LETTERS
https://doi.org/10.1038/s41564-020-0729-6
nature microbiology

Check for updates

Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing

Sydney B. Blattman[1,2,3,4], Wenyan Jiang[1,2,3,4], Panos Oikonomou[1,2,3] and Saeed Tavazoie[1,2,3] ✉

Despite longstanding appreciation of gene expression heterogeneity in isogenic bacterial populations, affordable and scalable technologies for studying single bacterial cells have been limited. Although single-cell RNA sequencing (scRNA-seq) has revolutionized studies of transcriptional heterogeneity in diverse eukaryotic systems[1–13], the application of scRNA-seq to prokaryotes has been hindered by their extremely low mRNA abundance[14–16], lack of mRNA polyadenylation and thick cell walls[17]. Here, we present prokaryotic expression profiling by tagging RNA in situ and sequencing (PETRI-seq)—a low-cost, high-throughput prokaryotic scRNA-seq pipeline that overcomes these technical obstacles. PETRI-seq uses in situ combinatorial indexing[11,12,18] to barcode transcripts from tens of thousands of cells in a single experiment. PETRI-seq captures single-cell transcriptomes of Gram-negative and Gram-positive bacteria with high purity and low bias, with median capture rates of more than 200 mRNAs per cell for exponentially growing *Escherichia coli*. These characteristics enable robust discrimination of cell states corresponding to different phases of growth. When applied to wild-type *Staphylococcus aureus*, PETRI-seq revealed a rare subpopulation of cells undergoing prophage induction. We anticipate that PETRI-seq will have broad utility in defining single-cell states and their dynamics in complex microbial communities.

Recent developments in high-throughput scRNA-seq technology have enabled the rapid characterization of cellular diversity within complex eukaryotic tissues[1–13]. Despite these advances, comparable tools for bacteria have lagged behind due to numerous technical challenges (Supplementary Fig. 1). Current massively parallel eukaryotic scRNA-seq methods typically require custom microfluidics to coencapsulate a single cell with a uniquely barcoded bead in a compartment, often a droplet[5,6,8] or microwell[4,7]. These approaches rely on two key properties of many eukaryotic cells, specifically that they are easily lysed with detergent to release their RNA and that their polyadenylated mRNAs can be effectively captured by beads coated with poly(dT) primers. Adaptation of these approaches for bacteria is thwarted by the presence of a thick prokaryotic cell wall[17], which makes lysis challenging, and the lack of polyadenylated mRNAs for effective capture. Given these considerations, we identified in situ combinatorial indexing[18] as an alternative basis on which to develop high-throughput prokaryotic scRNA-seq. Two conceptually similar eukaryotic methods—single-cell combinatorial indexing RNA sequencing (sci-RNA-seq)[11,13] and split-pool ligation-based transcriptome sequencing (SPLiT-seq)[12]—rely on cells as compartments for barcoding, which abrogates the need for cell lysis in droplets or microwells. These methods are also amenable to reverse transcription (RT) with random hexamers rather than poly(dT) primers[12]. With only pipetting steps and no complex instruments, individual transcriptomes of hundreds of thousands of fixed cells are uniquely labelled by multiple rounds of splitting, barcoding and pooling in microplates.

Here we present prokaryotic expression profiling by tagging RNA in situ and sequencing (PETRI-seq)—a high-throughput, affordable and easy-to-perform scRNA-seq method that is capable of profiling the gene expression states of tens of thousands of wild-type Gram-positive (*Staphylococcus aureus* USA300) and Gram-negative (*Escherichia coli* MG1655) cells. PETRI-seq (Fig. 1) consists of three experimental components: cell preparation, split-pool barcoding and library preparation, which are described in Extended Data Fig. 1a–c and the Methods. Cells grown in liquid culture were briefly pelleted before overnight fixation with 4% formaldehyde. We confirmed that centrifugation and fixation did not alter the bulk transcriptome (Extended Data Fig. 2a–c). Cells were next resuspended in 50% ethanol, which has been used previously for prokaryotic in situ PCR as a storage solution[19], although we have yet to test cellular and RNA integrity after long-term storage. Ethanol did not substantially change the cDNA yield from in situ RT (Extended Data Fig. 2d). Lysozyme (for *E. coli*; Extended Data Fig. 2e) or lysostaphin (for *S. aureus*) was subsequently added to permeabilize cells for in situ RT. Cells were next treated with DNase to remove background genomic DNA. We confirmed in situ DNase activity using quantitative PCR (qPCR; Extended Data Fig. 2f) and verified DNase inactivation (Extended Data Fig. 2g,h). DNase treatment did not significantly alter the bulk transcriptome (Extended Data Fig. 2i) or RNA integrity (Extended Data Fig. 2j). Before proceeding to RT, cells were imaged to confirm that they were intact (Extended Data Fig. 2k) and counted.

In the next stage, we performed split-pool barcoding. Cells were distributed across a microplate for RT with barcoded random hexamers specific to each well. After RT, cells were pooled and redistributed across new microplates for two rounds of barcoding by ligation to the cDNA. We reduced the length of the overhang for each ligation relative to the eukaryotic protocol[12] without compromising ligation efficiency (Extended Data Fig. 2l). This enabled us to perform 75 cycles of sequencing rather than 150, thereby reducing sequencing cost by around 50% (Supplementary Table 1b). After three rounds of barcoding, cells contained cDNA labelled with 1 out of nearly 1 million possible three-barcode combinations (BCs). We counted and lysed ~10,000 cells for library preparation. The number of cells was chosen to ensure a low multiplet frequency, which is the percentage of non-empty BCs that contain more than one cell[20]. For 10,000 cells, the expected multiplet frequency on the basis of a Poisson distribution is 0.56%.

[1]Department of Biological Sciences, Columbia University, New York City, NY, USA. [2]Department of Systems Biology, Columbia University, New York City, NY, USA. [3]Department of Biochemistry and Molecular Biophysics, Columbia University, New York City, NY, USA. [4]These authors contributed equally: Sydney B. Blattman, Wenyan Jiang. ✉e-mail: st2744@columbia.edu

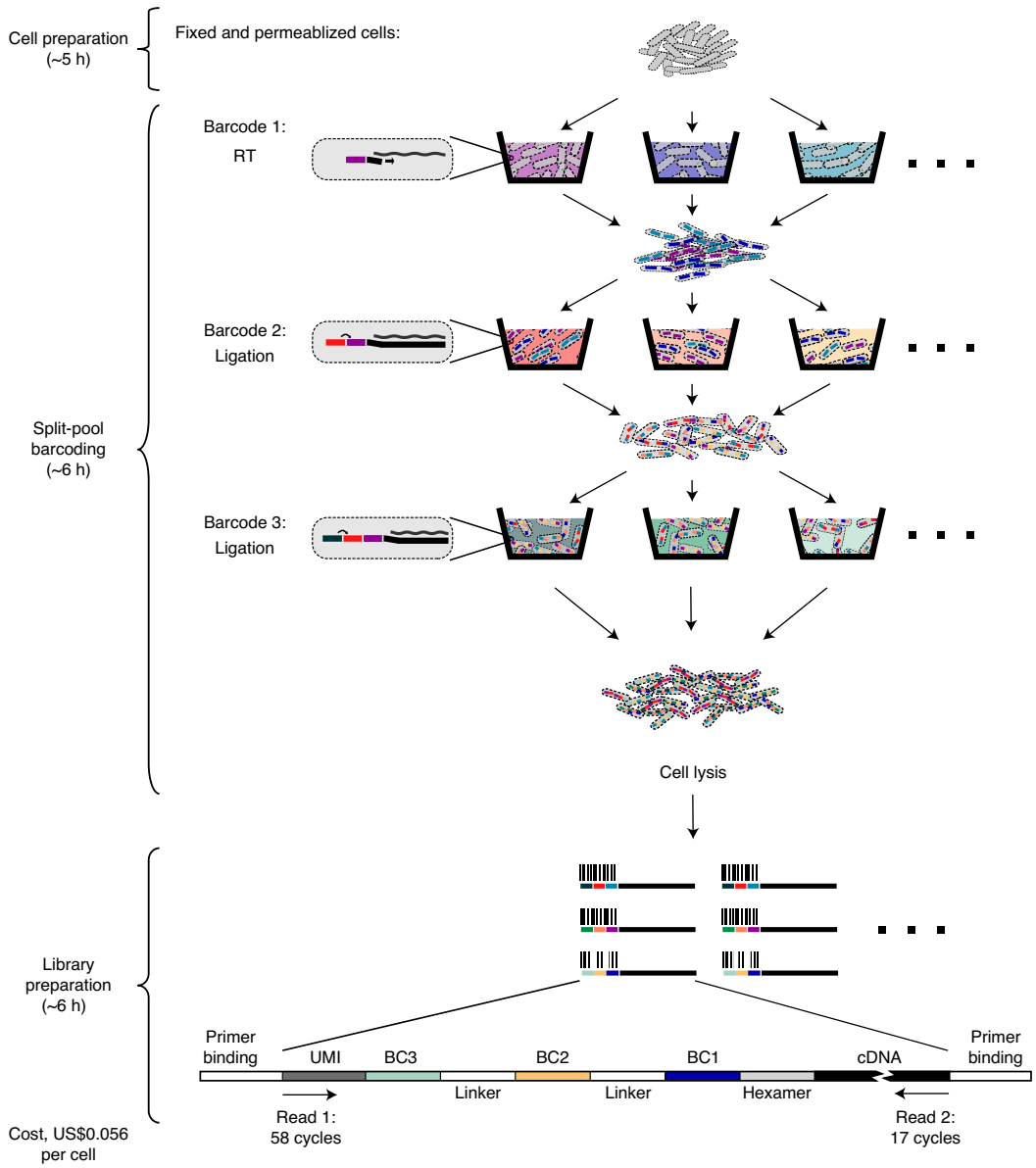


Fig. 1 | Overview of PETRI-seq. PETRI-seq includes three parts—cell preparation, split-pool barcoding and library preparation. During cell preparation, cells are prepared for in situ reactions by fixation (formaldehyde) and permeabilization (lysozyme/lysostaphin). During split-pool barcoding, cells are split across 96-well plates three times for three rounds of barcoding by RT and two ligations. After barcoding, cells are lysed to release cDNA, which is subsequently prepared for paired-end Illumina sequencing. Each cDNA fragment in the library includes a UMI and three barcodes, which are all sequenced in read 1. The UMI is a sequence of seven degenerate nucleotides that can distinguish between unique transcripts and PCR duplicates. The three barcodes comprise a BC, which enables reads to be grouped by their cell of origin. In read 2, the cDNA is sequenced.

Finally, cDNA was prepared for Illumina sequencing. We used AMPure XP beads to purify cDNA from cell lysates (Extended Data Fig. 2m). AMPure purification is faster and less costly (Supplementary Table 1c) than primer biotinylation and streptavidin purification used previously in SPLiT-seq[12]. Next, to generate double-stranded cDNA, we compared second-strand synthesis[21] and limited-cycle PCR after template switching[2]. Second-strand synthesis had a significantly higher yield (Extended Data Fig. 2n,o). We then performed tagmentation followed by PCR using the transposon-inserted sequence and the overhang upstream of the third barcode as primer sequences, thereby preventing amplification of any undigested genomic DNA. The libraries were sequenced and analysed using the pipeline described in Extended Data Fig. 1d–g and the Methods to generate a count matrix of operons by BC. We then set a threshold on the basis of total unique molecular identifiers (UMIs)[22] per BC to distinguish cells from the background signal (Extended Data Fig. 1h,i).

To demonstrate the ability of PETRI-seq to capture transcriptomes of single cells, we performed a species-mixing experiment involving three populations of cells in the exponential phase—green fluorescent protein (GFP)- and red fluorescent protein (RFP)-expressing *E. coli* and wild type *S. aureus* (Fig. 2a). From 14,975 sequenced BCs, we observed that BCs were highly species-specific with 99.8% clearly assigned to one species (Fig. 2b). We calculated an overall multiplet frequency of 1.5% after accounting for multiplets of the same species and non-equal representation of the two species[20]. Although this frequency exceeds the Poisson expectation of 0.85%, it is comparable to contemporary

eukaryotic methods[8,13]. Furthermore, within the *E. coli* population, BCs were highly strain-specific with 98.7% of plasmid-containing cells assigned to a single population (GFP or RFP expressing; Fig. 2c). Multiplet frequency is the probability of multiple cells travelling together during barcoding either by physical interaction or by chance; however, additional factors, such as barcoded free molecules released by occasional cell lysis, may compromise single-cell purity. This type of intercellular contamination has been described for eukaryotic scRNA-seq[23,24]. To assess the contamination rate (the probability that a UMI in a single cell is derived from other cells) for PETRI-seq, we first excluded species-mixed multiplets and then found that BCs assigned as *E. coli* included a mean of 0.23% *S. aureus* UMIs (Extended Data Fig. 3a, right), while BCs assigned as *S. aureus* also included a mean of 0.23% *E. coli* UMIs (Extended Data Fig. 3b, right). After correcting for alignment ambiguities (Extended Data Fig. 3e,f,i,j) and relative representation of the two species in the library, we calculated that 0.19–0.36% of UMIs in a PETRI-seq transcriptome were probably derived from other cells.

Performing molecular reactions inside of cells raises the possibility that RNA capture could be biased by specific cellular contexts. Previous results in eukaryotic cells revealed a capture bias against rRNAs during in situ RT[12], which is mildly recapitulated in our data (Extended Data Fig. 4a,b). For exponential-phase *E. coli*, 15% of sense PETRI-seq UMIs mapped to mRNA (Extended Data Fig. 4a, pie chart), whereas only 5% of bulk sense reads mapped to mRNA (Extended Data Fig. 4c, pie chart). Despite the capture bias against rRNA, we observed strong correlations between combined single-cell transcriptomes from PETRI-seq and bulk cDNA libraries prepared using standard RT for both *E. coli* and *S. aureus* (Fig. 2d,e). We also observed that reads mapped across the entire length of operons with minor bias against the 3′ end (Supplementary Fig. 2a), which was at least partially expected from our library preparation protocol (Supplementary Fig. 2b). Our single-cell transcriptomes were reproducible, as shown by the strong correlation between the aggregated transcriptomes of GFP-expressing *E. coli* cells from two independent libraries (Fig. 2f).

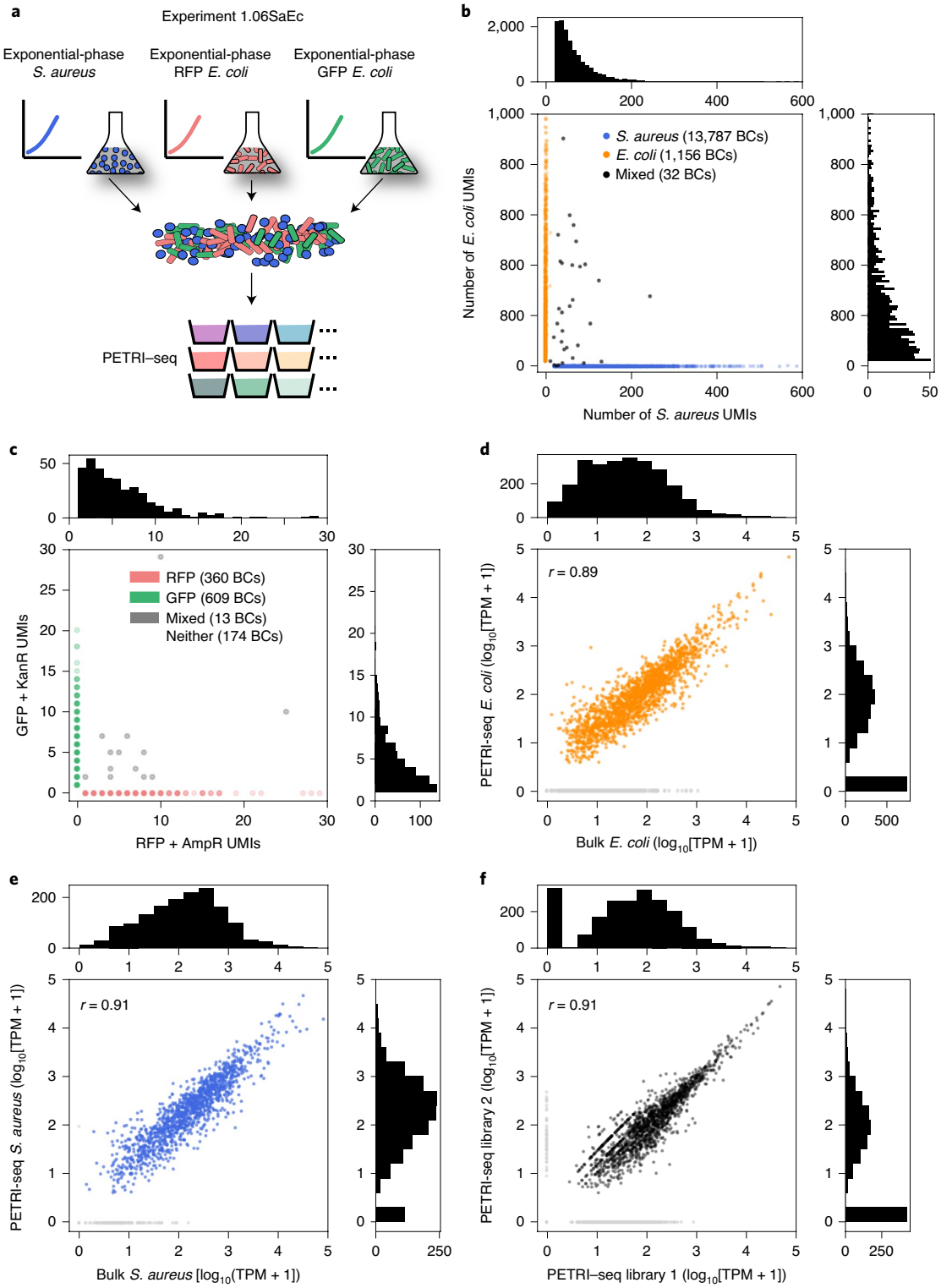
Having confirmed that PETRI-seq captured transcriptomes of single cells with high purity and low bias, we next sought to determine the ability of PETRI-seq to distinguish between cells in different growth states. In experiment 1.10, we mixed *E. coli* cells in two growth phases to create a population resembling naturally arising transcriptional heterogeneity. The mixed population consisted of GFP-expressing exponential-phase and anhydrotetracycline (aTc)-induced RFP-expressing stationary-phase *E. coli* (Fig. 3a). We applied unsupervised dimensionality reduction (principal component analysis (PCA))[25] to visualize the low-dimensional structure that underlies the diversity of transcriptional states. For the PCA, we considered only cells containing at least 15 mRNAs to avoid spurious effects from cells with extremely low mRNA content (Extended Data Fig. 1j,k). Without considering plasmid genes, we observed robust separation of two populations along PC1. We then used the plasmid genes to classify these populations as RFP-containing stationary-phase and GFP-containing exponential-phase cells (Fig. 3b, bottom). We found that 98.5% of all plasmid-containing cells were on the expected side of an empirically chosen threshold line, and the threshold line predicted RFP cells with a 98.59% true positive rate (TPR) to the left of the line and GFP cells with a 98.53% TPR to the right. Of the 7,387 cells analysed, 61% did not contain any plasmid transcripts; their growth states were therefore ambiguous at first (grey points in PCA). However, using the PC1 threshold, we predicted that 92.2% of these were stationary-phase cells. Over-representation of stationary-phase cells in the ambiguous population was not surprising, as plasmid expression in stationary-phase cells was generally lower than in exponential-phase cells. Importantly, separation of the two transcriptional states was similarly robust in another biological replicate (Extended Data Fig. 5a) or when operon counts were normalized using sctransform[26], an alternative method (Extended Data Fig. 5b). Finally, we investigated expression patterns for operons and Gene Ontology (GO) terms and found that many expected trends related to the transition from exponential growth to stationary phase (Fig. 3b, Extended Data Fig. 5c). For example, *rpoS*, which encodes the stationary-phase sigma-factor[27], and *dps*, which encodes a DNA-binding protein highly expressed in stationary phase[28], were upregulated along PC1, as expected, in the direction of stationary-phase cells (Fig. 3b, top). Consistent with induction of the stringent response[29], stationary-phase cells showed a large-scale reduction in ribosomal protein expression as well as an increase in the expression of amino acid biosynthetic operons (Fig. 3b, middle).

