

An Improved Bacterial Single-cell RNA-seq Reveals Biofilm Heterogeneity

Xiaodan Yan, Hebin Liao, Chenyi Wang, Chun Huang, Wei Zhang, Chunming Guo, Yingying Pu 

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Medical Research Institute, Wuhan University, Wuhan, China • Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China • Translational Medicine Research Center, North Sichuan Medical College, Nanchong, China • Center for Life Sciences, School of Life Sciences, Yunnan University, Kunming, China • Department of Immunology, Hubei Province Key Laboratory of Allergy and Immunology, State Key Laboratory of Virology and Medical Research Institute, Wuhan University School of Basic Medical Sciences, Wuhan, China

 https://en.wikipedia.org/wiki/Open_access

 Copyright information

Reviewed Preprint

v3 • November 26, 2024

Revised by authors

Reviewed Preprint

v2 • November 15, 2024

Reviewed Preprint

v1 • July 24, 2024

eLife Assessment

This work introduces an **important** new method for depleting ribosomal RNA from bacterial single-cell RNA sequencing libraries, demonstrating its applicability for studying heterogeneity in microbial biofilms. The findings provide **convincing** evidence for a distinct subpopulation of cells at the biofilm base that upregulates PdeI expression. Future studies exploring the functional relationship between PdeI and c-di-GMP levels, along with the roles of co-expressed genes within the same cluster, could further enhance the depth and impact of these conclusions.

<https://doi.org/10.7554/eLife.97543.3.sa3>

Abstract

In contrast to mammalian cells, bacterial cells lack mRNA polyadenylated tails, presenting a hurdle in isolating mRNA amidst the prevalent rRNA during single-cell RNA-seq. This study introduces a novel method, Ribosomal RNA-derived cDNA Depletion (RiboD), seamlessly integrated into the PETRI-seq technique, yielding RiboD-PETRI. This innovative approach offers a cost-effective, equipment-free, and high-throughput solution for bacterial single-cell RNA sequencing (scRNA-seq). By efficiently eliminating rRNA reads and substantially enhancing mRNA detection rates (up to 92%), our method enables precise exploration of bacterial population heterogeneity. Applying RiboD-PETRI to investigate biofilm heterogeneity, distinctive subpopulations marked by unique genes within biofilms were successfully identified. Notably, PdeI, a marker for the cell-surface attachment subpopulation, was observed to elevate cyclic diguanylate (c-di-GMP) levels, promoting persister cell formation. Thus, we address a persistent challenge in bacterial single-cell RNA-seq regarding rRNA abundance, exemplifying the utility of this method in exploring biofilm heterogeneity. Our method effectively tackles a long-standing issue in bacterial scRNA-seq: the overwhelming abundance of rRNA. This advancement significantly enhances our ability to investigate the intricate heterogeneity within biofilms at unprecedented resolution.

Introduction

Biofilms, contributing to approximately 80% of chronic and recurrent microbial infections in the human body¹, are complex microbial ecosystems characterized by a diverse array of bacterial cells existing in various physiological states^{1–3}. This heterogeneity within biofilms is influenced by multiple factors, including bacterial interactions (such as competition, symbiosis, and parasitism) that contribute to stable community structures^{4,5}, environmental factors (like host environment, aquatic conditions, and nutrient concentrations), and spatial organization. Bacteria occupying different positions within the biofilm perform distinct roles⁶: some mediate material exchange between cells and with the extracellular matrix, others facilitate complex communication systems between bacteria and with the host through signaling molecules, while certain bacteria participate in various energy conversion processes. This intricate division of labor not only contributes to bacterial heterogeneity but also enhances the biofilm's overall resistance to various stresses⁷. The resulting cellular and functional diversity reflects the complex nature of the biofilm ecosystem, allowing it to adapt and persist in challenging environments. However, the study of biofilms faces significant limitations, primarily stemming from challenges in investigating heterogeneity within a bacterial population^{8,9}. Single-cell RNA-seq emerges as a promising avenue for addressing this heterogeneity^{10–18}. Expanding on established protocols for cell fixation and permeabilization which facilitate in-cell barcoding while avoiding cell lysis, combinatorial barcoding-based bacterial scRNA-seq techniques, such as Prokaryotic Expression profiling by Tagging RNA in situ and sequencing (PETRI-seq)¹² and Microbial Split-Pool Ligation Transcriptomics (microSPLIT)¹⁴, have been developed. Nevertheless, these methods encounter challenges in terms of low transcript recovery rates due to overwhelmingly abundant rRNA, restricting the comprehensive analysis of within-population heterogeneity. In comparison to mammalian cells¹⁹, the absence of mRNA polyadenylated tails in bacteria necessitates an alternative approach for isolating mRNA (~5%) from the significantly more abundant rRNA (~95%). Here, by integrating a Ribosomal RNA derived cDNA Depletion protocol (RiboD) into a PETRI-seq, we developed RiboD-PETRI-seq that efficiently eliminates rRNA reads, thereby significantly improving mRNA detection rates and enabling exploration of within-population heterogeneity.

Results

In the RiboD protocol, we designed a set of probe primers that spans all regions of the bacterial rRNA sequence (Table S1). The core principle behind our probe design is twofold: the 3'-end of the probes are reverse complementary to the r-cDNA sequences, allowing for specific recognition of r-cDNA, while the 5'-end complements a biotin-labeled universal primer. This design enables the probes to be bound to magnetic beads, facilitating the separation of r-cDNA-probe-bead complexes from the rest of the library. Following template switching and RNaseH treatment on the barcoded cDNA from lysed cells to eliminate hybridized RNA, the library of probe primers and biotin-labeled universal primers is introduced to facilitate adequate hybridization. Pre-treated Streptavidin magnetic beads are then added to the hybridized rRNA-derived cDNA. The mRNA-derived cDNA remains in the supernatant and is collected for subsequent library construction and sequencing (Fig. 1A). To assess the efficiency of single-cell capture in RiboD-PETRI, we calculated the multiplet frequency¹² using a Poisson distribution based on our sequencing results (see details in Methods). The multiplet frequency for RiboD-PETRI ranges from 1.16% to 3.35% (Table S2), indicating the technique's capability to effectively capture transcriptomes at the single-cell level.

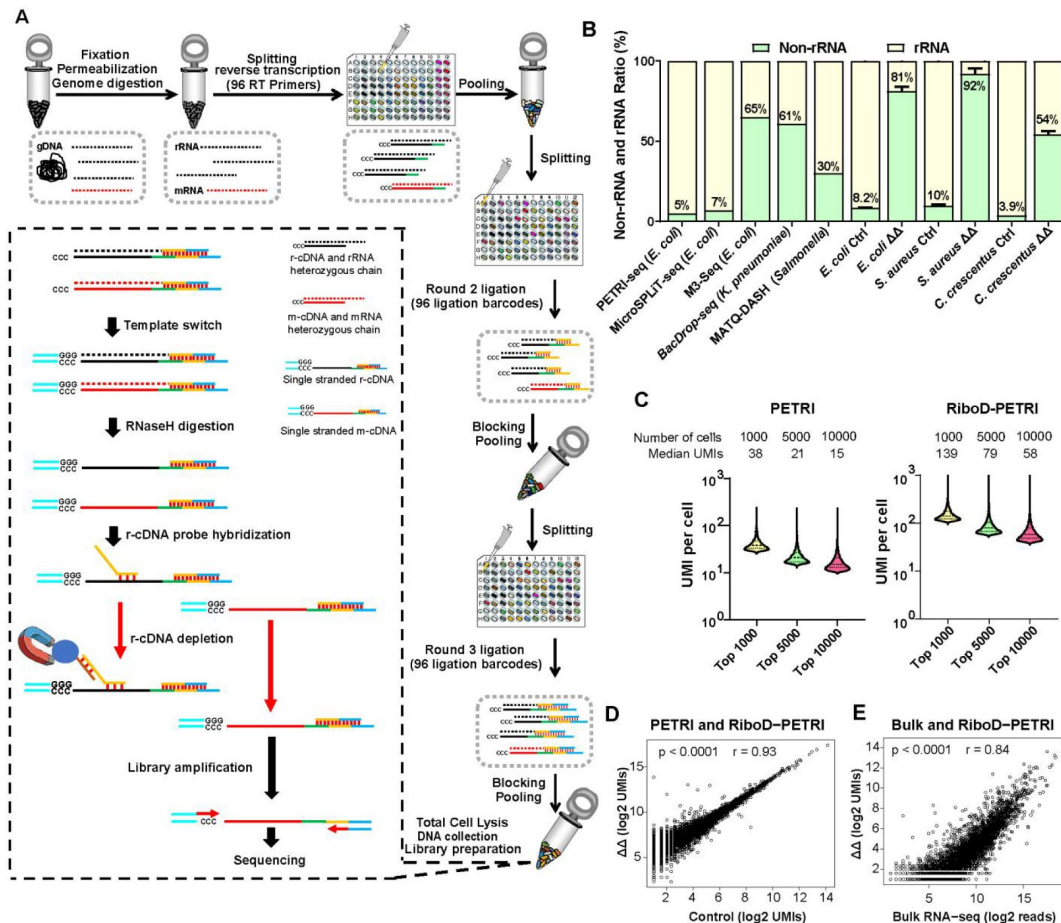


Figure 1.

Development of RiboD-PETRI and validation of its technical performance in studying population heterogeneity

(A) Graphic summary of the RiboD-PETRI method illustrating the incorporation of RiboD after cell pooling and lysis in PETRI-seq. The RiboD protocol is represented by the dashed-line box. In this box, firstly, we perform template-switching oligonucleotides (TSOs) in the mixture of heterozygous chain, then we remove the RNA strand using RNaseH, at this point the system contains r-cDNA and m-cDNA single-stranded mixture. Then we add the r-cDNA probe, which specifically binds to the r-cDNA. The probes are then bound to magnetic beads, allowing the r-cDNA-probe-bead complexes to be separated from the rest of the library. And then we remove the r-cDNA that is attached to the probe by Streptavidin magnetic beads. We then performed amplification of the libraries and sent them for sequencing. We designed separate probe sets for *E. coli*, *C. crescentus*, and *S. aureus*. Each set was specifically constructed to be reverse complementary to the r-cDNA sequences of its respective bacterial species. This species-specific approach ensures high efficiency and specificity in rRNA depletion for each organism. (B) Comparison of non-rRNA (tRNA, mRNA and other non-rRNA) and rRNA UMI counts ratio among different bacterial scRNA-seq methods. Data from PETRI-seq (*E. coli*), MicroSPLIT-seq (*E. coli*), M3-seq (*E. coli*) cited from previous studies. Error bars represent standard deviations of biological replicates. The “ΔΔ” label represents the RiboD-PETRI protocol; The “Ctrl” label represents the classic PETRI-seq protocol we performed. (C) Comparison of UMI counts per cell between RiboD-PETRI (Table S7) and PETRI (Table S8) at the same unsaturated sequencing depth. (D) Assessment of the effect of rRNA depletion on transcriptional profiles. The Pearson correlation coefficient (r) of UMI counts per gene ($\log_2$ UMIs) between RiboD-PETRI (Table S7) and PETRI (Table S9) was calculated for 3790 out of 4141 total genes, excluding those with zero counts in either library. Each point represents a gene. (E) Evaluation of the correlation between RiboD-PETRI (Table S7) data and bulk RNA-seq (Table S10) results. The Pearson correlation coefficient (r) of UMI counts per gene ($\log_2$ UMIs) among RiboD-PETRI data and the reads per gene ($\log_2$ reads) of bulk RNA-seq data was calculated for 3814 out of 4141 total genes, excluding those with zero counts in either library. Each point represents a gene. All data presented in Fig.1C, D, E were from our own sequencing experiments.

To assess the performance of RiboD-PETRI, we designed a comprehensive assessment of rRNA depletion efficiency under diverse physiological conditions, specifically contrasting exponential and stationary phases. This approach allows us to understand how these different growth states impact rRNA depletion efficacy. Additionally, we included a variety of bacterial species, encompassing both gram-negative and gram-positive organisms, to ensure that our findings are broadly applicable across different types of bacteria. By incorporating these variables, we aim to provide insights into the robustness and reliability of the RiboD-PETRI method in various biological contexts. The results highlight a substantial enhancement in rRNA-derived cDNA depletion, with mRNA ratio increases from 8.2% (Ctrl, the PETRI-seq we performed) to 81% (RiboD-PETRI) for *E. coli* from exponential phase, from 10% (Ctrl) to 92% for *S. aureus* from stationary phase, and from 3.9% (Ctrl) to 54% for *C. crescentus* from exponential phase (Fig. 1B; Table S3). Additionally, we compared our findings with other reported methods (Fig. 1B; Table S4). The original PETRI-seq¹² protocol, which does not include an rRNA depletion step, exhibited an mRNA detection rate of approximately 5%. The MicroSPLiT-seq¹⁴ method, which utilizes Poly A Polymerase for mRNA enrichment, achieved a detection rate of 7%. Similarly, M3-seq¹⁶ and BacDrop-seq¹⁵, which employ RNase H to digest rRNA post-DNA probe hybridization in cells, reported mRNA detection rates of 65% and 61%, respectively. MATQ-DASH²⁰, which utilizes Cas9-mediated targeted rRNA depletion, yielded a detection rate of 30%. smRandom-seq utilizes a CRISPR-based rRNA depletion technique, reduced the rRNA proportion from 83% to 32%, increasing the mRNA proportion from 16% to 63%²¹. BaSSSh-seq's employs a rational probe design for efficient rRNA depletion, though specific efficiency was not reported²². Among these, RiboD-PETRI demonstrated superior performance in mRNA detection while requiring the least sequencing depth. With equivalent sequencing depth, RiboD-PETRI demonstrates a significantly enhanced Unique Molecular Identifier (UMI) counts detection rate compared to PETRI-seq alone (Fig. 1C). This method recovered approximately 20,175 cells (92.6% recovery rate) with ≥ 15 UMIs per cell with a median UMI count of 42 per cell, which was significantly higher than PETRI-seq's recovery rate of 17.9% with a median UMI count of 20 per cell (Figure S1A, B), indicating the number of detected mRNA per cell increased prominently. Notably, this enhancement was achieved while maintaining mRNA profiles consistent with non depleted samples ($r = 0.93$; Fig. 1D) and show a significant correlation with profiles from the traditional bulk RNA-seq method ($r = 0.84$; Fig. 1E).

We subsequently investigated the transcriptome coverage of RiboD-PETRI across different physiological states and bacterial species. For exponential phase *E. coli* cells, we sequenced a library with 60,000 cells, recovering approximately 30,004 cells (50% recovery), each with ≥ 15 UMIs (Fig. 2A and S1C). This analysis revealed 99.86% transcriptome-wide gene coverage across the cell population. The method achieved an average of 128.8 UMIs per single cell, with a median UMI count of 102 per cell. Further examination of high-quality cells showed varying levels of detection: the top 1,000, 5,000, and 10,000 cells exhibited median UMI counts of 462, 259, and 193, respectively (Fig. 2B), and median gene detection of 362, 236, and 188, respectively (Fig. 2C). These high-performing cells demonstrate the upper limits of the method's capabilities. For stationary phase *S. aureus* cells, we sequenced a library with 30,000 cells, recovering approximately 9,982 cells (33.3% recovery), each with ≥ 15 UMIs (Fig. S2A). Analysis showed 99.96% transcriptome-wide gene coverage across the cell population. At the single-cell level, we observed an average of 153.8 UMIs and a median of 142 UMIs. Top high-quality cells exhibited the following median UMI counts: 378 (top 1,000 cells), 207 (top 5,000 cells), and 167 (top 8,000 cells) (Fig. S2B). These cells also demonstrated median gene detection of 308, 194, and 158 genes, respectively (Fig. S2C). For exponential phase *C. crescentus* cells, we sequenced a library with 30,000 cells, recovering approximately 13,897 cells (46.3% recovery), each with ≥ 15 UMIs (Fig. S2G). Analysis showed 99.64% transcriptome-wide gene coverage across the cell population. At the single-cell level, we observed an average of 439.7 UMIs and a median of 182 UMIs. Top high-quality cells demonstrated the following median UMI counts: 2,190 (top 1,000 cells), 662 (top 5,000 cells), and 225 (top 10,000 cells) (Fig. S2H). These cells also exhibited median gene detection of

1,262, 529, and 219 genes, respectively (**Fig. S2I**). These results underscore RiboD-PETRI's ability to capture a wide range of transcripts across varying cell qualities and species, providing a comprehensive view of gene expression at the single-cell level.

