Microbial single-cell RNA sequencing by split-pool barcoding

Authors: Anna Kuchina[1,†,*], Leandra M. Brettner[2,3,†], Luana Paleologu[4,5], Charles M. Roco[2], Alexander B. Rosenberg[1], Alberto Carignano[1], Ryan Kibler[6], Matthew Hirano[1], R. William DePaolo[3,7], Georg Seelig[1,8,9*].

Affiliations:

[1]Department of Electrical and Computer Engineering, University of Washington, Seattle, WA, USA.

[2]Department of Bioengineering, University of Washington, Seattle, WA, USA.

[3]Center for Microbiome Sciences and Therapeutics, School of Medicine, University of Washington, Seattle, WA, USA

[4]Department of Microbiology, University of Washington, Seattle, WA, USA.

[5]Department of Biology, University of Washington, Seattle, WA, USA.

[6]Biological Physics, Structure, and Design, University of Washington, Seattle, WA, USA.

[7]Gasteroenterology, School of Medicine, University of Washington, Seattle, WA, USA.

[8]Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA, USA.

[9]Paul G. Allen School of Computer Science and Engineering, University of Washington, Seattle, WA, USA.

*Correspondence to: kuchina.anna@gmail.com (A.K.), gseelig@uw.edu (G.S.)

†These authors contributed equally to this work.

Abstract:

Single-cell RNA-sequencing (scRNA-seq) has become an essential tool for characterizing multi-celled eukaryotic systems but current methods are not compatible with bacteria. Here, we introduce microSPLiT, a low cost and high-throughput scRNA-seq method that works for gram-negative and gram-positive bacteria and can resolve transcriptional states that remain hidden at a population level. We applied microSPLiT to >25,000 *Bacillus subtilis* cells sampled from different growth stages, creating a detailed atlas of changes in metabolism and lifestyle. We not only retrieve detailed gene expression profiles associated with known but rare states such as competence and PBSX prophage induction, but also identify novel and unexpected gene expression states including heterogeneous activation of a niche metabolic pathway in a subpopulation of cells. microSPLiT empowers high-throughput analysis of gene expression in complex bacterial communities.

Main Text:

Gene expression in bacteria is highly heterogeneous even in isogenic populations grown under the same lab conditions. Using bet-hedging strategies, bacteria differentiate into subpopulations that assume different roles for the survival of the community (*1*, *2*). For example, gene expression programs governing developmental and stress-response states such as competence or antibiotic resistance may switch on stochastically in a small number of single cells (*3*–*5*). Population level gene expression measurements are insufficient to resolve such rare states which, to date, have been characterized only in tractable model systems and through single-cell methods such as fluorescence microscopy that can only measure a limited set of reporter genes at a time (*6*).

Single-cell RNA-seq (scRNA-seq) methods developed for use with eukaryotic cells can provide comprehensive gene expression profiles for tens of thousands of cells. However, although the need for microbial scRNA-seq has been recognized, early attempts have been limited to relatively small cell numbers because of the technical challenges associated with adapting scRNA-seq technology to microbes (**Figure 1A**)(*7–9*). Specifically, bacteria have very low mRNA content, typically about two orders of magnitude less than human cells (*8*). Separation of mRNA from rRNA is challenging because bacterial mRNA is not polyadenylated. Bacteria have a wide diversity of cell walls and membranes which interfere with the lysis or permeabilization steps required for scRNA-seq and their small size can hinder microfluidic single-cell isolation. Here, we present microSPLiT (Microbial Split-Pool Ligation Transcriptomics), a scRNA-seq platform that can overcome these challenges and become a transformative technology for microbiology research.

Developing microSPLiT.

SPLiT-seq, which labels the cellular origin of RNA through combinatorial barcoding, provides a starting point for bacterial single cell transcriptomics (*10*). In SPLiT-seq, cells are fixed, permeabilized and mRNA is converted to cDNA through in-cell reverse transcription (RT) with barcoded poly-T and random hexamer primers in a multi-well format. Cells are then pooled, randomly split into a new 96-well plate, and a well-specific barcode is appended to the cDNA through ligation. This split-ligation-pool cycle is repeated and a fourth, optional barcode is added during sequencing library preparation to ensure that each cell acquires a unique barcode combination with very high likelihood (**Figure 1B**).

Because SPLiT-seq does not require cell isolation, it is compatible with a wide range of cell shapes and sizes. Moreover, because SPLiT-seq already uses random hexamer primers in addition to poly-T primers for RT, we reasoned that it might be suitable for detecting bacterial mRNA. However, a direct application of the mammalian SPLiT-seq protocol to bacteria, not surprisingly, resulted in low total UMI counts (<100 max UMIs/cell, median 0 mRNA reads/cell) and a bias toward gram-negative over gram-positive bacteria (**Figure S1**).

Next, we set out to optimize sample processing steps for bacteria. We took advantage of SPLiT-seq's multiplexing capabilities to test a wide range of approaches to cell wall removal and membrane permeabilization. We settled on a combination of a mild detergent, Tween-20, and lysozyme, as that treatment protocol demonstrated the best capture for both the gram positive and gram negative bacteria tested. (**Figure 1B, inset, and Supplementary Table S1**). Then, we compared different methods that would be compatible for in-cell mRNA enrichment. We tested polyadenylation with *E.coli* Poly(A) Polymerase I (PAP) which was previously used to preferentially polyadenylate mRNA (*11*), 5'-phosphate-dependent exonuclease ("Terminator", Epicentre) treatment and reverse transcription with ribosomal RNA-specific probes followed by RNaseH-mediated degradation (**Figure S2 and Supplementary Table S1**). We found that the treatment of fixed and permeabilized cells with PAP resulted in the highest (about 2.5-fold, or approximately 7% of total RNA) enrichment of mRNA reads (**Figure 1B, inset, and Figure S2**). We also optimized the fixation protocol as well as the downstream enzymatic reaction conditions (**Supplementary Table S1 and Materials and Methods).** Notably, we found that RT resulted in cell clumping and that mild sonication after this step was necessary to reliably obtain single cell suspensions (**Figure S3**).

microSPLiT generates high-quality single-cell RNA-seq data.

