

Supplementary Information for:

**Bacterial single-cell RNA sequencing captures biofilm transcriptional heterogeneity and differential responses to immune pressure**

Lee E. Korshoj<sup>1,\*</sup> and Tammy Kielian<sup>1,\*\*</sup>

<sup>1</sup>Department of Pathology, Microbiology, and Immunology, University of Nebraska Medical Center, Omaha, NE, USA

**Correspondence:**

\*lee.korshoj@unmc.edu

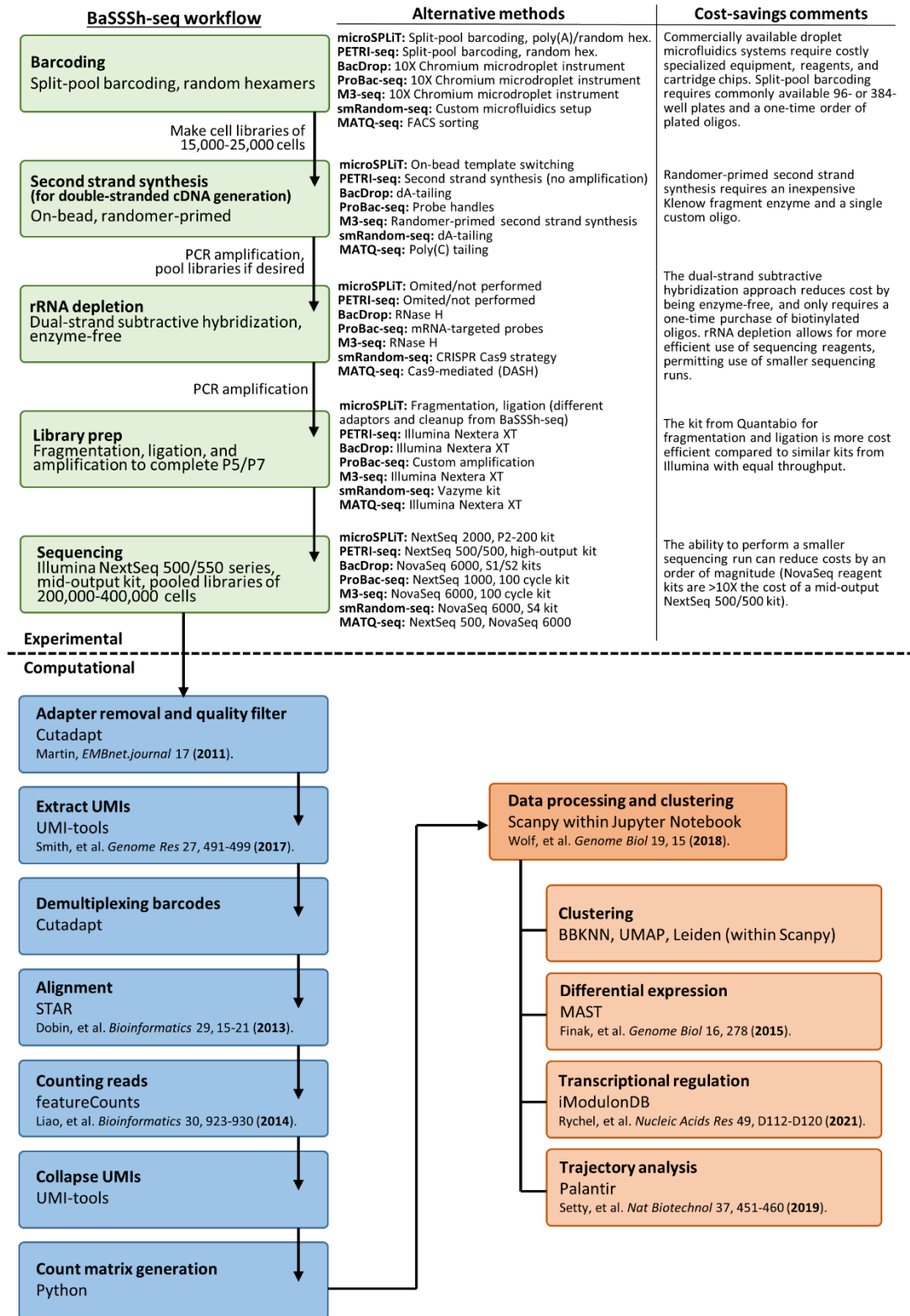
\*\*tkielian@unmc.edu

**Supplementary Table 1. Oligos used in BaSSSh-seq**

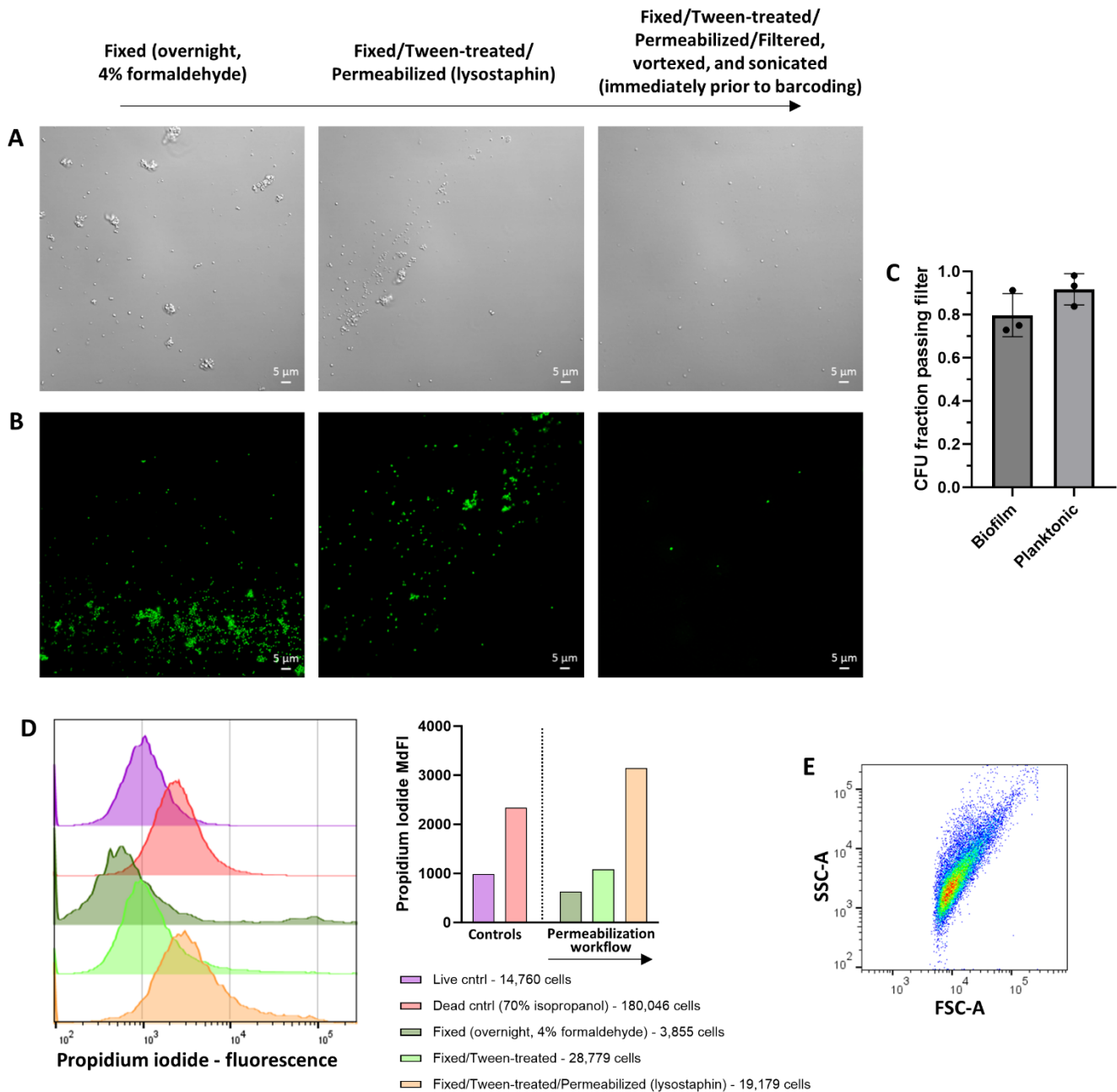
Oligos	Source	Sequence or location
Round_1_barcoding	IDT	Supplementary Data 13
Round_2_barcoding	IDT	Supplementary Data 14
Round_3_barcoding	IDT	Supplementary Data 15
Round2_linker	IDT	CCACAGTCTCAAGCACG
Round3_linker	IDT	TACGCCGATGCGAAACATCG
Round2_blocking (blocking after adding barcode 2 via ligation – blocks non-bound Round2_linker)	IDT	CGTGCTTGAGACTGTGG
Round2_blocking_HP (blocking after adding barcode 2 via ligation – blocks open barcode 2-linker constructs)	IDT	ACTGTGGACGTTAGGCAGGACCTAACGT
Round3_blocking (blocking after adding barcode 3 via ligation – blocks non-bound Round3_linker)	IDT	CGATGTTTCGCATCGGCGTA
Round3_blocking_HP (blocking after adding barcode 3 via ligation – blocks open barcode 3-linker constructs)	IDT	CATCGGCGTATGCAATCGGACCTCGATTGCA
S <sup>3</sup> _randomer (second strand synthesis randomer)	IDT	AAGCAGTCCTATCAACGCTCCACGANNNCCNNNB
PCR_P1 (amplification primer for the S <sup>3</sup> _randomer-modified end)	IDT	AAGCAGTCCTATCAACGCTCCAC
PCR_P2 (amplification primer for the end of barcode 3)	IDT	CAGACGTGTGCTCTTCCGATCT
rRNA_dep_Fwd (set of rRNA depletion oligos for the forward strand)	IDT	Supplementary Data 16
rRNA_dep_Rev (set of rRNA depletion oligos for the reverse strand)	IDT	Supplementary Data 17
Adapter_duplex_Top (top oligo for ligation during library prep)	IDT	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG*T, * denotes phosphorothioate bond
Adapter_duplex_Bott (bottom oligo for ligation during library prep)	IDT	CTGTCTCTTATACACATCTGACGCTGCCGACGA
Seq_i501 (P5 end with index i501)	IDT	AATGATACGGCGACCACCGAGATCTACACTAGATCGCTCGT CGGCAGCGTCAGATGTGTATAAGAGACAG
Seq_i502 (P5 end with index i502)	IDT	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTTCGTC GGCAGCGTCAGATGTGTATAAGAGACAG
Seq_i701 (P7 end with index i701)	IDT	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAGTGACTGGA GTTTCAGACGTGTGCTCTTCCGATCT
Seq_i702 (P7 end with index i702)	IDT	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTGACTGGA GTTTCAGACGTGTGCTCTTCCGATCT
gyrB_Fwd	IDT	GGAATCGGTGGCGACTTTGA
gyrB_Rev	IDT	CCATCCACATCGGCATCAGT
23S_Fwd	IDT	GGGAGGACCATCTCCTAAGGC
23S_Rev	IDT	AGGCACGCCATCACCCATTAAC
16S_Fwd	IDT	GCGAAAGCCTGACGGAGCAA
16S_Rev	IDT	GCGCTTTACGCCCAATAATTCCGG

