 2 increased 53x from the start of induction, while RFP expression in population 3 increased 123x in just 4.6 h (“x” vs dots). **h–i**, Example of gating strategy, using cells shown in panel a. First, cells are selected by SSC-Area vs. FSC-Area to exclude outliers (h). Then, singlets are selected as the upper distribution on FSC-Height vs. FSC-Area (i). Bottom right shows percent of cells taken with each step.

Article



Extended Data Fig. 4 | Individual replicates for samples in Fig. 2. UMAPs for all individual samples represented in Fig. 2. Each includes 1700 representative cells except third *metG** replicate (e). **a**, UMAPs for biological replicates in Fig. 2d,f. **Right**: For reference, kill curve shows expected proportion of persisters based on survival after ~45 min of ampicillin treatment. **b**, Biological replicates included in Fig. 2f,h. **d**, Replicates included in Fig. 2f,j. The first two plots are technical replicates, and the third is a biological replicate. **e**, Biological replicates included in Fig. 2f. **Right**: For reference, kill curve shows expected proportion of persisters based on survival after ampicillin treatment.

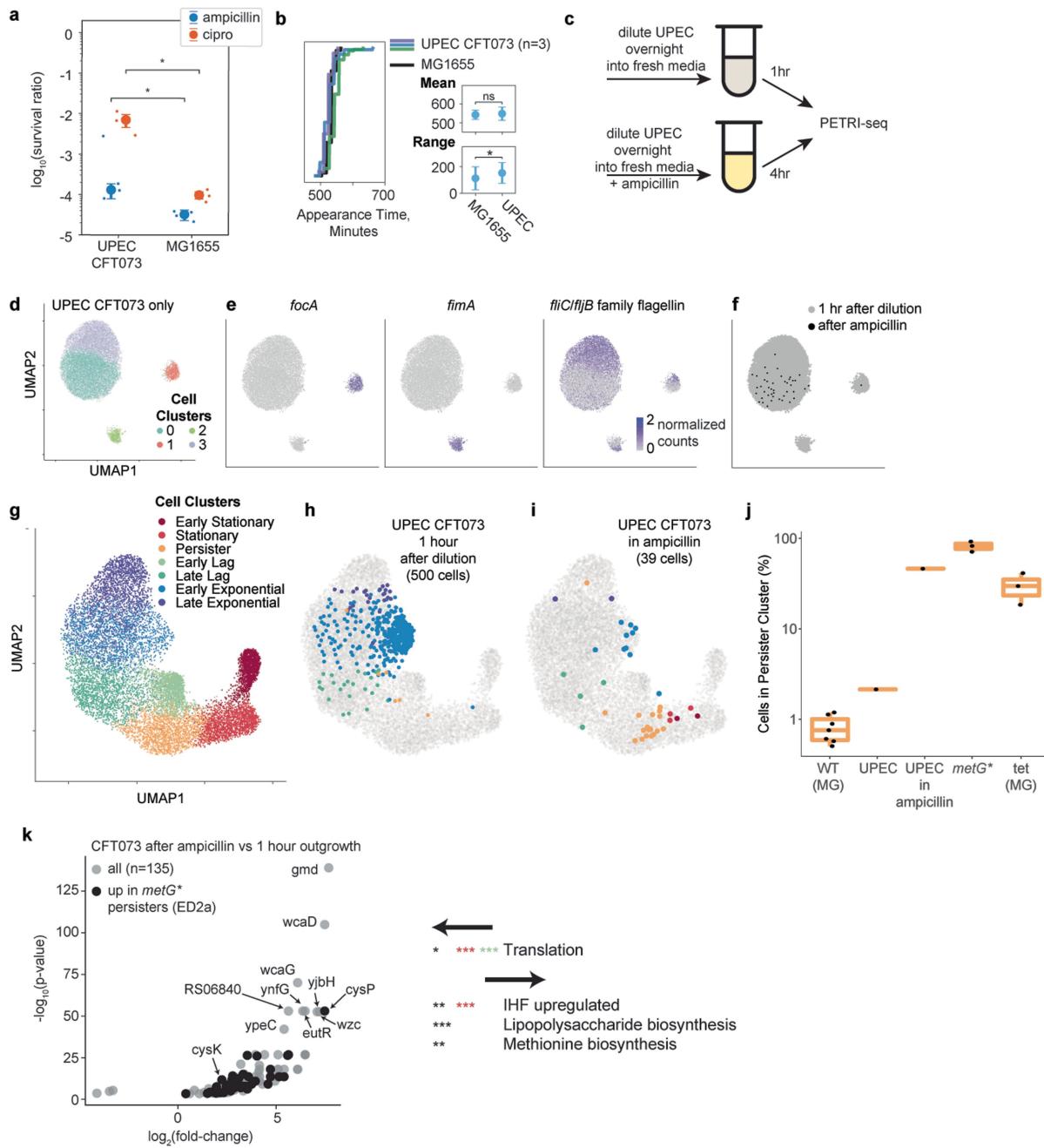
proportion of persisters based on survival after ~45 min of ampicillin treatment. **c**, Biological replicates included in Fig. 2f,h. **d**, Replicates included in Fig. 2f,j. The first two plots are technical replicates, and the third is a biological replicate. **e**, Biological replicates included in Fig. 2f. **Right**: For reference, kill curve shows expected proportion of persisters based on survival after ampicillin treatment.



Extended Data Fig. 5 | Further demonstration of persister cluster cells in *hipA7* and 6-day wildtype cells; supplement to Fig. 2. **a**, PCA and clustering of *hipA7* cells alone. **Top:** density of cells along PC1. **b**, Mean expression of *metG** persister markers in *hipA7* cell clusters. **c**, mRNA UMs per cell for *hipA7* clusters. **d**, Overlay of *hipA7* clusters on full UMAP. **e**, PCA of 6-day wildtype cells alone. Cells are colored by clusters as defined in Fig. 1f (66% early exponential, 7% persister, 22% late lag cluster). **Top and right:** cell densities along PC1 (top) or PC2 (right). **f**, PCA and clustering of 6-day wildtype cells, one *metG** population (2 h after dilution; 4361 cells) and 1-day wildtype cells (1 h after dilution; 7690 cells). Two cell clusters are shown with their density along PC1 (top). **g**, **Top:** PCA as in **f** but split by sample type. **Bottom:** Percent of cells in cluster 2 by sample type. 6-day wildtype cells are higher in cluster 2 than 1-day wildtype cells. **h**, mRNA UMs per cell (split by clusters in **f**) for 6-day wildtype cells. **i**, UMAP of all cell populations in which every cell is downsampled to 30

