Massively-parallel Microbial mRNA Sequencing (M3-Seq) reveals heterogenous behaviors in bacteria at single-cell resolution

Authors: Bruce Wang[2], Aaron E. Lin[1,2], Jiayi Yuan[1], Matthias D. Koch [1,2], Britt Adamson [1,2] *, Ned S. Wingreen [1,2 *], Zemer Gitai [1]*

Affiliations:

[1] Department of Molecular Biology, Princeton University, Princeton, NJ 08544, USA

[2] Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA

*Corresponding authors. Emails: badamson@princeton.edu (B.A.); wingreen@princeton.edu (N.S.W.); zgitai@princeton.edu (Z.G.)

Abstract: Bacterial populations are highly adaptive, enabling them to respond to and surviving in shifting environments or stresses. Yet, how these single-cell organisms vary and organize their behavior to tolerate stressors is poorly understood. This is because many bacterial subpopulations are rare and cannot be readily discovered by existing single-cell sequencing methods due to limitations in cell number and sequencing depth. Here we develop Massively-parallel Microbial mRNA sequencing (M3-Seq), which addresses these challenges by using combinatorially-indexed cells to overload droplets in combination with RNA amplification and post-hoc rRNA depletion. In a single M3-Seq experiment, we profile hundreds of thousands of bacterial cells from multiple species under a wide range of conditions. In addition to validating our approach and findings, we exploit the scale of M3-Seq to make several unexpected discoveries, including new insights into bet hedging strategies in stress responses, bacterial responses to antibiotics, and host responses to phage infection.

Introduction

Bacteria have a remarkable ability to survive and adapt in diverse, changing environments. This remains true even if a population of bacteria originates from a single parent bacterium. Individual cells can specialize at the transcriptional level, allowing the population as a whole to flourish in the face of unpredictable environmental stressors. These specializations can manifest in morphologically distinct states such as the sporulation phenotype in certain Gram-positive organisms such as *Bacillus subtilis*, or morphologically-indistinguishable but functionally distinct states such as the rare (1 in 10,000) antibiotic-resistant "persister" phenotypes in species such as *Staphylococcus aureus* and *Escherichia coli* (*1–3*).

To recognize the full range of specializations within bacterial populations, it is necessary to measure how single cells orchestrate the expression of multiple genes over different stressors. Single-cell RNA Sequencing (scRNA-seq) has transformed how researchers study phenotypic variation in eukaryotic systems via the ability to profile the transcription of all the genes in single cells (*4–6*). Despite the blazing pace at which single-cell studies have been developed for eukaryotic cells, there have only recently been a handful of techniques developed for bacterial cells (*7–11*). These pioneering methods represented crucial steps towards the systematic study of single-cell transcriptomes in bacteria but are still limited in coverage, scale, and application (Fig. 1A).

Previous efforts to profile single bacterial cells have relied on in-situ combinatorial plate-based indexing (PETRI-Seq and micro-Split) (*7*, *8*), isolating single cells in single wells (MATQ-Seq) (*10*), or hybridizing mRNA transcripts with specific probes followed by imaging (parSeqFISH) (*11*), or droplet-based indexing for sequencing (*9*). An advantage to combinatorial plate-based and single-well based indexing methods is that they are unbiased in the genes detected and can profile multiple conditions; however, current approaches have only been extended from a hundred (*10*) to around twenty thousand cells (*7*, *8*) because of the limited number of available indices and the abundance of ribosomal RNA (rRNA) (95%) in the final library. These approaches also typically use multiple in-situ reactions on un-amplified mRNA transcripts, creating a bottleneck for mRNA capture that decreases overall mRNA detection. Alternatively, transcript probe-based approaches do not capture rRNA and have a higher transcript detection efficiency, but require *a priori* selection of dedicated probes for each species and gene studied and can thus either study hundreds of thousands of cells and conditions for a handful of genes (*11*), or thousands of cells for all the genes within one condition (*9*). Each of these approaches requires trade-offs to be made amongst the number of cells, conditions, and mRNAs captured (Fig. 1A), limiting their ability to detect rare bacterial subpopulations and understand how these subpopulations vary across different classes of stressors.

Here, we report massively parallel microbial mRNA single-cell sequencing (M3-Seq), which addresses the existing limitations of bacterial single-cell RNA sequencing to profile hundreds of thousands of bacterial cells and tens of samples in a single experiment while maintaining a high mRNA capture rate. M3-Seq achieves this scale by sequentially employing plate-based in-situ barcoding using random primers followed by droplet-based barcoding and a novel post-hoc rRNA depletion scheme. We harnessed the power of M3-Seq to discover the independence of distinct phage induction programs in *B. subtilis*, the differentiation of a bet-hedging subpopulation in *E. coli*, heterogeneities in antibiotic responses in multiple species, and host-pathogen dynamics upon phage infection.

M3-Seq captures rRNA-depleted whole transcriptomes of single bacterial cells

To expand the scale and scope of scRNA-seq in bacteria, we reasoned that we could perform combinatorial indexing followed by droplet-based indexing to index a large number of single cells, and then apply RNA amplification and rRNA depletion to enrich for mRNA transcripts. We designed our platform as follows (Fig. 1B): Bacterial cells grown under conditions of interest are fixed (overnight incubation with 4% formaldehyde), washed, and permeabilized with lysozyme (similar to (*7*, *8*)). Cells are then subjected to two rounds of indexing: first, using in-situ reverse transcription in plates with random priming to mark individual transcripts with one cell barcode (BC1) and a unique molecular identifier (UMI); then, using a commercially available droplet-based, single-cell kit (Chromium Next GEM Single Cell ATAC, 10X Genomics) to ligate another cell barcode (BC2) onto the resulting cDNA molecules. After indexing, mRNA sequence enrichment is performed by depleting cDNA molecules containing abundant rRNA sequences, and the resulting gene expression libraries are sequenced (Fig. S1).

First, we verified that we could barcode a large number of bacterial cells by combinatorial-fluidic indexing, which was originally developed for mammalian cells (*12*). In principle, a minimal number of indices (<300) in the first indexing round, multiplied by the ~100,000 potential barcodes introduced during droplet-based second round of indexing, should lead to hundreds of thousands of unique cell indices. However, bacterial cells are wildly different from mammalian cells—they have thick cell walls that require enzymatic digestion, and are much smaller — so we optimized a number of different conditions based on (*7*–*9*). To implement this approach in bacteria, we evaluated bacterial cell loading into droplets and used that information to calculate expected index collision rates across different numbers of round-one indices (Fig. S2A). While standard loading protocols suggest a maximum of 10,000 cells per droplet-generation run to minimize index collision rate, distributions of bacterial cells in droplets suggest that using less than 300 round-one indices would enable hundreds of thousands of cells to be loaded in a single run of the system with <1% collision rate (Fig. S2B).

We sought to verify that we could generate single-cell transcriptomes at massive scale with our approach. We mixed equal numbers of cells from different bacterial strains (*B. subtilis* 168 and *E. coli* MG1655) at exponential and stationary phases, performed combinatorial indexing using 96 indices on the combined samples, and then loaded the droplet-generating system with 100,000 cells (experiment BW1 in Table S2). Without round-one indexing, we would expect 15.7% of all sequence-able droplets to yield an index collision (Fig. S2B). Consistent with these expectations, indexing the data with only droplet indices yielded a 12.7% collision rate (Fig. S2C). However, by incorporating the combinatorial indices in our analysis, the collision rate decreased by an order of magnitude to 0.7% (Fig. S2D). Our platform can therefore attain single-cell resolution at cell numbers an order of magnitude larger than previously achieved for bacteria (Fig. S2C-D).

We next examined the transcriptomes provided by these data. Average RNA-Seq profiles were similar to those obtained from bulk RNA-Seq (r = 0.59) (Fig S2E). However, consistent with other methods that use in-situ reverse transcription (*7*, *8*), most of our reads aligned to rRNAs, and consequently, the mRNA read depth per cell was low. Indeed, of roughly 1,000-2,000 reads per cell, only ~100 (3-10%) aligned to mRNA, and the rest aligned to rRNA (90-97%) (Fig. S2F-G). This problem could be overcome by sequencing to a greater depth (~40,000 reads per cell), but this approach becomes cost-prohibitive as the number of cells increases.

To profile bacterial transcriptomes more deeply across many cells, we therefore implemented rRNA depletion in conjunction with RNA amplification. Previous work has

suggested that rRNA depletion within fixed cells decreases mRNA capture efficiency (*7*). Therefore, we developed a custom approach to amplify total RNA after combinatorial-fluidic indexing and barcoding, and then deplete rRNA in solution (Fig. 1B, Fig. S3A, B). First, we fragment the gene expression cDNA library with Tn5 transposase and amplify the library with a primer containing a T7 RNA polymerase promoter. Then, we convert the resulting T7-tailed, double-stranded DNA library to single-stranded RNA using *in vitro* transcription. Next, we hybridize with rRNA-specific DNA probes and then treat the RNA library with RNase H to digest rRNA-DNA hybrids. Finally, we reverse transcribe the rRNA-depleted RNA library back into DNA using an end-specific primer that preserves the single-cell indices. This approach enables rRNA depletion while limiting bottlenecking of informative sequences. In the full M3-Seq pipeline, we recovered transcriptomes with 66.9% of *E. coli* reads and 80.8% of *B. subtilis* reads aligning to mRNA, representing a >20-fold increase from our previous experiment (Fig. 1C, Fig. S3B).

Finally, we applied M3-Seq to evaluate different stages of growth and treatment in *E. coli* and *B. subtilis* ("BW2", Table S2). We recovered single bacterial transcriptome resolution (1.7% doublet rate, Fig. 1D), with a range of 515–984 median UMIs per cell (298–371 median genes) for *B. subtilis* (blue), and 100–211 median UMIs per cell (75–151 median genes) for *E. coli* MG1655 in exponentially growing cells. Compared to state-of-the-art bacterial scRNA-Seq technologies that profiled at similar conditions (*7*–*9*), we capture twice as many UMIs for *B. subtilis*, and roughly the same number of UMIs for *E. coli* (Fig. 1E, Fig.S4A). Moreover, these data revealed a 25-fold increase in mRNA read coverage compared to if there was no rRNA depletion (Fig. 1D). To quantify how our rRNA-depletion method improves the sensitivity of detecting single bacterial transcriptomes, we evaluated the average number of captured genes per cell as a function of sequencing reads per cell. We observed that the un-depleted library requires an order of magnitude more reads (~15X) per cell to achieve the same number of genes detected as that of the depleted single-cell libraries (Fig. 1E). Finally, to ensure that the depletion scheme does not mask the biological signal of the data, we compared the rRNA depleted single-cell libraries to rRNA-depleted bulk libraries of *E. coli* grown in similar conditions and found good agreement in expression between the two (r=0.72) (Fig. S4B). Thus, by combining our post-hoc rRNA depletion with droplet overloading, M3-Seq provides biologically meaningful, mRNA-enriched transcriptomes at single-cell resolution at scales which have not been previously achieved for bacteria.

Early stationary phase *E. coli* contain a rare, acid-tolerant subpopulation as a bet-hedging strategy

We first set out to study exponential phase and early stationary phase, a transition point between rapid cell division versus slow growth as nutrients are depleted from the environment. We grew the lab strain *E. coli* MG1655, and the probiotic strain *E. coli* Nissle 1917 to exponential (OD=0.3) and early stationary phase (OD=2.8, OD=2.6) respectively, and used M3-Seq to profile the cultures. Similar to previous studies, cells in exponential and stationary phase could clearly be distinguished and identified by known biological processes such as glucose metabolism in exponential phase, and secondary carbon metabolism in early stationary phase (Fig. S4C-E). When we looked more closely at the early stationary phase population, we found that both strains of *E. coli* in that growth stage exhibited distinct subpopulations (Fig. 2A). A gene ontology analysis revealed that in addition to the main early stationary phase cluster (cluster 1 in MG1655, Fig. 2A), one subpopulation was dominated by induction of a prophage (cluster 3,

Fig. 2A), and one subpopulation was dominated by genes implicated in intracellular pH elevation and glutamate catabolism (cluster 2, Fig. 2A-B). The most strongly expressed genes in this cluster were *gadA* and *gadB* (Fig. 2A, Fig S5B). GadA and GadB are glutamate decarboxylases that are well conserved among enteric bacteria (Fig. S5A) and de-acidify the cellular cytoplasm by consuming a proton in the process of decarboxylating glutamate to GABA (*13–15*). Indeed, we verified that these genes are essential for acid tolerance by experimental manipulation. Consistent with previous studies, when we deleted the *gadABC* genes in MG1655, the bacterial population was unable to survive acid treatment (Fig. S5C). These observations prompted us to ask: if *gadABC* genes are essential for acid tolerance, what is the functional significance of only having a subpopulation of bacteria express the *gad* genes?

