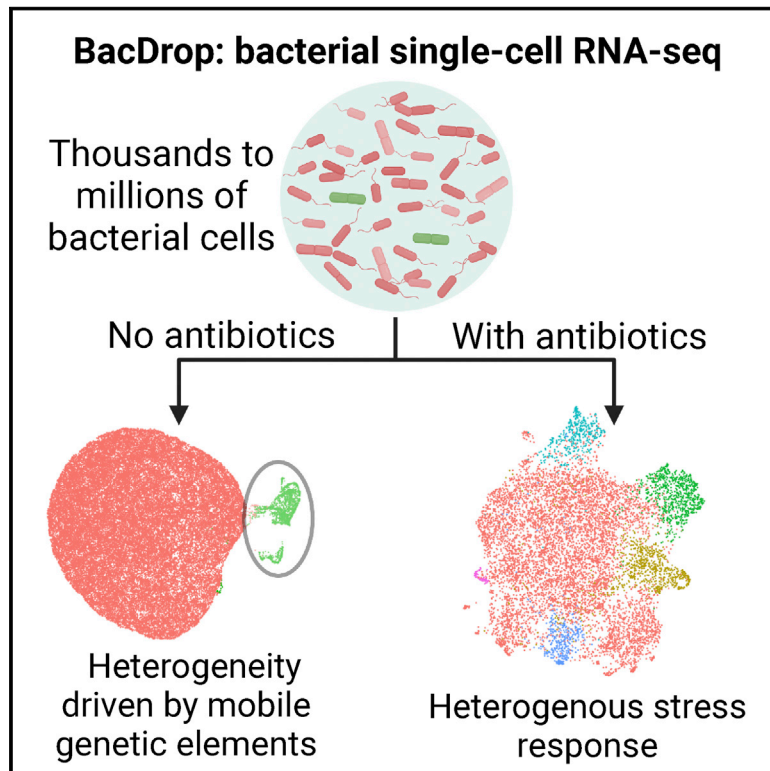


Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states

Graphical abstract



Authors

Peijun Ma, Haley M. Amemiya, Lorrie L. He, ..., Roby P. Bhattacharyya, Christopher S. Smillie, Deborah T. Hung

Correspondence

hung@molbio.mgh.harvard.edu

In brief

BacDrop is a droplet-based technology for single-cell RNA-seq in bacteria that can be scaled to millions of bacterial cells or hundreds of samples and was used to elucidate transcriptionally distinct bacterial subpopulations associated with varying phenotypic outcomes linked to antibiotic resistance and persistence.

Highlights

- We developed BacDrop, a droplet-based technology for bacterial single-cell RNA-seq
- BacDrop revealed a rare bacterial cellular state driven by expressions of mobile genetic elements
- BacDrop revealed distinct cellular states resulting from antibiotic perturbation
- Cell states associated with the emergence of antibiotic resistance or persistence were identified

Article

Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states

Peijun Ma,^{1,2,3,4} Haley M. Amemiya,^{1,2,3,4} Lorrie L. He,¹ Shivam J. Gandhi,^{1,3,6} Robert Nicol,¹ Roby P. Bhattacharyya,^{1,5} Christopher S. Smillie,^{1,3,6} and Deborah T. Hung^{1,2,3,4,7,*}

¹The Broad Institute of MIT and Harvard, Cambridge, MA, USA

²Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA

³Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA

⁴Department of Genetics, Harvard Medical School, Boston, MA, USA

⁵Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA, USA

⁶Department of Medicine, Harvard Medical School, Boston, MA, USA

⁷Lead contact

*Correspondence: hung@molbio.mgh.harvard.edu

<https://doi.org/10.1016/j.cell.2023.01.002>

SUMMARY

We introduce BacDrop, a highly scalable technology for bacterial single-cell RNA sequencing that has overcome many challenges hindering the development of scRNA-seq in bacteria. BacDrop can be applied to thousands to millions of cells from both gram-negative and gram-positive species. It features universal ribosomal RNA depletion and combinatorial barcodes that enable multiplexing and massively parallel sequencing. We applied BacDrop to study *Klebsiella pneumoniae* clinical isolates and to elucidate their heterogeneous responses to antibiotic stress. In an unperturbed population presumed to be homogeneous, we found within-population heterogeneity largely driven by the expression of mobile genetic elements that promote the evolution of antibiotic resistance. Under antibiotic perturbation, BacDrop revealed transcriptionally distinct subpopulations associated with different phenotypic outcomes including antibiotic persistence. BacDrop thus can capture cellular states that cannot be detected by bulk RNA-seq, which will unlock new microbiological insights into bacterial responses to perturbations and larger bacterial communities such as the microbiome.

INTRODUCTION

Single-cell RNA sequencing (scRNA-seq) has led to important discoveries in mammalian systems, resulting in great appreciation of the transcriptional heterogeneity of cell types and cell states.^{1–8} Although heterogeneity is essential for bacterial communities, bacterial scRNA-seq is underdeveloped due to longstanding technical challenges, including bacterial lysis, the absence of polyadenylated tails on messenger RNA (mRNA), and the paucity of mRNA molecules in a single bacterial cell,⁹ which collectively lead to degraded transcriptome coverage and quality relative to mammalian systems.

Recently, several bacterial scRNA-seq approaches have been described, including plate-based methods such as microSPLIT, PETRI-seq, and MATQ-seq,^{10–13} and probe-based methods such as par-seqFISH.^{14,15} These plate-based scRNA-seq technologies enable genome-wide scRNA-seq, but they have to date been limited by the numbers of cells (scale) that can be studied. Additionally, sequencing reads of these plate-based methods are dominated (>90%) by the overwhelmingly abun-

dant rRNA, wasting a large portion of the sequencing investment. In contrast, the probe-based methods avoid the problem of rRNA and have improved scale. However, they require prior knowledge of the genome(s) of interest to enable probe design and generation, limiting the numbers of genes that can be queried. Due to these limitations, studies using these previously reported methods have mainly focused on between-population heterogeneity, with a focus on proof of principle, whereas within-population heterogeneity, which requires the characterization of large number of cells from a single population at genome level, has not been extensively described.

One of the fundamental lessons from eukaryotic scRNA-seq is that because of the co-variation structure of gene expression within and between cells, profiling larger number of cells shallowly is a more favorable experimental design than profiling a small number of cells deeply and can better recover the statistical properties of cell populations and gene programs.¹⁶ This is especially the case when molecular techniques limit the profile of each cell to a relatively low fraction of transcripts, randomly sampled from all transcripts in the cell. Combined with powerful

algorithms, numerous scRNA-seq studies in eukaryotic systems have demonstrated that novel cell types, states, dynamic trajectories, gene programs, and even features like spatial locations and cell interactions can be comprehensively recovered by such massively parallel methods, whereas per-cell coverage plays a less important role.^{1–4,6,16–26} Experimental and computational analyses have also demonstrated that low coverage and/or low sequencing depth is sufficient for effective cell clustering, detection of rare populations, identification of biomarkers when analyzing a large number of cells, and conducting genetic “Perturb-seq” screens.^{25–29} The underlying principle of this observation is that scRNA-seq remains a sampling strategy, with only a portion of cells sampled from a population (scale) and a portion of RNA molecules sampled from a cell (coverage).¹⁶ Although higher coverage enables deep profiling of some cells and genes, it lacks the statistical power to recapitulate the phenotypic landscape of the population when the scale is limited. In contrast, because gene expression is structured due to shared regulatory mechanism, when a large number of cells are analyzed, even lower coverage of each cell allows the recovery of shared patterns such as clustering of similar cells to identify distinct subpopulations or ordering cells by pseudotime to recover temporal trajectories.^{16,26,30} Thus, to capture the phenotypic landscape of the population and its statistical properties and distributions, rather than what is happening to one particular gene in any one particular cell, scale is more critical.

This same biological and statistical principle is more important in single-cell transcriptional analysis of bacteria due to the inherent paucity of mRNA molecules in a single bacterial cell.⁹ Additionally, if applied to samples with high complexity or diversity, large numbers of cells are critical to ensure the capture of enough cells from any individual rare subpopulation or species, such as persister or heteroresistant subpopulations^{9,31–33} or the microbiome, mirroring the requirements in large-scale mammalian projects such as the Human Cell Atlas.³⁴ Therefore, methods for increasing bacterial scRNA-seq scale will be essential for studies in microbial systems.

Here, we report the development and application of BacDrop, a droplet-based genome-wide massively parallel bacterial scRNA-seq technology. BacDrop has the flexibility to investigate a wide range of numbers of cells in one experiment, from thousands to millions of single bacterial cells, without requiring prior knowledge of the genome or probe design. BacDrop also includes a universal and efficient rRNA-depletion step, reducing sequencing costs by at least 10-fold while simultaneously increasing information content. Furthermore, we demonstrated that BacDrop works on a variety of bacterial species, including gram-negative *Escherichia coli*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and gram-positive bacterium *Enterococcus faecium*.

In this study, we applied BacDrop to assess any within-population heterogeneity of *K. pneumoniae* and characterize its responses to antibiotics. *K. pneumoniae* is one of the leading threats in the antibiotic resistance crisis, especially resistance to carbapenems, the last-resort antibiotic used to treat the most resistant infections.³⁵ Under stable and dynamic conditions (i.e., without and with antibiotic perturbations), we identified important features of heterogeneity that have been

masked by bulk analysis. In the absence of antibiotic perturbation, we found within-population heterogeneity driven predominantly by mobile genetic elements (MGEs), which promote the evolution of antibiotic resistance, thus demonstrating novel subpopulation structure and function in a previously presumed homogeneous population. After antibiotic perturbation, BacDrop revealed transcriptionally distinct subpopulations that are associated with different phenotypic outcomes including decreased antibiotic efficacy and persister formation. With this demonstration of the power of BacDrop to illuminate heterogeneity both in static bacterial populations and their dynamic responses to perturbations, we propose that BacDrop has the potential to transform our understanding of bacterial survival, adaptation, and evolution, with numerous potential applications including the characterization of complex communities.

RESULTS

BacDrop: A bacterial droplet-based massively parallel scRNA-seq technology

We developed BacDrop based on droplet-based microfluidics technology, which has the advantage of achieving much higher scale than plate-based technologies, leveraging the 10x Genomics platform, a reliable and commercially available platform for scRNA-seq.²⁴ Two of the unique features of bacteria compared with mammalian cells are the challenge of lysing the bacterial cell wall, which requires harsher lysis conditions while preserving RNA integrity, and the absence of mRNA polyadenylated tails, thus requiring an alternative approach to isolate mRNA (~5%) from the vastly more abundant rRNA (~95%). To overcome these problems, we adapted a previously reported cell fixation and permeabilization protocol to avoid cell lysis prior to droplet encapsulation^{10,12} and implemented universal rRNA and genomic DNA (gDNA) depletion steps within the permeabilized, fixed cells using RNase H and DNase I, respectively (Figures 1A and S1A).

Droplet-based single-cell approaches require Poisson loading at mean droplet occupancy values (λ) far less than 1 to ensure that multiple cells are not encapsulated within a single droplet¹ and thus do not share the same barcode, limiting numbers of cells that can be loaded in each microfluidic channel. We overcame this problem by utilizing a multistep barcoding strategy, wherein cells are uniquely identified using the combination of two barcodes (Figure 1): a “plate barcode” (CB1) as a pre-indexing step corresponding to one of a possible 384 different barcodes achieved via reverse transcription (RT) and a second “droplet barcode” (CB2) that uniquely identifies each droplet by utilizing the commercially available single-cell kit from 10x Genomics (see [method details](#)) and loaded 3–6 cells per droplet. Together, CB1 and CB2 constitute the cell barcode. This pre-indexing strategy has recently been successfully applied to mammalian droplet-based scRNA-seq.³⁶ By pre-indexing, we were able to load each droplet with multiple cells, increasing the scale significantly from a λ of ~0.3 for conventional droplet scRNA-seq experiments to a λ of >1 for our approach with multiple cells in each drop. Additionally, template switching is less efficient for second strand cDNA synthesis from bacterial than mammalian mRNA due to the lack of proper modifications at

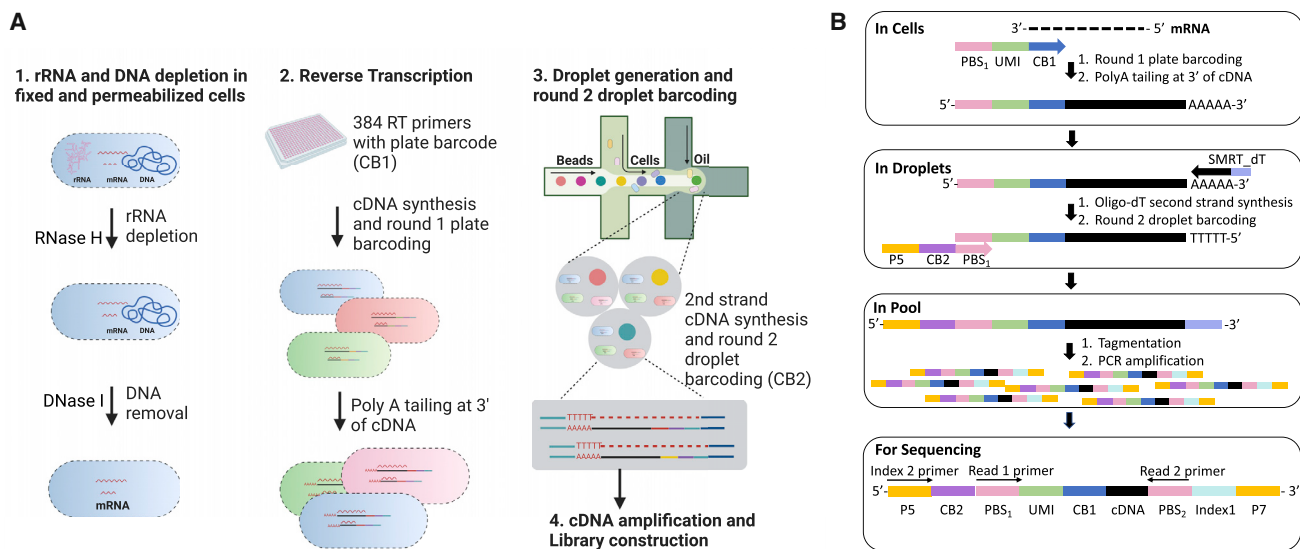


Figure 1. BacDrop: a bacterial droplet-based massively parallel scRNA-seq technology

(A) BacDrop workflow. Following cell fixation and permeabilization, rRNA and gDNA is depleted from cells in bulk. Then, CB1 and unique molecular identifiers (UMIs) are added to the 5' end of cDNA via RT reactions (round 1 plate barcoding) in 96- or 384-well plates. After round 1 plate barcoding, all cells are pooled, and cDNA is polyadenylated at the 3' end using terminal transferase, followed by droplet generation and round 2 droplet barcoding. The 3' poly(A) tail of cDNA enables second strand synthesis using oligo-dT primers. Round 2 droplet barcoding is achieved via second strand cDNA synthesis and 4 capturing cycles by barcoded primers (on 10x gel beads) in droplets. The successfully captured cDNA contains UMIs, CB1 and CB2, as well as adaptor sequences at both 5' and 3' end. Each cell is identified by a combination of CB1 and CB2.

(B) Scheme of two rounds of cell barcoding and library construction. The RT primer is composed of a partial primer binding sequence (PBS) for Illumina sequencing, UMI (8 bp), CB1 (13 bp), and 6-bp random sequence for RT priming. The 3' of cDNA was polyadenylated in cells after RT. A specific number of cells (thousands to millions) were then encapsulated, followed by 2nd strand cDNA synthesis and round 2 barcoding in droplets to attach CB2 to the double-stranded cDNA. An adaptor (SMRT) sequence was added to the polyA end of the cDNA to enable cDNA amplification. After cDNA purification and amplification, tagmentation and PCR enrichment was performed to generate Illumina sequencing libraries.

the 5' end of bacterial mRNA.^{37,38} We thus used terminal transferase (TdT) to append poly(A) tails to the 3' end of cDNA to facilitate second strand cDNA synthesis inside droplets.

Determining the technical performance of BacDrop

To characterize the performance of BacDrop, we tested the efficiency of rRNA depletion, the effect of fixation on gene expression profiles, and the congruence of transcriptional profiles with bulk RNA-seq libraries, barcode collision rates, and generalizability to different bacterial species. After rRNA depletion, the fraction of mRNA reads increased from ~5% to 50%–90% of the total aligned reads, while preserving mRNA profiles relative to non-depleted samples ($R^2 = 0.81$; Figures 2A, 2B, and S1B). In addition, we confirmed that mRNA profiles were not affected by cell fixation in bulk experiments ($R^2 = 0.97$; Figure 2C) and that BacDrop produces transcriptional profiles that are well correlated with those generated by the traditional bulk RNA-seq method ($R^2 = 0.91$; Figure 2D).