In order to validate microSPLiT performance on a mixture of gram-positive and gram-negative organisms, we grew *Escherichia coli* MW1255 and *Bacillus subtilis* PY79 cells to OD0.5 and

subjected half of each culture to a brief 47C heat-shock. We performed a microSPLiT experiment on both samples, using the first barcode as a sample identifier. We prepared and sequenced a cDNA library from 2677 total bacteria from heat-shocked and control treatments and aligned the reads to a combined *B. subtilis-E. coli* genome. 99.2% of the putative single cell transcriptomes were unambiguously assigned to a single species (**Figure 1C**). We sampled a median of 237 unique mRNA transcripts per cell for *E. coli* and 376 for *B. subtilis*, or approximately 5-10% of the estimated total mRNA (*12*). In total, we detected 3717 genes for *E. coli* and 3476 genes for *B. subtilis* (**Figure 1D**).

Next, we tested whether microSPLiT could detect transcriptional responses to heat shock (**Figure 1E**). Unsupervised clustering identified five distinct clusters which were visualized by t-distributed stochastic neighbor embedding (t-SNE) (**Figure 1F**). The first barcode identified two pairs of clusters corresponding to the heat treated and control cultures, and gene expression analysis within each pair further labeled them as corresponding to *B. subtilis* and *E. coli* cells (**Figure S4A**). Analysis of genes enriched within each cluster revealed classical heat shock genes differentially expressed in each of the E.C. and *B. subtilis* heat treated clusters. We detected induction of genes for abundant class I-IV and VI *B. subtilis* heat shock genes such as *groEL*, *dnaK, clpC* and *htpG* operons (**Figures 1G, S4B-C**) as well as both of the major chaperone systems of E.C. (*groEL* and *dnaK* operons, **Figures 1H, S4B-C**) (*13*, *14*).

Unexpectedly, we found an additional small cluster, representing *E. coli* cells from both control and heat treated samples that expressed a dramatically different signature of DEAD-box helicase *deaD* induction as well as cold shock genes *cspA-G* consistent with a transcriptional response to cold (**Figures 1F and S4D**) (*15*). This subpopulation of *E. coli* might be displaying a very rapid response to cold from a brief cold centrifugation step performed as the first step in sample preparation before formaldehyde fixation.

Transcriptional patterns during *B. subtilis* growth in rich medium.

Next, we applied microSPLiT to capture transcriptional states across the *B. subtilis* growth curve in a rich medium (LB). In total, we sampled ten optical density (OD) points along the growth curve of the laboratory strain PY79 ranging from OD 0.5 (early exponential phase) to 6.0 (early stationary phase). Four of the OD points were sampled in both replicates of microSPLiT while the rest were sampled only in one of the experiments (**Figure 2A**). In both experiments, the first barcode was used to record sample identity (i.e. OD). The data from two different experiments are consistent and produced a combined dataset of 25,214 cells (**Figures 2B, S5 and S6**). Unsupervised clustering of the combined datasets revealed 14 clusters, most of which contained cells predominantly from a single OD (**Figure 2B**). The most notable exceptions are two smaller clusters that contain cells from multiple ODs: a cluster with cells from OD2-3.2 that differentially express myo-inositol metabolism pathway genes, and a very small cluster containing only 36 cells from 5 different OD points uniquely expressing genes associated with the defective PBSX prophage (**Figures 2B and S7-S10**).

Alternative sigma factors are the primary regulators of the prokaryotic RNA polymerase specificity and thus directly shape transcriptional changes in response to environmental conditions. To understand whether microSPLiT could capture variation in sigma factor utilization across different growth stages, we averaged activity of genes regulated by each sigma factor, recording, for each cluster, both the percentage of cells expressing at least one gene regulated by a given sigma factor and the average intensity of gene expression (**Figure 2C**). We note that while averaged intensity could also be obtained from bulk RNA-seq, only single-cell methods can reveal information about cell-to-cell variation in regulon activity.

Consistent with expectations, we observe that the housekeeping σ^A activity is highest at OD0.5, while the activity of general stress response sigma factor σ^B rises as cells begin to exit from exponential phase (OD1.3-1.7) and then declines as cells approach stationary. Sporulation sigma factors σ^F, σ^J and σ^K were induced at later ODs but in a small fraction of cells, consistent with the previously reported heterogeneous initiation of sporulation (*3*, *16*). The extracellular function (ECF) sigma factors σ^M, σ^W and σ^X implicated in maintaining cell envelope function reached maximal activity at OD 1.0 in a large proportion of cells, consistent with reports of their basal activity in logarithmic phase in non-stressed cells (*17*). Meanwhile, the remaining ECF sigma factors σ^V, raising defenses against lytic endoglycosidases, and σ^Y increased in activity towards later OD points in a small subpopulation of cells, similar to the sporulation sigma factors. σ^I and σ^H activities, regulating heat response and post-exponential behavior respectively, peak in cluster 8 which represents a subgroup of cells at OD 1.7. In contrast, σ^B and σ^D regulating general stress response and motility are most active in cluster 7, a second distinct subgroup of OD 1.7 cells. Finally, σ^L implicated in utilization of arginine, acetoin and fructose as well as regulation of the cold shock response peaks in cells at OD 2-2.8 represented in cluster 10, most likely due to the highly enriched acetoin utilization genes in this cluster.