Bulk measurements have established that expression of the *gad* genes is induced in stationary-phase cultures of *E. coli* grown in complex media such as LB, but the heterogeneity of *gad* expression during the transition into stationary phase has not been previously studied. Previous work found heterogeneous expression of *gadA/gadB* in response to antifolate antibiotics, and of *gadX* (an activator of *gadA/gadB*) in exponentially growing cells (*16*, *17*). To confirm our M3-Seq findings, we transformed MG1655 with P_{gadB}-GFP, a plasmid encoding a fluorescent reporter controlled by the *gadB* promoter, and imaged the reporter strains under the same growth condition used for single-cell sequencing (Fig. 2C). We saw a heavy tail of GFP expression in the population, indicating that a small subpopulation of cells activated the *gadB* promoter (Fig. 2C).

We hypothesized that during entry to stationary phase, enteric bacteria asynchronously activate the *gad* genes, perhaps as a bet-hedging strategy to protect some cells against strong acid stress. To test this hypothesis, we developed an imaging-based acid-recovery assay to track individual cells in response to acid treatment (Fig. 2D). We first grew the P_{gadB}-GFP reporter strain to early stationary phase and imaged the population to obtain the GFP fluorescence distribution as a proxy for *gadB* expression in single cells (Fig. 2E). We experimentally acidified the culture from pH 7.5 to 3.0 for 1 hour using hydrochloric acid, and then performed live-cell imaging on an agarose pad made with fresh LB to correlate cell viability and P_{gadB}-GFP expression. A cell was considered "viable" if the cell underwent a division during the 8-hour imaging period after acid treatment (Movie 1).

The intensity of P_{gadB}-GFP in viable cells was much higher than the pre-treatment population average, indicating that cells that had induced *gadB* before acid treatment were pre-disposed to survive acid treatment (Fig. 2E). To further explore the possibility that surviving cells activated the *gadB* promoter before stress (and not in response to strong acid stress), we imaged the reporter strain at early stationary phase under an agarose pad made with acidic LB (pH 3.0) (Movie 2) and found that GFP fluorescence intensity steadily decreased in bacterial cells (Fig. S5D). This suggests that under sudden strong acid stress, *E. coli* cannot activate and translate acid resistance proteins, and must rely on the existing set of translated proteins that cells possess before the stress.

A significant cost for expressing high levels of the *gad* genes that are protective of acid stress would explain the benefit of only having a subpopulation of cells express these genes. To test this hypothesis, we cloned the *gadBC* operon into an inducible-overexpression vector and performed growth curve assays in LB across a range of inducer concentrations (*18*, *19*). We found that, in contrast to overexpressing GFP, overexpression of *gadBC* at even low induction levels (1-10μM, 50-1000x less than typical induction concentrations) causes a growth defect (Fig. 2F, Fig S5E). These data demonstrate how M3-Seq can uncover rare bacterial

subpopulations with important functional consequences such as acid tolerance. Specifically, in early stationary phase, a subpopulation of *E. coli* may activate the *gad* genes to protect against strong acid stresses that they may experience when passing through acidic environments like the stomach. Because the acid tolerance gained by *gad* expression comes at the cost of decreased growth of those cells, it may benefit the population to not induce these genes in all cells in case the acid stress does not materialize.

DNA damaging antibiotics independently induce two distinct prophages in *B. subtilis*

In addition to bet-hedging strategies, how single bacterial cells vary their responses to antibiotic-treatment remains an important question in the field. Historically, it has been challenging to profile such cells at scale due to the large number of possible bacterial species and antibiotics. Leveraging the massively parallel scale of M3-Seq, we grew bacterial cultures of *B. subtilis*, *E. coli*, and *Pseudomonas aeruginosa* to early exponential phase (OD=0.3) and treated the bacteria with each of 8 antibiotics of 4 different mechanisms (DNA damaging agents, cell wall inhibtors, 30S ribosome, 50S ribosome) at 2 times the minimum inhibitory concentration (2X MIC) (Fig. S6A) for 90 minutes. We sequenced a total of 29 different conditions in a single experiment, recovering a total of 263,760 single bacterial transcriptomes from a million loaded cells, representing an order of magnitude more cells than previous methods.

B. subtilis cells exhibited a wide variety of transcriptional states upon treatment with DNA damaging antibiotics (ciprofloxacin and nalidixic acid) (Fig. 3A). Upon clustering the data and identifying marker genes associated with each cluster, we found that clusters 4, 5, and 6 had distinct sets of strongly co-expressed genes that belonged to the PBSX or SPβ prophages (Fig. 3B). PBSX and SPβ are the two prophages carried in the *B. subtilis* 168 genome (Fig. 3C). These prophages are latent bacterial viruses that reside within the host genome, and can be induced under conditions that induce the SOS response such as UV light or other DNA damaging antibiotics such as mitomycin C (*20*). Previous single-cell studies have also found that the PBSX prophage is induced in a small fraction of exponentially growing *B. subtilis* PY79 (*7*). However, it was previously unclear whether prophage induction is stochastic or whether prophages induce in the cells with the greatest damage. Similarly, it remained unknown if there is crosstalk between the two phage induction programs.

The depth of coverage afforded by M3-Seq enabled us to observe that upon treatment with DNA damaging antibiotics, the PBSX and SPβ prophages induce in subsets of the treated population (Fig. 3B-D). We observed a single subpopulation (cluster 4) expressing PBSX prophage genes (Fig. 3B-D), and two subpopulations (clusters 5 and 6) expressing genes from the SPβ prophage. These two subpopulations represent different induction stages of the SPβ prophage (Fig. 3B), which is known to have a two-tiered induction strategy in which the late-stage genes are transcribed in a single operon using the SPβ RNA polymerase YonO (*21*). Cluster 5 expresses the early-stage genes of SPβ, and cluster 6 expresses a subset of the early-stage genes and the late-stage genes of SPβ. We validated the heterogeneity in prophage induction identified by M3-Seq by performing single-molecule FISH (smFISH) on ciprofloxacin-treated cells using two different fluorescently labeled sets of probes that hybridize to the most strongly expressed PBSX or SPβ genes (Fig. 3E).

In our M3-Seq data, we noted that while most cells either induced one or the other prophage, some cells induced both prophages (Fig. 3F). To determine if there is cross-talk between the prophages (e.g. if one phage represses the other's induction), or if the prophages might induce purely as a function of the magnitude of DNA damage in the host cell (in which

case the cells with the most damage would induce both phages), we computed the rates of induction for each prophage by thresholding on the percentage of prophage UMIs within the total UMIs detected per cell. We found that PBSX induces in ~17% of cells and SPβ induced in ~14% of cells. The observed rate of co-induction (2.44%) closely matched the rate that would be expected if the two subpopulations induced independently (2.47%), suggesting that the induction of the two prophages in DNA damaged *B. subtilis* 168 is stochastic and independent. We observed no significant cross-repression, nor a positive correlation such as would be expected if the individual cells with the greatest damage had the greatest likelihood of inducing both prophages.

Bacteriostatic antibiotics induce greater diversity of transcriptional states than bactericidal antibiotics in *E. coli*

The heterogeneity of bacterial responses to antibiotics has emerged as an area of great interest. However, to date these studies have primarily been limited to responses to bactericidal antibiotics, as they produce readily-measurable single-cell persistence and tolerance phenotypes (*2*, *22*–*24*). In contrast, little is known about heterogeneity in the response to bacteriostatic antibiotics.

Interestingly, M3-Seq revealed that *E. coli* MG1655 cells treated with bacteriostatic ribosomal antibiotics (tetracycline or chloramphenicol) exhibit a wide variety of transcriptional states corresponding to activation of mobile genetic element (prophages, transposons) (Fig. S6B-C). Some of these small subpopulations could represent rare alternative strategies that *E. coli* use to tolerate or respond to antibiotic treatment or other stresses.

To understand how the ability to detect alternative transcriptional states increases with the number of cells sequenced, we first needed a metric independent from our clustering analysis that captures the degree of transcriptional variability within a population. We found in our data that certain principal components had heavy tails, and cells in these tails were assigned as members of unique subpopulations in our clustering analysis (Fig. S7A). We therefore reasoned that we could assess the magnitude of transcriptional variability in a population by computing the kurtosis (a measure of how heavy the tails of a distribution are) for each principal component (Fig. S7B-C) (*25*).

To estimate our sensitivity to detect rare cell subpopulations, we performed rarefaction analysis in which we down-sampled the cells per experiment and ranked the kurtosis of the principal components of gene expression in each sample (Fig. 4A). We observed that the rare cell populations (such as the populations expressing *insI-2,* Fig. S7D) were undetectable at the limited cell numbers enabled by previously existing methods (~1,000 – 10,000 cells) (Fig. 4A). Furthermore, we found that the kurtosis of the "heaviest" tailed principal components monotonically increased with increasing cell numbers up to the actual number of cells in our experiment (75,000 cells), suggesting that sequencing even more cells could potentially enable discovery of even rarer subpopulations.

Next, we sought to test if the diversity of transcriptional states seen in bacteriostatic antibiotic treated cells was specific to the target of the drug (30S/50S ribosome), or the cidality of the drug. To test this, we compared single-cell transcriptomes of MG1655 treated with bactericidal ribosome inhibitors (gentamycin and erythromycin) and MG1655 treated with the above bacteriostatic ribosome inhibitors (down-sampled to the same number of cells as the bactericidal population) (Fig. 4B). We found that bactericidal drugs do not cause as much

variability in transcriptional states as bacteriostatic drugs (by comparing the kurtosis of the heaviest tailed principal components), suggesting that bacteriostatic drugs specifically induce the activation of alternative transcription programs in *E. coli*. Previous studies of antibiotic persistence and tolerance have drawn significant attention to the importance of rare subpopulations in the response to antibiotics. However, antibiotic persistence and tolerance assays depend on cell death, a measure which is not relevant for bacteriostatic drugs. Far less is known about the strategies that subpopulations of bacteria utilize in response to bacteriostatic antibiotics. Our use of M3-Seq to reveal that bacteria also have heterogeneous responses to clinically used bacteriostatic antibiotics should thus motivate future efforts in this understudied area.

M3-Seq enables the study of host-pathogen interactions at the single-cell level

Because M3-Seq can detect prophage gene expression within single cells, we reasoned it might also be able to detect the gene expression of new phage infections in single cells, thus providing a genomic platform to study how phages and host bacteria interact, in a manner similar to recent eukaryotic host-viral studies (*26*, *27*). Previous bulk transcriptomic measurements in the well-studied *E. coli*-λ system have claimed that *E. coli* host does not specifically respond to λ phage induction (*28*, *29*). However, because phage infection and induction are known to be heterogenous even in well-mixed environments, we hypothesized that bulk measurements could not accurately capture the host-pathogen dynamics that occur during infection (*30–32*). To precisely study how *E. coli* and λ behave at the single-cell level during infection, we infected *E. coli* MG1655 at exponential phase (OD=0.3) with λ phage at saturating multiplicity of infection (MOI=100) and sampled the cultures at 30- and 90-minutes post infection (Fig. 4C.I). We aligned our reads to a combined genome that includes both the *E. coli* and λ genes and found that there was a distinct subpopulation of cells expressing high levels of late-stage λ genes (Fig. 4C.II, Fig. 4D), indicating that these cells were undergoing lytic infection. We found that we could still differentiate the subpopulation by performing the analysis with only λ genes, but this subpopulation was not as strongly differentiated when we performed the analysis using only *E. coli* genes (Fig. 4C). This result suggested that the lytic subpopulation in our data was primarily differentiated by phage gene expression.