**Supplementary Table 2. Reagents used in BaSSSh-seq and related experiments**

Reagent	Source	Product number
Anti-Ly6G MicroBeads, UltraPure, mouse	Miltenyi Biotec	130-120-337
SUPERase-In RNase Inhibitor (20 U/ $\mu$ L)	ThermoFisher	AM2696
Lysostaphin	AMBI Products LLC	LSPN-50 (Ambicin L)
dNTPs	NEB	N0447L
Maxima H Minus Reverse Transcriptase (200 U/ $\mu$ L)	ThermoFisher	EP0753
Cell strainers (1 $\mu$ m and 10 $\mu$ m)	pluriSelect	43-50001-13 43-50010-03
Cell strainer connector	pluriSelect	41-50000-03
T4 DNA Ligase (400,000 U/mL)	NEB	M0202L
Proteinase K, recombinant, PCR grade (~20 mg/mL)	ThermoFisher	EO0492
Hydrophilic Streptavidin Magnetic Beads	NEB	S1421S
Klenow Fragment (3'→5' exo-)	NEB	M0212S
KAPA HiFi HotStart ReadyMix	Roche	07958927001
SPRIselect	Beckman Coulter	B23318
sparQ DNA fragment and library prep kit	Quantabio	95194-024
NEBNext Library Quant Kit for Illumina	NEB	E7630L
Enzytec Liquid D-Glucose	r-biopharm	E8140
Enzytec Liquid Acetic acid	r-biopharm	E8226
Enzytec Liquid D-/L-Lactic acid	r-biopharm	E8240
5-cyano-2,3-ditoyl tetrazolium chloride (CTC)	ThermoFisher	B34956
6-chloromethyl-2',7'-dichlorodihydrofluorescein diacetate (CM-H <sub>2</sub> DCFDA)	ThermoFisher	C6827
CellTracker Deep Red Dye	ThermoFisher	C34565



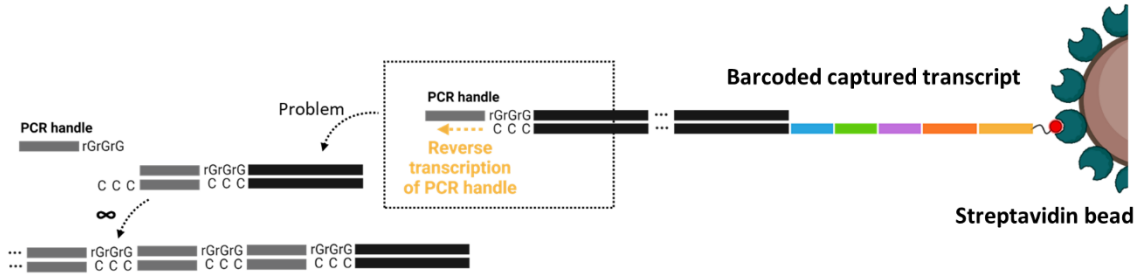
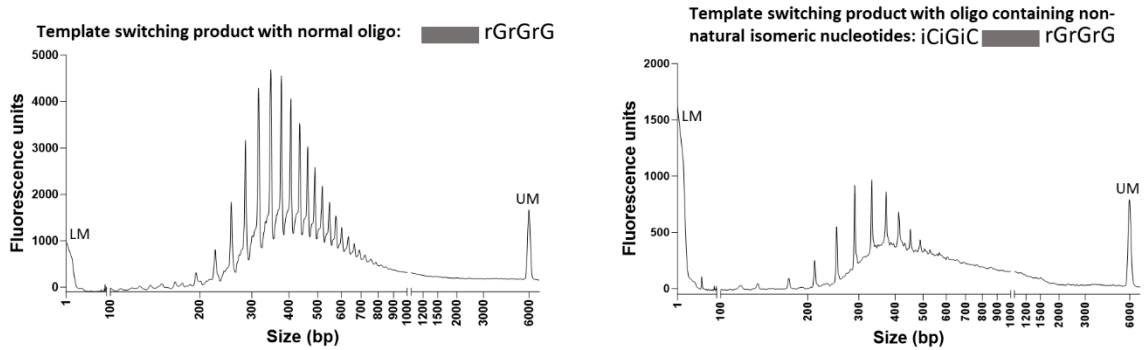
**Supplementary Figure 1 | BaSSSh-seq workflow.** Experimental overview of BaSSSh-seq and its computational analysis. Comparisons are provided to alternative bacterial scRNA-seq methods<sup>1-7</sup> and key aspects of cost-savings.



### Supplementary Figure 2 | Single cell isolation and permeabilization for barcoding.

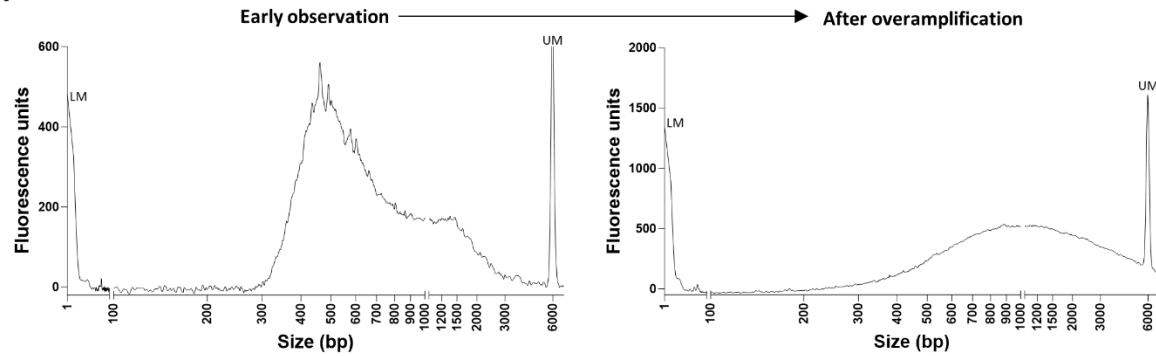
Representative (A) brightfield and (B) fluorescence (SYTO 24-stained cells) images of *S. aureus* biofilm cells at three stages of the BaSSSh-seq workflow prior to barcoding: (left) after overnight fixation; (middle) after Tween treatment and lysostaphin permeabilization of fixed cells; and (right) after filtering, vortexing, and sonicating the fixed and permeabilized cells immediately prior to the first reverse transcription round of barcoding. Note brightfield (A) and fluorescence (B) images are not of corresponding regions. Images were captured using a Zeiss 710 laser scanning microscope with a 60X oil lens, and represent at least five similarly captured areas. (C) Quantifying cell losses from filtering of biofilm and planktonic samples. Filtering consists of passing cells through consecutive 10  $\mu$ m and 1  $\mu$ m cell strainers, as described in the Methods. Data are presented as mean  $\pm$  standard deviation from 3 biological replicates. (D) Flow cytometry quantification of *S. aureus* permeabilization by propidium iodide uptake at different

stages of the BaSSSh-seq workflow prior to barcoding, namely after overnight fixation, Tween treatment, and lysostaphin permeabilization. Results are presented from one experiment. (E) Representative FSC-A vs. SSC-A pseudocolor plot, from which all cells were taken for histogram analysis in (D). Source data are provided as a Source Data file.

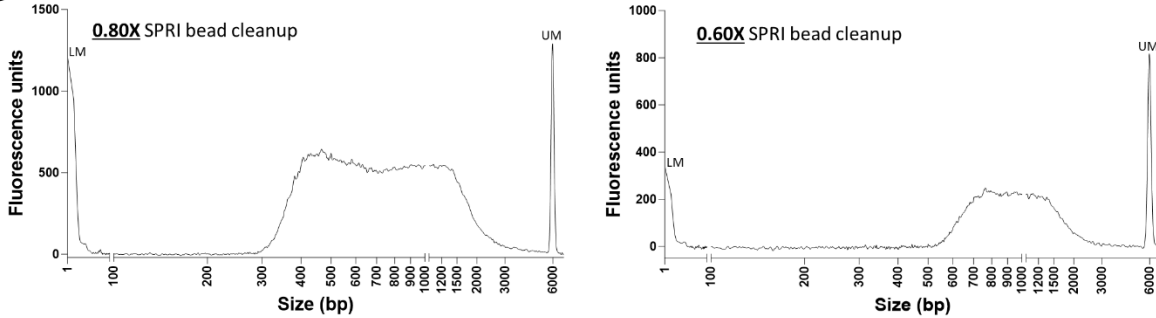
**A****B**

**Supplementary Figure 3 | Difficulties with template switching for generating double-stranded cDNA.** (A) Illustration of the concatamerization issue when performing template switching with low concentrations of starting material. Ultimately, the concatamers overrun the sequencing output. (B) Representative observation of the concatamerization issue in fragment analyses with (left) a normal unmodified template switching oligo, and (right) a template switching oligo containing non-natural isomeric nucleotides at the 5' end as previously described to suppress concatamerization<sup>8</sup>. The modified oligo alleviated but did not solve the concatamerization issue. All size distributions were measured on a 5200 Fragment Analyzer (Agilent). Lower markers (LM) are indicated at 1 bp, and upper markers (UM) are indicated at 6,000 bp. Source data are provided as a Source Data file.

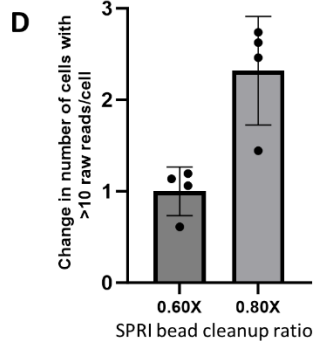
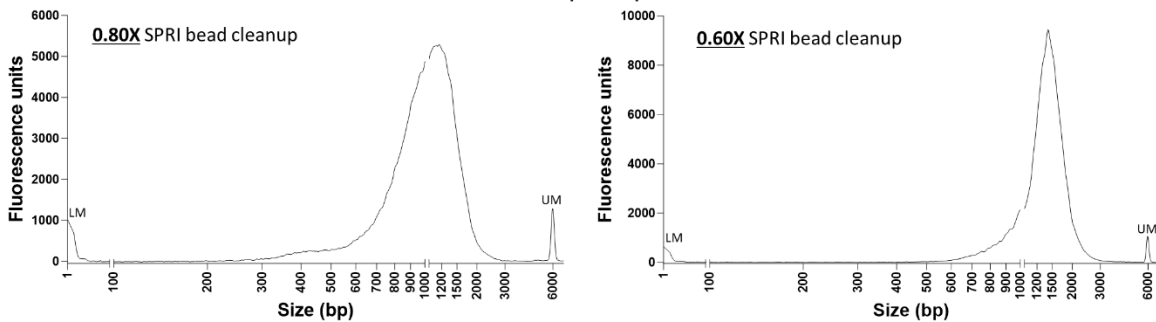
**A** Insufficient removal of free/unbound round 3 barcode oligos and/or too many cells in the library (second strand synthesis product)



**B** Second strand synthesis product



**C** rRNA depletion product

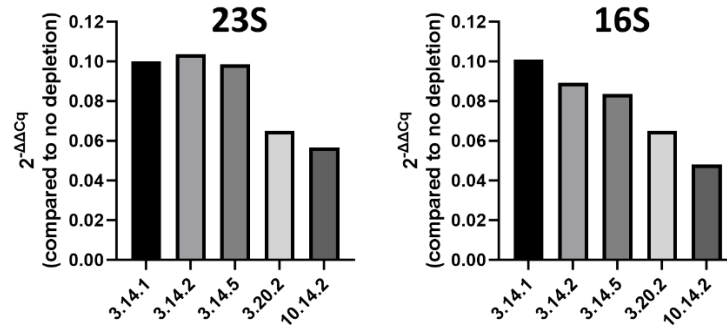


**Supplementary Figure 4 | Experimental considerations during second strand synthesis for generating double-stranded cDNA.** (A) Examples of size distributions from fragment analysis after second strand synthesis if there was insufficient removal of free/unbound round 3 barcode oligos or too many cells in the library. If amplification is stopped early, the size distribution is biased towards smaller fragments, and if amplification is prolonged the size distribution becomes

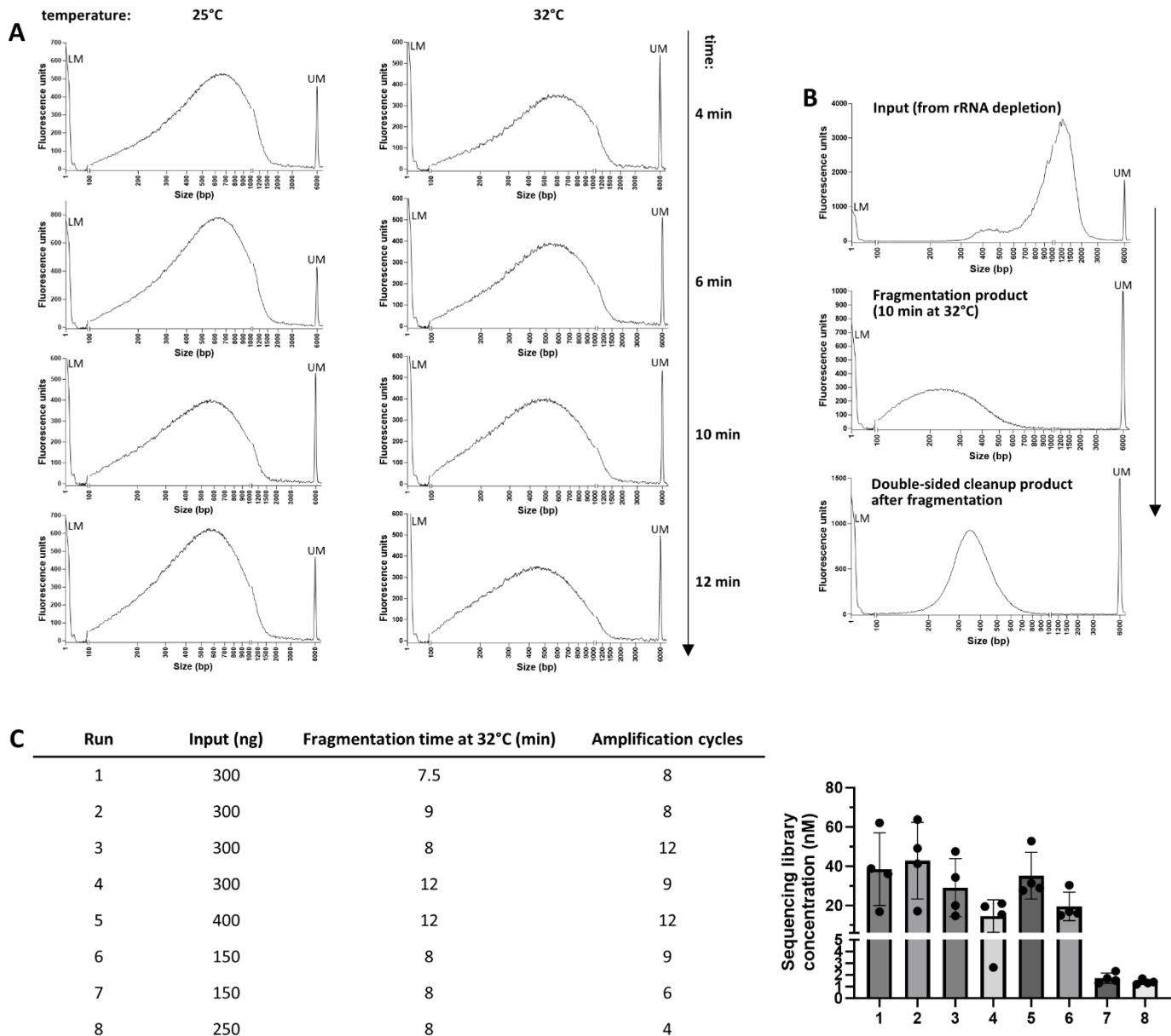


larger and raises the baseline past the upper 6,000 bp marker. (B) Representative size distributions from fragment analysis after second strand synthesis using a 0.80X (left) or 0.60X (right) SPRIselect bead cleanup ratio. (C) Representative size distributions from fragment analysis after rRNA depletion using a 0.80X (left) or 0.60X (right) SPRIselect bead cleanup ratio. (D) Number of cells detected in the sequencing output with > 10 raw reads/cell, normalized to the number detected with a 0.60X SPRIselect bead cleanup ratio used throughout second strand synthesis and rRNA depletion. The larger cleanup ratio allows for > 2X more detected cells on average. Data are presented as mean +/- standard deviation from 4 biological replicates. Source data are provided as a Source Data file.

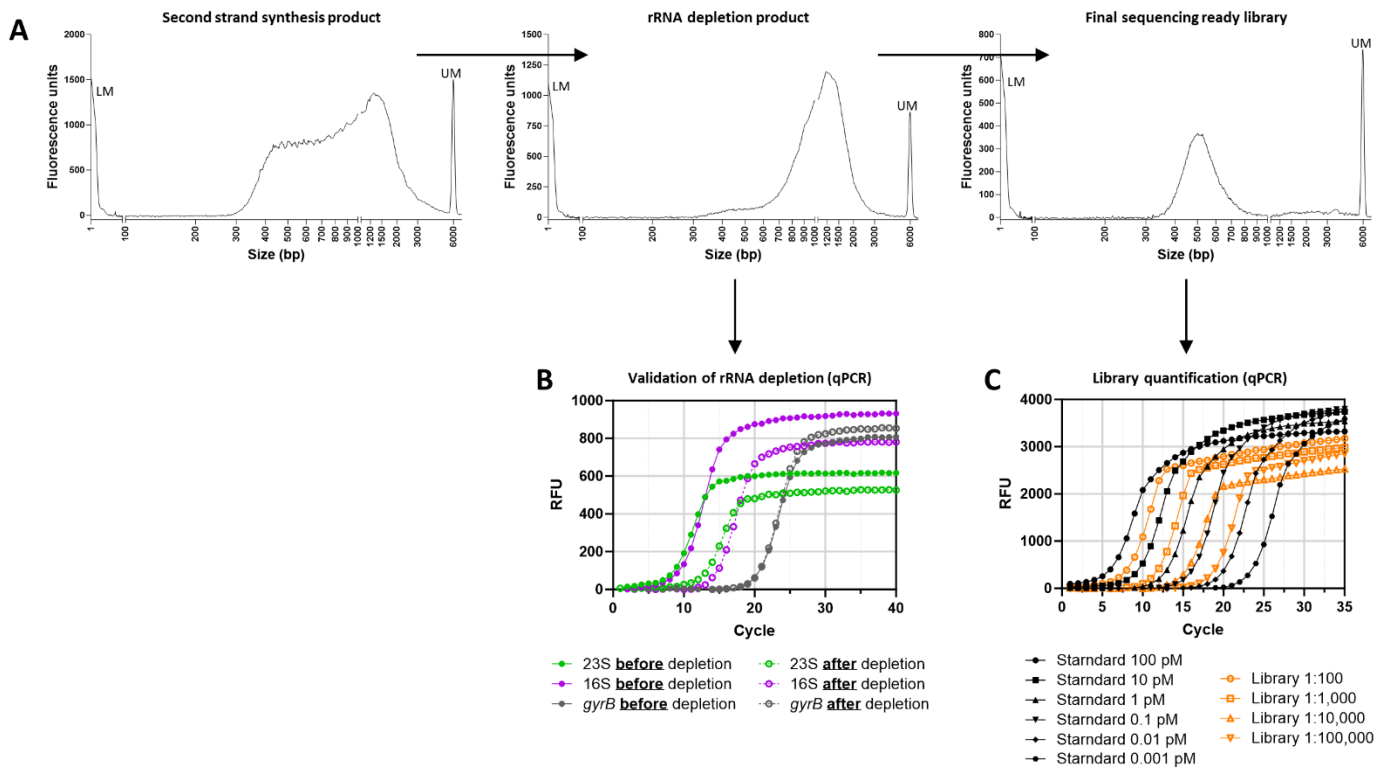
Trial	Oligo:RNA	Bead:Oligo	SSC
3.14.1	3	14	1X
3.14.2	3	14	2X
3.14.5	3	14	5X
3.20.2	3	20	2X
10.14.2	10	14	2X



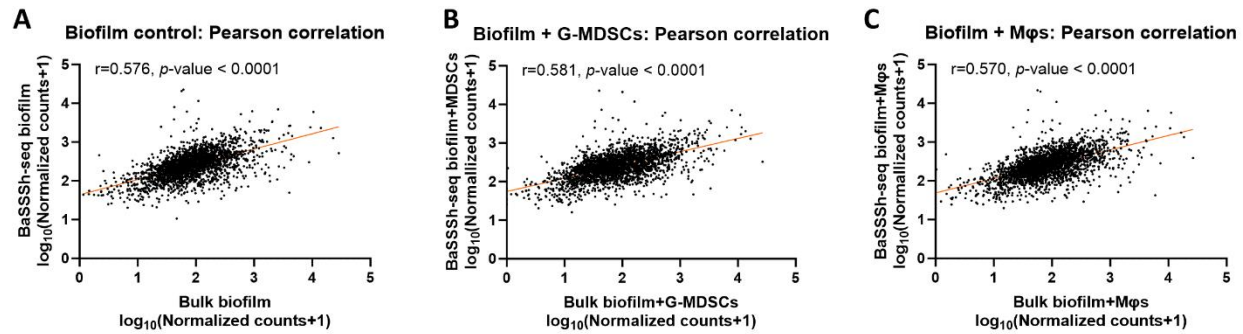
**Supplementary Figure 5 | Experimental considerations during rRNA depletion.** Levels of 23S (left) and 16S (right) rRNA depletion for each of the conditions listed in the table, with results presented from one experiment. Larger ratios for oligo:RNA, bead:oligo, and buffer (SSC) lead to more robust depletion; however, economic considerations must be made as quantities of oligos and streptavidin beads increase accordingly. Source data are provided as a Source Data file.



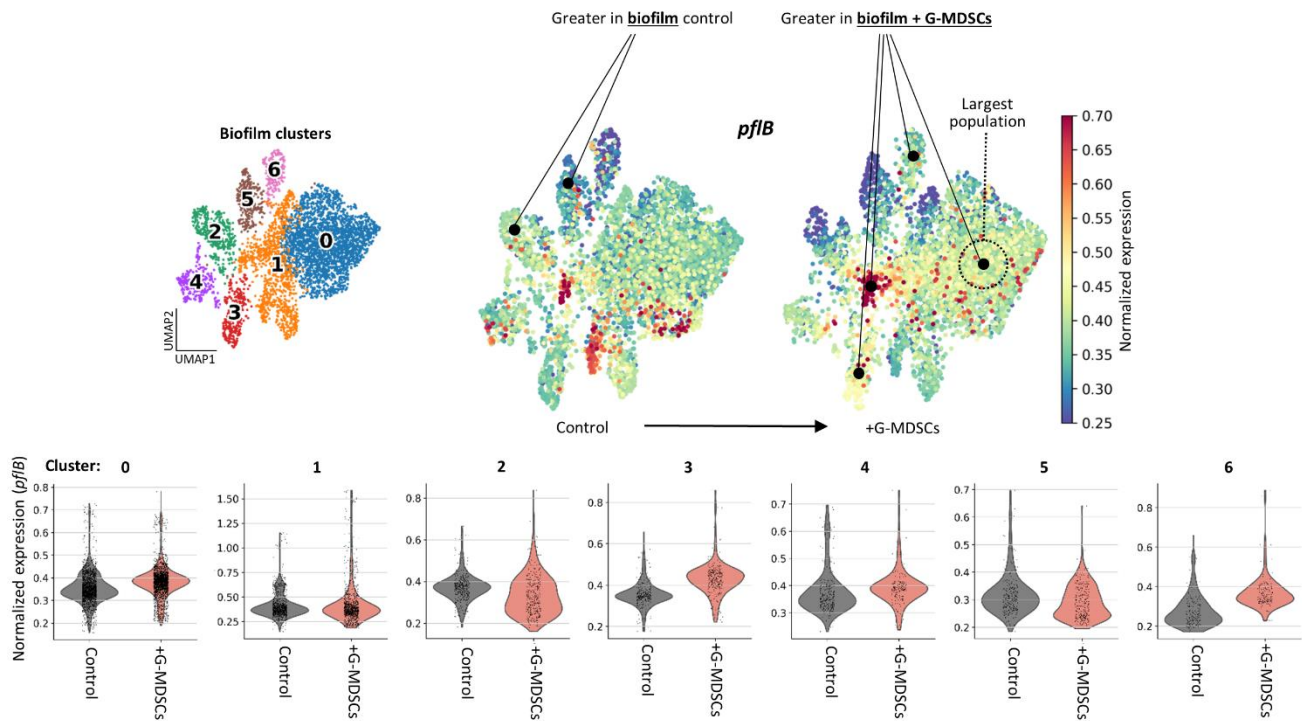
**Supplementary Figure 6 | Experimental considerations during library prep.** (A) Time and temperature of fragmentation can be tuned, with higher temperatures and longer times leading to smaller fragment sizes. (B) Representative examples of size distributions from fragment analysis for the fragmentation step of library prep, including the input (from rRNA depletion) and after double-sided SPRIselect bead cleanup. (C) Representative selection of how metrics for library prep input, fragmentation time, and the number of amplification cycles relate to final sequencing library concentration. Concentrations should be kept near the minimum amount required to load the sequencer to ensure full library diversity is represented, as seen in example runs 7 and 8. Data are presented as mean +/- standard deviation from 4 biological replicates. Source data are provided as a Source Data file.



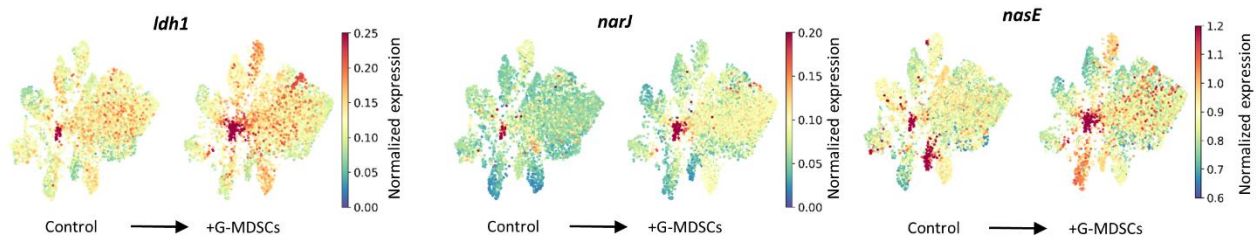
**Supplementary Figure 7 | BaSSSh-seq quality control steps.** (A) Representative size distributions from fragment analysis during each of the quality control checks in the BaSSSh-seq process. Check points occur after second strand synthesis (left), rRNA depletion (middle), and library prep (right). (B) Representative validation of rRNA depletion by qPCR. Sample aliquots from before and after rRNA depletion were analyzed for 23S and 16S rRNA abundance in relation to *gyrB* as a control. Curves were used to calculate  $2^{-\Delta\Delta Cq}$  values for each sample (0.038 and 0.026 for 23S and 16S, respectively in the example). (C) Representative library quantification by qPCR. Libraries were diluted and analyzed in comparison to a standard curve. Source data are provided as a Source Data file.



**D** Gene *pflB*, upregulated following immune cell co-culture in Bertrand et al. *Infection and Immunity*, 2022

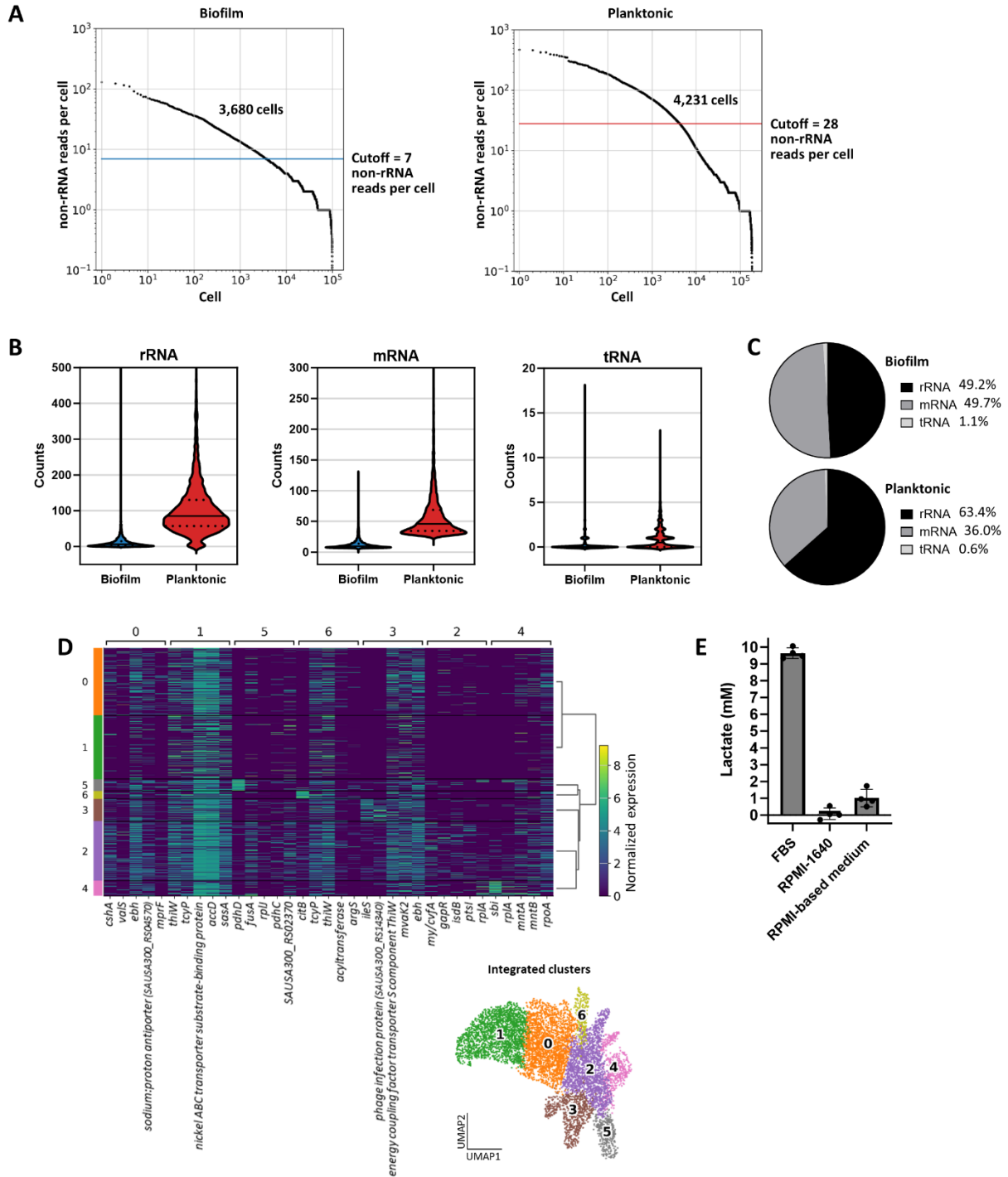


**E** Other genes identified as upregulated following immune cell co-culture in Bertrand et al. *Infection and Immunity*, 2022



**Supplementary Figure 8 | Correlation of BaSSh-seq with bulk RNA-seq, and insights gained from single-cell resolution.** (A, B, C) Transcriptomic profiles from pseudobulk analysis of single-cell RNA-seq with BaSSh-seq significantly correlate with those captured in a prior report (Bertrand et al., doi: 10.1128/iai.00428-22)<sup>9</sup> using bulk RNA-seq for biofilm control (no immune cells) (A), biofilm + G-MDSC co-culture (B), and biofilm + MΦ co-culture (C). Pearson correlations performed with normalized counts per gene. (D) For the case of biofilm + G-MDSC

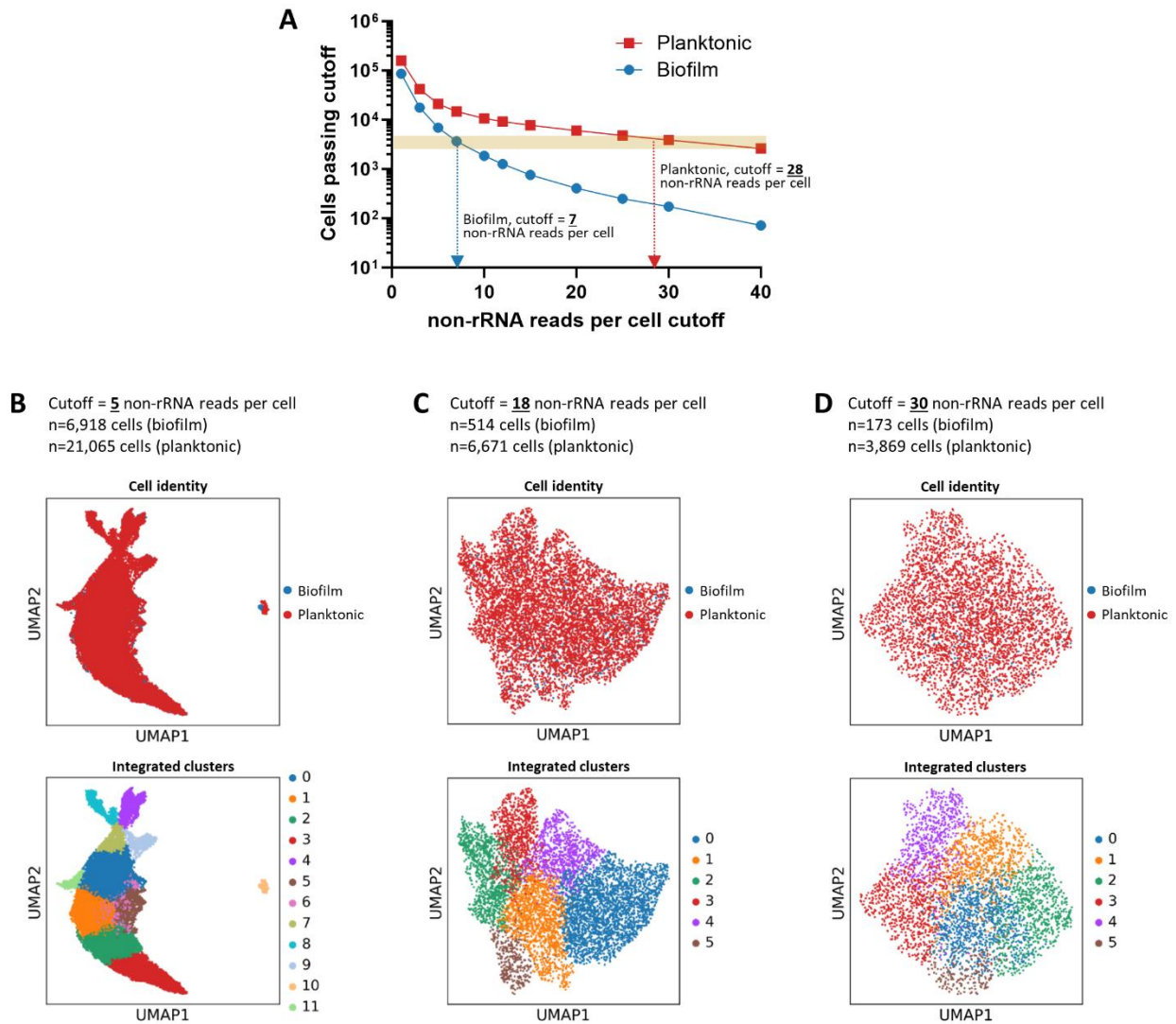
co-culture, Bertrand et al. identified biofilm upregulation of fermentative genes, most prominently *pflB*, that promoted biofilm persistence. Using BaSSSh-seq single-cell datasets for the biofilm control and biofilm co-cultured with G-MDSCs, not all biofilm clusters uniformly upregulate *pflB* in the presence of immune pressure. While many clusters showed *pflB* upregulation, some clusters showed down-regulation. These details are lost in bulk RNA-seq. (E) Similar to *pflB*, other fermentation-related genes found to be upregulated following G-MDSC co-culture with biofilm from the bulk RNA-seq analysis in Bertrand et al. (*ldh1*, *narJ*, *nasE*) are not uniformly upregulated across all biofilm clusters. Source data are provided as a Source Data file.



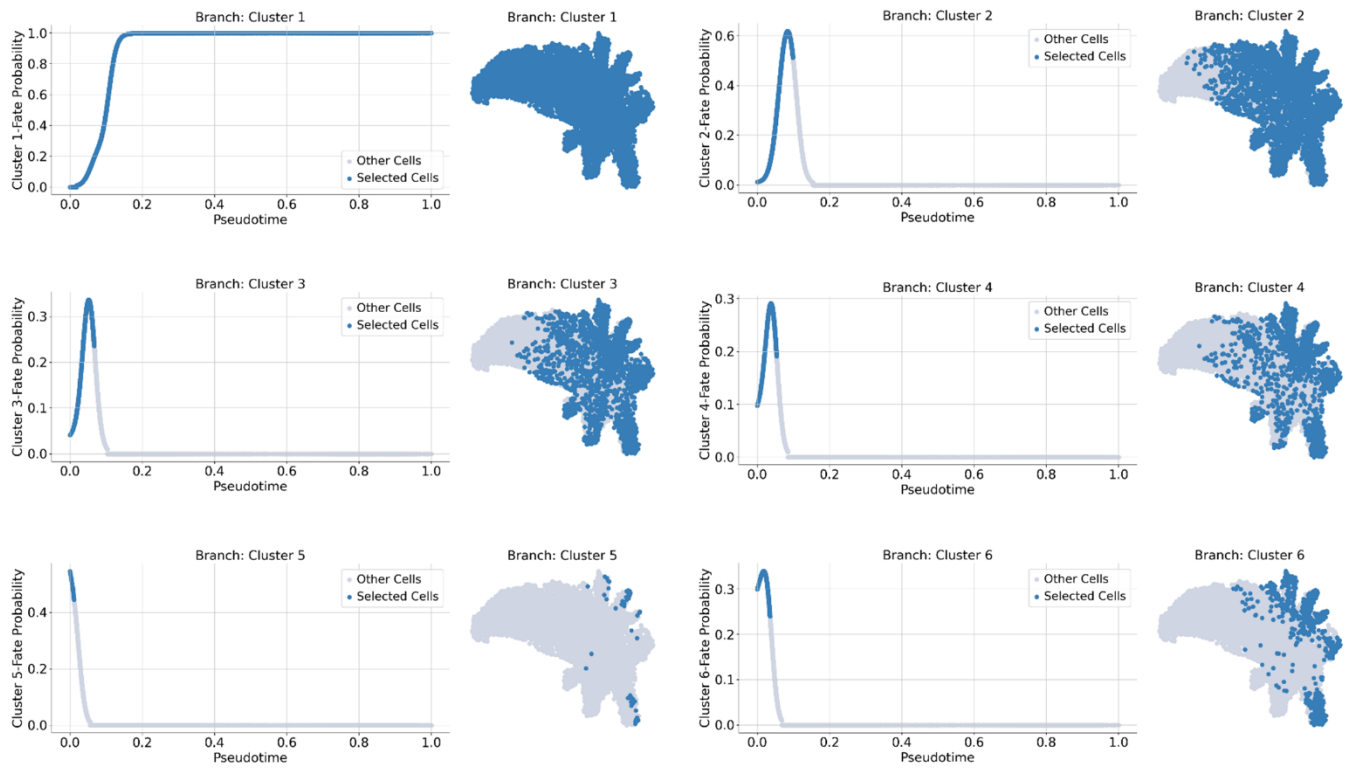
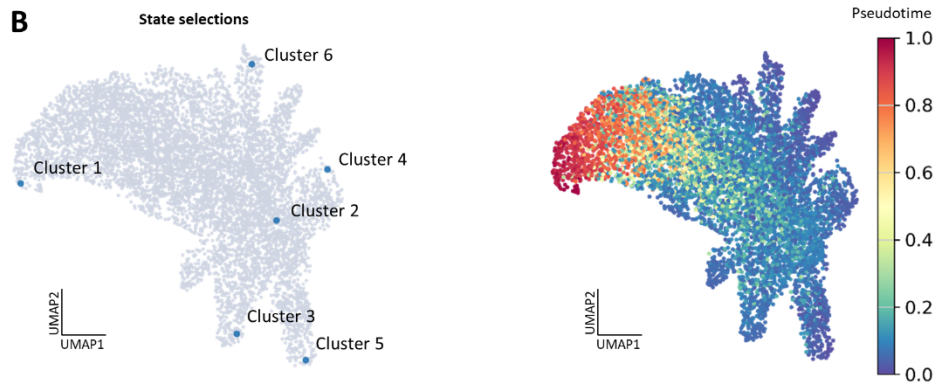
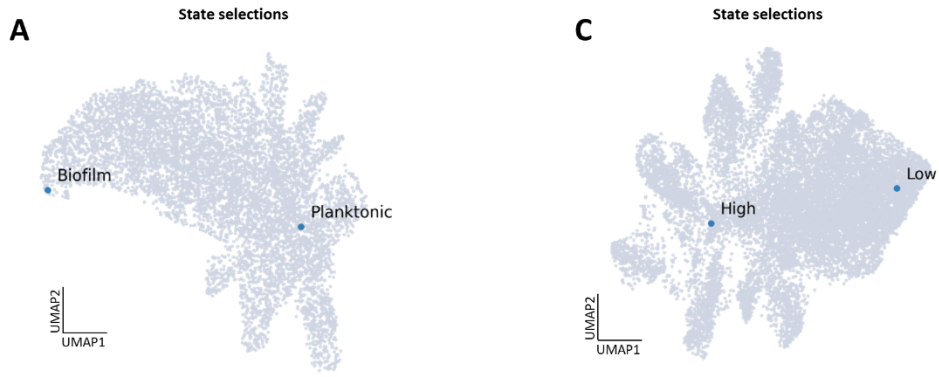
**Supplementary Figure 9 | Filtering, metrics, and clustering when comparing biofilm and planktonic growth with BaSSSh-seq.** (A) Filtering sequenced cells based on the number of non-rRNA (mRNA and tRNA) reads per cell. Cells are sorted in terms of decreasing reads per cell. Numbers of cells carried through for analysis are noted on the plots. (B) Counts of rRNA, mRNA, and tRNA for filtered cells from biofilm and planktonic samples. Solid lines indicate the median, and dotted lines reflect upper and lower quartiles. (C) Percentages of rRNA, mRNA, and tRNA for filtered cells from biofilm and planktonic samples. (D) Marker gene heatmap arranged

by dendrogram relation for the integrated biofilm and planktonic samples, with UMAP on the bottom right (UMAP originally defined in Figure 1C). (E) Lactate levels in fetal bovine serum (FBS), RPMI-1640, and the RPMI-based medium used for *S. aureus* culture containing 10% FBS. Data are presented as mean +/- standard deviation from 4 biological replicates. Source data are provided as a Source Data file.

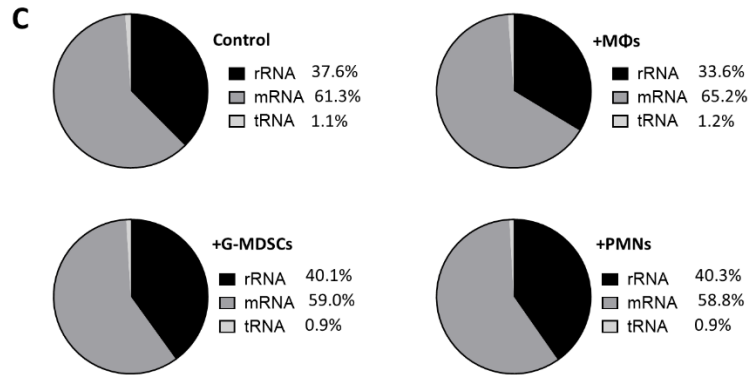
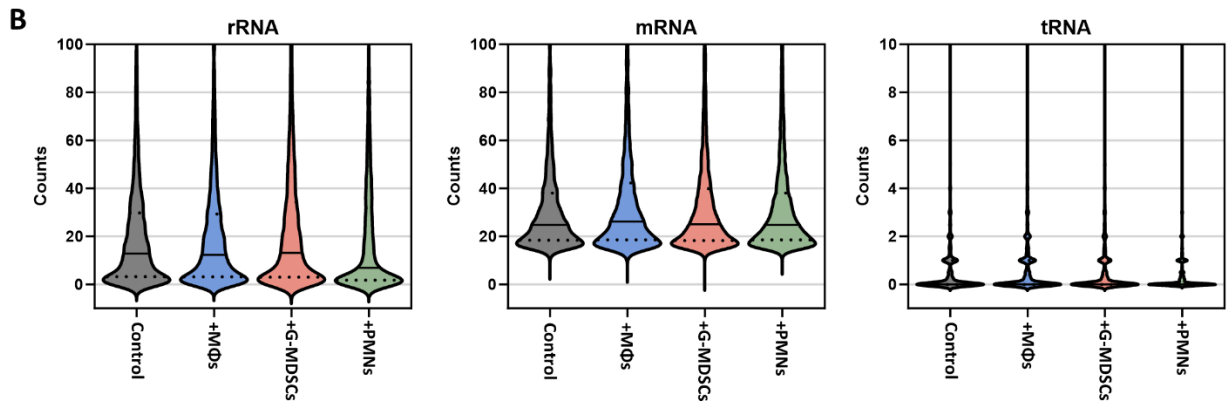
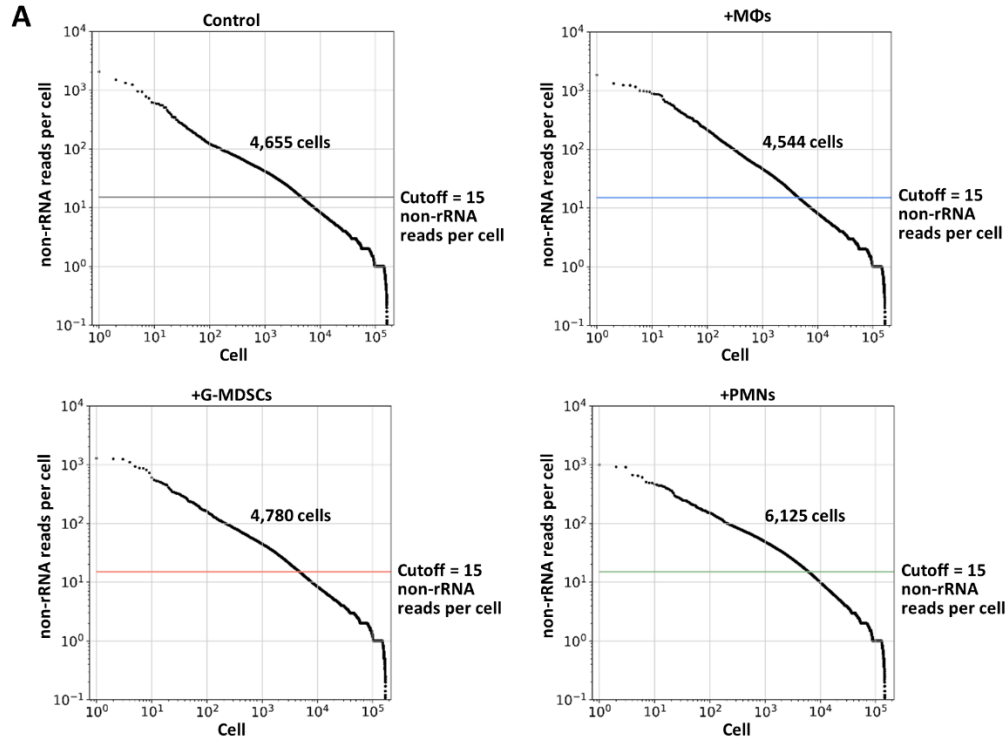




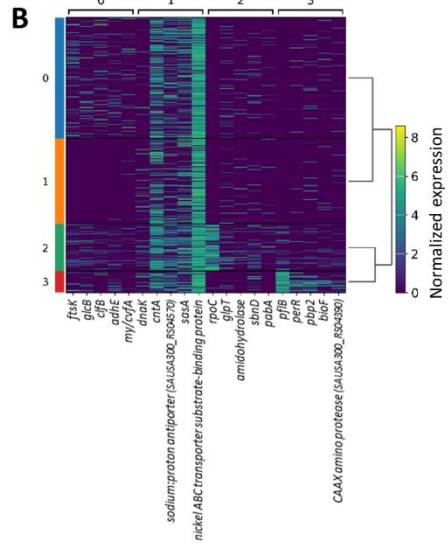
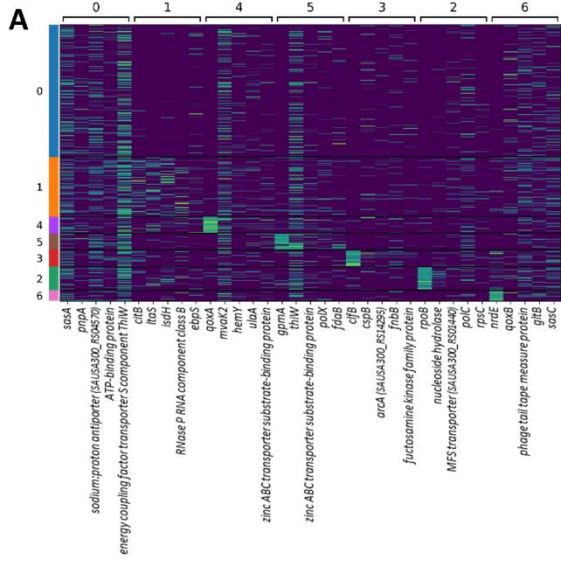
**Supplementary Figure 10 | Determining filter cutoffs for biofilm-planktonic growth comparisons.** (A) Cell numbers passing the non-rRNA reads per cell filter for biofilm and planktonic samples, as the cutoff is increased. Arrows denote where cutoffs were determined for biofilm (7 non-rRNA reads per cell) and planktonic (28 non-rRNA reads per cell) to yield comparable cell numbers. Source data are provided as a Source Data file. (B, C, D) Integrated clustering (performed with the same parameters as the clustering in Figure 2B-C) over a range of three different cutoffs that are equally applied to both the biofilm and planktonic datasets: (B) 5 non-rRNA reads per cell, (C) 18 non-rRNA reads per cell, and (D) 30 non-rRNA reads per cell. When equal cutoffs are applied, the planktonic dataset contains between 3-20X more cells than the biofilm dataset. This is problematic because differential expression analyses performed with unbalanced cell numbers have been shown to skew true positive and false positive identifications.



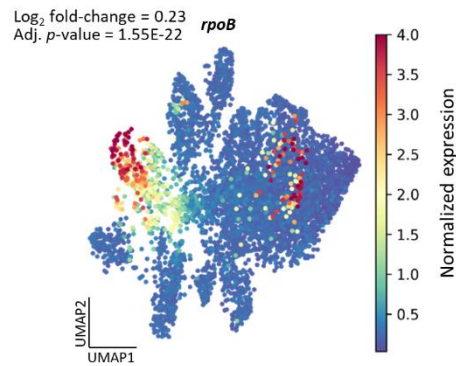
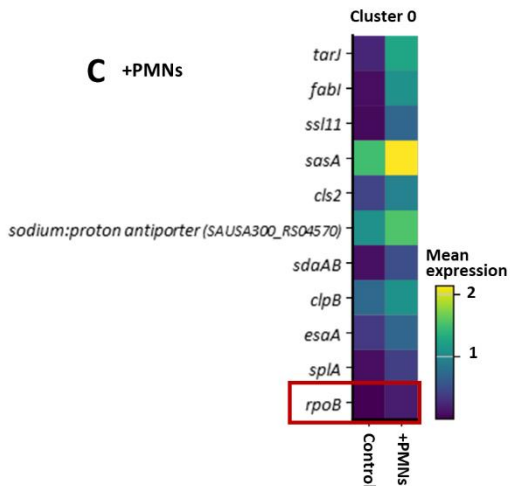
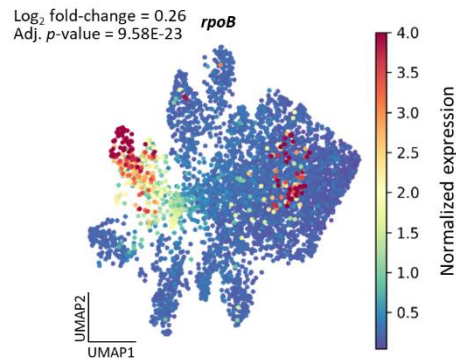
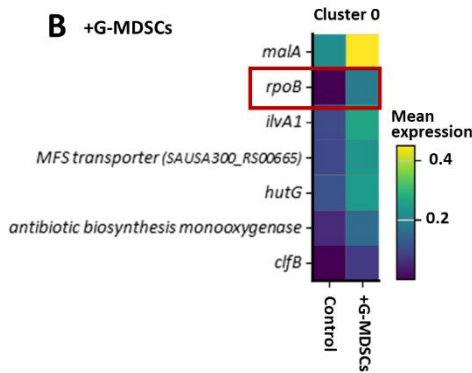
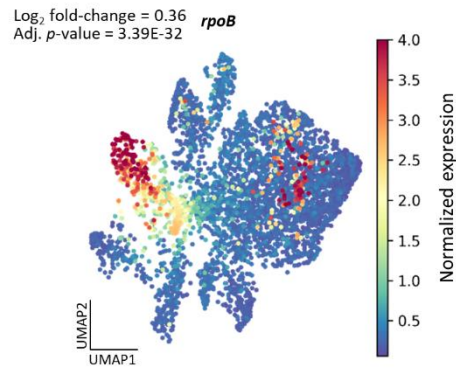
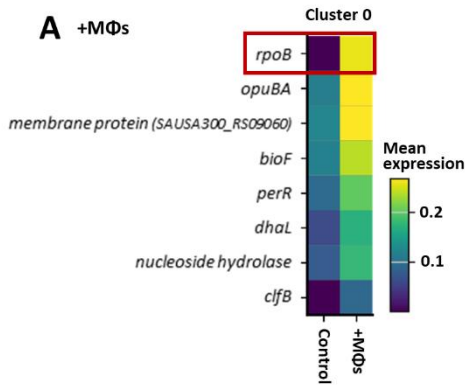
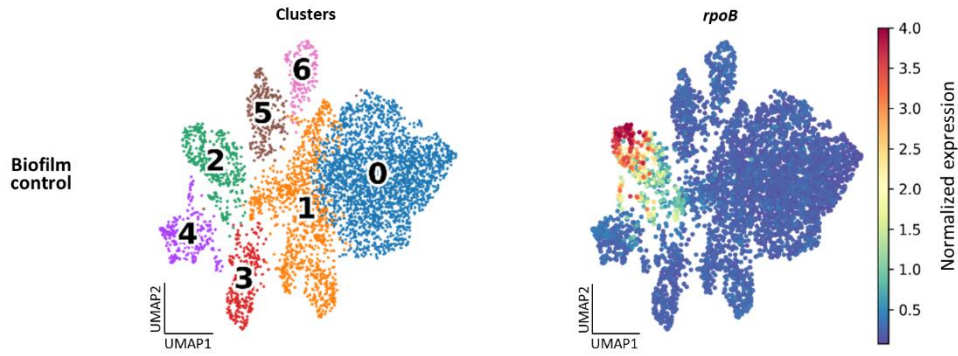
**Supplementary Figure 11 | Trajectory analysis details.** (A) Cell selections denoting the *terminal\_states* parameters in the Palantir<sup>10</sup> trajectory algorithm, for both 'Biofilm' and 'Planktonic' for the integrated biofilm and planktonic samples (trajectory in Figure 3E). (B) Exploring alternative trajectories within the biofilm and planktonic dataset. (Top left) Cell selections denoting the *terminal\_states* parameters in the Palantir trajectory algorithm, for all clusters radiating outward from cluster 2 (clusters 3, 4, 5, and 6) alongside of the previously defined states (clusters 1 and 2 representing core biofilm and planktonic groups, respectively). (Top right) The resulting trajectory over pseudotime, converging towards the biofilm cells in cluster 1, similar to Figure 3E. (Bottom) The alternative clusters (clusters 3, 4, 5, and 6) showed moderate probability as alternative branches of the overall trajectory early in pseudotime, but the analysis converged towards the biofilm cells in cluster 1 later in pseudotime. (C) Cell selections denoting the *terminal\_states* parameters in the Palantir trajectory algorithm, selected to represent 'High' activity and 'Low' activity for the biofilm-leukocyte co-culture samples (trajectory in Figure 6C).



**Supplementary Figure 12 | Filtering and metrics for biofilm-leukocyte co-culture experiments with BaSSSh-seq.** (A) Filtering sequenced cells based on the number of non-rRNA (mRNA and tRNA) reads per cell. Cells are sorted in terms of decreasing reads per cell. Numbers of cells carried through for analysis are noted on the plots. (B) Counts of rRNA, mRNA, and tRNA for filtered cells from co-culture samples. Solid lines indicate the median, and dotted lines reflect upper and lower quartiles. (C) Percentages of rRNA, mRNA, and tRNA for filtered cells from co-culture samples. Source data are provided as a Source Data file.



**Supplementary Figure 13 | Clustering for biofilm-leukocyte co-culture experiments with BaSSSh-seq.** Marker gene heatmaps arranged by dendrogram relation for the biofilm control (A), and independently clustered biofilm + MΦs (B), biofilm + G-MDSCs (C), and biofilm + PMNs (D), with UMAPs below (UMAPs originally defined in Figure 4B-C). For comparative analysis, all co-cultured biofilm cells were projected onto the control UMAP in Figure 6.





**Supplementary Figure 14 | Transcriptional evidence of a potential biofilm persister cell population.** Differential expression was performed using the MAST<sup>11</sup> algorithm for cluster 0 on each of the biofilm-leukocyte co-cultures compared to the biofilm control. Each co-culture condition displayed a set of genes significantly upregulated within cluster 0, a small subset of which are displayed in the matrix plots for +MΦs (A), +G-MDSCs (B), or +PMNs (C). RNA polymerase (*rpoB*) was consistently upregulated across each condition, which is further illustrated in UMAP overlays of *rpoB* expression displayed to the right of each matrix plot, with the biofilm control shown at the top along with cluster identities for reference.

## SUPPLEMENTARY REFERENCES

1. Blattman, S.B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat Microbiol* **5**, 1192-1201 (2020).
2. Homberger, C., Hayward, R., Barquist, L. & Vogel, J. Improved Bacterial Single-Cell RNA-Seq through Automated MATQ-Seq and Cas9-Based Removal of rRNA Reads. *mBio* **14** (2023).
3. Kuchina, A. et al. Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371** (2021).
4. Ma, P. et al. Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell* **186**, 877-891 e814 (2023).
5. McNulty, R. et al. Probe-based bacterial single-cell RNA sequencing predicts toxin regulation. *Nat Microbiol* **8**, 934-945 (2023).
6. Wang, B. et al. Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. *Nat Microbiol* **8**, 1846-1862 (2023).
7. Xu, Z. et al. Droplet-based high-throughput single microbe RNA sequencing by smRandom-seq. *Nat Commun* **14**, 5130 (2023).
8. Kapteyn, J., He, R., McDowell, E.T. & Gang, D.R. Incorporation of non-natural nucleotides into template-switching oligonucleotides reduces background and improves cDNA synthesis from very small RNA samples. *BMC Genomics* **11**, 413 (2010).
9. Bertrand, B.P. et al. Role of *Staphylococcus aureus* Formate Metabolism during Prosthetic Joint Infection. *Infect Immun* **90**, e0042822 (2022).
10. Setty, M. et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol* **37**, 451-460 (2019).
11. Finak, G. et al. MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* **16**, 278 (2015).