To obtain an even finer-grained picture of the transcriptional programs during exponential growth and entry to stationary phase, we inferred the activity profiles of select transcriptional regulators (TR) from expression of the genes in their respective regulons (**Figure 2D**) (*18*). This analysis revealed pronounced changes in regulation of carbon utilization, stress responses, metal uptake, developmental decisions and more. The main transition state regulator AbrB becomes inactive after OD1.0, indicating that the preferred carbon sources such as glucose start getting depleted. In response, cells begin to activate transcriptional programs to utilize different alternative carbon sources. Toward the intermediate growth stages, the cells start to sense carbon, nitrogen and phosphate limitation. Clusters 7 and 8 (OD 1.7) display a striking change in carbon metabolism indicated by the strong expression of genes repressed by the carbon catabolite control proteins CcpA, CcpC and CcpN. Similarly, the regulator of nitrogen assimilation TnrA becomes activated at OD1.7, while PhoP, regulating the phosphate metabolism, becomes active at three different growth stages: OD1.0, OD1.7 and later on at OD6.0. In addition, cells respond to metal deficiency, switching off the negative regulators of iron and manganese uptake Fur and MntR after OD1.0. Finally, we can observe cellular response to a variety of intrinsic and cell-envelope stresses, as well as temporal activation patterns of a battery of developmental regulators including ComA, SinI, DegU, Rok, Spo0A and others. Surprisingly, we also observe an upregulation of ComK-regulated genes in a high proportion of cells in the early ODs which is not consistent with the primary role of this transcription factor in a rare developmental state of competence. It could be explained by the fact that a large cohort of ComK-induced genes are involved in metabolism and DNA repair and can be activated by other regulators. These data show that microSPLiT not only captures known regulatory programs but also reveals intriguing heterogeneity in a wide range of these pathways.

Central carbon metabolism changes and differential expression of TCA cycle enzymes

Given the dramatic changes in regulation of carbon utilization observed in our TR analysis, we turned to a more comprehensive examination of carbon metabolism genes enriched in each cluster (**Figures S7-S9 and Figure 3**). When glucose and other preferred sugars (discussed below) are present, they are converted to pyruvate during glycolysis, which is the primary metabolic route under excess of these sugars. In these conditions promoting rapid growth, pyruvate is then converted to lactate, acetate, acetoin, and other by-products through fermentation. Upon depletion of preferred sugars, cells redirect the fermentation by-products to

be metabolized in the TCA cycle generating additional adenosine triphosphate (ATP) and carbon dioxide.

In the exponential phase (Clusters 0 and 1, OD0.5) we find high expression of *ptsG,* a glucose permease which transports and phosphorylates glucose. The enzymes in the *gapA* operon constituting the metabolic pathway from glyceraldehyde-3P to phosphoenolpyruvate (PEP): *gapA, pgk, pgm* and *eno* (*19*) were upregulated in this cluster (**Figure 3A,B**).

Next, in clusters 3 and 4 (OD1.0) we observe transcriptional patterns suggesting an increase in flux from pyruvate either being converted to lactate by *ldh* (*20*) which is then exported via a malate antiporter *mleN* (*21*) or converted to acetate via intermediates by *pdhAD* (*22*) and *ackA* (*23*) (**Figures 3A,B**). These observations are consistent with previous reports of transient medium acidification via acetate production during rapid fermentative *B. subtilis* growth (*24*).

At OD1.7 cells appear to undergo a dramatic transition from glycolysis to gluconeogenesis with multiple genes from the gluconeogenetic pathway activated in clusters 7 and 8. There are two glyceraldehyde-3-phosphate dehydrogenases in *B. subtilis*: GapA and GapB, mediating the flux of carbon either from glucose to the TCA cycle or vice versa (*25*). The glucose and intermediates generated by the gluconeogenetic pathway under conditions of glucose limitation are then used for synthesis of necessary structural constituents. We observe the switch from GapA to GapB expression in clusters 7 and 8 along with an upregulation of most of the TCA cycle enzymes. We also find a different pattern of pyruvate production and utilization, together with catabolism of acetoin, another fermentation product, and additional nutrient fluxes into the TCA cycle (**Supplementary Text and Figures 3A-C, S7-S9**).

Interestingly, the TCA cycle enzymes which are collectively enriched in the entire OD1.7 sample display differential transcriptional abundances between clusters 7 and 8 (**Figures 3A, B**). For instance, cells in cluster 7 express more aconitase (*citB*), 2-oxoglutarate dehydrogenase (*odhAB*), and succinate dehydrogenase (*sdhABC*), while cells from cluster 8 display higher expression of genes in the *citZ-icd-mdh* operon and *sucCD* (succinyl-CoA synthetase). CitB was reported to destabilize the *citZ-icd-mdh* mRNA upon citrate accumulation or iron limitation (*26*). Thus, differential transcript abundance of the TCA cycle enzymes could reflect the regulatory interactions between the respective genes.

Transient activation of alternative carbon utilization pathways in distinct subpopulations

Complementing changes in the core carbon metabolism, we observe expression of pathways responsible for uptake and utilization of a variety of different carbon sources. As the preferred sources of carbon are depleted, the major repressor of alternative carbon utilization pathways CcpA becomes inactive, permitting the cells to catabolize alternative carbon sources including a variety of carbohydrates (*27*) (**Figure 3**). We find that the activation and suppression of these pathways happen in varying proportions of the cells in each OD sample and appear to follow a temporal order (**Figures 3 and S7-S11**). The preferred carbon sources for *B. subtilis* are glucose along with other Group A sugars (sucrose, fructose, and mannitol) and malate (*28*, *29*). While the latter is directly metabolized in the tricarboxylic acid (TCA) cycle, the former sugars are converted to one of the glucose metabolic intermediates. At OD0.5, in addition to glucose-specific *ptsG*, we observe increased expression of *mtlA-D* and *glpK,D,F* responsible for utilization of mannitol and glycerol, respectively. Cells in clusters 3 and 4 (OD1.0) activate catabolism of mannose and aryl-β-glucosides (*manA,P* and *bglH,P*). In cluster 7 (OD1.7) we observe the upregulation of genes for utilization of glucomannan (*gmuA-F*) and the ribose transporter (*rbsA-D*), while the gene for utilization of ribose *rbsK*, curiously, is upregulated much earlier in cluster 3. Finally, at even later ODs three additional alternative carbon source utilization programs switch on. Cluster 9

comprising cells from ODs 1.7, 2.0, 2.8, and 3.2 is defined by the expression of genes implicated in the most common stereoisomer of inositol, myo-inositol catabolism (*iolABCDEFGHIJ,* further "*iolAJ*"), while cluster 10 (OD2.0) is enriched for genes responsible for utilization of acetoin (*acoABCL*). Finally, cluster 11, representing a range of ODs from 3.2 to 5.3, differentially expresses genes for melibiose utilization (*melA, melCDE*).