mRNA UMs per cell. Cells with less than 30 mRNA UMs were omitted from this analysis. For all other analysis, cells were downsampled to 38 mRNA UMs per cell, but cells with at least 10 were kept. **j**, Percent of cells in persister cluster for clusters in **i**. Results confirm that clusters are based on expression patterns, not different mRNA counts per cell (* $p < 0.05$ one-sided Mann-Whitney U test; WT: n = 7; *hipA7*: n = 3; WT 6-day: n = 2; tet: n = 3; *metG** 1 hr: n = 3; all others: n = 1). Center line, boxes, and whiskers show median, 25th/75th percentile, and minimum/maximum value, respectively. **k**, Like *metG** persisters, *hipA7* and tetracycline-treated cells in the persister cluster express PspF-upregulated genes ($p < 0.01$; iPAGE). WT 6-day persisters do not ($p > 0.05$). **l**, Like *metG** persisters, *hipA7* persisters do not express early lag-specific pathways. WT 6-day persisters express only oligopeptide transport genes ($p < 0.05$; iPAGE), though not as highly as early lag. Mean expression shown by cell cluster (colored dots) and persister type (colored bars).

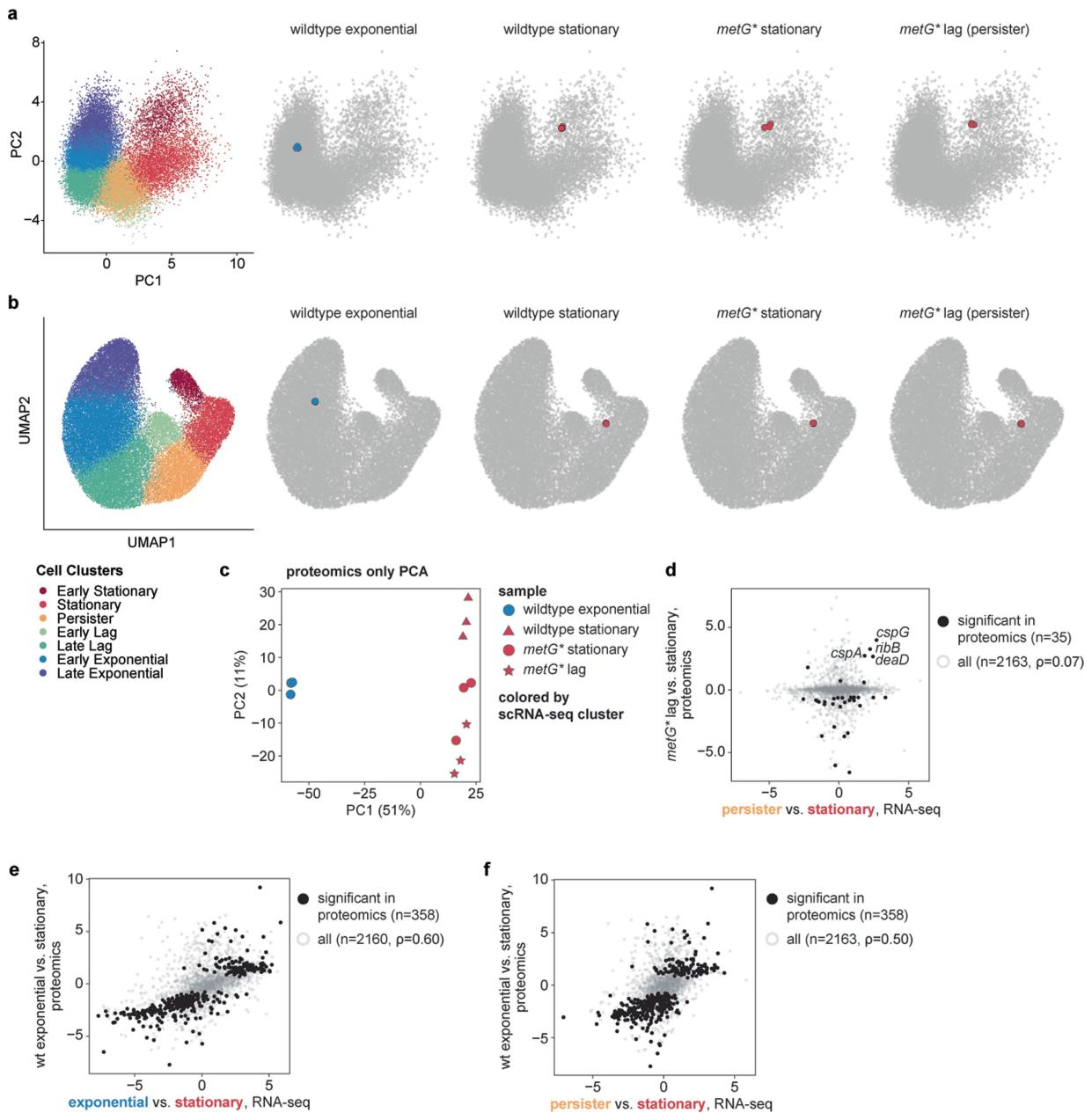
Article



Extended Data Fig. 6 | Identification of persister state for UPEC strain CFT073.

a, UPEC CFT073 exhibits higher antibiotic survival than MG1655 (“wild-type” elsewhere in this paper) (one-sided Mann-Whitney U test; cipro: $n = 3, p = 0.04$; ampicillin: $n = 4, p = 0.02$). Mean and standard deviation are shown. For UPEC in ampicillin, an outlier is shown and included in statistics but not in standard deviation (error bar). **b**, *Left*: cumulative distribution of colony appearance times. *Right*: Range (*bottom*) of appearance times for CFT073 is significantly higher than MG1655 (one-sided Mann-Whitney U test; * $p = 0.03$; CFT073: $n = 13$; MG1655: $n = 18$). Mean (of mean or range) and standard deviation for replicate populations are shown. **c**, Schematic of experiments in d-k. **d**, UMAP of all CFT073 cells sequenced ($n = 17,917$). Unsupervised clustering found 4 clusters. **e**, Marker genes identified by cluster ($p < 10^{-100}$, two-sided Mann-Whitney U test). No marker genes were found for cluster 0. **f**, Cells after ampicillin on UMAP ($n = 43$; 41 in cluster 0, 1 in cluster 1, 2 in cluster 3). **g**, Unsupervised clustering and UMAP of CFT073 cells combined with all MG1655 cells. 7 clusters were annotated as in Fig. 1f. 2000 representative cells