We sought to further improve mRNA capture and evaluate the power of PETRI-seq to distinguish different transcriptional states. To this end, we barcoded exponential- and stationary-phase *E. coli* cells separately during RT before pooling them for subsequent steps in experiment 2.01 (Fig. 3c). First, by further permeabilizing cells with detergent before ligation and using a higher concentration of ligation primers (Extended Data Fig. 6a–d), we substantially improved the capture in experiment 2.01 compared with the previous experiments (Extended Data Figs. 4d–f and 5a,d). Specifically, experiment 2.01 captured a median of 227 and 27 UMIs aligned to mRNA (hereafter referred to as mRNA UMIs) per exponential- and stationary-phase *E. coli* cell, respectively (Fig. 3d), corresponding to a median of 103 or 24 operons represented per cell (Fig. 3e). Previous studies have similarly found fewer RNAs in stationary-phase relative to exponential-phase *E. coli* cells[30]. On the basis of estimates that single exponentially growing *E. coli* cells

Fig. 2 | PETRI-seq captures transcriptomes of single *E. coli* and *S. aureus* cells with high purity and low bias. a, Schematic of the species-mixing experiment (experiment 1.06SaEc). Exponential-phase *S. aureus* and *E. coli* cells were grown separately, and then mixed for PETRI-seq after cell preparation. The *E. coli* cells included two populations; exponential-phase RFP-expressing *E. coli* and exponential-phase GFP-expressing *E. coli* were grown separately, and then mixed for cell preparation and PETRI-seq. **b**, Species mixing plot for *E. coli* and *S. aureus* on the basis of total UMIs per BC, including rRNA. BCs were assigned to a single species if more than 90% of UMIs mapped to that species and fewer than 20 UMIs mapped to the other species. The numbers of *S. aureus* and *E. coli* cells with the corresponding number of total UMIs are shown (top and right, respectively). BCs with fewer than 20 total UMIs were omitted. The multiplet frequency is 1.5%. **c**, Quantification of BC collisions within the *E. coli* population by plasmid mRNAs. Cells without plasmid genes (Neither) are omitted. BCs were assigned to a single cell type when more than 90% of plasmid UMIs matched a single plasmid. The numbers of RFP BCs and GFP BCs with the corresponding number of plasmid UMIs are shown (top and right, respectively). **d**, Correlation between mRNA abundances from PETRI-seq versus a bulk library prepared from fixed *E. coli* cells. The Pearson correlation coefficient (r) was calculated for 1,873 out of 2,617 total operons, excluding those with zero counts in either library (grey points). If all operons are included, $r = 0.78$. **e**, Correlation between mRNA abundances from PETRI-seq versus a bulk library prepared from fixed *S. aureus* cells. Pearson's r was calculated for 1,395 out of 1,510 total operons, excluding those with zero counts in either library (grey points). If all operons are included, $r = 0.89$. **f**, Correlation between two biological replicate libraries of exponential-phase GFP-expressing *E. coli* prepared by PETRI-seq. Pearson's r was calculated for 1,714 out of 2,617 total operons, excluding those with zero counts in either library (grey points). If all operons are included, $r = 0.78$. For all correlations (**e**–**g**), PETRI-seq TPM was calculated from UMIs, and bulk TPM was calculated from reads.

contain 2,000–8,000 mRNAs[14–16], we estimated that our capture rate is approximately 2.5–10%. For *S. aureus*, we captured a median of 43 mRNA UMIs per cell (Extended Data Fig. 6e). *S. aureus* cells may contain fewer mRNAs than *E. coli* cells, possibly due to their smaller cell size and genome[31], although technical differences may affect capture. Importantly, we confirmed that optimized PETRI-seq continued to capture single cells with high purity (Extended Data Fig. 7), comparable to eukaryotic scRNA-seq techniques[23,24], and robustly discriminate single *E. coli* cells by growth state (Fig. 3f). Comparison of the subpopulations in experiments 1.10 and 2.01 corroborated the single-cell purity of PETRI-seq (Extended Data Fig. 8).

Given around 20–200 mRNA UMIs captured per average bacterial cell, we anticipate that future PETRI-seq studies will benefit



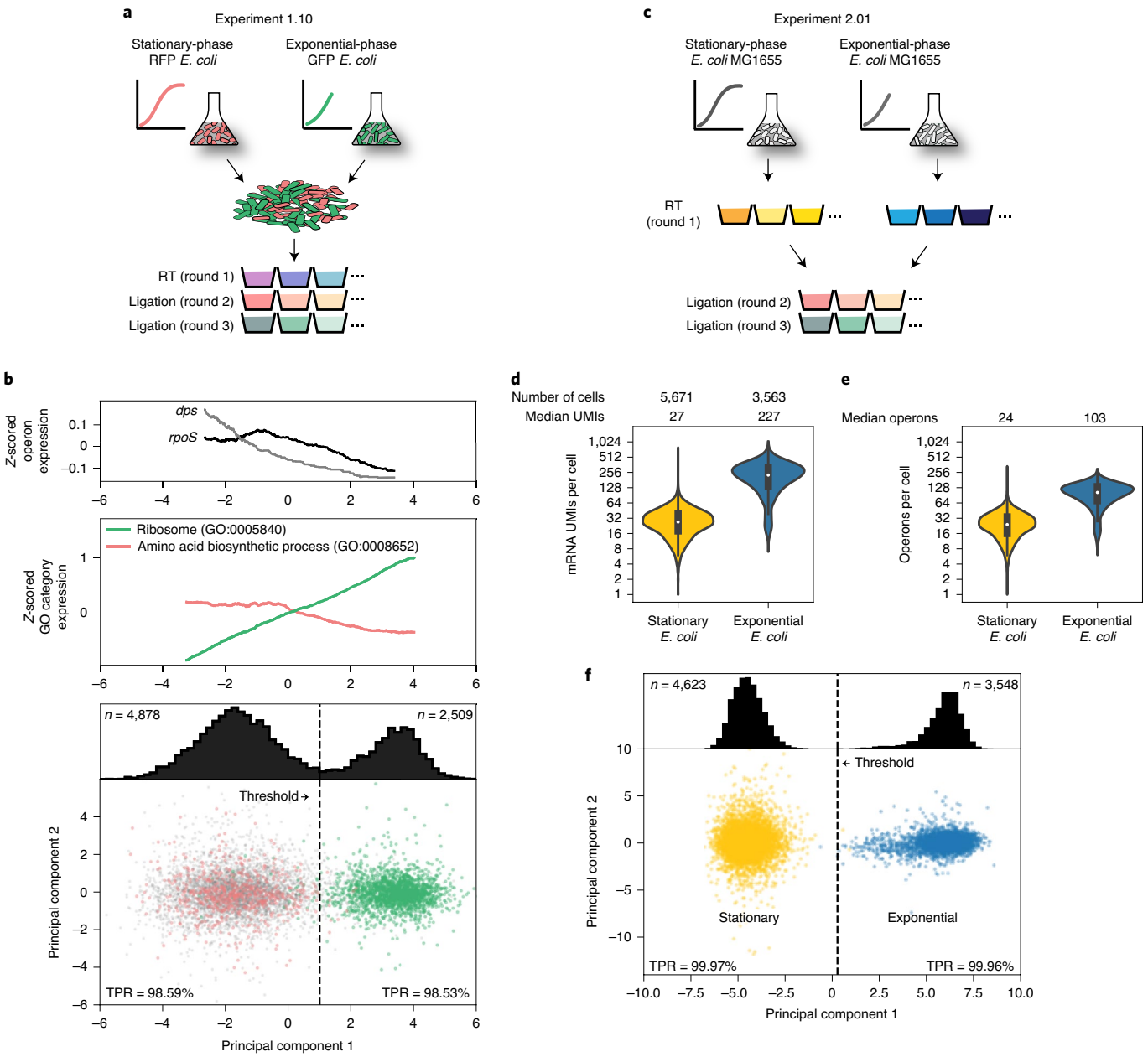


Fig. 3 | PCA distinguishes between exponential- and stationary-phase single *E. coli* cells through mRNA expression patterns. a, Schematic of experiment 1.10. Stationary-phase RFP-expressing *E. coli* and exponential-phase GFP-expressing *E. coli* were grown separately, and then mixed for cell preparation and PETRI-seq. **b**, Bottom, PCA of cells from experiment 1.10. RFP-expressing cells are shown in red and GFP-expressing cells in green. The grey points indicate ambiguous cells (no plasmid UMIs). TPR (see Methods) was calculated for RFP cells to the left of the threshold line (PC1 = 1.0) and GFP cells to the right of the threshold line. The TPR refers to the probability that a given cell to the left or right of the threshold is RFP-expressing or GFP-expressing, respectively. The distribution of all cells across PC1 (including ambiguous) is shown above; 7,387 cells were included (4,878 below threshold and 2,509 above). Middle, gene expression for GO terms associated with exponential-to-stationary-phase transition. The moving average (size, 1,200 cells) of the *z*-scored expression of operons within the GO term is shown. Expression was *z*-score-transformed for each gene and then for each GO term. Both GO terms are significantly correlated with PC1 before calculating moving averages (Spearman rank, $P < 10^{-70}$). Top, expression of genes that are involved in exponential-to-stationary-phase transition along PC1. The moving average (size, 2,400 cells) of the z-scored operon expression is shown. Both operons were significantly correlated with PC1 before calculating moving averages (Spearman rank, *dps*, $P = 10^{-29}$; *rpoS*, $P = 0.003$; FDR < 0.01). **c**, Schematic of experiment 2.01. Exponential-phase wild-type *E. coli* and stationary-phase wild-type *E. coli* were prepared independently and barcoded separately during round 1 (RT). Exponential-phase *S. aureus* cells (not shown) were prepared independently and combined with both cell types before RT for downstream quantification of multiplet frequency and intercellular contamination. **d**, Distributions of mRNA UMIs captured per stationary- or exponential-phase cell in experiment 2.01. **e**, Distributions of operons per stationary- or exponential-phase *E. coli* cell in experiment 2.01. Cell numbers were the same as shown in **d**. The box plots within the violin plots in **d** and **e** show the interquartile range (black box) and the median (white circle). **f**, PCA of exponential- and stationary-phase cells (experiment 2.01). The distribution of all cells across PC1 is shown above. The threshold line at PC1 = 0.28 results in a TPR of 99.97% (stationary; left) or 99.96% (exponential; right); 8,171 cells were included (4,623 below the threshold and 3,548 above the threshold).

from aggregation of similar cells to define consensus states for subpopulations within heterogenous communities. As a demonstration, we generated consensus transcriptomes by aggregating the expression counts from varying numbers of single cells at either exponential (Extended Data Fig. 9a,b) or stationary (Extended Data Fig. 9c,d) phase. As expected, correlations with independently prepared bulk libraries from cells in the same growth state increased as more cells were included. Notably, the correlations were stronger and increased at a greater rate for single-cell/bulk libraries of cells in the same state (Extended Data Fig. 9b,d), indicating that the aggregated single cells were asymptotically approaching a transcriptome reflecting their growth state.

A key advantage of scRNA-seq compared with bulk methods is the ability to characterize rare subpopulations that exhibit distinct gene expression programs. We applied PCA to 6,663 *S. aureus* single-cell transcriptomes generated using PETRI-seq and detected a small subpopulation that diverged from the rest of the population along PC1 (Extended Data Fig. 10a, red points). The eight operons most highly correlated with PC1 were lytic genes of prophage ϕSA3usa[32,33] (Extended Data Fig. 10b,c), indicating that PC1 might be capturing rare prophage induction in the *S. aureus* culture. Within the small subpopulation, 3 cells exhibited substantial upregulation of phage lytic transcripts, reaching approximately 80% of these single-cell transcriptomes (Extended Data Fig. 10d). The remaining 25 cells contained fewer than 10% phage transcripts. In further analysis of the heterogeneity in gene expression across the entire *S. aureus* population, we found that, for most operons, transcriptional noise[14] (σ^2/μ^2) inversely scaled with mean expression (μ) and followed a Poisson expectation ($\mu=\sigma^2$; Extended Data Fig. 10e), as described in previous single-cell studies[34,35]. *SAUSA300_1933-1925*—a phage lytic operon that encodes a putative phage tail and structural genes—clearly diverged and exhibited higher noise than expected from the mean (Extended Data Fig. 10e), recapitulating its hypervariability in expression that was found using PCA. Similar analysis in *E. coli* discovered candidate operons displaying high transcriptional noise (Extended Data Fig. 10f,g) that warrant independent validation by methods such as smFISH[34,36]. One of these, *fimAICDFGH*, which encodes type I fimbriae, is known to exhibit population-level phase-variable expression due to promoter inversion[37]. As such, PETRI-seq can detect rare cells that occupy distinct transcriptional states and genes that display high transcriptional heterogeneity within a population.

With a straightforward experimental pipeline that requires no advanced equipment and a per-cell cost of US$0.056, PETRI-seq is an efficient and affordable (Supplementary Table 1) method for high-throughput single-cell RNA sequencing of bacterial populations. We sequenced around 30,000 *E. coli* and *S. aureus* cells with high single-cell purity and found that aggregated transcriptomes from single cells were well correlated with bulk RNA-seq libraries. PETRI-seq assigned more than 98% of single cells within isogenic *E. coli* populations to their correct growth phases (that is, stationary or exponential phase). Moreover, the high throughput capacity of PETRI-seq was vital for detecting a rare subpopulation undergoing prophage induction in 0.04% of *S. aureus* cells. This has important clinical implications, as prophage induction is intimately linked to bacterial pathogenesis[38,39].

Optimization of mRNA capture and library preparation (see the 'Future directions for optimization' section in the Methods) will probably further improve the sensitivity of PETRI-seq and decrease its cost. Since our initial deposit of an earlier version of this manuscript at *bioRxiv*[40], and during its formal review, Kuchina et al. deposited a manuscript at *bioRxiv*[41] reporting a conceptually similar split-pool-based bacterial scRNA-seq method in which in situ polyadenylation was utilized to capture mRNAs. It will be of great interest to compare these methods and further improve the performance of PETRI-seq. We anticipate that PETRI-seq will be a highly useful tool with broad applications, such as characterization of rare, clinically important populations (such as persisters[42,43]) and high-resolution capture of native microbial communities, including unculturable components, which is a major challenge in microbiology[44].