Our results affirm RiboD-PETRI's reliability in capturing the bacterial single-cell transcriptome, providing ample coverage and sensitivity for various species. To provide a thorough evaluation of our sequencing depth and library quality, we performed sequencing saturation analysis on our sequencing samples. The findings reveal that our sequencing saturation is greater than 90% (**Fig. S1D-F**), indicating that our sequencing depth is sufficient to capture the diversity of most transcripts.

We further investigated its ability to consistently identify within-population heterogeneity across different bacterial species and growth conditions. In the exponential phase of *E. coli*, we recovered 1,464 cells and identified three major subpopulations (**Fig. 2D**), with 17 cells (1.2%) in a unique subpopulation characterized by pentose and glucuronate interconversions (**Fig. 2E, F**) and the marker genes of cluster 2 included *yfgM*, *glpG*, *scpA*, *elyC*, *ptsA*, *carB*, *actP* and *pgpB* (**Fig. 2G**). For the expression levels of marker gene shown in **Fig. 2E**, violin plots have been created to offer a more comprehensive view of the distribution across different cell populations (**Fig. S4**). In stationary phase *S. aureus* cells, we recovered 9,386 cells and found six major subpopulations (**Fig. S2D**), with 437 cells (4.7%) in a distinct subpopulation named cluster 4. The marker genes of cluster 4 included KQ76-13335, KQ76-00740 and KQ76-11725 (**Fig. S5**). In the stationary phase of *C. crescentus* cells, we recovered 5728 cells and identified four major subpopulations (**Fig. S2J**), with 603 cells (10.5%) in a unique subpopulation named cluster 3. The marker genes of cluster 3 included CCNA-00259, CCNA-03402, CCNA-02361 and CCNA-03119 (**Fig. S6**). These findings highlight RiboD-PETRI's consistent ability to unveil within-population heterogeneity across different cell physiology and bacterial species (**Fig. 2H, I** and **S2E, F, K, L**), crucial for understanding bacterial population complexity. While RiboDPETRI consistently detects potential heterogeneity, further experimental validation would be required to confirm the biological significance of the observations.

We next focused on exploring biological heterogeneity of a biofilm at the early stage of development by utilizing the static biofilm system²³. *E. coli* cells were cultured in microtiter dishes overnight, adhered cells were fixed for RiboD-PETRI processing in duplicate experiments. For these two replicates, we sequenced libraries containing 20,000 and 40,000 cells, recovering 5,244 and 11,344 cells, which corresponded to recovery rates of 26% and 28%, respectively. The correlations between detected reads and unique molecular identifiers (UMIs) were found to be 0.87 and 0.90 for the two replicates, respectively (**Fig. S3A**). The median UMI counts for the recovered cells were 34 and 52 (**Fig. 3A**). After screening, the final datasets comprised 1,621 and 3,999 cells for the two replicates, respectively. While no significant batch effects were observed, we applied batch correction as a precautionary measure (**Fig. S3B-D**). In replicate 1, each cell was sequenced with an average of 1,563 reads, while in replicate 2, the average was 2,034 reads (**Table S5**), yielding median UMI counts of 283.5 and 239 per cell, respectively (**Fig. 3B**). For gene detection, the median counts were 219 and 193 genes per cell for the respective replicates (**Fig. 3C**). Additionally, UMAP visualization was employed to illustrate the distribution of cellular UMI numbers, revealing heterogeneity among populations that was independent of UMI counts (**Fig. S3E, F**). Unsupervised clustering analysis identified four major subpopulations in each replicate, with a consistently identified rare subpopulation (2.6%/2.1%) as cluster 2, driven by cell envelope genes (**Fig. 3D-F**). Marker genes for this cluster included *yffO*, *lptE*, *rdgB*, *pdeI*, *sstT*, *fixA*, *yjjG*, *rlmI*, *accC* and *yaiA* (**Fig. 3G, H** and **S7**).

PdeI, identified among marker genes, was predicted as a phosphodiesterase enzyme hydrolyzing c-di-GMP, a vital bacterial second messenger^{24,25} (**Fig. 4A, B**). However, our comprehensive structural analysis revealed a more complex and novel role for PdeI. While PdeI contains an intact EAL domain typically associated with c-di-GMP degradation, it also possesses a divergent GGDEF

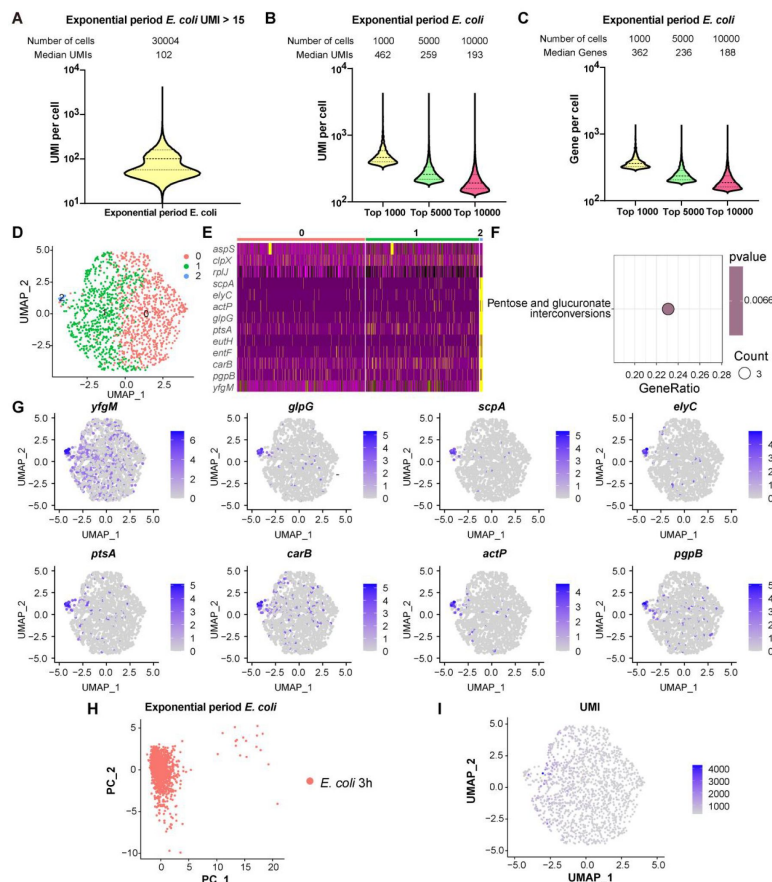


Figure 2.

Comprehensive Analysis of single-cell mRNA Transcriptomic Profiles in Exponential Phase *E. coli* using RiboD-PETRI

(A) The number of UMIs detected per cell in recovered cells in exponential period *E. coli* (≥ 15 UMIs/cell). The cells are ranked from highest to lowest based on the number of detected UMIs, and cells with ≥ 15 UMIs are selected for plotting. The median number of UMIs is calculated for these selected cells. (B) Distribution of mRNA UMIs captured per cell in RiboD-PETRI data of exponential period *E. coli*, presented as violin plots showing the upper quartile, median, and lower quartile lines. The cells are ranked from highest to lowest based on the number of UMIs detected. Then, specific numbers of cells (indicated above the panel) are selected for plotting. The median number of UMIs is calculated for these selected cells. (C) The number of genes detected per cell in exponential period *E. coli*. The cells are ranked from highest to lowest based on the number of genes detected. Then, specific numbers of cells (indicated above the panel) are selected for plotting. The median number of genes is calculated for these selected cells. (D) Uniform Manifold Approximation and Projection (UMAP) visualization of *E. coli* bacteria during the exponential phase. Data were filtered for cells with UMIs between 200 and 5,000, resulting in 1,464 cells. Each dot represents a cell. (E) Heatmap illustrating the normalized gene expression levels of marker genes in different clusters of exponential period *E. coli*. Marker genes with relatively high expression levels are depicted in yellow, while lower expression levels are shown in purple. Each row represents a gene, and each column represents a cell. (F) Functional enrichment analysis of marker genes of exponential period *E. coli* in cluster 2. Marker genes were selected based on screening criteria of p-value < 0.001 and $\log_2$ fold change (FC) > 0.2 . The color blocks in these figures represent the p-values of the data points. The color scale ranges from red to blue. Red colors indicate smaller p-values, suggesting higher statistical significance and more reliable results. Blue colors indicate larger p-values, suggesting lower statistical significance and less reliable results. Count is the number of genes enriched into this pathway. (G) Expression levels of marker genes in cluster 2 during the 3-hour exponential period of *E. coli* overlaid on the UMAP plot. Cells with high expression levels are depicted in blue. Marker genes were selected based on a p-value greater than 0.001 and a $\log_2$ FC greater than 3. (H) Principal Component Analysis (PCA) performed on screened data of exponential phase *E. coli*. The resulting scatterplots show heterogeneity among the populations, with each point representing a cell. (I) Distribution of UMIs on the UMAP results for exponential phase *E. coli*. UMAP results reveal heterogeneity among populations, with each point representing a cell and color shading indicating UMI counts (Table S11).

domain, generally linked to c-di-GMP synthesis (Fig. S8). This dual-domain architecture suggested potential complex regulatory roles. To validate PdeI's function, we created a PdeI-BFP fusion construct under the native pdeI promoter, integrated with a ratiometric c-di-GMP sensing system²⁶ in *E. coli*. Confocal microscopy revealed PdeI as a membrane protein (Fig. 4C). Single-cell level monitoring showed cell-to-cell variability in c-di-GMP levels and PdeI expression, with a positive correlation observed (Fig. 4D), indicating PdeI upregulated c-di-GMP synthesis rather than degradation. This finding was confirmed by HPLC LC-MS/MS, which showed an approximately elevenfold increase in c-di-GMP concentration in the PdeI overexpression strain compared to the control strain (Fig. 4E). These results align with previous studies showing that a point mutation (G412S) in PdeI's divergent GGDEF domain in a strain lacking PdeH, the major phosphodiesterase in *E. coli*, resulted in decreased c-di-GMP levels²⁷. Our additional experiments with a PdeI(G412S)-BFP mutation strain showed constant c-di-GMP levels despite increasing BFP fluorescence, serving as a proxy for PdeI(G412S) expression levels (Fig. 4D). These results, combined with the presence of a CHASE (cyclases/histidine kinase-associated sensory) domain in PdeI, strongly suggest that PdeI functions as a membrane-associated sensor that integrates environmental signals with c-di-GMP production under complex regulatory mechanisms. This discovery challenges the initial prediction of PdeI as solely a phosphodiesterase and highlights its novel role as a c-di-GMP synthetase, contributing significantly to our understanding of bacterial signaling pathways. It's worth noting that while the other marker genes in this cluster are co-expressed, our analysis indicates that they do not have a significant impact on biofilm formation or a direct relationship with c-di-GMP or PdeI.

Confocal laser scanning microscopy provided further insights into the spatial distribution of PdeI-positive cells within the biofilm structure. In the PdeI-BFP fusion strain, PdeI-BFP positive cells, characterized by elevated c-di-GMP levels, were predominantly located at the bottom of the static biofilm (Fig. 4F). This localization corresponds to the region of cell-surface attachment, aligning with our hypothesis that PdeI functions as a membrane-associated sensor integrating environmental signals with c-di-GMP production through complex regulatory mechanisms²⁸. In contrast, in the control strain where BFP was expressed alone under arabinose-induced promoter, BFP-positive cells were observed to be distributed throughout the entire biofilm community (Fig. 4G). This distinct spatial distribution pattern between the PdeI-BFP fusion and the BFP-only control strains provides compelling evidence for PdeI's specific role in biofilm formation, particularly at the biofilm-surface interface. The concentration of PdeI-positive cells at the bottom of the biofilm suggests that PdeI may be especially crucial in the initial stages of biofilm formation, potentially responding to surface-associated cues to modulate c-di-GMP levels and promote attachment. The uniform distribution of BFP in the control strains suggests that the localization of PdeI-BFP is not affected by BFP labelling. These observations further underscore the complex and nuanced role of PdeI in bacterial signaling and biofilm development, highlighting the importance of considering cell-to-cell heterogeneity in understanding the function of regulatory proteins in microbial communities.

To investigate the association of the PdeI-high cluster with bacterial drug tolerance in the early stages of biofilm development, we isolated PdeI-high cells using flow cytometry (Fig. 4H, I) and subjected them to an ampicillin antibiotic killing assay to determine their persister frequency. Our results revealed that the PdeI-high population produced a significantly higher ratio of persister cells (~7.3%) compared to the whole biofilm population (~0.6%). Notably, cells expressing high levels of BFP alone or PdeI(G412S)-BFP showed no increase in persister ratios (Fig. 4J). This finding suggests that the increased persistence is specifically linked to PdeI activity. Time-lapse imaging during the antibiotic killing process consistently demonstrated that persisters primarily originated from PdeI-GFP positive cells (Fig. 4K and Movie S1). These PdeI-GFP positive cells, displaying characteristics of dormancy, survived ampicillin treatment for 6 hours without visible growth or division. Upon antibiotic removal and replacement with fresh growth medium, the PdeI-GFP positive persister cells resumed activity, elongating, dividing, and forming new microcolonies (Fig. 4K and Movie S1). This dynamic behavior provides direct visual evidence

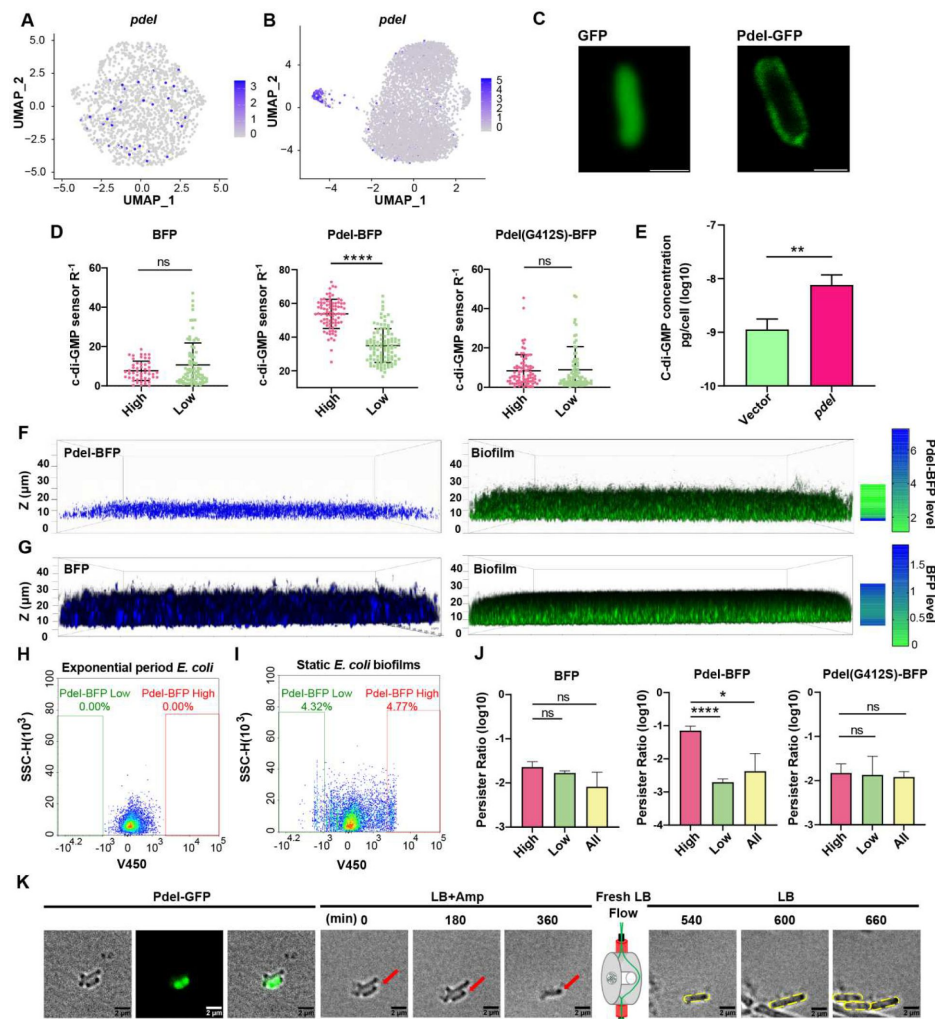


Figure 4.