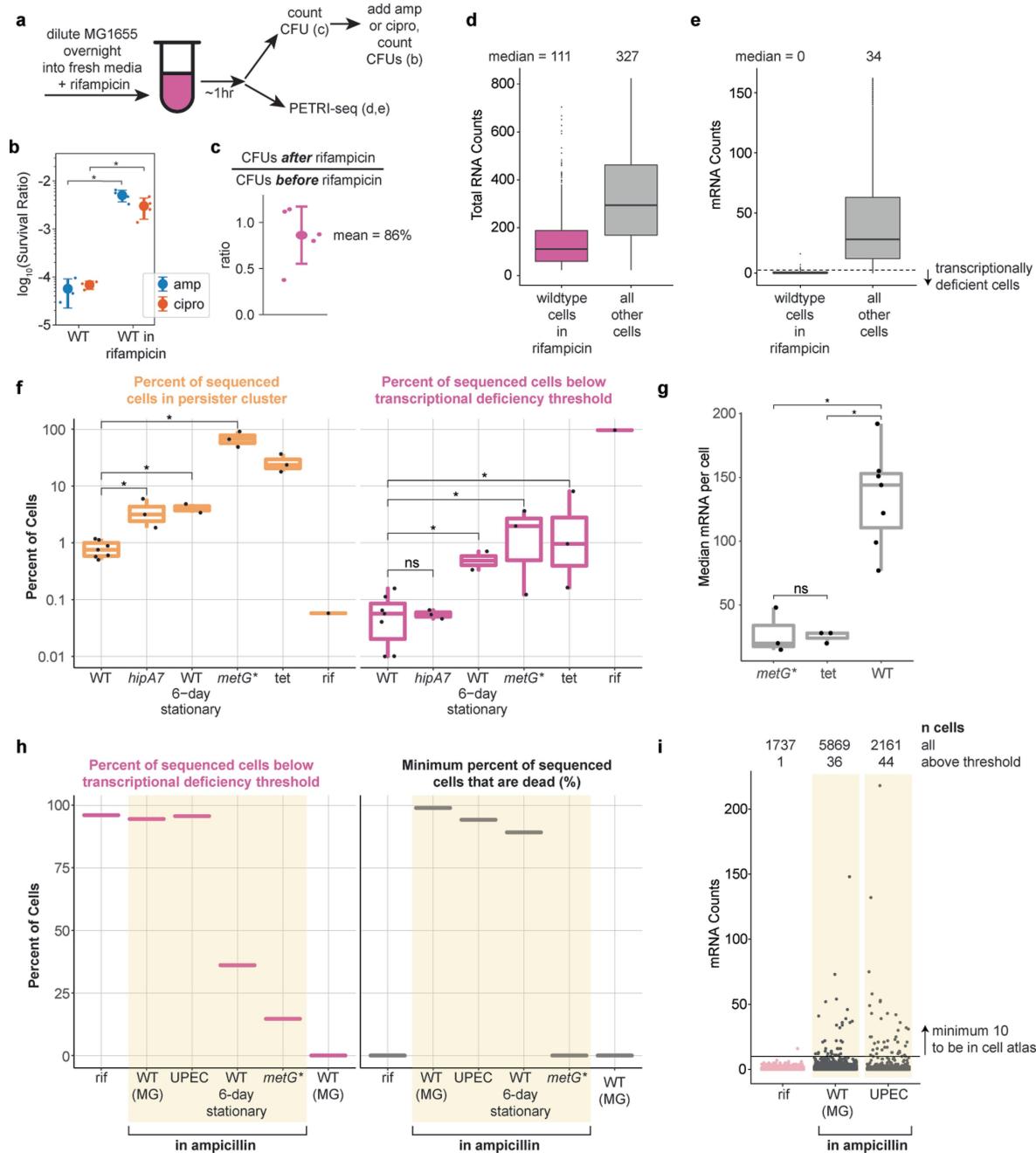
shown per cluster. **h-i**, CFT073 cells overlaid and colored by cell cluster. 500 representative cells (**h**) or 39 cells (**i**) are shown, each from 1 biological replicate. **j**, Quantification of the percent of cells in the persister cluster (WT MG: $n = 7$; CFT073 UPEC (1 hr): $n = 1$; CFT073 UPEC (amp): $n = 1$; *metG**: $n = 3$; tet: $n = 3$). Center line, boxes, and whiskers show median, 25th/75th percentile, and minimum/maximum value, respectively. **k**, *Left*: Volcano plot showing genes differentially expressed between CFT073 cells after ampicillin vs. after dilution from overnight. The 10 genes most significantly upregulated are labeled, as is common persister marker *cysK*. Only significant genes are shown (two-sided Mann-Whitney U; FDR = 0.01). Black dots indicate genes expressed in *metG** persisters ($p_{overlap} = 0.0005$; hyper-geometric test). Genes upregulated in tetracycline also overlap ($p = 0.003$). *Right*: selected pathways are listed that are significantly under- ($p < 0.05$; iPAGE) or overexpressed ($p < 0.01$; iPAGE) after ampicillin treatment (** $p < 0.0005$, ** $p < 0.005$, * $p < 0.05$; red and green indicate pathways also differentially expressed in stationary or tetracycline, respectively).



Extended Data Fig. 7 | Bulk proteomics supports translational deficiency of *metG persisters.** **a**, Projecting proteomes with single-cell transcriptomes shows that *metG** persister proteomes cluster with stationary phase proteomes and transcriptomes. *Left*: 40,000 representative cells are shown and colored by cell cluster. *Right*: Bulk proteomes on the same PCA ($n = 3$ biological replicates each panel) colored by cluster. **b**, Similar to panel a but UMAP. **c**, PCA of proteomes only. Unlike scRNA-seq, persister proteomes are similar to stationary cells and do not appear in a transitional state between stationary and exponential phase. **d**, Differential protein analysis vs. differential RNA analysis for persisters vs. stationary. Upregulated genes by RNA mostly are not