We confirmed the limited nature of the host response to lytic infection by computing the silhouette scores of cells undergoing lysis under all 3 combinations of host and pathogen genes *(E. coli* genes only, λ genes only, *E. coli* + λ genes) *(*Fig. 4D). The silhouette score is a "goodness of clustering" metric that quantifies how similar cells in a cluster of interest are to each other compared to cells outside the cluster of interest (*33*). In our application, we compute the silhouette scores across the subsets of genes in the principal component space to see if the lytic cluster has distinguishable transcriptional signatures on both the host (*E. coli*) and the pathogen (λ) side. We found that the group of cells undergoing lysis had a strong silhouette score when aligned to genomes that included λ genes, but the signal was drastically decreased when those cells were aligned only to the *E. coli* genome. Specifically, the set of genes involved in host response are genes that respond to the downstream effects of lysis, such as dNTP depletion from phage DNA replication (Fig. S8A-B). This suggests that despite the production of hundreds of foreign virions during late-stage lysis, the *E. coli* host genome does not mount a specific transcriptional response in defense when infected during exponential growth (Fig. S8A-B). Despite the intense study of *E. coli* infection by λ for so many years, we believe this is the first

study to demonstrate that bacterial hosts undergoing lytic infection do not mount a specific transcriptional response to lambda phage.

To see if λ had detectable transcriptional programs distinct from the late-stage lytic cycle, we clustered the cells aligned to the combined genome and then examined λ gene expression across the clusters (Fig 4C.III). As expected, the lytic cluster (cluster 4) contained upregulated late-stage λ genes (*S, R, Rz, F*). Interestingly, cluster 1 also upregulated genes in the *nin* region (*ninB-ninI*), which are nonessential λ genes of unknown function in the lytic operon (Fig. 4E). During the early stages of the λ lytic cycle, transcription from the early lytic PR promoter is terminated at different locations in the *nin* region until enough of the λ antiterminator N accumulates to allow the RNA polymerase to read through the terminators and transcribe another antiterminator Q, which in turn enables transcription of the downstream genes (Fig. S8B) (*29, 34, 35*). We hypothesize that cluster 1 represents cells undergoing the early stages of the lytic cycle, during which the PR promoter is active but there are insufficient quantities of antiterminator N, so that only genes in the *nin* region are transcribed. This suggests that during lytic infection, infected cells exist in two transcriptional states: a less studied state in which levels of antiterminator are still low resulting in high *nin* gene expression and low late-stage gene expression, and the more familiar state when levels of antiterminators (N, Q) pass a critical threshold to promote high late-stage gene expression and foreign virion production. Having validated M3-Seq with an extremely well-studied phage whose results could be readily contextualized based on prior studies, we anticipate that this approach will be of great utility to the rapidly expanding field of host-phage competition.

Discussion

Even though bacteria often live as genetically identical populations of single cells, they can readily adapt to new environments due to transcriptional heterogeneity. Identifying and then characterizing these rare subpopulations requires the ability to sequence large numbers of single bacterial cells experiencing a diverse array of experimental manipulations. Here we describe the development of M3-Seq, a two-step procedure of combinatorial-fluidic indexing and post-hoc ribosomal RNA depletion, which simultaneously enables scale in the number of cells profiled (260,000), breadth in the stressors that can be profiled in a single experiment (29), and a high mRNA detection efficiency (100-1,000 UMIs per cell) (Fig. 1A).

Sequencing a large number of cells, and achieving high transcriptome coverage per cell, are both critical to discovering rare subpopulations in biologically meaningful contexts. For example, little is known about the strategies that rare subpopulations of bacteria utilize in response to bacteriostatic antibiotics. After sequencing >75,000 cells using M3-Seq, we successfully identified a number of rare subpopulations. However, rarefaction analysis indicated that we would not have discerned these populations using fewer than 50,000 cells per sample. The scale of M3-Seq also enables analysis of the cross-interactions or independence of subpopulations, as we demonstrated with *B. subtilis* PBSX and SPβ prophage induction.

Why do rare bacterial subpopulations exist among a genetically identical bacterial population? One reason may be that transcriptional heterogeneity may act as a bet-hedging strategy in response to environmental variation, which has been challenging to study with previous methods. We discovered a rare acid-tolerant subpopulation expressing the *gad* genes in *E. coli*. Through genetic manipulation and orthogonal validation, we found that *gad* expressing bacteria could survive strong acid treatment but were less fit in standard growth conditions. Yet, our M3-Seq data show that a rare *gad*+ population exists at early stationary phase, before the

environment becomes acidified, which supports a bet-hedging model. Many questions remain: what mechanisms generates this transcriptional heterogeneity? How do varying environments change the presence of this subpopulation? How prevalent is this bet-hedging strategy in nature?

M3-Seq holds several advantages over existing methods, but also has a few limitations. Pioneering combinatorial indexing-based methods had outstanding scale and UMI capture for the first of their kind, but still had an overabundance of rRNA reads in the final library, limiting the scale of their applications (*7*, *8*). Furthermore, the mRNA detection efficiency was hampered by the multiple consecutive in-situ enzymatic reactions in the protocols. On the other hand, the probe-based, droplet-based approach contained no rRNA reads and offered an outstanding mRNA detection efficiency but could not be readily scaled to multiple conditions or larger numbers of cells (*9*). Imaging based methods captured multiple conditions and many cells but could only capture up to a hundred genes at a time (*11*). In contrast to imaging-based methods, M3-Seq cannot capture the spatial information of single cells, and thus requires further development for studying systems such as biofilms. Finally, we note that another recently published study combines many of the same elements of our method, albeit without post-hoc mRNA depletion and more in-situ enzymatic steps leading to a lower mRNA detection efficiency (*36*). These emerging studies are indicative of the excitement in the field to push the number of cells and conditions profiled with a high capture rate to discover novel bacterial subpopulations.

Looking forward, we see multiple biological systems for which our technology is ripe to be applied to understand and discover new phenomena at the single-cell level, including host-pathogen interactions and mixed-species microbial communities. When we infected *E. coli* with λ phage, we captured both pathogen (λ) and host (*E. coli*) mRNA transcripts in individual cells. We anticipate that applying our method in targeted experiments may provide new insights into how bacteria mobilize immunity mechanisms in response to phage infection. Moreover, this application need not be restricted to bacterial cells. Because of the generality of using random primers and the rRNA depletion scheme, our method can also be employed to study how mammalian cells respond to infection by intracellular pathogens, and how these infecting pathogens respond to host factors. Finally, based on our species mixing experiments, M3-Seq can also be profitably extended to study mixed-species bacterial communities, where we could use our technology to uncover new niches and strategies that emerge within a microbial community.

References

1. T. Dörr, K. Lewis, M. Vulić, SOS response induces persistence to fluoroquinolones in Escherichia coli. *PLoS Genet.* **5**, e1000760 (2009).
2. N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, S. Leibler, Bacterial persistence as a phenotypic switch. *Science (80-.).* **305**, 1622–1625 (2004).
3. F. Peyrusson, H. Varet, T. K. Nguyen, R. Legendre, O. Sismeiro, J. Y. Coppée, C. Wolz, T. Tenson, F. Van Bambeke, Intracellular Staphylococcus aureus persisters upon antibiotic exposure. *Nat. Commun.* **11** (2020), doi:10.1038/s41467-020-15966-7.
4. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*. **161**, 1202–1214 (2015).
5. A. P. Patel, I. Tirosh, J. J. Trombetta, A. K. Shalek, S. M. Gillespie, H. Wakimoto, D. P. Cahill, B. V. Nahed, W. T. Curry, R. L. Martuza, D. N. Louis, O. Rozenblatt-Rosen, M. L. Suvà, A. Regev, B. E. Bernstein, Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-.).* **344**, 1396–1401 (2014).
6. A. B. Rosenberg, C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. T. Graybuck, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, G. Seelig, Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science (80-.).* **360**, 176–182 (2018).
7. A. Kuchina, L. M. Brettner, L. Paleologu, C. M. Roco, A. B. Rosenberg, A. Carignano, R. Kibler, M. Hirano, R. W. DePaolo, G. Seelig, Microbial single-cell RNA sequencing by split-pool barcoding. *Science*. **371** (2021), doi:10.1126/SCIENCE.ABA5257.
8. S. B. Blattman, W. Jiang, P. Oikonomou, S. Tavazoie, Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat. Microbiol.* **5**, 1192–1201 (2020).
9. R. McNulty, D. Sritharan, S. Liu, S. Hormoz, A. Z. Rosenthal, *bioRxiv*, in press, doi:10.1101/2021.03.10.434868.
10. F. Imdahl, E. Vafadarnejad, C. Homberger, A. E. Saliba, J. Vogel, Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat. Microbiol.* **5**, 1202–1206 (2020).
11. D. Dar, N. Dar, L. Cai, D. K. Newman, Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science (80-.).* **373** (2021), doi:10.1126/SCIENCE.ABI4882/SUPPL_FILE/SCIENCE.ABI4882_SM_TABLE_S2.ZIP.
12. P. Datlinger, A. F. Rendeiro, T. Boenke, T. Krausgruber, D. Barreca, C. Bock, Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing 1, doi:10.1101/2019.12.17.879304.
13. M.-P. Castanie-Cornet, T. A. Penfound, † Dean Smith, J. F. Elliott, J. W. Foster, Control of Acid Resistance in Escherichia coli. *J. Bacteriol.* **181**, 3525–3535 (1999).
14. C. Feehily, K. A. G. Karatzas, Role of glutamate metabolism in bacterial responses towards acid and other stresses. *J. Appl. Microbiol.* **114**, 11–24 (2013).
15. A. He, S. R. Penix, P. J. Basting, J. M. Griffith, K. E. Creamer, D. Camperchioli, M. W. Clark, A. S. Gonzales, J. S. Chávez Erazo, N. S. George, A. A. Bhagwat, J. L. Slonczewski, Acid evolution of Escherichia coli K-12 eliminates amino acid

decarboxylases and reregulates catabolism. *Appl. Environ. Microbiol.* **83** (2017), doi:10.1128/AEM.00442-17/SUPPL_FILE/ZAM999117871SD1.XLSX.