Heterogeneous activation of myo-inositol catabolism pathway at intermediate growth stages

Inositol is an abundant resource in soil, and *B. subtilis* is able to subsist on inositol as its sole carbon source (*30*). While LB medium is not typically expected to contain myo-inositol (further "inositol"), heterogeneous inositol utilization pathway activation is observed in a small (3-15%) subpopulation in both of our independent LB growth experiments (cluster 9, see **Figure 3A**, **Figure 4A-B** and **Figure S12**). It has been shown that the inositol catabolism intermediate, 2-deoxy-5-keto-D-gluconic acid 6-phosphate (DKGP), is responsible for the pathway induction (*30*). We hypothesize that the trace amounts of inositol may be present in the LB medium, potentially from the yeast extract since yeast is capable of inositol production as a precursor to the essential membrane component, phosphatidylinositol.

There are three operons involved in inositol utilization, *iolT* (main transporter), *iolRS* (the first gene is a repressor and the second is a likely dehydrogenase), and *iolAJ* (metabolic enzymes), with *iolC* producing and *iolJ* cleaving the pathway-activating DKGP intermediate. *iolRS* and *iolAJ* are normally transcribed by σ^A through divergent transcription (*31*). In the absence of the inducer (I), IolR supresses transcription of all three operons (*32*). In addition, CcpA represses the *iolAJ* operon in the presence of glucose (*33*) (**Figure 4C**). Interestingly, the pathway suppressor *iolR* gene is more broadly expressed both outside and inside of cluster 9 (**Figure 4A-B** and **Figure S12**).

Next, we asked whether our data could be explained by the underlying gene regulatory network architecture (**Figure 4D**). Within the complex system of interactions, we observe two topological features capable of amplifying small molecular variations. An *iolAJ*-mediated incoherent feed-forward loop can generate a pulse of DKGP, while IolT can generate a positive-feedback loop by transporting inositol into the cell. These features of the network can explain the difference in clusters, as probed with a mathematical model (**Figure S13**). Due to the small proportion of cells expressing the main catabolic operon *iolAJ*, this particular metabolic behavior could only be reliably detected on a single-cell level.

Motility, antimicrobials production, stress response, and metal ion import

B. subtilis is known to exhibit a variety of behaviors to enhance survival in adverse conditions including, but not limited to, production of degradative enzymes and antimicrobials, secretion and uptake of siderophores, three types of motility (swimming, swarming, and sliding), natural competence, and sporulation (*34*). Many if not most of these behaviors are not displayed by all cells at the same time (*35*), thus providing good targets for single-cell interrogation.

Bacteria universally produce peptide and small molecule antimicrobials that are meant to target both closely and distantly related organisms (**Figure 5A**) (*36*). We observe the expression of endoA toxin/antitoxin (*ndoA/I*) and bacilysin (*bacA-G*) more broadly prior to OD2.0 (**Figure 5A and Figure S14)**. Subtilosin (*sboA,X-albA-F*), bacillaene (*pksC-R*), and plipastatin (*ppsA-E*) production, on the other hand, are detected in a greater proportion of cells after OD2.0. All three of these peptides are broad spectrum antimicrobials (*37–39*) which suggests that under nutrient limitation, *B. subtilis* becomes competitive to preserve its access to the available resources, even

in the absence of other species. Additionally, we also see a rise in spore killing factor (SKF) and spore delay protein (SDP) in the last three ODs (**Figure S15**).

B. subtilis has two morphological phenotypes during active growth: filamentous, sessile cells and smaller motile cells (*40*, *41*). In liquid media, the bacteria primarily swim which requires the expression of the flagellin protein (*hag*) and the *fla-che* operon (**Figures 5B**). We detected expression of these genes in 25 to 50% of cells in every cluster from early to intermediate growth stages before the fraction noticeably declines at late ODs, consistent with reports of heterogeneity of the motility operon expression at mid-exponential phase (*42*). The *B. subtilis* population is expected to be similarly differentiated into surfactin-producing and extracellular matrix-producing bacteria as cell density increases (*42*). We see the fraction of cells producing *srfA-D* genes gradually increase, slightly dipping at OD 1.7 and 2.0, and finally reaching almost 100% detection at OD 6.0, consistent with the PY79 strain having defective matrix production genes that cannot negatively regulate *srfA-D* expression (*43*).

Cellular stress response pathways exemplified by GroEL chaperonin and ClpC proteases peak at OD1.7, the same time as the cells switch from glycolysis to gluconeogenesis (**Figure 5C**). The ClpP associated proteases (*clpP,C,X,E*), McsA and McsB kinases (*mcsA,B*), and chaperonins (*groEL,ES*) are all involved in the unfolded protein response (*44*). A transient increase in the regulatory sigma factor, σ^B, inducing expression of these genes, has been reported during normal exponential growth and attributed to intrinsic cellular stresses (*45*). *clpP* mutants have been linked to slower glucose consumption and overproduction of TCA cycle metabolites during rapid fermentative growth, which suggests their participation in glycolysis and may explain their highest expression during the early growth stages (*46*).