We confirmed that the two-step barcoding strategy, which was used to enable loading of multiple cells per droplet, resulted in minimal cell barcode collision rates. We conducted round 1 plate barcoding using 384 RT primers and round 2 droplet barcoding containing on average 6 cells per droplet, although with a range of cell numbers per droplet, on a mixed population of two distinguishable species, *K. pneumoniae* and *P. aeruginosa*. Of 44,140 cells that passed quality control at a sequencing depth

of ~1,000 reads per cell, 42,893 cells (97.2%) had >99% of UMIs aligning to a single species, whereas the other 1,247 cells (2.8%) had >1% of UMIs aligning to both species (Figure 2E). The resulting barcode collision rate of 6.6% compares with published methods for eukaryotic droplet scRNA-seq that ranges from 0.36% to 11.3%, depending on the numbers of cells loaded in droplets.^{1,36} We performed all subsequent experiments with <3 cells per droplet, which would yield a library containing maximally a million cells in each 10x channel with an even lower collision rate.

To demonstrate the generalizability of BacDrop to different bacterial species and the ability of BacDrop to differentiate among multiple species, we applied BacDrop to the gram-negative *E. coli*, *K. pneumoniae*, *P. aeruginosa*,³⁹ and the gram-positive *E. faecium*⁴⁰ (Figure 2F; Table S2). Each strain was separately labeled with a unique set of CB1 (Table S1), allowing us to track and validate the accuracy of species identification in downstream analyses. CB1-barcoded cells were then pooled for round 2 droplet barcoding and library construction. All four species were distinguished in the analysis using Seurat and the RNase H-based rRNA depletion worked efficiently across all four species (Figures 2F, S1C, and S1D). Cells from *E. faecium* fell into several subpopulations distinguished by differential expression of two highly expressed housekeeping genes (*ef-tu* and *ef-g*; log₂ fold change ~0.6). We noted that *P. aeruginosa* cell numbers were underrepresented among the four species

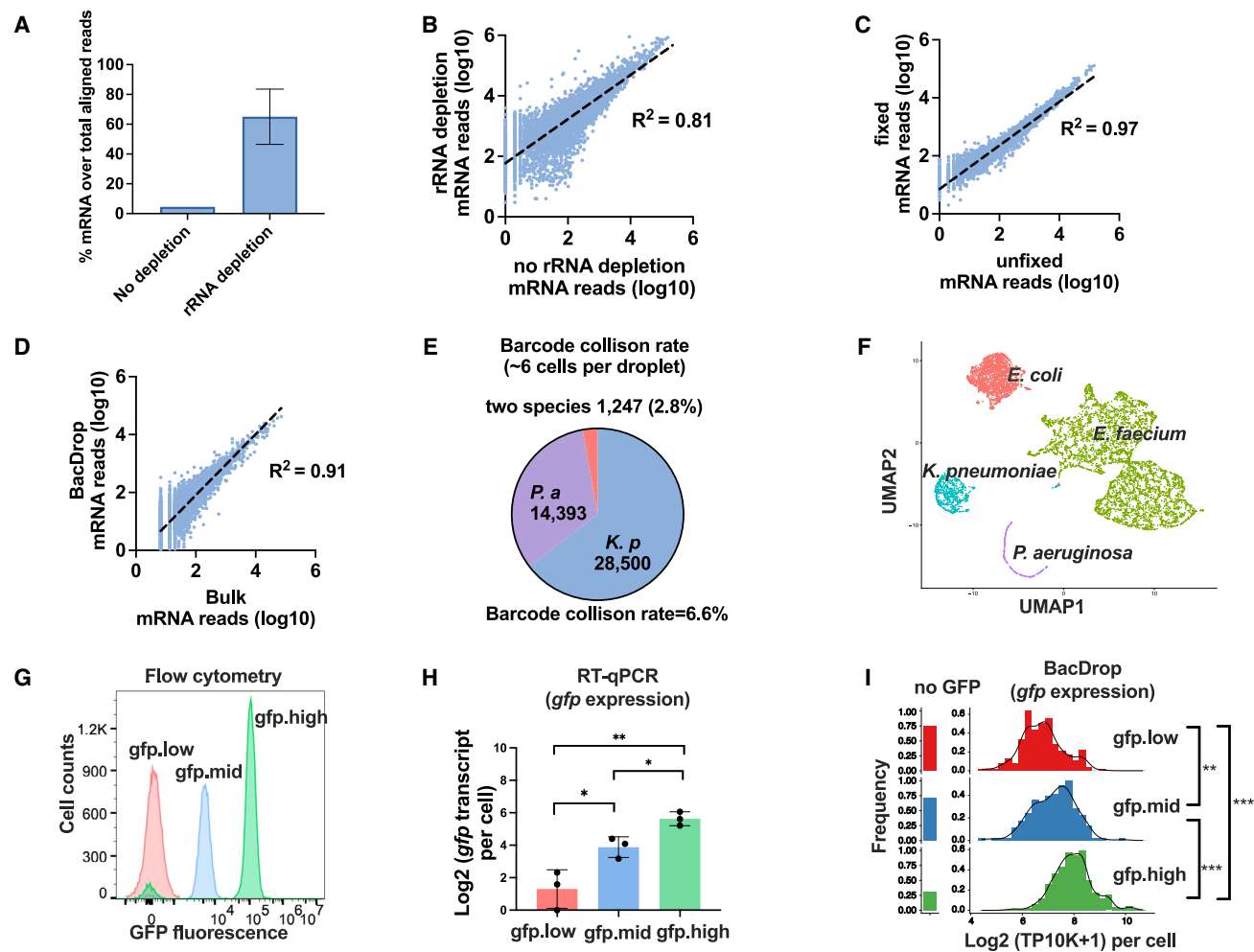


Figure 2. Validation and technical performance of BacDrop

(A) rRNA depletion inside fixed and permeabilized cells significantly increases the percentage of total reads that align to mRNA in BacDrop libraries of *K. pneumoniae*. The average percentage of reads aligned to mRNA genes was calculated from 10 independent BacDrop libraries of *K. pneumoniae* (5 without rRNA depletion and 5 with rRNA depletion), and error bars were plotted as the standard deviation.

(B) rRNA depletion does not affect transcriptional profiles. BacDrop libraries were constructed using in *K. pneumoniae* samples with or without rRNA depletion, and a linear regression model was fitted to the mRNA counts from each library ($R^2 = 0.81$).

(C) Cell fixation and permeabilization does not affect transcriptional profiles. Bulk RNA-seq results derived from Trizol-extracted RNA samples of *K. pneumoniae* versus RNA derived from fixed and permeabilized *K. pneumoniae* cells were highly correlated ($R^2 = 0.97$).

(D) Single-cell BacDrop results are highly correlated with bulk RNA-seq results ($R^2 = 0.91$) when analyzed in bulk mode (without cell barcode extraction).

(E) BacDrop has low barcode collision rates in an experiment where 2 million bacterial cells, mixed with *K. pneumoniae* and *P. aeruginosa* cells, were loaded into one 10x channel (~6 cells per droplet). About 2.8% of the cells were assigned to two species, resulting in a 6.6% barcode collision rate.

(F) BacDrop was performed on 4 different bacterial species including *E. coli*, *K. pneumoniae*, *P. aeruginosa*, and *E. faecium*. At the sequencing depth of 500 reads per cell, approximately 5,000 cells of *E. faecium*, 2,500 cells of *E. coli*, 1,000 cells of *K. pneumoniae*, and 300 cells of *P. aeruginosa* passed the analysis threshold (see STAR Methods). Uniform manifold approximation and projection (UMAP) of this mixed population shows the separation of different species, colored by species identity.

(G and H) Testing the sensitivity of BacDrop using three GFP strains of *E. coli*. The expression levels of *gfp* in these three *E. coli* strains were confirmed via flow cytometry (G). The mean numbers of *gfp* transcripts per cell were estimated via RT-qPCR (H). Three biological replicates were performed, and error bars were plotted as standard deviation. The two-tailed Student's t test was used for statistical analysis.

(I) Roughly 3,300 cells from each *gfp* strain were mixed to create a heterogeneous population and a BacDrop library was generated. The *gfp* expression levels were calculated using \log_2 -transformed value of transcript per 10,000 reads (\log_2 (TP10K+1)) per cell from the BacDrop results. Compared with the RT-qPCR results (H), BacDrop showed a good sensitivity for the *gfp*.high strain. The difference between *gfp*.mid and *gfp*.low is less distinct but statistically significant ($p < 0.005$). The Wilcoxon signed-ranks test was used for the statistical analysis.

See also Figure S1.

in the sequenced library and determined that their loss before round 2 droplet barcoding could be accounted for by inefficient recovery following centrifugation due to their small size (Figures S1E and S1F). This technical consideration was corrected in subsequent experiments by optimizing centrifugation speeds for each species/sample and by only pooling samples just prior to round 2 droplet barcoding to ensure equal number of cells from each sample are loaded into the 10x channel.

We next examined the transcriptome coverage of BacDrop in *E. coli* and *K. pneumoniae* by generating two small libraries containing either ~10,000 cells of *E. coli* or ~12,000 cells of *K. pneumoniae* (Figures S1G and S1H). Both libraries were sequenced with ~80,000 reads per cell, and we recovered roughly 4,000 or 6,000 cells (40%–50% cell recovery rates), containing at least 15 mRNA genes per cell, respectively. We detected the expression of ~70% genes of the *E. coli* genome and ~80% genes of the *K. pneumoniae* genome when we analyzed all cells together (Tables S3 and S4). At the single-cell level, we detected an average of 90 and 88 mRNA genes per cell (only including cells in which at least 15 mRNA genes detected), respectively, which is comparable with other reported bacterial scRNA-seq methods.^{10,12,13} We also generated a large library from ~1 million *K. pneumoniae* cells and sequenced with ~5,000 reads per cell. We recovered ~60,000 cells with at least 15 mRNA genes per cell and detected the expression of 96% genes of the entire genome when we analyzed all cells together (Table S5). At the single-cell level, we detected an average of 30 mRNA genes per cell across all 60,000 cells with at least 15 mRNA genes detected. However, from this large library, the top 3,000 high-quality cells had an average of 127 mRNA genes detected per cell (Figure S1I). Since this large library was only sequenced with ~5,000 reads per cell, we expect to detect a higher number of mRNA genes per cell with increased sequencing depth, as illustrated with the smaller libraries.

Finally, we assessed BacDrop's sensitivity to distinguish different expression levels of a gene. We created a heterogeneous population containing three *E. coli* strains constitutively expressing *gfp* at different levels (Table S2). We used flow cytometry to confirm the differing expression levels of *gfp* in these strains and estimated the mRNA copy numbers of *gfp* in each of these strains using RT-qPCR⁹ (Figures 2G and 2H). The estimated mRNA copy numbers per cell were 1–5 for the *gfp*.low strain, 9–30 for the *gfp*.mid strain, and 30–70 for the *gfp*.high strain. We then quantified *gfp* expression in single cells from the BacDrop results (Figure 2I) and found that *gfp* expression levels estimated from BacDrop are statistically different among three strains ($p < 0.005$), which is consistent with the RT-qPCR results. Notably, BacDrop did not detect the expression of *gfp* in certain fractions of cells from all three GFP strains, which is likely a true biological phenomenon given that the bimodality in GFP protein levels previously described in *E. coli*⁴¹ was also observed here with flow cytometry (Figure 2G).

Validating BacDrop's ability to identify subpopulations of a single bacterial isolate

Since our initial results showed that BacDrop is a robust and reliable technology with sufficient coverage and sensitivity that can be applied across different numbers of cells and species, we

next investigated whether BacDrop could also reproducibly identify subpopulations of cells within the same isolate. We chose to explore biological heterogeneity in the antibiotic-susceptible clinical isolate *K. pneumoniae* MGH66 in the absence and presence of antibiotic perturbations (Figure 3A; Table S2). We split one MGH66 culture ($OD_{600} \sim 0.2$) into four identical cultures, one of which was left untreated, whereas the other three were treated with an antibiotic with a different mechanism of action including: inhibition of cell wall synthesis (meropenem), DNA synthesis (ciprofloxacin), and protein synthesis (gentamicin). This experiment was performed in duplicate for two biological replicates (replicate 1 and replicate 2). For each replicate, bulk RNA-seq on samples collected by using the same treatment schemes revealed distinct cellular responses to each antibiotic relative to the untreated control. As expected (Figure 3B), ciprofloxacin induced genes involved in the major DNA damage response (SOS response), e.g., *recA*, whereas gentamicin treatment induced a group of heat-shock chaperone proteins, e.g., *ibpB*. In contrast, at 30 min, minimal transcriptional responses were observed in bulk from meropenem treatment, under the condition used ($OD_{600} \sim 0.2$, meropenem concentration 2 $\mu\text{g}/\text{mL}$), which is consistent with previous observations.⁴²

To validate BacDrop, we applied round 1 plate barcoding to each of these 4 samples separately using one of 4 distinct sets of CB1 (96 different CB1s corresponded to each sample, Table S1) that allowed us to confirm the original identity of these 4 samples and associate them with the corresponding antibiotic exposure. We then mixed all cells from the different treatments for round 2 droplet barcoding and library construction (Figure 3A). Each of the replicate libraries contained roughly 1 million cells. Replicate 1 was sequenced with ~5 billion paired-end reads (~5,000 reads per cell) and replicate 2 with ~3 billion paired-end reads (~3,000 reads per cell).

To confirm BacDrop's robustness and reproducibility, we compared the two replicate experiments. For the depth of sequencing performed for each experiment, we obtained ~80,000 cells with ≥ 15 mRNA genes per cell, with an average of ~30 unique mRNA genes detected per cell. No strong batch effect was observed between the replicates. Treatment with the different antibiotics resulted in cell clustering based on their treatment conditions (Figures 3C–3F). The overlap between the meropenem-treated and untreated samples (Figure S2) was consistent with their bulk RNA-seq results that suggest relatively minimal transcriptional response to meropenem, at least on the population level (Figure 3B). Across both replicates, 72% and 56% of the ciprofloxacin-treated cells belonged to the SOS-response cluster, whereas 72% and 66% of the gentamicin-treated cells belonged to the heat-shock response cluster, with good reproducibility across the two replicates (Figure 3E). These experiments confirmed that BacDrop could reproducibly identify population heterogeneity.

BacDrop reveals within-population heterogeneity with subpopulations driven largely by the expression of MGEs

We next analyzed in depth the untreated culture of MGH66 at the single-cell level to understand whether it contained any previously unrecognized within-population heterogeneity in

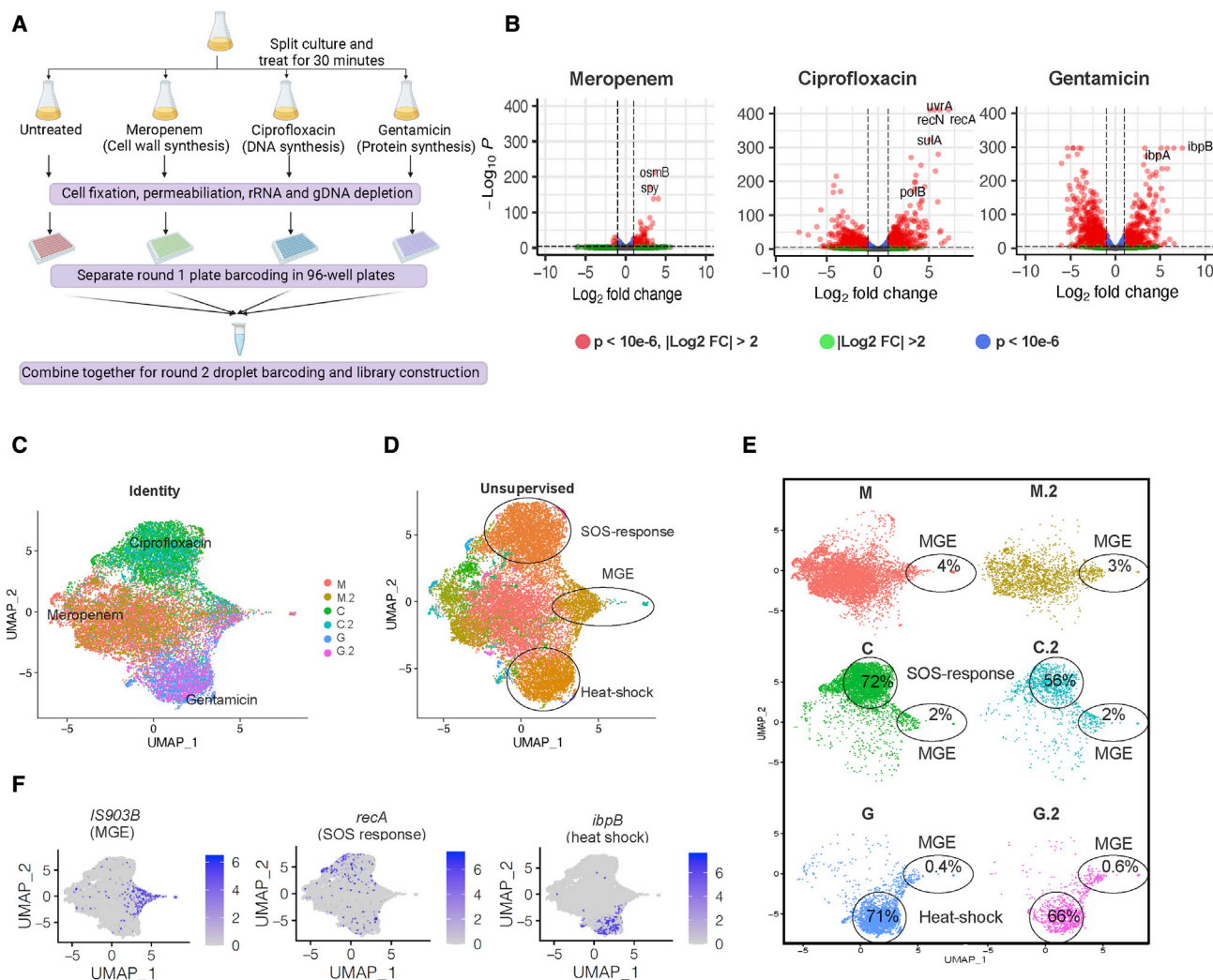


Figure 3. Validating BacDrop's ability to distinguish subpopulations based on distinct responses to different antibiotic treatments

(A) Creation of a BacDrop library containing cells of the same bacterial strain under 4 different conditions, including treatment of meropenem, ciprofloxacin, gentamicin, and untreated control. Cells were collected and processed separately until after round 1 plate barcoding. The four samples were then pooled for round 2 droplet barcoding and library construction. Two biological replicates were performed.