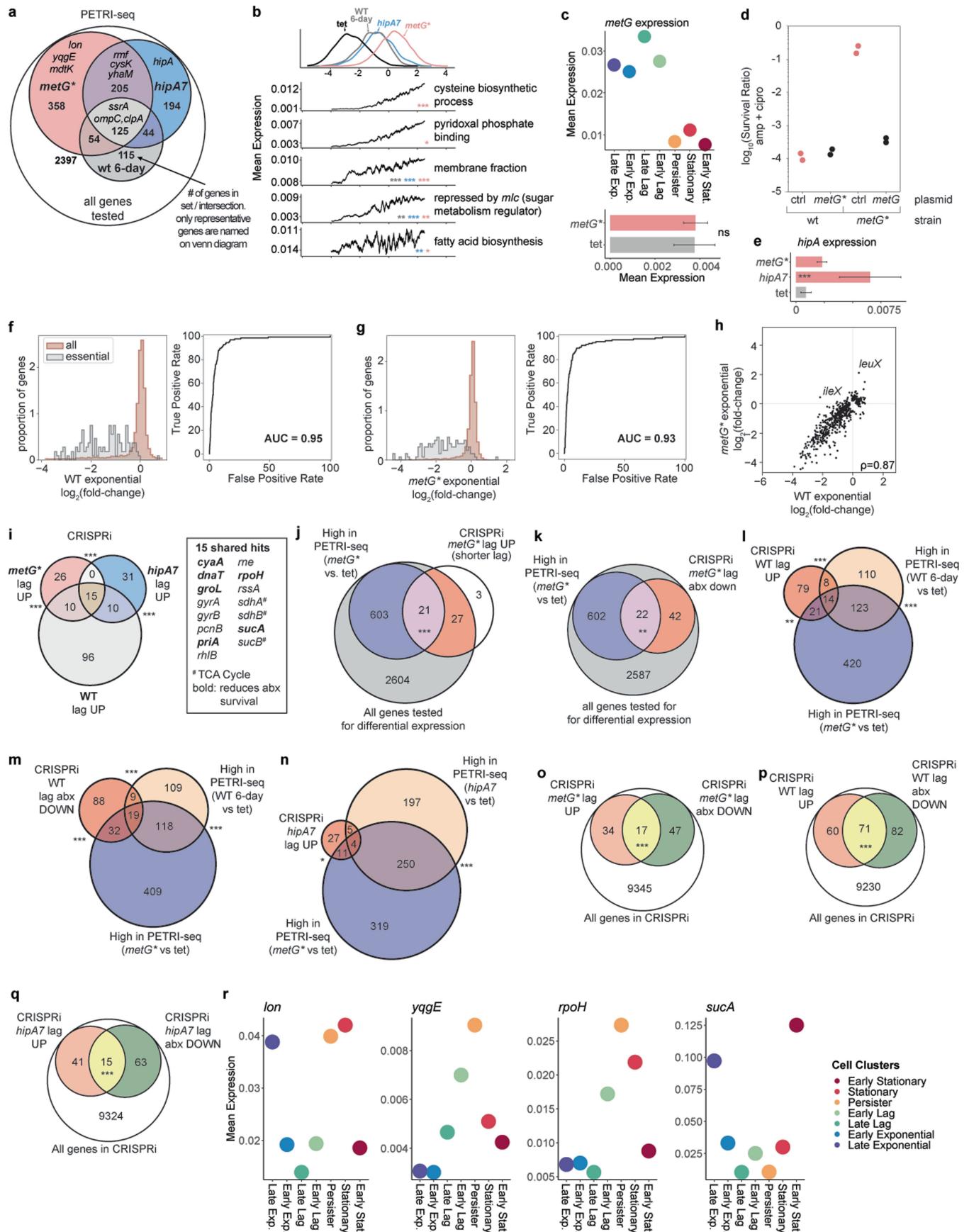
upregulated proteins. Only 7 proteins are upregulated in *metG** lag (persisters) vs. *metG** stationary; 3 are cold-shock genes (*cspA*, *cspG*, *deAD*) known to be upregulated by low translation^{93,94}. **e,f**, Differential protein analysis vs. differential RNA analysis for exponential or persister vs. stationary cells. Proteins expressed in exponential phase (e) but also with transcripts in persisters (f), corroborating that persister transcriptomes are in a transitional state between stationary and exponential. In d-f, significant proteins are defined by FDR = 0.05 (Methods). Source data for proteomics comparisons in d-f are included in Supplementary Table 2.

Article



Extended Data Fig. 8 | Transcriptional inhibition does not recapitulate the persister transcriptional state. **a**, Schematic of experiment. **b**, Rifampicin increases antibiotic tolerance of wildtype cells (one-sided Mann-Whitney U test; * $p = 0.03$; WT (no rifampicin): $n = 3$; WT in rifampicin: $n = 4$ biological replicates). Mean and standard deviation are shown. **c**, Rifampicin is bacteriostatic ($n = 5$ biological replicates). Mean and standard deviation are shown. **d**, Cells in rifampicin ($n = 664$) have a median of 111 RNAs per cell, which is ~ 3 -fold fewer than the median captured for all other conditions ($n = 154,714$ cells). **e**, Cells in rifampicin have extremely low mRNA content. In the same cells as **d**, the mRNA count is 0 in rifampicin. The dotted line defines a threshold for “transcriptional deficiency” (rifampicin mean + $2\text{SD} = 2.9$). In **d** and **e**, center line, boxes, and whiskers show median, 25th/75th percentile, and minimum/maximum non-outlier value, respectively. Outliers are upper/lower quartile +/- 1.5*interquartile range. **f**, Percent of all cells in the persister cluster (left) or below the transcriptional deficiency threshold (right) for rifampicin-treated and populations in Fig. 2f (* $p < 0.05$ one-sided Mann-Whitney U test; WT: $n = 7$; *hipA7*: $n = 3$; WT 6-day: $n = 2$; tet: $n = 3$; *metG**: $n = 3$; rif: $n = 1$). Center line, boxes,

and whiskers show median, 25th/75th percentile, and minimum/maximum value, respectively. Though WT 6-day, *metG**, and tetracycline-treated populations have more transcriptionally deficient cells, cells in the persister cluster (left) are >8 -fold more abundant. **g**, Translational deficiency is sufficient to explain decreased mRNAs per *metG** persister cell. *metG** and tetracycline-treated cells express fewer mRNAs than untreated wildtype cells (“WT”; $p = 0.01$, one-sided Mann-Whitney U test). Center line, boxes, and whiskers show median, 25th/75th percentile, and minimum/maximum value, respectively. **h**, Dead cells give the same signal as alive, transcriptionally deficient cells. *Left*: After ampicillin, the percent of sequenced cells below the transcriptional deficiency threshold increases ($n = 1$ for all sample types). *Right*: Minimum percent of sequenced cells that are dead (CFU counts / number of sequenced cells). Only minimum can be reported because of variable cell loss during PETRI-seq. **i**, After rifampicin, only 1 (0.06%) cell has enough mRNA to be included in the cell atlas. After ampicillin, wildtype MG1655 or UPEC CFT073 populations include 0.6% or 2.0% cells above the threshold. These are likely alive persister cells.