Methods

Experimental methods. *Bacterial strains and growth conditions. E. coli* MG1655 was routinely grown in MOPS EZ rich defined medium (M2105, Teknova). The plasmid pBbE2A-RFP was a gift from J. Keasling[45] (Addgene plasmid 35322). RFP was induced with 20 nM anhydrotetracycline hydrochloride (233131000, Acros Organics). GFP was expressed from plasmid p_{rplN}-GFP[46]. Plasmid-containing MG1655 cells were grown in appropriate antibiotics (50 μg ml^{-1} kanamycin, 100 μg ml^{-1} carbenicillin). *S. aureus* USA300 (ref. [32]) was routinely grown in trypticase soy broth (TSB) medium (211825, BD). All of the bacterial strains were grown at 37 °C and shaken at 300 r.p.m.

Custom primers used in this study. All of the single-tube primers are shown in Supplementary Table 2. All of the primer sequences for 96-well split-pool barcoding are shown in Supplementary Table 3. Primers were purchased from Integrated DNA Technologies.

Preparation of annealed ligation oligos. The ligation primers used in round 2 and round 3 ligation reactions (Fig. 1) were prepared by annealing barcode oligos to specific linker oligos (SB83 and SB80). In experiment 2.01—our most optimized version of PETRI-seq—we used 4× more annealed ligation primers compared with earlier versions of the protocol. Thus, the quantities of reagents provided hereafter are for the 4× protocol (that is, experiment 2.01) and quantity of reagents for the 1× protocol are provided in parenthesis.

Round 2 barcode oligos (Supplementary Table 3) were diluted to 100 μM (20 μM). Round 3 barcode oligos were diluted to 70 μM (20 μM). Linker oligo SB83 was diluted to 100 μM (20 μM). Linker oligo SB80 was diluted to 70 μM (20 μM). To anneal round 2 barcode oligos to linker oligos, a 96-well PCR plate (AB0600, Thermo Fisher Scientific) was prepared by adding 3.52 μl (4.4 μl) of diluted SB83, 2.64 μl (0.8 μl) water and 3.84 μl (4.8 μl) of each round 2 barcode oligo to each well. To anneal round 3 barcode oligos, a 96-well PCR plate was prepared by adding 6.6 μl (4.4 μl) of diluted SB80, 7.2 μl (4.8 μl) of each round 3 barcode oligo and 0 μl (0.8 μl) water (that is, water was added only for the 1× protocol) to each well. Oligos were annealed by heating the plate to 95 °C for 3 min and then decreasing the temperature to 20 °C at a ramp speed of −0.1 °C s^{-1}.

Oligos SB84 and SB81 were also annealed (to form an intramolecular hairpin) before blocking by heating 50 μl or 80 μl, respectively, of each 400 μM (100 μM) oligo to 94 °C and slowly decreasing the temperature to 25 °C.

Cell preparation for PETRI-seq. For sequencing and qPCR measurements, cells were grown overnight then diluted into fresh medium (1:100 for *S. aureus, E. coli* MG1655, and *E. coli* MG1655 harbouring p_{rplN}-GFP, 1:50 for *E. coli* MG1655 harbouring pBbE2A-RFP) with inducer and antibiotics when applicable. For exponential-phase cells, *E. coli* and *S. aureus* cultures were grown for approximately 2 h until reaching an optical density at 600 nm (OD_{600}) of 0.4 or 0.9, respectively. Exponential-phase *E. coli* cells were used for all of the qPCR optimization experiments. For RFP-expressing stationary-phase cells, *E. coli* MG1655 cells harbouring pBbE2A–RFP were grown for an additional 3 h until the culture reached an OD_{600} of ~4 (experiment 1.06, $OD_{600}=4$; experiment 1.10, $OD_{600}=3.68$). For wild-type *E. coli* MG1655 stationary-phase cells, *E. coli* cells were diluted 1:100 and grown for ~3.75 h to an OD_{600} of ~4 (experiment 2.01, $OD_{600}=3.87$). Different cell types grown separately were then mixed as described below.

For the combined exponential-phase *E. coli* library (experiment 1.06SaEc), 3.5 ml of exponential-phase GFP *E. coli* was combined with 3.5 ml of exponential-phase RFP *E. coli*. The *S. aureus* library was prepared separately from 7 ml of exponential-phase cells. For the libraries of exponential-phase GFP *E. coli* combined with stationary-phase RFP *E. coli* (experiments 1.06 and 1.10), 3 ml of exponential-phase GFP cells was added to ~300 μl of stationary-phase RFP cells. For experiment 2.01, 7 ml of exponential-phase wild-type *E. coli* and 7 ml of stationary-phase wild-type *E. coli* were independently fixed. Before fixation, cells were pelleted at 5,525*g* (Allegra 25R centrifuge, Beckman Coulter) for 2 min at 4 °C. Spent medium was removed, and cells were resuspended in 7 ml of ice-cold 4% formaldehyde (F8775, Millipore Sigma) in PBS (P0195, Teknova). This suspension was rotated at 4 °C for 16 h on a Labquake Shaker (415110, Thermo Fisher Scientific).

The next day, fixed cells were centrifuged at 5,525*g* (Allegra 25R centrifuge, Beckman Coulter) for 10 min at 4 °C. The supernatant was removed, and the pellet was resuspended in 7 ml PBS supplemented with 0.01 U μl^{-1} SUPERase In RNase Inhibitor (AM2696, Invitrogen), hereafter referred to as PBS-RI. Cells were centrifuged again at 5,525*g* for 10 min at 4 °C and then resuspended in 700 μl PBS-RI. Subsequent centrifugations for cell preparation were all performed at 7,000*g* (5415D centrifuge, Eppendorf) for 8–10 min at 4 °C. Cells were centrifuged, then resuspended in 700 μl 50% ethanol (2716, Decon Labs) in PBS-RI. Cells were next washed twice with 700 μl PBS-RI, and then resuspended in 105 μl of 100 μg ml^{-1} lysozyme (90082, Thermo Fisher Scientific) or 40 μg ml^{-1} lysostaphin (LSPN-50, AMBI) in TEL-RI (100 mM Tris pH 8.0 (AM9856, Invitrogen), 50 mM EDTA (AM9261, Invitrogen) and 0.1 U μl^{-1} SUPERase In RNase inhibitor (10×

more than in PBS-RI)). Cells were permeabilized for 15 min at room temperature (~23 °C). After permeabilization, cells were centrifuged then washed with 175 µl PBS-RI then resuspended in 175 µl PBS-RI; 100 µl was taken for subsequent steps and centrifuged, and the remaining 75 µl was discarded. Cells were resuspended in 40 µl DNase-RI buffer (4.4 µl 10× reaction buffer, 0.2 µl SUPERase In RNase inhibitor, 35.4 µl water). DNase I (4 µl; AMPD1, Millipore Sigma) was added, and cells were incubated at room temperature for 30 min. To inactivate the DNase I, 4 µl of stop solution was added, and the cells were heated to 50 °C for 10 min with shaking at 500 r.p.m. (Multi-Therm, Benchmark Scientific); 50 °C, rather than 70 °C, was used to avoid cell lysis. After DNase inactivation, cells were pelleted, washed twice with 100 µl PBS-RI and then resuspended in 100 µl 0.5× PBS-RI. Cells were counted using a haemocytometer (DHC-S02, INCYTO).

Split-pool barcoding for PETRI-seq. For RT, round 1 primers (Supplementary Table 3) were diluted to 10 µM, then 2 µl of each primer was aliquoted across a 96-well PCR plate. A reaction mix was prepared for RT with 240 µl 5× 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). Cells (3×10^7) were added to this mix. For species-mixed libraries, *E. coli* and *S. aureus* cells were combined at this point. Water was added to bring the volume of the reaction mixture to 960 µl. The reaction mixture (8 µl) was added to each well of the 96-well plate already containing RT primers, making the final volume in each well 10 µl. The plate was sealed and 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 then held at 4 °C.

After RT, the 96 reactions were pooled into one tube. At this point, detergent was added to the pooled cells for experiment 2.01, our most optimized version of PETRI-seq. Specifically, 5% Tween-20 was diluted 125× to a final concentration of 0.04%. We measured the volume of the pooled cells to determine this exact volume. Cells were incubated on ice for 3 min, then PBS-RI was added to bring the final concentration of Tween-20 to 0.01% (for example, we added 2,508 µl to the 836 µl sample, splitting the samples into multiple Eppendorf tubes for centrifugation). Cells were then centrifuged at 10,000*g* for 20 min at 4 °C. The supernatant was removed. For experiments without detergent, cells were centrifuged (10,000*g* for 20 min at 4 °C) immediately after pooling. Without detergent, a cell pellet was not visible after this centrifugation, but with detergent a cell pellet was visible.

For the round 2 ligation, cells were then resuspended in 600 µl 1× T4 ligase buffer (M0202L, NEB) supplemented with 0.1 U µl^{-1} SUPERase In RNase inhibitor. The following additional reagents were added to make a master mix: 7.5 µl water, 37.5 µl 10× T4 ligase buffer, 16.7 µl SUPERase In RNase Inhibitor, 5.6 µl BSA (B14, Thermo Fisher Scientific) and 27.9 µl T4 ligase, making the final volume 695.2 µl; 5.76 µl of this mix was added to each well of a 96-well plate containing 2.24 µl of annealed round 2 ligation oligos (see the 'Preparation of annealed ligation oligos' section) for a final volume of 8 µl. Ligation reactions were performed for 30 min at 37 °C. After this incubation, 2 µl of blocking mix (37.5 µl of 400 µM SB84 (100 µM for the 1× protocol), 37.5 µl of 400 µM SB85 (100 µM for the 1× protocol), 25 µl 10× T4 ligase buffer and 150 µl water) was added to each well, and reactions were incubated for an additional 30 min at 37 °C. Cells were then pooled into a single tube.

The following reagents were added to the pooled cells for round 3 barcoding for experiment 2.01 (the most optimized protocol): 46 µl 10× T4 ligase buffer, 12.65 µl T4 ligase and 115 µl water; 8.51 µl of this mixture was added to each well of a 96-well plate containing 3.49 µl annealed round 3 ligation oligos (see the 'Preparation of annealed ligation oligos' section).

Alternatively, for the 1× protocol, the following reagents were added to the pooled cells: 15.6 µl water, 48 µl 10× T4 ligase buffer and 13.2 µl T4 ligase; 8.64 µl of this mixture was added to each well of a 96-well plate containing 3.36 µl of annealed round 3 ligation primers (see the 'Preparation of annealed ligation oligos' section).

The plate was incubated for 30 min at 37 °C. After ligation, 10 µl of round 3 blocking mix (72 µl of 400 µM SB81 (100 µM for the 1× protocol), 72 µl of 400 µM SB82 (100 µM for the 1× protocol), 120 µl 10× T4 ligase buffer, 336 µl water and 600 µl 0.5 M EDTA) was added to each well. Cells were then pooled into a single tube. When detergent was used (in the most optimized protocol, that is, experiment 2.01), Tween-20 was added to a final concentration of 0.01%. With or without detergent, cells were then centrifuged at 7,000*g* for 10 min at 4 °C. When including detergent, cells were resuspended in 500 µl TEL-RI + 0.01% Tween-20. Without detergent, cells were resuspended in 50 µl TEL-RI (because cell retention is very poor in large volumes without detergent). At this stage, additional washing may be advantageous to reduce any contamination from ambient cDNA (Extended Data Figs. 3 and 7), although we have yet to test this. This suspension was centrifuged at 7,000*g* for an additional 10 min at 4 °C, the supernatant was removed and the cells were resuspended in 30 µl TEL-RI. Cells were counted using a haemocytometer. Aliquots of ~10,000 cells were taken and diluted in 50 µl lysis buffer (50 mM Tris pH 8.0, 50 mM EDTA, 200 mM NaCl (AM9759, Invitrogen) and 0.5% Triton X-100). Proteinase K (5 µl of 20 mg ml^{-1}; AM2548, Invitrogen) was added to the cells in lysis buffer. Cells were lysed for 1 h at 55 °C with shaking at 750 r.p.m. (Multi-Therm). Lysates were stored at −80 °C.

Library preparation for PETRI-seq. Library preparation steps after cell lysis and before PCR amplification should be performed with care. This is because, without amplification, every barcoded cDNA molecule that originates from a single cell is non-recoverable if lost. In other words, any loss of cDNA results in a reduction in total UMI capture per cell.

Lysates were purified with AMPure XP beads (A63881, Beckman Coulter) at a ratio of 1.8× (~99 µl beads). cDNA was eluted in 20 µl of water. Water (14 µl), 4 µl NEBNext Second Strand Synthesis Reaction Buffer and 2 µl NEBNext Second Strand Synthesis Enzyme Mix (E6111S, NEB) were added to the purified cDNA. This reaction (40 µl) was incubated at 16 °C for 2.5 h. The resulting double-stranded cDNA was purified with AMPure XP beads at a ratio of 1.8× (~72 µl beads). Double-stranded cDNA was eluted in 20 µl water and used immediately for tagmentation or stored at −20 °C.

Double-stranded cDNA was tagmented and amplified using the Nextera XT DNA Library Preparation Kit (FC-131–1096, Illumina). We followed the manufacturer's protocol with the following modified reagent volumes and primers: 25 µl TD, 20 µl cDNA, 5 µl ATM, 12.5 µl NT, 2.5 µl N70x (Nextera Index Kit v2 Set A, TG-131-2001, Illumina), 2.5 µl i50x (E7600S, NEB), 20 µl water and 37.5 µl NPM. Libraries were amplified for 8 cycles according to the manufacturer's protocol. After 8 cycles, 5 µl was removed, added to a qPCR mix (0.275 µl EvaGreen (31000, Biotium), 0.11 µl ROX Low Reference Dye (KK4602, Kapa Biosystems) and 0.115 µl water) and further cycled on a qPCR machine. qPCR amplification was used to determine the exponential phase of amplification, which occurred after 11 cycles for experiments 1.06SaEc and 1.10 and after 8 cycles for experiment 2.01. The remaining PCR product (which was not removed for qPCR) was thermocycled for an additional 11 or 8 cycles, resulting in a total of 19 or 16 PCR cycles. Products were purified with AMPure XP beads at a ratio of 1× and eluted in 30 µl water. The concentration of the library was measured using the Qubit dsDNA HS Assay Kit (Q32854, Invitrogen) and the Agilent Bioanalyzer High Sensitivity DNA Kit (5067–4626, Agilent). Although we used a 1× ratio of AMPure beads for the libraries presented here, we note that, after sequencing, a substantial fraction of molecules was too short to be assigned to a BC and/or aligned to the genome (Supplementary Table 4). A lower ratio of AMPure beads or an additional round of purification might be helpful to reduce the abundance of these wasted reads.

Libraries were sequenced for 75 cycles using the NextSeq 500/550 High Output Kit v2.5 (20024906, Illumina). Cycles were allocated as follows: 58 cycles read 1 (UMI and barcodes), 17 cycles read 2 (cDNA), 8 cycles index 1 and 8 cycles index 2.

Modifications tested to optimize PETRI-seq. To test fixing cells immediately from cultures without centrifugation, ice-cold 5% formaldehyde in PBS was added directly to cells in spent medium to bring the final concentration of formaldehyde up to 4%. Cell preparation with no lysozyme or no DNase was performed by simply omitting the enzyme and using water to replace that volume.

Template switching was performed by adding 2.5 µl 100 µM SB14, 20 µl Maxima H Minus 5× buffer, 10 µl dNTPs, 2.5 µl SUPERase In RNase Inhibitor, 2 µl Maxima H Minus Reverse Transcriptase, 3 µl water and 20 µl betaine (J77507VCR, Thermo Fisher Scientific) to 40 µl of AMPure purified lysate. SB14 was heated to 72 °C for 5 min before combining the above reagents. The reaction was incubated at 42 °C for 90 min and then heat-inactivated at 85 °C for 5 min. The reaction was purified with AMPure XP beads at a 1.8× ratio and eluted in 30 µl. The purified cDNA was then amplified by setting up the following PCR reaction: 10 µl 5× PrimeSTAR GXL Buffer, 0.1 µl of 10 µM SB86, 0.1 µl of 10 µM SB15, 1 µl PrimeSTAR GXL Polymerase (R050B, Takara Bio), 1 µl dNTPs and 8 µl water. The reaction was heated to 98 °C for 1 min and then thermocycled 10 times (98 °C for 10 s, 60 °C for 15 s and 68 °C for 6 min). The products were purified using AMPure XP beads at a 1.8× ratio and eluted in 30 µl. The DNA concentration was measured using the Qubit dsDNA HS Assay Kit, and tagmentation was performed according to the manufacturer's protocol using the appropriate primers (described above for standard PETRI-seq).

For the library 1.06SaEc-replicate (Supplementary Table 4), we included an RT clean-up step as part of library preparation. RT clean-up was performed in the same manner as template switching, but SB14 (TSO) was not added. After incubating the reaction at 42 °C for 90 min and then heat inactivating at 85 °C for 5 min, reaction components were added for second-strand synthesis (70 µl water, 20 µl NEB second strand buffer and 10 µl NEB second strand enzyme). Second strand synthesis was performed as described and the double-stranded cDNA was used as an input for tagmentation. Although RT clean-up resulted in a broader size distribution on the bioanalyzer after tagmentation (not shown), it did not change the yield of PETRI-seq and was therefore not used for other libraries.