B. subtilis needs both manganese and iron to grow (*47*), and the two ions often antagonize each other's regulators (*48*). Upon iron limitation, *B. subtilis* produces and secretes siderophores including bacillibactin (*dhbA*) together with the ABC transporter (*feuA*) (**Figure 5D**) (*49*, *50*). Both genes show increased expression at OD1.7, similar to the stress response proteins above. Manganese, on the other hand, is transported into the cells via a separate ABC transporter (*mntA-D*) and a proton symporter (*mntH*) (*48*). It is required for the successful transition into stress related states such as biofilm formation and sporulation (*51*), which likely explains the increase in manganese transport related genes detected at later ODs.

microSPLiT quantifies a rare stress response

Cluster 13 (36 cells, or 0.142% of total cells, representing ODs between 0.5 and 2.8) contains a rare subpopulation of cells expressing PBSX prophage genes (**Figure 6A**). The PBSX element is a defective prophage that is non-infectious but upon induction causes the release of phage-like particles (*52*) containing 13 kb of random fragmented chromosomal DNA (*53*, *54*). Prophage gene expression is induced by DNA damage mediated by the SOS response (*52*, *55*)(**Figure 6B**). We identified eleven host genes with known or putative functions expressed in the PBSX prophage cluster (**Figure 6C**). Five of these genes have previously been shown to be induced only in PBSX-harboring strains of B. S. after DNA damage (*53*). The rest, including a chemoreceptor (*mcpC*), an ATP-binding cassette transporter (*liaL*), a cell wall binding protein (*ykuG*), an ammonium transporter (*amtB*), a sucrose-6-phospate hydrolase (*sacA*), and a regulatory protein of homologous recombination (*recX*), have not previously been linked to prophage induction (**Supplementary Text**). These genes could have been induced by the damage sustained by the cells which also caused PBSX prophage activation. Alternatively, these genes could have been activated in response to the chromosomal fragmentation and membrane vesicle formation caused by the prophage induction.

microSPLiT detects a rare stochastically induced developmental state

Under stress or nutrient limitation, a small fraction (2-5%) of *B. subtilis* cells undergoes stochastic transient differentiation into a state of natural competence, characterized by the ability to uptake extracellular DNA and integrate it into the chromosome (**Figure 6D**) (*56*). The master transcriptional regulator of competence ComK is activated via a positive feedback loop, inducing expression of a suite of >165 genes involved in a variety of cellular processes in addition to DNA uptake (*57*). Competence is expected to naturally occur under nutrient limitation. We thus separately subclustered the last two OD points (OD5.3 and 6.0). UMAP embedding revealed a small cluster (62 cells, or 4.6% of cells at OD 5.3 and 6) expressing a distinct transcriptional signature of the competent state, or K-state (**Figure 6E-F**). The most enriched gene was *comGA*, as expected from prior transcriptomic data, followed by the succinyl-CoA synthetase (*sucCD*) operon which has been shown to be induced in competent cells (*24*, *57*, *58*). We also see strong enrichment of other genes encoding the DNA uptake machinery: *comF* and *comE* operons, as well as the response regulator RapH which represses sporulation development in competent cells by dephosphorylating the phosphorelay component Spo0F (*59*). As the cells in the competent state arrest replication to integrate the incoming DNA into the chromosome, we observe the upregulation of genes necessary for processing of internalized ssDNA such as DNA recombination and repair protein RecA along with single-strand DNA binding proteins SsbA and SsbB consistent with prior reports (*57*). We also observe enrichment of several genes related to DNA processing such as *topA* encoding topoisomerase A and *holA*, delta subunit of DNA polymerase III which is a part of the replisome. These genes are not in the annotated ComK regulon but were tentatively identified in the microarray data comparing gene expression between *mecA* strain in which essentially all cells express ComK and a double mutant *mecA comK* strain (*57*). This coordinated upregulation of *topA* and *holA* is consistent with the model of RecA binding to the SsbA\SsbB coated ssDNA and forming a complex with the replisome during competence (*60*). In addition, we found four genes not previously linked to the competent state: *ywfM* (unknown), *hemQ* (coproheme decarboxylase), *tlpC* (an orphan membrane-bound chemotaxis receptor), and *trmF,* a folate- and FAD-dependent tRNA methyltransferase (**Supplementary Text**).

Discussion

We applied microSPLiT to *B. subtilis* cells growing in liquid rich medium, conditions which are not usually associated with abundant cellular heterogeneity. Nevertheless, we found a large variety of subpopulations displaying differential gene expression of select metabolic, stress response or developmental pathways. In particular, we identified a metabolic pathway, myo-inositol catabolism, which was strikingly activated only in a fraction of cells from several later OD points in a distinct temporal fashion. We expect microSPLiT to have even more utility in identifying heterogeneous cell states in an environment more conducive to bacterial differentiation, such as in multi-species biofilms and natural microbiota.

We were able to detect subpopulations of cells as rare as 0.142%, pointing to microSPLiT's potential to uncover physiologically relevant rare cell states, such as persistence, that are hard to study by bulk or low-throughput methods. For many such states, the regulators are not well known and consequently, the reporters or mutants producing the desired state at a higher frequency cannot be engineered. Even for states that are better understood and can be artificially induced, such as prophage induction by UV irradiation, microSPLiT is useful since the transcriptional signatures it produces are state-specific and free of artifacts introduced by the perturbation.

In order to use microSPLiT on complex natural communities, the protocol will likely need to be further optimized, particularly the permeabilization and mRNA enrichment steps. Since cell wall and membrane composition vary significantly among bacteria, any permeabilization treatment is

likely to work better for some species than others. However, alternate treatment for different subsamples may still provide optimal results. In addition, we experienced lower mRNA counts from bacteria in stationary phase as opposed to logarithmic growth phase (**Figure S16**), consistent with slower growth rate and smaller cell size at this stage. Although the resulting data were still sufficient to reliably identify rare cell states such as the K-state, further improvement of the protocol will be instrumental to increase sensitivity for applications to slower dividing bacteria or challenging environmental conditions. Still, we expect microSPLiT to provide an exciting new dimension to studies of bacterial gene expression heterogeneity and community behavior facilitated by the method's potential scalability to millions of bacterial cells and single-cell resolution without need for constructing reporters.