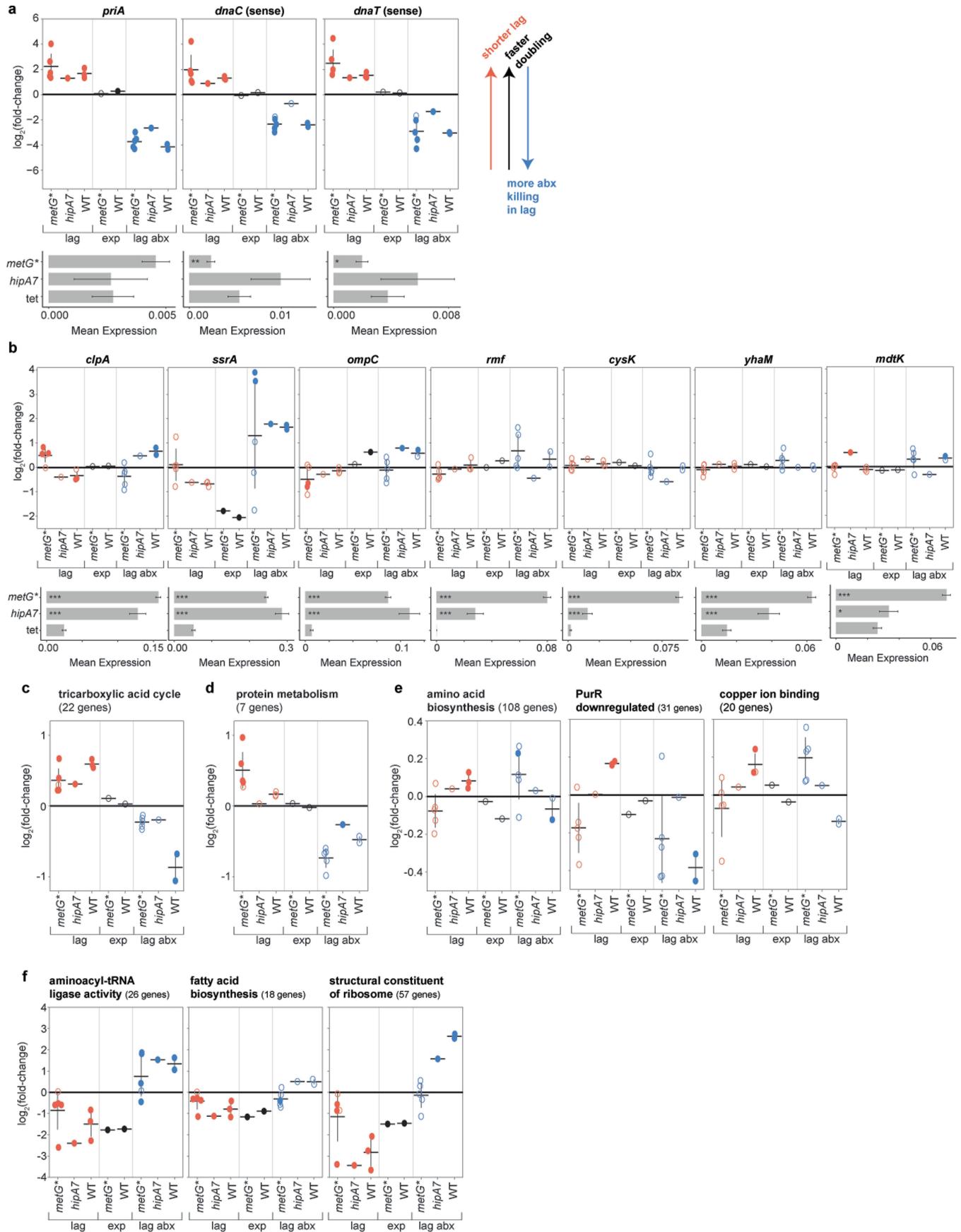
Extended Data Fig. 9 | See next page for caption.

Article

Extended Data Fig. 9 | Gene expression across persister types and CRISPRi.

a, Overlap between genes overexpressed in *metG**, *hipA7*, and 6-day wildtype persisters vs. tetracycline-treated cells (two-sided Mann-Whitney U test, FDR = 0.1). Selected genes are annotated, and the number of genes in each region is noted. All two-way overlaps are highly significant (hyper-geometric test; $p < 10^{-40}$). The most significant markers shared by all persister types are *ssrA*, *ompC*, and *clpA*. *metG** markers (Extended Data Figs. 1p, 3a–e) are shown. **b**, Same as Fig. 3c for remaining significant pathways. **c**, Expression of *metG* by cell cluster (*top*) or in *metG** persisters vs. tetracycline-treated cells (*bottom*; $p > 0.05$; two-sided Mann-Whitney U test). Mean and standard error shown. **d**, *metG** is a hypomorphic allele. Expression of wildtype *metG* in the mutant *metG** background restores wildtype persistence levels ($n = 2$). Expression of mutant *metG** gene in wildtype background has no effect. **e**, *hipA* is highly expressed in *hipA7* persisters relative to tetracycline treatment (** $p < 0.0004$; two-sided Mann-Whitney U test). Mean and standard error shown. **f,g**, Essential

gene crRNAs are depleted after wildtype (a) and *metG** (b) exponential outgrowth (a: $n = 4422$, b: $n = 4408$ genes [126 essential]). **h**, Correlation between gene perturbations in wildtype exponential versus *metG** exponential outgrowth. Only genes significant in least one condition are shown ($n = 572$). tRNAs *ileX* and *leuX* are labeled as they visibly deviate. **i, Left:** Overlap between genes that shorten lag times when targeted by CRISPRi. All two-way overlaps are highly significant (hyper-geometric test; $p < 10^{-20}$). Note that the stringency for *hipA7* is lower here than in Fig. 4c. **Right:** 15 hits shared between all. Bolded also reduce antibiotic survival in all genotypes. **j–q**, Venn diagrams of PETRI-seq and CRISPRi hits. Significant overlaps indicated (** $p < 0.0005$, ** $p < 0.005$, * $p < 0.05$; one-sided hypergeometric test). **r**, Expression by cell cluster of genes in Fig. 4d,e. Fig. 4e shows expression of each gene after splitting the persister cluster by cell type (*metG**, *hipA7*, tetracycline-treated). Here, the mean expression for all cells in the cluster is shown for broad context.