For experiment 1.08 (Supplementary Table 4), we included a longer RT reaction (~2 h) using the following thermocycling protocol: 50 °C for 10 min; then 10 cycles of 8 °C for 12 s, 15 °C for 45 s, 20 °C for 45 s, 25 °C for 5 min, 42 °C for 1 min and 50 °C for 2 min.

qPCR quantification after in situ DNase or in situ RT. For qPCR quantification after in situ RT, cells were counted before RT, and then the in situ RT reaction described above (scaled to one 50 µl reaction) was set up with equal cell numbers for each condition and technical replicate. A random hexamer (SB94) or a gene-specific primer (SB10) was used as an RT primer. After RT, cells were centrifuged at 7,000*g* for 10 min and then washed in 50 µl PBS-RI. After one wash, cells were resuspended in 50 µl lysis buffer, and 5 µl of proteinase K was added. Cells were lysed for 1 h at 55 °C with shaking at 750 r.p.m. For qPCR quantification after

in situ DNase treatment, cells were washed twice after DNase treatment, as described for PETRI-seq cell preparation and then lysed.

Unpurified lysates were diluted 50× (except for ethanol versus no ethanol experiments, in which lysates were diluted 10×) in water and heated to 95 °C for 10 min to inactivate proteinase K. Diluted lysates were then used directly in qPCR with either Kapa 2× MasterMix Universal (KK4602, Kapa Biosystems) or Power SYBR Green Master Mix (4368706, Applied Biosystems). For quantification of genomic DNA after DNase treatment or quantification of cDNA after RT with random hexamers, qPCR primers SB5 and SB6 were used, and relative abundances were calculated on the basis of an experimentally determined amplification efficiency of 88%, which corresponded to an amplification factor of 1.88. Thus, relative abundance referred to $1.88^{-\Delta C_t}$, where ΔC_t was the difference between the C_t value of each sample and a calibrator C_t. For RT with the gene-specific primer, qPCR primers SB12 and SB13 were used, as SB12 anneals to the gene-specific primer (SB10). The experimentally determined amplification factor for these primers was 1.73. To quantify cDNA yield, the abundance of a matched sample with no RT (processed equivalently but RT enzyme omitted) was subtracted from each measurement. All replicates were technical replicates, which were treated independently during and after the condition tested.

qPCR quantification of ligation efficiency. To test barcode ligation with a 16-base linker relative to a 30-base linker, approximately 1 µg of purified RNA (bulk) was used for RT with either SB110 or SB114 (used as a positive control). RT was performed as described for in situ RT, scaled to 50 µl. cDNA was then purified with AMPure XP beads. SB113, the primer to be ligated, was annealed either to SB111 (30 bases) or SB83 (16 bases). The annealed primers (2.24 µl) were then used in a 10 µl ligation reaction. The products were purified with AMPure XP beads. To quantify the proportion of ligated product, qPCR was performed with SB86 and SB13, which amplifies only the ligated product, as SB86 anneals to the ligated overhang, or SB115 and SB13, which amplifies all RT product, as SB115 anneals to the RT primer overhang. $\Delta\Delta C_t$ was calculated for the two primer sets with RT product from SB114 as a reference ($\Delta\Delta C_t = \Delta C_t$(experimental, ligated) − ΔC_t(control, SB114 RT); $\Delta C_t = C_t$(SB86,SB13) − C_t(SB115,SB13)). SB114 includes primer sites for both SB86 and SB115, so it mimics ligation with 100% efficiency.

Test of DNase inactivation by incubating cells with exogenous DNA. After DNase treatment, inactivation and two PBS-RI washes (described above), cells were resuspended in 20 µl PBS-RI; 6 µl was then removed and added to 1 µl DNase reaction buffer, 1 µl water and 2 µl of a 775 bp PCR product (800 ng). As a control, 1 µl DNase I was added instead of 1 µl water. The reactions were incubated for 1 h, after which 1 µl of stop solution was added. The cells were centrifuged for 10 min at 7,000*g*. The supernatants were then heated to 70 °C for 10 min to inactivate DNase; 5 µl of each reaction was run on a gel.

Bulk library preparation. To prepare bulk samples from fixed cells (Fig. 2d,e, Extended Data Fig. 9), 25 µl (~10^7 cells) was taken after PETRI-seq cell preparation and just before in situ RT. These cells were centrifuged and resuspended in 50 µl lysis buffer supplemented with 5 µl proteinase K. Cells were lysed at 55 °C for 1 h with shaking at 750 r.p.m. (Multi-Therm). RNA was then purified from lysates using the Norgen Total RNA Purification Plus Kit (48300, Norgen Biotek). Buffer RL (300 µl) was added to the lysate before proceeding to the total RNA purification protocol. Alternatively, the standard bulk RNA sample (Extended Data Fig. 2b,c) was prepared by centrifuging a cell culture at 5,525*g* for 2 min at 4 °C and then resuspending cells in 1 ml of PBS-RNAprotect (333 µl RNAprotect Bacteria Reagent (76506, Qiagen), 666 µl PBS). For immediate RNA stabilization by RNAprotect (Extended Data Fig. 2b), 2 ml of RNAprotect was immediately added to 1 ml of exponential-phase *E. coli* cells. For immediate RNA stabilization by flash-freezing (Extended Data Fig. 2b), 330 µl 60% glycerol was added to 1 ml exponential-phase *E. coli* cells, and cells were flash-frozen in ethanol and dry ice (<1 min). Frozen cells were retained at −80 °C overnight, and then thawed, centrifuged and resuspended in PBS-RI. For all three protocols, after resuspending cells in RNAprotect or PBS-RI, cells were then pelleted again, and RNA was prepared using the Norgen Total RNA Purification Plus Kit according to the manufacturer's instructions for Gram-negative bacteria.

Purified RNA from either protocol was treated with DNase I in a 50 µl reaction consisting of 2–5 µg RNA, 5 µl DNase Reaction Buffer, 5 µl DNase and water. Reactions were incubated at room temperature for 30–40 min. Reactions were purified by adding 300 µl buffer RL and proceeding according to the Norgen total RNA purification protocol. Total RNA was depleted of rRNA using the Gram-Negative Ribo-Zero rRNA Removal Kit (MRZGN126, Illumina), purified by ethanol precipitation and resuspended in 10 µl water. For RT, 6 µl RNA was combined with 4 µl Maxima H Minus 5× Buffer, 2 µl dNTPs, 0.5 µl SUPERase In RNase Inhibitor, 1 µl SB94, 0.5 µl Maxima H Minus Reverse Transcriptase, 4 µl betaine and 2 µl water. The reaction was thermocycled 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, 85 °C for 5 min and then held at 4 °C. For second-strand synthesis, 14 µl water, 4 µl NEBNext Second Strand Synthesis Reaction Buffer and 2 µl NEBNext Second Strand Synthesis Enzyme Mix were added directly to the RT mix. This reaction was incubated at 16 °C for 2.5 h. Double-stranded cDNA was purified with AMPure XP beads at a 1.8× ratio (~72 µl beads) and eluted in 30 µl water. Purified cDNA was used for tagmentation using the Nextera XT kit according to the manufacturer's protocol. Bulk libraries were purified twice with AMPure XP beads at a 0.9× ratio. The resulting libraries were quantified and sequenced as described for PETRI-seq libraries above.

Growth curves. Overnight cultures were grown as described above and then diluted 1:100 into 1 ml EZ Rich Defined Media with or without 20 nM aTc. Antibiotics were added for plasmid-containing strains. For each condition, 100 µl of diluted cells were aliquoted into 4 wells of a 96-well plate. The plate was incubated at 37 °C with shaking on the plate reader (Synergy Mx, Biotek). OD_{600}, GFP and RFP were measured every 10 min.

Computational methods. *Barcode demultiplexing, cell selection and alignment.* Cutadapt[47] was used to trim low-quality read 1 and read 2 sequences with a phred score of less than ten. Umi_tools[48] was used in paired-end mode to extract the seven-base UMI sequence from the beginning of read 1. Read pairs were then grouped on the basis of their three barcode sequences using the cutadapt demultiplex feature. FASTQ files were first demultiplexed by barcode 3, requiring that matching sequences were anchored at the beginning of the read, overlapped at 21 positions ('--overlap 21', including downstream linker (GGTCCTTGGCTTCGC)), and had no more than 1 mismatch relative to the barcode assignment (-e 0.05). As part of demultiplexing, the barcode and linker sequence were trimmed in read 1. For barcode 2, cutadapt was used to locate barcode sequences with the expected downstream linker, allowing for no more than 1 mismatch (-e 0.05 --overlap 20) and requiring the barcode at the beginning of the read. The barcode and linker sequences were trimmed. Next, reads were demultiplexed by barcode 1, requiring the barcode at the beginning of the read and allowing 1 mismatch but no indels. The final output after demultiplexing was a set of read 1 and read 2 FASTQ files where each file corresponded to a three-barcode combination (BC). The knee method[5] was used to identify BCs for further processing. In brief, each BC was sorted by descending total number of reads, and then the cumulative fraction of reads for each BC was plotted. As the yield per BC could be better assessed later after collapsing reads to UMIs, an inclusive threshold was used at this stage to select BCs for downstream processing, which enabled more-precise cell selection after downstream processing (Extended Data Fig. 1f). Cutadapt was then used to trim and discard read 2 sequences containing barcode 1 or the linker sequence. Note that, at this point, all necessary information was contained in the read 2 FASTQ files, so further processing did not consider the read 1 files. Next, cDNA sequences were aligned to reference genomes using the backtrack algorithm in the Burrows–Wheeler Alignment tool, bwa[49], allowing for a maximum edit distance of 1 for assigned alignments.

Annotating features and grouping PCR duplicates by shared UMI. FeatureCounts[50] was used to annotate operons on the basis of the alignment position. Operon sequences were obtained from RegulonDB[51] and ProOpDB[52] for *E. coli* and *S. aureus*, respectively. As featureCounts uses an XT SAM file tag for annotation, the bwa XT tag was first removed from all SAM files using a python script. The resulting BAM files after featureCounts were used as input for the group function of umi_tools with the '--per-gene' option in directional mode[48]. The directional algorithm is a network-based method that identifies clusters of connected UMI sequences to group as single UMIs. The result was a set of BAM files with UMI sequences corrected on the basis of probable errors from sequencing or amplification. A python script was used to collapse reads to UMIs. Reads with the same BC, error-corrected UMI and operon assignment were grouped into a single count. With 4^7 possible UMIs, we confirmed that the expected rate of UMI collisions (different molecules with the same UMI) was low by implementing a correction on the basis of the Poisson expectation of collisions[53]. As this correction had a negligible effect, we did not include it for other analyses. Reads that mapped to multiple optimal positions were omitted, except for rRNA alignments, for which multiple alignments were expected. The distribution of number of reads per UMI for all UMI–BC–operon combinations was plotted to establish a threshold below which UMIs were excluded (Extended Data Fig. 1g). Filtered UMIs were used to generate an operon by BC count matrix. Anti-sense transcripts were removed. BCs with fewer than a threshold of total UMIs were then removed (Extended Data Figs. 1j,i and 6h,i). GNU Parallel[54] was used to execute many of the above processes more efficiently.

Bulk sequencing libraries. For bulk sequencing libraries, only read 2 was used for alignment to mimic single-cell methods. Bulk sequencing libraries were preprocessed to remove adaptors using cutadapt[47]. Trimmomatic[55] was then used to remove leading or trailing bases below quality phred 33 quality 3 and discard reads shorter than 14 bases. Surviving reads were aligned using the backtrack algorithm in bwa[49] with a maximum edit distance of 1. Reads with more than one optimal alignment position were removed. FeatureCounts[50] was used to generate a matrix of operon counts for the bulk libraries. To compare single-cell libraries generated by PETRI-seq to bulk samples, the UMI counts for a given set of BCs (for example GFP-expressing *E. coli*) were summed for all operons. A count matrix was then

generated as described for bulk libraries. To calculate TPM, raw counts were divided by the length of the operon in kb. Each length-adjusted count was then divided by the sum of all adjusted counts divided by 1 million.

Calculating multiplet frequency. The multiplet frequency was defined as the fraction of non-empty BCs corresponding to more than one cell. To calculate the predicted multiplet frequency, the proportion of predicted BCs with 0 cells was calculated on the basis of a Poisson process: $P(0) = \frac{\lambda^0}{0!}e^{-\lambda}$, the proportion of BCs with 1 cell was calculated: $P(1) = \frac{\lambda^1}{1!}e^{-\lambda}$, the proportion with greater than 0 cells was calculated: $P(\geq 1) = 1 - P(0)$ and the proportion with greater than 1 cell was calculated: $P(\geq 2) = 1 - P(1) - P(0)$. Finally, the multiplet frequency was calculated: $\frac{P(\geq 2)}{P(\geq 1)}$. λ was the fraction of cells relative to total possible BCs—for example, $\frac{10{,}000\ \text{cells}}{96 \times 96 \times 96\ \text{barcodes}} = 0.011 = \lambda$. The experimental multiplet frequency was computed from the species-mixing experiment as described for populations with unequal representation of two species[20].

PCA analysis. rRNA and all plasmid genes (*RFP*, *GFP*, *AmpR*, *KanR* and *tetR*) were first removed from the count matrix. Operons with 5 or fewer total counts in the library were also removed. Cells with fewer than 15 mRNAs were removed (Extended Data Fig. 1j,k). Total operon counts for each cell were normalized by dividing each count by the total number of counts for that cell and then multiplying the resulting value by the geometric mean[13] of the total mRNA counts for each cell. The scaled values were then log-transformed after adding a pseudocount to each. For each operon, expression values were scaled to *z*-scores[56]. Principal components were computed using scikit-learn in python.

To normalize counts using sctransform in Seurat[26], first rRNA and all plasmid genes were removed from the count matrix. Operons with 10 or fewer total counts, and cells with fewer than 15 mRNAs were also removed. A Seurat object was created in R from the resulting matrix, and sctransform was applied. The resulting scaled counts were used as input for PCA.

TPR was calculated as follows, using red cells to the left of a threshold line as an example: $\frac{\frac{n_{rl}}{n_r}}{\frac{n_{rl}}{n_r} + \frac{n_{gl}}{n_g}}$, where n_{rl} is the number of red cells left of threshold, n_r is the total number of red cells, n_{gl} is the number of green cells left of threshold and n_g is total number of green cells.

Computing moving averages of gene expression along PC1. Using a custom Python script, the cells in the normalized, log-transformed *z*-scored gene matrix were sorted by PC1. The rolling function in the pandas package was then used to compute rolling averages of the size indicated for each figure. Win_type was set to 'None'. The corresponding PC1 coordinate was the moving average of the PC1 values. Moving averages for GO terms were computed as described, except that the *z*-scored sum of *z*-scored counts for all operons in the GO term was used to calculate the moving average instead of expression from a single operon. In cases in which multiple genes from the same operon were included in a GO term, only one gene was included. The significance of expression trends was determined by the Spearman rank correlation between the operon or GO term expression and PC1, before calculating a moving average. FDR was determined using the Benjamini–Hochberg procedure[57].

Computing operon noise. Noise was defined as σ^2/μ^2, where σ is s.d. and μ is mean. Noise and mean were calculated for all operons with at least 5 raw counts (UMIs) in the dataset (either *S. aureus* or *E. coli*). Count matrices were normalized by cell and multiplied by the geometric mean of total UMIs per cell in the library (but not log-transformed) before computing noise and mean. Operons with mean expression of <0.002 after normalization were excluded. To calculate a *P* value for the divergence of *SAUSA300_1933-1925* (Extended Data Fig. 10e) or candidate hypervariable *E. coli* operons (Extended Data Fig. 10f), a line was fit to the log-scaled noise versus log-scaled mean of the data. The residuals of the experimental data to the best-fit line were calculated and *z*-scored. The *P* value was determined on the basis of a normal distribution of the *z*-scored residuals. For the *E. coli* dataset, cells with BC2 22, 49 or 69 were removed because, in rare cases, these barcodes misaligned to an operon, resulting in the appearance of hypervariability in gene expression.

Future directions for optimization. We anticipate that the following modifications would further improve the final mRNA capture of PETRI-seq. During the library-preparation step of PETRI-seq, subjecting double-stranded cDNA to conventional tagmentation with both N5 and N7 adaptors (Illumina Nextera XT) incurs a twofold decrease in mRNA capture. This is because only one of the adaptors (N7 in our case) could be subsequently amplified, leading to the loss of all molecules tagmented by N5. Thus, modified tagmentation using a single adaptor (N7 only), as demonstrated previously[13], could prevent this twofold loss.

Second, capture may be improved by further increasing primer and enzyme concentrations during the ligation steps and/or using a hairpin ligation[13] instead of an intermolecular linker. For example, increasing the concentration of round 3 ligation oligos by 4-fold alone increased mRNA capture by 2.7-fold in both exponential- and stationary-phase *E. coli* cells (Extended Data Fig. 6a). Our preliminary results also indicate that adding polyethylene glycol to the third round of ligation increases capture by 30% (not shown).

Given that rRNAs comprise >95% of total RNA species in many bacteria, we reason that rRNA depletion could substantially improve mRNA capture and sequencing efficiency. We propose four such strategies here. First, rRNA degradation through hybridization has been demonstrated for bulk RNA sample preparations[58], in which rRNAs are hybridized with a comprehensive set of short complementary DNA oligos, followed by RNase H treatment. Second, mRNA capture might be improved by designing RT primers with sequences biased against rRNA[59], thereby directing reagents preferentially towards mRNA. Third, in situ 5′-phosphate-dependent exonuclease treatment could be used to preferentially degrade processed RNAs, the majority of these being rRNAs[60], before RT. Whereas these three strategies aim to deplete rRNAs in situ, the fourth strategy is applied during library preparation. Specifically, abundance-based normalization by melting and rehybridizing the double-stranded cDNA library followed by duplex-specific nuclease treatment[61] can be used to deplete double-stranded DNAs that encode rRNAs. In developing these rRNA-depletion strategies, it will be important to ensure that the depletion is specific by comparing the depleted and non-depleted transcriptomes.