Figures:

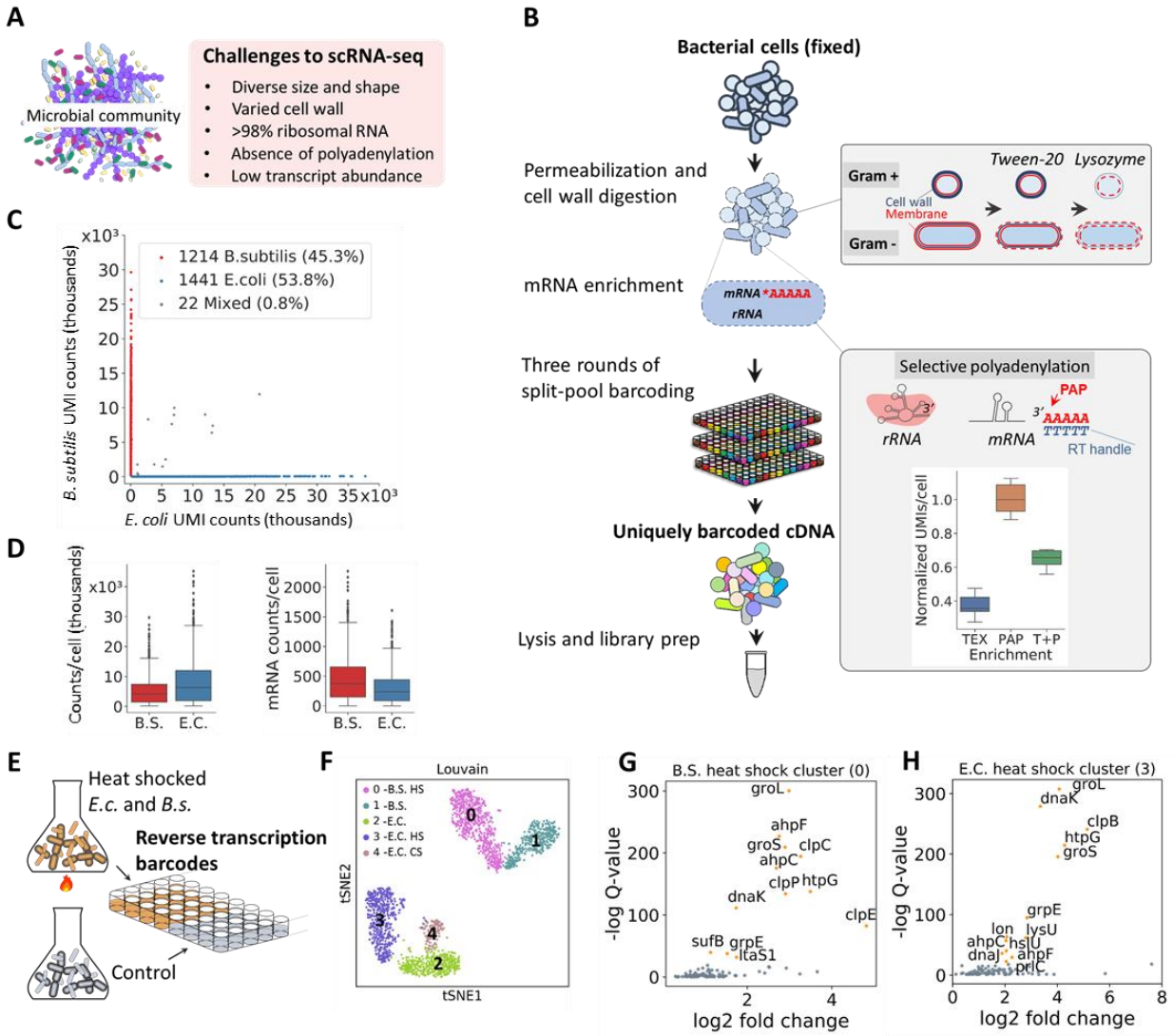

Fig. 1. microSPLiT development and validation. (A) Challenges associated with single-cell RNA sequencing of bacteria. (**B**) microSPLiT method summary. Fixed gram-positive and gram-negative bacterial cells are permeabilized with a mild detergent Tween-20 and their cell walls are degraded by lysozyme. The mRNA is then polyadenylated in-cell with *E.coli* Poly(A) Polymerase I (PAP). Inset: comparison of mRNA read counts (normalized) obtained from PAP-treated cells compared with cells treated with Terminator 5' phosphate-dependent exonuclease (TEX) and both methods consecutively (T+P). The cellular RNA then undergoes three rounds of combinatorial barcoding including in-cell reverse transcription (RT) and two in-cell ligation reactions, followed by lysis and library preparation. (**C**) Barnyard plot for the *E. coli* and *B. subtilis* species-mixing experiment. Each dot corresponds to a putative single-cell transcriptome. UMI – unique molecular identifier. (**D**) Total (in thousands) and mRNA UMI counts per cell for both species. (**E**) Summary of the heat shock experiment showing the first barcode as sample identifier. (**F**) t-stochastic neighbor embedding (t-SNE) of the data from heat shock experiment showing distinct clusters. "HS" – heat shocked. (**G**) and (**H**) Genes enriched in the B.S and E.C. heat shock clusters, respectively.