Functional Investigation of Marker Gene *PdeI* in Static *E. coli* Biofilm

(A & B) UMAP plots showing the distribution of *pdeI* in single-cell data of exponential period *E. coli* (A) and static *E. coli* biofilm (B). Each dot represents a cell colored by normalized expression levels of genes. (C) Subcellular localization of PdeI-GFP and GFP. Scale bar, 1 μ m. (D) c-di-GMP levels (R^{-1} score) in *E. coli* cells with different BFP, PdeI-BFP, PdeI(G412S)-BFP expression levels (low or high), under the control of the *pdeI* native promoter, in static *E. coli* biofilm. c-di-GMP levels are measured using the c-di-GMP sensor system integrated into *E. coli* cells. R^{-1} score was determined using the fluorescent intensity of mVenusNB and mScarlet-I in the system. The fluorescent intensity is measured by flow cytometry. (E) Determination of cellular concentrations of c-di-GMP by HPLC-MS/MS in cells overexpressing PdeI under the control of arabinose promoter, with 0.002% arabinose induction for 2 h (n=3). (F & G) Localization of PdeI-high cells in the biofilm matrix. Cells expressing PdeI-BFP under the control of the *pdeI* native promoter were grown in a glass-bottom cell culture dish and stained with SYTO™ 24 for bacterial DNA. Cells expressing BFP under the control of arabinose promoter, with 0.00001% arabinose induction for 24h in a glass-bottom cell culture dish and stained with SYTO™ 24 for bacterial DNA. (H & I) Heterogeneous expression of PdeI in single-cell data of exponential period *E. coli* (H) and *E. coli* in static *E. coli* biofilm (*E. coli* 24h static culture) (I). Biofilm cells with high or low expression levels of PdeI-BFP were sorted by flow cytometry. (J) Persister counting assay using 150 μ g/ml ampicillin on cells with high or low expression levels of BFP, PdeI-BFP and PdeI(G412S)-BFP from static *E. coli* biofilm, sorted by flow cytometry. These strains were under the control of the *pdeI* native promoter. (K) Time-lapse images of the persister assay observed under a microscope. Static biofilm cells of the PdeI-GFP strain were spotted on a gel pad and treated with 150 μ g/ml ampicillin in LB broth. Images were captured over 6 hours at 37 $^{\circ}$ C, followed by the replacement of fresh LB broth to allow persister cell resuscitation. Scale bar, 2 μ m. Error bars represent standard deviations of biological replicates. Significance was ascertained by unpaired Student's *t*-test; Statistical significance is denoted as **P* < 0.05, ***P* < 0.01, ****P* < 0.001, and *****P* < 0.0001.

of the persister phenotype associated with PdeI-high cells. These findings strongly suggest that c-di-GMP, a molecule whose intracellular levels are upregulated by PdeI, plays a significant role in generating a persister subpopulation during the early stages of biofilm development. The mechanism by which elevated c-di-GMP levels contribute to antibiotic tolerance may involve modulation of cellular metabolism or activation of stress response pathways, leading to a state of dormancy that enables survival under antibiotic stress. This discovery not only enhances our understanding of the link between biofilm formation and antibiotic tolerance but also identifies PdeI as a potential target for strategies aimed at combating persistent bacterial infections.

Discussion

In this study, we introduce RiboD-PETRI, an enhanced bacterial scRNA-seq method that offers a cost-effective (Table S6), equipment-free, and high-throughput solution. By incorporating a probe hybridization-based rRNA-derived cDNA depletion protocol, our approach efficiently removes rRNA reads and significantly improves mRNA detection rates, enabling a more comprehensive exploration of within-population heterogeneity. At \$0.0049 per cell, RiboD-PETRI is substantially more economical than the original PETRI-seq (\$0.056 per cell), making it an attractive option for budget-conscious researchers. The method demonstrates improved mRNA detection, versatility across various bacterial species and growth conditions, preservation of transcriptome profiles consistent with non-depleted samples ($r = 0.93$) and traditional bulk RNA-seq methods ($r = 0.84$), high transcriptome coverage (>99%), and robust single-cell resolution with median UMI counts ranging from 102 to 182 per cell across different species and conditions.

The application of RiboD-PETRI to investigate biofilm heterogeneity exemplifies its potential for exploring complex biological systems. Our analysis of early-stage biofilm development uncovered a rare subpopulation (2.1-2.6%) characterized by cell envelope genes, including the previously uncharacterized gene *pdeI*. Further investigation revealed PdeI's novel role as a c-di-GMP synthetase rather than a phosphodiesterase, challenging initial predictions and contributing significantly to our understanding of bacterial signaling pathways. Moreover, we demonstrated that PdeI-high cells exhibit increased antibiotic tolerance, with a significantly higher proportion of persister cells compared to the general biofilm population. This finding establishes a link between elevated c-di-GMP levels, regulated by PdeI, and the generation of antibiotic-tolerant subpopulations during early biofilm development. The spatial distribution of PdeI-positive cells at the bottom of the static biofilm, corresponding to the cell-surface attachment region, supports our hypothesis that PdeI functions as a membrane-associated sensor integrating environmental signals with c-di-GMP production. This localization pattern suggests PdeI's crucial role in the initial stages of biofilm formation, potentially responding to surface-associated cues to modulate c-di-GMP levels and promote attachment. While other marker genes in this cluster are co-expressed, our analysis indicates they do not significantly impact biofilm formation or directly relate to c-di-GMP or PdeI.

In conclusion, RiboD-PETRI represents a significant advancement in bacterial scRNA-seq methodology. Its ability to uncover hidden variations within bacterial populations, as demonstrated in our biofilm analysis, underscores its potential impact on advancing our understanding of microbial behavior and population dynamics. By providing a cost-effective and efficient tool for exploring bacterial heterogeneity, RiboD-PETRI opens new avenues for research in microbiology, potentially leading to novel insights into antibiotic resistance, biofilm formation, and other critical areas of bacterial biology.

Materials and Methods

Resource availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Yingying Pu (yingyingpu@whu.edu.cn).

Materials availability

Plasmids generated in this study are available from the lead contact upon request.

Data and code availability

Sequencing data, the processing data and original code have been deposited to GEO repository (GSE260458).

Bacterial strains and growth conditions

The bacterial strains used in this study included *Escherichia coli* strains MG1655, *Caulobacter crescentus* NA1000 and *Staphylococcus aureus* strain ATCC. *E. coli* cultures were grown in Luria Broth (LB) medium. For the biofilm setup, bacterial cultures were grown overnight. The next day, we diluted the culture 1:100 in a petri dish. We added 2ml of LB medium to the dish. If the bacteria contain a plasmid, the appropriate antibiotic needs to be added to LB. The petri dish was then incubated statically in a growth chamber for 24 hours. After incubation, we performed imaging directly under the microscope. The petri dishes used were glass-bottom dishes from Biosharp (catalog number BS-20-GJM), allowing for direct microscopic imaging without the need for cover slips or slides. This setup allowed us to grow and image the biofilms *in situ*, providing a more accurate representation of their natural structure and composition. *Caulobacter crescentus* strain NA1000 was grown in peptone yeast extract (PYE) medium. And *Staphylococcus aureus* strain ATCC 25923 was grown in Mueller-Hinton Broth (MHB) medium. All bacterial strains were routinely grown at 37 °C and 220 rpm. To maintain plasmids, when necessary, media were supplemented with chloramphenicol (25 µg/ml) or kanamycin sulfate (50 µg/ml). For arabinose-induction system expression experiments, 0.002% or 0.00002% arabinose was supplemented in the medium.

Strains construction

The construction of recombinant plasmids was performed using the 2 x MultiF Seamless Assembly Mix (ABclonal, RK21020). For the detection of c-di-GMP levels using c-di-GMP sensor and the detection of persister, the *PdeI* gene, *pdeI*, along with its native promoter (250bp), was fused with either *gfp* or *bfp* and cloned into the pBAD backbone. The original promoter region of the pBAD vector was removed to avoid any potential interference. This construction allows the expression of the BFP, PdeI-BFP and PdeI(G412S)-BFP fusion proteins to be driven by *pdeI*'s native promoter, thus maintaining its physiological control mechanisms. And the BFP coding sequence was fused to the *pdeI* gene to create the PdeI-BFP fusion construct. Besides, for membrane localisation and localisation in biofilm community, the *pdeI-gfp* and *pdeI-bfp* with native promoter of *pdeI* were cloned into the pBAD backbone. For the control group, *bfp* and *gfp* was cloned into the pBAD backbone under the control of Arabinose-induction system. For HPLC LC-MS/MS analysis, *pdeI* and empty vector were cloned into the pBAD backbone, induced by arabinose. GFP and BFP were used in different experiments. GFP was used for imaging and time-lapse imaging to observe persister cell growth. BFP was used for cell sorting and detecting the proportion of persister cells. For the c-di-GMP sensor (Addgene: #182291), the plasmid origin was replaced with the p15A ori.

RiboD-PETRI

Cell preparation

E. coli MG1655 cells were cultured overnight and subsequently diluted at a ratio of 1:100 into fresh LB medium and grown statically for 24 h at 37 °C. For 3 h exponential period *E. coli* sample, *E. coli* MG1655 cells were grown overnight and then diluted 1:100 into fresh LB medium and grown for 3 h at 37 °C and 220 rpm. *Caulobacter crescentus* strain NA1000 cells were grown overnight and then diluted 1:100 into fresh Mueller-Hinton Broth (MHB) medium and grown for 9 h at 37 °C and 220 rpm. And *Staphylococcus aureus* strain ATCC 25923 cells were grown overnight and then diluted 1:100 into fresh peptone yeast extract (PYE) medium and grown for 3 h at 37 °C and 220 rpm. All the culture was vigorously shaken using a vortex, and the cells were then centrifuged at 5,000g for 2 minutes at 4 °C. The pellet was resuspended in 2 ml of ice-cold 4% formaldehyde (F8775, Millipore Sigma, diluted into PBS). These suspensions were rotated at 4 °C for 16 hours.

Cell permeabilization

1 ml of fixed cells were centrifuged at 5,000g for 5 minutes at 4 °C, then resuspended in 1 ml washing buffer (100 mM Tris-HCl pH 7.0, 0.02 U/μl SUPERase-In RNase Inhibitor, AM2696, Invitrogen). After another centrifugation at 5,000g for 5 minutes at 4 °C, the supernatant was removed. The pellet was then resuspended in 250 μl permeabilization buffer (0.04% Tween-20 in PBS-RI, PBS with 0.01 U/μl SUPERase-In RNase Inhibitor) and incubated on ice for 3 minutes. 1 ml cold PBS-RI was added, and the cells were centrifuged at 5,000g for 5 minutes at 4 °C. The pellet was resuspended in 250 μl Lysozyme Mix (250 μg/ml Lysozyme or 5 pg/ml Lysostaphin for *S. aureus* in TEL-RI buffer, comprising 100 mM Tris pH 8.0 (AM9856, Invitrogen), 50 mM EDTA (AM9261, Invitrogen), and 0.1 U/μl SUPERase In RNase Inhibitor). The samples were incubated at 37 °C and mixed gently every minute. Then 1 ml cold PBS-RI was added immediately, and cells were centrifuged at 5,000g for 5 min at 4 °C. The cells underwent another wash with 1 ml cold PBS-RI. Subsequently, cells were resuspended in 40 μl DNaseI-RI buffer (4.4 μl 10× reaction buffer, 0.2 μl SUPERase In RNase inhibitor, 35.4 μl H₂O), followed by addition of 4 μl DNaseI (AMPD1, Millipore Sigma). The samples were incubated for 30 min at room temperature and mixed gently every 5 minutes. 4 μl Stop Solution was added, and the samples were incubated for 10 minutes at 50 °C with gentle mixing every minute. Following centrifugation at 5,000g for 10 minutes at 4 °C, cells were washed twice with 0.5 ml cold PBS-RI. Finally, cells were resuspended in 200 μl cold PBS-RI, and their count and integrity were assessed using the ACEA NovoCyte flow cytometer with a 100× oil immersion lens.

Primer preparation

For round 1 RT, round 2 and round 3 ligation reactions, all primers design and preparation as previously described¹². All primers were purchased from Sangon Biotech (Table S1). For ligation primers preparation, mixtures were prepared as follows: 31.1 μl each R2 primer (100 μM), 28.5 μl SB83 (100 μM) and 21.4 μl H₂O, were splitted to 2.24 μl for one sample. Mixtures containing 63.2 μl each R3 primer (70 μM) and 58 μl SB8 (70 μM), were splitted to 3.49 μl for one sample. Before use, ligation primers were incubated as follows: 95 °C for 3 min, then decreasing the temperature to 20 °C at a ramp speed of -0.1 °C s⁻¹, 37 °C for 30 min. For blocking mix preparation, 50 μl primer SB84 (400 μM) and 80 μl primer SB81 (400 μM) were incubated as follows: 94 °C for 3 min, then decreasing the temperature to 25 °C at a ramp speed of -0.1 °C s⁻¹, 4 °C for keeping. Round 2 blocking primers were mixed as follows: 37.5 μl 400 μM SB84, 37.5 μl 400 μM SB85, 25 μl 10× T4 ligase buffer, 150 μl H₂O. Round 3 blocking primers were mixed as follows: 72 μl 400 μM SB81, 72 μl 400 μM SB82, 120 μl 10× T4 ligase buffer, 336 μl H₂O, 600 μl 0.5 M EDTA.

Round 1 RT reaction

About 3×10^7 cells were introduced into an RT reaction mix composed of 240 μL 5 $\times$ RT buffer, 24 μL dNTPs (N0447L, NEB), 12 μL SUPERase In RNase Inhibitor and 24 μL Maxima H Minus Reverse Transcriptase (EP0753, Thermo Fisher Scientific), 132 μL PEG8000 (50%). Nuclease-free water was added to achieve a total reaction volume of 960 μL , and the mixture was thoroughly mixed by vortexing. Subsequently, 8 μL of the reaction mixture was dispensed into each well of a 96-well plate, where 2 μL of each RT primer had been added previously. The sealed 96-well plate was inverted repeatedly for thorough mixing, followed by a brief spin. The plate was then incubated as follows: 50 °C for 10 min, 8 °C for 12 s, 15 °C for 45 s, 20 °C for 45 s, 30 °C for 30 s, 42 °C for 6 min, 50 °C for 16 min, and finally held at 4 °C. After the RT process, all 96 reactions were pooled into one tube. 75 μL of 0.5% Tween-20 was added, and the reactions were incubated on ice for 3 minutes. Cells were centrifuged at 7,000g for 10 minutes at 4 °C and then resuspended in 0.4 ml PBS-RI. Thirty-two microliters of 0.5% Tween-20 was added, and the cells underwent centrifugation at 7,000g for 10 minutes at 4 °C.

Round 2 ligation reaction

Cells were resuspended in 500 μL 1 $\times$ T4 ligase buffer, followed by the addition of 107.5 μL PEG8000, 37.5 μL 10 $\times$ T4 ligase buffer, 16.7 μL SUPERase In RNase Inhibitor, 5.6 μL BSA, and 27.9 μL T4 ligase (M0202L, NEB). The reaction solution was thoroughly mixed by vortexing. Subsequently, 5.76 μL of the reaction mixture was dispensed into each well of a 96-well plate, where 2.24 μL of each round 2 ligation primer had been added previously. The sealed 96-well plate was inverted repeatedly for thorough mixing and then subjected to a short spin. The plate was incubated at 37 °C for 45 minutes. Following this, 2 μL of round 2 blocking mix was added to each well and incubated at 37 °C for an additional 45 minutes. All 96 reactions were pooled into one tube after incubation.