(B) Bulk RNA-seq results of cells exposed to the same antibiotic conditions as in (A). The abundance for each treated condition and comparison between the treated and untreated cultures are shown as well as significantly upregulated and downregulated genes from each treatment (performed in triplicates).

(C) UMAP plot based on the original identity of the 6 samples treated with meropenem (M and its replicate M.2), ciprofloxacin (C and its replicate C.2), and gentamicin (G and its replicate G.2).

(D) Unsupervised UMAP showed three clusters with significantly ($p < 0.05$) higher expression of genes in the SOS-response pathway, heat-shock response, and genes encoding an IS903B transposase (MGE).

(E) No strong batch effect was observed between the two biological replicates with the same treatment conditions.

(F) Expression of a representative gene from each cluster was highlighted on the UMAP. The purple color bars represent the normalized expression of a gene across all cells analyzed.

See also [Figure S2](#).

transcriptional states. From this untreated condition, we recovered ~50,000 cells with at least 15 unique mRNA genes detected in each cell, and we identified two major subpopulations in both replicates using an unsupervised clustering approach (see [Figures 4A](#) and [S3A](#) and [method details](#)). Although most cells fell into one major homogeneous subpopulation, 2,191 cells (~4.5%) fell into the MGE subpopulation driven by *IS903B* transposase genes ([Figure 4B](#)), which has 83 copies in MGH66

genome. (In fact, this MGE subpopulation is present in all 8 samples, untreated and antibiotic-treated replicates, suggesting both that the presence of this MGE population is a robust phenomenon and that BacDrop is reproducible; [Figures 3D–3F](#) and [4A](#).)

Previously, we had functionally shown that MGH66 and several other *K. pneumoniae* isolates have high-level transposon insertional mutagenesis activity, which contributes to their high frequencies of carbapenem-resistance acquisition⁴³; however,

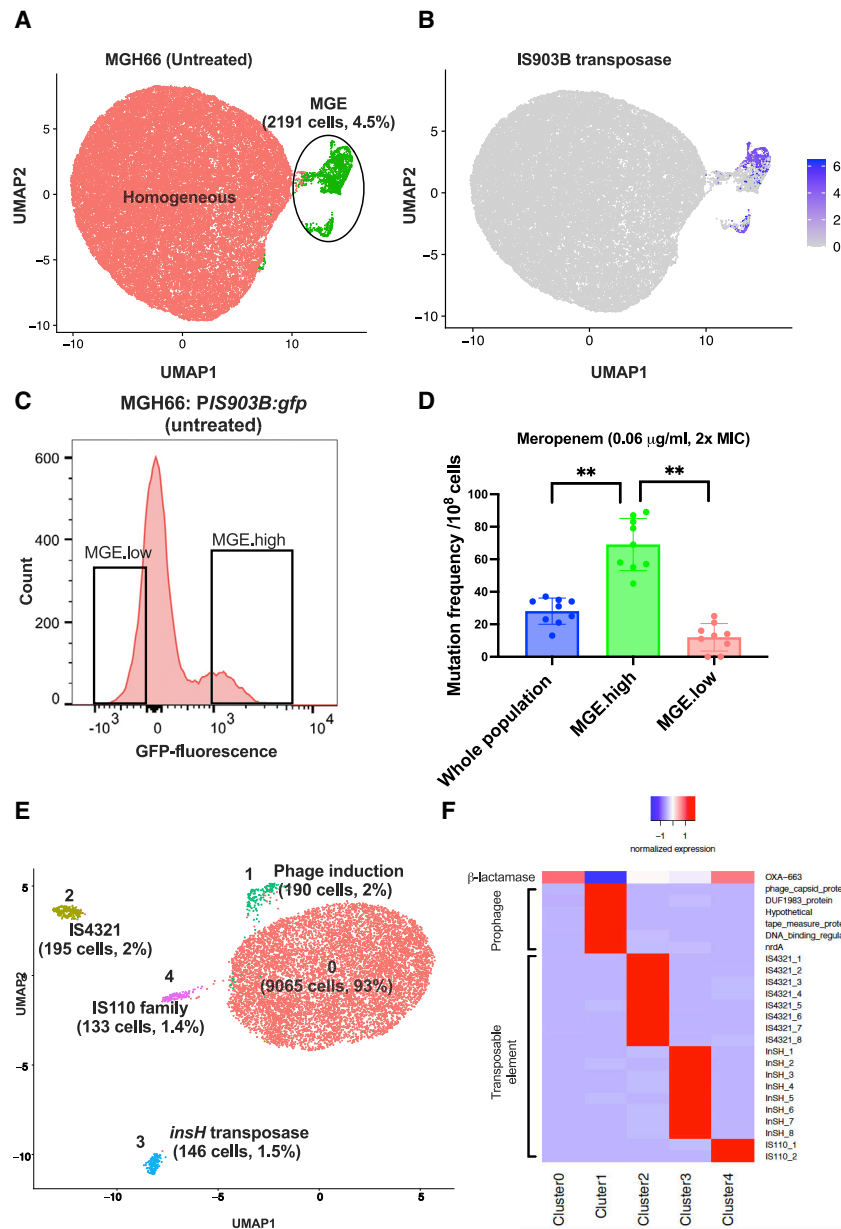


Figure 4. BacDrop reveals within-population heterogeneity driven largely by MGE

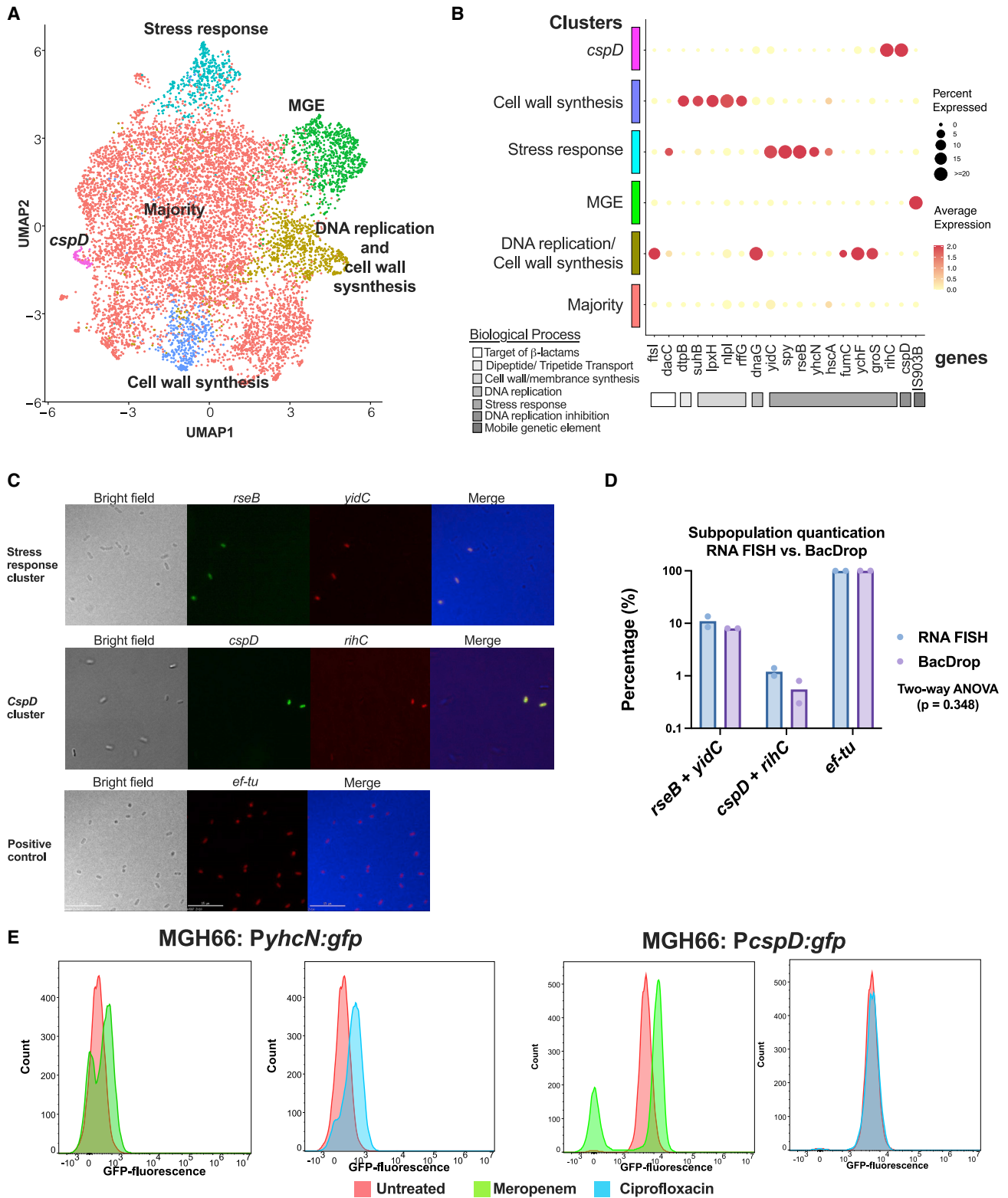
(A and B) In untreated culture of MGH66, a population showing high-level expression of *IS903B* transposase (MGE, 4.5%; green) was detected. (C and D) Flow cytometry of reporter MGH66 strains expressing GFP driven by the promoter of *IS903B* (MGH66:*PIS903B:gfp*) shows a heterogeneous expression pattern. The MGE.high population (~10% of the whole population) and MGE.low population (~10% of the whole population) were sorted into MHB medium without antibiotics, and mutation frequencies (D) were measured under meropenem treatment. Experiments in (C) and (D) were repeated with nine biological replicates. Error bars were plotted as the standard deviation. The Student's *t* test was used for statistical analysis. (E and F) MGE-driven subpopulations were detected in another *K. pneumoniae* clinical isolate BIDMC35. UMAP (E) and heatmap (F) shows 4 subpopulations differing from the majority population (cluster 0; red) of BIDMC35. Clusters 2, 3, and 4 are each driven by the high expression of a different transposase gene. In cluster 1, nearly all highly expressed genes belong to a prophage in BIDMC35 genome. Expression levels of all genes are normalized to expression in cluster 0 (F). See also [Figures S3](#) and [S4](#).

and measured their mutation frequencies under meropenem treatment using a modified fluctuation analysis.⁴³ The MGE.high population had at least 7 times higher mutation frequencies than the MEG.low population under meropenem treatment ($p = 0.002$) (Figure 4D), confirming our hypothesis that resistance is more likely to emerge from this subpopulation highly expressing MGE genes and revealing phenotypic consequences of the different subpopulations.

To examine the robustness of the MGE subpopulation in our datasets and to compare the relative values of deeper sequencing of fewer cells versus more shallow sequencing of more cells, we

bulk RNA-seq had failed to show elevated expression of any transposon genes in these strains. Here, the existence of this subpopulation with high-level transposon expression provides a possible explanation for the strain's elevated carbapenem-resistance frequencies, with resistance likely emerging from this small subpopulation. To test this hypothesis, we engineered a MGH66 reporter strain expressing the green fluorescent protein⁴⁴ driven by the promoter of one copy of the *IS903B* transposase genes (MGH66: *PIS903B:gfp*). Consistent with the BacDrop results, we observed heterogeneous expression pattern of *gfp* in this reporter strain by flow cytometry (Figure 4C). We then performed fluorescence-activated cell sorting (FACS) and sorted the MGE.high (high GFP expression) and MGE.low (low GFP expression) populations (Figures 4C, 4D, and S3B)

analyzed the untreated samples in replicate 1 and replicate 2 libraries separately, along with a smaller BacDrop library containing ~3,000 cells of MGH66 similarly collected but sequenced more deeply. We sequenced this smaller library to obtain 80,000 reads per cell with recovery of ~2,000 cells with at least 15 mRNA genes per cell, and collectively, an average of 85 mRNA genes detected per cell of these ~2,000 cells. Analysis of this smaller library detected the same MGE population identified in the larger cell population, albeit less distinctly (Figure S4). In contrast, both the replicates of the larger libraries, even when sequenced only to a depth to obtain ~30 mRNA genes per cell, identified an additional small subpopulation (0.25%–0.36%) featuring high-level expression of maltose transport genes that was not identified in the smaller library, despite its deeper



(legend on next page)

sequencing and higher coverage (Figure S4). Together, these results show that analyzing larger numbers of cells can help to reveal heterogeneity within a bacterial population, which is consistent with analyses in eukaryotic systems.^{16,17,19,25,26,28,29} Although the coverage can be higher when fewer cells were analyzed, increasing the scale, rather than the coverage, resulted in the identification of a rare population.

To determine whether MGE-driven subpopulations are unique to MGH66, we applied BacDrop to another *K. pneumoniae* clinical isolate (BIDMC35) (Table S2). BIDMC35 is a carbapenem-resistant isolate in which carbapenem resistance results from a transposon disruption of the major porin gene *ompK36* and the transposon-mediated high-level expression of a β -lactamase gene *bla*_{OXA-663}.⁴⁵ We again observed MGE-driven subpopulations, as in MGH66. Analyzing 9,748 BIDMC35 single cells that passed the analysis threshold at a sequencing depth of \sim 2,000 reads per cell (Figures 4E and 4F), we identified three clusters each driven by a unique transposon gene, including cluster 2 driven by an *IS4321* family transposase (195 cells, 2%), cluster 3 driven by the *insH* transposase (146 cells, 1.5%), and cluster 4 driven by an *IS110* family transposase (133 cells, 1.4%). Together with the observation in MGH66, it reinforces the finding that variable expression of MGEs may be one of the major drivers of population heterogeneity.

In BIDMC35, besides these MGE-driven subpopulations, we observed another unique subpopulation (190 cells, 2%) driven by the 30- to 320-fold higher expression of a group of prophage genes (Figures 4E and 4F; Table S6), compared with the rest of the populations, indicating that this cluster of cells was likely undergoing spontaneous phage induction. This observation is similar to the phage-induction subpopulation reported in the microSPLIT study.¹² Additionally, we found the expression of *bla*_{OXA-663} is significantly lower in this phage-induction subpopulation (Figure 4F), possibly caused by spontaneous phage induction.

BacDrop reveals heterogeneous stress responses to antibiotic exposure

Finally, to determine if a perturbed population might have heterogeneous dynamic transcriptional responses, we analyzed single-cell responses after different antibiotic exposures (Figures 3A and S5). Although ciprofloxacin or gentamicin resulted in clear transcriptional responses following treatment,

there was no obvious heterogeneity of response at the single-cell level under the conditions examined. In contrast, meropenem treatment, which only induced minimal transcriptional responses in bulk (Figure 3B), despite impacting cell killing at 30 min (Figure S5A), demonstrated heterogeneous responses on the single-cell level. Meropenem treatment induced four interesting subpopulations with distinct molecular responses, in addition to the MGE subpopulations (Figures 5A, 5B, and S5B; Table S7). These subpopulation clusters were characterized by co-upregulation of genes involved in: (1) the stress response (e.g., *rseB*, *yidC*, and *yhcN*), (2) cell wall/membrane synthesis (e.g., *nlpI* and *lpxH*), (3) cell wall synthesis and DNA replication (e.g., *dnaG* and *ftsI*), or (4) cold shock response (e.g., *pnp* and *cspD*). Notably, although CspD was initially identified as a cold shock protein, it has also been reported to be a DNA-binding toxin that inhibits DNA synthesis and induces the formation of persisters in *E. coli*.⁴⁶

Having identified genes highly expressed in specific subpopulations after meropenem treatment (Figure 5B; Table 1), we used RNA fluorescence *in situ* hybridization (FISH) to validate the existence of the “stress response” and “cspD-expressing” subpopulations by probing for two genes highly co-expressed in each cluster (*rseB* and *yidC*, stress-response cluster; *cspD* and *rihC*, *cspD*-expressing cluster) (Figure 5C). RNA FISH confirmed both the existence and proportion of cells assigned to the clusters identified by BacDrop (Figures 5C and 5D).