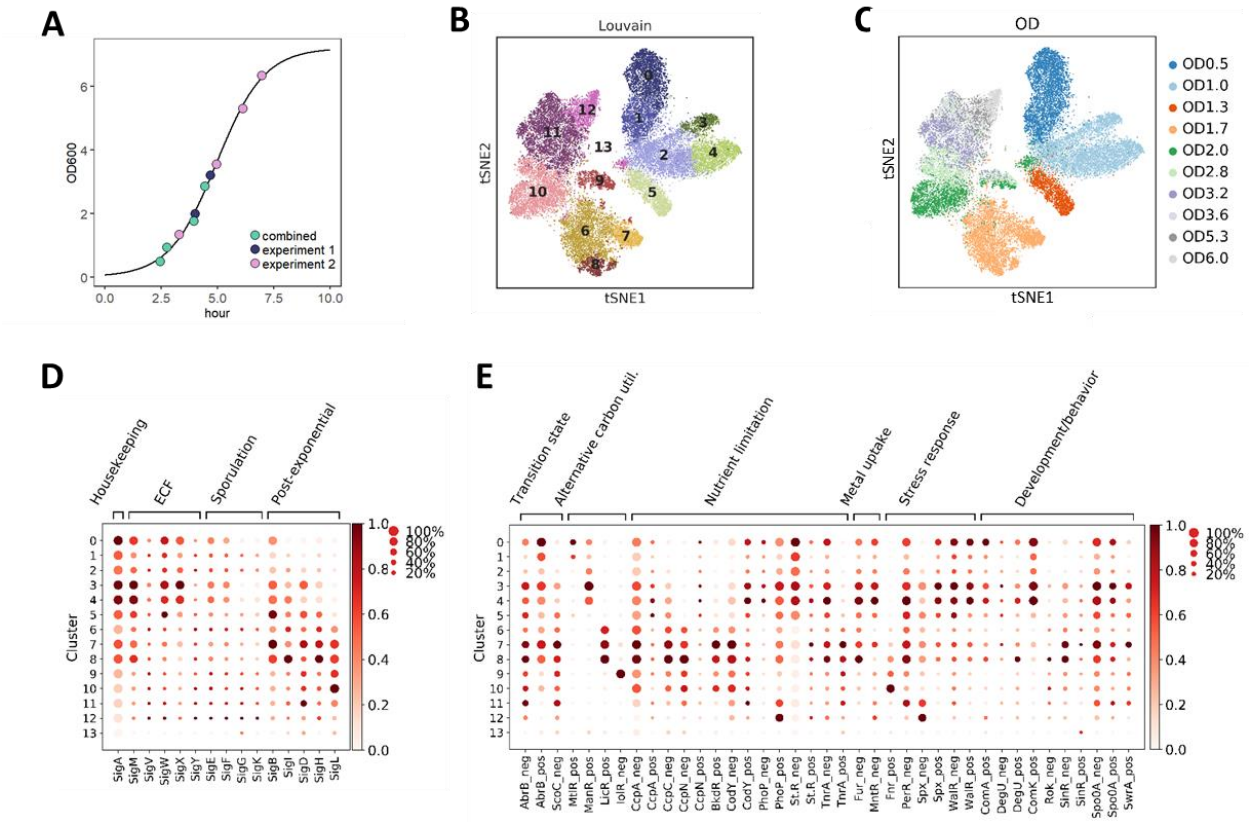

Fig. 2. microSPLiT detects global transcriptional states during *B. subtilis* growth. (A) Optical density (OD) points sampled in two experiments overlayed on the growth curve from experiment 2 (see Supplementary Text). (**B**) t-SNE embedding of the combined growth curve data colored by cluster. A total of 14 clusters were found using Louvain clustering. (**C**) Inferred normalized sigma factor activity for each cluster. The size of each dot indicates the proportion of cells in the cluster in which the sigma factor is active, while the color indicates the average activity. (**D**) Inferred normalized activity of select transcriptional regulators per cluster. “Neg” indicates that activity was calculated for the genes known to be negatively regulated by this TR, and “pos” indicates the activity was calculated for the genes positively regulated by the given TR.