In addition to optimizing the mRNA capture rate, further reduction in cost and time will improve the PETRI-seq workflow. We have preliminary results indicating that DNase treatment may not be necessary (not shown). However, we have not yet determined whether omitting the DNase buffer incubation or heat inactivation would alter cell permeability. Without DNase treatment, cell preparation time would be reduced by ~1.5 h.

Finally, we have shown that, in experiment 2.01, ~1–5% of UMIs within a single-cell transcriptome are probably derived from other cells (Extended Data Fig. 7d). This cross-contamination, which may be the result of ambient cDNA released from cells during or after barcoding, might be reduced by more thorough cell washing before lysis. Cross-contamination may also be reduced by preparing lysates with fewer cells, thereby reducing the likelihood of barcode collisions with ambient cDNA (or other cells). PCR may also be a source of cross-contamination through chimaera formation or priming by residual barcodes. This type of contamination may be reduced by thorough washing before lysis (to remove free barcodes) or by optimizing the parameters of the PCR. Computationally, we also showed that a more stringent alignment reduces the level of apparent cross-contamination resulting from incorrect alignment (Extended Data Fig. 3e,f), but more stringent alignment results in a decrease in captured UMIs per cell (Extended Data Fig. 3c,d,g,h). Future studies could use longer reads (that is, 150-cycle Illumina Nextseq) to eliminate ambiguities in alignment without sacrificing capture rate.

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

Data availability

Raw data have been submitted to the Gene Expression Omnibus under accession number GSE141018. Source data are also provided for all figures. All of the figures except for Fig. 1 include original data. An overview of all of the experiments is provided in Supplementary Table 4. A count matrix for the three primary PETRI-seq experiments is provided in Supplementary Table 6.

Code availability

Relevant code for this manuscript is available from the corresponding author on request; current PETRI-seq code and protocols are available at https://tavazoielab.c2b2.columbia.edu/PETRI-seq/.

Received: 26 November 2019; Accepted: 23 April 2020;
Published online: 25 May 2020

References

1. Tang, F. et al. mRNA-seq whole-transcriptome analysis of a single cell. *Nat. Methods* **6**, 377–382 (2009).
2. Ramsköld, D. et al. Full-length mRNA-seq from single-cell levels of RNA and individual circulating tumor cells. *Nat. Biotechnol.* **30**, 777–782 (2012).
3. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
4. Fan, H. C., Fu, G. K. & Fodor, S. P. A. Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* **347**, 1258367 (2015).
5. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
6. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
7. Bose, S. et al. Scalable microfluidics for single-cell RNA printing and sequencing. *Genome Biol.* **16**, 120 (2015).
8. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).

9. Picelli, S. Single-cell RNA-sequencing: the future of genome biology is now. *RNA Biol.* **14**, 637–650 (2016).
10. Sheng, K., Cao, W., Niu, Y., Deng, Q. & Zong, C. Effective detection of variation in single-cell transcriptomes using MATQ-seq. *Nat. Methods* **14**, 267–270 (2017).
11. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661–667 (2017).
12. Rosenberg, A. B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science* **360**, 176–182 (2018).
13. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
14. Taniguchi, Y. et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
15. Bartholomäus, A. et al. Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos. Trans. R. Soc. A* **374**, 20150069 (2016).
16. Moran, M. A. et al. Sizing up metatranscriptomics. *Isme J.* **7**, 237–243 (2013).
17. de Lange, N., Tran, T. M. & Abate, A. R. Electrical lysis of cells for detergent-free droplet assays. *Biomicrofluidics* **10**, 024114 (2016).
18. Amini, S. et al. Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing. *Nat. Genet.* **46**, 1343–1349 (2014).
19. Hodson, R. E., Dustman, W. A., Garg, R. P. & Moran, M. A. In situ PCR for visualization of microscale distribution of specific genes and gene products in prokaryotic communities. *Appl. Environ. Microbiol.* **61**, 4074–4082 (1995).
20. Bloom, J. D. Estimating the frequency of multiplets in single-cell RNA sequencing from cell-mixing experiments. *PeerJ* **6**, e5578 (2018).
21. Okayama, H. & Berg, P. High-efficiency cloning of full-length cDNA. *Mol. Cell. Biol.* **2**, 161–170 (1982).
22. Kivioja, T. et al. Counting absolute number of molecules using unique molecular identifiers. *Nat. Methods* **9**, 72–74 (2012).
23. Yang, S. et al. Decontamination of ambient RNA in single-cell RNA-seq with DecontX. *Genome Biol.* **21**, 57 (2020).
24. Young, M. D. & Behjati, S. SoupX removes ambient RNA contamination from droplet based single-cell RNA sequencing data. Preprint at *bioRxiv* https://doi.org/10.1101/303727 (2020).
25. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417–441 (1933).
26. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
27. Gentry, D. R., Hernandez, V. J., Nguyen, L. H., Jensen, D. B. & Cashel, M. Synthesis of the stationary-phase sigma factor σ^s is positively regulated by ppGpp. *J. Bacteriol.* **175**, 7982–7989 (1993).
28. Almirón, M., Link, A. J., Furlong, D. & Kolter, R. A novel DNA-binding protein with regulatory and protective roles in starved *Escherichia coli*. *Genes Dev.* **6**, 2646–2654 (1992).
29. Traxler, M. F. et al. The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli*. *Mol. Microbiol.* **68**, 1128–1148 (2008).
30. Chen, H., Shiroguchi, K., Ge, H. & Xie, X. S. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.* **11**, 781 (2015).
31. Vargas-Garcia, C. A., Ghusinga, K. J. & Singh, A. Cell size control and gene expression homeostasis in single-cells. *Curr. Opin. Syst. Biol.* **8**, 109–116 (2018).
32. Diep, B. A. et al. Complete genome sequence of USA300, an epidemic clone of community-acquired meticillin-resistant *Staphylococcus aureus*. *Lancet* **367**, 731–739 (2006).
33. Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. & Wheeler, D. L. GenBank. *Nucleic Acids Res.* **35**, D21–D25 (2007).
34. Saint, M. et al. Single-cell imaging and RNA sequencing reveal patterns of gene expression heterogeneity during fission yeast growth and adaptation. *Nat. Microbiol.* **4**, 480–491 (2019).
35. Grün, L., Kester, L. & Oudenaarden, A. Validation of noise models for single-cell transcriptomics. *Nat. Methods* **11**, 637–640 (2014).
36. Raj, A., van den Bogaard, P., Rifkin, S. A., van den Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).
37. Abraham, J. M., Freitag, C. S., Clements, J. R. & Eisenstein, B. I. An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **82**, 5724–5727 (1985).
38. Deutsch, D. R. et al. Extra-chromosomal DNA sequencing reveals episomal prophages capable of impacting virulence factor expression in *Staphylococcus aureus*. *Front. Microbiol.* **9**, 1406 (2018).
39. Balasubramanian, S., Osburne, M. S., BrinJones, H., Tai, A. K. & Leong, J. M. Prophage induction, but not production of phage particles, is required for lethal disease in a microbiome-replete murine model of enterohemorrhagic *E. coli* infection. *Plos Pathog.* **15**, e1007494 (2019).
40. Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. Preprint at *bioRxiv* https://doi.org/10.1101/866244 (2019).
41. Kuchina, A. et al. Microbial single-cell RNA sequencing by split-pool barcoding. Preprint at *bioRxiv* https://doi.org/10.1101/869248 (2019).
42. Brauner, A., Fridman, O., Gefen, O. & Balaban, N. Q. Distinguishing between resistance, tolerance and persistence to antibiotic treatment. *Nat. Rev. Microbiol.* **14**, 320–330 (2016).
43. Girgis, H. S., Harris, K. & Tavazoie, S. Large mutational target size for rapid emergence of bacterial persistence. *Proc. Natl Acad. Sci. USA* **109**, 12740–12745 (2012).
44. Franzosa, E. A. et al. Sequencing and beyond: integrating molecular 'omics' for microbial community profiling. *Nat. Rev. Microbiol.* **13**, 360–372 (2015).
45. Lee, T. S. et al. BglBrick vectors and datasheets: a synthetic biology platform for gene expression. *J. Biol. Eng.* **5**, 12 (2011).
46. Zaslaver, A. et al. A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat. Methods* **3**, 623–628 (2006).
47. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBNet* **17**, 10–12 (2011).
48. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modelling sequencing errors in unique molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
50. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
51. Santos-Zavaleta, A. et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* **47**, D212–D220 (2019).
52. Taboada, B., Ciria, R., Martinez-Guerrero, C. E. & Merino, E. ProOpDB: Prokaryotic Operon DataBase. *Nucleic Acids Res.* **40**, D627–D631 (2012).
53. Fu, G. K., Hu, J., Wang, P. & Fodor, S. P. A. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc. Natl Acad. Sci. USA* **108**, 9026–9031 (2011).
54. Tange, O. *GNU Parallel 2018* (Ole Tange, 2018).
55. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
56. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
57. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
58. Huang, Y., Sheth, R. U., Kaufman, A. & Wang, H. H. Scalable and cost-effective ribonuclease-based rRNA depletion for transcriptomics. *Nucleic Acids Res.* **48**, e20 (2020).
59. Armour, C. D. et al. Digital transcriptome profiling using selective hexamer priming for cDNA synthesis. *Nat. Methods* **6**, 647–649 (2009).
60. He, S. et al. Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nat. Methods* **7**, 807–812 (2010).
61. Zhulidov, P. A. et al. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Res.* **32**, e37 (2004).

Acknowledgements

We thank the members of the Tavazoie laboratory for discussions and comments on early drafts of the manuscript; and P. Sims for suggestions during the early development of PETRI-seq. S.T. is supported by award no. 5R01AI077562 from the National Institutes of Health. S.B.B. is supported by a National Science Foundation Graduate Research Fellowship (no. DGE 16-44869). W.J. is supported by a fellowship from the Jane Coffin Childs Fund.

Author contributions

W.J., S.B.B. and S.T. conceived the study. S.B.B., W.J. and S.T. designed experiments. S.B.B. and W.J. performed experiments and data analysis. P.O. assisted with computational analysis. S.B.B., W.J. and S.T. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

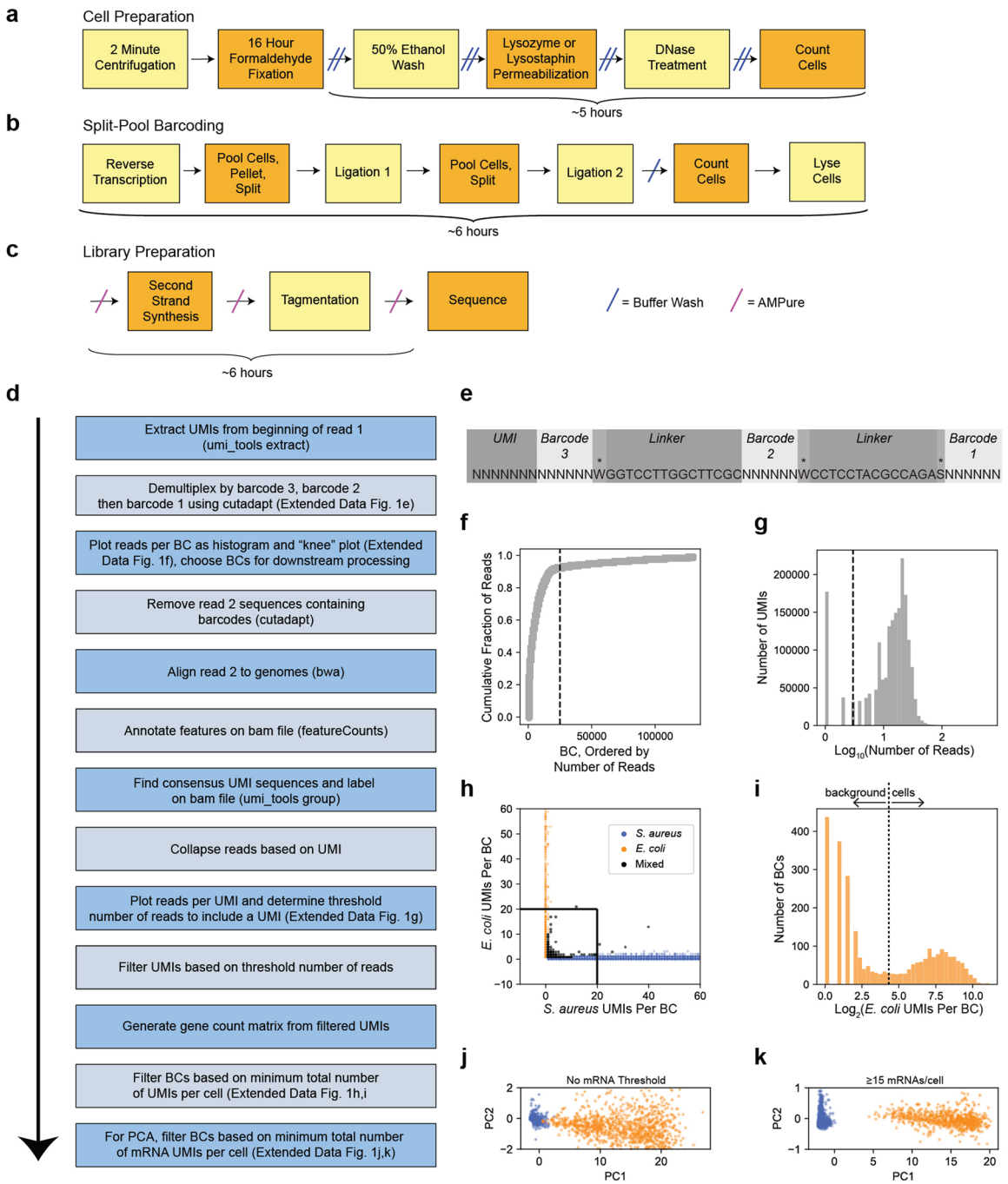
Extended data is available for this paper at https://doi.org/10.1038/s41564-020-0729-6.

Supplementary information is available for this paper at https://doi.org/10.1038/s41564-020-0729-6.

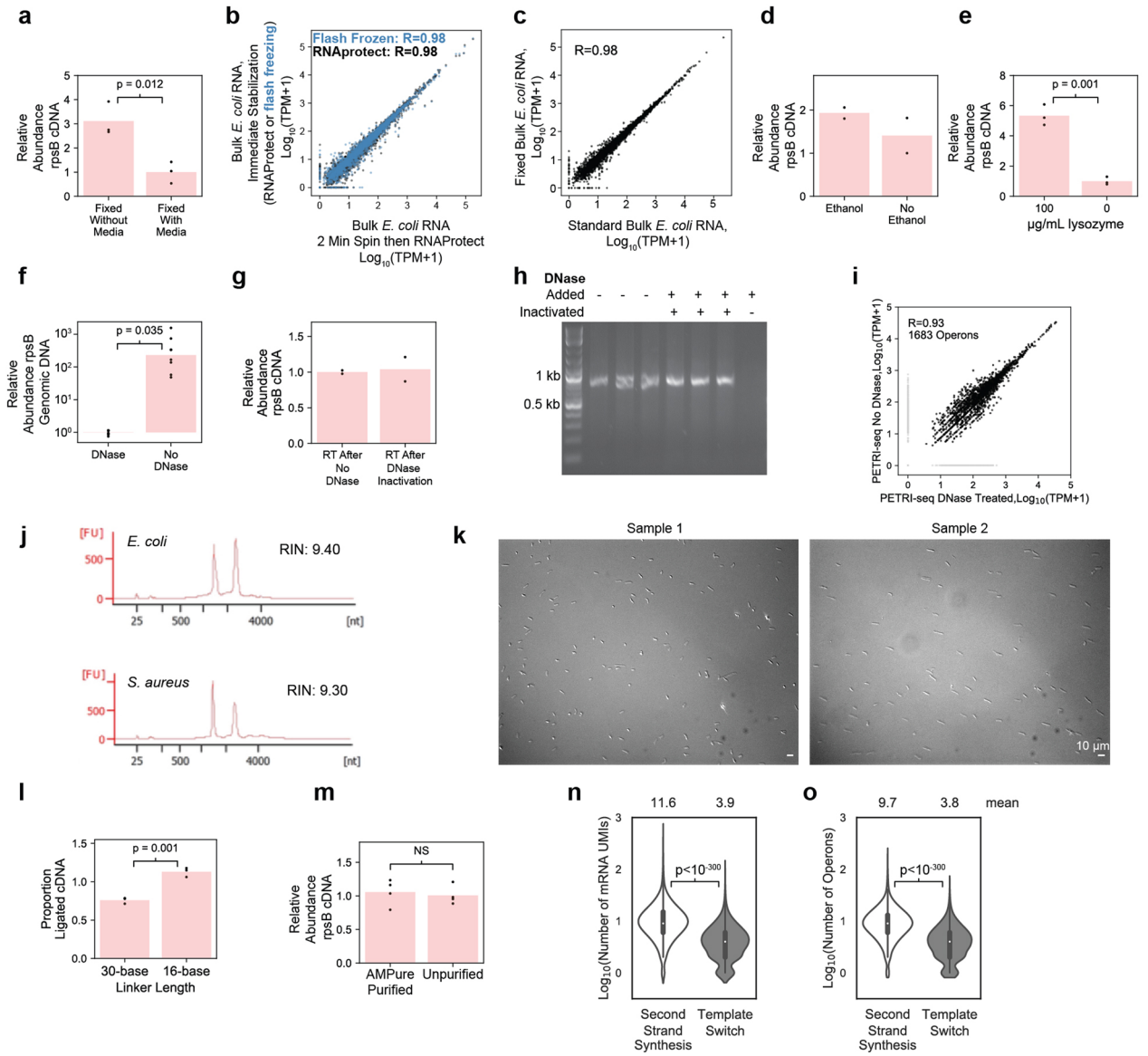
Correspondence and requests for materials should be addressed to S.T.