We also validated BacDrop using fluorescence cytometry to independently quantify cell numbers in the stress response and *cspD*-expressing clusters. We engineered MGH66 reporter strains expressing GFP⁴⁴ driven by the promoter of *yhcN*, a stress-response gene (MGH66:PyhcN:*gfp*), or *cspD* (MGH66:PcspD:*gfp*). Consistent with BacDrop results, meropenem but not ciprofloxacin induced heterogeneous expression of both *yhcN* and *cspD* (Figure 5E). We confirmed that the cells with reduced fluorescence of MGH66:PcspD:*gfp* (GFP-low) were live and not simply dying using a live-dead stain and plating for colony forming units (Figure S6). Interestingly, the GFP-low population had lower fluorescence than the untreated population, suggesting that there may be suppression of *cspD* expression in this population. Of note, the fractions of cells highly expressing these two genes were greater by flow cytometry than by BacDrop; this would be consistent with the fact that RNA and protein levels are not

Figure 5. BacDrop reveals heterogeneous responses to meropenem exposure

(A) Besides the subpopulation highly expressing the *IS903B* transposase genes (MGE), meropenem treatment induced heterogeneous responses, including a stress-response subpopulation, a cell wall synthesis subpopulation, a DNA replication and cell wall synthesis subpopulation, and a *cspD*-expressing subpopulation.

(B) Dot plot showing the expression of genes that are significantly different among clusters and the percentage of cells expressing these genes in each cluster.

(C) Validation of subpopulations identified in the meropenem-treated MGH66 using RNA FISH with double marker genes from the stress-response subpopulation (*rseB* [green] + *yidC* [red]) and *cspD*-expressing subpopulation (*cspD* [green] + *rihC* [red]). A probe targeting the housekeeping gene *ef-tu* was used as the positive control to show that more than 99% of cells were successfully permeabilized and hybridized. Subpopulations co-expressing double marker genes were identified. The scale bar size is 15 μ m.

(D) Across 20 fields of view, the RNA FISH results showed that \sim 1% of cells co-expressed *cspD* and *rihC*, and \sim 10% of cells co-expressed *rseB* + *yidC*. These results were statistically consistent with the BacDrop result in which \sim 0.6% of cells co-expressed *cspD* and *rihC*, and \sim 8% of cells co-expressed *rseB* + *yidC*. Two-way ANOVA was performed for the statistical analysis ($p = 0.348$). This experiment was repeated twice, and data were plotted separately from two replicates.

(E) Flow cytometry of reporter MGH66 strains expressing GFP driven by the promoter of *yhcN* (MGH66:PyhcN:*gfp*; left) or *cspD* (MGH66:PcspD:*gfp*; right) shows a heterogeneous response to meropenem (green) but not to ciprofloxacin (blue) treatment, relative to untreated control (red).

See also Figure S5 and Table 1.

Table 1. Marker genes identified in clusters of the meropenem-treated sample

Gene name	Gene product	Function	Biological pathways	Cluster	Log ₂ fold change	p value
<i>ftsI</i>	PBP3	target of β-lactams	cell wall synthesis	DNA replication and cell wall synthesis	5.2957	<1.00E–300
<i>dacC</i>	PBP6	target of β-lactams	cell wall synthesis	stress response	3.7036599	1.81E–40
<i>dtpB</i>	DtpB	transportation of β-lactams	dipeptide/tripeptide transport	cell wall synthesis	6.281594	<1.00E–300
<i>suhB</i>	SuhB	membrane protein transport and insertion	cell wall/membrane synthesis	cell wall synthesis	4.406598	1.39E–125
<i>lpxH</i>	LpxH	lipid A biosynthesis	cell wall/membrane synthesis	cell wall synthesis	7.605374	<1.00E–300
<i>nlpI</i>	NlpI	cell wall synthesis, bind to PBP4 and PBP1A	cell wall/membrane synthesis	cell wall synthesis	3.487722	2.54E–87
<i>rffG</i>	RffG	LPS synthesis	cell wall/membrane synthesis	cell wall synthesis	4.280285	5.97E–107
<i>dnaG</i>	DnaG	DNA replication	DNA replication	DNA replication and cell wall synthesis	4.490715	<1.00E–300
<i>yidC</i>	YidC	membrane damage response	stress response	stress response	4.0687447	8.25E–100
<i>spy</i>	Spy	membrane damage response	stress response	stress response	4.0142941	2.33E–98
<i>rseB</i>	RseB	membrane damage response	stress response	stress response	5.1874707	2.26E–218
<i>yhcN</i>	YhcN	cytoplasmic acid stress	stress response	stress response	6.1654092	<1.00E–300
<i>hscA</i>	HscA	cold shock	stress response	stress response	3.2262902	5.75E–26
<i>fumC</i>	FumC	oxidative stress	stress response	DNA replication and cell wall synthesis	5.304257	<1.00E–300
<i>ychF</i>	YchF	oxidative stress	stress response	DNA replication and cell wall synthesis	5.286679	<1.00E–300
<i>groS</i>	GroES	stress response	stress response	DNA replication and cell wall synthesis	4.510794	2.41E–261
<i>rihC</i>	RihC	DNA damage response	stress response	<i>cspD</i> expressing	6.16122	<1.00E–300
<i>cspD</i>	CspD	persister formation	DNA replication inhibition	<i>cspD</i> expressing	6.342385	<1.00E–300
<i>IS903B</i>	IS903B	insertion sequence	mobile genetic elements	MGE	4.463117	<1.00E–300

necessarily well correlated in single cells,⁹ as one copy of mRNA can produce one to hundreds of copies of proteins and mRNA is less stable than protein. Nevertheless, the heterogeneous responses were observed by both methods of measurement.

Taken together, under meropenem perturbation, we observed strong heterogeneous responses driven by various stress-response pathways that had been previously masked in bulk RNA-seq results. Rather than uniformly turning on a specific stress-response pathway in all cells, a diverse range of stress responses appears to be induced in different subpopulations, which could potentially contribute to heterogeneous cell fates such as cell lysis or antibiotic tolerance.

BacDrop identifies a subpopulation with reduced meropenem efficacy and increased persisters

Given CspD's reported role in inducing the formation of antibiotic-tolerant persisters in *E. coli*,⁴⁶ we wondered whether antibiotic-tolerant cells might make up a greater percentage of the GFP-high than the GFP-low subpopulation, with cells in which *cspD* had been induced surviving preferentially under meropenem treatment. Because CspD plays a role in persister formation rather than maintenance, we anticipated that some cells might

induce *cspD* and thus make high levels of GFP, become antibiotic tolerant, then turn off CspD expression resulting in reduced expression of GFP as the GFP is degraded. Despite this possibility, we nevertheless hypothesized that the GFP-high subpopulation might still be enriched for tolerant cells compared with the GFP-low subpopulation. We performed FACS with MGH66:P*cspD::gfp* coupled with dead-cell stain, after a 30-min exposure to meropenem, and sorted live cells into GFP-high and GFP-low (Figure 6A) subpopulations directly into liquid media with and without meropenem. (We confirmed that all sorted cells have similar survival rates on LB agar plates without antibiotics; Figure S6.) Indeed, the GFP-high subpopulation was enriched for meropenem tolerant cells compared with the GFP-low subpopulation, with evidence of a persister population in the GFP-high but not the GFP-low subpopulations (Figure 6B). We verified that no genetic mutations were acquired by these persister cells via whole genome sequencing. Consistent with this observation, we also observed ~100 times more persister cells in a MGH66 strain over-expressing *cspD*, although the overexpression of *cspD* did not reduce antibiotic susceptibility as reflected in the minimal inhibitory concentration of meropenem (Figures 6C and 6D). Together, these results confirmed

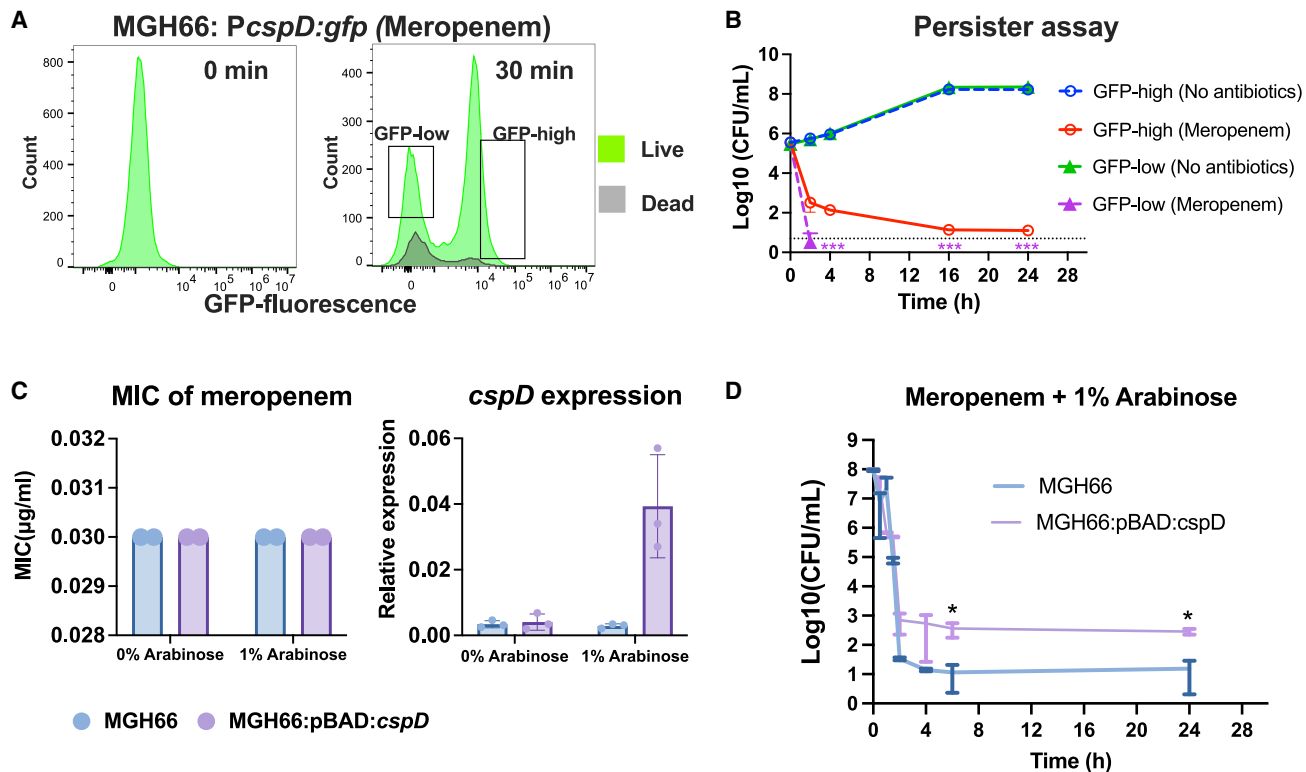


Figure 6. *cspD*-expressing cluster was enriched with persister cells

(A) MGH66:*PcspD:gfp* cells were analyzed by FACS at time 0 and 30 min after exposure to meropenem (2 µg/mL). After 30 min, live cells from GFP-low and GFP-high subpopulations were identified and sorted.

(B) GFP-low and GFP-high subpopulations sorted from meropenem-treated MGH66:*PcspD:gfp* cells differed in their response to meropenem. GFP-low and GFP-high subpopulations were sorted directly into media with no antibiotic or with meropenem (2 µg/mL). Samples were then taken over time and plated on solid agar to enumerate CFU. Persisters only emerged from the GFP-high subpopulation. The limit of detection is indicated by black-dashed line. Three asterisks from the treated GFP-low subpopulation indicate that no persister was observed.

(C and D) Overexpression of *cspD* in MGH66 increased numbers of persisters but did not affect the susceptibility of meropenem. *cspD* driven by the arabinose inducible promoter pBAD was transformed into MGH66. (C) Induction of *cspD* with 1% arabinose did not affect the minimal inhibitory concentration (MIC) of meropenem. (D) When *cspD* was induced with 1% arabinose and treated with meropenem at 2 µg/mL (purple), the numbers of persister cells were significantly greater at 6- and 24-h time points (asterisk shows significance at 6-h and 24-h time points) compared with the culture without arabinose induction under 2 µg/mL meropenem treatment (blue).

All experiments were repeated 3 times. Error bars were plotted as the standard deviation. The Student's t test was used for statistical analysis.

See also Figure S6.

the role that *cspD* plays in persister formation in a subpopulation of *K. pneumoniae* MGH66.

DISCUSSION

We report a novel bacterial scRNA-seq method, BacDrop, that is robust and reproducible and leverages droplet-based technology to enable the massively parallel profiling of the transcriptional program of thousands to millions of single bacterial cells. In line with analyses in mammalian systems,^{16,25–29} we demonstrated that studying large numbers of cells can robustly identify population heterogeneity including rare subpopulations even with the relatively sparse per cell coverage afforded by single-cell technologies (Figure S4). We applied BacDrop to study the naturally occurring heterogeneity in a heretofore presumed uniform culture of bacteria and its heterogeneous responses to perturbation. Due to limited scale, previous genome-wide sin-

gle-cell transcriptional studies on smaller numbers of cells (hundreds to thousands) have demonstrated between-population heterogeneity by artificially mixing different populations.^{10–12} Here, we characterized both a stable and a dynamic population of cells derived from the same bacterial isolate and demonstrated a diversity of states within the stable population and heterogeneous responses in the dynamic population after perturbation with antibiotics.

Using BacDrop, we report the observation of within-population heterogeneity driven predominantly by the expression of MGEs in the presumed uniform cultures of *K. pneumoniae*. Although MGEs drive genetic diversity by their relatively random movement in the genome, we find that the expression levels of individual MGEs are also highly variable within populations of MGH66 and BIDMC35 (Figure 4). Whether this heterogeneity in expression is due to genotypic variation resulting from the movement of genetic elements such as transposon insertions

or phase variation, to transcriptional changes in response to stochastic/or local microenvironmental cues, or to epigenetic mechanisms remains to be understood. However, the consequences of such variation can be significant. We demonstrated that the high-level expression of MGE genes found only in a subpopulation of MGH66 contribute to the strain's propensity to become carbapenem resistant, consistent with our previous finding that high-level transposon mutagenesis plays an important role in high-frequency evolution of carbapenem resistance.⁴³

Meanwhile, in response to a perturbation such as exposure to the important antibiotic meropenem, BacDrop revealed a wide range of transcriptional responses that are masked in bulk RNA-seq. Interestingly, a diverse range of stress responses appear to be induced in different subpopulations rather than a single or uniform response occurring in all cells; this subpopulation diversity could potentially contribute to heterogeneous cell fates or phenotypic outcomes, including cell lysis or antibiotic tolerance. It remains to be understood whether this heterogeneous stress response occurring in subpopulations is a general phenomenon in response to stressors or is specific to certain types of stresses such as meropenem, which may be working in a more pleiotropic manner than is assumed, thereby eliciting a wide range of stress responses. Meanwhile, we examined one such subpopulation defined by high-level expression of *cspD*, a gene encoding a toxin that inhibits DNA replication and induces the formation of persisters but that is not significantly upregulated in bulk RNA-seq results. Using both FISH and flow cytometry, we confirmed that *cspD* is induced in a subpopulation after meropenem treatment. Moreover, we found a higher survival rate under meropenem treatment in the *cspD*-induced subpopulation, pointing to its role in antibiotic tolerance within this subpopulation and more generally, to the importance of certain responses in some subpopulations in surviving the lethal effects of antibiotics.

We have developed BacDrop to be a versatile method to characterize the transcriptional programs of thousands to millions of bacterial cells in multiple species. To demonstrate its utility, we validated BacDrop in four different pathogenic bacterial species. We expect that BacDrop should be easily adaptable to more species, including commensal bacterial strains of the microbiome and other environmental species. Therefore, BacDrop will enable the identification of both heterogeneity in cell types (species) in a mixed community or cell states in a single strain under stable or dynamic conditions. We thus propose that BacDrop will become a powerful tool for a broad range of studies, including studies focused on elucidating phenotypic heterogeneity, understanding bacterial interactions in microbial communities, dissecting host-pathogen interactions, expanding our knowledge of the microbiome beyond genomes and bulk metabolomics, and investigating the emergence of antibiotic resistance, persistence, and tolerance.