Round 3 ligation reaction

A mixture comprising 89 μL H₂O, 26 μL PEG8000, 46 μL 10 $\times$ T4 ligase buffer, and 12.65 μL T4 ligase was prepared and thoroughly mixed by vortexing. Subsequently, 8.51 μL of the reaction mixture was dispensed into each well of a 96-well plate, where 3.49 μL of each round 3 ligation primer had been added previously. The sealed 96-well plate was inverted repeatedly for thorough mixing and then subjected to a brief spin. The plate was incubated at 37 °C for 45 minutes. Following this, 10 μL of round 3 blocking mix was added to each well and incubated at 37 °C for an additional 45 minutes. All 96 reactions were combined into one tube after incubation.

Cells lysis

42 μL of 0.5% Tween-20 was added, and cells were centrifuged at 7,000g for 10 minutes at 4 °C. The cells underwent two washes using 200 μL TEL-RI containing 0.01% Tween-20, each time centrifuged at 7,000g for 10 minutes at 4 °C. Subsequently, cells were resuspended in 30 μL TEL-RI buffer. Cell counting and integrity checks were performed using the ACEA NovoCyte flow cytometer with a 100 $\times$ oil immersion lens. A moderate amount of cells was then added to the lysis buffer (50 mM Tris pH 8.0, 25 mM EDTA, 200 mM NaCl, 0.5% Triton X-100), and 5 μL of proteinase K (AM2548, Invitrogen) was introduced. Samples were incubated at 55 °C for 60 minutes and gently mixed every minute.

Library construction

To facilitate template switching, lysates were purified with VAHTS DNA Clean Beads (N411, Vazyme) at a ratio of 2.0 $\times$, and cDNA was eluted in 12 μL of water. The purified cDNA was then combined with 4 μL of 5 $\times$ RT buffer, 1 μL of dNTPs (N0447L, NEB), 0.5 μL of SUPERase In RNase Inhibitor, 0.5 μL of Maxima H Minus Reverse Transcriptase, and 2 μL of the TSO²⁹ primer (100

mM, **Table S1** [↗](#)). This reaction solution underwent incubation as follows: 25 °C for 30 min, 42 °C for 90 min, 85 °C for 5 min, and then held at 4 °C. Subsequently, 1 µl of RNaseH was added, and the reaction solution was incubated at 37 °C for 30 min. The cDNA was purified once again with VAHTS DNA Clean Beads at a ratio of 2.0× and eluted in 13 µl of H₂O. The integrity of the cDNA was assessed using primers TSO-2 and R1 or R2 or R3 by qPCR (**Table S1** [↗](#)).

Ribosomal RNA derived cDNA depletion (RiboD)

We developed a set of cDNA probe primers to selectively deplete r-cDNA (**Table S1** [↗](#)). These probe primers possess the ability to specifically hybridize with r-cDNA and also hybridize with a biotin-labeled universal primer. In the reaction, 5 µl of r-cDNA probe primers (10 µM), 2.5 µl of 10× hybridization buffer (Tris-HCl pH 8.0 100 mM, NaCl 500 mM, EDTA pH 8.0 10 mM), and 5 µl of biotin primer (10 µM) were added to 12.5 µl of purified cDNA. The reaction solution underwent incubation as follows: 95 °C for 2 min, followed by a temperature decrease to 20 °C at a ramp speed of −0.1 °C s^{−1}, and then held at 37 °C for 30 min. Subsequently, 20 µl of Streptavidin magnetic beads (BEAVER, 22307) was washed twice using 1 ml of 1× B&W buffer (Tris-HCl pH 7.5 10 mM, EDTA 1 mM, NaCl 1M, Tween-20 0.05%) and resuspended in 25 µl of 2× B&W buffer. Twenty-five microliters of washed Streptavidin magnetic beads were added to 25 µl of annealed cDNA. The reaction solution was incubated at room temperature for 30 min with gentle mixing per minute. Following this, the reaction solution tube was placed into a magnetic stand to collect the supernatant. The cDNA depleted of r-cDNA was purified using VAHTS DNA Clean Beads at a ratio of 2.0× and eluted in 12.5 µl of H₂O. The depletion of r-cDNA could be repeated using the above protocol, and ultimately, the cDNA was eluted in 20 µl of H₂O. We designed separate probe sets for *E. coli*, *C. crescentus*, and *S. aureus*. Each set was specifically constructed to be reverse complementary to the r-cDNA sequences of its respective bacterial species. This species-specific approach ensures high efficiency and specificity in rRNA depletion for each organism.

Library amplification and sequencing

To the 20 µl cDNA solution, the following components were added: 2.4 µl R3 primer (10 mM, **Table S1** [↗](#)), 2.4 µl TSO-2 primer (10 mM, **Table S1** [↗](#)), 40 µl 2× KAPA HIFI mix (KAPA, 2602), 1.6 µl Sybr green (25×), 0.8 µl MgCl₂ (0.1 M), and 12.8 µl H₂O. This PCR reaction solution was placed in a thermocycler and incubated with the following parameters: 98 °C for 45 s, followed by cycling of 98 °C for 15 s, 60 °C for 30 s, and 72 °C for 60 s. Cycling continued on a qPCR machine until the reaction approached saturation. PCR products were then purified using VAHTS DNA Clean Beads at a ratio of 0.9× and eluted in 25 µl of H₂O. Finally, the purified PCR products underwent end repair and adaptor ligation using the VAHTS Universal DNA Library Prep Kit for Illumina V3 (Vazyme, ND607).

Bulk RNA-seq library construction

Total RNA of the samples was extracted utilizing the Bacteria RNA Extraction Kit (R403-01, Vazyme). Subsequently, the RNA underwent mRNA enrichment (N407, Vazyme), fragmentation, cDNA synthesis, and library preparation using the VAHTSTM Total RNA-seq (H/M/R) Library Prep Kit for Illumina® (NR603, Vazyme).

Bioinformatics analysis Methods

Single-Cell Analysis

The sequencing data underwent processing into matrices using scripts and a pipeline as previously described¹² [↗](#) in Python 2.7.15, with some modifications (the detailed original code and all the data were deposited in the GEO repository). After the count tables were made, subsequent analysis of single-cell data was conducted using Seurat³⁰ [↗](#) package (version 4.3.0; <http://satijalab.org/seurat/> [↗](#)) in R (<https://www.r-project.org/> [↗](#)). Since there were two replicates of static *E. coli*

biofilm, these two datasets were merged into one SeuratObject and batch effects were removed. However, the samples for exponential period *E. coli*, *S. aureus* and *C. crescentus* only had one sample, so they did not need this process. At the beginning of doing the scRNA-seq analysis, we screened the data of all samples. For preprocessing of static *E. coli* biofilm data, cells were filtered with UMI per cell more than 100 and less than 2000 for replicate 1 and replicate 2 to obtain 1621 and 3999 cells, respectively. For data of exponential period *E. coli*, the data was screened for cells with UMIs greater than 200 and less than 5000 to obtain 1464 cells. The screening criteria of *S. aureus* were cells with UMIs greater than 15 and less than 1000 and genes greater than 30 ($1000 > \text{UMIs} > 15$, $\text{gene counts} > 30$). The screening criteria of *C. crescentus* were cells with UMIs greater than 200 and less than 5000 and gene counts greater than 30 ($5000 > \text{UMIs} > 200$, $\text{gene counts} > 30$). After screening, all the data were normalized using a scale factor of 10000 through a global-scaling normalization method called “LogNormalize”. Highly variable features were then identified, returning 500 features per dataset. Then we combine the data of the two replicates of static *E. coli* biofilm into a single SeuratObject by FindIntegrationAnchors and IntegrateData functions. Then all the data underwent scaling using the ScaleData function, followed by dimension reduction through principal component analysis. To avoid subtle batch effects influencing downstream analyses, we removed batch effects using RunHarmony³¹ for the two replicates of static *E. coli* biofilm data. Then a graph-based clustering approach was employed in all data to identify clusters of gene expression programs using the Louvain algorithm (Seurat 4.3.0). The dims we choosed were 6. And the resolution was 0.3 for *C. crescentus* and *S. aureus* or 0.4 for *E. coli* data. Marker genes for each cluster were computed using the Wilcoxon Rank-sum test. Specifically, marker genes for each cluster were initially obtained using the FindMarkers function of Seurat. Then we performed pathway enrichment analysis of marker gene by clusterProfiler function³² within R. For transcriptome-wide gene coverage across the cell population, we counted the number of genes expressing at least one UMI. Then we calculated the percentage of these genes out of all the genes in each bacterium.

Comparison of scRNA-seq with Bulk RNA-Seq

The bulk RNA-seq clean data reads were mapped to the *E. coli* MG1655 k12 genome (EnsemblBacteria Taxonomy ID: 511145) using the BWA aligner software (v0.7.17-r1188, <https://github.com/lh3/bwa.git>). Sam files were converted to bam files using samtools (v1.9). The mapping results were counted by featureCounts (v2.0.1, <https://github.com/topics/featurecounts>) to generate expression results. Single-cell and bulk transcriptomes of *E. coli* were compared by computing the Pearson correlation of $\log_2$ reads per gene of bulk RNA-seq and $\log_2$ UMI per gene of scRNA-seq.

Sequencing saturation of the libraries

To assess sequencing saturation, we generated five subsamples from the single-cell sequencing data, representing 20%, 40%, 60%, 80%, and 100% of the total data. Each subsample was analyzed independently following the previously described single-cell sequencing data analysis process. We then created gene expression matrices for each subsample based on the analysis results. The number of UMIs or genes was counted for each cell detected in each subsample. Next, we sorted the cells in descending order based on their UMI counts or gene counts and selected different numbers of cells from this sorted list, starting from those with the highest UMI or gene counts. For each selection of cells, we calculated the median number of UMIs or genes. Finally, we created a line graph representing the median UMI counts and genes using GraphPad Prism 9 software, allowing us to visualize the sequencing saturation across the different subsamples. In addition, we used the saturation calculation method of 10× Genomics to further detect the saturation of the data. The formula for calculating this metric is as follows: Sequencing Saturation = $1 - (\text{n_deduped_reads} / \text{n_reads})$. Given the differences between RiboD-PETRI and 10× Genomics datasets, we have adapted the calculation as follows: n_deduped_reads: The number of UMIs as a measure of unique reads. n_reads: The total number of confidently mapped reads.

Multiplet frequency determination

Determination of the multiplet frequency was essential in assessing the efficiency of single-cell capture in RiboD-PETRI. This frequency is defined as the probability that a non-empty barcode corresponds to more than one cell. To calculate it, we used a Poisson distribution-based approach involving several key steps. Initially, we calculated the proportion of barcodes corresponding to zero cells using the formula $P(0) = \frac{\lambda^0}{0!}e^{-\lambda}$. Then, we determined the proportion for one cell,

$P(1) = \frac{\lambda^1}{1!}e^{-\lambda}$, and derived the proportions for more than zero cells $P(\geq 1) = 1 - P(0)$ and more than one cell $P(\geq 2) = 1 - P(1) - P(0)$. These values allow for the calculation of the multiplet frequency as $\frac{P(\geq 2)}{P(\geq 1)}$. The parameter λ plays a vital role in this model, representing the ratio of the number of cells to the total number of possible barcode combinations.

Flow cytometry sorting of bacteria and analysis

All samples were measured using a Beckman CytoFLEX SRT flow cytometer with a 70 μ m nozzle, using normal saline as sheath fluid. During the 24-hour static biofilm growth phase, strains labeled with BFP, PdeI-BFP, PdeI(G412S)-BFP or c-di-GMP sensors were washed and resuspended in sterile PBS. Microorganisms were identified based on forward scatter (FSC) and side scatter (SSC) parameters. Cells were sorted into distinct groups according to their fluorescence intensity, with V450 used for BFP, FITC for mVenusNB, and ECD for mScarlet-I. The resulting data were subsequently analyzed using FlowJo V10 software (Tree Star, Inc.).

Antibiotic killing and persister counting assay

Cells sorted by flow cytometry were resuspended in fresh LB broth supplemented with 150 μ g/ml ampicillin. The suspension was then incubated at 37 °C for 3 hours with continuous shaking at 220 rpm. To determine the initial cell count, an aliquot of the cell suspension was taken before the ampicillin challenge, serially diluted, and plated on LB agar plates for colony-forming unit (CFU) enumeration. These plates were incubated overnight at 37 °C. Following the ampicillin challenge, cells were harvested by centrifugation, washed once with sterile PBS to remove residual antibiotic, and resuspended in fresh PBS. This suspension was then serially diluted and plated on LB agar plates for post-challenge CFU counts. The persister ratio was calculated as the number of CFUs after ampicillin challenge divided by the number of CFUs before challenge. All experiments were performed in triplicate, with the results presented as means $\pm$ standard deviations of three independent biological replicates.

Microscopy

Bright-field and fluorescence imaging

Inverted microscopes, specifically the Nikon Eclipse Ti2 and Leica Stellaris 5 WLL, were employed for imaging, utilizing different lasers for illumination: a 405 nm laser for BFP and a 488 nm laser for GFP. Fluorescence emission signals were captured using an sCMOS camera (pco.edge 4.2 bi). Dedicated filter sets corresponding to the spectral characteristics of each fluorophore were utilized. Image analysis was performed with ImageJ software (Fiji). For the analysis of the c-di-GMP sensor, the ratio of mVenusNB to mScarlet-I (R) displayed a negative correlation with c-di-GMP concentration. Consequently, the value of R^{-1} demonstrated a positive correlation with c-di-GMP concentration.

Time-lapse imaging

To investigate the processes of antibiotic killing and bacterial resuscitation, cells labeled with PdeI-GFP during the 24-hour static growth phase were collected and washed twice with PBS. These cells were then imaged on a gel pad composed of 3% low melting temperature agarose in PBS, which was prepared as a gel island in the center of the FCS3 chamber. The cells were observed under either bright-field or epifluorescence illumination. Following the imaging, the gel pad was surrounded by LB broth containing 150 pg/ml ampicillin, and the cells were incubated for 6 hours at 35 °C. Fresh LB broth was subsequently flushed in, and the growth medium was refreshed every 3 hours, ensuring sufficient recovery time for the cells.

Determination of c-di-GMP concentration by HPLC-MS/MS

The determination of c-di-GMP concentration by HPLC-MS/MS involved a series of steps. Initially, MG1655 Aara Δ BAD::*pdel* and MG1655 Δ ara pBAD::empty-vector strains were grown to mid-exponential growth phase, followed by induction with 0.002% arabinose. After a 2-hour incubation period, cells were harvested and washed with PBS. The washed cells were then rapidly frozen using liquid nitrogen.

Simultaneously, another portion of washed cells was stained with SYTOTM24 and quantified using flow cytometry. The determination of c-di-GMP concentration was conducted by Wuhan Lixinheng Technology Co. Ltd. through high-pressure liquid chromatography-tandem mass spectrometry (HPLC-MS/MS). In the experiment, first, for cell samples, addition of 500 μ L of extract solvent (precooled at -20 °C, acetonitrile-methanol-water, 2:2:1), the samples were vortexed for 30 s, homogenized at 38 Hz for 4 min, and sonicated for 5 min in ice-water bath. The homogenate and sonicate circle were repeated for 3 times, followed by incubation at -20 °C for one hour and centrifugation at 12000 rpm and 4 °C for 15 min. An 80 μ L aliquot of the clear supernatant was transferred to an auto-sampler vial for LC-MS/MS analysis. The UHPLC separation was carried out using an Waters ACQUITY H-class plus UPLC System, equipped with Agilent ZORBAX Eclipse Plus C18 column (2.1 mm $\times$ 150 mm, 1.8 μ m). An Waters Xevo TQ-XS triple quadrupole mass spectrometer, equipped with an electrospray ionization (ESI) interface, was applied for assay development. All strains were assayed in biological triplicates, and the measured values were converted into intracellular c-di-GMP concentrations (pg) per cell.