Reprints and permissions information is available at www.nature.com/reprints.

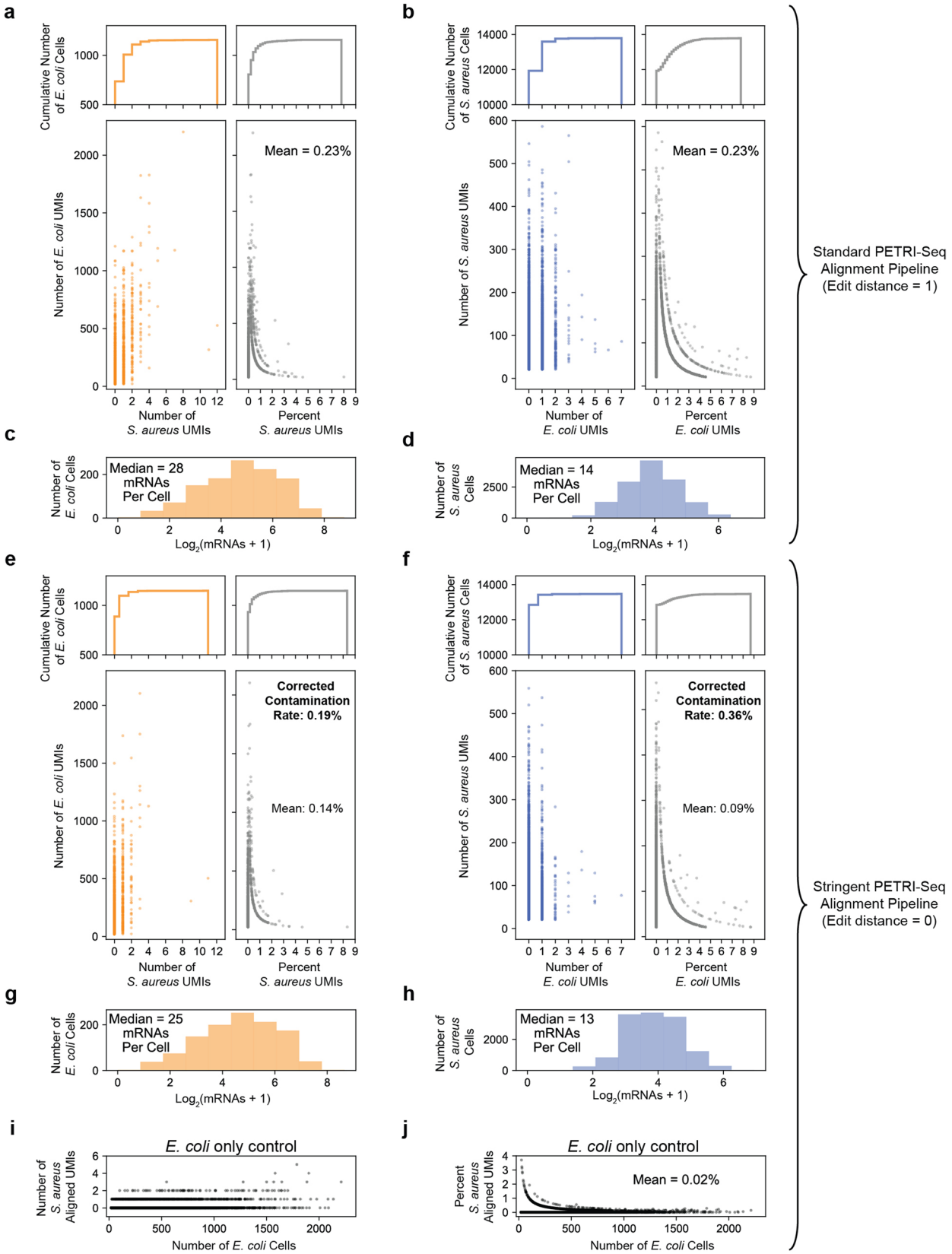
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Extended Data Fig. 1 | Experimental and computational pipelines for PETRI-seq. a–c, Experimental pipeline for PETRI-seq. PETRI-seq libraries can be prepared in just 2.5 days. (**a**) Detailed schematic of steps for cell preparation, which is started at the end of day 1 and finished on day 2. (**b**) Detailed schematic of steps for split-pool barcoding, which is entirely done on day 2. (**c**) Detailed schematic of steps for library preparation, which can be completed (up to sequencing) on day 3 (or later, if preferred). **d**, Computational pipeline for PETRI-seq analysis after sequencing. **e**, Structure of contig elements in read 1 after Illumina sequencing of PETRI-seq. To reduce the length of the sequence, barcodes overlap by one base (indicated by asterisk) with the adjacent linker sequence. **f**, Representative 'knee plot' used to select BCs for further analysis. The threshold line at 25,000 BCs is inclusive to facilitate additional filtering after collapsing PCR duplicates to UMIs. **g**, Representative histogram of reads per UMI. A threshold line was set for each library. For this library, only UMIs with more than 3 reads were kept for downstream analysis. Threshold line at $\log_{10}(3)$. **h**, Species mixing plot with all BCs containing >0 UMIs for library 1.06SaEc. BCs with fewer than 20 UMIs per cell were removed from further analysis. Line segments at x = 20 and y = 20. **i**, Distribution of *E. coli* BCs from species mixing plot in **h**. BCs above the threshold line were used for further analysis and considered single *E. coli* cells. Threshold line at $\log_2(20)$. **j**,**k**, PCAs of *E. coli* (orange) and *S. aureus* (blue) BCs from library 1.06SaEc. For calculation of principal components, rRNA operons were omitted and counts were normalized and scaled as described in methods. In **j**, all *S. aureus* and *E. coli* BCs with greater than 20 total UMIs and greater than 0 mRNAs are included (13,786 *S. aureus*, 1,153 *E. coli*). In **k**, only BCs with greater than or equal to 15 mRNA UMIs are included (6,683 *S. aureus*, 800 *E. coli*). For 100% of *S. aureus* BCs, PC1 < 0.05, and for 100% of *E. coli* BCs, PC1 > 4.

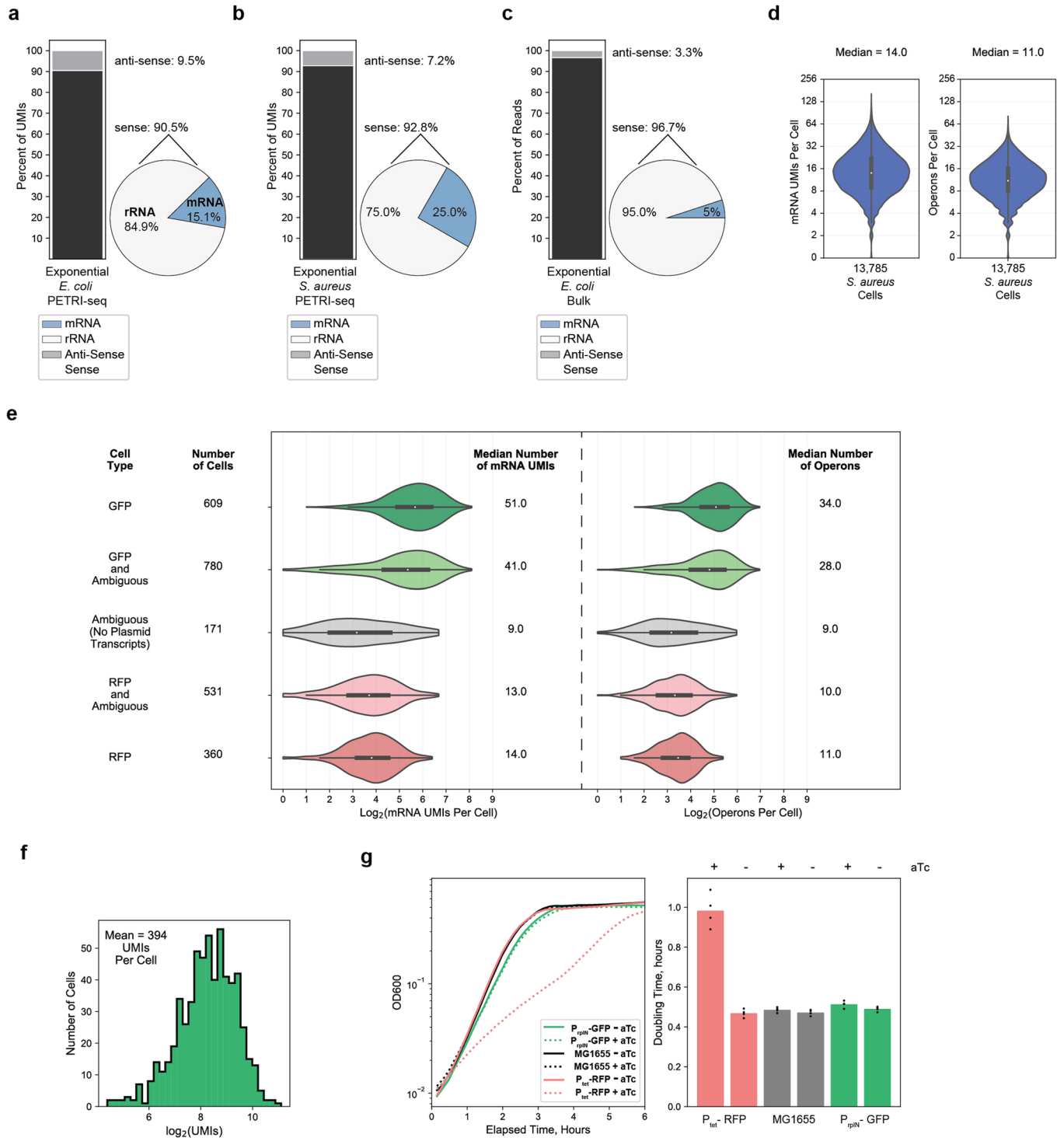


Extended Data Fig. 2 | Development and preliminary optimization of PETRI-seq. a, qPCR after *in situ* RT with random hexamers shows higher yield of *rpsB* cDNA from fixation without media (pelleting before) than fixation with media (formaldehyde added to culture) [n = 3 technically independent samples (dots), p = 0.012, 2-sided t-test]. Bars show mean abundance. **b**, Transcriptome stabilized by RNAprotect after 2-minute spin was highly correlated with transcriptomes stabilized immediately by either RNAprotect or flash freezing. Pearson's r is reported. **c**, RNA purified from *E. coli* cells after 16-hour 4% formaldehyde fixation ('Fixed Bulk') was highly correlated with non-fixed RNA ('Standard Bulk'). 2,617 operons included. Pearson's r is reported. **d**, qPCR after *in situ* RT with *rpsB*-specific primer (SB10) showed similar yield when cells were resuspended in 50% ethanol (n = 2 technically independent samples). **e**, qPCR after *in situ* RT with random hexamers shows improved yield of *rpsB* cDNA after lysozyme treatment (n = 3 technically independent samples [dots], p = 0.001, 2-sided t-test). Bars show mean abundance. **f**, qPCR after DNase treatment or incubation with only DNase buffer confirmed *in situ* DNase treatment efficacy (n = 8 technically independent samples [dots], p = 0.035, 2-sided t-test). Bars show mean abundance. **g**, qPCR after *in situ* RT with *rpsB*-specific primer (SB10) confirmed DNase inactivation, as yield was unchanged (n = 2 technically independent samples [dots]). Bars show mean proportion. **h**, Gel of 775-bp PCR fragment after 1-hour incubation with DNase-treated cells confirmed DNase inactivation. *Right-most lane:* DNase was directly added to PCR product. Experiment conducted one time. **i**, Aggregated PETRI-seq UMIs from DNase-treated and untreated libraries were highly correlated. Pearson's r reported. **j**, Bioanalyzer traces of RNA purified after *in situ* DNase treatment and cell lysis (methods). **k**, Imaging after *E. coli* cell preparation. Images for all libraries looked similar (n = 8). **l**, qPCR after bulk RT and ligation (methods) confirmed effective ligation with a 16-base linker. Minor increase (1.5×) in ligation efficiency was detected (p = 0.001, n = 3 technically independent samples [dots], 2-sided t-test). Bars show mean proportion. **m**, qPCR after *in situ* RT showed cDNA retention after AMPure purification (n = 4 technically independent samples, p = 0.69, 2-sided t-test). Bars show mean abundance. **n**,**o**, Second-strand synthesis yielded more mRNAs and operons per cell ($p < 10^{-300}$, 2-sided Mann-Whitney U) than template switching. 10,000 BCs are included from unoptimized PETRI-seq (Experiment 1.08). Boxplots within violins show interquartile range (black box) and median (white circle).

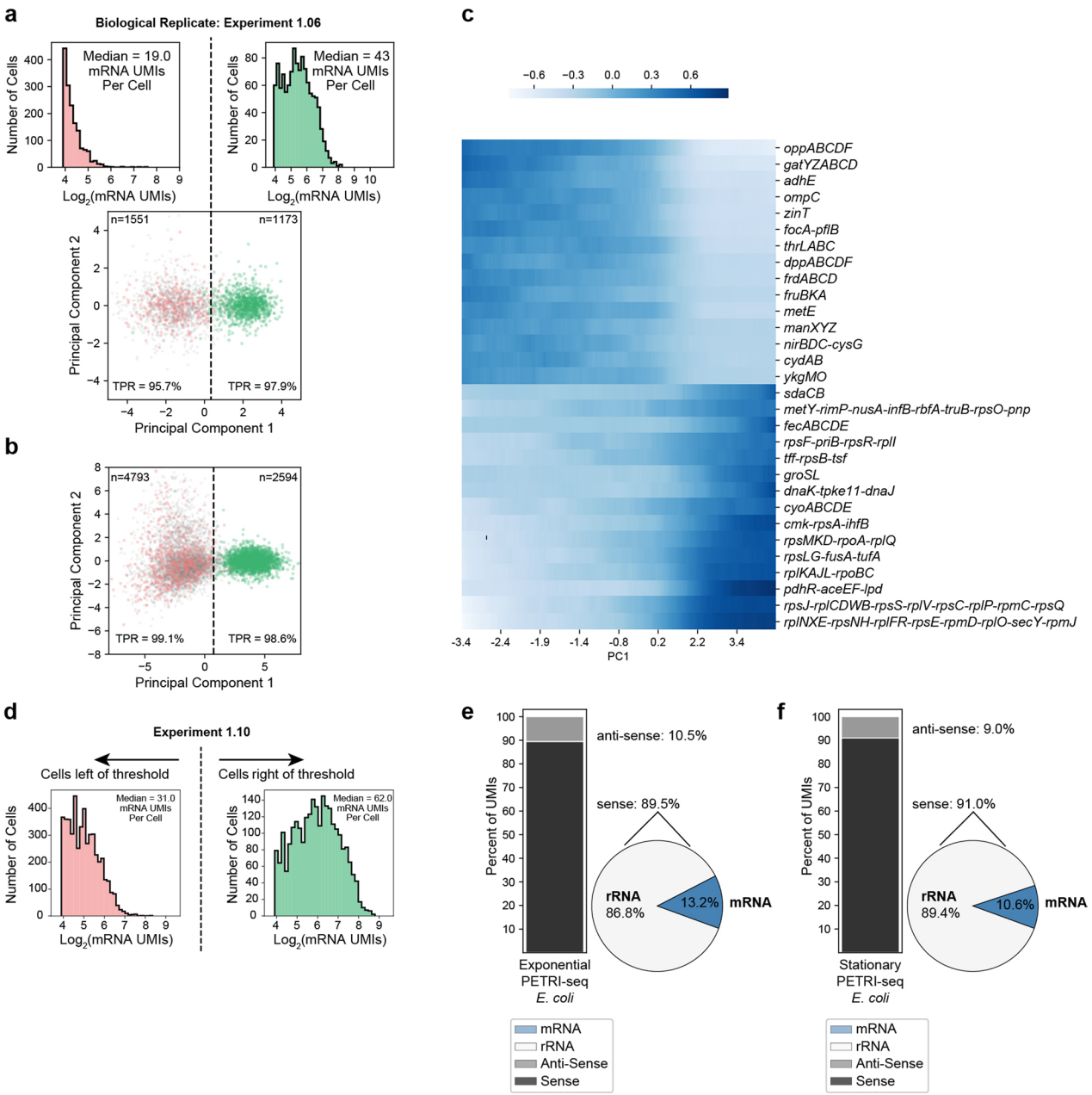


Extended Data Fig. 3 | See next page for caption.