Limitations of the study

Although BacDrop has the capacity to study millions of cells simultaneously, we are currently limited by sequencing costs which impact sequencing depth and the information (i.e., genes per cell) that can be extracted per cell. Nevertheless, we expect

that sequencing costs will continue decreasing and BacDrop will be a powerful platform for large-scale single-cell experiments in bacteria. Although we have demonstrated the ability of BacDrop to work on four major bacterial pathogenic species, including the gram-negative *E. coli*, *K. pneumoniae*, and *P. aeruginosa*, and the gram-positive *E. faecium* (Figure 2F), we nevertheless recognize that there will be phylogenetic biases in performance, including, as we show, variability in centrifugation in permeabilized cells and also other factors, e.g., permeabilization reagents and concentrations. We thus anticipate that limitations will be present when applying BacDrop to complex bacterial communities such as microbiota samples, with the ability to characterize the single-cell transcriptional programs of millions of cells without requiring any prior knowledge of the genomes of interest outweighing more modest technical limitations of individual species.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Bacterial strains
- METHOD DETAILS
 - Cell fixation and permeabilization
 - In-cell rRNA depletion and gDNA removal
 - In-cell reverse transcription, round 1 cell barcoding and sample multiplexing
 - In-cell cDNA 3' poly-A tailing
 - Droplet generation
 - Second strand cDNA synthesis and round 2 droplet barcoding in droplets
 - Breaking emulsions and cDNA purification
 - cDNA enrichment
 - Illumina sequencing library construction
 - Illumina sequencing of BacDrop libraries
 - Bulk library construction
 - Quality control for rRNA and gDNA depletion
 - Assessing cell recovery rates during cell permeabilization
 - Saturation curve for coverage analysis
 - BacDrop experiments to assess the numbers of cells required to detect extremely rare populations
 - Estimation of mRNA copy numbers in *E. coli* GFP strains using RT-qPCR
 - Killing kinetics
 - Construction of GFP reporter strains
 - Cell sorting of untreated MGH66: *PIS903B:gfp* and measurement of mutation frequencies
 - Flow cytometry of the meropenem-treated samples
 - RNA FISH
 - Whole genome sequencing of the persister cells of MGH66

- Processing of sequencing data
- Analysis of BacDrop data
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.cell.2023.01.002>.

ACKNOWLEDGMENTS

We thank Jonathan Livny, Noam Shores, and Eugenio Mattei for discussions on the data analysis, Zohar Bloom-Ackermann for the suggestions on the flow cytometry, Anne Clatworthy and Thulasi Warriar for the comments on the manuscript, and Kevin Gosselin and Ming Pan for their discussion on the droplet generation steps. We thank Aviv Regev for her expert advice and comments on the manuscript. We thank the Microbial Omics Core at the Broad Institute for generating bulk RNA-seq libraries and the Flow Cytometry Core Facility at the Broad Institute for cell sorting. This publication was supported in part by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award 5R01AI117043-05 to D.T.H. and by a generous gift from Anita and Josh Bekenstein.

AUTHOR CONTRIBUTIONS

P.M. and D.T.H. designed the study and wrote the manuscript. P.M. and L.L.H. performed the experiments. P.M., H.M.A., and S.J.G. performed data analysis. R.P.B. provided extensive input and suggestions on the design of this study and manuscript preparation. R.N. provided extensive input and suggestions on the development of BacDrop. C.S.S. provided extensive input and suggestions on the data analysis, cluster identification, and manuscript revision. All authors have read and approved the manuscript.

DECLARATION OF INTERESTS

D.T.H., P.M., and H.M.A. have filed an U.S. Patent Application (application no. 17/819,034) based on this work.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: July 24, 2022

Revised: November 15, 2022

Accepted: January 3, 2023

Published: January 27, 2023

REFERENCES

1. Macosko, E.Z., Basu, A., Satija, R., Nemes, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214. <https://doi.org/10.1016/j.cell.2015.05.002>.
2. Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C., Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240. <https://doi.org/10.1038/nature12172>.
3. Shalek, A.K., Satija, R., Shuga, J., Trombetta, J.J., Gennert, D., Lu, D., Chen, P., Gertner, R.S., Gaublomme, J.T., Yosef, N., et al. (2014). Single-cell RNA-seq reveals dynamic paracrine control of cellular variation. *Nature* 510, 363–369. <https://doi.org/10.1038/nature13437>.
4. Shapiro, E., Biezuner, T., and Linnarsson, S. (2013). Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630. <https://doi.org/10.1038/nrg3542>.
5. Zeisel, A., Muñoz-Manchado, A.B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., et al. (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* 347, 1138–1142. <https://doi.org/10.1126/science.aaa1934>.
6. Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161, 1187–1201. <https://doi.org/10.1016/j.cell.2015.04.044>.
7. Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, 377–382. <https://doi.org/10.1038/nmeth.1315>.
8. Grün, D., Lyubimova, A., Kester, L., Wiebrands, K., Basak, O., Sasaki, N., Clevers, H., and van Oudenaarden, A. (2015). Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525, 251–255. <https://doi.org/10.1038/nature14966>.
9. Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emilii, A., and Xie, X.S. (2010). Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538. <https://doi.org/10.1126/science.1188308>.
10. Blattman, S.B., Jiang, W., Oikonomou, P., and Tavazoie, S. (2020). Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat. Microbiol.* 5, 1192–1201. <https://doi.org/10.1038/s41564-020-0729-6>.
11. Imdahl, F., Vafadarnejad, E., Homberger, C., Saliba, A.E., and Vogel, J. (2020). Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat. Microbiol.* 5, 1202–1206. <https://doi.org/10.1038/s41564-020-0774-1>.
12. Kuchina, A., Brettner, L.M., Paleologu, L., Roco, C.M., Rosenberg, A.B., Carignano, A., Kibler, R., Hirano, M., DePaolo, R.W., and Seelig, G. (2021). Microbial single-cell RNA sequencing by split-pool barcoding. *Science* 371, eaba5257. <https://doi.org/10.1126/science.aba5257>.
13. Homberger, C.H., Regan, Barquist, L., and Vogel, J. (2022). Improved bacterial single-cell RNA-seq through automated MATQ-seq and Cas9-based removal of rRNA reads. Preprint at bioRxiv. <https://doi.org/10.1101/2022.11.28.518171>.
14. McNulty, R., Sritharan, D., Liu, S., Hormoz, S., and Rosenthal, A.Z. (2021). Droplet-based single cell RNA sequencing of bacteria identifies known and previously unseen cellular states. Preprint at bioRxiv. <https://doi.org/10.1101/2021.03.10.434868>.
15. Dar, D., Dar, N., Cai, L., and Newman, D.K. (2021). Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* 373, eabi4882. <https://doi.org/10.1126/science.abi4882>.
16. Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. *Nature* 541, 331–338. <https://doi.org/10.1038/nature21350>.
17. Heimberg, G., Bhatnagar, R., El-Samad, H., and Thomson, M. (2016). Low dimensionality in gene expression data enables the accurate extraction of transcriptional programs from shallow sequencing. *Cell Syst.* 2, 239–250. <https://doi.org/10.1016/j.cels.2016.04.001>.
18. Shekhar, K., Lapan, S.W., Whitney, I.E., Tran, N.M., Macosko, E.Z., Kowalczyk, M., Adiconis, X., Levin, J.Z., Nemes, J., Goldman, M., et al. (2016). Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. *Cell* 166, 1308–1323.e30. <https://doi.org/10.1016/j.cell.2016.07.054>.
19. Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretzky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776–779. <https://doi.org/10.1126/science.1247651>.
20. Olah, M., Menon, V., Habib, N., Taga, M.F., Ma, Y., Yung, C.J., Cimpean, M., Khairallah, A., Coronas-Samano, G., Sankowski, R., et al. (2020). Single cell RNA sequencing of human microglia uncovers a subset associated

- with Alzheimer's disease. *Nat. Commun.* **11**, 6129. <https://doi.org/10.1038/s41467-020-19737-2>.
21. Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J.Z., Menon, M., He, L., Abdurrob, F., Jiang, X., et al. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* **570**, 332–337. <https://doi.org/10.1038/s41586-019-1195-2>.
22. Del-Aguila, J.L., Li, Z., Dube, U., Mihindukulasuriya, K.A., Budde, J.P., Fernandez, M.V., Ibanez, L., Bradley, J., Wang, F., Bergmann, K., et al. (2019). A single-nuclei RNA sequencing study of Mendelian and sporadic AD in the human brain. *Alzheimers Res. Ther.* **11**, 71. <https://doi.org/10.1186/s13195-019-0524-x>.
23. Habib, N., Avraham-Davidi, I., Basu, A., Burks, T., Shekhar, K., Hofree, M., Choudhury, S.R., Aguet, F., Gelfand, E., Ardlie, K., et al. (2017). Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958. <https://doi.org/10.1038/nmeth.4407>.
24. Zheng, G.X., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049. <https://doi.org/10.1038/ncomms14049>.
25. Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., et al. (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058. <https://doi.org/10.1038/nbt.2967>.
26. Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Aron, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17. <https://doi.org/10.1016/j.cell.2016.11.038>.
27. Wang, X., He, Y., Zhang, Q., Ren, X., and Zhang, Z. (2021). Direct comparative analyses of 10X Genomics chromium and smart-seq2. *Genomics Proteomics Bioinformatics* **19**, 253–266. <https://doi.org/10.1016/j.gpb.2020.02.005>.
28. Zhang, X., Xu, C., and Yosef, N. (2019). Simulating multiple faceted variability in single cell RNA sequencing. *Nat. Commun.* **10**, 2611. <https://doi.org/10.1038/s41467-019-10500-w>.
29. Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F., et al. (2014). Quantitative assessment of single-cell RNA-sequencing methods. *Nat. Methods* **11**, 41–46. <https://doi.org/10.1038/nmeth.2694>.
30. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386. <https://doi.org/10.1038/nbt.2859>.
31. Ackermann, M. (2015). A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* **13**, 497–508. <https://doi.org/10.1038/nrmicro3491>.
32. Dewachter, L., Fauvar, M., and Michiels, J. (2019). Bacterial heterogeneity and antibiotic survival: understanding and combatting persistence and heteroresistance. *Mol. Cell* **76**, 255–267. <https://doi.org/10.1016/j.molcel.2019.09.028>.
33. Andersson, D.I., Nicoloff, H., and Hjort, K. (2019). Mechanisms and clinical relevance of bacterial heteroresistance. *Nat. Rev. Microbiol.* **17**, 479–496. <https://doi.org/10.1038/s41579-019-0218-1>.
34. Regev, A., Teichmann, S.A., Lander, E.S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., et al. (2017). The human cell atlas. *eLife* **6**, e27041. <https://doi.org/10.7554/eLife.27041>.
35. Centers for Disease Control and Prevention (2019). Antibiotic resistance threats in the United States. <https://www.cdc.gov/drugresistance/pdf/threatsreport/2019-ar-threats-report-508.pdf>.
36. Datlinger, P., Rendeiro, A.F., Boenke, T., Senekowitsch, M., Krausgruber, T., Barreca, D., and Bock, C. (2021). Ultra-high-throughput single-cell RNA sequencing and perturbation screening with combinatorial fluidic indexing. *Nat. Methods* **18**, 635–642. <https://doi.org/10.1038/s41592-021-01153-z>.
37. Wulf, M.G., Maguire, S., Humbert, P., Dai, N., Bei, Y., Nichols, N.M., Corrêa, I.R., and Guan, S. (2019). Non-templated addition and template switching by Moloney murine leukemia virus (MMLV)-based reverse transcriptases co-occur and compete with each other. *J. Biol. Chem.* **294**, 18220–18231. <https://doi.org/10.1074/jbc.RA119.010676>.
38. Vasilyev, N., Gao, A., and Serganov, A. (2019). Noncanonical features and modifications on the 5'-end of bacterial sRNAs and mRNAs. *Wiley Interdiscip. Rev. RNA* **10**, e1509. <https://doi.org/10.1002/wrna.1509>.
39. Stover, C.K., Pham, X.Q., Erwin, A.L., Mizoguchi, S.D., Warrener, P., Hickney, M.J., Brinkman, F.S., Hufnagle, W.O., Kowalik, D.J., Lagrou, M., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* **406**, 959–964. <https://doi.org/10.1038/35023079>.
40. Lebreton, F., Manson, A.L., Saavedra, J.T., Straub, T.J., Earl, A.M., and Gilmore, M.S. (2017). Tracing the enterococci from Paleozoic origins to the hospital. *Cell* **169**, 849–861.e13. <https://doi.org/10.1016/j.cell.2017.04.027>.
41. Choi, P.J., Cai, L., Frieda, K., and Xie, X.S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**, 442–446. <https://doi.org/10.1126/science.1161427>.
42. Bhattacharyya, R.P., Bandyopadhyay, N., Ma, P., Son, S.S., Liu, J., He, L.L., Wu, L., Khafizov, R., Boykin, R., Cerqueira, G.C., et al. (2019). Simultaneous detection of genotype and phenotype enables rapid and accurate antibiotic susceptibility determination. *Nat. Med.* **25**, 1858–1864. <https://doi.org/10.1038/s41591-019-0650-9>.
43. Ma, P., He, L.L., Pironti, A., Laibinis, H.H., Ernst, C.M., Manson, A.L., Bhattacharyya, R.P., Earl, A.M., Livny, J., and Hung, D.T. (2021). Genetic determinants facilitating the evolution of resistance to carbapenem antibiotics. *eLife* **10**, e67310. <https://doi.org/10.7554/eLife.67310>.
44. Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M.G., and Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods* **3**, 623–628. <https://doi.org/10.1038/nmeth895>.
45. Ma, P., Laibinis, H.H., Ernst, C.M., and Hung, D.T. (2018). Carbapenem resistance caused by high-level expression of OXA-663 beta-lactamase in an OmpK36-deficient *Klebsiella pneumoniae* clinical isolate. *Antimicrob. Agents Chemother.* **62**, e01281–e01218. <https://doi.org/10.1128/AAC.01281-18>.
46. Kim, Y., and Wood, T.K. (2010). Toxins Hha and CspD and small RNA regulator Hfq are involved in persister cell formation through MqsR in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **391**, 209–213. <https://doi.org/10.1016/j.bbrc.2009.11.033>.
47. Shishkin, A.A., Giannoukos, G., Kucukural, A., Ciulla, D., Busby, M., Surka, C., Chen, J., Bhattacharyya, R.P., Rudy, R.F., Patel, M.M., et al. (2015). Simultaneous generation of many RNA-seq libraries in a single reaction. *Nat. Methods* **12**, 323–325. <https://doi.org/10.1038/nmeth.3313>.
48. Wiegand, I., Hilpert, K., and Hancock, R.E. (2008). Agar and broth dilution methods to determine the minimal inhibitory concentration (MIC) of antimicrobial substances. *Nat. Protoc.* **3**, 163–175. <https://doi.org/10.1038/nprot.2007.521>.
49. Wright, E.S., Yilmaz, L.S., Corcoran, A.M., Ökten, H.E., and Noguera, D.R. (2014). Automated design of probes for rRNA-targeted fluorescence in situ hybridization reveals the advantages of using dual probes for accurate identification. *Appl. Environ. Microbiol.* **80**, 5124–5133. <https://doi.org/10.1128/AEM.01685-14>.
50. Skinner, S.O., Sepúlveda, L.A., Xu, H., and Golding, I. (2013). Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat. Protoc.* **8**, 1100–1113. <https://doi.org/10.1038/nprot.2013.066>.

51. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
52. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9, e112963. <https://doi.org/10.1371/journal.pone.0112963>.
53. Smith, T., Heger, A., and Sudbery, I. (2017). UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* 27, 491–499. <https://doi.org/10.1101/gr.209601.116>.
54. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930. <https://doi.org/10.1093/bioinformatics/btt656>.
55. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. *Cell* 177, 1888–1902.e21. <https://doi.org/10.1016/j.cell.2019.05.031>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial strains		
MGH66	Ma et al. ⁴⁵	GenBank: GCA_000694555.1
BIDMC35	Ma et al. ⁴⁵	GenBank: GCA_000567225.2
UCI38	Ma et al. ⁴⁵	GenBank: GCA_000566805.1
<i>E. coli</i> 10B	NEB	GenBank: GCF_000008865.2
PAO1	Stover et al. ³⁹	GenBank: GCF_000006765.1
EnGen0052-E3346	Lebreton et al. ⁴⁰	GenBank: GCA_000322285.1
<i>gfp</i> .low	Zaslaver et al. ⁴⁴	N/A
<i>gfp</i> .mid	Zaslaver et al. ⁴⁴	N/A
<i>gfp</i> .high	Zaslaver et al. ⁴⁴	N/A
MGH66: <i>PyhcN:gfp</i>	This study	N/A
MGH66: <i>PcpsD:gfp</i>	This study	N/A
MGH66: <i>PIS903B:gfp</i>	This study	N/A
MGH66:pBAD: <i>cspD</i>	This study	N/A
Recombinant DNA		
pUA139	Zaslaver et al. ⁴⁴	N/A
pUA139- <i>PyhcN:gfp</i>	This study	N/A
pUA139- <i>PcpsD:gfp</i>	This study	N/A
pUA139- <i>PIS903B:gfp</i>	This study	N/A
Oligos		
SMRT_dT, 2nd strans synthesis after TdT treatment with dATP	IDT	AAGCAGTGGTATCAACGCAGAGTTTTTTTTTTTTTTTTV
SMRT_PCR, used for cDNA amplification	IDT	AAGCAGTGGTATCAACGCAGAGT
P5, P5 partial primer, used for cDNA amplification (paired with SMRT_PCR) and for library construction (paired with Nextera index primers N7**)	IDT	AATGATACGGCGACCACCGAGA
<i>cspD</i> _F, Forward primer used to amplify the promoter of <i>cspD</i> from MGH66 and ligate into pUA139.	IDT	ACTGGGATCCATGCAGATGGCGGTTTAATCGCAT
<i>cspD</i> _R, Reverse primer used to amplify the promoter of <i>cspD</i> from MGH66 and ligate into pUA139.	IDT	ACTGCTCGAGATGCCAT ACT TCG ACA TCC TTC GT
<i>YhcN</i> _F, Forward primer used to amplify the promoter of <i>yhcN</i> from MGH66 and ligate into pUA139.	IDT	ACTGGGATCCATGCAAGCGTTGTGATGAGAACAT
<i>YhcN</i> _R, Reverse primer used to amplify the promoter of <i>yhcN</i> from MGH66 and ligate into pUA139.	IDT	ACTGCTCGAGATGCGAT GTT CAC CTC GTC GAA TC
RNA FISH oligos	IDT	Table S2
Deposited data		
BacDrop data	This paper	GEO: GSE180237
Bulk RNA-seq data	This paper	GEO: GSE180237
MGH66 persister WGS data	This paper	GEO: GSE180237

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Key reagent		
NxGen RNase Inhibitor	Lucigen	30281-2
Lysozyme	ThermoFisher	90082
NEBNext rRNA Depletion Kit	NEB	E7850X
DNase I	Sigma	AMPD1
SUPERase-In RNase inhibitor	ThermoFisher	AM2696
Maxima H Minus reverse transcriptase	ThermoFisher	EP0753
DTT	ProMega	P1171
dNTPs	NEB	N0447L
Terminal transferase (TdT)	NEB	M0315L
dATP	NEB	N0446S
SUPERase-In RNase inhibitor	ThermoFisher	AM2696
Chromium Next GEM Chip H	10X Genomics	1000161
Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit	10X Genomics	1000176
KAPA HiFi HotStart ReadyMix	Roche	KK2602
5% FC-40 oil	RAN Biotechnologies	008-FluoroSurfactant-5wtF
Buffer EB	Qiagen	19086
5X SYBR Green	VWR	12001-796
Illumina Nextera XT DNA library preparation kit	Illumina	FC-131-1096
Nextera index 1 primer	Illumina	FC-131-2001
AMPure XP beads	Beckman Coulter	A63881

RESOURCE AVAILABILITY

Lead contact

Please direct requests for resources and reagents to the lead contact: Deborah T. Hung (hung@molbio.mgh.harvard.edu)

Materials availability

Plasmids generated in this study are available from the [lead contact](#) upon request.