Quantification and statistical analysis

Statistical analysis was conducted using GraphPad Prism 9 software for Windows. The significance of results was determined using a two-tailed Student's *t*-test. Error bars in the data representation indicate the standard deviations of the mean from a minimum of three independent experiments. A significance threshold of $P < 0.05$ was applied throughout the analysis. To denote significant differences in the results, a system of asterisks was used: * for $P < 0.05$, ** for $P < 0.01$, *** for $P < 0.001$, and **** for $P < 0.0001$. This comprehensive approach ensured a thorough and statistically sound analysis of the c-di-GMP concentration in the studied bacterial strains.

Additional Declarations

The authors declare no competing interests.

Author Contributions

Conceptualization: Y.Z., H.L., Y.P.; Methodology: H.L., X.Y., C.W., C.H.; Investigation: H.L., X.Y., C.W., C.H.; Visualization: H.L., C.W.; Bioinformatics analysis: X.Y.; Clinical stain isolation: C.G.; Supervision: Y.P.; Writing - original draft: H.L., Y.P.; Writing - review & editing: H.L., Y.P.

Competing Interest Statement

Authors declare that they have no competing interests.

Classification

Biological Sciences

Acknowledgements

We thank Prof. Fan Bai (Peking University) for valuable discussions. We thank Drs. Jidong Xing and Ziyang Liu for support in bioinformatics. We also thank the members of our laboratory for helpful discussion. This work is supported by the grants to Y.P. from the National Key R&D Program of China (2021YFC2701602), the National Natural Science Foundation of China (31970089), Science Fund for Distinguished Young Scholars of Hubei Province (2022CFA077), Major Project of Guangzhou National Laboratory (GZNL2024A01023), and the Fundamental Research Funds for the Central Universities (2042022dx0003). This work is also supported by the grants to Y.Z. for the National Science Fund for Distinguished Young Scholars (82025011) and C.G. from the Natural Science Foundation of Yunnan Province of China (202001BB050005). We also thank all the staff in the Core Facilities of Medical Research Institute at Wuhan University and the Core Facilities at School of Life Sciences at Peking University for their technical support.

Supplementary Figures

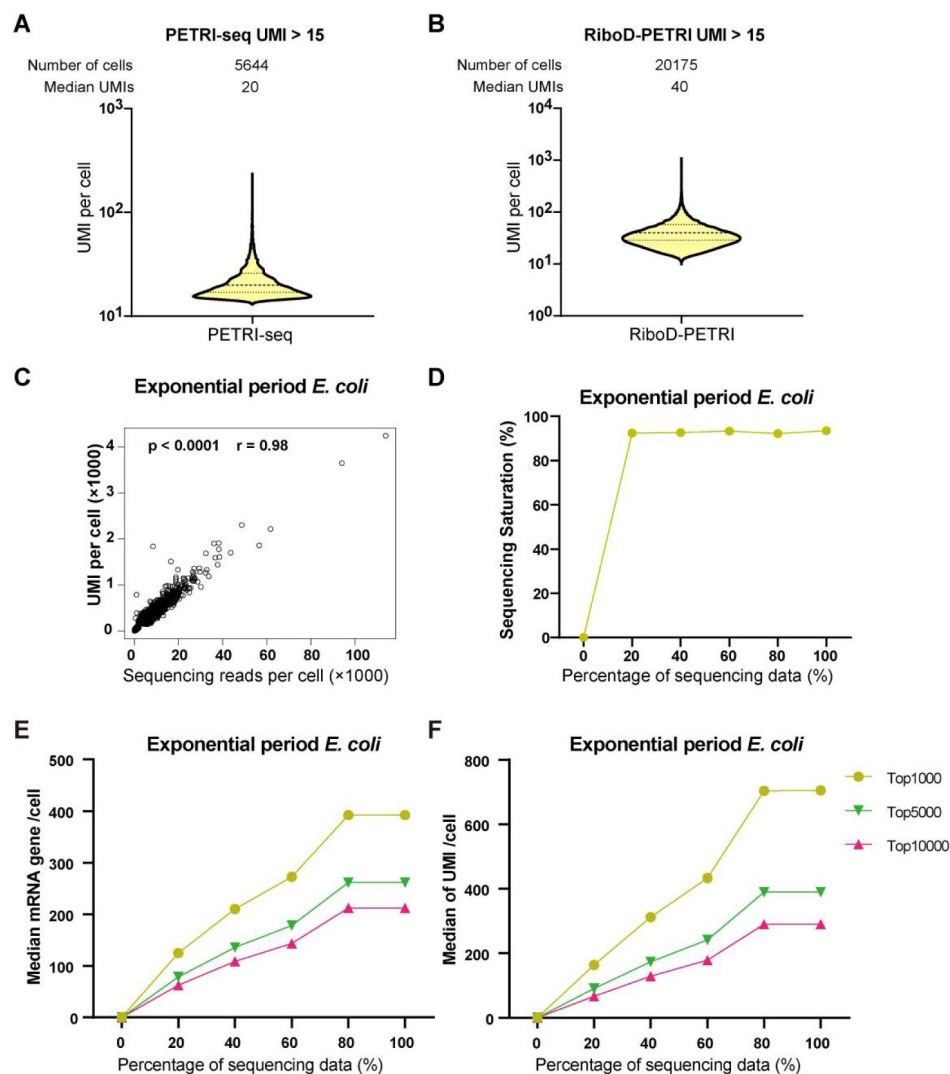


Figure S1.

(A, B) The number of UMIs detected per cell in recovered cells in different samples (≥ 15 UMIs/cell): (A) PETRI, (B) RiboD-PETRI at the same unsaturated sequencing depth. The cells are ranked from highest to lowest based on the number of detected UMIs, and cells with ≥ 15 UMIs are selected for plotting. The median number of UMIs is calculated for these selected cells. (C) Scatterplot illustrating the relationship between reads per cell and counts of UMIs per cell detected from exponential phase *E. coli* data. Each dot represents a cell. (D) Sequencing saturation of data of exponential period *E. coli* (3h). We extracted 20%, 40%, 60%, 80% and 100% of the data and further tested their saturation using the saturation calculation method of 10 $\times$ Genomics. (E & F) Sequencing saturation analysis. We took 20%, 40%, 60%, 80% and 100% of the sequencing data for single-cell analysis and counted the number of genes and UMIs for each cell in these data. The cells were then sorted from largest to smallest values, and cells were taken to count the median number of genes (E) and UMIs (F).

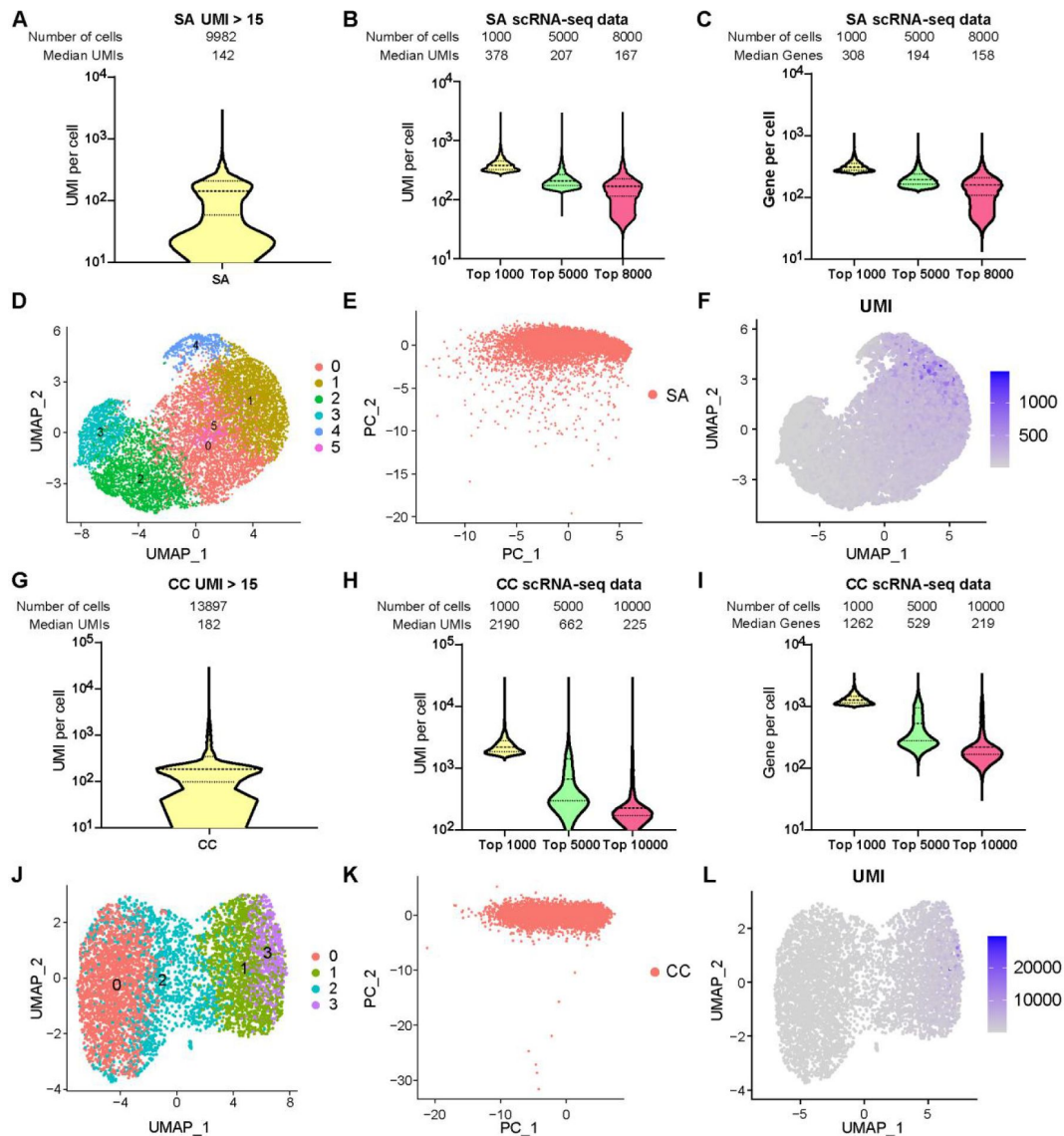


Figure S2.

Comprehensive Single-Cell Transcriptomic Analysis of *S. aureus* and *C. crescentus* using RiboD-PETRI

Technical Application of RiboD-PETRI in *S. aureus* (SA) (A-F), cultured for 9 hours in MHB medium at 37 °C (Table S14) and *C. crescentus* (CC) (G-L), incubated at 37 °C for 3 hours (Table S15). (A, G) The number of UMIs detected per cell in different samples (≥ 15 UMIs/cell): (A) *S. aureus* (SA) and (G) *C. crescentus* (CC). (B, H) Distribution of mRNA UMIs captured per cell in RiboD-PETRI data of (B) *S. aureus* (SA) and (H) *C. crescentus* (CC), presented as violin plots showing the upper quartile, median, and lower quartile lines. The cells are ranked from highest to lowest based on the number of UMIs detected. Then, specific numbers of cells (indicated above the panel) are selected for plotting. The median number of UMIs is calculated for these selected cells. (C, I) The number of genes detected per cell in different samples (C) *S. aureus* and (I) *C. crescentus*. The cells are ranked from highest to lowest based on the number of genes detected. Then, specific numbers of cells (indicated above the panel) are selected for plotting. The median number of genes is calculated for these selected cells. "SA" denotes *S. aureus*, and "CC" denotes *C. crescentus*. (D, J) UMAP visualization of (D) *S. aureus* and (J) *C. crescentus*, demonstrating the ability of RiboD-PETRI to distinguish population heterogeneity. (E, K) Normalized and Principal Component Analysis (PCA) performed on screened data of (E) *S. aureus* and (K) *C. crescentus*. The resulting scatterplots show heterogeneity among the populations, with each point representing a cell. (F, L) Distribution of UMIs on the UMAP results for (F) *S. aureus* and (L) *C. crescentus*. UMAP results reveal heterogeneity among populations, with each point representing a cell and color shading indicating UMI counts.

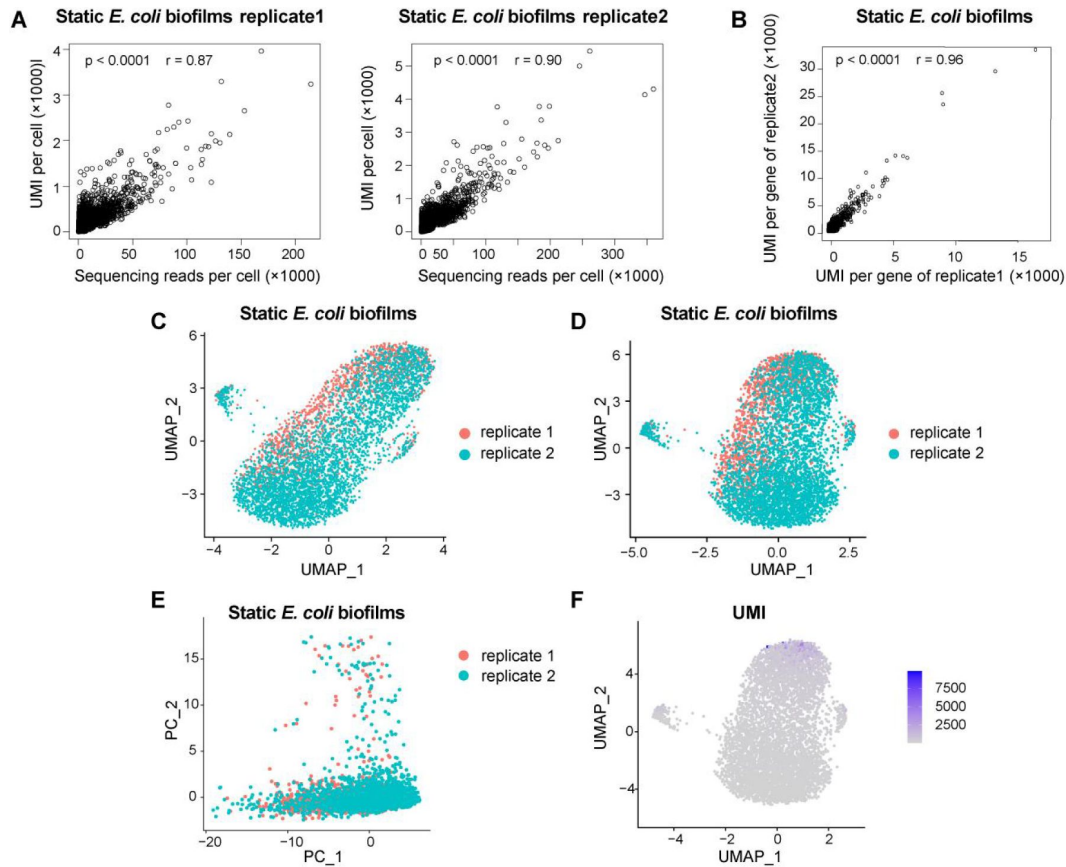


Figure S3.

Evaluation of Transcriptomic Consistency and Batch Effect Analysis in Static Biofilm *E. coli* Samples

(A) Scatterplot demonstrating the relationship between reads per cell and counts of UMIs per cell detected from static biofilm *E. coli* data. Two replicates of the sample are included. (B) Calculation of the Pearson correlation coefficient (r) of UMI counts per gene between replicate 1 and replicate 2 of static biofilm *E. coli*. The analysis involved 4,062 out of 4,141 total genes, with a significant correlation (p -value < 0.0001 , $r = 0.96$), indicating good replication between samples. Each dot represents a gene. (C) Before batch effects were removed, UMAP plot based on the original identity of static biofilm *E. coli* samples (replicate 1 and replicate 2). Each dot represents a cell, with red indicating replicate 1 and green indicating replicate 2. (D) After batch effects were removed using Harmony, UMAP plot based on the original identity of static biofilm *E. coli* samples (replicate 1 and replicate 2). (E) Principal Component Analysis (PCA) performed on screened data of two replicates of static biofilm *E. coli*. The resulting scatterplots show heterogeneity among the populations, with each point representing a cell. (F) Distribution of UMIs on the UMAP results for two replicates of static biofilm *E. coli*. UMAP results reveal heterogeneity among populations, with each point representing a cell and color shading indicating UMI counts.