16. N. M. V. Sampaio, C. M. Blassick, V. Andreani, J.-B. Lugagne, M. J. Dunlop, Dynamic gene expression and growth underlie cell-to-cell heterogeneity in *Escherichia coli* stress response. *Proc. Natl. Acad. Sci.* **119** (2022), doi:10.1073/PNAS.2115032119.
17. K. Mitosch, G. Rieckh, T. Bollenbach, Noisy Response to Antibiotic Stress Predicts Subsequent Single-Cell Survival in an Acidic Environment. *Cell Syst.* **4**, 393-403.e5 (2017).
18. H. Chen, S. Venkat, J. Wilson, P. McGuire, A. L. Chang, Q. Gan, C. Fan, Genome-Wide Quantification of the Effect of Gene Overexpression on Escherichia coli Growth. *Genes (Basel).* **9** (2018), doi:10.3390/GENES9080414.
19. M. Kitagawa, T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga, H. Mori, Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. *DNA Res.* **12**, 291–299 (2005).
20. S. Krogh, S. T. Jørgensen, K. M. Devine, Lysis genes of the Bacillus subtilis defective prophage PBSX. *J. Bacteriol.* **180**, 2110–2117 (1998).
21. D. Forrest, K. James, Y. Yuzenkova, N. Zenkin, Single-peptide DNA-dependent RNA polymerase homologous to multi-subunit RNA polymerase. *Nat. Commun. 2017 81*. **8**, 1–8 (2017).
22. K. Lewis, Persister Cells. *http://dx.doi.org/10.1146/annurev.micro.112408.134306*. **64**, 357–372 (2010).
23. T. K. Wood, S. J. Knabel, B. W. Kwan, Bacterial persister cell formation and dormancy. *Appl. Environ. Microbiol.* **79**, 7116–7121 (2013).
24. K. R. Allison, M. P. Brynildsen, J. J. Collins, Metabolite-enabled eradication of bacterial persisters by aminoglycosides. *Nat. 2011 4737346*. **473**, 216–220 (2011).
25. K. P. Balanda, H. L. Macgillivray, Kurtosis: A Critical Review (2012), doi:10.1080/00031305.1988.10475539.
26. D. Kotliar, A. E. Lin, J. Logue, T. K. Hughes, N. M. Khoury, S. S. Raju, M. H. Wadsworth, H. Chen, J. R. Kurtz, B. Dighero-Kemp, Z. B. Bjornson, N. Mukherjee, B. A. Sellers, N. Tran, M. R. Bauer, G. C. Adams, R. Adams, J. L. Rinn, M. Melé, S. F. Schaffner, G. P. Nolan, K. G. Barnes, L. E. Hensley, D. R. McIlwain, A. K. Shalek, P. C. Sabeti, R. S. Bennett, Single-Cell Profiling of Ebola Virus Disease In Vivo Reveals Viral and Host Dynamics. *Cell*. **183**, 1383-1401.e19 (2020).
27. M. Y. Hein, J. S. Weissman, Functional single-cell genomics of human cytomegalovirus infection. *Nat. Biotechnol. 2021 403*. **40**, 391–401 (2021).
28. R. E. Osterhout, I. A. Figueroa, J. D. Keasling, A. P. Arkin, Global analysis of host response to induction of a latent bacteriophage. *BMC Microbiol.* **7**, 82 (2007).
29. X. Liu, H. Jiang, Z. Gu, J. W. Roberts, High-resolution view of bacteriophage lambda gene expression by ribosome profiling. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 11928–11933 (2013).
30. F. St-Pierre, D. Endy, Determination of cell fate selection during phage lambda infection. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20705–20710 (2008).
31. L. Zeng, S. O. Skinner, C. Zong, J. Sippy, M. Feiss, I. Golding, Decision Making at a Subcellular Level Determines the Outcome of Bacteriophage Infection. *Cell.* **141**, 682–691 (2010).

32. L. Imamovic, E. Ballesté, A. Martínez-Castillo, C. García-Aljaro, M. Muniesa, Heterogeneity in phage induction enables the survival of the lysogenic population. *Environ. Microbiol.* **18**, 957–969 (2016).
33. P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* **20**, 53–65 (1987).
34. S.-W. C. Cheng, D. L. Courtt, D. I. Friedman, Transcription Termination Signals in the nin Region of Bacteriophage Lambda: Identification of Rho-Dependent Termination Regions (1995).
35. S. R. Casjens, R. W. Hendrix, Bacteriophage lambda: Early pioneer and still relevant. *Virology*. **479**–**480**, 310–330 (2015).
36. P. Ma, H. M. Amemiya, L. L. He, S. J. Gandhi, R. Nicol, R. P. Bhattacharyya, C. S. Smillie, D. T. Hung, *bioRxiv*, in press, doi:10.1101/2022.08.01.502326.
37. J. B. Lugagne, H. Lin, M. J. Dunlop, DeLTA: Automated cell segmentation, tracking, and lineage reconstruction using deep learning. *PLOS Comput. Biol.* **16**, e1007673 (2020).
38. M. Büttner, Z. Miao, F. A. Wolf, S. A. Teichmann, F. J. Theis, A test metric for assessing single-cell RNA-seq batch correction. *Nat. Methods 2018 161*. **16**, 43–49 (2018).

Acknowledgments: We thank the Genomics Core Facility of the Lewis-Sigler Institute for Integrative Genomics performing the sequencing, the Adamson, Gitai, and Wingreen lab for their input.

Competing interests: BA was a member of a ThinkLab Advisory Board for, and holds equity in, Celsius Therapeutics. ZG is the founder of ArrePath. The remaining authors declare no competing interests.

Data and materials availability: Sequencing data will be deposited on GEO; all analysis and demultiplexing scripts will be uploaded onto https://github.com/brwaang55/m3seq_scripts. Any raw image files will be available upon request.

Supplementary Materials

Materials and Methods

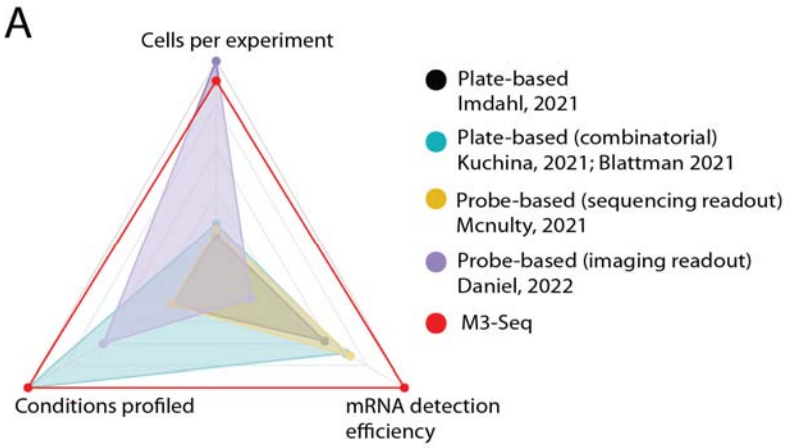
Supplementary Text

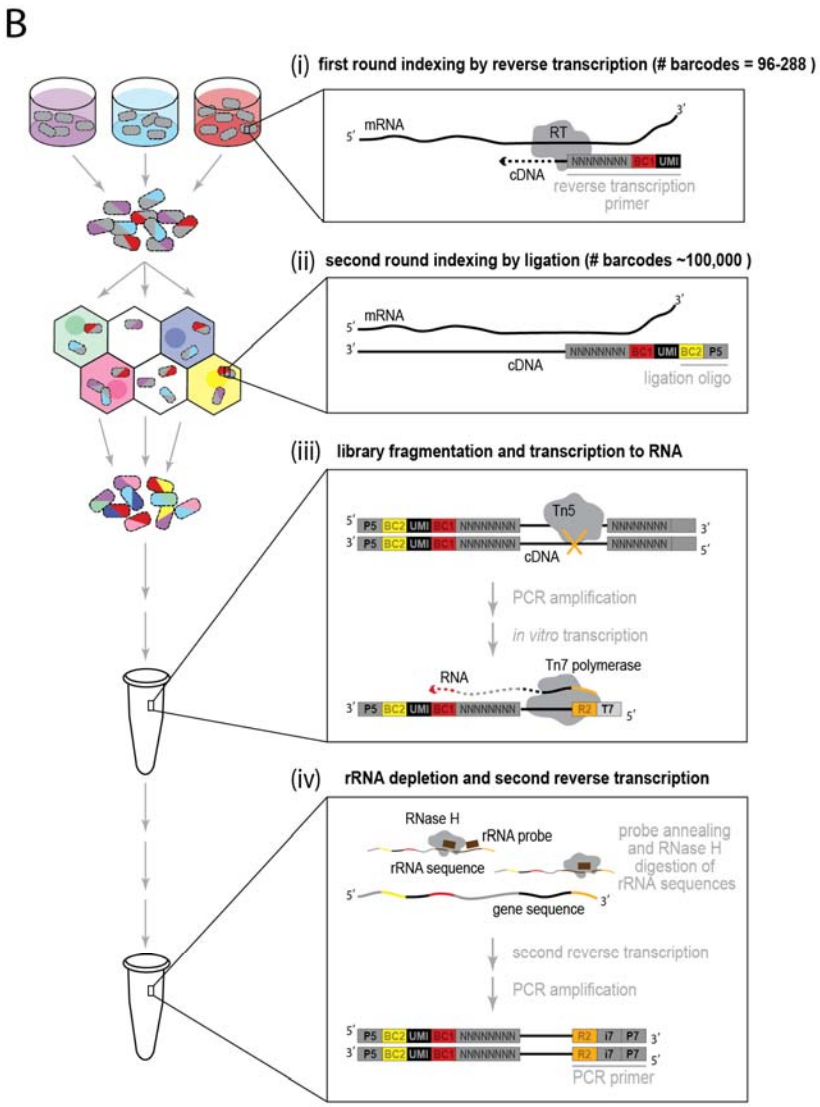
Figs. S1 to S8

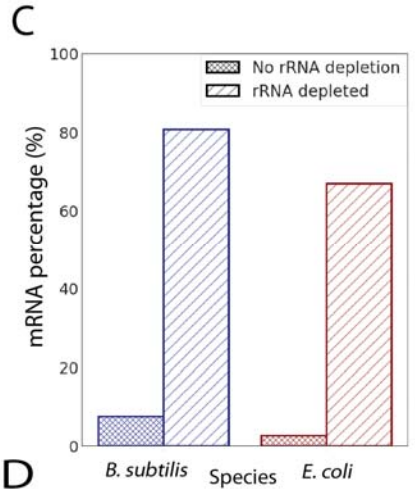
Tables S1 to S3

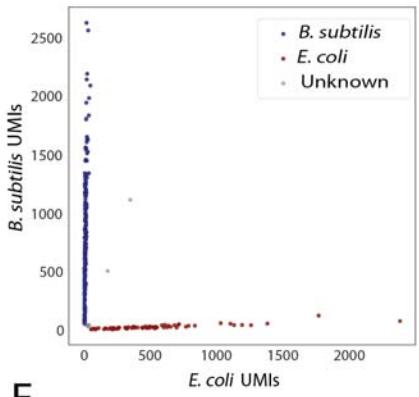
References (*1–38*)

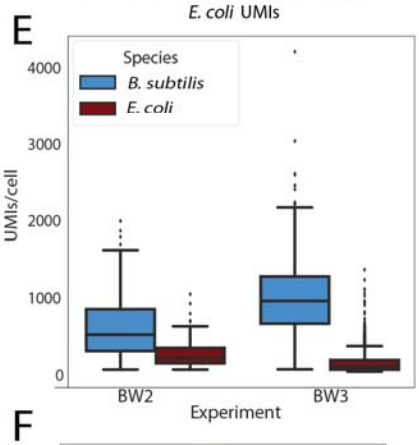
Movies 1 to S











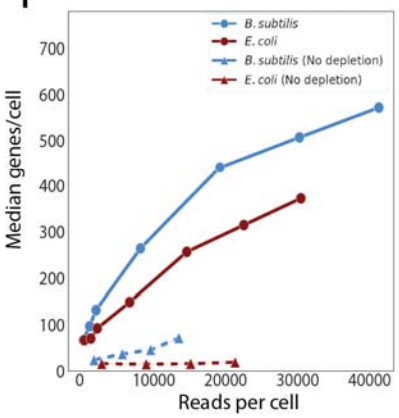


Fig. 1. M3-Seq Overview. A. Overview of existing methods and reported number of cells, conditions, and mRNA transcripts per cell that were detected in one experiment. Current approaches require sacrifices to be made on at least one of the three axes. M3-Seq simultaneously addresses all three limitations to maximize the chances of discovering novel bacterial subpopulations. **B.** Schematic of M3-Seq: (i) RNA molecules in fixed, permeabilized cells are barcoded and reverse transcribed in situ in a 96-well plate. (ii) Cells are collected after the first barcoding step, and loaded onto the 10x scATAC-Seq platform, where they undergo another round of barcoding in droplets. (iii) Fixed cells are then collected, lysed, and the cDNA library is prepared following [1] (in the final step, a T7-Nextera promoter is used to amplify the library). The full cDNA constructs are converted to RNA using T7 RNA polymerase. (iv) The pool of RNA is depleted of ribosomal reads by hybridizing the RNA library with specific DNA probes that are complementary to the rRNA sequences, and then reverse transcribed back to DNA using a primer at the 5' end (P5) of the construct. The final library is then constructed with another round of PCR using the P5 and P7 primers. **C.** "Barnyard" experiment where a mixture of *B. subtilis* (blue) and *E. coli* (red) are processed together using our method. Each point corresponds to a unique combination of plate and droplet barcodes, ideally representing a single cell. These "cells" were colored and assigned to a particular species if more than 90% of the corresponding UMIs belonged to that species. **D.** mRNA percentages in single-cell samples with and without rRNA depletion. Both *B. subtilis* (blue) and *E. coli* exhibit >20x enrichment of mRNA via rRNA depletion. **E.** mRNA UMI counts for the two species at exponential phase across two experiments. For *B. subtilis* (blue), we recover a median of 515 ± 245 (median absolute deviation) UMIs in the first experiment using M3-Seq ("BW2"), and 953± 310 in the second experiment using M3-Seq ("BW3"). For *E. coli* MG1655 (red), we recover 211 ± 85 (median absolute deviation) UMIs in "BW2", and 100±47 UMIs in "BW3". **F.** Median genes detected per *B. subtilis* (blue) or *E. coli* (red) cell as a function of number of total reads in the final sequencing sample. Depleting the sequencing library of rRNA reads (solid curves) enables an order of magnitude deeper detection of single bacterial transcriptomes than un-depleted (dashed curves) libraries.