Data and code availability

Sequencing data and the processed counting matrix have been deposited to GEO repository (GEO: GSE180237).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Bacterial strains

Bacterial strains used in this study are listed in [Table S2](#). *K. pneumoniae*, *E. coli*, and *P. aeruginosa* were cultured in Luria-Bertani (LB) medium or Mueller-Hinton Broth (MHB) with shaking at 37 °C. *E. faecium* was cultured in Todd-Hewitt broth (THB) with shaking at 37 °C. For the GFP experiment, *E. coli* strains expressing *gfp* driven by different promoters were cultured separately in LB at 37 °C. Early exponential growth phase cells (OD₆₀₀ ~0.2) were collected and fixed immediately. For the antibiotic treatment experiment, *K. pneumoniae* clinical isolate MGH66 was cultured in MHB at 37 °C until early exponential phase. Then the culture was diluted to OD₆₀₀ 0.05 in MHB. After growing for two doublings (~40 min, OD₆₀₀ ~0.2), the cultures were split into four equal volume cultures. One culture was left untreated, while the other three were treated with relevant antibiotic at breakpoint concentrations set by the Clinical and Laboratory Standards Institute (CLSI): 2 µg/mL for meropenem, 2.5 µg/mL for ciprofloxacin and 4 µg/mL for gentamicin. After 30 min, 7 mL of cells were collected from each of these cultures, and immediately proceeded with the cell fixation protocol. For the bulk RNA-seq experiments, samples were collected using the same treatment schemes, and three biological replicates were included in each condition. For the BIDMC35 experiment, BIDMC35 was cultured in LB medium at 37 °C until early exponential phase (OD₆₀₀ ~0.2). 7 mL cells were then collected and immediately proceed with the cell fixation protocol.

METHOD DETAILS

Cell fixation and permeabilization

All reagents ([key resources table](#)) for cell fixation and permeabilization were kept ice-cold. Up to 10 billion bacterial cells grown at specified conditions were collected by centrifuging at 5525 x g for 10 min at 4 °C. The supernatant was removed and cell pellets were resuspended in 7 mL fresh, ice-cold 4% formaldehyde (Sigma, CN 47608) in 1x PBS and incubated with shaking overnight at 4 °C (Digital Platform Rocker Shaker, VWR). Following overnight fixation, cells were centrifuged at 5525 x g for 10 min at 4 °C. The supernatant was removed and cells were resuspended in 7 mL PBS-RI (1x PBS supplemented with 0.1 U/μL NxGen RNase inhibitor (Lucigen, CN 30281)). Cells were centrifuged again at 5525 x g for 10 min at 4 °C and pellets were resuspended in 700 μL PBS-RI. Subsequent centrifugations were carried out at 7000 x g for 5 minutes at 4 °C. We noticed that some species, e.g., *P. aeruginosa*, do not pellet very well at this centrifugation speed ([Figures S1E and S1F](#)). Thus, we recommend optimizing the centrifugation speed for the specific species of interest. Cells were centrifuged and resuspended in 700 μL 50% Ethanol in PBS-RI. Cells were then washed twice with PBS-RI. After the second wash, cells were resuspended in 1 mL 100 mM Tris-HCL (pH 7.5) supplemented with 0.1 U/μL NxGen RNase inhibitor. Cells were diluted 100x and quantified using a hemocytometer (VWR, CN 102966).

We found that the number of cells in a permeabilization reaction is critical for achieving sufficient permeabilization. Insufficient permeabilization may result in inefficient rRNA depletion and gDNA removal. For 1 permeabilization reaction, up to 40 million cells were centrifuged and resuspended in 250 μL 0.04% Tween-20 in 1x PBS. If more cells are desired, multiple parallel reactions can be set up. Immediately following a 3-minute incubation on ice, 1 mL cold PBS-RI was added, and cells were spun down and resuspended in 200 μL lysozyme mix (100 mM Tris (pH 8.0), 50 mM EDTA pH 8.0, 0.25 U/μL NxGen RNase Inhibitor, 2.5 mg/mL lysozyme). Cells were then incubated at 37 °C for 15 minutes. After the incubation, 1 mL PBS-RI was added and cells were washed twice with 175 μL PBS-RI. After the second wash, cells were resuspended in 150 μL PBS (without RNase inhibitor added) and cell concentrations were measured by diluting cells 100 times and counting using a hemocytometer.

In-cell rRNA depletion and gDNA removal

Immediately after cell permeabilization, up to 40 million cells were centrifuged and resuspended in 11 μL nuclease-free H₂O. The cell number is critical for achieving efficient rRNA depletion ([Figure S1D](#)). 2 μL NEBNext Bacterial rRNA depletion solution and 2 μL Probe Hybridization Buffer (NEB, CN E7850) were mixed with cells on ice. The hybridization was conducted per the following (lid temperature set to 55 °C): 50 °C for 2 minutes, ramp down to 22 °C at 0.1 °C/second, and hold at 22 °C for 5 minutes. Probe hybridization was immediately followed by RNase H digestion by mixing the probe-hybridized cells with 2 μL RNase H reaction buffer, 2 μL Thermostable RNase H (NEB, CN E7850), and 1 μL nuclease-free H₂O, followed by a 30-minute incubation at 50 °C (lid temperature set to 55 °C). The 20 μL reaction was centrifuged and resuspended in 10 μL DNase-RI buffer (1 μL DNase I reaction buffer (Sigma, CN AMPD1), 1 μL DNase I (Sigma, CN AMPD1), 0.025 μL NxGen RNase inhibitor (Lucigen, CN 30281), 8 μL nuclease-free H₂O). The reaction was incubated at room temperature for 30 minutes, and the DNase treatment was stopped by adding 1 μL Stop Solution (50 mM EDTA) and incubating at 50 °C for 10 minutes. After the incubation, cells were centrifuged and washed twice with 100 μL PBS-RI. After the second wash, cells were resuspended in 20 μL 0.5x PBS supplemented with 1 U/μL SUPERase-In RNase inhibitor and used immediately for in-cell reverse transcription.

In-cell reverse transcription, round 1 cell barcoding and sample multiplexing

The round 1 plate barcoding and sample multiplexing is achieved via RT reactions in 384- or 96-well plates. 384 RT primers ([Table S1](#)) containing UMI sequences and round 1 plate barcodes (CB1) were synthesized at Integrated DNA Technologies at 100 μM concentration. The primers were diluted with ddH₂O to a working concentration of 25 μM, and 2.5 μL of each primer was aliquoted into individual wells of the 384- or 96-well plates. The rRNA and gDNA depleted cells was diluted with nuclease-free H₂O supplemented with 1 U/μL SUPERase-In RNase inhibitor and added to the plate containing RT primers (1 μL cells per well). Then the cell-primer mix was incubated at 55 °C for 5 minute and immediately put on ice. For each well, the RT master mix, containing 0.25 μL DTT (100 mM), 0.25 μL dNTP (10 mM each), 0.25 μL SUPERase-In, 1 μL RT buffer, 0.25 μL Maxima H Minus reverse transcriptase (Thermo Scientific, CN EP0753), was added and the RT reaction was incubated as follows (set lid temperature to 60 °C): 22 °C for 30 min, 50 °C for 10 min, 3 cycles of [8 °C for 12 s, 15 °C for 45 s, 20 °C for 45 s, 30 °C for 30 s, 42 °C for 2 min, 50 °C for 3 min], 50 °C for 5 min, hold at 4 °C.

In-cell cDNA 3' poly-A tailing

After RT, cells were recovered and pooled from 384- or 96-well plate. For multiplexed samples, we suggest to only pool cells from the same sample together, generating individual pools for each sample. This will allow more flexibility to adjust cell numbers of different samples during droplet generation. Then the pooled cells were centrifuged and each pool was resuspended 40 μL nuclease-free H₂O supplemented with 1 U/μL SUPERase-In. The poly-A tailing reaction was set up as the following: 38 μL cells, 5 μL 10x terminal transferase buffer, 1 μL dATP (100 mM) (NEB, N0446S), 1 μL SUPERase-In, 5 μL terminal transferase (TdT) (NEB, CN M0315L). The reaction was incubated at 37 °C for 1 hour. Then 10 μL 0.2 M EDTA was added to each 50 μL reaction and incubated at

room temperature for 10 minutes. Cells were then centrifuged and resuspended in 10 μ L nuclease-free H₂O supplemented with 1 U/ μ L SUPERase-In. Cells were diluted and concentrations were quantified.

Droplet generation

The Chromium Next GEM Chip H (10x Genomics, PN 1000161) and Chromium Next GEM Single Cell ATAC Library & Gel Bead Kit (10x Genomics, PN 1000176) was used for the droplet generation. The unused wells were filled with 70 μ L (row 1), 50 μ L (row 2), and 40 μ L (row 3) 50% glycerol solution. The desired number of cells was diluted to 33.75 μ L with nuclease-free H₂O supplemented with 1 U/ μ L SUPERase-In. Right before loading the chip, 33.75 μ L cells were mixed with a PCR master mix containing 37.5 μ L KAPA HiFi HotStart ReadyMix (Roche, CN KK2602), 2.25 μ L second strand synthesis primer SMRT_dT (10 μ M) ([key resources table](#)), and 1.5 μ L Reducing Agent B (10x Genomics, PN 2000087). The chip was loaded with 70 μ L PCR-cell mix (row 1), 50 μ L Gel Beads (row 2), and 40 μ L partitioning oil (row 3), and run on the Chromium system. Approximately 100 μ L droplet emulsion was obtained in row 3.

Second strand cDNA synthesis and round 2 droplet barcoding in droplets

To increase thermostability of the droplets, we split each 100- μ L emulsion into 4 25- μ L reactions into PCR tubes (USA Scientific, CN 1402-4700). In each reaction, 25 μ L 5% FC-40 oil (RAN Biotechnologies, CN 008-FluoroSurfactant-5wtF) was added to the bottom and 50 μ L mineral oil (Sigma, CN M5904) was added on the top. The 2nd strand cDNA synthesis and round 2 cell barcoding was performed as follows: 95 °C for 30 s, 39 °C for 5 min, 65 °C for 10 min; then 4 cycles of [98 °C for 20 s, 62 °C for 15 s, 72 °C for min], 72 °C for 5 min, hold at 4 °C.

Breaking emulsions and cDNA purification

After the round 2 cell barcoding was finished, the mineral oil and FC-40 oil was removed from the PCR tube, being careful and not to remove the middle layer which contains the emulsion. The emulsion was then combined. In cases where only a small number of cells are desired for library construction, the emulsion can be kept separately but all downstream reaction volume should be reduced accordingly. Each 100 μ L emulsion can then be broken by adding 125 μ L Recovery Agent (10x Genomics, PN 2000087). The tubes were inverted 10 times and centrifuged briefly to ensure that all droplets were broken. 125 μ L Recovery Agent/Partitioning Oil (pink) from the bottom of the tube was then removed. cDNA was first purified using Dynabeads. In brief, for each reaction, a mix containing 182 μ L cleanup buffer (10x Genomics, PN 2000088), 8 μ L Dynabeads MyOne SILANE (10x Genomics, PN 2000048), 5 μ L Reducing Agent B, and 5 μ L Nuclease-free water was added. After mixing and incubating at room temperature for 10 minutes, samples were placed on a magnetic separator and washed twice with freshly prepared 80% ethanol. After removing ethanol from the second wash, each sample was eluted in 40.5 μ L elution buffer that was prepared by mixing 98 μ L Buffer EB (Qiagen, CN 19086), 1 μ L 10% Tween 20 (Teknova, CN T0710), and 1 μ L Reducing Agent B. Then 40 μ L of the elution was transferred to a fresh PCR tube and subjected to a 0.6x Cleanup with AMPure XP beads (Beckman Coulter, CN A63881). The cDNA was then eluted in 30 μ L nuclease-free water.

cDNA enrichment

Before the cDNA enrichment, a qPCR reaction, containing 1 μ L cDNA, 5 μ L KAPA HiFi HotStart Ready Mix, 0.3 μ L primer P5 (10 μ M), 0.3 μ L primer SMRT_PCR (10 μ M) ([key resources table](#)), 2 μ L 5x SYBR green (VWR, CN 12001-796), and 1.4 μ L nuclease-free water, was set up to determine cycle numbers of the cDNA enrichment in a real-time thermocycler using the following program: 98 °C for 3 min, 30 cycles of [98 °C for 20 s, 67 °C for 20 s, 72 °C for 3 min], 72 °C for 5 min, hold at 4 °C. The cycle numbers at which the qPCR reaction reaches early exponential amplification phase was determined as the cycle numbers for cDNA enrichment. For cDNA enrichment, 25 μ L of cDNA was mixed with 125 μ L KAPA HiFi HotStart Ready Mix, 7.5 μ L primer P5 (10 μ M), 7.5 μ L primer SMRT_PCR (10 μ M), and 85 μ L nuclease-free water. The cDNA was enriched using the same program as the qPCR reaction, using the cycle numbers determined from the qPCR reaction. The enriched cDNA was purified using 0.6x AMPure XP beads and eluted in 50 μ L nuclease-free water.

Illumina sequencing library construction

The Illumina Nextera XT DNA library preparation kit (Illumina, CN FC-131-1096) was used to prepare sequencing libraries with these following modifications: ~ 2 ng enriched cDNA and 4 μ L ATM was used for each 50 μ L tagmentation reaction. Following tagmentation, 2.5 μ L primer P5 (5 μ M) and 2.5 μ L Index 1 primer (N7**, Illumina, CN FC-131-2001) was used for the PCR enrichment. The PCR reaction was removed from the thermocycler and immediately put on ice after 5 cycles of PCR amplification. Then a qPCR reaction, containing 5 μ L Nextera XT PCR reaction, 3 μ L NPM, 1 μ L primer P5 (5 μ M), 1 μ L Index 1 primer, 3 μ L SYBR green, and 2 μ L nuclease-free water, was set up to determine the remaining cycle numbers of the library enrichment using the following program: 72 °C for 3 min, 95 °C for 30s, 25 x [95 °C for 10 s, 55 °C for 30 s, 72 °C for 30 s], 72 °C for 5 min, hold at 10 °C. The cycle numbers at which the qPCR reaction reaches one third of the fluorescence saturation was determined as the remaining cycle numbers for library enrichment. Then the remaining 45 μ L library enrichment PCR reaction was put back to a thermocycler and amplified using the same program with cycle numbers determined from the qPCR. The libraries were then subjected to a 0.6x cleanup with AMPure XP beads and eluted in 25 μ L nuclease-free water.

Illumina sequencing of BacDrop libraries

Libraries were diluted to desired concentrations and sequenced on the Illumina NovaSeq 6000 platform with standard sequencing primers, using the following specifications: Read 1: 60 bp; Read 2: 39 bp; Index 1: 8 bp; Index 2: 16 bp. Depending the scale of the experiment and the sequencing depth desired, NovaSeq 6000 SP (Illumina #20027464), S1 (Illumina #20012865), or S2 (Illumina #20012862) reagent kits were used.