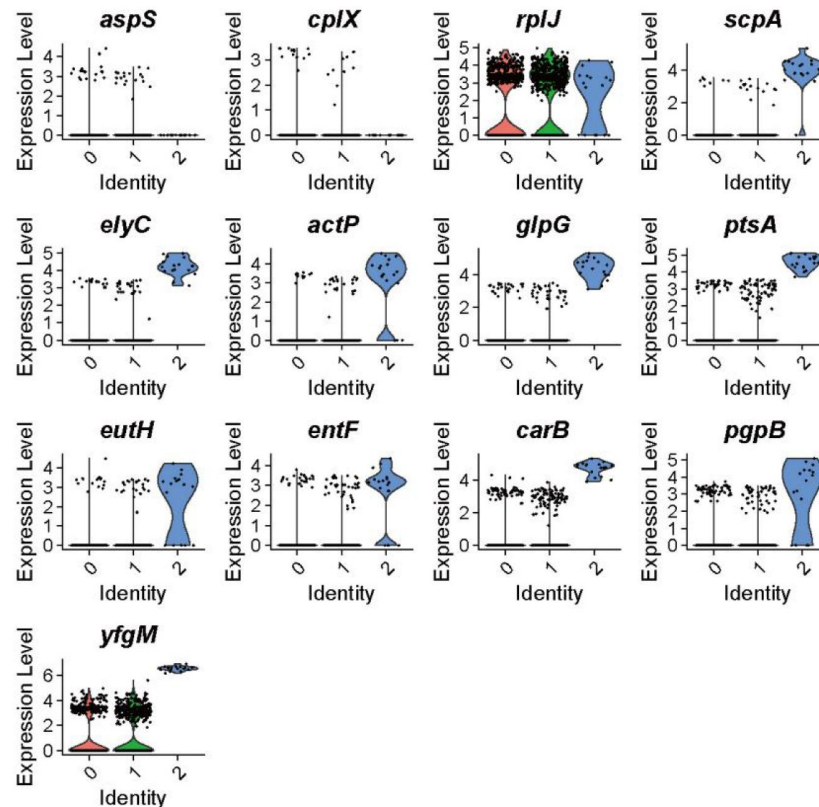


Figure S4.

Profiling of Marker Genes in exponential phase *E. coli* culture by RiboD-PETRI

Expression levels of diverse marker genes across distinct clusters in exponential phase *E. coli* culture, visualized through violin plots. Each individual dot represents a single cell, demonstrating the high-resolution, single-cell nature of the RiboD-PETRI analysis.

Figure S5.

Marker Genes Identified in stationary phase *S. aureus* culture by RiboD-PETRI

Expression levels of different marker genes across different clusters in stationary phase *S. aureus* culture overlaid on the UMAP plot. Marker genes were selected based on a p-value greater than 0.001 and a $\log_2$ FC greater than 0.2. Each dot represents a cell and color shading indicating UMI counts.

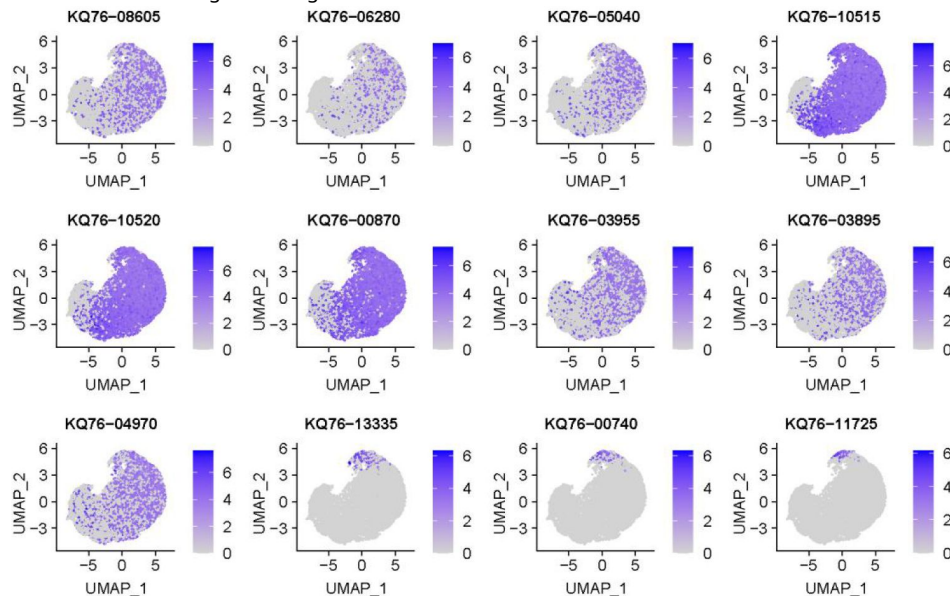


Figure S6.

Marker Genes Identified in exponential phase *C. crescentus* culture by RiboD-PETRI

Expression levels of different marker genes across different clusters in exponential phase *C. crescentus* culture overlaid on the UMAP plot. Marker genes were selected based on a p-value greater than 0.001 and a $\log_2$ FC greater than 0.2. Each dot represents a cell and color shading indicating UMI counts.

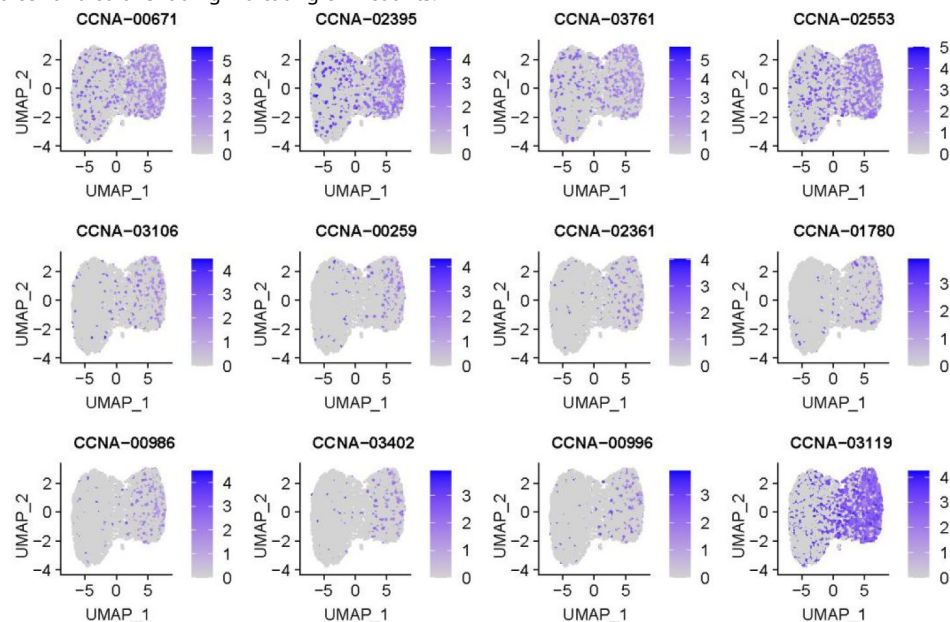


Figure S7.

Marker Genes Identified in static *E. coli* biofilms by RiboD-PETRI

Expression levels of different marker genes across different clusters in static *E. coli* biofilms overlaid on the UMAP plot. Marker genes were selected based on a p-value greater than 0.001 and a $\log_2$ FC greater than 3. Each dot represents a cell and color shading indicating UMI counts.

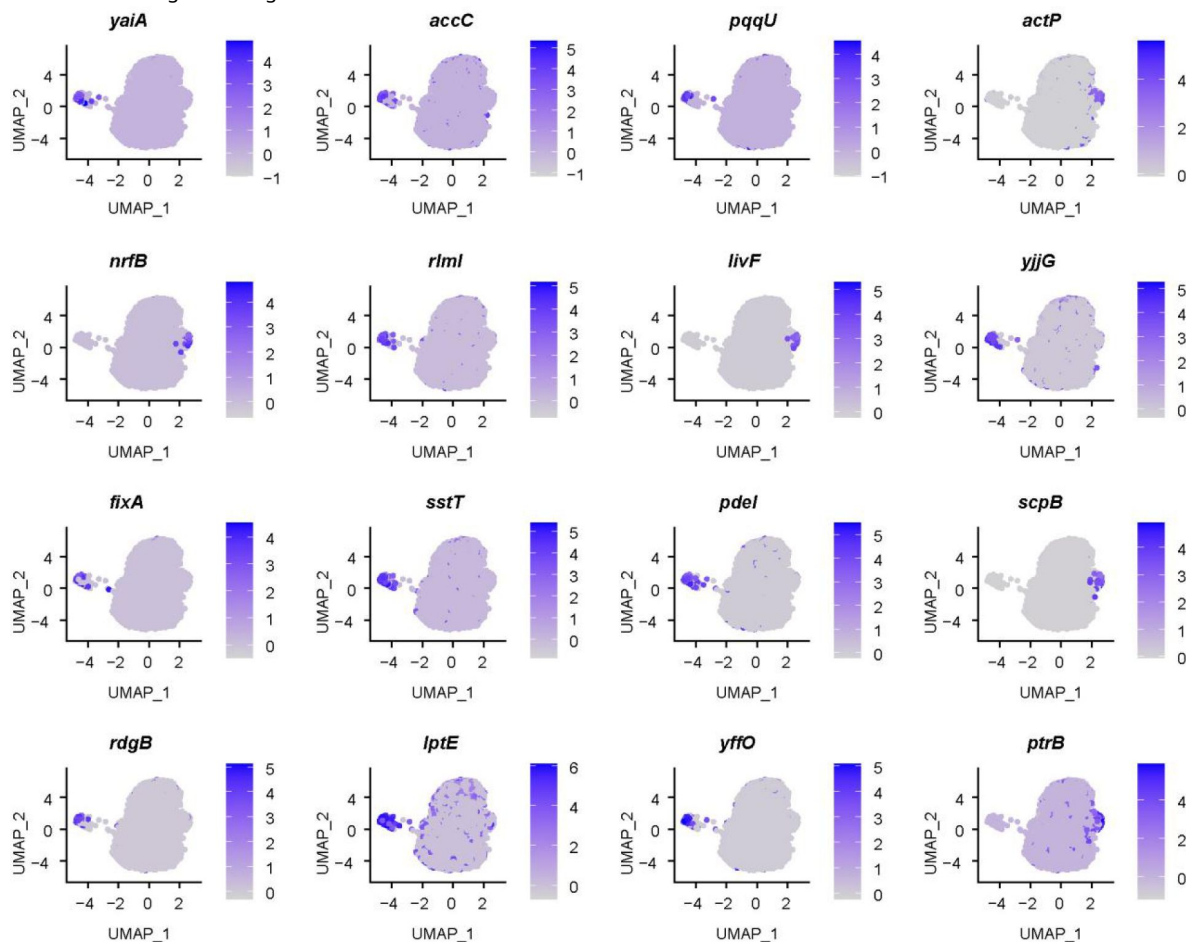
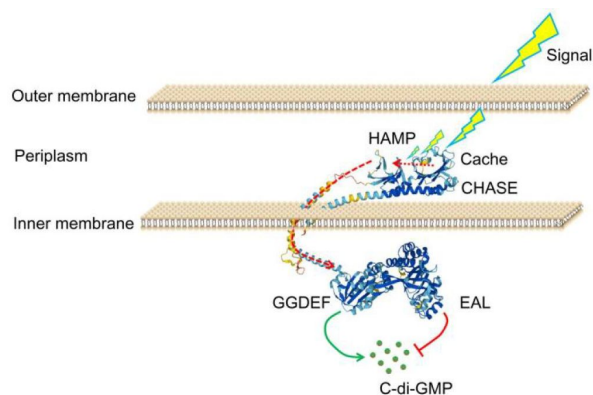


Figure S8.

Schematic chart for the structure of *E. coli* PdeI



References

- 1 Costerton J. W., Stewart P. S., Greenberg E. P. (1999) **Bacterial biofilms: a common cause of persistent infections** *Science* **284**:1318–1322
- 2 Evans C. R., et al. (2023) **Spatial heterogeneity in biofilm metabolism elicited by local control of phenazine methylation** *Proceedings of the National Academy of Sciences of the United States of America* **120**
- 3 Stewart P. S., Franklin M. J. (2008) **Physiological heterogeneity in biofilms** *Nature reviews. Microbiology* **6**:199–210
- 4 Flemming H. C., et al. (2023) **The biofilm matrix: multitasking in a shared space** *Nature reviews. Microbiology* **21**:70–86
- 5 Shokeen B., Dinis M. D. B., Haghighi F., Tran N. C., Lux R. (2021) **Omics and interspecies interaction** *Periodontology 2000* **85**:101–111
- 6 Qian P. Y., Cheng A., Wang R., Zhang R. (2022) **Marine biofilms: diversity, interactions and biofouling** *Nature reviews. Microbiology* **20**:671–684
- 7 Momeni B. (2018) **Division of Labor: How Microbes Split Their Responsibility** *Current biology : CB* **28**:R697–R699
- 8 Cheng J., et al. (2023) **Biofilm heterogeneity-adaptive photoredox catalysis enables red light-triggered nitric oxide release for combating drugresistant infections** *Nature communications* **14**
- 9 Spormann A. M. (2008) **Physiology of microbes in biofilms** *Current topics in microbiology and immunology* **322**:17–36
- 10 Shapiro E., Biezuner T., Linnarsson S. (2013) **Single-cell sequencing-based technologies will revolutionize whole-organism science** *Nature reviews. Genetics* **14**:618–630
- 11 Tang F., et al. (2009) **mRNA-Seq whole-transcriptome analysis of a single cell** *Nature methods* **6**:377–382
- 12 Blattman S. B., Jiang W., Oikonomou P., Tavazoie S. (2020) **Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing** *Nature microbiology* **5**:1192–1201
- 13 Imdahl F., Vafadarnejad E., Homberger C., Saliba A. E., Vogel J. (2020) **Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria** *Nature microbiology* **5**:1202–1206
- 14 Kuchina A., et al. (2021) **Microbial single-cell RNA sequencing by split-pool barcoding** *Science*
- 15 Ma P., et al. (2023) **Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states** *Cell* **186**:877–891

- 16 Wang B., et al. (2023) **Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection** *Nature microbiology* **8**:1846–1862
- 17 McNulty R., et al. (2023) **Probe-based bacterial single-cell RNA sequencing predicts toxin regulation** *Nature microbiology* **8**:934–945
- 18 Lu J., et al. (2023) **scRNA-seq data analysis method to improve analysis performance** *IET nanobiotechnology* **17**:246–256
- 19 Maynard A., et al. (2020) **Therapy-Induced Evolution of Human Lung Cancer Revealed by Single-Cell RNA Sequencing** *Cell* **182**:1232–1251
- 20 Homberger C., Hayward R. J., Barquist L., Vogel J. (2023) **Improved Bacterial Single-Cell RNA-Seq through Automated MATQ-Seq and Cas9-Based Removal of rRNA Reads** *mBio* **14**
- 21 Xu Z., et al. (2023) **Droplet-based high-throughput single microbe RNA sequencing by smRandom-seq** *Nature communications* **14** <https://doi.org/10.1038/s41467-023-40137-9>
- 22 Korshoj L. E., Kielian T. (2024) **Bacterial single-cell RNA sequencing captures biofilm transcriptional heterogeneity and differential responses to immune pressure** *bioRxiv : the preprint server for biology* <https://doi.org/10.1101/2024.06.28.601229>
- 23 Merritt J. H., Kadouri D. E., O'Toole G. A. (2011) **Growing and Analyzing Static Biofilms** *Current protocols in microbiology*
- 24 Yu Z., et al. (2023) **Gas and light: triggers of c-di-GMP-mediated regulation** *FEMS microbiology reviews* **47**
- 25 Li S., et al. (2023) **c-di-GMP inhibits the DNA binding activity of H-NS in Salmonella** *Nature communications* **14**
- 26 Vrabioiu A. M., Berg H. C. (2022) **Signaling events that occur when cells of Escherichia coli encounter a glass surface** *Proceedings of the National Academy of Sciences of the United States of America* **119** <https://doi.org/10.1073/pnas.2116830119>
- 27 Reinders A., et al. (2016) **Expression and Genetic Activation of Cyclic Di-GMP-Specific Phosphodiesterases in Escherichia coli** *J Bacteriol* **198**:448–462
- 28 Lacanna E., Bigosch C., Kaefer V., Boehm A., Becker A. (2016) **Evidence for Escherichia coli Diguanylate Cyclase DgcZ Interlinking Surface Sensing and Adhesion via Multiple Regulatory Routes** *J Bacteriol* **198**:2524–2535
- 29 Picelli S., et al. (2013) **Smart-seq2 for sensitive full-length transcriptome profiling in single cells** *Nature methods* **10**:1096–1098
- 30 Hao Y., et al. (2021) **Integrated analysis of multimodal single-cell data** *Cell* **184**:3573–3587
- 31 Korsunsky I., et al. (2019) **Fast, sensitive and accurate integration of single-cell data with Harmony** *Nature methods* **16**:1289–1296