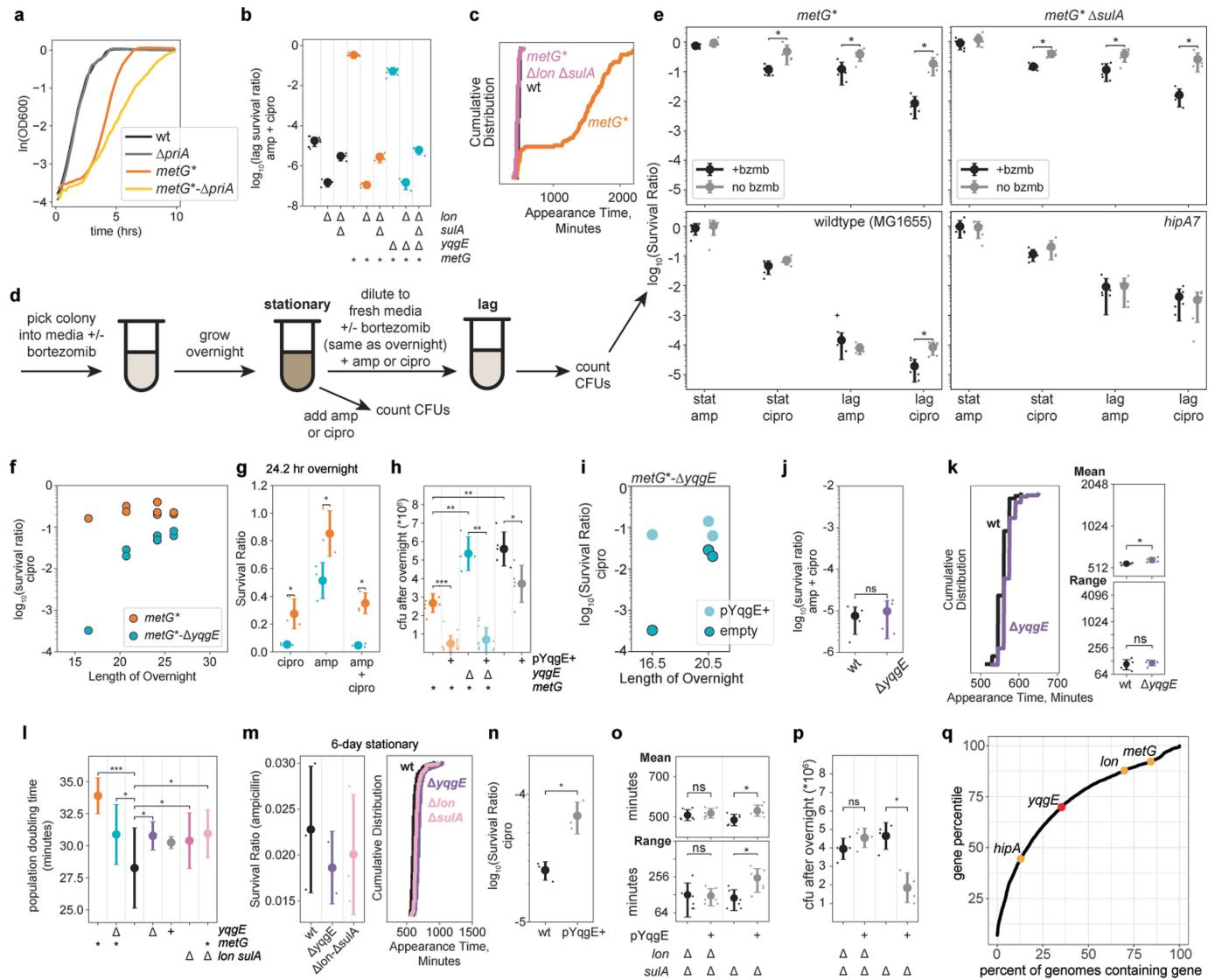


Extended Data Fig. 10 | See next page for caption.

Article

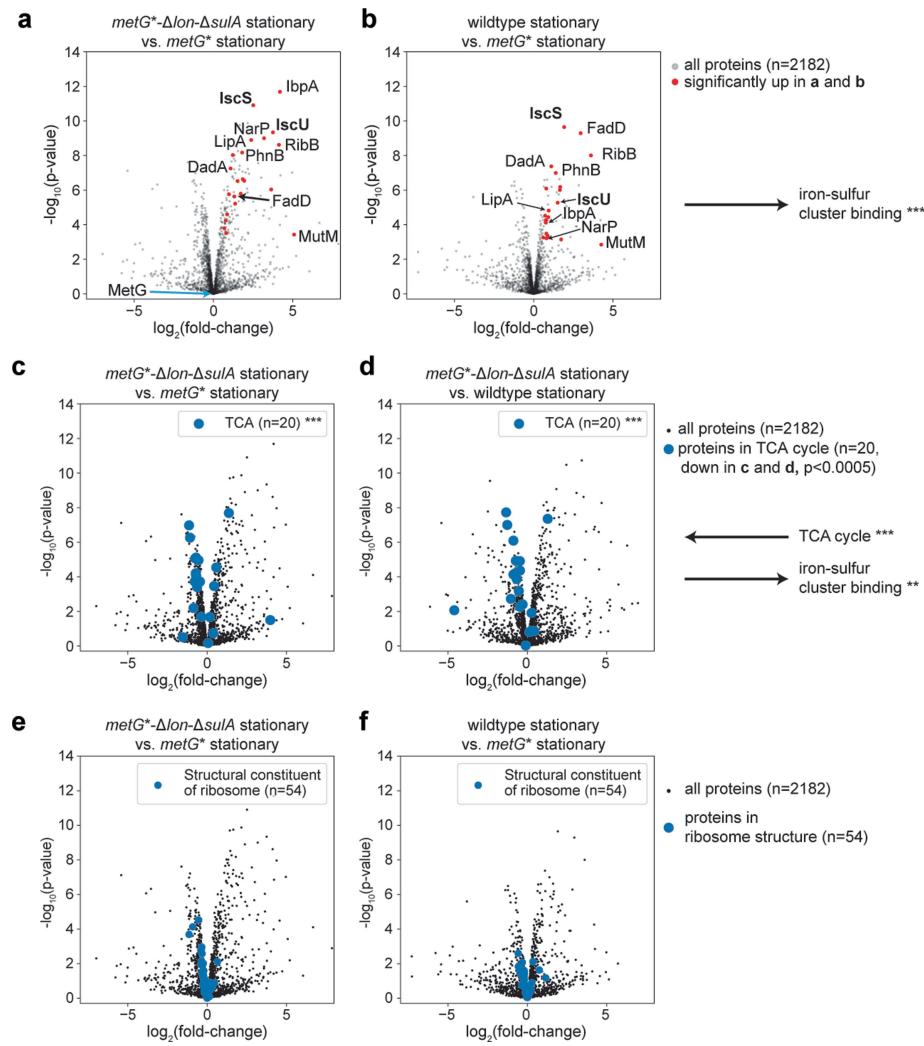
Extended Data Fig. 10 | CRISPRi and expression data for selected genes and pathways. **a**, As in Fig. 4d,e but for additional genes *priA*, *dnaC*, and *dnaT* (all components of the primosome³⁵). For *dnaC* and *dnaT* (essential genes), sense crRNAs are shown. **b**, As in panel a but for selected persister marker genes found by scRNA-seq (Extended Data Fig. 9a). **c**, Repression of tricarboxylic acid (TCA) cycle genes reduces *metG**, *hipA7*, and WT lag times. **d**, Repression of protein metabolism genes reduces *metG** lag time. **e**, CRISPRi knockdown of 3 pathways shortens wildtype lag time. **f**, CRISPRi for pathways expressed in *metG** persisters (identified in Fig. 3c, Extended Data Fig. 9b). Pathways shown

may be important for persister recovery, as repression lengthens lag times (red dots), although they also affect growth (black dots). Other gene sets (cysteine biosynthetic process, pyridoxal phosphate binding, membrane fraction, genes repressed by *m/c*) do not show significant effects after CRISPRi perturbation. In **c–f**, each dot shows mean fold-change replicate; solid circles indicate significant enrichment or depletion (iPAGE; $p < 0.01$; $n = 5$ both *metG** lag, $n = 3$ wildtype lag, $n = 2$ wildtype lag antibiotic, $n = 1$ both *hipA7*, $n = 1$ both exponential). Horizontal line and error bars shows mean and standard deviation of replicates.



Extended Data Fig. 11 | Additional characterization of persistence drivers, supplement to Fig. 5. **a**, priA deletion does not reduce metG^* lag time ($n = 2$ biological replicates per genotype). **b**, Lag phase antibiotic survival for various genotypes. Mean and standard deviation shown. **c**, Cumulative distribution of colony appearance times. lon deletion reverses metG^* phenotype. **d–e**, Effect of bortezomib on antibiotic survival. As outlined in **d**, antibiotics were added either to the overnight ("stat") or the diluted culture ("lag"). **e**, Survival ratio after antibiotic (* $p = 0.03125$, one-sided Mann-Whitney U test, $n = 6$ biological replicates; mean and standard deviation shown; '-' symbol shows excluded outlier). **f**, Survival vs. length of overnight as in Fig. 5c but for ciprofloxacin only. **g**, Lag phase antibiotic survival for $\text{metG}^* \Delta yggE$ cells after 24-hour overnight (one-sided Mann-Whitney U; * $p < 0.04$; $n = 3$). Mean and standard deviation shown. **h**, pYggE+ decreases CFUs after overnight growth (* $p < 0.05$; one-sided Mann-Whitney U tests). Mean and standard deviation shown. Strains without pYggE expressed pRFP+ as a control. **i**, pYggE+ increases lag phase antibiotic survival of $\text{metG}^* \Delta yggE$ (one-sided Mann-Whitney U test; $p = 0.04$). **j**, Deletion of $yggE$ ($\Delta yggE$) has no effect on lag phase survival in wildtype background after 17-hour overnight growth (one-sided Mann-Whitney U, $n = 4$). Mean and standard deviation shown. **k**, Representative distributions of colony appearance times. Right: Statistics with more replicates (one-sided Mann-Whitney U test; * $p < 0.05$; $n = 4$). $\Delta yggE$ cells have a marginally later mean appearance time than wildtype cells, which can be explained by slightly longer doubling time (panel l). Range (heterogeneity) does not increase. Mean and standard deviation shown. **l**, Doubling times of relevant mutant and deletion

strains [*** $p < 0.0005$; * $p < 0.05$; one-sided Mann-Whitney U tests; $n = 16$ (metG^*), $n = 13$ ($\text{metG}^* \Delta yggE$), $n = 25$ (wt), $n = 7$ ($\Delta yggE$), $n = 3$ (wt-pYggE+), $n = 9$ (wt- $\Delta \text{lon} \Delta \text{sulA}$; $\text{metG}^* \Delta \text{lon} \Delta \text{sulA}$)]. Doubling time was calculated during the exponential growth phase using OD600 measurements from a plate reader. Mean and standard deviation shown. **m**, lon or $yggE$ deletion in wildtype background have no effect on lag phase persistence rates after 6-day stationary phase. Left: Lag phase ampicillin survival. Mean and standard deviation shown. Right: Colony appearance times ($n = 1$ for each genotype). **n**, Overexpression of $yggE$ (pYggE+) increases lag phase survival in wildtype cells after 16-hour overnight ($p = 0.04$; one-sided Mann-Whitney U test; $n = 3$). Mean and standard deviation shown. **o**, Lon is required for extended lag by $yggE$ overexpression. pYggE+ does not affect the mean (top) or range (bottom) of lag times in the $\Delta \text{lon} \Delta \text{sulA}$ background (ns $p > 0.05$; * $p < 0.05$; $n = 6$). Each small dot shows summary statistics for a single biological replicate population, while the large dot and error bar show the mean and standard deviation across populations. **p**, Lon is required for $yggE$ overexpression to decrease CFUs after overnight growth (ns $p > 0.05$; * $p = 0.01$; one-sided Mann-Whitney U test; $n = 4$). Mean and standard deviation are shown. In o-p, strains without pYggE expressed pRFP+ as a control. **q**, Cumulative distribution of E. coli genes based on the number of species in which they are found. We searched for homologs of all E. coli genes across a dataset of 2421 diverse taxa (including 39 phyla; Methods). Query genes were ranked by the percent of genomes in which they were found. $yggE$ was detected in 35% of genomes, which is the 70th percentile of E. coli genes. Additional relevant genes shown for reference.