Bulk library construction

To construct sequencing libraries from bacterial cultures, cell pellets collected from 1 mL early exponential phase cultures were re-suspended in 500 μ L TRIzol Reagents (ThermoFisher, CN 15596026) and frozen at -80 C for at least 20 min. Cells were then thawed and mixed with 250 μ L of 0.1 mm diameter Zirconia/Silica beads (BioSpec Products), and lysed mechanically via bead-beating for 90 second at 10 m/sec on a FastPrep (MP Bio). After addition of 0.1 mL chloroform, each sample tube was mixed thoroughly by inversion, incubated for 3 minutes at room temperature, and centrifuged at 12,000 xg for 15 minutes at 4°C. The aqueous phase was mixed with an equal volume of 100% ethanol, transferred to a Direct-zol spin column (Zymo Research, CN R2051), and RNA was extracted according to the Direct-zol protocol. The sequencing libraries were then generated using the RNAtag-Seq protocol.⁴⁷

To construct sequencing libraries from fixed cells, or fixed and permeabilized cells, 20 μ L cells were pelleted and resuspended 20 μ L lysis buffer (50 mM Tris pH 8.0, 200 mM NaCl, 25 mM EDTA pH 8.0) supplemented with 1.6 μ L proteinase K (50 mg/mL). Cells were lysed at 55 °C for 1 hour, then RNA was purified using 1.5x AMPure RNAClean XP beads (Beckman Coulter, CN A63987) and eluted in 20 μ L nuclease-free water. The sequencing libraries were then generated using the RNAtag-Seq protocol.

Quality control for rRNA and gDNA depletion

The efficiency of rRNA and gDNA depletion can be assessed by lysing and extracting RNA from permeabilized cells. As shown in [Figure S1A](#), without rRNA and gDNA depletion, three peaks were clearly seen by running the extracted RNA on Agilent TapeStation RNA 1000 high sensitivity tape. The three peaks are 16s rRNA (~1000 bp), 23 rRNA (~2000 bp), and gDNA (> 4000 bp). When an efficient rRNA and gDNA depletion was achieved, these three peaks will be absent. Due to the capacity of the rRNA depletion kit (10 μ g total RNA maximal), the rRNA depletion efficiency is a function of cell numbers ([Figure S1D](#)). In BacDrop protocol, we used 4×10^7 cell for each rRNA depletion reaction, resulting in a ~80% depletion efficiency.

Assessing cell recovery rates during cell permeabilization

In the experiment in [Figure S1E](#), cell numbers were quantified by counting cells using a hemacytometer before and after each centrifugation step to calculate the cell recovery rates. Equal numbers of cells from each species were mixed before round 1 plate barcoding (RT). Due to different cell sizes, these four species had different recovered rates at each centrifugation step at 7,000 x g. After RT and PolyA tailing (right before round 2 droplet barcoding), the cells recovered from each species differed significantly.

Saturation curve for coverage analysis

Three libraries were used to generate the saturation curve to assess the coverage of BacDrop ([Figures S1G–S1I](#)). The first library containing 10,000 cells of *E. coli* was sequenced at 80,000 reads per cell. Roughly ~4,000 cells were recovered with an average of 90 mRNA genes detected per cell. The second library containing 12,000 cells of *K. pneumoniae* was sequenced at 80,000 reads per cell. Roughly ~6,000 cells were recovered with an average of 88 mRNA genes detected per cell. The third library containing ~1 million cells of *K. pneumoniae* was sequenced at 5,000 reads per cell. Top 3,000 cells were analyzed with an average of 127 mRNA genes detected per cell. For the first two libraries, sequencing reads were randomly subsampled to ~40,000, ~20,000, ~4,000 reads per cell, and the analysis was repeated to calculate numbers of mRNA genes detected per cell. For the third library, sequencing reads were randomly subsampled to ~4,000, ~3,000, ~2,000 reads per cell, and the analysis was repeated to calculate numbers of mRNA genes detected per cell.

BacDrop experiments to assess the numbers of cells required to detect extremely rare populations

Two libraries (Replicate 1 and Replicate 2) were constructed with ~250k untreated MGH66 cells. We sequenced Replicate 1 with ~5,000 reads/cell and Replicate 2 with ~3,000 reads/cell, and recovered ~40k cells and ~10k cells from Replicate 1 and Replicate 2, respectively. In these two libraries, a subpopulation highly expressing maltose transport genes (i.e, *lamB*) were detected. However, this subpopulation was not detected in the third library containing only ~3,000 untreated MGH66 cells that were collected at the same condition. The MGE subpopulation was detected in all libraries.

Estimation of mRNA copy numbers in *E. coli* GFP strains using RT-qPCR

The estimation of mRNA copy numbers was performed using a protocol modified from a previous study.⁹ In brief, a *gfp* gBlock fragment was synthesized at Integrated DNA technology. A serial dilution was performed to create *gfp* dsDNA standards ranging from 1 to 10^{10} molecules/ μ L. One μ L of the standard was used in a 10 μ L qPCR reaction to generate the standard curve. 10 ng RNA from each GFP strain (~66,666 cells with the assumption that there is 0.15 pg RNA per bacterial cell) was converted into cDNA and subjected to qPCR reaction together with the *gfp* dsDNA standards. GFP mRNA copy numbers were estimated by mapping to the standard curve.

Killing kinetics

At each time points, cells under antibiotic treatments and the untreated samples were diluted and plated on LB agar plates without antibiotics. CFUs from each condition were enumerated and normalized to the CFUs of the untreated samples at the same time points to calculate the killing rates for each time points.

Construction of GFP reporter strains

We amplified the promoter region of *cspD* and *yhcN* from MGH66 and ligated it into pUA139⁴⁴ using the BamHI and XhoI sites, resulting in pUA139-*PcspD* and pUA139-*PyhcN*. For the promoter of IS903B, we synthesized the promoter sequence at IDT with BamHI and XhoI restriction digestion sites as the following sequences and ligated it into pUA139: GGATCCAGAAATTCTCTGTTCCATGGTAGATTAATAAGTCCCCAACATTTAAATATACAGGATAATCTAAATATTACTTCGTTCTTATCCTTAATAAATGGCAAATTTTCATTTAATTTATTTTCAAATTATTCTGATGCATGAGTTACCCTATAATTTACACATAAAGAAGGCTTTGTTGAATAAATCGAACTTTTGCTGAGTTGCTCGAG.

The construct was then transformed into *E. coli* DH5 α and the plasmids were extracted and verified using Sanger sequencing. Electrocompetent cells of MGH66 was made as previously reported.⁴³ The extracted plasmids were then transformed into MGH66 via electroporation, generating MGH66: *PIS903B:gfp*, MGH66: *PcspD:gfp* and MGH66: *PyhcN:gfp*.

Cell sorting of untreated MGH66: *PIS903B:gfp* and measurement of mutation frequencies

For cell sorting of MGH66: *PIS903B:gfp*, the top 10% cells with high-level GFP expression and bottom 10% of cells with low-level GFP expression were sorted into MHB medium without any antibiotics. This 10% sorting window was determined as a compromise between GFP levels and the time spent on sorting. Ideally, we could sort 1% of the population with the highest GFP expression, but this would take hours to sort enough cells for the experiments to measure mutation frequencies, which might alter transcript levels or cell physiology. We found 10% is a feasible time frame for us to get enough cells with sufficient GFP expression for subsequent experiments. The sorted cells were then used as the starting culture for measuring the mutation frequencies. We used a modified fluctuation analysis to measure mutation frequencies as previously described.⁴³ In brief, ~100 sorted cells were seeded into each well of 384-well plates, followed by incubating at 37°C for 3 hours. After the incubation, cells from three randomly chosen wells were taken and plated for cell counting. The rest of the wells were added meropenem at 0.06 μ g/mL (2x MIC) and treated overnight at 37°C in a humidified chamber. The 2nd day morning, wells with resistant mutants growing up were counted to calculate the mutation frequencies. An unsorted culture was included as a control. This experiment was repeated 3 times with 2 biological replicates each time.

Flow cytometry of the meropenem-treated samples

The cells for flow cytometry were prepared using the same scheme as the meropenem-treated sample for BacDrop. After treating with meropenem at 2 μ g/mL for 30 minutes, cells were diluted into PBS at the final concentration of 10⁶ cells/mL and immediately run through flow cytometer. For the flow cytometry including dead-cell staining, 10 μ L propidium iodide (1 mg/mL) (Invitrogen, cat# P1304MP) was mixed with 1 mL PBS. Cells after treatment were then diluted in the staining buffer at the final concentration of 10⁶ cells/mL, and incubated in the dark at room temperature for 15 minutes followed by running through flow cytometer. To sort GFP-low and GFP-high populations, cells were run through fluorescence-activated cell sorting (FACS). We sorted ~10⁶ cells into 2 mL LB without or with meropenem (2 μ g/mL) from each population. A 50 μ L aliquot from cells sorted in LB medium without antibiotics was immediately diluted and plated on LB agar plates without antibiotics. The colony forming units (CFU) on LB agar plates were enumerated to calculate the survival rates. The rest of the cells was used to measure MICs using the standard microdilution protocol⁴⁸ or proceeded with the persister assay. For the persister assay, cells were cultured at 37 °C with shaking. At each time points, a 50 μ L aliquot was taken, diluted and plated on LB agar plates without antibiotics. Three replicates were performed in each experiment.

RNA FISH

RNA FISH probes were designed using Design Probes tool of DECIPHER.⁴⁹ 3 - 5 probes were designed for each target gene ([key resources table](#)). Probes were synthesized at Integrated DNA technologies and labeled with either Cy3 or Alex488 dye at the 5'. Probes of the same gene were pooled together and diluted into 10 μ M stocks. Meropenem-treated MGH66 cells were fixed and permeabilized using the same protocol as the cell fixation and permeabilization steps in BacDrop. The hybridization was carried out in 40% hybridization buffer at 50 °C overnight and the following washing steps of washing were performed as described previously.⁵⁰ Cells were imaged using the DeltaVision widefield deconvolution imaging system with a 60x objective. ImageJ was used for image data analysis and cell quantification ([Table S8](#)).

Whole genome sequencing of the persister cells of MGH66

Genomic DNA was isolated using DNeasy Blood and Tissue Kits (Qiagen, cat. # 69504) and quantified using Qubit dsDNA HS Assay Kit (Invitrogen, cat. # Q32851). WGS libraries were made using Nextera XT DNA library preparation kit (Illumina, cat. # FC-131-1096). Then the samples were sequenced using the MiSeq or NextSeq system with 300 cycles, pair-ended. For each strain sequencing, depth was set at approximately 100x coverage. BWA mem version 0.7.12⁵¹ and Pilon v1.23, using default settings,⁵² were used

to align reads against a reference genome assembly and to identify variants, respectively. SNP positions having mapping quality less than 10 ($MQ < 10$) were not considered.

Processing of sequencing data

To build input matrices with gene-barcode information, we created a pipeline that pulls UMI, barcodes 1 and 2. In experiments where multiple samples were pooled for droplet barcoding, a demultiplexing step based on CB1 was performed to parse different samples. Threshold of valid cell barcodes, removal invalid cell barcodes, and final count table generation was performed using UMI-tools⁵³ (<https://doi.org/10.1101/gr.209601.116>). Alignments were performed using BWA⁵¹ and annotation of bam files were done using FeatureCounts.⁵⁴

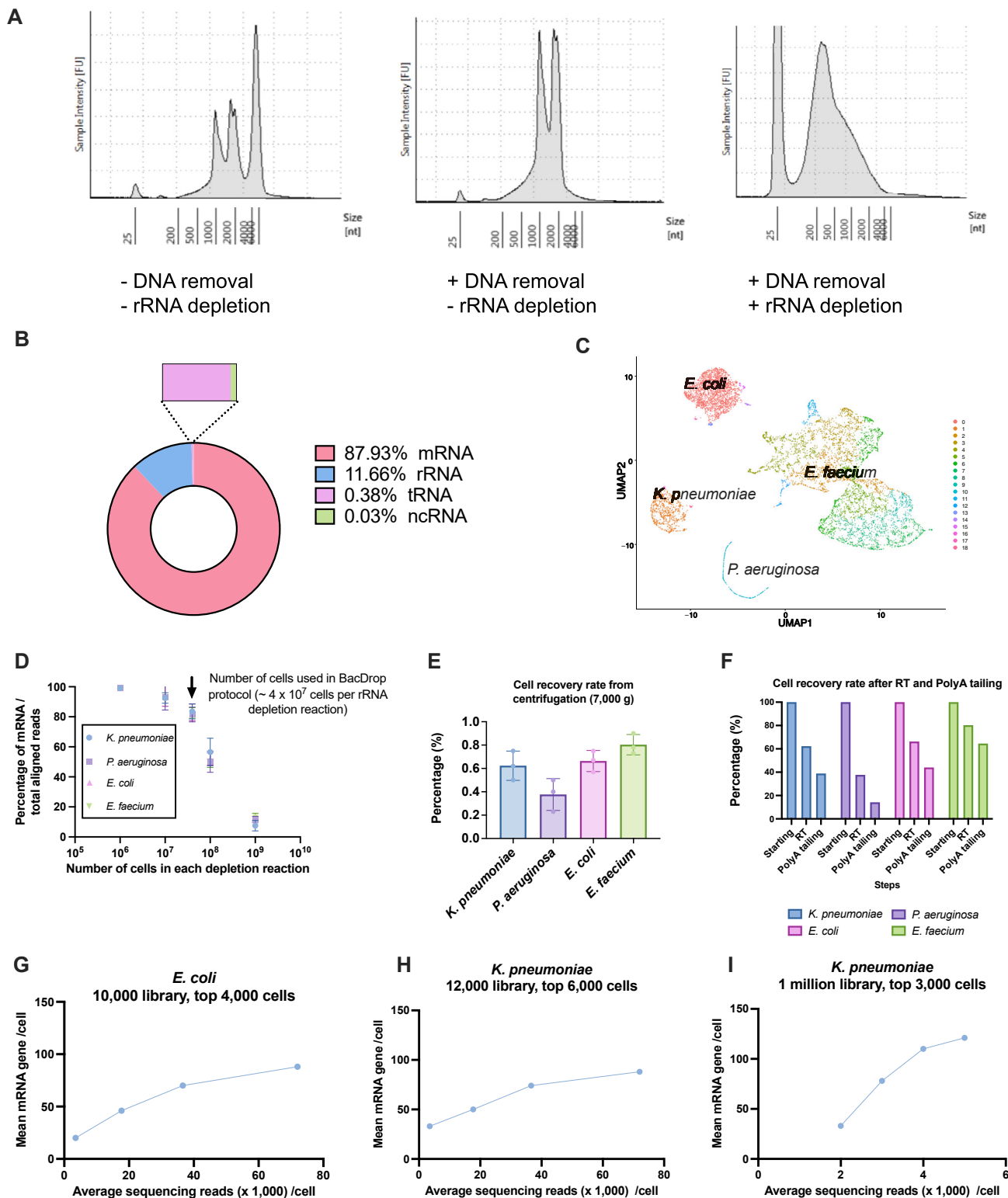
Analysis of BacDrop data

Once the count tables were made, we used the standard workflow of the R package Seurat 3⁵⁵ (<https://doi.org/10.1038/nbt.4096>, <https://doi.org/10.1016/j.cell.2019.05.031>) (v.3.2.2). We excluded genes that were not expressed in any cells in the dataset, and excluded cells that had fewer than 10 or 15 genes detected and cells that had abnormally high numbers of mRNA detected. For experiments done using MGH66, we identified three genes that showed consistently high-level expression in the majority of cells and in various conditions: WP-004174069.1, WP-004174069.1-2, and WP-002920103.1. We use these three genes as internal controls and removed cells with a normalized expression of any of these three genes that is less than 50. The standard Seurat workflow prior to clustering was used including global normalization, feature selection, and scaling of gene expression. We used the top 2000 highly variable genes as input features for clustering analysis and downstream annotation. The Seurat packages FindNeighbors and FindClusters were used for clustering at a resolution of 0.5. Uniform manifold approximation and projection (UMAP) was utilized for visualization of clustering. For marker identification and annotation of clusters, Seurat's FindMarkers tool was used with a requirement that the markers were expressed in 25% of the cells present in the dataset. For some rare population detection, the 25% criterion was removed. From the FindMarkers results, we consider genes with \log_2 fold changes that are greater than 2, and adjusted p-value less than 0.05 as significantly differentiated genes. If a cluster did not contain any genes that pass this threshold, we did not consider this as a significant cluster.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis for BacDrop experiments and bulk RNA-seq experiments were performed in R (v.3.2.2) and detailed analysis was described in relevant results and method sections. For all other experiments, including RT-qPCR, RNA-FISH, FACS, the quantification of mutation frequencies, and persister assays, the statistical analyses were performed in Prism 9 (version 9.3.1). Statistical details, including numbers of samples, types of statistical analyses, and precision measure can be found in figure legends and results sections. No methods were used to determine whether the data met assumptions of the statistical approach.

Supplemental figures



(legend on next page)

Figure S1. Development and performance of BacDrop, related to Figure 2 and STAR Methods

(A) rRNA and gDNA depletion are implemented in BacDrop. The three peaks are 16S rRNA (~1,000 bp), 23S rRNA (~2,000 bp), and gDNA (>4,000 bp).

(B) RNA species and their percentage in a BacDrop library of *K. pneumoniae*.

(C–F) Validation of BacDrop in *K. pneumoniae*, *P. aeruginosa*, *E. coli*, and *E. faecium*. (C) Unsupervised cell clustering separates these four species into distinct clusters.