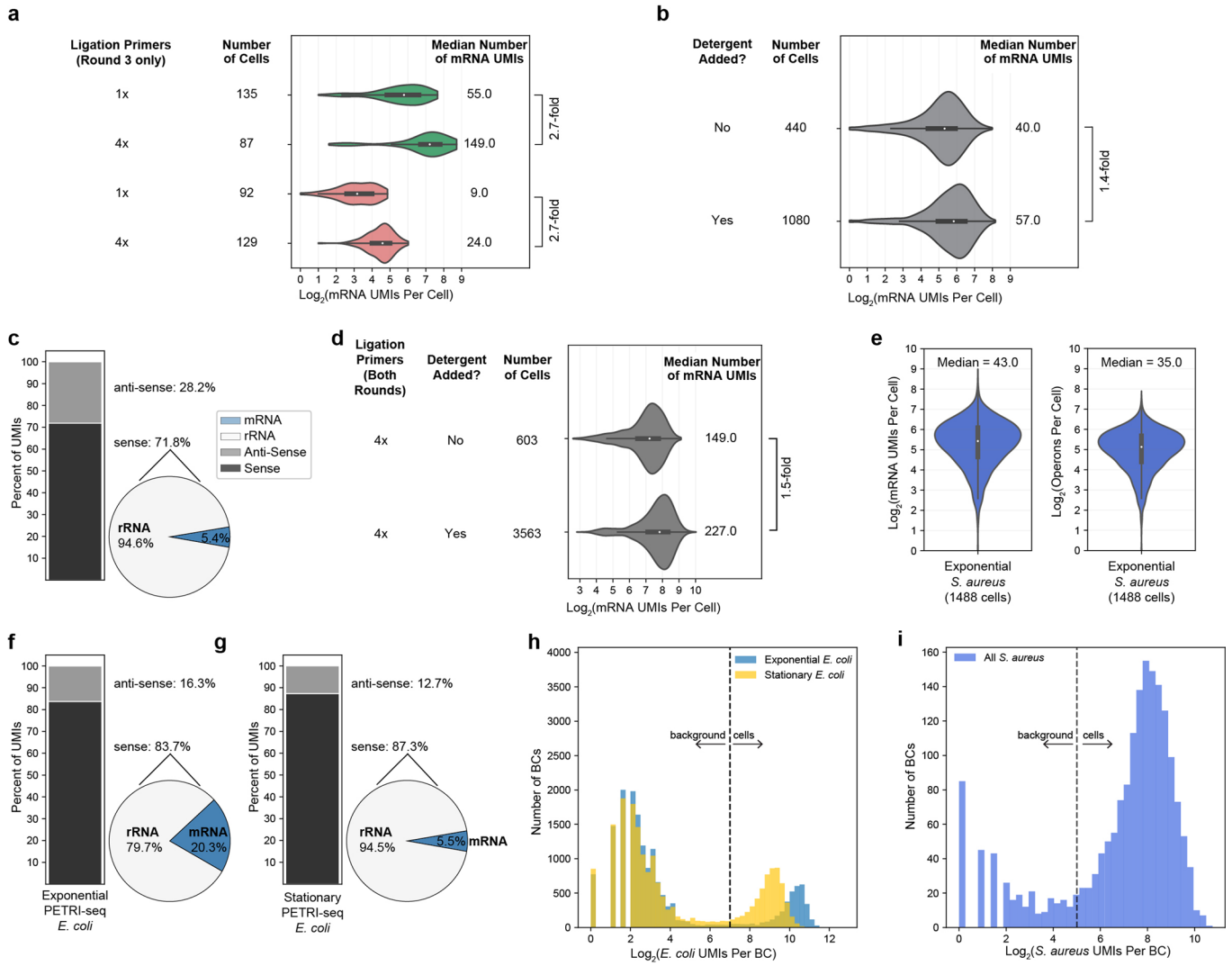
Extended Data Fig. 3 | Quantification of intercellular contamination using E. coli and S. aureus cells. After defining single *E. coli* and *S. aureus* cells (Fig. 2b, Experiment 1.06SaEc), we examined levels of cross-contamination within single cells. Similar analysis for Experiment 2.01 is shown in Extended Data Fig. 7c, d. **a**, Quantification of *S. aureus*-aligned UMIs assigned to *E. coli* cells after standard PETRI-seq alignment (edit distance ≤1). Reads mapping equally well to both species are discarded. *Bottom*: Scatterplots of *E. coli* UMIs vs. absolute (*left*) or percent (*right*) *S. aureus* UMIs assigned to each *E. coli* cell. *Top*: Cumulative distributions corresponding to scatterplots. **b**, Quantification of *E. coli*-aligned UMIs assigned to *S. aureus* cells after standard alignment. *Bottom*: Scatterplots of *S. aureus* UMIs vs absolute (*left*) or percent (*right*) *E. coli* UMIs assigned to each *S. aureus* cell. *Top*: Cumulative distributions corresponding to scatterplots. **c**, mRNAs per *E. coli* cell in **a**. **d**, mRNAs per *S. aureus* cell in **b**. **e**,**f**, Same analysis as (**a**,**b**) but using more stringent alignment (edit distance = 0) to better understand source of contamination. **g**, mRNAs per *E. coli* cell in **e**. **h**, mRNAs per *S. aureus* cell in **f**. **i**,**j**, To further understand the impact of alignment on apparent cross-contamination, we used stringent alignment to map UMIs for a library of only *E. coli* (Experiment 1.10). Total UMIs (**i**) or percent of UMIs (**j**) assigned to *S. aureus* were determined after stringent alignment for a PETRI-seq library prepared with only *E. coli*. *S. aureus* UMIs are computational artifacts. *E. coli* cells include a mean of 0.02% *S. aureus* aligned UMIs, indicating that the majority of interspecies contamination observed in **e** is not caused by incorrect alignment. To quantify contamination, we needed to correct percentages of inter-species alignment based on species abundance in the library (25% of UMIs aligned to *E. coli*, 75% *S. aureus*) to predict the percent of UMIs in a given single-cell derived from any other cell (whether or not the same species). We predict a 'corrected contamination rate', or percent of UMIs in a single-cell transcriptome derived from another cell, of 0.19-0.36% ($\frac{0.14}{0.75} = 0.19; \frac{0.09}{0.25} = 0.36$).



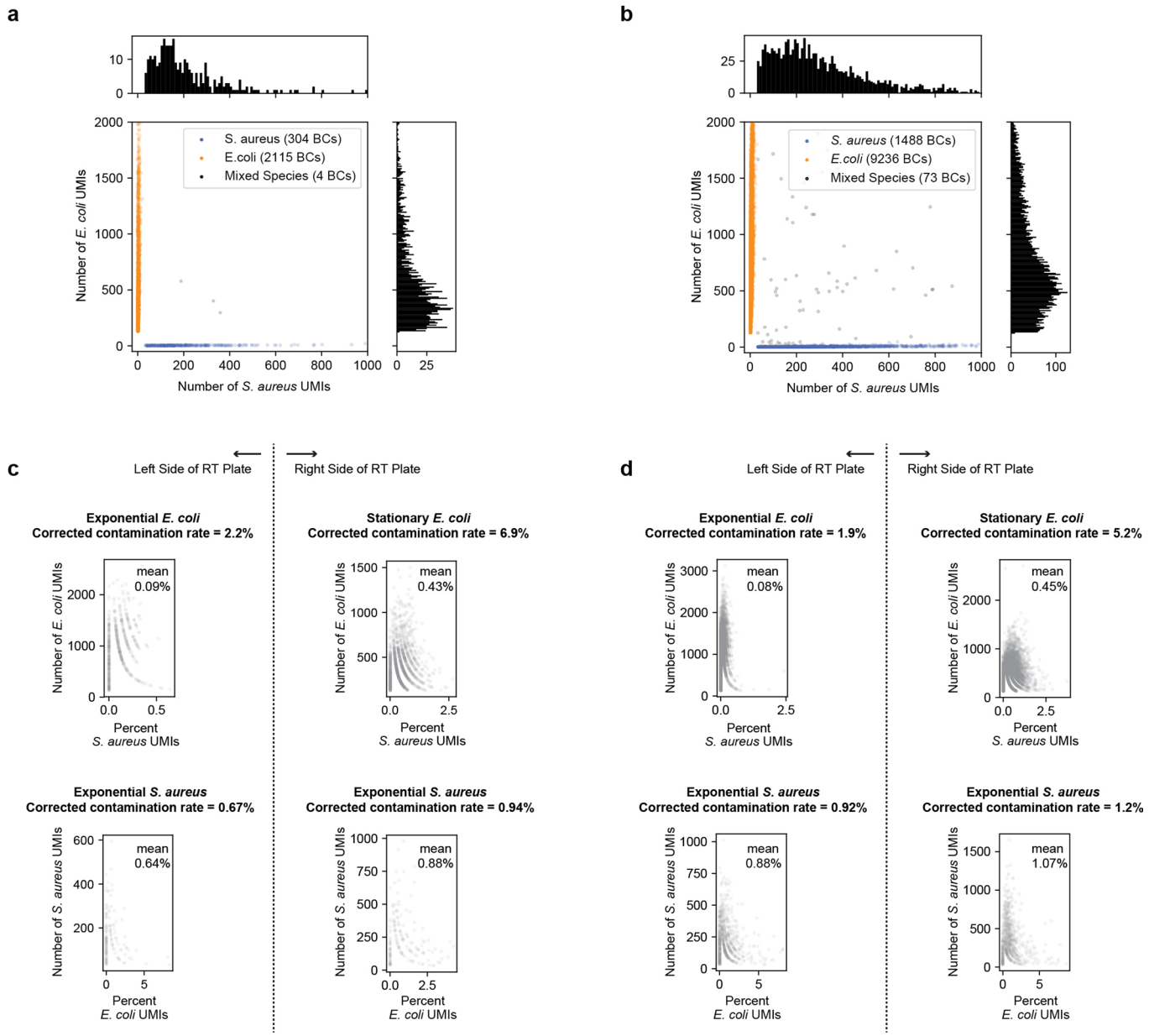
Extended Data Fig. 4 | Further evaluation of PETRI-Seq for E. coli and S. aureus in Experiment 1.06SaEc. a,**b**,**c**, Breakdown of total aligned UMIs (**a**,**b**) or reads (**c**) per cell for PETRI-seq exponential GFP- and RFP-expressing *E. coli* (**a**), PETRI-seq exponential *S. aureus* (**b**), and bulk exponential wild-type *E. coli* (**c**). *Left:* Stacked bar shows breakdown of sense and anti-sense alignments. *Right*: Pie shows breakdown of rRNA and mRNA alignments within the sense fraction. **d**, Distributions of mRNA UMIs (*left*) and operons (*right*) per *S. aureus* cell. 13,785 cells are included. 2 cells were omitted as they contained zero mRNAs. Boxplots within violins show interquartile range (black box) and median (white circle). **e**, Distributions of mRNA UMIs (*left*) and operons (*right*) per *E. coli* cell in five sub-populations, including GFP cells (contain GFP plasmid transcripts), RFP cells (contain RFP plasmid transcripts), ambiguous cells (contain no plasmid transcripts), and either RFP or GFP *and* ambiguous cells. Three ambiguous cells classified as *E. coli* in Fig. 2B were omitted as they contained zero mRNAs. Boxplots within violins show interquartile range (black box) and median (white circle). **f**, Distribution of total RNAs per GFP-containing exponential *E. coli* cell. 609 cells are included. **g**, *Left*, growth curves for P_{rplN}-GFP, P_{tet}-RFP, and MG1655 (no plasmid) cells with and without aTc. *Right*, doubling times calculated from the growth curves. P_{tet}-RFP had a significantly longer doubling time than all other strains/conditions when induced with aTc (n=4, p=2.2 * 10^{-5}, 2.5 * 10^{-5}, 2.1 * 10^{-5}, 3.6 * 10^{-5}, 2.6 * 10^{-5} [for each sample moving left to right], 2-sided t-test), which might explain fewer mRNA UMIs in these cells.



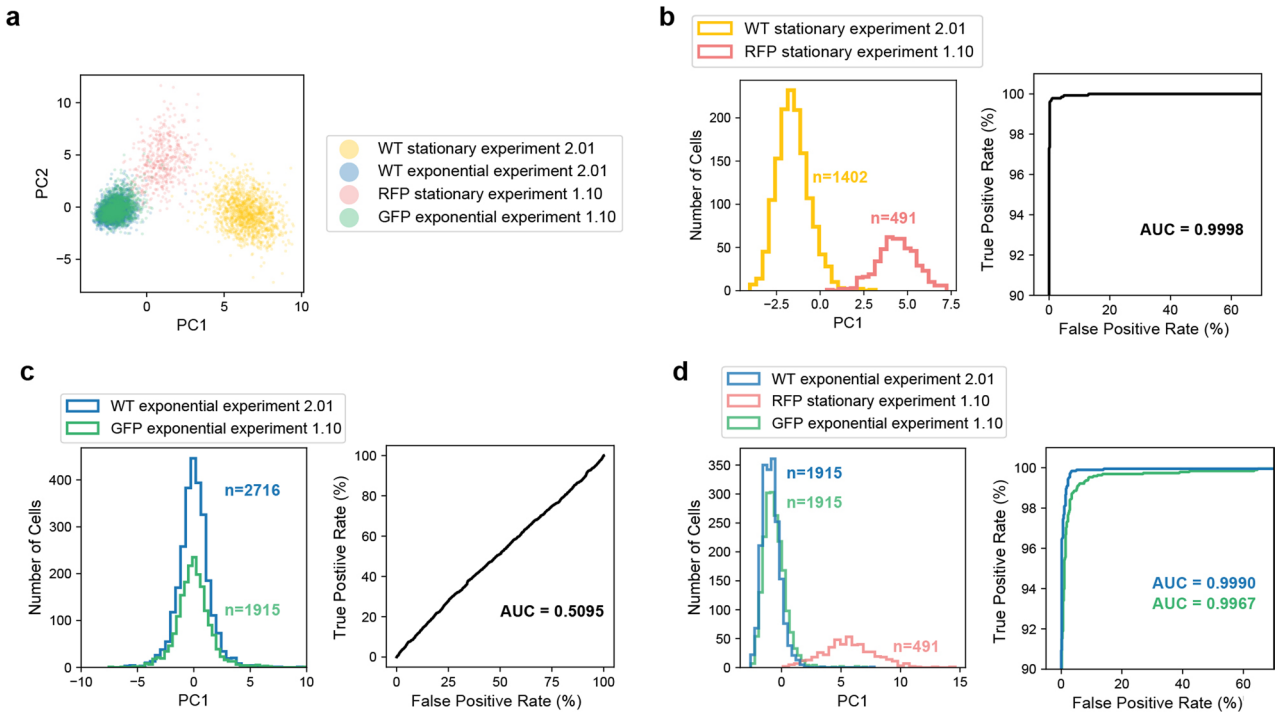
Extended Data Fig. 5 | Further evaluation of growth phase characterization by PETRI-seq. a, PCA of Experiment 1.06 (biological replicate of 1.10) shows that PETRI-seq can reproducibly distinguish between stationary and exponential cells by projecting cells onto the principal components calculated from the first library (*bottom*). 2,724 cells are included. 1,551 cells are left of the threshold (PC1=0.34), and 1,173 cells are right of the threshold. mRNA UMIs captured per cell on either side of the threshold line are shown (*top*). **b**, PCA as in Fig. 3b, but UMI counts were normalized using sctransform[26]. **c**, Expression along PC1 (Fig. 3b, Experiment 1.10) of operons with the most positive or negative PC1 loadings (z-scored moving average, size=1,000 cells). **d**, Distribution of mRNA UMIs per cell (Experiment 1.10) on either side of the threshold line in Fig. 3b. Grey cells (without plasmid UMIs) are included. Only cells with greater than 14 mRNA UMIs per cell were included, as cells with fewer were excluded from the PCA. 4,878 cells are left of the threshold, and 2,509 cells are right of the threshold. **e**,**f**, Breakdown of total aligned UMIs per cell for Experiment 1.10 for cells above and below the PC1 threshold in Fig. 3b. In **e**, Exponential *E. coli* (above the threshold) are shown and in **f**, stationary *E. coli* (below the threshold) are shown. *Left:* Stacked bar shows breakdown of sense and anti-sense alignments. *Right*: Pie shows breakdown of rRNA and mRNA alignments within the sense fraction.