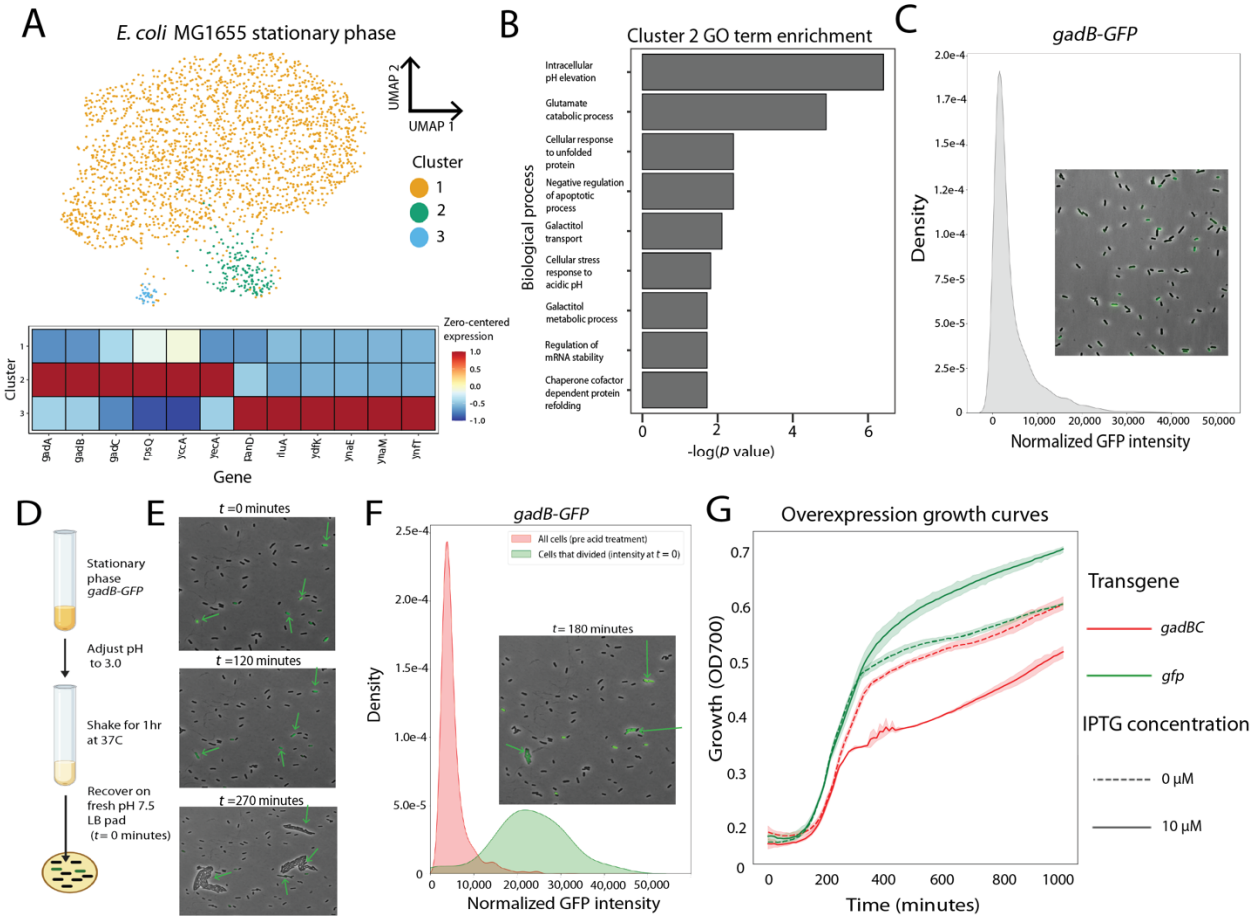


Fig. 2. M3-Seq reveals an acid-tolerant bet-hedging subpopulation in early stationary phase *E. coli*: A. UMAP embeddings of single *E. coli* MG1655 transcriptomes from cells at early stationary phase (OD=2.8) colored by cluster identity. Centered and normalized expression of marker genes for each cluster is displayed below the embedding. **B.** GO-term enrichment analysis for marker genes in cluster 2. The *p*-values for each biological process are log10 transformed and multiplied by -1 such that the strongest enriched biological processes have the highest score. **C.** Normalized fluorescence distribution of *E. coli* expressing P_{gadB}-GFP. The inset is a representative image of *E. coli* MG1655 transformed with P_{gadB}-GFP grown to OD=2.8. **D.** Schematic of acid-stress recovery assay. *E. coli* cells expressing P_{gadB}-GFP are grown to early stationary phase in LB media (OD = 2.8) and imaged. The pH of the culture is then adjusted to 3.0 using 12N HCl. The culture is allowed to shake for another hour, after which an aliquot of the cells is placed onto a fresh LB-agarose pad (t = 0) and imaged over 9 hours. **E.** Representative images of *E. coli* cells expressing P_{gadB}-GFP on a fresh LB-agarose pad after one hour of strong acid treatment. The arrows point to the viable cells that divided after an hour of strong acid stress. **F.** Result of the recovery assay. Kernel density estimates of the fluorescence distribution of all cells before acid treatment, and density estimates of the initial fluorescence distribution of cells (at t = 0) that survived the acid treatment and divided over the course of 8 hours. The inset is a representative composite overlay of the cells 180 minutes after the start of recovery. **G.** Growth curves of *E. coli* MG1655 transformed with a GFP overexpression plasmid (green), and MG1655 transformed with a *gadBC* overexpression plasmid (red). Strains were grown either with 10μM of IPTG (dotted curves), or no inducer (solid curves) for 1000 minutes. We note that induction of *gadBC* (red) using 10μM of inducer creates a growth defect in the

overexpression strain compared to the uninduced case, whereas inducing GFP at (green) with the same amount of inducer does not lead to any noticeable growth defect.

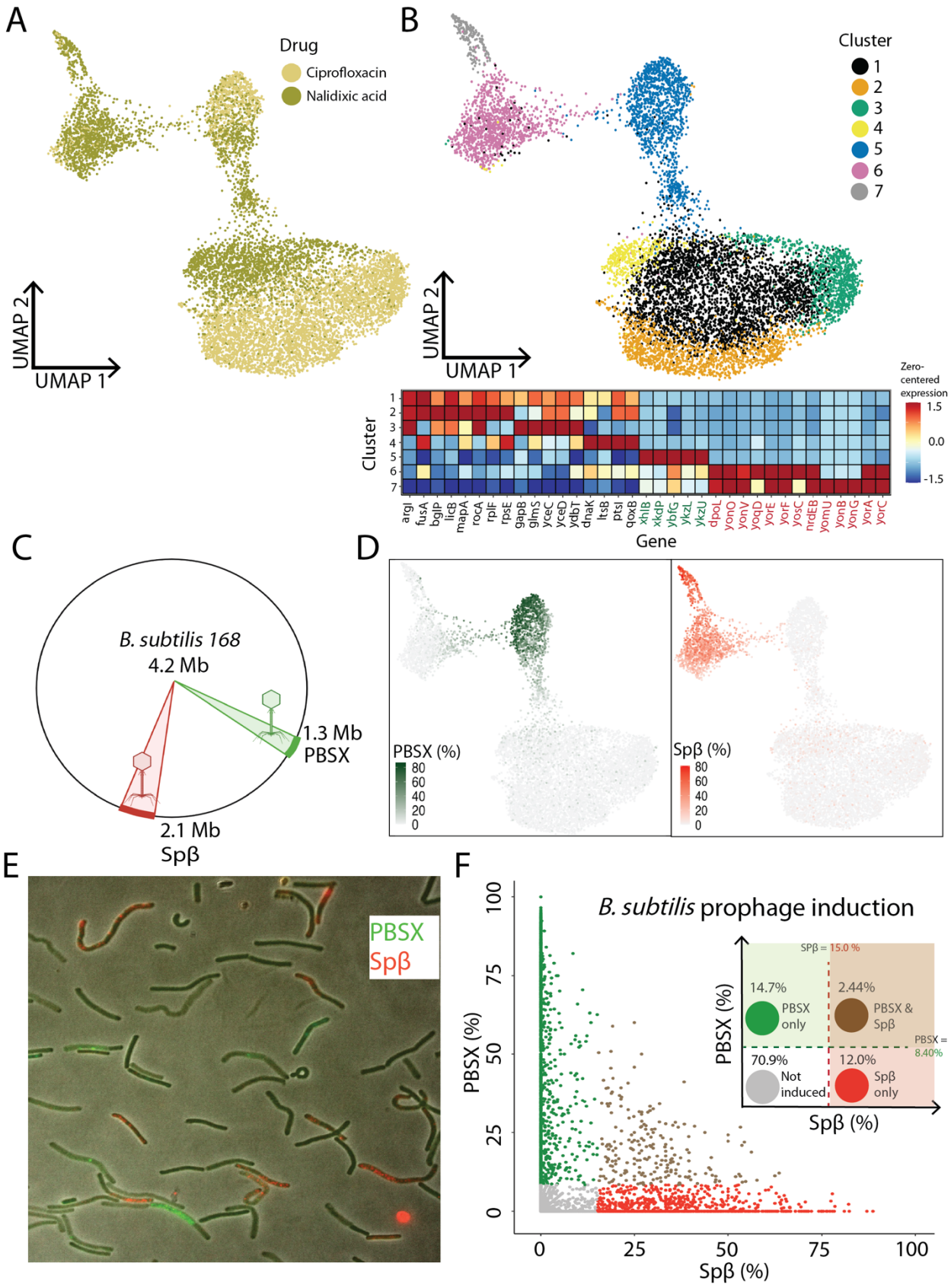


Fig. 3. M3-Seq enables detection of heterogeneous activation of prophages in *B. subtilis* in response to DNA damage: A. Uniform manifold approximation and projection (UMAP) embeddings of single *B. subtilis* transcriptomes from cells at exponential phase (OD = 0.3) treated for 90 minutes with DNA damaging antibiotics (ciprofloxacin and nalidixic acid). **B.** UMAP embedding of the treated cells colored by the assigned cluster, and heatmap of marker gene expression for each cluster. Six clusters were detected, with PBSX and Spβ prophage genes upregulated in clusters 3,4,5. **C.** Schematic of the location of the two prophages in the *B. subtilis* genome. **D.** UMAP data from **A**, **B** colored by relative expression of PBSX (Left) and Spβ (Right) prophage genes as a percentage of total single-cell transcriptome. **E.** Dual color smFISH image of *B. subtilis* treated with ciprofloxacin for 90 minutes. Probes hybridizing to PBSX genes were labeled with a green fluor, and probes hybridizing to Spβ genes were labeled with a red fluor (see *Methods*). **F.** Co-expression of Spβ (*x*-axis) and PBSX (*y*-axis) genes in each single cell. Cells were identified as inducing PBSX prophage if the relative expression of PBSX genes in each cell was greater than the 10^{th} percentile (8.4%) of PBSX prophage gene expression in cluster 5, and inducing Spβ if the relative expression of Spβ was greater than the 10^{th} percentile (15.0%) expression of Spβ genes in cluster 6. The induction probabilities of each prophage in the population are denoted as above, with the expected independent co-induction probability of 2.47% calculated by multiplying the observed PBSX and Spβ probabilities.

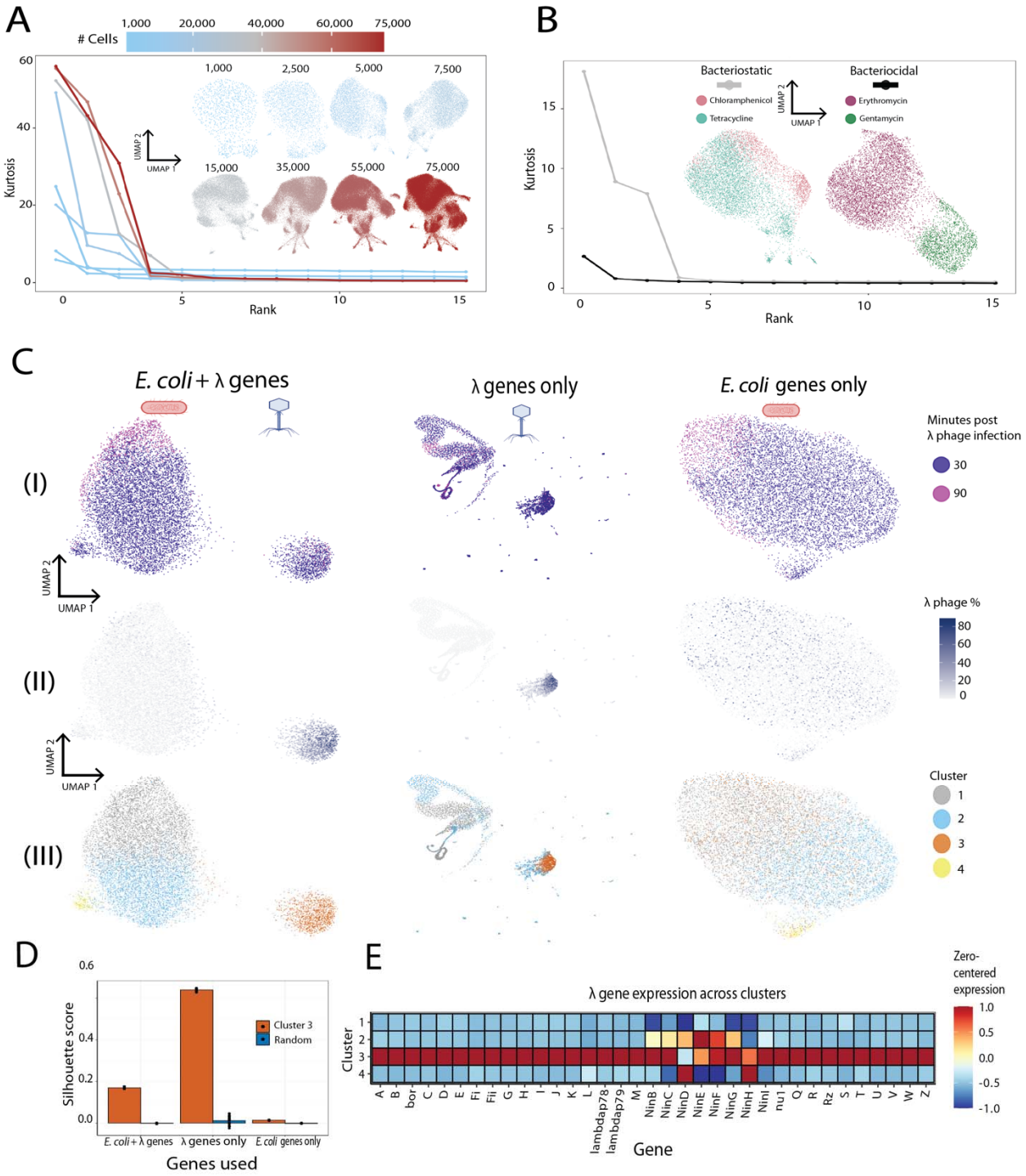


Fig. 4. M3-Seq provides the scale and technical capacity to discover new bacterial responses to antibiotics and phage infection: A. Cell rarefaction experiments of bacteriostatic antibiotic treated *E. coli* MG1655 cells. We sampled from a range of 1,000 to 75,000 cells, performed principal component analysis, and ranked the kurtosis of each principal component. Each curve (colored by the number of cells used) corresponds to the kurtosis of the 15 principal components with the highest kurtosis within that sample. Using the principal components, we also computed the UMAP embedding from each down-sampled matrix (inset, also colored by

cells used in that experiment). The top row of embeddings corresponds to the scale afforded by current technologies, while the bottom row represents the scale enabled by M3-Seq. **B.** UMAP embedding of cells treated with different modes of ribosome-inhibiting antibiotics (bacteriostatic versus bactericidal), along with the ranked kurtosis curves of the top 15 principal components for each mode of ribosome inhibitor. **C.** UMAP embeddings of λ phage infected *E. coli* MG1655. The first column are embeddings using only *E. coli* genes, the second are embeddings using only λ genes, and the third are embeddings using both *E. coli* and λ genes. In the first row (I) each point in the embeddings is colored by the timepoint at which the infected cells were sampled. In the second row (II), each point in the embeddings is colored by relative percentage of λ UMIs in that single cell. In the third row (III), cells are colored by cluster (which are assigned using the combined *E. coli* and λ genomes). Cluster 4 corresponds to infected cells undergoing late-stage lytic infection. **D.** Silhouette scores of the lytic cluster (cluster 3) and of a random sample of cells across each alignment. Silhouette scores were computed using the principal components of gene expression. We note that the silhouette score of the lytic cluster drastically decreases with the removal of the λ genes, indicating that the lysis signal is primarily driven from the phage. **E.** Heatmap of detected λ phage genes in each cluster. We note that while most of the known λ genes are upregulated in the late-stage lytic cluster (cluster 3), a particular set of λ genes (*ninB-H)* are also upregulated in a distinct cluster (cluster 2).

Materials and Methods

Experimental methods

Bacterial strains and growth conditions for BW1. *B. subtilis* 168 and *E. coli* (MG1655) were streaked out from a frozen glycerol stock onto an LB plate and grown overnight at 37°C. Following a night of growth, a single colony was picked and inoculated into 5 mL of LB broth and grown shaking at 250 RPM overnight at 37°C. The next morning, the overnight culture was diluted (1:100 for *E. coli*, 1:25 for *B. subtilis*) into multiple tubes 5 mL of fresh LB media in a 30 mL tube grown shaking at 250 RPM. Cells were harvested once at OD=0.6, and again 4 hours post dilution. The volume of cells was normalized so that 1 OD of cells was sampled and fixed at each step. Cells were immediately spun down for 5 minutes at 5,000 g at 4°C, resuspended in 4 mL of freshly made 4% formaldehyde. The resuspended cells were rotated overnight at 4°C until the next morning.

Bacterial strains and growth conditions for BW2. *B. subtilis* 168 and *E. coli* (MG1655) were streaked out from a frozen glycerol stock onto an LB plate and grown overnight at 37°C. Following a night of growth, a single colony was picked and inoculated into 5 mL of LB broth and grown shaking at 250 RPM overnight at 37°C. The next morning, the overnight culture was diluted (1:100 for *E. coli*, 1:25 for *B. subtilis*) into 35 mL of fresh LB media in a 250mL Erlenmyer flask and grown shaking at 250RPM. Upon reaching OD = 0.3, 5 mL of cells were split into tubes containing 2X the minimum inhibitory concentration of antibiotics (Ciprofloxacin or Cefazolin, 2 tubes), or no drug (2 tubes). The cells in the no drug tubes were sampled once at OD = 0.6, and again 120 minutes after the split. The cells in the tubes with drugs were sampled 20 minutes post-split (T20), and again at 120 minutes post-split (T360). The volume of cells was normalized so that 1 OD of cells was sampled and fixed at each step. Cells were immediately spun down for 5 minutes at 5,000g at 4°C, resuspended in 4 mL of freshly made 4% formaldehyde. The resuspended cells were rotated overnight at 4°C until the next morning.

Bacterial strains and growth conditions for BW3. *B. subtilis* 168 and *E. coli* (MG1655 and Nissle) were streaked out from a frozen glycerol stock onto an LB plate and grown overnight at 37°C. Following a night of growth, a single colony was picked and inoculated into 5 mL of LB broth and grown shaking at 250 RPM overnight at 37°C. The next morning, the overnight culture was diluted (1:100 for *E. coli*, 1:25 for *B. subtilis*) into 35 mL of fresh LB media in a 250mL Erlenmyer flask and grown shaking at 250RPM. Upon reaching OD = 0.3, 5 mL of cells were split into tubes containing 2X the minimum inhibitory concentration of antibiotics (Ciprofloxacin or Cefazolin), or no drug. The cells in the no drug tubes were sampled once at OD = 0.6, and again 360 minutes after the split. The cells in the tubes with drugs were sampled 90 minutes post-split (T90), and again at 360 minutes post-split (T360). The volume of cells was normalized so that 1 OD of cells was sampled and fixed at each step. Cells were immediately spun down for 5 minutes at 5,000g at 4°C, resuspended in 4 mL of freshly made 4% formaldehyde. The resuspended cells were rotated overnight at 4°C until the next morning.

Bacterial strains and growth conditions for BW4

B. subtilis 168, *E. coli* MG1655, and *P. aeruginosa* PA14 were streaked out from a frozen glycerol stock onto an LB plate and grown overnight at 37°C. Following a night of growth, a single colony was picked and inoculated into 5 mL of LB broth and grown shaking at 250 RPM overnight at 37°C. The next morning, the overnight culture was diluted (1:100 for *E. coli*, 1:25 for *B. subtilis*, 1:50 for *P. aeruginosa*) into 35 mL of fresh LB media in a 250mL Erlenmyer flask and grown shaking at 250 RPM. Upon reaching OD = 0.3, 4mL of cells were split into tubes containing 2X the minimum inhibitory concentration of antibiotics (gentamycin, tetracycline, erythromycin, chloramphenicol, cefazolin, cycloserine, ciprofloxacin, or nalidixic acid), λ phage at MOI=100 (for *E. coli*), or no drug. The cells in the tubes were sampled and had their absorbance read 90 minutes post-split (T90). The volume of cells was normalized so that 1 OD of cells was sampled and fixed at each step. Cells were then prepared in the same manner as with BW1,2,3.

Cell preparation. Following an overnight fixation, cells were prepared for scRNA-Seq as previously described[2]. Briefly, cells were first spun down for 10 minutes at 5,000g at 4°C. Cells were then resuspended in 0.5 mL of PBS-RI, which comprises of PBS + 0.01 U/µL SUPERase-IN RNase Inhibitor (Invitrogen, AM2696). Cells were spun

down again for 10 minutes at 5,000g at 4°C and resuspended in 300 µL of 1X PBS-RI, and 300 µL of 100% ethanol. Following the first permeabilization, cells were spun down for 8 minutes at 7,000g at 4°C, and washed twice with 200 µL of PBS-RI. After this final wash, cells were permeabilized by resuspension in 45 µL of 2.5mg/mL lysozyme solution dissolved in TEL-RI buffer, comprised of 100mM Tris pH 8.0, 50mM EDTA, 0.1U/µL SUPERase-IN RNase Inhibitor, and incubated at 30°C for 15 minutes. Cells were then spun down and washed twice in 100 µL of PBS-RI. After the final wash, cells were resuspended in 100 µL of 0.5X PBS-RI, and counted and examined with a hemocytometer (INCYTO DHC-S02).