(D) rRNA-depletion efficiency in these four species. (E and F) Cell losses in the mixed-species experiment are due to different cell recovery rates from centrifugation steps. Experiments in (D) and (E) were repeated 3 times. Error bars are plotted as standard deviation.

(G–I) Assessment of the transcriptome coverage of BacDrop libraries, see [method details](#).

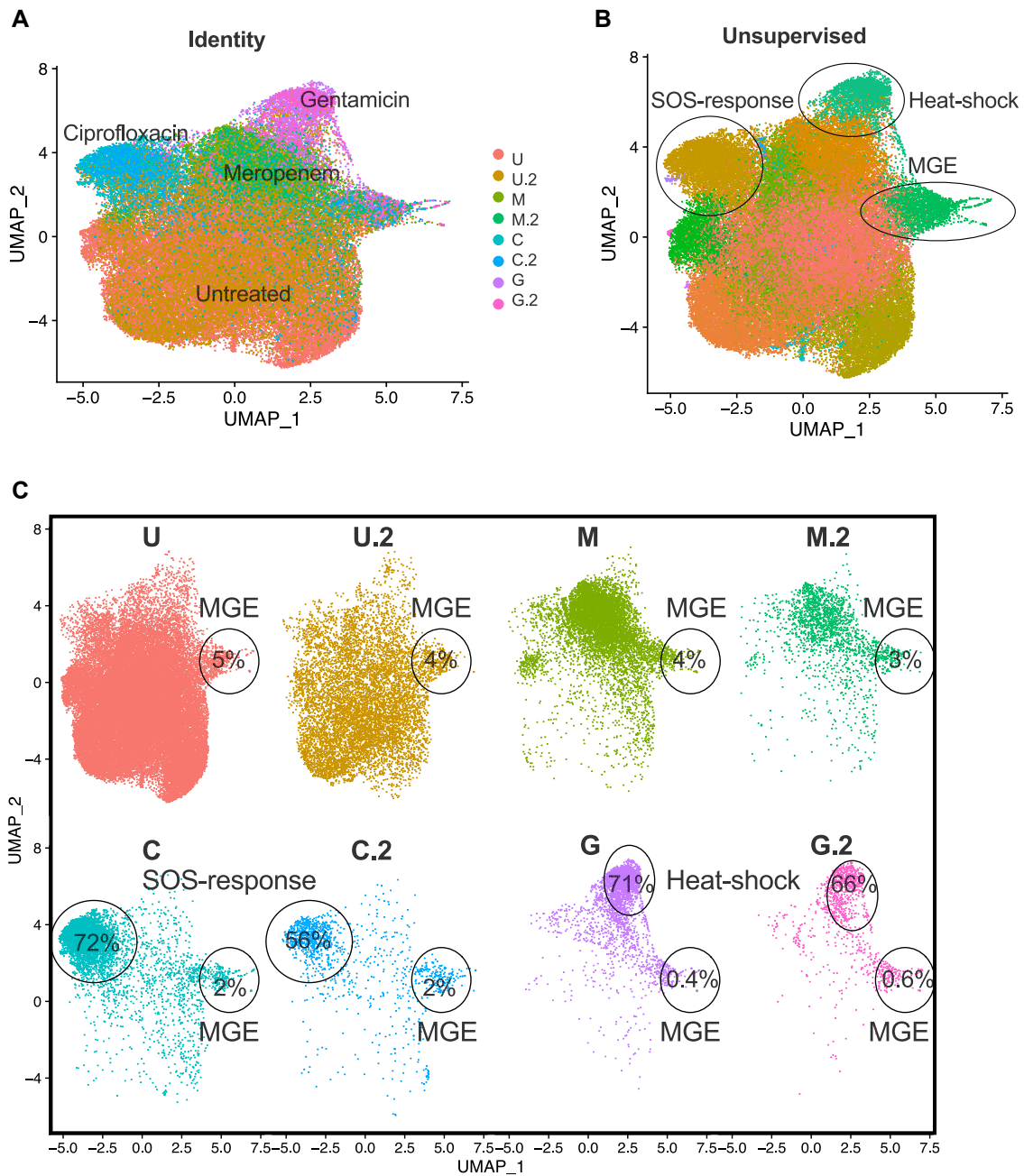


Figure S2. Analysis of MGH66 experiment including the antibiotic-treated samples and untreated samples, related to Figure 3

(A) UMAP plotted based on the original identity of these 8 samples. Cells were separated well based on their treatment.

(B) Unsupervised UMAP showed three clusters with significantly ($p < 0.05$) higher expression of genes in the SOS-response pathway, heat-shock response, and genes encoding an IS903B transposase (MGE).

(C) No strong batch effect was observed between the two biological replicates with the same treatment conditions.

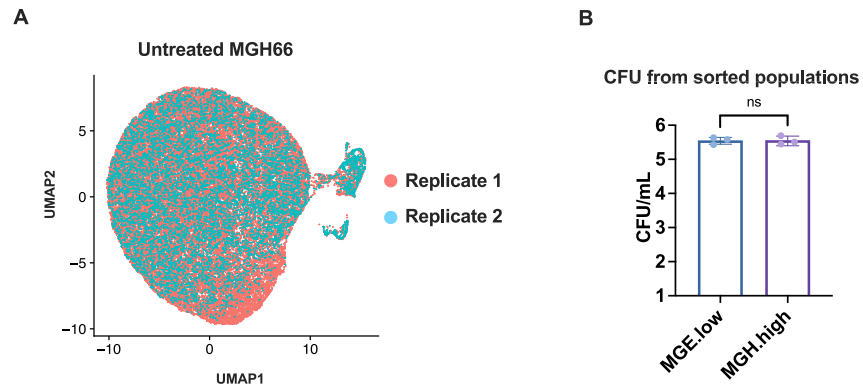
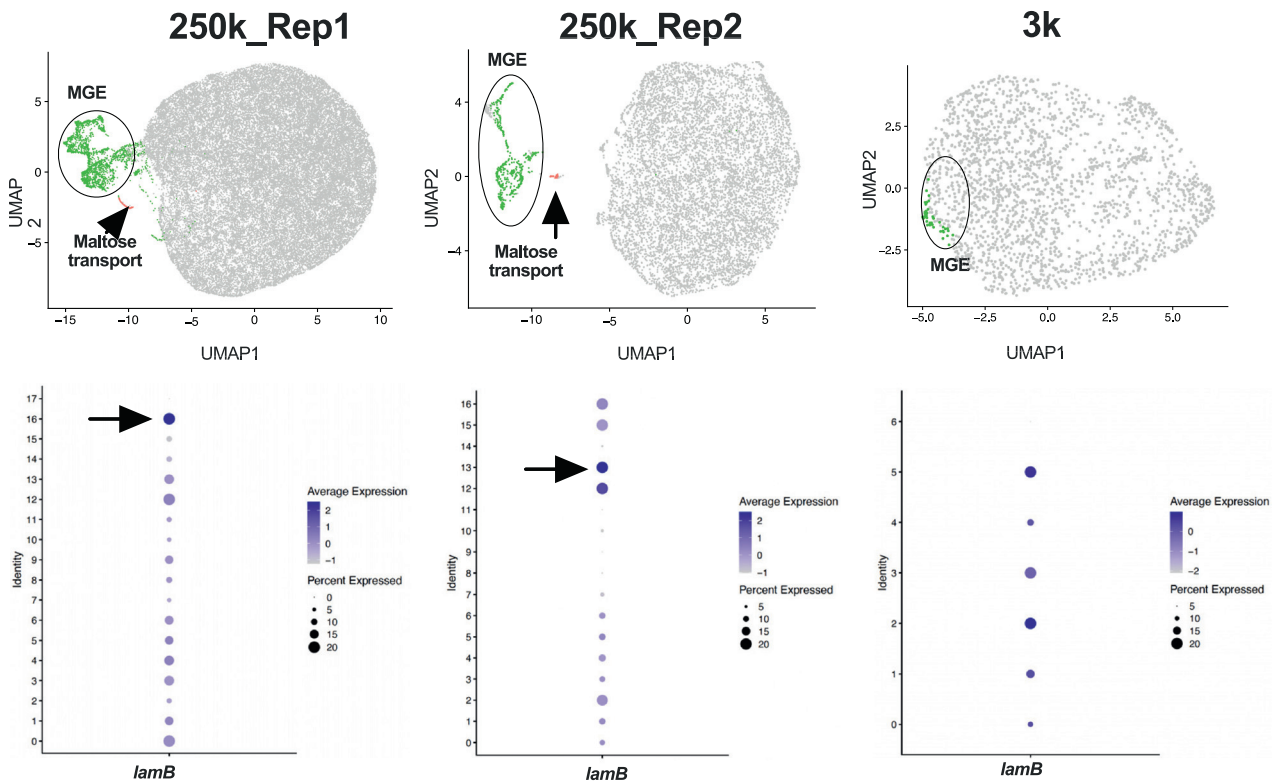


Figure S3. Analysis of the untreated MGH66 samples, related to Figure 4

(A) UMAP plotted based on the original identity of these two samples (replicate 1 and replicate 2).

(B) Cells sorted from MGE.high and MGE.low had similar survival rates ($p = 0.95$).

This experiment was run in triplicates and error bars were plotted as standard deviation. The Student's *t* test was used for statistical analysis ($p = 0.95$).



BacDrop Library	No. of cells in the library	No. of cells passed analysis threshold	Average mRNA genes per cell	MGE subpopulation	Maltose-transport subpopulation
250k_Rep1	~250k	40k	30	5%	0.25%
250k_Rep2	~250k	10k	30	5.50%	0.36%
3k	~3k	2k	85	1.80%	Not detected

Figure S4. A large number of cells is beneficial for the detection of small population, related to Figure 4 and STAR Methods

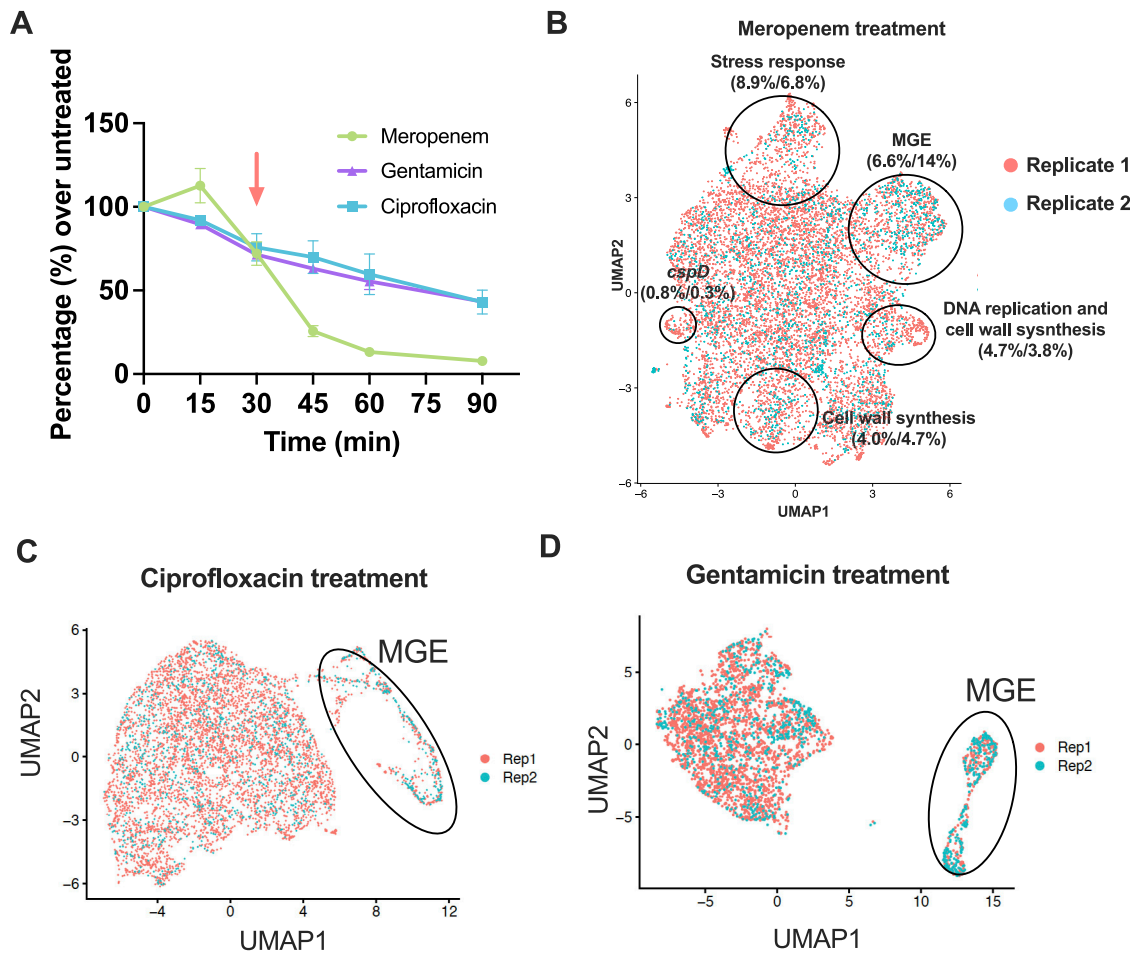


Figure S5. Analysis of the antibiotic-treated samples, related to Figures 3 and 5 and STAR Methods

(A) Killing kinetics of MGH66 treated with meropenem (green), ciprofloxacin (blue) and gentamicin (purple). The cells were collected at 30 min under the treatment (red arrow) for BacDrop experiments. The experiment was repeated 3 times and the error bars are plotted as standard deviation.

(B–D) Analysis of the antibiotic-treated samples based on the replicates. UMAP plotted based on the original identity of two replicates (replicate 1 and replicate 2). (B) Meropenem, (C) ciprofloxacin, (D) gentamicin. For the meropenem-treated samples, fractions of cells in identified clusters were calculated separately from two replicates and indicated in the parentheses.

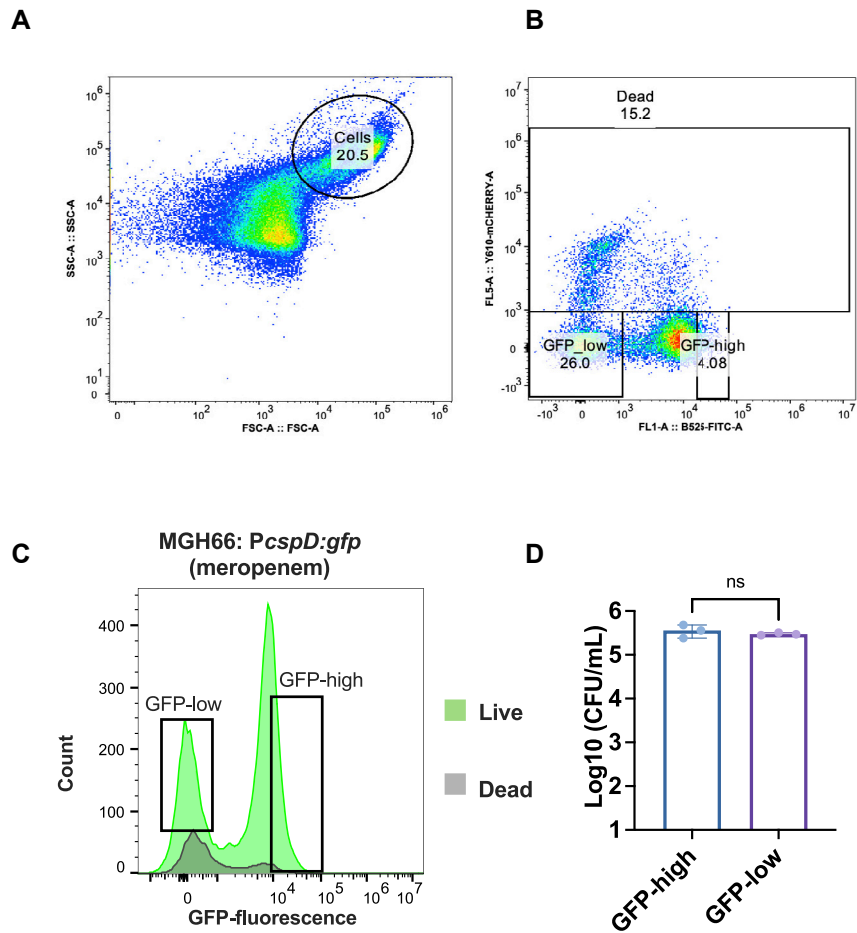


Figure S6. Validate the expression of *cspD* in MGH66: *PcpsD:gfp* strain using FACS sorting coupled with live/dead staining, related Figure 6 and STAR Methods

(A) Gating for cells to exclude cell debris.

(B and C) Dead cells were detected using mCherry red fluorescence, and GFP fluorescence of all cells was detected using green fluorescence. Roughly ~15.2% of cells are dead in the gated population. The gating of GFP-low and GFP-high population is shown.

(D) 10^6 live cells from the GFP-high and GFP-low subpopulations were sorted into 1 mL LB medium, and immediately plated 100 μ L on LB agar plates without any antibiotics.

This experiment was run in triplicates and error bars were plotted as standard deviation. The Student's t test was used for statistical analysis ($p = 0.43$).