- 32 Yu G., Wang L. G., Han Y., He Q. Y. (2012) **clusterProfiler: an R package for comparing biological themes among gene clusters** *Omics : a journal of integrative biology* **16**:284–287

Author information

Xiaodan Yan*

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Medical Research Institute, Wuhan University, Wuhan, China, Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China

*These authors contributed equally

Hebin Liao*

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Medical Research Institute, Wuhan University, Wuhan, China, Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China, Translational Medicine Research Center, North Sichuan Medical College, Nanchong, China

*These authors contributed equally

Chenyi Wang

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Medical Research Institute, Wuhan University, Wuhan, China, Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China

Chun Huang

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Medical Research Institute, Wuhan University, Wuhan, China, Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China

Wei Zhang

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Medical Research Institute, Wuhan University, Wuhan, China, Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China

Chunming Guo

Center for Life Sciences, School of Life Sciences, Yunnan University, Kunming, China

Yingying Pu[†]

The State Key Laboratory Breeding Base of Basic Science of Stomatology & Key Laboratory of Oral Biomedicine Ministry of Education, School & Hospital of Stomatology, Medical Research Institute, Wuhan University, Wuhan, China, Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China, Department of Immunology, Hubei Province

Key Laboratory of Allergy and Immunology, State Key Laboratory of Virology and Medical Research Institute, Wuhan University School of Basic Medical Sciences, Wuhan, China

For correspondence: yingyingpu@whu.edu.cn

[†]Lead contact

Editors

Reviewing Editor

Dominique Soldati-Favre

University of Geneva, Geneva, Switzerland

Senior Editor

Dominique Soldati-Favre

University of Geneva, Geneva, Switzerland

Reviewer #1 (Public review):

Summary:

In this manuscript, Yan and colleagues introduce a modification to the previously published PETRI-seq bacterial single cell protocol to include a ribosomal depletion step based on a DNA probe set that selectively hybridizes with ribosome-derived (rRNA) cDNA fragments. They show that their modification of the PETRI-seq protocol increases the fraction of informative non-rRNA reads from ~4-10% to 54-92%. The authors apply their protocol to investigating heterogeneity in a biofilm model of *E. coli*, and convincingly show how their technology can detect minority subpopulations within a complex community.

Strengths:

The method the authors propose is a straightforward and inexpensive modification of an established split-pool single cell RNA-seq protocol that greatly increases its utility, and should be of interest to a wide community working in the field of bacterial single cell RNA-seq.

Comments on revised version:

The reviewers have responded thoughtfully and comprehensively to all of my comments. I believe the details of the protocol are now much easier to understand, and the text and methods have been significantly clarified. I have no further comments.

<https://doi.org/10.7554/eLife.97543.3.sa2>

Reviewer #2 (Public review):

Summary:

This work introduces a new method of depleting the ribosomal reads from the single-cell RNA sequencing library prepared with one of the prokaryotic scRNA-seq techniques, PETRI-seq. The advance is very useful since it allows broader access to the technology by lowering the cost of sequencing. It also allows more transcript recovery with fewer sequencing reads. The authors demonstrate the utility and performance of the method for three different model species and find a subpopulation of cells in the *E. coli* biofilm that express a protein, PdeI, which causes elevated c-di-GMP levels. These cells were shown to be in a state that promotes persister formation in response to ampicillin treatment.

Strengths:

The introduced rRNA depletion method is highly efficient, with the depletion for *E. coli* resulting in over 90% of reads containing mRNA. The method is ready to use with existing PETRI-seq libraries which is a large advantage, given that no other rRNA depletion methods were published for split-pool bacterial scRNA-seq methods. Therefore, the value of the method for the field is high. There is also evidence that a small number of cells at the bottom of a static biofilm express PdeI which is causing the elevated c-di-GMP levels that are associated with persister formation. This finding highlights the potentially complex role of PdeI in regulation of c-di-GMP levels and persister formation in microbial biofilms.

Comments on revised version:

The authors edited the manuscript thoroughly in response to the comments, including both performing new experiments and showing more data and information. Most of the major points raised between both reviewers were addressed. The authors explained the seeming contradiction between c-di-GMP levels and PdeI expression.

<https://doi.org/10.7554/eLife.97543.3.sa1>

Author response:

The following is the authors' response to the previous reviews.

eLife Assessment

This work presents an important method for depleting ribosomal RNA from bacterial single-cell RNA sequencing libraries, enabling the study of cellular heterogeneity within microbial biofilms. The approach convincingly identifies a small subpopulation of cells at the biofilm's base with upregulated PdeI expression, offering invaluable insights into the biology of bacterial biofilms and the formation of persister cells. Further integrated analysis of gene interactions within these datasets could deepen our understanding of biofilm dynamics and resilience.

Thank you for your valuable feedback and for recognizing the importance of our method for depleting ribosomal RNA from bacterial single-cell RNA sequencing libraries. We are pleased that our approach has convincingly identified a small subpopulation of cells at the base of the biofilm with upregulated PdeI expression, providing significant insights into the biology of bacterial biofilms and the formation of persister cells.

We acknowledge your suggestion for a more comprehensive analysis of multiple genes and their interactions. While we conducted a broad analysis across the transcriptome, our decision to focus on the heterogeneously expressed gene PdeI was primarily informed by its critical role in biofilm biology. In addition to PdeI, we investigated other marker genes and noted that *lptE* and *sstT* exhibited potential associations with persister cells. However, our interaction analysis revealed that LptE and SstT did not demonstrate significant relationships with c-di-GMP and PdeI based on current knowledge. This insight led us to concentrate on PdeI, given its direct relevance to biofilm formation and its close connection to the c-di-GMP signaling pathway.

We fully agree that other marker genes may also have important regulatory roles in different aspects of biofilm dynamics. Thus, we plan to explore the expression patterns and potential functions of these genes in our future research. Specifically, we intend to conduct more extensive gene network analyses to uncover the complex regulatory mechanisms involved in biofilm formation and resilience.

Public Reviews:**Reviewer #1 (Public review):***Summary:*

*In this manuscript, Yan and colleagues introduce a modification to the previously published PETRI-seq bacterial single cell protocol to include a ribosomal depletion step based on a DNA probe set that selectively hybridizes with ribosome-derived (rRNA) cDNA fragments. They show that their modification of the PETRI-seq protocol increases the fraction of informative non-rRNA reads from ~4-10% to 54-92%. The authors apply their protocol to investigating heterogeneity in a biofilm model of *E. coli*, and convincingly show how their technology can detect minority subpopulations within a complex community.*

Strengths:

The method the authors propose is a straightforward and inexpensive modification of an established split-pool single cell RNA-seq protocol that greatly increases its utility, and should be of interest to a wide community working in the field of bacterial single cell RNA-seq.

We sincerely thank the reviewer for their thoughtful and positive evaluation of our work. We appreciate the recognition of our modification to the PETRI-seq bacterial single-cell RNA sequencing protocol by incorporating a ribosomal depletion step. The significant increase in the fraction of informative non-rRNA reads, as noted in the reviewer's summary, underscores the effectiveness of our method in enhancing the utility of the PETRI-seq approach. We are also encouraged by the reviewer's acknowledgment of our ability to detect minority subpopulations within complex biofilm communities. Our team is committed to further validating and optimizing this method, and we believe that RiboD-PETRI will contribute meaningfully to the field of bacterial single-cell transcriptomics. We hope this innovative approach will facilitate new discoveries in microbial ecology and biofilm research.

Reviewer #2 (Public review):*Summary:*

*This work introduces a new method of depleting the ribosomal reads from the single-cell RNA sequencing library prepared with one of the prokaryotic scRNA-seq techniques, PETRI-seq. The advance is very useful since it allows broader access to the technology by lowering the cost of sequencing. It also allows more transcript recovery with fewer sequencing reads. The authors demonstrate the utility and performance of the method for three different model species and find a subpopulation of cells in the *E. coli* biofilm that express a protein, PdeI, which causes elevated c-di-GMP levels. These cells were shown to be in a state that promotes persister formation in response to ampicillin treatment.*

Strengths:

*The introduced rRNA depletion method is highly efficient, with the depletion for *E. coli* resulting in over 90% of reads containing mRNA. The method is ready to use with existing PETRI-seq libraries which is a large advantage, given that no other rRNA depletion methods were published for split-pool bacterial scRNA-seq methods. Therefore, the value of the method for the field is high. There is also evidence that a small number of cells at the bottom of a static biofilm express PdeI which is causing the elevated c-di-GMP levels that are associated with persister formation. This finding highlights the potentially*

complex role of PdeI in regulation of c-di-GMP levels and persister formation in microbial biofilms.

Weaknesses:

Given many current methods that also introduce different techniques for ribosomal RNA depletion in bacterial single-cell RNA sequencing, it is unclear what is the place and role of RiboD-PETRI. The efficiency of rRNA depletion varies greatly between species for the majority of the available methods, so it is not easy to select the best fitting technique for a specific application.

Thank you for your insightful comments regarding the place and role of RiboD-PETRI in the landscape of ribosomal RNA depletion techniques for bacterial single-cell RNA sequencing. We appreciate the opportunity to address your concerns and clarify the significance of our method.

We acknowledge that the field of rRNA depletion in bacterial single-cell RNA sequencing is diverse, with many methods offering different approaches. We also recognize the challenge of selecting the best technique for a specific application, given the variability in rRNA depletion efficiency across species for many available methods. In light of these considerations, we believe RiboD-PETRI occupies a distinct and valuable niche in this landscape due to following reasons: 1) Low-input compatibility: Our method is specifically tailored for the low-input requirements of single-cell RNA sequencing, maintaining high efficiency even with limited starting material. This makes RiboD-PETRI particularly suitable for single-cell studies where sample quantity is often a limiting factor. 2) Equipment-free protocol: One of the unique advantages of RiboD-PETRI is that it can be conducted in any lab without the need for specialized equipment. This accessibility ensures that a wide range of researchers can implement our method, regardless of their laboratory setup. 3) Broad species coverage: Through comprehensive probe design targeting highly conserved regions of bacterial rRNA, RiboD-PETRI offers a robust solution for samples involving multiple bacterial species or complex microbial communities. This approach aims to provide consistent performance across diverse taxa, addressing the variability issue you mentioned. 4) Versatility and compatibility: RiboD-PETRI is designed to be compatible with various downstream single-cell RNA sequencing protocols, enhancing its utility in different experimental setups and research contexts.

In conclusion, RiboD-PETRI's unique combination of low-input compatibility, equipment-free protocol, broad species coverage, and versatility positions it as a robust and accessible option in the landscape of rRNA depletion methods for bacterial single-cell RNA sequencing. We are committed to further validating and improving our method to ensure its valuable contribution to the field and to provide researchers with a reliable tool for their diverse experimental needs.

Despite transcriptome-wide coverage, the authors focused on the role of a single heterogeneously expressed gene, PdeI. A more integrated analysis of multiple genes and/or interactions between them using these data could reveal more insights into the biofilm biology.

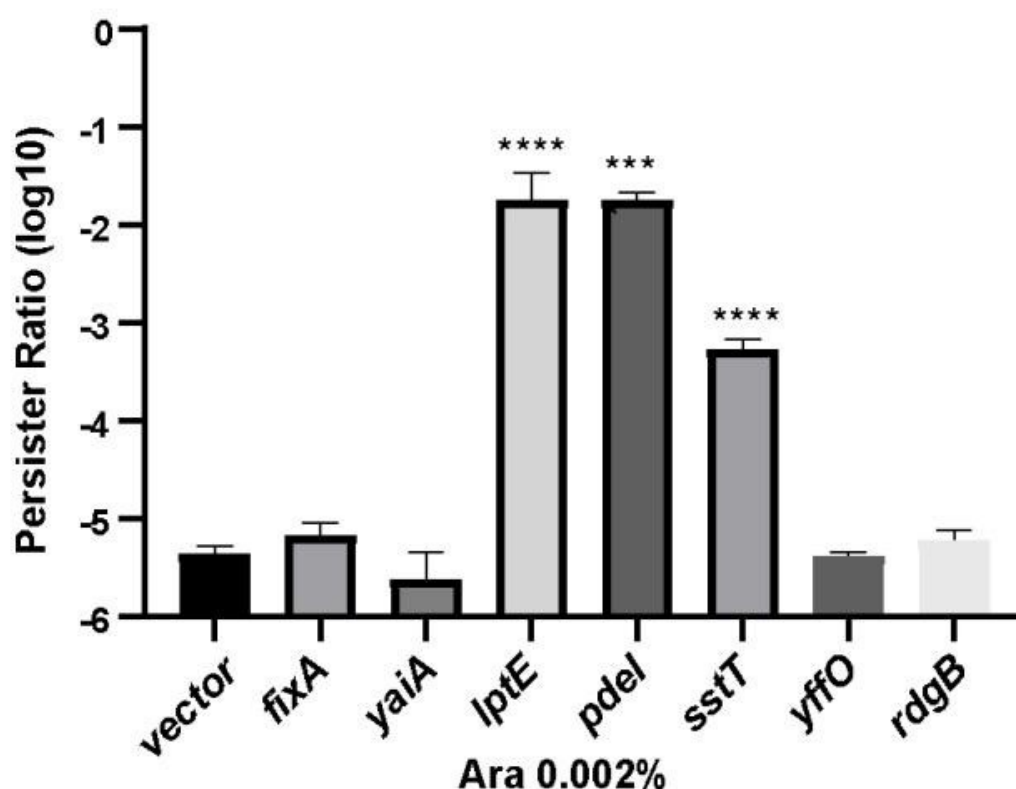
Thank you for your valuable feedback. We understand your suggestion for a more comprehensive analysis of multiple genes and their interactions. While we indeed conducted a broad analysis across the transcriptome, our decision to focus on the heterogeneously expressed gene PdeI was primarily based on its crucial role in biofilm biology. Beyond PdeI, we also conducted overexpression experiments on several other marker genes and examined their phenotypes. Notably, the *lptE* and *sstT* genes showed potential associations with persister cells. We performed an interaction analysis, which revealed that *LptE* and *SstT* did

not show significant relationships with c-di-GMP and PdeI based on current knowledge. This finding led us to concentrate our attention on PdeI. Given PdeI's direct relevance to biofilm formation and its close connection to the c-di-GMP signaling pathway, we believed that an in-depth study of PdeI was most likely to reveal key biological mechanisms.

We fully agree with your point that other marker genes may play regulatory roles in different aspects. The expression patterns and potential functions of these genes will be an important direction in our future research. In our future work, we plan to conduct more extensive gene network analyses to uncover the complex regulatory mechanisms of biofilm formation.

Author response image 1.

The proportion of persister cells in the partially maker genes and empty vector control groups. Following induction of expression with 0.002% arabinose for 2 hours, a persister counting assay was conducted on the strains using 150 µg/ml ampicillin.



The authors should also present the UMIs capture metrics for RiboD-PETRI method for all cells passing initial quality filter (≥ 15 UMIs/cell) both in the text and in the figures. Selection of the top few cells with higher UMI count may introduce biological biases in the analysis (the top 5% of cells could represent a distinct subpopulation with very high gene expression due to a biological process). For single-cell RNA sequencing, showing the statistics for a 'top' group of cells creates confusion and inflates the perceived resolution, especially when used to compare to other methods (e.g. the parent method PETRI-seq itself).

Thank you for your valuable feedback regarding the presentation of UMI capture metrics for the RiboD-PETRI method. We appreciate your concern about potential biological biases and the importance of comprehensive data representation in single-cell RNA sequencing analysis.

We have now included the UMI capture metrics for all cells passing the initial quality filter (≥ 15 UMIs/cell) for the RiboD-PETRI method. This information has been added to both the main text and the relevant figures, providing a more complete picture of our method's performance across the entire range of captured cells. These revisions strengthen our manuscript and provide readers with a more complete understanding of the RiboD-PETRI method in the context of single-cell RNA sequencing.

Recommendations for the authors:

Reviewer #1 (Recommendations for the authors):

The reviewers have responded thoughtfully and comprehensively to all of my comments. I believe the details of the protocol are now much easier to understand, and the text and methods have been significantly clarified. I have no further comments.

Reviewer #2 (Recommendations for the authors):

The authors edited the manuscript thoroughly in response to the comments, including both performing new experiments and showing more data and information. Most of the major points raised between both reviewers were addressed. The authors explained the seeming contradiction between c-di-GMP levels and PdeI expression. Despite these improvements, a few issues remain:

- Despite now depositing the data and analysis files to GEO, the access is embargoed and the reviewer token was not provided to evaluate the shared data and accessory files.

Please note that although the data and analysis files have been deposited to GEO, access is currently embargoed. To evaluate the shared data and accessory files, you will need a reviewer token, which appears to have not been provided.

To gain access, please follow these steps:

Visit the GEO accession page at: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE260458>

In the designated field, enter the reviewer token: ehipgqiohhcvjev

- Despite now discussing performance metrics for RiboD-PETRI method for all cells passing initial quality filter (≥ 15 UMIs/cell) in the text, the authors continued to also include the statistics for top 1000 cells, 5,000 cells and so on. Critically, Figure 2A-B is still showing the UMI and gene distributions per cell only for these select groups of cells. The intent to focus on these metrics is not quite clear, as selection of the top few cells with higher UMI count may introduce biological biases in the analysis (what if the top 5% of cells are unusual because they represent a distinct subpopulation with very high gene expression due to a biological process). I understand the desire to demonstrate the performance of the method by highlighting a few select 'best' cells, however, for single-cell RNA sequencing showing the statistics for a 'top' group of cells is not appropriate and creates confusion, especially when used to compare to other methods (e.g. the parent method PETRI-seq itself).

We appreciate your insightful feedback regarding our presentation of the RiboD-PETRI method's performance metrics. We acknowledge the concerns you've raised and agree that our current approach requires refinement. We have revised our analysis to prominently feature metrics for all cells that pass the initial quality filter (≥ 15 UMIs/cell) (Fig. 2A, Fig. 3A, Supplementary Fig. 1A, B and Supplementary Fig. 2A, G). This approach provides a more representative view of the method's performance across the entire dataset, avoiding potential biases introduced by focusing solely on top-performing cells.

We recognize that selecting only the top cells based on UMI counts can indeed introduce biological biases, as these cells may represent distinct subpopulations with unique biological processes rather than typical cellular states. To address this, we have clearly stated the potential for bias when highlighting select 'best' cells. We also provided context for why these high-performing cells are shown, explaining that they demonstrate the upper limits of the method's capabilities (lines 139). In addition, when comparing RiboD-PETRI to other methods, including the parent PETRI-seq, we ensured that comparisons are made using consistent criteria across all methods.

By implementing these changes, we aim to provide a more accurate, unbiased, and comprehensive representation of the RiboD-PETRI method's performance while maintaining scientific rigor and transparency. We appreciate your critical feedback, as it helps us improve the quality and reliability of our research presentation.

- Line 151 "The findings reveal that our sequencing saturation is 100% (Fig. S1B, C)" - I suggest the authors revisit this calculation as this parameter is typically very challenging to get above 95-96%. The sequencing saturation should be calculated from the statistics of alignment themselves, i.e. the parameter calculated by Cell Ranger as described here <https://kb.10xgenomics.com/hc/en-us/articles/115003646912-How-is-sequencing-saturation-calculated> :

"The web_summary.html output from cellranger count includes a metric called "Sequencing Saturation". This metric quantifies the fraction of reads originating from an already-observed UMI. More specifically, this is the fraction of confidently mapped, valid cell-barcode, valid UMI reads that are non-unique (match an existing cell-barcode, UMI, gene combination).

The formula for calculating this metric is as follows:

$$\text{Sequencing Saturation} = 1 - (n_{\text{deduped_reads}} / n_{\text{reads}})$$

where

$n_{\text{deduped_reads}}$ = Number of unique (valid cell-barcode, valid UMI, gene) combinations among confidently mapped reads.

n_{reads} = Total number of confidently mapped, valid cell-barcode, valid UMI reads.

Note that the numerator of the fraction is $n_{\text{deduped_reads}}$, not the non-unique reads that are mentioned in the definition. $n_{\text{deduped_reads}}$ is a degree of uniqueness, not a degree of duplication/saturation. Therefore we take the complement of $(n_{\text{deduped_reads}} / n_{\text{reads}})$ to measure saturation."

We appreciate your insightful comment regarding our sequencing saturation calculation. The sequencing saturation algorithm we initially employed was based on the methodology used in the BacDrop study (PMID: PMC10014032, <https://pmc.ncbi.nlm.nih.gov/articles/PMC10014032/>).

We acknowledge the importance of using standardized and widely accepted methods for calculating sequencing saturation. As per your suggestion, we have recalculated our sequencing saturation using the method described by 10x Genomics. Given the differences between RiboD-PETRI and 10x Genomics datasets, we have adapted the calculation as follows:

- $n_{\text{deduped_reads}}$: We used the number of UMIs as a measure of unique reads.
- n_{reads} : We used the total number of confidently mapped reads.

After applying this adapted calculation method, we found that our sequencing saturation ranges from 92.16% to 93.51%. This range aligns more closely with typical expectations for sequencing saturation in single-cell RNA sequencing experiments, suggesting that we have captured a substantial portion of the transcript diversity in our samples. We also updated Figure S1 to reflect these recalculated sequencing saturation values. We will also provide a detailed description of our calculation method in the methods section to ensure transparency and reproducibility. It's important to note that this saturation calculation method was originally designed for 10× Genomics data. While we've adapted it for our study, we acknowledge that its applicability to our specific experimental setup may be limited.

We thank you for bringing this important point to our attention. This recalculation not only improves the accuracy of our reported results but also aligns our methodology more closely with established standards in the field. We believe these revisions strengthen the overall quality and reliability of our study.

- Further, this calculated saturation should be taken into account when comparing the performance of the method in terms of retrieving diverse transcripts from cells. I.e., if the RiboD-Petri dataset was subsampled to the same saturation as the original PETRI-seq dataset was obtained with, would the median UMIs/cell for all cells above filter be comparable? In other words, does rRNA depletion just decreases the cost to sequence to saturation, or does it provide UMI capture benefits at a comparable saturation?

We appreciate your insightful question regarding the comparison of method performance in terms of transcript retrieval diversity and the impact of saturation. To address your concerns, we conducted an additional analysis comparing the RiboD-PETRI and original PETRI-seq datasets at equivalent saturation levels besides our original analysis with equivalent sequencing depth.

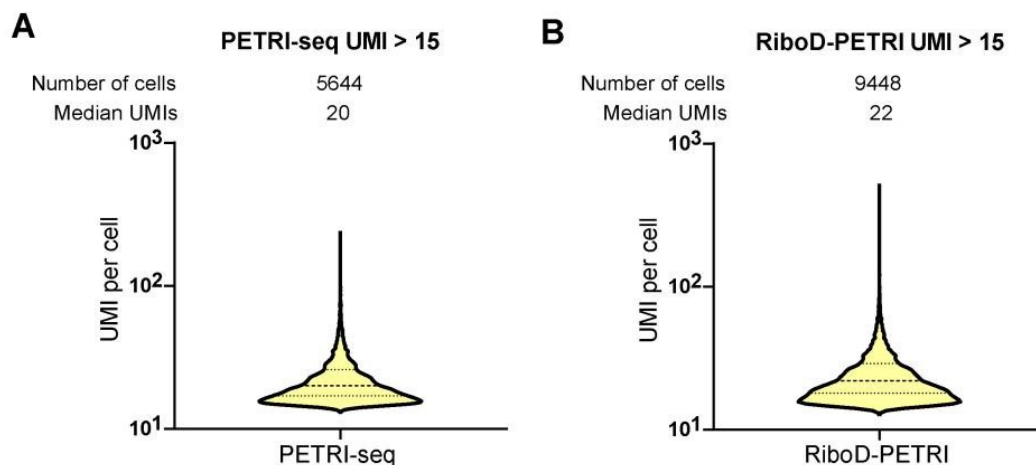
With equivalent sequencing depth, RiboD-PETRI demonstrates a significantly enhanced Unique Molecular Identifier (UMI) counts detection rate compared to PETRI-seq alone (Fig. 1C). This method recovered approximately 20175 cells (92.6% recovery rate) with ≥ 15 UMIs per cell with a median UMI count of 42 per cell, which was significantly higher than PETRI-seq's recovery rate of 17.9% with a median UMI count of 20 per cell (Figure S1A, B), indicating the number of detected mRNA per cell increased prominently.

When we subsampled the RiboD-PETRI dataset to match the saturation level of the original PETRI-seq dataset (i.e., equalizing the $n_deduped_reads/n_reads$ ratio), we found that the median UMIs/cell for all cells above the filter threshold was higher in the RiboD-PETRI dataset compared to the original PETRI-seq (as shown in Author response image 2). This observation can be primarily attributed to the introduction of the rRNA depletion step in the RiboD-PETRI method. Our analysis suggests that rRNA depletion not only reduces the cost of sequencing to saturation but also provides additional benefits in UMI capture efficiency at comparable saturation levels. The rRNA depletion step effectively reduces the proportion of rRNA-derived reads in the sequencing output. Consequently, at equivalent saturation levels, this leads to a relative increase in the number of $n_deduped_reads$ corresponding to mRNA transcripts. This shift in read composition enhances the capture of informative UMIs, resulting in improved transcript diversity and detection.

In conclusion, our findings indicate that the rRNA depletion step in RiboD-PETRI offers dual advantages: it decreases the cost to sequence to saturation and provides enhanced UMI capture benefits at comparable saturation levels, ultimately leading to more efficient and informative single-cell transcriptome profiling.

Author response image 2.

At almost the same sequencing saturation (64% and 67%), the number of cells exceeding the screening criteria (≥ 15 UMIs) and the median number of UMIs in cells in Ribod-PETRI and PETRI-seq data of exponential period *E. coli* (3h).



- *smRandom-seq* and *BaSSSh-seq* need to also be discussed since these newer methods are also demonstrating rRNA depletion techniques. (<https://doi.org/10.1038/s41467-023-40137-9> and <https://doi.org/10.1101/2024.06.28.601229>)

Thank you for your valuable feedback. We appreciate the opportunity to discuss our method, Ribod-PETRI, in the context of other recent advances in bacterial RNA sequencing techniques, particularly *smRandom-seq* and *BaSSSh-seq*.

Ribod-PETRI employs a Ribosomal RNA-derived cDNA Depletion (Ribod) protocol. This method uses probe primers that span all regions of the bacterial rRNA sequence, with the 3'-end complementary to rRNA-derived cDNA and the 5'-end complementary to a biotin-labeled universal primer. After hybridization, Streptavidin magnetic beads are used to eliminate the hybridized rRNA-derived cDNA, leaving mRNA-derived cDNA in the supernatant. *smRandom-seq* utilizes a CRISPR-based rRNA depletion technique. This method is designed for high-throughput single-microbe RNA sequencing and has been shown to reduce the rRNA proportion from 83% to 32%, effectively increasing the mRNA proportion four times (from 16% to 63%). While specific details about *BaSSSh-seq*'s rRNA depletion technique are not provided in the available information, it is described as employing a rational probe design for efficient rRNA depletion. This technique aims to minimize the loss of mRNA during the depletion process, ensuring a more accurate representation of the transcriptome.

Ribod-PETRI demonstrates significant enhancement in rRNA-derived cDNA depletion across both gram-negative and gram-positive bacterial species. It increases the mRNA ratio from 8.2% to 81% for *E. coli* in exponential phase, from 10% to 92% for *S. aureus* in stationary phase, and from 3.9% to 54% for *C. crescentus* in exponential phase. *smRandom-seq* shows high species specificity (99%), a minor doublet rate (1.6%), and a reduced rRNA percentage (32%). These metrics indicate its efficiency in single-microbe RNA sequencing. While specific performance metrics for *BaSSSh-seq* are not provided in the available information, its rational probe design approach suggests a focus on maintaining mRNA integrity during the depletion process.

RiboD-PETRI is described as a cost-effective (\$0.0049 per cell), equipment-free, and high-throughput solution for bacterial scRNA-seq. This makes it an attractive option for researchers with budget constraints. While specific cost information is not provided, the efficiency of smRandom-seq is noted to be affected by the overwhelming quantity of rRNAs (>80% of mapped reads). The CRISPR-based depletion technique likely adds to the complexity and cost of the method. Cost and accessibility information for BaSSSh-seq is not provided in the available data, making a direct comparison difficult.

All three methods represent significant advancements in bacterial RNA sequencing, each offering unique approaches to the challenge of rRNA depletion. RiboD-PETRI stands out for its cost-effectiveness and demonstrated success in complex systems like biofilms. Its ability to significantly increase mRNA ratios across different bacterial species and growth phases is particularly noteworthy. smRandom-seq's CRISPR-based approach offers high specificity and efficiency, which could be advantageous in certain research contexts, particularly where single-microbe resolution is crucial. However, the complexity of the CRISPR system might impact its accessibility and cost-effectiveness. BaSSSh-seq's focus on minimizing mRNA loss during depletion could be beneficial for studies requiring highly accurate transcriptome representations, although more detailed performance data would be needed for a comprehensive comparison. The choice between these methods would depend on specific research needs. RiboD-PETRI's cost-effectiveness and proven application in biofilm studies make it particularly suitable for complex bacterial community analyses. smRandom-seq might be preferred for studies requiring high-throughput single-cell resolution. BaSSSh-seq could be the method of choice when preserving the integrity of the mRNA profile is paramount.

In conclusion, while all three methods offer valuable solutions for rRNA depletion in bacterial RNA sequencing, RiboD-PETRI's combination of efficiency, cost-effectiveness, and demonstrated application in complex biological systems positions it as a highly competitive option in the field of bacterial transcriptomics.

We have revised our discussion in the manuscript according to the above analysis (lines 116-119)

- Ctrl and Delta-Delta abbreviations are used in main text but not defined there (lines 107-110).

Thank you for your valuable feedback. We have now defined the abbreviations "Ctrl" and "Delta-Delta" in the main text for clarity.

- The utility of Figs 2E and 3E is questionable - the same information can be conveyed in text.

Thank you for your thoughtful observation regarding Figures 2E and 3E. We appreciate your feedback and would like to address the concerns you've raised.

While we acknowledge that some of the information in these figures could be conveyed textually, we believe that their visual representation offers several advantages. Figures 2E and 3E provide a comprehensive visual overview of the pathway enrichment analysis for marker genes, which may be more easily digestible than a textual description. This analysis was conducted in response to another reviewer's request, demonstrating our commitment to addressing diverse perspectives in our research.

These figures allow for a systematic interpretation of gene expression data, revealing complex interactions between genes and their involvement in biological pathways that might be less apparent in a text-only format. Visual representations can make complex data more

accessible to readers with different learning styles or those who prefer graphical summaries. Additionally, including such figures is consistent with standard practices in our field, facilitating comparison with other studies. We believe that the pathway enrichment analysis results presented in these figures provide valuable insights that merit inclusion as visual elements. However, we are open to discussing alternative ways to present this information if you have specific suggestions for improvement.

<https://doi.org/10.7554/eLife.97543.3.sa0>