Extended Data Fig. 12 | Proteomics of *Δlon* cells implicate iron-sulfur cluster assembly and TCA regulation in Lon-mediated persistence.

a–b, Volcano plots showing proteins differentially expressed between stationary *metG^{*}-Δlon-ΔsulA* cells vs. stationary *metG^{*}* cells (a) or stationary wildtype cells vs. stationary *metG^{*}* cells (b). Red dots indicate proteins significantly upregulated in both comparisons ($n = 22$; FDR = 0.05; Methods), suggesting these proteins are stabilized by *lon* deletion and more likely to be degraded in *metG^{*}* than wildtype cells. Top differentially expressed proteins are labeled and include the iron-sulfur cluster assembly proteins IscU and IscS. IscS has been found previously as a target of Lon⁵⁶, and the yeast homolog of IscU has been previously identified as a target of Lon⁹⁶. On the right, iron-sulfur cluster binding proteins are listed as the only protein set significantly dysregulated in both a and b ($p = 0.00004$; iPAGE). **c–d**, *lon* deletion results in lower abundance

of TCA cycle proteins. Volcano plots show differential protein abundance between stationary *metG^{*}-Δlon-ΔsulA* cells vs. stationary *metG^{*}* cells (c) or stationary *metG^{*}-Δlon-ΔsulA* cells vs. stationary wildtype cells (d). TCA cycle proteins are significantly enriched among proteins with negative fold-change ($p = 0.00009$ *metG^{*}*, 0.00001 wt; iPAGE). Reduction of TCA cycle proteins by absence of Lon has been seen previously⁵⁶. In Extended Data Fig. 10c, we also show functionally that loss of TCA cycle expression decreases lag time and persistence. As in a and b, iron-sulfur cluster binding proteins are enriched in both panels c and d among proteins with positive fold-change ($p = 0.002$ vs. wt, iPAGE). **e–f**, Ribosomal proteins are not stabilized by *lon* deletion (e) or depleted in *metG^{*}* cells (f). Instead, they are modestly reduced by *lon* deletion ($p = 0.03$ [e]). Source data for all panels is included in Supplementary Table 2.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Sequencing was done on a NextSeq 500 at the Columbia Genome Center. Fastq files were generated using bcl2fastq (v2.20.0.422). Scanlag was run using ScanningManager software (https://github.com/IritLevin/ScanningManager). Proteomics was done as described in Methods by the Proteomics and Macromolecular Crystallography Core at Columbia University Irving Medical Center.
Data analysis	Cutadapt 1.18, UMI-tools 0.5.5, bwa 0.7.17, featureCounts 1.6.3, PEAR 0.9.11 and trimmomatic 0.33 were used for processing sequencing data. Seurat 4.1.1 was used for scRNA-seq analysis. EdgeR 3.36.0 and limma 3.50.3 were used for differential expression analysis. Custom code for CRISPRi processing and statistics calculations were implemented in Python 2.7.17 or 3.6.9. Flow cytometry graphs were generated in FlowJo 10.8.1. MaxDIA and DEP 1.22.0 were used for processing proteomics data.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Sequencing data has been uploaded to the Gene Expression Omnibus (GEO) under accession number GSE229435.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender No human studies were performed.

Reporting on race, ethnicity, or other socially relevant groupings No human studies were performed.

Population characteristics No human studies were performed.

Recruitment No human studies were performed.

Ethics oversight No human studies were performed.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

Behavioural & social sciences

Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

For antibiotic and lag time assays, at least three replicates were used to enable statistics by Mann-Whitney U tests.

For scRNA-seq, the number of cells was chosen to minimize multiplet frequency and to enable differential expression testing (at least 1000 cells based on previous comparisons between single cells and bulk RNA-seq [Blattman, Jiang et al 2020: 10.1038/s41564-020-0729-6]). Cell counts by experiment are listed in Supplementary Table 3. For a given condition, scRNA-seq was carried out on 2-3 biological replicate populations. We found this sufficient as there are many cells in each sample, and replicates were very consistent.

For CRISPR screens, replicates are listed in Supplementary Table 5. However, each individual crRNA (average 100 per gene) can be treated as a replicate to enable robust statistics on a single screen. To be thorough and to check reproducibility, we included 5 replicate metG* screens and 3 wildtype screens. The top hits are consistent, which is why we determined 1 replicate screen to be sufficient for hipA7.

Data exclusions

No data were excluded.

Replication

For scRNA-seq experiments, metG* and wildtype cells were sampled in multiple experiments conducted on different days. hipA7 and 6-day wildtype cells were sequenced from 2-3 biological replicates. Cells in tetracycline were sequenced for 2 technical replicates. metG* cells in ampicillin were sequenced for 2 biological replicates. All replicates are consistent and are shown in Extended Data Fig. 4. We include one replicate of CFT073 cells without or with ampicillin, of MG1655 cells in ampicillin, and of MG1655 in rifampicin.

Replicate CRISPR screens are listed in Supplementary Table 5. 5 metG* screens, 3 wildtype screens, and 1 hipA7 screen were carried out.

For antibiotic survival and lag time assays, the number of biological replicates is noted in the figure legend or source data. Statistics are given only when n>2.

Randomization

Randomization is not relevant to this study because comparisons were made for the different conditions and genotypes of the same parental bacterial strain. Biological replicates (individual colonies) were included to account for variation.

Blinding

Researchers were not blinded during the study because experiment planning and execution were carried out by one person.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	Antibodies
<input checked="" type="checkbox"/>	Eukaryotic cell lines
<input checked="" type="checkbox"/>	Palaeontology and archaeology
<input checked="" type="checkbox"/>	Animals and other organisms
<input checked="" type="checkbox"/>	Clinical data
<input checked="" type="checkbox"/>	Dual use research of concern
<input checked="" type="checkbox"/>	Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	ChIP-seq
<input type="checkbox"/>	Flow cytometry
<input checked="" type="checkbox"/>	MRI-based neuroimaging

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Overnight or exponential growing bacterial cells were pelleted and resuspended in PBS for analysis.

Instrument

S3e Cell Sorter (12007058, Bio-Rad)

Software

ProSort (BioRad) was used during data collection. FlowJo 10.8.1 was used for further analysis and plotting.

Cell population abundance

Antibiotic survival and lag times were measured for sorted populations.

Gating strategy

Gating strategy is shown in Extended Data Fig. 3h-i. First, cells are selected by SSC-Area vs. FSC-Area to exclude outliers (h). Then, singlets are selected as the upper distribution on FSC-Height vs. FSC-Area (i).

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.