Plate-based indexing. Fixed and permeabilized cells were split into wells of a 96 well plate, each containing a single indexing primer (2.5 µL/well, 20µM). To each well, we added 312,500 cells, 0.25 µL of Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific, EP0753), 0.25 µL of dNTPs at an original concentration of 10 µM (NEB, N0447L), 2.5µL of 5X Maxima H Minus Reverse Transcription Buffer, 0.125µL RNase-OUT (Thermo Fisher Scientific, 10777019), and PEG8000 to a final concentration of 7.5%, Tween-20 to a final concentration of 0.02%, and nuclease free water up to 10 µL . Reactions were then incubated as follows to perform first-round indexing by reverse transcription: 50°C for 10 minutes, 8°C for 12 seconds, 15°C for 45 seconds, 20°C for 45 seconds, 30°C for 40 seconds, 42°C for 6 minutes, 50°C for 50 minutes, and hold at 4° C. Samples were then pooled together and spun for 20 minutes at 7,000 g to isolate processed cells. Cells were then washed in 0.5 X PBS-RI and resuspended in 75 µL of 1X Ampligase buffer (Lucigen, A0102K). Pooled cells were counted and examined on the hemocytometer and diluted for loading onto the Chromium Controller (10x Genomics). The cell loading for each experiment indicated in Supplementary Table 2. Methods in this section were adapted from single-cell combinatorial fluidic indexing procedures.

Loading cells into microfluidic droplets. Cells were prepared for loading onto the Chromium scATAC platform v1.1 (10X Genomics 1000176). After counting, pooled cells were aliquoted with and mixed with 19µL 10X Ampligase Buffer, 2.3µL U/µL Ampligase (Lucigen A0102K), 1.5 µL Reducing agent B (10x Genomics 2000087), 2.3µL of 100 µM bridge oligo oDS025, and nuclease free water up to 75 µL. The mixture was kept on ice and loaded onto the Chromium Next GEM Chip H (10x Genomics, 1000162) with gel beads from the NextGem scATAC platform (10x Genomics, 1000176). To create emulsions, we followed the 10X Next GEM scATAC protocol[MOU1] (CG000209 Rev A). Briefly, the microfluidic chip was prepared by adding 70µL of cell mixture to wells in row 1, 50µL Next GEM scATAC beads to wells in row 2, and 40µL of partitioning oil to wells in row 3. Additionally, 50% glycerol was added to all unused lanes (40µL of 50% glycerol was added to unused lanes in row 3, 50µL to unused lanes in row 2, and 70µL to unused lanes in row 1). The chip runs on the Chromium Controller (10x Genomics) with the Next GEM Chip H program. This step partitions the cells and uniquely indexed gel beads into droplets. Methods in this section were adapted from single-cell combinatorial fluidic indexing procedures[1].

Droplet-based indexing . After transferring 100µL of each emulsion mixture to a clean reaction tube, second-round indexing was performed by ligation. Briefly, emulsions were incubated at 98°C for 30 seconds and 59°C for 2 minutes in 12 cycles. Emulsions were broken by adding 125µL Recovery Agent (10x Genomics) and pipetting up the hydrophobic phase . Cells were then reverse crosslinked and lysed by adding 10µL of 10X Lysis-T (250mM EDTA, 2M NaCl, 10% Triton X-100) and 4µL of proteinase K (NEB, P8107S) and incubating at 55°C for 1 hour. After lysis, DNA:RNA hybrid libraries were isolated with the following procedure: (1) 200µL of Dynabead Cleanup Mix, which consists of 182 µL Cleanup Buffer (10X Genomics, 2000088), 9µL Dynabeads MyOne Silane (Thermo Fisher Scientific, 37002D), 5µL Reducing Agent B (10X Genomics, no catalog no.) , and 5µL of nuclease free water, was added to each sample; (2) samples were mixed by pipetting (10x); (3) samples were incubated at room temperature for at least 10 minutes; (4) beads were isolated from samples using a magnetic stand and washed 2 times with 200µL 80% ethanol; (5) hybrid libraries were then eluted in 40µL of elution buffer (Qiagen, 19086).
Second strand synthesis. The eluted single stranded library was stripped of RNA by adding 2µL of RNAse H (NEB M0297L), 4 µL of 10X RNase H buffer (NEB B0297S) and incubating for 30 minutes at 37°C. The reaction was purified with a 1.8X SPRI, where the final eluate volume was 25 µL. To perform second strand synthesis, we used a modified version from [8], where we added 8µL of 5X Maxima H- Reverse Transcription Buffer, 4 µL of 10 µM dNTP's, 2.5 µL of Klenow Fragment (3' -> 5' exo -, NEB M0212L), 5 µL 50% PEG 8000, and 1.5 µL of 100 µM S^3 randomer (oBW140). The reaction was incubated at 37°C for 60 minutes, cleaned with a 1.8X SPRI, and eluted in 30 µL of Nuclease free water. The full length, double stranded library was amplified using PCR by adding 30 µL of 2X Q5 High Fidelity Master Mix (NEB M0492L), 0.4 µL of 100 µM oDS028, and 0.4 µL of 100 µM

oBW170. We amplified the library using the following protocol: 98°C for 30 seconds, 14 cycles of 98°C for 20 seconds, 65°C for 30 seconds, 72°C for 3 minutes. Following the first round of PCR, the reaction was cleaned twice once using a 1.2X SPRI reaction, each time eluting in 40 µL. This was to ensure primer dimers were properly removed. The resulting samples were the gene expression (GEX) libraries.

Library fragmentation using Tn5 transposase. We prepared the following 5X Tn5 reaction buffer: 50mM TAPS (Sigma, T9659-100G), 25mM MgCl2. We assembled i7-only transposomes according to established protocols[2]. Briefly, 10 µL of 100 µM oDS029 and 10 µL of 100 µM oDS30 were mixed and annealed using the following temperature program: 95°C for 2 minutes, followed by a 0.1°C/second ramp down to 4°C. Annealed oligos were then diluted with 80 µL of nuclease free water (final concentration, 10 µM) and, after 10 µL of 100% glycerol was added to the mixture, 8µL of the oligo-glycerol sample was mixed with 2µL of EZ-Tn5 (Lucigen, TNP92110) and incubated at 25°C for 40 minutes. The resulting i7 transposomes were stored at -20°C.

After construction, gene expression libraries were quantified (Qubit HS dsDNA kit) and fragmented in multiple reactions with the following components: 10ng gene expression library sample, 4 µL of 5x Tn5 buffer, 1µL of i7 transposome, and water up to 20uL. Reactions were incubated at 55°C for 10 minutes and then inactivated with 1µL of 20% SDS at 55°C for 10 minutes. Following inactivation , reactions were purified using a 1.2X SPRI reaction (elution volume, 25 µL). The resulting samples were the fragmented GEX libraries.

Second library amplification and *in vitro* transcription. Fragmented GEX libraries were mixed with 25 µL of 2X Q5 Master Mix, 0.4 µL of 100 µM oBW170, 0.4 µL of 100 µM oBW168 and amplified using the following protocol: 72°C for 3 minutes, 98°C for 30 seconds, 9 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds, a final incubation at 72°C for 5 minutes, and hold at 4°C. Resulting samples were purified with a 1.2X SPRI reaction (elution volume,40 µL) and converted into RNA by *in vitro* transcription. Briefly, 100ng of amplified libraries were mixed with 8µL 5X Transcription Buffer (Thermo Fisher Scientific, EP0112), 6µL of 2.5 mM rNTPs (NEB, N0466L), 1.5 µL of T7 RNA Polymerase (Thermo Fisher Scientific, EP0112), and 1µL of RNase-Out. Reactions weres incubated at 37°C for 2 hours, after which DNA templates were digested with 3µL DNase I (NEB, M0303L) and 3µL 10X DNase I buffer (NEB, B0303S) at 37°C for 15 minutes. RNA was then purified using a 2X SPRI reaction (elution volume, 25 µL). These samples were the *in vitro* transcribed GEX libraries.

Ribosomal RNA depletion. To enrich mRNA reads within a DNA library that was constructed using random priming, we developed an in-house approach to deplete ribosomal reads. Probes hybridizing to ribosomal RNA sequences of the bacterial species that we used of interest were designed using software from [9]. Multiple reactions containing the following were prepared (adapted from [9]): 500ng of in vitro transcribed RNA, 3µg of rRNA probes, 0.6µL 5M NaCl, 1.5 µL 1M Tris-HCl, and Nuclease free water up to 15 µL. In vitro transcribed RNA was hybridized to the DNA probes using the following temperature program: 95°C for 2 minutes, and 0.1°C/second ramp down to 25°C, 25°C for 5 minutes. Following rRNA probe hybridization, 6µL RNase mix (3µL of 10x RNase H buffer (NEB B0297), 2µL of Thermostable RNase H (NEB M0523S), 1µL of RNase H) were added to each tube. The reactions were incubated for 45 minutes at 50°C to digest the rRNA-DNA hybrids. Following rRNA digestion, the DNA probes were degraded by adding 3µL of 10x DNase I buffer, 3µL of DNase I, and incubating for 45 minutes at 37°C. The rRNA-depleted RNA library was purified with a 2x SPRI reaction and eluted in 25 µL of nuclease free water.

Final library prep. To recover the original DNA library from the depleted RNA library, we used the end specific P5 primer to ensure the full DNA molecule gets reverse transcribed. To each tube of purified RNA, the following reagents were added: 8µL Maxima H Minus Reverse Transcription Buffer, 1µL Maxima H Minus Reverse Transcriptase, 1µL RNase Out, 6µL 2.5mM dNTPs, 0.4 µL 100 µM oBW170, 0.2µL 100 µM oBW171. The reaction was incubated in the thermocycler with the following temperature program: 50°C for 10 minutes, 8°C for 12 seconds, 15°C for 45 seconds, 20°C for 45 seconds, 30°C for 40 seconds, 42°C for 6 minutes, 50°C for 18 minutes, and hold at 4°C.

Following reverse transcription, the reaction was purified with a 1.2X SPRI and eluted in 25 µL of nuclease free water. The reverse transcribed DNA reactions were then indexed using a final indexing PCR. For each reaction that needed to be indexed, 25 µL of reverse transcribed DNA was mixed with 25 µL Q5 High Fidelity Master Mix, 0.4 µL 100 µM oBW501, and 0.4 µL 100 µM of a unique P7 index primer. The reactions were amplified with the following temperature program: 98°C for 30 seconds, 9 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds, a final incubation at 72°C for 5 minutes, and hold at 4°C.

The resulting final DNA library was purified twice with a 0.8X SPRI, and then quality controlled on the Qubit and Bioanalyzer. We also checked the concentration and sequencability of the DNA library using qPCR with the following primers: oBW170/oBW176, oBW141/oBW176. We note that this final qPCR step is essential as it lets us know if we need to redo some steps of the procedure by checking for the percentage of the sequencable reads in the library. Following the final qPCR, libraries were diluted to 5nM, and sequenced with the NovaSeq SP 100 cycle kit (Illumina 20028401) using the following read structure: 26bp Read 1, 30bp i5 index, 8bp i7 index, 74bp Read2.

Fluorescent in-situ hybridization (FISH). To enable cost effective detection of multiple different RNAs in the same cells, we closely followed the framework used in [10–12]. Briefly, multiple primary probes hybridizing to an mRNA of interest are first designed. These probes contain a constant 20nt flanking sequence that allows for hybridization of a fluorescent secondary probe. This allows us to avoid the cost of ordering multiple fluorescent primary probes to tile our gene of interest.

Primary probes for fluorescent in-situ hybridization for RNA sequences of interest were designed using [9]. For each RNA transcript of interest, we designed at least 10 different probes hybridizing to different regions of that transcript. A 20nt sequence was added to the 3' end of each probe to allow for hybridization of the fluorescent readout probes. Primary probes for each gene were mixed at an equimolar ratio such that the final concentration of DNA molecules was 100 µM. Fluorescent readout probes were ordered following [12].