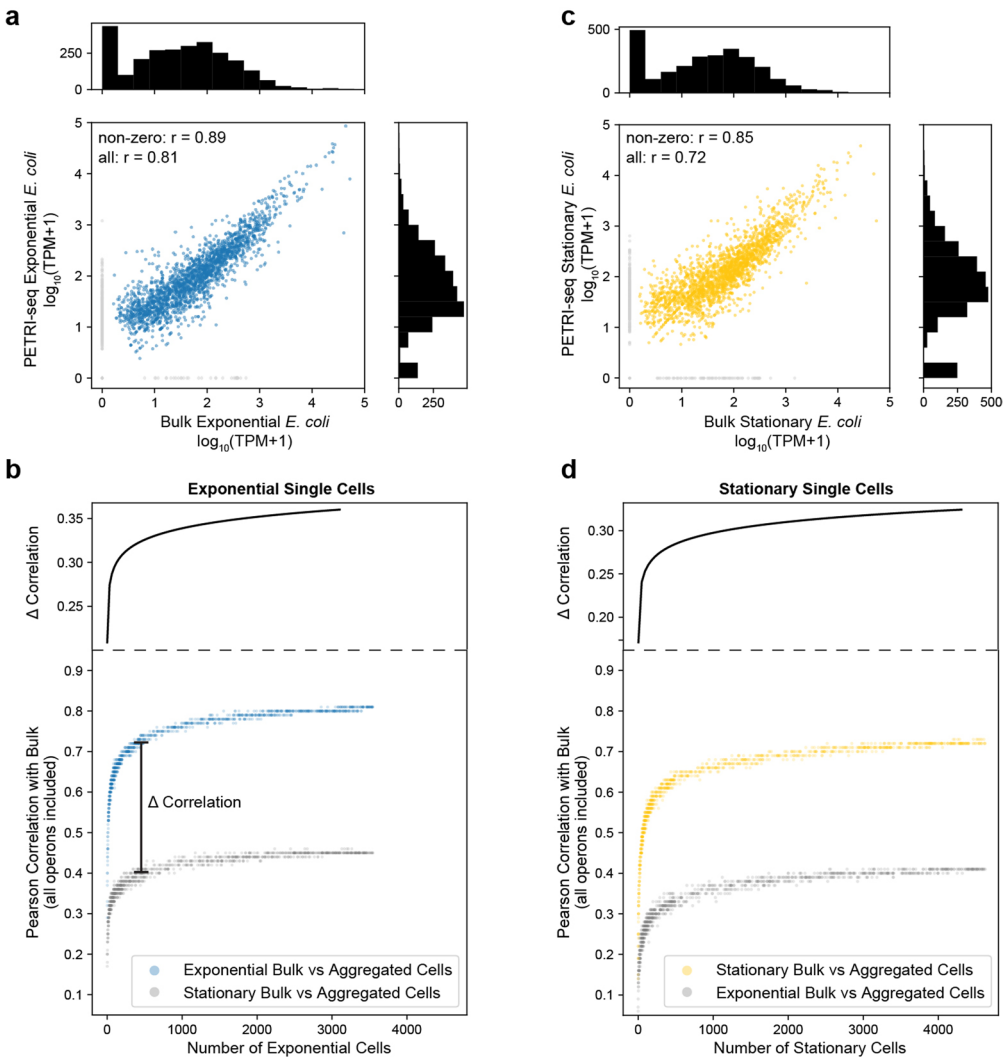
Extended Data Fig. 6 | Additional optimization of PETRI-seq by increasing ligation primer concentration and adding detergent during barcoding. **a**, Increasing the concentration of round 3 ligation primers by 4x relative to previous experiments (1.06SaEc and 1.10) increases mRNA UMIs per cell 2.7-fold for GFP-expressing exponential (green) and RFP-expressing stationary *E. coli* cells (red). Boxplots within violins show interquartile range (black box) and median (white circle). **b**, Adding detergent (tween-20) to cells before ligation 1 and after ligation 3 increased mRNA UMIs per cell 1.4-fold relative to original PETRI-seq for wild-type exponential *E. coli* cells. Boxplots within violins show interquartile range (black box) and median (white circle). **c**, With 10x more RT primer relative to original PETRI-seq, we observed a shift in the breakdown of sense/anti-sense and mRNA/rRNA UMIs. *Left*: Stacked bar shows breakdown of sense and anti-sense alignments. *Right*: Pie shows breakdown of rRNA and mRNA alignments within the sense fraction. Proportions of anti-sense RNAs and sense rRNAs are significantly increased. We hypothesized that any condition effectively increasing the intracellular concentration of RT primers could lead to this undesirable shift. For this reason, detergent was only ever added after RT to avoid further permeabilizing cells and increasing the effective concentration of RT primer. **d**, Combining detergent treatment and increased ligation primer (for both rounds) resulted in higher mRNA capture for wild-type exponential *E. coli* cells. Detergent again increased mRNA UMIs per cell (1.5-fold). Boxplots within violins show interquartile range (black box) and median (white circle). **e**, Optimized PETRI-seq (4x ligation primer, detergent treatment) resulted in *S. aureus* transcriptomes with a median of 43 mRNA UMIs per cell (*left*) and 35 operons per cell (*right*). Boxplots within violins show interquartile range (black box) and median (white circle). **f**,**g**, Breakdown of total aligned UMIs per cell for optimized PETRI-seq (Experiment 2.01) for exponential (**f**) and stationary *E. coli* (**g**). *Left:* Stacked bar shows breakdown of sense and anti-sense alignments. *Right*: Pie shows breakdown of sense rRNA and mRNA alignments. **h**,**i**, Distributions of total UMIs per *E. coli* (**h**) and *S. aureus* (**i**) BCs in Experiment 2.01. Given higher capture, we imposed higher thresholds for distinguishing cells from background than used previously (Extended Data Fig. 1i). *E. coli* BCs with more than 128 total UMIs (threshold line in **h**) and *S. aureus* BCs with more than 32 total UMIs (threshold line in **i**) were considered cells.



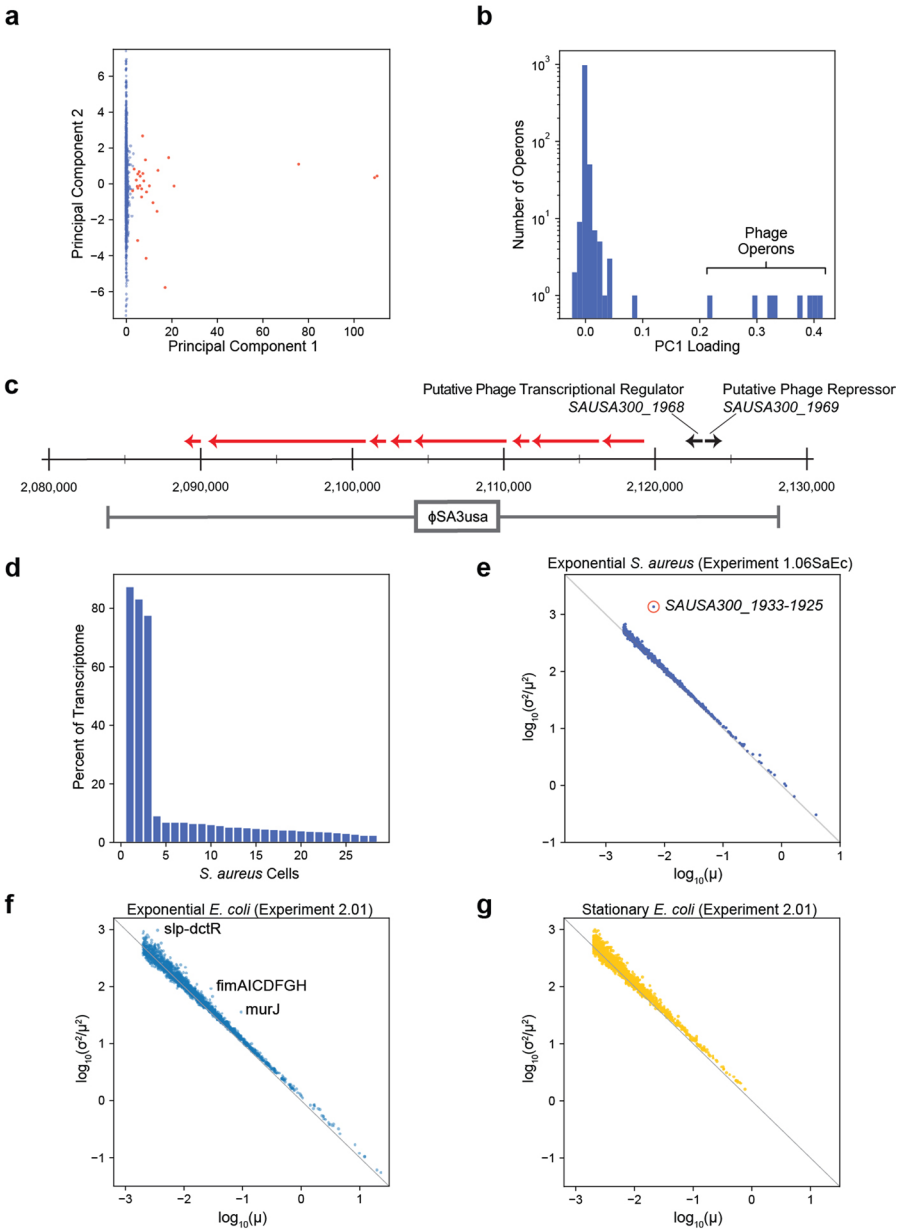
Extended Data Fig. 7 | Multiplet frequency and intercellular contamination for optimized PETRI-seq. a, Species mixing plot for PETRI-seq with 4x ligation primers and no detergent. The multiplet frequency is 0.7%, which is 5-fold higher than the Poisson expectation of 0.14% for 2,423 BCs. **b**, Species mixing plot for PETRI-seq with 4x ligation primers and detergent (Experiment 2.01). The multiplet frequency is 2.8%, which is 4.7-fold higher than the Poisson expectation of 0.6% for 10,797 BCs. This indicates that compared to no detergent, detergent treatment did not significantly increase multiplet frequency relative to the Poisson expectation. In (a,b), *E. coli* BCs with > 128 total UMIs and *S. aureus* BCs with > 32 total UMIs were included. **c**,**d**, Quantification of cross-contamination for PETRI-seq with 4x ligation primers and no detergent (**c**, same experiment as **a**) or 4x ligation primers and detergent (**d**, Experiment 2.01 as in **b**). Scatterplots show the percent of total UMIs for each cell aligned to the incorrect species. Reads were aligned using the stringent alignment (edit distance = 0) described in Extended Data Fig. 3. *Top left:* Percent of *S. aureus* UMIs in exponential *E. coli* cells (based on first round barcode). *Top right:* Percent of *S. aureus* UMIs in stationary *E. coli* cells (based on first round barcode). *Bottom left:* Percent of *E. coli* UMIs in *S. aureus* cells barcoded with exponential *E. coli* (based on first round barcode). *Bottom right:* Percent of *E. coli* UMIs per *S. aureus* cell barcoded with stationary *E. coli* (based on first round barcode). As described in Extended Data Fig. 3, we used these inter-species contamination rates to predict a corrected contamination rate (including intra-species contamination). Though higher than the contamination rates observed in the previous species mixing experiment (Extended Data Fig. 3e, f), these rates are comparable to previous findings for eukaryotic scRNA-seq methods[23,24] and are not affected by detergent treatment (**c** vs. **d**). Furthermore, we anticipate that contamination could be reduced by additional washing prior to cell lysis (see 'Future directions for optimization' in Methods).



Extended Data Fig. 8 | Comparison of plasmid-labeled (Experiment 1.10) and RT-labeled (Experiment 2.01) mixed growth stage libraries reveals minimal cross-contamination between E. coli cells barcoded together. In Experiment 2.01, exponential and stationary cells were prepared separately and then barcoded independently during RT. In contrast, the RFP-expressing stationary cells and GFP-expressing exponential cells barcoded in Experiment 1.10 were combined for fixation and barcoded together, resulting in more opportunity for cross-contamination. Experiment 2.01 is thus a useful reference to quantify this cross-contamination. To account for differences in the capture efficiency for the two experiments, cells were down-sampled to 30 mRNA UMIs. **a**, PCA for all 4 cell types reveals that the two stationary populations are biologically distinct, possibly because they were grown independently to slightly different ODs, and RFP cells were induced with aTc. In contrast, the two exponential populations appear very similar. **b**, PC1 was calculated using only the stationary cells from both experiments. *Right*: The receiver operating characteristic (ROC) shows that PC1 is a strong classifier of the two states. **c**, PC1 was calculated using only exponential cells from both experiments. *Right*: The ROC shows that PC1 is a weak classifier of the two exponential states with performance similar to random assignment (Area Under the ROC Curve [AUC]=0.5). **d**, PC1 was calculated using wild-type exponential cells from Experiment 2.01, GFP-expressing exponential cells from Experiment 1.10, and RFP-expressing stationary cells from Experiment 1.10 in order to quantify cross-contamination between the GFP and RFP cells using the wild-type exponential cells from Experiment 2.01 as a reference. *Right*: ROC shows that PC1 is a strong classifier of exponential and stationary cells. The probability that the PC1 value of a wild-type exponential cell is lower than the PC1 value of a stationary RFP cell is 99.9% (AUC = 0.999), while the probability that the PC1 value of a GFP exponential cell is lower than the PC1 value of a stationary RFP cell is 99.67% (AUC = 0.9967). Thus, for the GFP exponential cells, 23 out of 10,000 cell pairs (1 exponential, 1 stationary) will be incorrectly ranked due to cross-contamination in the GFP cells. Finally, we confirmed that in the original library for Experiment 1.10, the relative representation of UMIs from exponential and stationary cells were roughly equal (50.3% stationary, 45.6% exponential), indicating that the cross-contamination analysis for the GFP exponential population would be reciprocal for the RFP stationary population.



Extended Data Fig. 9 | Defining consensus transcriptional states of sub-populations by aggregating single-cell transcriptomes. a, Correlation between mRNA abundances from 3,547 aggregated wild-type exponential cells (Experiment 2.01) vs. bulk preparation from fixed exponential wild-type *E. coli* cells. The Pearson correlation coefficient (r) was calculated for 2,150 out of 2,612 total operons, excluding those with zero counts in either library (grey points), or for all 2,612 operons. Bulk library was prepared from the same cells as the PETRI-seq library. **b**, *Bottom*: The correlation between the aggregated mRNA counts of single exponential cells (PETRI-seq) and the bulk exponential library increases as more single cells are included. Correlations were calculated from $\log_{10}(\text{TPM}+1)$ for each sample. *Top*: Difference between top curve and bottom curve in plot below, based on best-fit lines ($y = \ln(x) + b$, $r > 0.98$). **c**, Correlation between RNA abundances from 4,627 aggregated wild-type stationary cells (Experiment 2.01) vs. bulk preparation from fixed wild-type stationary *E. coli* cells. The Pearson correlation coefficient (r) was calculated for 2,050 out of 2,612 total operons, excluding those with zero counts in either library (grey points), or for all 2,612 operons. Bulk library was prepared from the same cells as the PETRI-seq library. **d**, *Bottom*: The correlation between the aggregated mRNA counts of single stationary cells (PETRI-seq) and the bulk stationary library increases as more single cells are included. Correlations were calculated from $\log_{10}(\text{TPM}+1)$ for each sample. *Top*: Difference between top curve and bottom curve in plot below, based on best-fit lines ($y = \ln(x) + b$, $r > 0.98$).



Extended Data Fig. 10 | PETRI-seq detects rare transcriptional states and candidate genes with highly variable expression. a, PCA detects rare transcriptional states among 6,663 *S. aureus* cells. A small sub-population of 28 cells (*red*) expressed operons from the φSA3usa phage. **b**, Distribution of PC1 loadings for all operons included in the *S. aureus* analysis. Eight operons from the φSA3usa phage have the highest PC1 loadings. **c**, Map of genomic region[33] surrounding φSA3usa in the genome of *S. aureus* strain USA300. Red arrows indicate phage operons upregulated along PC1. **d**, Percent of mRNA UMIs mapped to the φSA3usa phage for the 28 cells containing phage UMIs. Three cells are composed of >77% phage transcripts. **e**, Noise (σ^2/μ^2) versus mean (μ) for operon expression within an *S. aureus* population of 6,663 cells. 676 operons are included. The circled operon (red) is *SAUSA300_1933-1925*, which deviated significantly from the rest of the distribution (z-score = 20.6 [determined by residuals from linear regression (see methods)], $p = 10^{-94}$, FDR < 0.01). **f**,**g**, Noise (σ^2/μ^2) versus mean (μ) for operon expression in either exponential (**f**) or stationary (**g**) *E. coli* populations from Experiment 2.01. 1,960 operons are included in (**f**) and 1,219 operons in (**g**). Five operons significantly (FDR < 0.01, z-scores determined by residuals from linear regression [see methods]) deviated from the other operons in (**f**): *sip-dctR* (z-score = 7.3, $p = 3*10^{-13}$), *murJ* (z-score = 6.7, $p = 3*10^{-11}$, *fimAICDFGH* (z-score = 5.4, $p = 7*10^{-8}$), *mdtL* (z-score = 4.8, $p = 1*10^{-6}$), *rnhA* (z-score = 4.6, $p = 4*10^{-6}$). *fimAICDFGH*, which encodes the type I fimbriae system, has been shown previously to exhibit population-level phase variation that is mediated by transcriptional control[37]. In (e-g), lines at y = -x indicate Poisson noise where $\sigma^2 = \mu$. Operon counts were normalized for each cell before plotting. Operons with fewer than 6 raw total UMIs and a mean less than 0.002 after normalization were excluded.

nature research

Corresponding author(s): Saeed Tavazoie

Last updated by author(s): Apr 11, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Sequencing was done on a NextSeq 500 at the Columbia Genome Center. Fastq files were generated using bcl2fastq.
Data analysis	Cutadapt 1.18, UMI-tools 0.5.5, bwa 0.7.17, featureCounts 1.6.3, and trimmomatic 0.33 were used for data processing. Custom scripts were implemented in Python 2.7.15 (collapsing reads to UMIs, PCA, figure generation) or R 3.6.1 (sctransform). All relevant code is available upon request.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw data has been uploaded to the Gene Expression Omnibus (GEO) under accession number GSE141018. All figures except Figures 1, S1, S2 include original data.

Field-specific reporting

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

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

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

Life sciences study design

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

Sample size	Number of cells was chosen to minimize the multiplet frequency. Number of cells for each experiment is indicated in the relevant figures.
Data exclusions	BCs below a threshold of UMIs per cell (listed in table S4) were excluded from analysis. Thresholds were determined empirically by plotting distributions of UMIs per cell. For PCA, BCs with fewer than 15 mRNA UMIs were excluded to avoid spurious results. This threshold was set during data analysis.
Replication	Two independent replicate libraries of stationary/exponential E. coli cells were prepared months apart and confirmed reproducibility in capture rate and biological differences between the populations.
Randomization	na
Blinding	na

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging