

## **Title: Highly Multiplexed Spatial Transcriptomics in Bacteria**

**Authors:** Ari Sarfatis<sup>1,2†</sup>, Yuanyou Wang<sup>1,2†</sup>, Nana Twumasi-Ankrah<sup>1,2</sup>, Jeffrey R. Moffitt<sup>1,2,3\*</sup>

### 5 **Affiliations:**

<sup>1</sup>Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, MA 02115 USA.

<sup>2</sup>Department of Microbiology, Blavatnik Institute, Harvard Medical School, Boston, MA 02115 USA.

10 <sup>3</sup>Broad Institute of Harvard and MIT, Cambridge, MA 02142 USA.

†These authors contributed equally

\*Corresponding author. Email: [jeffrey.moffitt@childrens.harvard.edu](mailto:jeffrey.moffitt@childrens.harvard.edu)

## Abstract:

Single-cell decisions made in complex environments underlie many bacterial phenomena. Image-based transcriptomics approaches offer an avenue to study such behaviors, yet these approaches have been hindered by the massive density of bacterial mRNA. To overcome this challenge, we  
5 combine 1000-fold volumetric expansion with multiplexed error robust fluorescence *in situ* hybridization (MERFISH) to create bacterial-MERFISH. This method enables high-throughput, spatially resolved profiling of thousands of operons within individual bacteria. Using bacterial-MERFISH, we dissect the response of *E. coli* to carbon starvation, systematically map subcellular RNA organization, and chart the adaptation of a gut commensal *B. thetaiotaomicron* to micron-  
10 scale niches in the mammalian colon. We envision bacterial-MERFISH will be broadly applicable to the study of bacterial single-cell heterogeneity in diverse, spatially structured, and native environments.

## Main Text:

### Introduction

Population-level bacterial dynamics often emerge from the heterogeneous behaviors of single cells. Notable examples include entry into and exit from antibiotic persistent states (1), bet hedging (2), virulence factor expression (3), and cellular specialization within biofilms (4). Recent advances in bacterial single-cell RNA-sequencing (scRNA-seq) offer an exciting avenue to study such phenomena by providing transcriptome-wide expression profiles for thousands of cells (5–14). Indeed, such methods have provided insights into antibiotic response (11, 13, 14), prophage activation (9, 11, 14), toxin expression (9, 10), sporulation (10), competence (9, 10), mobile genetic elements (11, 14), cell-cycle-dependent gene regulation (15), and functional heterogeneity within the rumen microbiome (16).

Missing from these studies is the natural spatial context in which many behaviors occur. Yet, spatial organization, across a range of length scales, is an essential modulator of bacterial dynamics. On the tens-of-micron-scale, spatial gradients in small molecule concentrations tune bacterial responses, define niches for commensal growth, or shape interactions in multi-species communities (17, 18). On the micron-scale, direct cell-to-cell contact controls effector protein delivery which mediates predation (19), self- versus non-self recognition (20), contact-dependent inhibition (21), and virulence (22). Finally, even sub-micron length scales are relevant, as growing evidence indicates that the bacterial transcriptome is internally organized with functional consequences (23). Unfortunately, such spatial information is lost during cell dissociation and RNA extraction in scRNA-seq; thus, current methods are not well suited for the study of such processes.

By contrast, image-based approaches to single-cell transcriptomics provide this spatial context by directly imaging and identifying RNAs within fixed cells in their native spatial environment (24–26). Moreover, by leveraging combinatorial optical barcodes to distinguish RNAs, these measurements can be massively multiplexed, producing spatially resolved, transcriptome-scale expression profiles that span intracellular to tissue-scale lengths (24–26). In eukaryotic systems such methods have mapped the intracellular RNA organization, explored regulatory networks, and defined, discovered, and charted cell types and states across a range of tissues (24–26). Unfortunately, current methods are not compatible with bacteria, as the massive density of bacterial RNA challenges combinatorial barcode detection. Non-combinatorial barcoding approaches can bypass this challenge, as recently illustrated with *Pseudomonas aeruginosa* (27), yet are not suitable for transcriptome-wide profiling.

Here we overcome this RNA density challenge and introduce a transcriptome-scale, image-based approach for bacterial, single-cell transcriptomics. This approach combines a bacterially optimized expansion microscopy toolbox with multiplexed error-robust fluorescent *in situ* hybridization (MERFISH) (28) and allows single-cell profiling of up to 80% of the transcriptome. We demonstrate that this technique—bacterial-MERFISH—accurately profiles 97, 1,057, or 1,930 operons with large detection efficiency, accuracy, and throughput in log-phase *Escherichia coli* (*E. coli*) cells. To highlight the discovery potential of this technique, we first profile *E. coli*

response to a carbon-source switch, revealing a heterogeneous sequential nutrient exploration program. Next, we chart the intracellular organization of the *E. coli* transcriptome, uncovering a previously unappreciated diversity in spatial patterning and a cooperative role for genome and proteome organization in shaping transcriptome organization. Finally, we map the adaptation of a human gut commensal—*Bacteroides thetaiotaomicron* (*B. theta*)—to the mouse colon, revealing micron-scale fine-tuning of gene expression based on local polysaccharide availability. More broadly, these measurements illustrate the potential for bacterial-MERFISH to reveal single-cell heterogeneity in a wide range of biological and spatial contexts.

## Results

### 10 ***Bacterial-MERFISH accurately profiles thousands of operons in E. coli***

MERFISH enables the identification of thousands of different mRNA molecules by using combinatorial, error-robust, fluorescent optical barcodes built from repetitive rounds of single-molecule FISH (smFISH) (28). However, to decipher barcodes, the fluorescent signal from different molecules must be optically resolvable. For conventional high-resolution optical microscopy, only a few molecules per  $\mu\text{m}^3$  can be distinguished (29). For eukaryotic systems, large transcriptome fractions can be targeted while satisfying this limit (28). By contrast, a log-phase *E. coli* cell contains  $\sim 8,000$  mRNA molecules in a cell volume of  $\sim 3 \mu\text{m}^3$  (30), producing a total mRNA density nearly three orders of magnitude greater than that resolvable with diffraction-limited imaging (Fig. 1A). Thus, mRNA density restricts the imaging of more than a small number of bacterial mRNAs (31) and is a substantial challenge to transcriptome-scale imaging.

Density reduction is a natural avenue to address this challenge. Indeed, image-based approaches to single-cell transcriptomics have achieved modest degrees of RNA density reduction by spreading the barcode signal over more imaging rounds (32, 33) or by leveraging expansion microscopy (34) to physically swell the sample (32, 35). While the approximately 10-fold reduction in RNA density achieved by these approaches was sufficient to extend image-based transcriptomes to whole-transcriptome-scale in eukaryotes (32, 33), such density reduction is still nearly two orders of magnitude insufficient for similar profiling in bacteria.

To overcome bacterial mRNA density, we leveraged recent advances in expansion microscopy (36, 37) to develop a bacterial-FISH-optimized expansion toolbox capable of up to 1000-fold volumetric expansion (Fig. 1B), complementing recent bacterial expansion methods developed for non-RNA targets and with modest degrees of expansion (38–43). Briefly, we grew *E. coli* to mid-log phase, fixed them with paraformaldehyde (PFA), digested the cell wall, expanded them in a Ten-fold Robust Expansion (TREx) gel (36), and then re-embedded the sample in a non-expanding, stabilizing gel (Fig. 1B; Materials and Methods). Using custom expansion-optimized staining protocols, we labeled samples with a 16S ribosomal RNA (rRNA) probe and a MERFISH probe set targeting 97 *E. coli* operons (Materials and Methods). We targeted operons rather than individual genes as the signals from different barcodes from genes on the same polycistronic mRNA would not be resolvable.

Expansion increased the width of *E. coli*  $3.9 \pm 0.9$ -fold (standard deviation [STD],  $n=4$ ; fig. S1A) consistent with a  $\sim 50$ -fold volumetric expansion (fig. S1B). Thus, we term this approach the 50X

expansion protocol. In expanded cells stained with the 97-operon library, individual fluorescent puncta were visible, and these molecules were identified with MERFISH (Fig. 1, C to H), indicating sufficient expansion to resolve the mRNA density of this targeted library. Interestingly, we noted ample number of molecules even in the absence of specific RNA-gel anchoring chemistries in PFA-fixed but not methanol-fixed expanded samples, suggesting a PFA-dependent, RNA-anchoring mechanism (fig. S1, C to I).

To further explore the multiplexing possible with the 50X protocol, we stained 50X-expanded *E. coli* with a MERFISH probe set targeting 1,057 operons, roughly 40% of the *E. coli* transcriptome (Materials and Methods). While the mRNA density was higher than that observed for the 97-operon measurement, individual puncta were still resolved and many molecules were identifiable with MERFISH (fig. S1, J to L). Nonetheless, we noticed an increased frequency of overlapping RNA signals (fig. S1K). To address this overlap, we developed an iterative expansion protocol that combined TREx (36) with a previous iterative strategy (37). Briefly, cells expanded and stabilized with the 50X protocol were expanded in a second TREx gel and then embedded in a second stabilizing gel (Fig. 1B; Materials and Methods). This protocol expanded the width of cells 11.1±1.7-fold (STD,  $n=4$ ; fig. S1A), consistent with a ~1,300-fold volumetric expansion (fig. S1B). Thus, we term this protocol the 1000X protocol. When cells were expanded with this protocol, the signal from individual RNAs and their identity were clearly distinguished when 1,057 operons were stained (Fig. 1, I to K). Notably, the 1000X protocol starts with 50X-expanded samples (Fig. 1B), facilitating exploration of the necessary expansion for a given sample.

Inspired by the ability to expand *E. coli* volumetrically by three orders of magnitude, we designed a MERFISH library that covers 80% of the transcriptome, corresponding to 1,930 operons (Materials and Methods). 1000X-expanded samples stained with this probe set showed clear single-molecule signals with limited overlap, and these RNAs could be identified with MERFISH (fig. S1, M to O), suggesting that 1000X expansion is sufficient to allow MERFISH profiling of a substantial fraction of the *E. coli* transcriptome.

Notably, during the early development of bacterial-MERFISH, we observed a low correlation between the abundance determined via bacterial-MERFISH and that of bulk RNA-sequencing for lowly expressed operons, suggesting a higher false-positive rate than that measured by our internal false-positive controls (Fig. 1L; Materials and Methods). This discrepancy suggested a bacterial-expansion-dependent source of false positives, which we reasoned might be due to probe binding to genomic DNA melted during expansion. Supporting this hypothesis, DNase treatment dramatically reduced this apparent false-positive rate, bringing it into agreement with that measured with internal controls (Fig. 1M).

To benchmark the performance of bacterial-MERFISH, we performed two replicate measurements of 97, 1,057, or 1,930 operons in combination with 50X or 1000X expansion in log-phase *E. coli*, segmented cells from these images, and partitioned RNAs into those cells (Fig. 1, F to K, and fig. S1, J to O). We observed strong correlation between these measurements and bulk RNA-sequencing across all conditions (Fig. 1, M to O, and fig. S2, A and B), indicating that bacterial-MERFISH can accurately profile RNA expression across four orders of magnitude in abundance.

Further supporting its accuracy, measurements with bacterial-MERFISH correlated strongly between biological replicates, expansion protocols, and multiplexing levels (fig. S2, C to E).

5 These measurements also revealed that the detection efficiency—the fraction of targeted molecules actually detected—of bacterial-MERFISH is comparable to or greater than the capture efficiencies reported for scRNA-seq (5–14) (Fig. 1P and fig. S2F; Materials and Methods). Values ranged from 7% for near-whole transcriptome measurements (1,930 operons) to as high as 50% for more  
10 targeted measurements (97 operons), with additional expansion providing an increase from 10% to 25% for 1,057 operons measurements (Fig. 1P). As expected, the number of mRNA counts and unique operons observed per cell varied based on the multiplexing and detection efficiency (Fig. 1, Q and R). Finally, bacterial-MERFISH can image large numbers of cells, comparable or greater  
15 than those characterized previously with scRNA-seq (5–16), despite the decreased imaging throughput due to expansion (fig. S2G). Importantly, with the ability to vary multiplexing and expansion, it is possible to balance different aspects of performance, e.g., the number of imaged cells versus the degree of expansion, to best suit the question. Collectively, these measurements indicate that bacterial-MERFISH is a high-performance, versatile image-based approach to bacterial single-cell transcriptomics, with low false-positive rates and large dynamic range, detection efficiencies, and throughput.

### 20 ***Bacterial-MERFISH reveals an asynchronous hierarchical response during a shift in carbon source***

One promise of single-cell bacterial transcriptomics is the ability to identify heterogeneous bacterial behaviors and to computationally resynchronize the desynchronized response of individual cells to environmental stimuli, unmasking dynamics obscured by population averages. One illustrative dynamic response is the diauxic shift caused by a switch in carbon source. Bacteria  
25 often consume different carbon sources in a preferential order regulated, in part, by carbon-catabolite repression (CCR) (44). A bacterial population grown in a mixture of two carbon sources will often consume the preferred source, pause growth in a diauxic shift, and then resume growth on the less preferred substrate (45). The diauxic shift is classically interpreted as the time required to express utilization machinery for the second sugar; however, recent single-cell studies suggest  
30 that this response is instead shaped by differential dynamics of sub-populations (46). Nonetheless, the diversity and transcriptional profiles of such sub-populations remain poorly defined.

Thus, we revisited the classic diauxic shift with bacterial-MERFISH. We grew *E. coli* in a minimal defined medium with a mixture of glucose and xylose (Materials and Methods). As expected, the culture grew logarithmically until glucose was exhausted, paused growth in a diauxic shift, and  
35 then resumed logarithmic growth on xylose before entering stationary phase once both sugars were exhausted (Fig. 2A). We harvested *E. coli* cells throughout this process and profiled the expression of 1,057 operons (Fig. 2, A to C) in 296,666 cells with 50X expansion—a multiplexing and expansion level set to balance transcriptome-wide profiling with the number of profiled cells. For experimental efficiency, multiple time points were combined and profiled in a single MERFISH  
40 measurement by labeling cells with barcoded 16S rRNA probes (27) optimized for expansion



protocols (Fig. 2B; Materials and Methods). Putative cells were segmented, RNA molecules assigned to these cells, and conditions demultiplexed with the 16S rRNA barcode (Fig. 2C; Materials and Methods).

Supporting these measurements, we found that the average RNA expression determined via MERFISH was strongly correlated between biological replicates and with bulk RNA-sequencing at different growth stages (fig. S3, A to D). Moreover, cell populations displayed expected expression patterns (Fig. 2D) with cells harvested during growth expressing operons associated with amino acid synthesis (e.g., *argG* and *ilvC*), translation (e.g., *tff-rpsB-tsf* and *cmk-rpsA-ihfB*), and aerobic respiration (e.g., *atpIBEFHAGDC*), and cells harvested during the diauxic shift and stationary phase expressing operons associated with stress response (e.g., *nlpD-rpoS*) and gluconeogenesis (e.g., *glpFKX* and *glpD*).

To explore the heterogeneity in cellular response to this carbon shift, we integrated the measurements from two biological replicates, visualized single-cell heterogeneity with Uniform Manifold Approximation and Projection (UMAP), and performed Leiden clustering to distinguish sub-populations (Fig. 2, E to I; Materials and Methods). Supporting our analysis, cells harvested at similar growth phases largely co-integrated (fig. S3E), yet, despite modest transcriptional differences, cells collected from log-phase growth in glucose or xylose were largely resolved (Fig. 2E and fig. S3F). This analysis revealed a rich diversity in behavior at the single-cell level. Cells were organized into two major groups, corresponding to cells taken from conditions of growth or non-growth (Fig. 2E), and were collectively sub-divided into 14 different clusters (Fig. 2F). These clusters had unique gene expression profiles (Fig. 2, G and H) and abundances across different conditions (Fig. 2I). Supporting cluster validity, individual clusters were observed across both replicates (Fig. 2I) and were each marked by multiple operons associated with related biological processes (Fig. 2H). Moreover, single-molecule FISH (smFISH) in unexpanded *E. coli* confirmed the co-expression of cluster markers (fig. S4).

During log-phase growth, we identified a diversity of sub-populations (Fig. 2, E to I), including clusters associated with translation and amino acid transport (G0 marked by *tff-rpsB-tsf* and *lysP*), nucleobase synthesis (G6 marked by *carAB*, *codBA*, and *purHD*), arginine synthesis (G7 marked by *argG*, *argD*, and *argCBH*), serine biogenesis (G4 marked by *serA* and *serC-aroA*), sulfate utilization (G11 marked by *cysDNC* and *cysJIH*), protein folding (G12 marked by *dnaKJ*, *groSL*, *hslVU*, *clpB*, and *htpG*), and motility and chemotaxis (G3 marked by *motRAB-cheAW*, *fliDST*, and *tar-tap-cheRBYZ*) (Fig. 2H). Intriguingly, many of these biochemical processes are required for growth in minimal media, yet our analysis revealed that only subsets of cells expressed high levels of these essential operons, suggesting a model in which the homeostatic, population-level expression of these pathways is produced not by uniform expression across all cells but rather by transient, coherent bursts of expression that time-average to required levels. This observation is consistent with transient bursts in promoter activity revealed with live-cell imaging (47, 48) and the presence of similar clusters in recent scRNA-seq measurements (10). Our measurements now suggest that such bursts are widespread and occur not at the individual operon level but rather upstream in common regulatory factors.

We also observed a similar degree of heterogeneity during the diauxic shift (Fig. 2, E to I). Interestingly, only a subset of clusters expressed operons associated with xylose utilization (N2 and N10 both marked by *xylAB*, *xylE*, *xylFGHR*, and *yagGH*) while most clusters in the diauxic shift expressed utilization operons associated with carbon sources not included in our medium (Fig. 2, E to I). These sources comprised mannose and glycerol (N1 marked by *manXYZ*, *glpABC*, and *glpTQ*); maltose (N9 marked by *malEFGH*, *malKM-lamB*, and *malPQ*); arabinose (N10 marked by *araE-ygeA*, *araFGH*, and *araBAD* in addition to xylose-associated operons); mannitol and glycolate (N8 marked by *mtLADR* and *glcDEFGBA*); and acetate and putrescine (N5 marked by *acs-yjcH-actP*, *puuAP-ymjE*, and *puuDRCBE*) (Fig. 2, H and I). N13, like G3, was marked by chemotaxis and motility operons. However, these clusters were distinguished by operons associated with chemotaxis towards peptides (G3; e.g., *tar-tap-cheRBYZ* and *tsr*) or sugars (N13; e.g., *trg* and *aer*), reflecting the differential nutrient needs between the conditions associated with these clusters (Fig. 2H).

We also noted an apparent temporal order in which specific carbon sources were explored. To investigate this ordering, we computationally resynchronized cells using a pseudotime analysis from glucose-log-phase growth through the diauxic shift (Materials and Methods). Not only did this analysis reproduce the order of sampling during the diauxic shift (Fig. 2, J and K), supporting the pseudotime ordering, it also revealed a temporal cascade of carbon-utilization operon expression (Fig. 2L). This cascade started with operons associated with glucose (e.g., *ptsG*) and then proceeded through operons associated with glycogen (e.g., *segA-pgm*); the glyoxylate cycle—related to acetate utilization—(e.g., *pdhR-aceEF-lpd* and *aceBAK*); mannose (e.g., *manXYZ*); galactitol and galactose (e.g., *gatYZABCD* and *mglBAC*); glycerol (e.g., *glpFKX*); maltose and fucose (e.g., *malT*, *malEFGH*, *malKM-lamB*, and *fucAO*); xylose (e.g., *xylFGHR* and *xylAB*); and then, finally, arabinose (e.g., *araFGH* and *araBAD*). The transient co-expression of xylose- and arabinose-utilization operons seen at late pseudotime values (Fig. 2L) may represent an overshoot in carbon-source progression due to the lag between transcription of xylose-utilization operons and the transition to the functional utilization of xylose. Importantly, the range of pseudotime values observed for each shift condition overlapped with those of other conditions (Fig. 2M), supporting a stochastic response model in which cells are desynchronized in their progression along this carbon-source hierarchy. Notably, this model does not assume that all cells explore all carbon sources.

Together these results reveal that the homeostatic levels of required pathways can be maintained by transient, coherent bursts of expression within entire regulons, and that, when faced with carbon starvation, *E. coli* adopts a responsive diversification strategy (49) in which the lack of glucose triggers a stochastic hierarchical progression along sugar utilization operons. Importantly, with the ability to characterize sub-populations and computationally resynchronize cells undergoing dynamic responses, bacterial-MERFISH could prove useful in dissecting the diverse regulatory mechanisms that encode such cellular heterogeneity.

**40 *Bacterial-MERFISH reveals a diversity of intracellular localization patterns for E. coli mRNAs***



Another advantage of image-based, single-cell transcriptomic methods is the ability to measure the intracellular transcriptome organization. Specific bacterial mRNAs are localized to the cytoplasm (50–52), membrane (50–52), poles (50, 52–55), chromosomal loci of origin (56), or the surface of intracellular organelles (57) as determined via low throughput mRNA imaging (50–57) or biochemical fractionation (50, 52). This organization has proposed functional roles in protein sorting, complex assembly, and mRNA turnover (23). However, as RNA localization has yet to be imaged at the transcriptome-scale, the extent and diversity of spatial organization remain unclear.

To explore the spatial organization of the *E. coli* transcriptome, we leveraged our 50X-expanded, 1,057-operon measurements of *E. coli* grown in LB. To define intracellular RNA localization, we mapped each mRNA molecule to its axial and radial position within each cell and normalized these coordinates to the cell length and width (Fig. 3A; Materials and Methods). We then computed the average axial and radial distributions for each mRNA across all measured cells (Fig. 3, B to D; Materials and Methods). Consistent with previous reports in *E. coli* (23), we identified mRNAs enriched in the cytosol (e.g., *dnaKJ*), at the membrane (e.g., *ptsG*), and toward the poles (e.g., *dnaB*) (Fig. 3, C and D). However, we also noticed a striking diversity of variations on these major patterns, with mRNAs enriched at different locations along the membrane, at multiple cytoplasmic locations, or in a single central focus (Fig. 3E). These patterns were reproduced between replicate measurements (fig. S5, A to C), with 1000X-expanded MERFISH measurements (fig. S5D), and with unexpanded smFISH (Fig. 3F).

To further categorize this spatial diversity, we leveraged a measure of pattern similarity to visualize and cluster spatial patterns (Materials and Methods). This analysis produced five major clusters of mRNA distributions (Fig. 3G) which showed some degree of continuous spatial variation within the clusters. The Cytoplasmic cluster comprised patterns including uniform filling of the cytoplasm (e.g., *dnaKJ* and *glnS*), multiple foci throughout the cell (e.g., *uxuAB* and *groSL*), or increased central density (e.g., *metG*). mRNAs with a strong central focus (e.g., *adhE* and *rnb*) defined the Midcell-cytosolic cluster. The Membrane cluster was defined by diffuse membrane-enriched patterns (e.g., *ptsG* and *cydAB*) or membrane enrichment at or adjacent to the poles (e.g., *yifK*, *kgtP* or *yojI*) while the Midcell-membrane cluster was defined by strong membrane enrichment only in the middle of the cell (e.g., *dtpA* and *pntAB*). Finally, the Polarized cluster was defined by diffuse (e.g., *mreBCD*) or sharp cytoplasmic foci (e.g., *metBL*) near the poles. Importantly, some mRNAs at cluster boundaries shared features similar to nearby clusters (e.g., *proS* and *proP*), underscoring a continuous variation in spatial patterns.

To investigate possible patterning mechanisms, we explored the correlation between clusters and mRNA features. Covariation with spatial distribution was modest for transcript length, GC content, abundance, and half-life (fig. S6, A to H) but strong for encoded protein location (Fig. 3, H to J, and fig. S6, I to N). Specifically, operons containing mRNAs that encode at least one inner-membrane protein were enriched in membrane-associated RNA localization clusters, whereas mRNAs that encode cytoplasmic proteins were enriched in clusters found within the cytosol (Fig. 3J). These observations are consistent with previous measurements for individual mRNAs (50) as well as transcriptome-wide mRNA groups (51); with the co-translational insertion mechanism of inner-membrane proteins, which would concentrate mRNAs at the membrane during translation

(58); and with the report of membrane-associated sequence features enriched in inner-membrane-protein-coding mRNAs (59, 60).

In parallel, the genomic locus from which the mRNA was transcribed also correlated strongly with spatial patterning. mRNAs within the Midcell-cytosolic or Midcell-membrane clusters were preferentially encoded from genomic loci near the terminus of replication (*terC*) while the genes for mRNAs in other clusters were depleted in this chromosomal region (Fig. 3K and fig. S6O). During conditions of fast growth, the *E. coli* chromosome is organized with *terC* in the cell center and *oriC* and the left and right chromosomal arms replicating near the poles. This structure is maintained for the majority of the cell cycle, with the polar *terC* macrodomain formed at the pole of a newly divided cell moving rapidly to the cell center (61). Indeed, the axial distribution of mRNAs averaged across all (Fig. 3, L and M) or portions (fig. S6P) of the cell cycle were consistent with this organization.

Our measurements now unify two previous models suggested for bacterial transcriptome organization—that either genomic (56) or proteomic (50–52) features dictate organization. Specifically, we show that both features play a role in the global organization of mRNAs, with protein location shaping cytoplasmic versus membrane enrichment and genomic feature shaping axial enrichment both within the cytoplasm or on the membrane. Nonetheless, we identified multiple mRNAs with spatial patterns that deviate from these global rules, including membrane enriched mRNAs that do not encode known inner-membrane proteins (fig. S7, A and B), cytoplasmic-enriched mRNAs that encode an inner-membrane protein (fig. S7, C and D) and many mRNAs with spatial patterns inconsistent with those predicted for their genomic loci (fig. S7, E to G). These exceptions raise the possibility that there are mRNA-specific localization mechanisms that remain to be discovered and that these exceptions, in addition to global patterns, might have functional significance. Importantly, bacterial-MERFISH offers a direct approach to measuring such patterns, which should greatly enable mechanistic and functional studies of the intracellular organization of the bacterial transcriptome.

### ***Bacterial-MERFISH reveals adaptation of a gut commensal to distinct niches in the colon***

Image-based approaches to single-cell transcriptomics also promise the ability to explore gene expression within complex, spatially structured environments. To explore this capability of bacterial-MERFISH, we leveraged germ-free mice monocolonized with the human gut commensal *B. theta* (Fig. 4A). As *Bacteroides* have the ability to harvest a remarkable diversity of polysaccharides—both from the rich dietary pool as well as those deposited onto the mucus layer by the host (62)—we reasoned that *B. theta* might modulate its polysaccharide utilization based on local polysaccharide availability, as suggested previously (63). To explore this possibility, we designed a MERFISH panel that covers the diverse importers (i.e., SusC proteins) within polysaccharide utilization loci (PUL), hybrid two-component systems (HTCS) often associated with the regulation of PUL expression, and a handful of genes associated with central metabolism and other bacterial functions, targeting 159 operons in total (Fig. 4B).

We then resected the colon from a monocolonized mouse, and the sample was methacarn-fixed, paraffin-embedded, and sectioned (Fig. 4A; Materials and Methods). In an unexpanded section, 16S rRNA staining revealed depletion of *B. theta* from the sterile inner mucus layer next to the host epithelial layer, enrichment near the outer mucus layer (64), and variable density throughout the colonic lumen (Fig. 4, A and C). We then expanded a section with a modified 50X-protocol and stained it for MERFISH. The 16S rRNA signal revealed that the distribution of *B. theta* was largely preserved despite a notable loss in dietary debris (Fig. 4, C and D; Materials and Methods). Critically, individual mRNAs could be identified by MERFISH in single *B. theta* cells (Fig. 4, E and F). Across this colonic cross section, we observed that the expression of many *B. theta* operons was spatially variable. Some mRNAs were expressed throughout the lumen (e.g., BT\_0364 and BT\_4671) while others were more prominently expressed near the mucus layer (e.g., BT\_3958 and BT\_2894) (Fig. 4G). Multi-color smFISH in unexpanded samples supported these expression patterns (fig. S8).

To quantify this spatial variation, we performed these measurements in six slices sampled from the colon of two mice. Each colon was fixed with one of two different methods to control for potential fixation-dependent artifacts (Materials and Methods), which were minor as evidenced by strong correlation between these measurements (fig. S9, A to C). As PUL expression can be low (fig. S8), we averaged gene expression over small spatial patches containing  $\sim 5$  *B. theta* cells and visualized transcriptional variation across patches using UMAP and diffusion maps (Fig. 4H and fig. S9, D to F). Individual genes expressed more prominently near the mucus layer or throughout the lumen defined different regions of this UMAP (fig. S9, D and E), and the diffusion analysis suggested that an important axis of variation, captured by the first diffusion component (DC1), is related to mucus proximity. Patches of low DC1 values were found enriched near the mucus whereas patches of high DC1 values were distributed more uniformly throughout the lumen (Fig. 4, H to K, and fig. S9F). While these observations reveal that mucus proximity is a major driver of *B. theta* expression variation, the presence of low-DC1 patches deeper into the lumen and many high-DC1 patches near the mucus, suggest that local spatial heterogeneity may produce similar scale variations. More broadly, our data suggest that other spatial covariates remain to be described, as some samples showed luminal regions with unique patterns of gene expression (fig. S9G) or low-DC1 expression signatures throughout the entire lumen (fig. S9, H to M).

To determine the operons that underlie the mucus-proximal adaptation of *B. theta* identified with DC1, we separated spatial patches into low-DC1 (mucus-associated) or high-DC1 (lumen-associated) compartments. We found 38 operons with statistically significant enrichment between these two compartments (Fig. 4L), with substantial overlap in the enriched operons when samples from the two fixation methods were analyzed separately (fig. S9, N to Q). Interestingly, there was a clear difference in the basic gene categories differentially expressed between the mucus- or lumen-associated compartments. 9 of the 15 lumen-associated operons were connected with central metabolism while none of the 23 mucus-associated operons had this functional annotation (Fig. 4L), suggesting that luminal *B. theta* may upregulate elements of central metabolism relative to *B. theta* closer to the mucus layer. By contrast, all 23 mucus-associated operons contain SusC or PUL-associated HTCS genes, as opposed to only 5 of the 15 lumen-associated operons, suggesting mucus-associated niches support harvesting of a greater polysaccharide diversity.

We next examined the known substrates for the PUL enriched in each compartment. Of the mucus-associated PUL with known substrates, 10 target host-mucus polysaccharides while only 1 targets dietary polysaccharides (Fig. 4L). By contrast, in the luminal compartment, 2 of the 3 PUL with known substrates target dietary polysaccharides (Fig. 4L). These observations are consistent with a greater availability of host-derived polysaccharides near the mucus, as expected, supporting our spatial analysis. Moreover, of the 23 PUL enriched in mucus-associated patches, 10 have unknown substrates (Fig. 4L). Given that we observe these operons enriched in mucus-associated patches, we predict that these PUL target host-derived polysaccharides or other carbohydrates enriched in this local environment. Our measurements complement a previous microdissection study (63) of the spatial variation of *B. theta* gene expression by providing a direct micron-scale measure of this variation and by extending the list of mucus-enriched PUL due, perhaps, to increased spatial resolution, improved sensitivity, or biological variability between the studies.

In total, these measurements reveal that bacteria fine tune gene expression to adapt to micron-scale niches in the gut. Excitingly, with bacterial-MERFISH it should now be possible to profile such micron-scale adaptation to a wide variety of complex environments.

## Discussion

Here we introduced bacterial-MERFISH, an image-based approach to single-cell transcriptomics that overcomes the massive mRNA density within bacterial cells by combining an optimized expansion microscopy toolbox with MERFISH. Bacterial-MERFISH offers complementary benefits to the growing suite of bacterial scRNA-seq methods (5–14). As a targeted method, it can sidestep abundant RNA (e.g., rRNA) challenges while providing an opportunity for high detection efficiency, which may prove essential for the many bacterial mRNAs that are very lowly expressed. In parallel, bacterial-MERFISH can be highly multiplexed, providing the ability to screen transcriptional changes with minimal prior knowledge of relevant targets. Bacterial-MERFISH can also image large numbers of cells, which may prove useful in the characterization of rare phenotypes. As an image-based technique, it naturally links gene expression to cell morphology or to intracellular molecular organization. Finally, with recent advances in all-optical readouts of pooled genetic screens (65, 66), it may now be possible to combine transcriptome-wide profiling with genome-wide perturbations in bacteria.

However, we anticipate that one of the most substantial advantages of bacterial-MERFISH will be the ability to profile bacterial behaviors *in situ*. Whether it is interactions between specialized cellular states within single-species biofilms or between different species in mixed communities, bacterial-MERFISH provides a means of linking spatial proximity, cellular micro-environment, and global architecture to gene expression. Studying bacteria in their native environment also bypasses the need for culture; thus, bacterial-MERFISH may offer an avenue for the *in situ* characterization of the diverse range of unculturable bacteria. Finally, as MERFISH can now target both eukaryotic and prokaryotic mRNAs, bacterial-MERFISH may allow the simultaneous profiling of host and bacterial gene expression, which may deepen our understanding of host-microbe interactions such as commensal colonization or pathogenic infection. More broadly, the

ability to directly profile single-cell bacterial transcriptomes in their native, complex environments may offer a new window into the substantial range of bacterial behaviors not well captured in a culture flask.

## 5 References and Notes

1. R. A. Fisher, B. Gollan, S. Helaine, Persistent bacterial infections and persister cells. *Nat. Rev. Microbiol.* **15**, 453–464 (2017).
2. L. P. Morawska, J. A. Hernandez-Valdes, O. P. Kuipers, Diversity of bet - hedging strategies in microbial communities—Recent cases and insights. *WIREs Mech. Dis.* **14**, e1544 (2022).  
10
3. L. Schröter, P. Dersch, Phenotypic diversification of microbial pathogens-cooperating and preparing for the future. *J. Mol. Biol.* **431**, 4645–4655 (2019).
4. J. Van Gestel, H. Vlamakis, R. Kolter, “Division of labor in biofilms: The ecology of cell differentiation” in *Microbial Biofilms* (ASM Press, Washington, DC, USA, 2015), pp. 67–  
15 97.
5. Y. Kang, M. H. Norris, J. Zarzycki-Siek, W. C. Nierman, S. P. Donachie, T. T. Hoang, Transcript amplification from single bacterium for transcriptome analysis. *Genome Res.* **21**, 925–935 (2011).
6. J. Wang, L. Chen, Z. Chen, W. Zhang, RNA-seq based transcriptomic analysis of single  
20 bacterial cells. *Integr. Biol. (Camb.)* **7**, 1466–1476 (2015).
7. F. Imdahl, E. Vafadarnejad, C. Homberger, A.-E. Saliba, J. Vogel, Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat. Microbiol.* **5**, 1202–1206 (2020).
8. S. B. Blattman, W. Jiang, P. Oikonomou, S. Tavazoie, Prokaryotic single-cell RNA  
25 sequencing by in situ combinatorial indexing. *Nat. Microbiol.* **5**, 1192–1201 (2020).
9. A. Kuchina, L. M. Brettner, L. Paleologu, C. M. Roco, A. B. Rosenberg, A. Carignano, R. Kibler, M. Hirano, R. W. DePaolo, G. Seelig, Microbial single-cell RNA sequencing by split-pool barcoding. *Science*, doi: 10.1126/science.aba5257 (2020).
10. R. McNulty, D. Sritharan, S. H. Pahng, J. P. Meisch, S. Liu, M. A. Brennan, G. Saxer, S.  
30 Hormoz, A. Z. Rosenthal, Probe-based bacterial single-cell RNA sequencing predicts toxin regulation. *Nat. Microbiol.* **8**, 934–945 (2023).
11. P. Ma, H. M. Amemiya, L. L. He, S. J. Gandhi, R. Nicol, R. P. Bhattacharyya, C. S. Smillie, D. T. Hung, Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell* **186**, 877-891.e14 (2023).



12. C. Homberger, R. J. Hayward, L. Barquist, J. Vogel, Improved bacterial single-cell RNA-seq through automated MATQ-seq and Cas9-based removal of rRNA reads. *MBio* **14**, e0355722 (2023).
- 5 13. Z. Xu, Y. Wang, K. Sheng, R. Rosenthal, N. Liu, X. Hua, T. Zhang, J. Chen, M. Song, Y. Lv, S. Zhang, Y. Huang, Z. Wang, T. Cao, Y. Shen, Y. Jiang, Y. Yu, Y. Chen, G. Guo, P. Yin, D. A. Weitz, Y. Wang, Droplet-based high-throughput single microbe RNA sequencing by smRandom-seq. *Nat. Commun.* **14**, 5130 (2023).
- 10 14. B. Wang, A. E. Lin, J. Yuan, K. E. Novak, M. D. Koch, N. S. Wingreen, B. Adamson, Z. Gitai, Single-cell massively-parallel multiplexed microbial sequencing (M3-seq) identifies rare bacterial populations and profiles phage infection. *Nat. Microbiol.* **8**, 1846–1862 (2023).
- 15 15. A. W. Pountain, P. Jiang, T. Yao, E. Homae, Y. Guan, K. J. C. McDonald, M. Podkowik, B. Shopsin, V. J. Torres, I. Golding, I. Yanai, Transcription–replication interactions reveal bacterial genome regulation. *Nature* **626**, 661–669 (2024).
- 15 16. M. Jia, S. Zhu, M.-Y. Xue, H. Chen, J. Xu, M. Song, Y. Tang, X. Liu, Y. Tao, T. Zhang, J.-X. Liu, Y. Wang, H.-Z. Sun, Single-cell transcriptomics across 2,534 microbial species reveals functional heterogeneity in the rumen microbiome. *Nat. Microbiol.*, doi: 10.1038/s41564-024-01723-9 (2024).
- 20 17. O. X. Cordero, M. S. Datta, Microbial interactions and community assembly at microscales. *Curr. Opin. Microbiol.* **31**, 227–234 (2016).
18. G. McCallum, C. Tropini, The gut microbiota and its biogeography. *Nat. Rev. Microbiol.*, doi: 10.1038/s41579-023-00969-0 (2023).
19. S. Thiery, C. Kaimer, The Predation Strategy of *Myxococcus xanthus*. *Front. Microbiol.* **11** doi: 10.3389/fmicb.2020.00002 (2020).
- 25 20. K. A. Gibbs, M. L. Urbanowski, E. P. Greenberg, Genetic determinants of self identity and social recognition in bacteria. *Science* **321**, 256–259 (2008).
21. S. K. Aoki, R. Pamma, A. D. Hernday, J. E. Bickham, B. A. Braaten, D. A. Low, Contact-dependent inhibition of growth in *Escherichia coli*. *Science* **309**, 1245–1248 (2005).
- 30 22. J. Zhou, H. Ma, L. Zhang, Mechanisms of virulence reprogramming in bacterial pathogens. *Annu. Rev. Microbiol.* **77**, 561–581 (2023).
23. M. Irastortza-Olaziregi, O. Amster-Choder, RNA localization in prokaryotes: Where, when, how, and why. *Wiley Interdiscip. Rev. RNA*, e1615 (2020).
24. A. Rao, D. Barkley, G. S. França, I. Yanai, Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).



25. X. Zhuang, Spatially resolved single-cell genomics and transcriptomics by imaging. *Nat. Methods* **18**, 18–22 (2021).
26. J. R. Moffitt, E. Lundberg, H. Heyn, The emerging landscape of spatial profiling technologies. *Nat. Rev. Genet.* **23**, 741–759 (2022).
- 5 27. D. Dar, N. Dar, L. Cai, D. K. Newman, Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution. *Science* **373**, eabi4882 (2021).
28. K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, X. Zhuang, RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
- 10 29. B. Huang, M. Bates, X. Zhuang, Super-resolution fluorescence microscopy. *Annu. Rev. Biochem.* **78**, 993–1016 (2009).
30. A. Bartholomäus, I. Fedyunin, P. Feist, C. Sin, G. Zhang, A. Valleriani, Z. Ignatova, Bacteria differently regulate mRNA abundance to specifically respond to various stresses. *Philos. Trans. A Math. Phys. Eng. Sci.* **374**, 20150069 (2016).
- 15 31. S. O. Skinner, L. A. Sepúlveda, H. Xu, I. Golding, Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat. Protoc.* **8**, 1100–1113 (2013).
32. C. Xia, J. Fan, G. Emanuel, J. Hao, X. Zhuang, Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19490–19499 (2019).
- 20 33. C.-H. L. Eng, M. Lawson, Q. Zhu, R. Dries, N. Koulena, Y. Takei, J. Yun, C. Cronin, C. Karp, G.-C. Yuan, L. Cai, Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* **568**, 235–239 (2019).
34. F. Chen, P. W. Tillberg, E. S. Boyden, Optical imaging. Expansion microscopy. *Science* **347**, 543–548 (2015).
- 25 35. G. Wang, J. R. Moffitt, X. Zhuang, Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* **8**, 4847 (2018).
36. H. G. J. Damstra, B. Mohar, M. Eddison, A. Akhmanova, L. C. Kapitein, P. W. Tillberg, Visualizing cellular and tissue ultrastructure using Ten-fold Robust Expansion Microscopy (TReX). *Elife* **11**, e73775 (2022).
- 30 37. D. Sarkar, J. Kang, A. T. Wassie, M. E. Schroeder, Z. Peng, T. B. Tarr, A.-H. Tang, E. D. Niederst, J. Z. Young, H. Su, D. Park, P. Yin, L.-H. Tsai, T. A. Blanpied, E. S. Boyden, Revealing nanostructures in brain tissue via protein decrowding by iterative expansion microscopy. *Nat. Biomed. Eng.* **6**, 1057–1073 (2022).

38. Y. Lim, A. L. Shiver, M. Khariton, K. M. Lane, K. M. Ng, S. R. Bray, J. Qin, K. C. Huang, B. Wang, Mechanically resolved imaging of bacteria using expansion microscopy. *PLoS Biol.* **17**, e3000268 (2019).
- 5 39. T. C. Kunz, R. Götz, M. Sauer, T. Rudel, Detection of Chlamydia developmental forms and secreted effectors by expansion microscopy. *Front. Cell. Infect. Microbiol.* **9**, 276 (2019).
40. R. Götz, T. C. Kunz, J. Fink, F. Solger, J. Schlegel, J. Seibel, V. Kozjak-Pavlovic, T. Rudel, M. Sauer, Nanoscale imaging of bacterial infections by sphingolipid expansion microscopy. *Nat. Commun.* **11**, 6173 (2020).
- 10 41. N. A. Rumyantseva, A. D. Vedyaykin, I. E. Vishnyakov, M. A. Khodorkovskii, Visualization of the intracellular structures of bacteria using expansion microscopy. *J. Phys. Conf. Ser.* **1697**, 012048 (2020).
42. T. C. Kunz, M. Rühling, A. Moldovan, K. Paprotka, V. Kozjak-Pavlovic, T. Rudel, M. Fraunholz, The expandables: Cracking the staphylococcal cell wall for expansion microscopy. *Front. Cell. Infect. Microbiol.* **11**, 644750 (2021).
- 15 43. D. Valdivieso González, J. Jara, V. G. Almendro-Vedia, B. Orgaz, I. López-Montero, Expansion microscopy applied to mono- and dual-species biofilms. *NPJ Biofilms Microbiomes* **9**, 92 (2023).
44. B. Görke, J. Stülke, Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat. Rev. Microbiol.* **6**, 613–624 (2008).
- 20 45. J. Monod, The growth of bacterial cultures. *Annu. Rev. Microbiol.* **3**, 371–394 (1949).
46. M. Ackermann, A functional perspective on phenotypic heterogeneity in microorganisms. *Nat. Rev. Microbiol.* **13**, 497–508 (2015).
47. J. M. Kim, M. Garcia-Alcala, E. Balleza, P. Cluzel, Stochastic transcriptional pulses orchestrate flagellar biosynthesis in *Escherichia coli*. *Sci. Adv.* **6** (2020).
- 25 48. N. M. V. Sampaio, C. M. Blassick, V. Andreani, J.-B. Lugagne, M. J. Dunlop, Dynamic gene expression and growth underlie cell-to-cell heterogeneity in *Escherichia coli* stress response. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2115032119 (2022).
49. O. Kotte, B. Volkmer, J. L. Radzikowski, M. Heinemann, Phenotypic bistability in *Escherichia coli*'s central carbon metabolism. *Mol. Syst. Biol.* **10**, 736 (2014).
- 30 50. K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, O. Amster-Choder, Translation-independent localization of mRNA in *E. coli*. *Science* **331**, 1081–1084 (2011).
51. J. R. Moffitt, S. Pandey, A. N. Boettiger, S. Wang, X. Zhuang, Spatial organization shapes the turnover of a bacterial transcriptome. *Elife* **5**, e13065 (2016).

52. S. Kannaiah, J. Livny, O. Amster-Choder, Spatiotemporal organization of the E. coli transcriptome: Translation independence and engagement in regulation. *Mol. Cell* **76**, 574–589.e7 (2019).
53. M. Pilhofer, M. Pavlekovic, N. M. Lee, W. Ludwig, K.-H. Schleifer, Fluorescence in situ hybridization for intracellular localization of nifH mRNA. *Syst. Appl. Microbiol.* **32**, 186–192 (2009).
54. V. T. dos Santos, A. W. Bisson-Filho, F. J. Gueiros-Filho, DivIVA-mediated polar localization of ComN, a posttranscriptional regulator of Bacillus subtilis. *J. Bacteriol.* **194**, 3661–3669 (2012).
55. S. Kannaiah, O. Goldberger, N. Alam, G. Barnabas, Y. Pozniak, A. Nussbaum-Shochat, O. Schueler-Furman, T. Geiger, O. Amster-Choder, MinD-RNase E interplay controls localization of polar mRNAs in E. coli. *EMBO J.*, doi: 10.1038/s44318-023-00026-9 (2024).
56. P. Montero Llopis, A. F. Jackson, O. Sliusarenko, I. Surovtsev, J. Heinritz, T. Emonet, C. Jacobs-Wagner, Spatial organization of the flow of genetic information in bacteria. *Nature* **466**, 77–81 (2010).
57. M. Mahbub, L. Hemm, Y. Yang, R. Kaur, H. Carmen, C. Engl, T. Huokko, M. Riediger, S. Watanabe, L.-N. Liu, A. Wilde, W. R. Hess, C. W. Mullineaux, mRNA localization, reaction centre biogenesis and thylakoid membrane targeting in cyanobacteria. *Nat. Plants* **6**, 1179–1191 (2020).
58. I. Collinson, R. A. Corey, W. J. Allen, Channel crossing: how are proteins shipped across the bacterial plasma membrane? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20150025 (2015).
59. J. Prilusky, E. Bibi, Studying membrane proteins through the eyes of the genetic code revealed a strong uracil bias in their coding mRNAs. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6662–6666 (2009).
60. P. Sarmah, W. Shang, A. Origi, M. Licheva, C. Kraft, M. Ulbrich, E. Lichtenberg, A. Wilde, H.-G. Koch, mRNA targeting eliminates the need for the signal recognition particle during membrane protein insertion in bacteria. *Cell Rep.* **42**, 112140 (2023).
61. X. Wang, D. Z. Rudner, Spatial organization of bacterial chromosomes. *Curr. Opin. Microbiol.* **22**, 66–72 (2014).
62. N. T. Porter, E. C. Martens, The critical roles of polysaccharides in gut microbial ecology and physiology. *Annu. Rev. Microbiol.* **71**, 349–369 (2017).
63. H. Li, J. P. Limenitakis, T. Fuhrer, M. B. Geuking, M. A. Lawson, M. Wyss, S. Brugiroux, I. Keller, J. A. Macpherson, S. Rupp, B. Stolp, J. V. Stein, B. Stecher, U. Sauer, K. D. McCoy, A. J. Macpherson, The outer mucus layer hosts a distinct intestinal microbial niche. *Nat. Commun.* **6**, 8292 (2015).

64. M. E. V. Johansson, J. M. H. Larsson, G. C. Hansson, The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. *Proc. Natl. Acad. Sci. U. S. A.* **108 Suppl 1**, 4659–4665 (2011).
- 5 65. G. Emanuel, J. R. Moffitt, X. Zhuang, High-throughput, image-based screening of pooled genetic-variant libraries. *Nat. Methods* **14**, 1159–1162 (2017).
66. M. J. Lawson, D. Camsund, J. Larsson, Ö. Baltekin, D. Fange, J. Elf, In situ genotyping of a pooled strain library after characterizing complex phenotypes. *Mol. Syst. Biol.* **13**, 947 (2017).
- 10 67. J. R. Moffitt, J. Hao, G. Wang, K. H. Chen, H. P. Babcock, X. Zhuang, High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 11046–11051 (2016).
68. J. R. Moffitt, D. Bambah-Mukku, S. W. Eichhorn, E. Vaughn, K. Shekhar, J. D. Perez, N. D. Rubinstein, J. Hao, A. Regev, C. Dulac, X. Zhuang, Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**, aau5324 (2018).
- 15 69. P. D. Karp, S. Paley, R. Caspi, A. Kothari, M. Krummenacker, P. E. Midford, L. R. Moore, P. Subhraveti, S. Gama-Castro, V. H. Tierrafria, P. Lara, L. Muñoz-Rascado, C. Bonavides-Martinez, A. Santos-Zavaleta, A. Mackie, G. Sun, T. A. Ahn-Horst, H. Choi, M. W. Covert, J. Collado-Vides, I. Paulsen, The EcoCyc database (2023). *EcoSal Plus*, eesp00022023 (2023).
- 20 70. E. C. Martens, H. C. Chiang, J. I. Gordon, Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* **4**, 447–457 (2008).
- 25 71. P. Cadinu, K. N. Sivanathan, A. Misra, R. J. Xu, D. Mangani, E. Yang, J. M. Rone, K. Tooley, Y.-C. Kye, L. Bod, L. Geistlinger, T. Lee, R. T. Mertens, N. Ono, G. Wang, L. Sanmarco, F. J. Quintana, A. C. Anderson, V. K. Kuchroo, J. R. Moffitt, R. Nowarski, Charting the cellular biogeography in colitis reveals fibroblast trajectories and coordinated spatial remodeling. *Cell* **187**, 2010-2028.e30 (2024).
- 30 72. R. I. Amann, B. J. Binder, R. J. Olson, S. W. Chisholm, R. Devereux, D. A. Stahl, Combination of 16S rRNA-targeted oligonucleotide probes with flow cytometry for analyzing mixed microbial populations. *Appl. Environ. Microbiol.* **56**, 1919–1925 (1990).
73. J. R. Moffitt, J. Hao, D. Bambah-Mukku, T. Lu, C. Dulac, X. Zhuang, High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14456–14461 (2016).
- 35 74. C. Zhang, J. S. Kang, S. M. Asano, R. Gao, E. S. Boyden, Expansion microscopy for beginners: Visualizing microtubules in expanded cultured HeLa cells. *Curr. Protoc. Neurosci.* **92**, e96 (2020).

75. J.-B. Chang, F. Chen, Y.-G. Yoon, E. E. Jung, H. Babcock, J. S. Kang, S. Asano, H.-J. Suk, N. Pak, P. W. Tillberg, A. T. Wassie, D. Cai, E. S. Boyden, Iterative expansion microscopy. *Nat. Methods* **14**, 593–599 (2017).
- 5 76. F. Chen, A. T. Wassie, A. J. Cote, A. Sinha, S. Alon, S. Asano, E. R. Daugharthy, J.-B. Chang, A. Marblestone, G. M. Church, A. Raj, E. S. Boyden, Nanoscale imaging of RNA with expansion microscopy. *Nat. Methods* **13**, 679–684 (2016).
- 10 77. Y. Wang, M. Eddison, G. Fleishman, M. Weigert, S. Xu, T. Wang, K. Rokicki, C. Goina, F. E. Henry, A. L. Lemire, U. Schmidt, H. Yang, K. Svoboda, E. W. Myers, S. Saalfeld, W. Korff, S. M. Sternson, P. W. Tillberg, EASI-FISH for thick tissue defines lateral hypothalamus spatio-molecular organization. *Cell* **184**, 6361–6377.e24 (2021).
78. T. Cordes, J. Vogelsang, P. Tinnefeld, On the mechanism of Trolox as antiblinking and antibleaching reagent. *J. Am. Chem. Soc.* **131**, 5018–5019 (2009).
- 15 79. M. B. Stead, A. Agrawal, K. E. Bowden, R. Nasir, B. K. Mohanty, R. B. Meagher, S. R. Kushner, RNAsnap<sup>TM</sup>: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic Acids Res.* **40**, e156 (2012).
80. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
- 20 81. S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, A. Kreshuk, Ilastik: Interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).
82. S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image contributors, scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014).
- 25 83. C. Tropini, T. K. Lee, J. Hsin, S. M. Desmarais, T. Ursell, R. D. Monds, K. C. Huang, Principles of bacterial cell-size determination revealed by cell-wall synthesis perturbations. *Cell Rep.* **9**, 1520–1527 (2014).
84. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, doi: 10.1186/s13059-017-1382-0 (2018).
- 30 85. I. Korsunsky, N. Millard, J. Fan, K. Slowikowski, F. Zhang, K. Wei, Y. Baglaenko, M. Brenner, P.-R. Loh, S. Raychaudhuri, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
- 35 86. W. Y. V. Lau, G. R. Hoad, V. Jin, G. L. Winsor, A. Madyan, K. L. Gray, M. R. Laird, R. Lo, F. S. L. Brinkman, PSORTdb 4.0: expanded and redesigned bacterial and archaeal protein subcellular localization database incorporating new secondary localizations. *Nucleic Acids Res.* **49**, D803–D808 (2021).

87. D. V. Klopfenstein, L. Zhang, B. S. Pedersen, F. Ramírez, A. Warwick Vesztrocy, A. Naldi, C. J. Mungall, J. M. Yunes, O. Botvinnik, M. Weigel, W. Dampier, C. Dessimoz, P. Flick, H. Tang, GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 10872 (2018).
- 5 88. C. Ausland, J. Zheng, H. Yi, B. Yang, T. Li, X. Feng, B. Zheng, Y. Yin, dbCAN-PUL: a database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Res.* **49**, D523–D528 (2021).
89. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).
- 10 90. E. Drula, M.-L. Garron, S. Dogan, V. Lombard, B. Henrissat, N. Terrapon, The carbohydrate-active enzyme database: functions and literature. *Nucleic Acids Res.* **50**, D571–D577 (2022).

**Acknowledgements:** We thank members of the Moffitt laboratory and S. Helaine for helpful discussion and a critical reading of the manuscript. We thank L. Bry, V. Yeliseyev, and M. Delany at the Massachusetts Host-Microbiome Center (MHMC) for providing monocolonized mice. Portions of this research were conducted on the O2 High Performance Compute Cluster, supported by the Research Computing Group, at Harvard Medical School. We thank the Dana-Farber/Harvard Cancer Center in Boston, MA, for the use of the Rodent Histopathology Core, which provided paraffin embedding and sectioning service. The Dana-Farber/Harvard Cancer Center is supported in part by an NCI Cancer Center Support Grant (P30CA06516). Access to this core and the MHMC were supported by the Harvard Digestive Disease Center (P30DK034854).

**Funding:**

National Institutes of Health grant R01GM143277 (JRM)

25 National Institutes of Health grant R21AI166230 (JRM)

Pew Biomedical Scholars Program (JRM)

ARPA-H (JRM)

**Author contributions:**

Conceptualization: AS, YW, JRM

30 Methodology: AS, YW, JRM

Investigation: AS, YW, NTA, JRM

Formal analysis: AS, YW, NTA, JRM

Funding acquisition: JRM

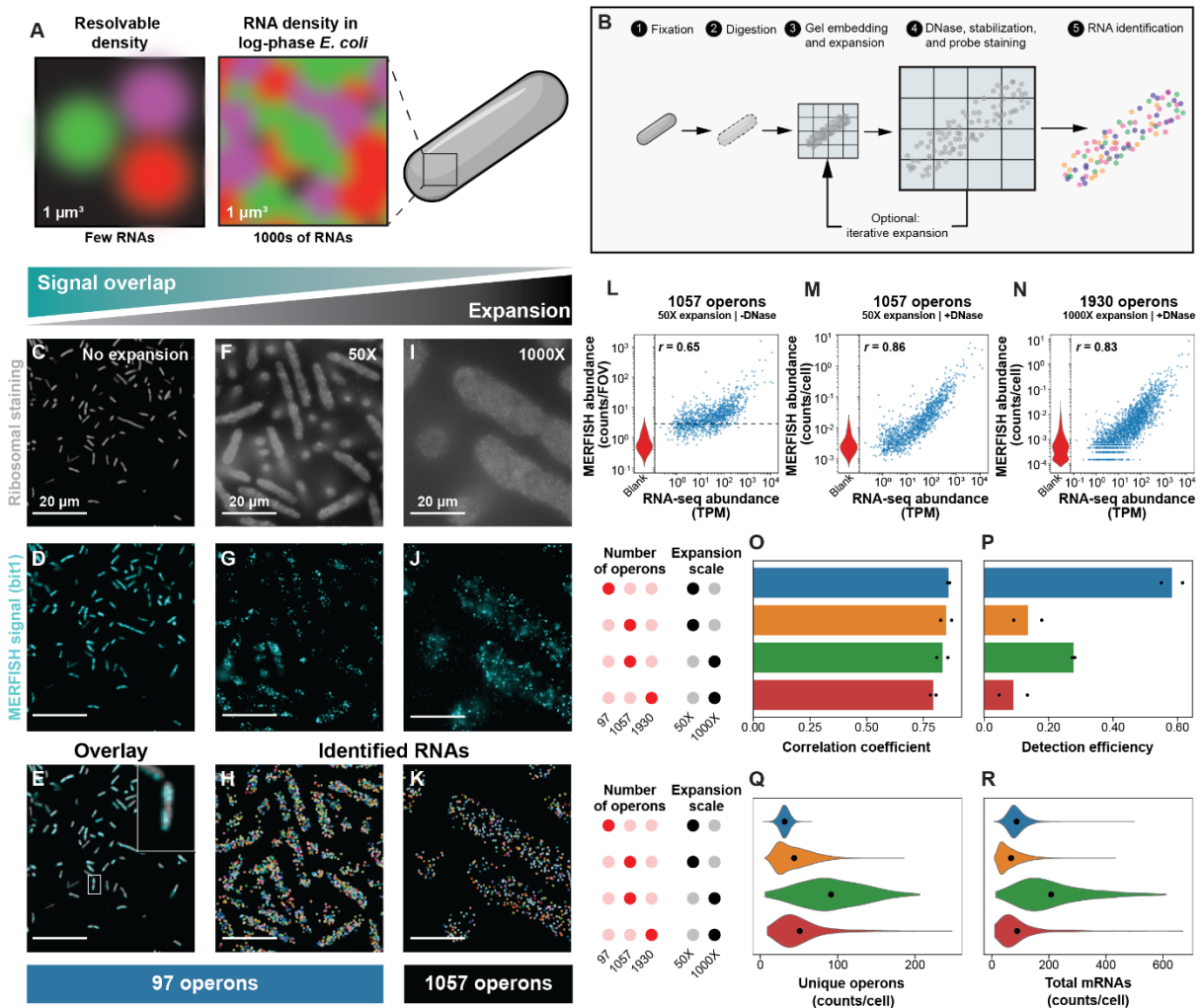
Supervision: JRM



Writing – original draft: AS, YW, NTA, JRM

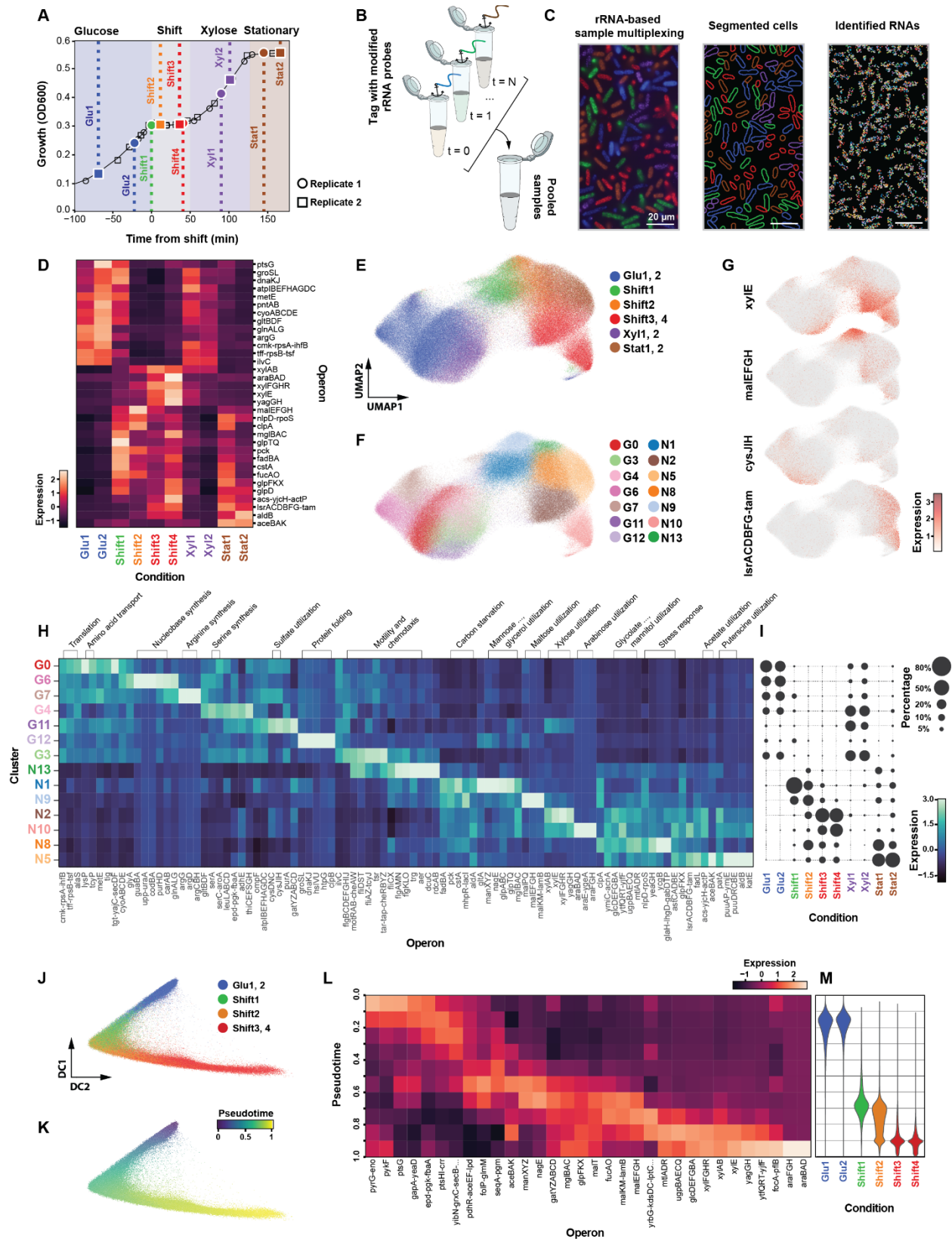
Writing – review & editing: AS, YW, NTA, JRM

**Competing interests:** JRM is a co-founder of, stakeholder in, and advisor for Vizgen, Inc. JRM is an inventor on patents associated with MERFISH applied for on his behalf by Harvard University and Boston Children’s Hospital. JRM’s interests were reviewed and are managed by Boston Children’s Hospital in accordance with their conflict-of-interest policies. All other authors declare that they have no competing interests.



**Fig. 1. Image-based expression profiling of thousands of operons in *E. coli* with bacterial-MERFISH.** (A) The RNA density resolvable with fluorescence microscopy and the estimated *E. coli* mRNA density. (B) Protocol for bacterial-MERFISH. (C-E) Images of fixed, log-phase *E. coli* grown in Luria Broth (LB) stained for the 16S ribosomal RNA (rRNA) (C), the first bit of a 97-operon MERFISH measurement (D), or overlay of the two (E). Scale bars: 20  $\mu$ m. (F,G) As in (C,D) but for 50X-expanded *E. coli*. (H) RNA identity (color) determined by MERFISH for cells in (F,G). (I-K) As in (F-H) but for 1000X-expanded *E. coli* stained with a 1,057-operon library. (L) Average mRNA copy number per field of view (FOV) for a 50X-expanded, 1,057-operon MERFISH measurement without DNase treatment versus bulk RNA-sequencing (RNA-seq) abundance from a matched culture. Dashed line: false-positive rate estimated from the average MERFISH abundance for the 100 mRNAs with the lowest RNA-seq abundance. TPM: transcripts per million reads. Left: distribution of MERFISH false-positive-controls ('blank'). (M,N) As in (L) but for 50X-expanded, 1,057-operon MERFISH (M) or 1000X-expanded, 1,930-operon MERFISH (N) with DNase treatment. *r*: Pearson correlation coefficient between logarithmic expression. (O,P) Correlation coefficient with bulk sequencing (O) or detection efficiency (P) for

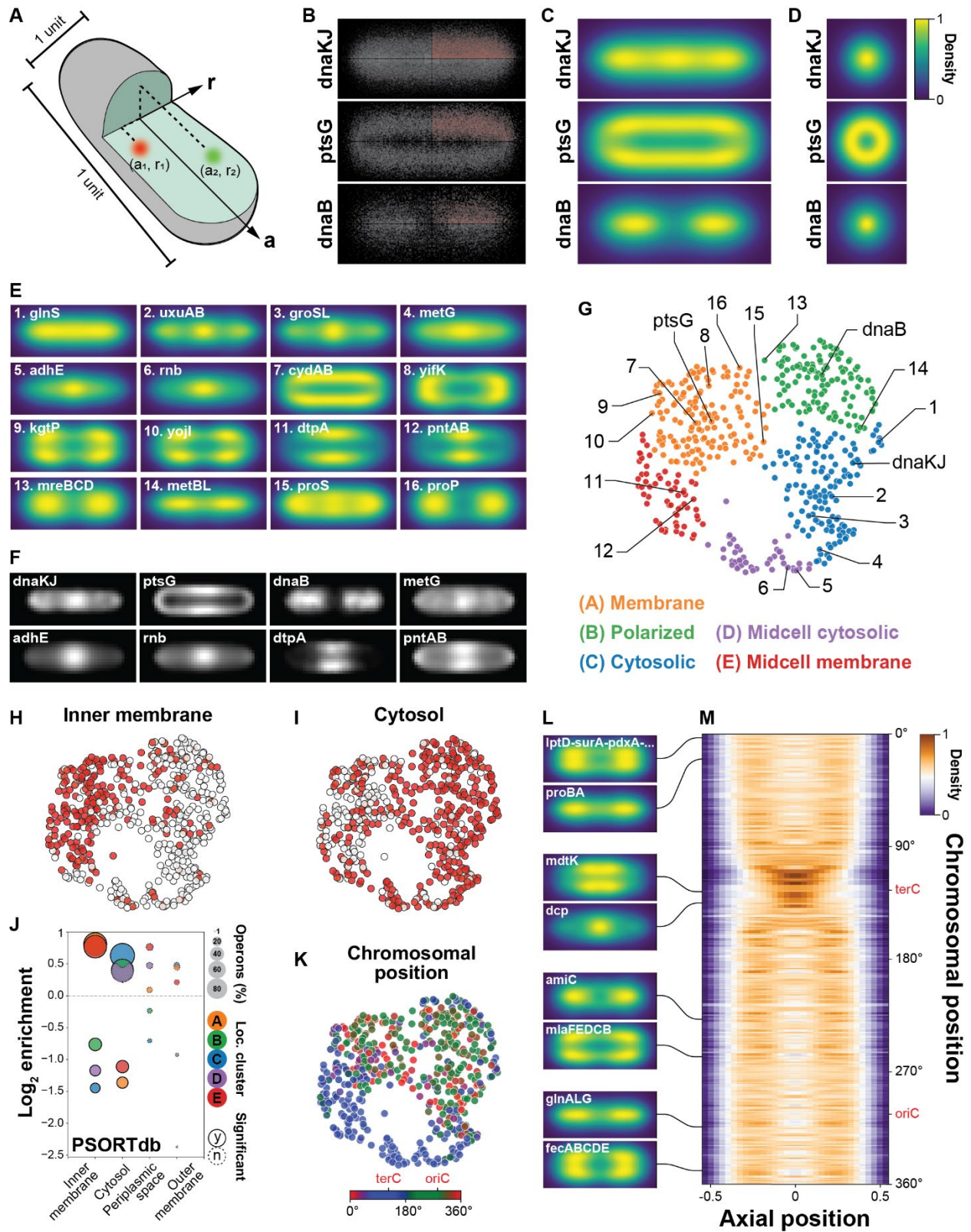
different MERFISH measurements. Bars represent the average between two replicates (markers). **(Q,R)** Distribution of unique operons (Q) or total mRNAs (R). Markers represent the average.



**Fig. 2. Bacterial-MERFISH reveals a stochastic hierarchical response to carbon starvation.**

(A) Optical density (OD) at 600 nm of *E. coli* growing in minimal defined medium containing a mixture of glucose and xylose. Markers: biological replicates, colored if sampled for MERFISH. (B) 16S rRNA sample multiplexing strategy. (C) rRNA image colored by identified sampling condition (left), segmented cell boundaries similarly colored (middle), and identified RNAs (color; right) for a 50X-expanded, 1,057-operon MERFISH measurement. rRNA colors are defined as in (A). Scale bars: 20  $\mu\text{m}$ . (D) Average gene expression measured in z-score for sampling conditions as listed in (A). (E-G) UMAP of cells colored by condition (E), Leiden cluster (F), or the natural log of normalized gene expression (G). Clusters are numbered in order of abundance and labeled based on growth ('G') or non-growth ('N'). (H,I) Marker gene expression measured in z-score (H) and relative fraction in each condition (I) for clusters. (J,K) Scatter plot of the first two diffusion components for cells collected from glucose log-phase or diauxic shift colored by condition (J) or pseudotime (K). (L) Average carbon-utilization operon expression measured in z-score at different pseudotimes. (M) Distribution of the pseudotime values by condition.

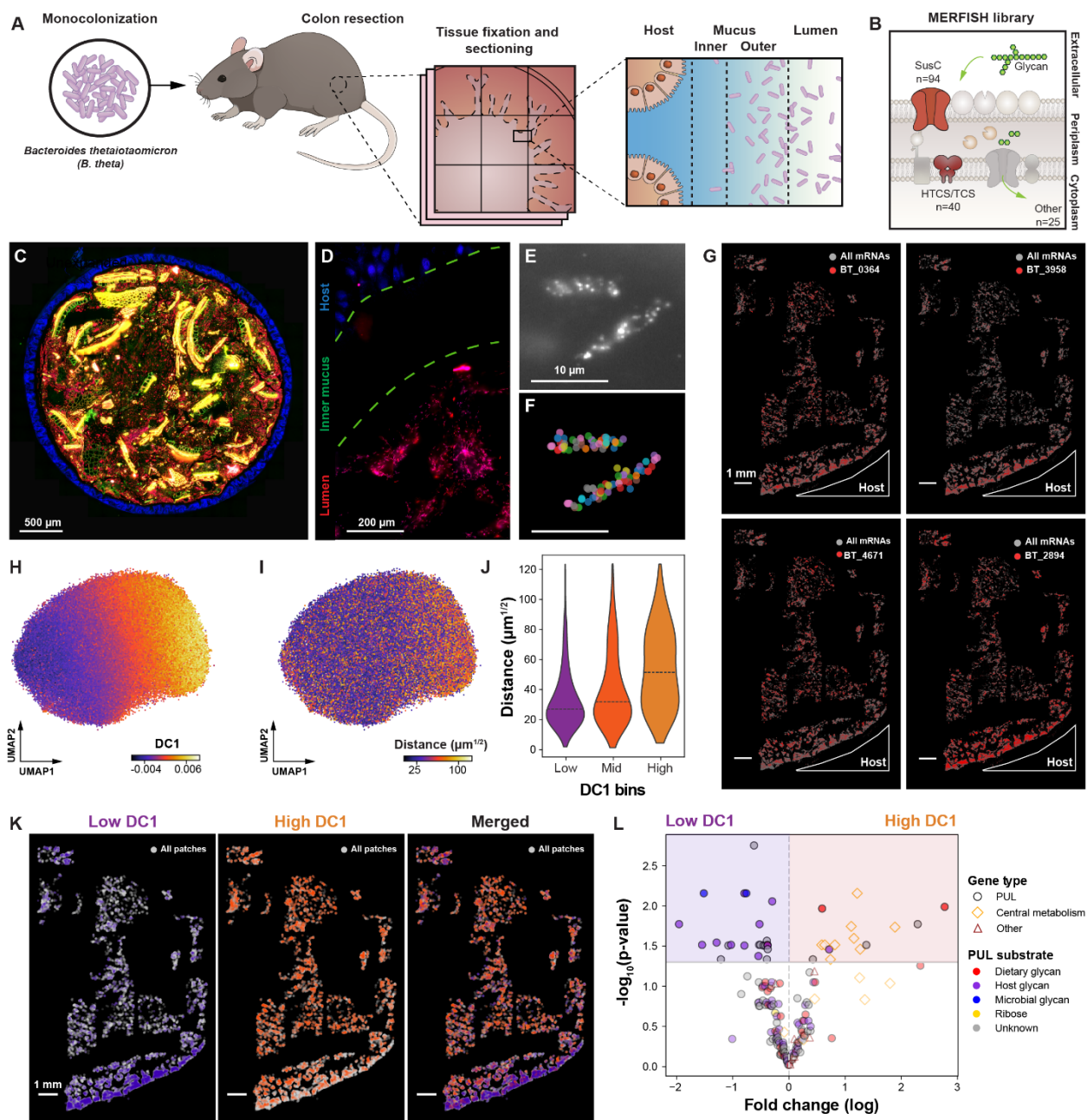




**Fig. 3. The *E. coli* transcriptome is organized in diverse patterns shaped by proteome and genome organization. (A) Mapping mRNAs to a normalized cellular coordinate system. (B)**



Location of *dnaKJ* (top), *ptsG* (middle), or *dnaB* (bottom) mRNAs measured with 50X-expanded, 1,057-operon MERFISH in LB log-phase *E. coli*. mRNAs were mapped to a positive axial midcell distance and radial coordinate (red upper quadrant) and mirrored (gray quadrants) for visualization and analysis. **(C,D)** Smoothed, normalized RNA density for the localizations plotted in **(B)** along the radial and axial directions **(C)** or radial directions **(D)**. **(E)** As in **(C)** for an illustrative mRNA set. **(F)** Averaged, normalized RNA signal for unexpanded smFISH of individual operons for a single z-plane. **(G)** UMAP of mRNA spatial distributions colored by cluster and marked by name or number for examples in **(C)** or **(E)**, respectively. **(H,I)** UMAP as in **(G)** colored by predicted location, inner membrane **(H)** or cytoplasm **(I)**, of the encoded proteins for each mRNA. **(J)** Enrichment of protein location label within each cluster. Color: cluster. Size: fraction of operons with label. Boundary: significance of enrichment ('y' is an FDR-corrected p-value less than 0.05). **(K)** UMAP as in **(G)** colored by chromosomal position from which each operon is transcribed. The origin (*oriC*) and terminus (*terC*) of replication are listed. **(L)** As in **(C)** for example mRNAs. **(M)** Normalized, averaged, axial mRNA density arrayed by the chromosomal location from which they were transcribed.



**Fig. 4. Adaptation of a gut commensal to micron-scale niches in the mammalian colon.** (A,B) The colon from *B. theta* monocolonized mice was dissected, fixed, sliced, and expanded (A), and MERFISH targeting 159 operons—comprising polysaccharide utilization loci (PUL), hybrid two-component systems (HTCS), and other operons—was performed (B). (C) Unexpanded colon section colored by 16S rRNA (red), DAPI (blue), and dietary debris autofluorescence (green). Yellow indicates autofluorescence in multiple channels. Scale bar: 500 µm. (D) 50X-expanded colonic section colored by DAPI (blue) and the first bit of a MERFISH measurement (magenta). Scale bar: 200 µm. (E,F) The first bit of a MERFISH measurement (E) and RNA identity (F) for two *B. theta* cells imaged in (D). Scale bars: 10 µm. (G) Spatial distribution of all (gray) or individual RNAs (red). Host tissue is marked. Scale bars: 1 mm. (H,I) UMAP of patch gene

expression colored by the first diffusion coefficient (DC1; H) or distance to host tissue (I). (J) Distribution of patch-host distance for patches with low, mid, and high DC1 values. Lines: median. (K) Spatial distribution of patches with low (blue), high (orange), or intermediate (gray) DC1 values for slice in (G). (L) Enrichment significance versus log-fold change for operon expression in high- versus low-DC1 patches. Marker: gene type. Marker fill color: PUL substrate, where known. The colored regions indicate an FDR-corrected p-values less than 0.05.