Cells in each condition of interest were grown, fixed, and permeabilized as described above. After the permeabilization step, cells were washed and resuspended in 600 µL primary hybridization buffer (40% Formamide (Thermo Fisher Scientific 15515026), 2X SSC (Invitrogen AM9673)) and aliquoted into 1.5 mL tubes. 1µL of 100 µM primary probe mix was added to each tube and hybridized overnight at 30°C in the dark. The next morning, cells were spun down at 7,000g for 8 minutes and resuspended in 200 µL wash buffer (30% Formamide (Thermo Fisher Scientific 15515026), 2X SSC (Invitrogen AM9673)). Cells were spun down for 8 minutes at 7,000g, resuspended again in 200 µL wash buffer, and incubated in the dark at room temperature for 30 minutes. Cells were then spun down at 7,000g for 8 minutes and resuspended in 100 µL secondary hybridization buffer (10% Formamide, 2X SSC, 10% Ficoll PM-400 (Sigma-Aldrich F5415-25 mL)). 0.5 µL of each 100 µM readout probe was added to the tubes, and incubated for 1 hour at 34°C. Following secondary hybridization, cells were spun down at 7,000g, and resuspended in wash buffer with 10 µg/mL DAPI (Thermo Fisher Scientific D1306). Cells were incubated for 20 minutes at room temperature, spun down at 7,000g, and resuspended in 100 µL of 2X SSC.

Cells were imaged on 1% agarose pads made with filtered water on a Nikon TiE microscope with a Plan Apo 100X objective, and Hanamatsu ORCAFlash4.0 camera. Images were analyzed using FIJI software.

Acid tolerance assay. A 25mL culture of *E. coli* (MG1655) or *E. coli* (MG1655 *ΔgadAΔgadBΔgadC)* was first grown to OD = 0.3 in a 125mL flask shaking at 250 RPM 37°C. After reaching OD = 0.3, the cultures were split in aliquots of 5mL to culture tubes and placed back onto the shaker to grow for another 6 hours until OD = 2.8. Cultures were then acidified to pH 3.0 using 12N HCl and returned to the shaker. 10 µL of the cultures was sampled at intermittent timepoints and serial diluted for CFU counting.

Acid recovery assay. A 25mL culture of *E. coli* (MG1655) transformed with P_{gadB}-GFP was first grown to OD = 0.3 in a 125mL flask shaking at 250 RPM 37°C. After reaching OD = 0.3, the cultures were split in aliquots of 5mL to culture tubes and placed back onto the shaker to grow for another 6 hours until OD = 2.8. At this point, 1 µL of the culture was imaged on an 1% agarose pad made with LB media to understand the distribution of GFP fluorescence in single cells. Cultures were then acidified to pH 3.0 using 12N HCl and returned to the shaker. Following an hour of acid stress, 1 µL of the acidified culture was transferred onto an 1% agarose pad made with fresh LB media to assess viability. Cells were imaged every 15 minutes to track and assess growth over time.

The resulting movies were analyzed by first segmenting the cells using Delta (*37*), and then using custom python scripts to extract the fluorescence distribution and assess viability. A cell was considered viable if it underwent a single division during the 8 hour imaging period.

Bulk RNA-Seq Library preparation. *E. coli* (MG1655) was grown as described above to OD = 0.6. 2mL of cells were spun down at 5,000g for 10 minutes, resuspended in 45 µL of 2.5mg/mL lysozyme solution (described above), and incubated at 37°C for 15 minutes. RNA was purified using the Qiagen RNeasy Mini Kit (Qiagen 74104) where the final eluate volume was 30 µL. The RNA was reverse transcribed by adding 5 µL Maxima H Minus Reverse Transcription Buffer, 0.5 µL Maxima H Minus Reverse Transcriptase, 0.5 µL RNase Out, 4 µL 2.5mM dNTPs, 0.4

µL 100 µM oBW121 and incubating using the following temperature program: 50°C for 10 minutes, 8°C for 12 seconds, 15°C for 45 seconds, 20°C for 45 seconds, 30°C for 40 seconds, 42°C for 6 minutes, 50°C for 50 minutes, and hold at 4°C.
Following reverse transcription, RNA was stripped from the reverse transcribed DNA by adding 2µL of RNase H and incubating the mixture at 37°C for another 30 minutes. The library was purified using a 1.2X SPRI and eluted in 25 µL nuclease free water. Second strand synthesis, PCR, and tagmentation were performed as described above. The first PCR was performed using primer pairs oBW154 and oDS28. Following tagmentation, the library was amplified 8 cycles as described above using oBW154 and oBW168. This library was used to test for different rRNA depletion strategies.

Cas9 based rRNA depletion. To test Cas9 based rRNA depletion, we first synthesized a pool of guide RNAs which cleave at different sites of the 5S, 16S, and 23S ribosomal RNAs. DNA templates for the guide RNAs were designed using software from [20]. The DNA templates were purchased as a pool from IDT, and amplified with PCR by first annealing at a 1:1 equimolar ratio mixing 1µL DNA template, 0.4 µL 100 µM oBW138, 0.4 µL 100 µM oBW139, 10 µL nuclease free water, 12.5 µL 2X Q5 High Fidelity Master Mix, and using the following temperature program: 98°C for 30 seconds, 35 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 45 seconds, a final incubation at 72°C for 5 minutes, and hold at 4°C. Following PCR, the DNA templates were purified using a 1.2X SPRI and used for in vitro transcription. Guide RNAs were transcribed using the NEB HiScribe kit (NEB E2040S) by mixing 100ng of DNA template, 2µL of 10X reaction buffer, 2µL 100mM ATP, 2µL 100mM GTP, 2µL 100mM CTP, 2µL 100mM UTP, 2µL T7 RNA Polymerase Mix, nuclease free water up to 20 µL, and incubated overnight at 37°C.

Following an overnight in vitro transcription, DNA template was digested by adding 3µL 10X DNase buffer, 2µL DNase I, and incubating for an additional 15 minutes at 37°C. Guide RNAs were purified using a 2X SPRI reaction and checked for purity by running on a 15% TBE-Urea Gel (Invitrogen EC6885BOX). Guide RNA concentration was quantified using the Broad Range RNA Qubit kit (Thermo Fisher Scientific Q10210).

To perform Cas9 based depletion in our most optimized condition, 2 ng of library was mixed with 1.5 µL NEB 3.1 buffer, and sgRNA and NEB cas9 at a 20,000:3,000:1 ratio of sgRNA:Cas9: DNA. The reaction was incubated at 37°C for 2 hours after which Cas9 was stripped from the DNA by adding in 1µL Proteinase K, 1µL 10% SDS, and incubating for 15 minutes at 50°C. The DNA library was purified with a 1.2X SPRI, eluted in 25 µL nuclease free water, and mixed with 25 µL 2X Q5 High Fidelity Master Mix, 0.4 µL 100 µM oBW170, and 0.4 µL 100 µM of a unique P7 index primer. The reactions were amplified with the following temperature program: 98°C for 30 seconds, 12 cycles of 98°C for 10 seconds, 65°C for 30 seconds, 72°C for 30 seconds, a final incubation at 72°C for 5 minutes, and hold at 4°C. Libraries were sequenced on the MiSeq Reagent Kit v2 (300 cycles) (Illumina MS-102-2002) using the following read structure: 26bp Read 1, 30bp i5 index, 8bp i7 index, 100bp Read2.

Quantifying cell loading in the 10X Microfluidic system. To quantify if single bacterial cells could be loaded into the 10X Microfluidic system, we first fixed 2mL of *E. coli* MG1655 cells grown to OD=0.4 overnight in 4 mL of 4% formaldehyde. Cells were prepared as described above up to after the first wash following permeabilization. Following the first wash, cells were incubated in 50 µL of 5 µM Sytox Green (Thermo Fisher Scientific S7020) for 15 minutes. After the incubation, cells were washed twice in 100 µL of PBS-RI, and then resuspended in 100 µL of 0.5X PBS-RI. Cells were counted, and then loaded onto the 10X Microfluidic system using the Chip A 5' kit.

Following droplet generation, 5uL of the mixture was transferred onto a glass coverslip, and imaged on a Nikon TiE microscope with a Plan Apo 20X objective, and Hanamatsu ORCAFlash4.0 camera. Cells in each droplet were then manually counted.

Computational Methods

Data Preprocessing. Raw basecalls were retrieved from the NovaSeq, and processed with a custom version of picard tools v2.19.2 following the pipeline described in [1]. Reads were aligned to a combinations of one or more of *B. subtilis* 168, *E. coli* MG1655, and *E. coli* Nissle genomes using STAR v2.76[14], and annotated with featureCounts v2.0.0[15]. Reads were filtered such that all the reads used for downstream analysis have mapQ score > 1, and mapped lengths greater than 20bp. Annotated and filtered reads were loaded into Python 3.7.6 where custom code was written to assign non-rRNA reads to combinations of droplet and plate barcodes in pandas.

After assigning reads to barcode combinations, we filtered out droplets barcodes in which a given droplet barcode had more than 8 associated plate barcodes, which corresponds to a barcode collision rate greater than 25%. We then split barcode combinations by condition (plate barcodes), and then performed another filtering step using the knee method for each condition[16]. We note that this step is important because bacteria in different conditions

have different amounts of mean mRNA expression. After the last filtering step, a cell/gene matrix was made where the entries of the matrix are the number of UMI's that we measured for that gene in a particular cell.

Single-Cell Analysis. Metrics for the scRNA-Seq results were compiled and plotted using custom scripts in Python 3.7.6. Downstream analysis of single cell data was performed using pipelines detailed in Seurat v4.0.3[17]. Data was first preprocessed filtering out genes that were expressed in less than 10 cells, and cells that expressed less than 10 UMI's. The data was then normalized by dividing the UMI counts in each cell by the total number of UMI's measured in that cell, multiplying by a scale factor of 100, adding a count of 1 to each entry, and then log-normalizing the scaled values. The normalized expression data was then scaled to have mean 0 and unit variance, and dimension reduced using principal component analysis. When necessary, the kurtosis of each principal component was computed by taking the matrix of cells by principal component coordinates, and then calling the "kurtosis" function.

Following principal component analysis, we computed a uniform manifold approximation representation[18] and a shared neighbor graph (SNN) using the first 10 principal components. We performed graph-based clustering on the shared neighbor graph to identify clusters of gene expression programs using the Louvain algorithm (algorithm 3 in Seurat 4.0.3)[19]. Marker genes for each cluster were computed using the Wilcoxon Rank-sum test. Further data analysis and plotting was performed using custom scripts in R.

Geneset enrichment analyses were performed using topGo (2.48.0). Briefly, marker genes for each cluster were first obtained using the FindMarkers function in Seurat. The list was filtered by taking genes with p value < 0.05. This list was then split into genes that were upregulated in the cluster, and genes that were downregulated. The two lists of genes were then used for biological process term enrichment using Fisher's exact test.

To compute silhouette scores, we took the PCA matrix and cluster outputs from Seurat, and used the silhouette score function from the KBET package(*38*).

Comparison with Bulk RNA-Seq. Bulk RNA-Seq data for exponentially growing *E. coli* was downloaded from (*8*) (GEO accession number GSE141018). Raw reads from the bulk data were aligned to the *E. coli* MG1655 genome and annotated as described above. Single-cell and bulk transcriptomes of exponential growing *E. coli* were compared by computing the Pearson correlation of $\log_{10}$ transcripts per million (TPM) of each gene between the two measurements. Genes with 0 counts in either dataset were first omitted. TPM for each gene in single cell data was then computed by adding a pseudocount of 1 to each gene, summing over the UMI counts for that gene across all cells, normalizing by gene length, and dividing by the sum of length normalized counts. TPM for bulk measurements were computed as previously described. The TPMs of the bulk and single-cell datasets were $\log_{10}$ transformed and used for plotting and correlation measurements.