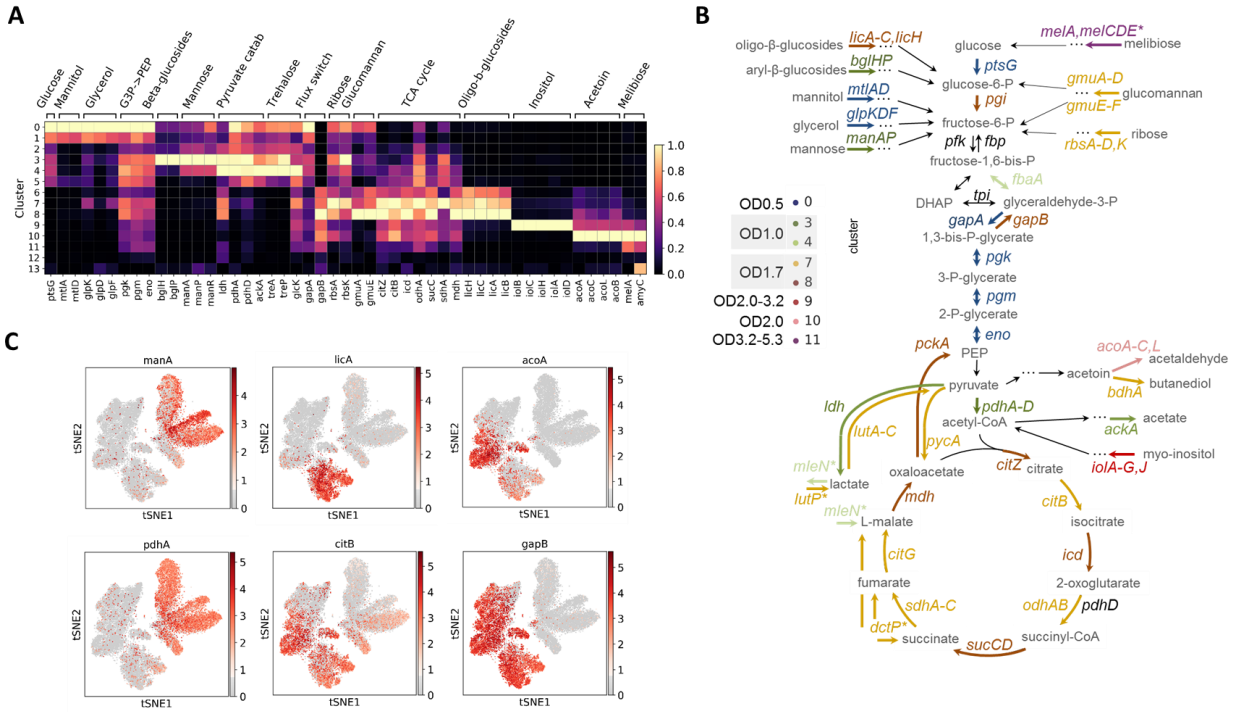

Fig. 3. Central carbon metabolism changes and alternative carbon sources utilization during *B. subtilis* growth. (A) Normalized expression of genes from select metabolic pathways and central carbon metabolism shown per cluster. Gene expression clearly shows distinct carbon utilization programs associated with different clusters and growth states. (**B**) Expression of select genes from (A) overlaid on the t-SNE plot to illustrate the differential patterns of activation. (**C**) Schematic of the central carbon metabolism pathway showing alternative carbon sources, metabolic products and genes in the pathway. The genes are color-coded according to the cluster they are highest expressed in.

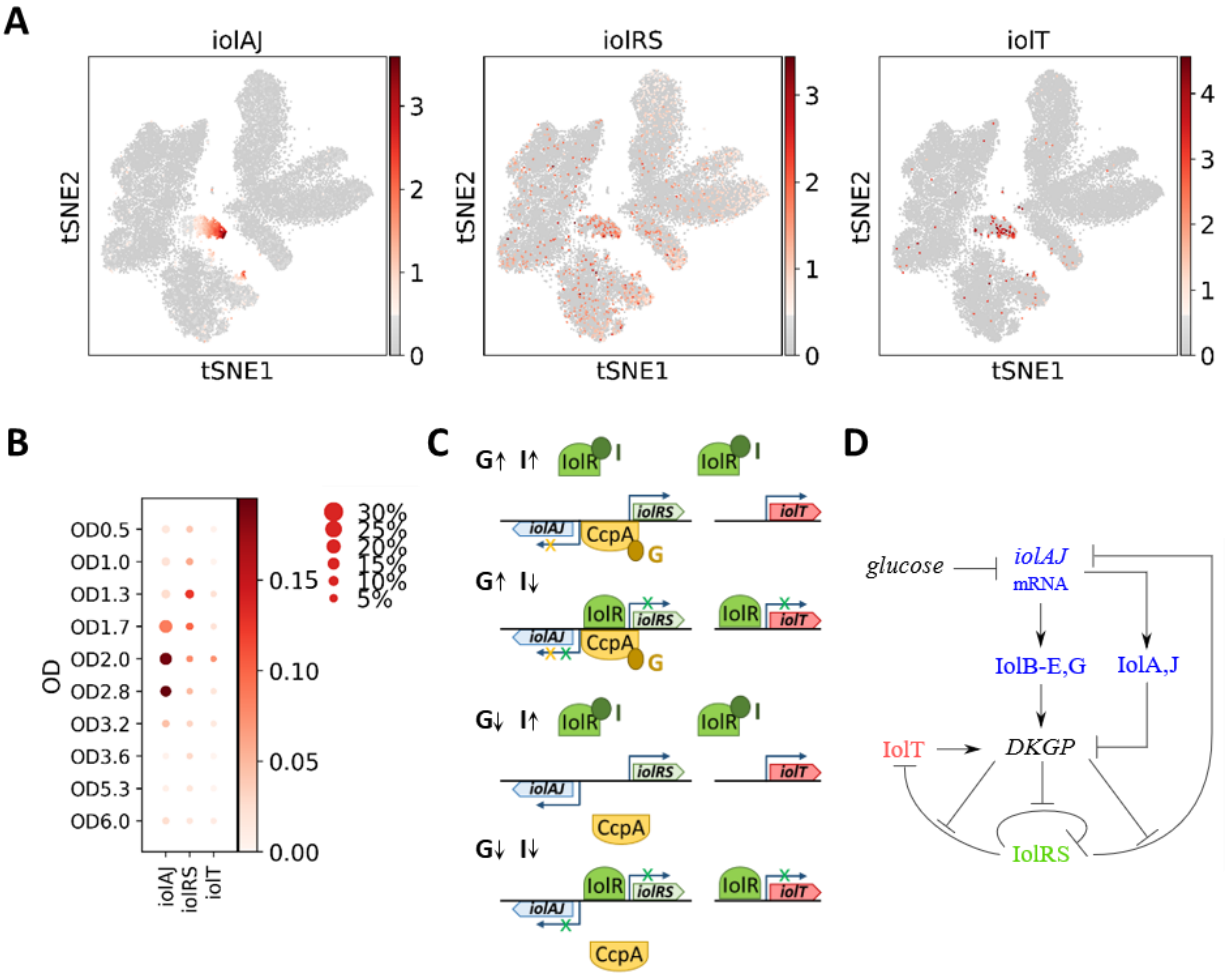

Fig. 4. Myo-inositol utilization pathway genes are expressed heterogeneously at intermediate growth stages. (A) Expression of each of the three inositol utilization operons, averaged across all genes in a given operon, and overlaid on the t-SNE plot. (**B**) Activities of the three inositol utilization operons across ODs. The size of each dot indicates the proportion of cells in each OD sample expressing any of the genes in the selected operon, while the color shows the average expression of the genes in a given operon. (**C**) Schematic of the inositol utilization operons structure along with the main regulators in different metabolic states. When both glucose (G) and inositol (I) are present in the cell, the IolR repression of all three operons is relieved, but CcpA still represses the *iolAJ* operon preventing it from metabolizing inositol (we show the signaling intermediates bound to the repressors also as G and I for simplicity). As glucose levels drop, CcpA no longer inhibits transcription of *iolAJ*, and if inositol is present, it relieves the repression of *iolRS* and *iolT*, inducing the expression of the entire metabolic network. Meanwhile, in the absence of inositol, all of the operons remain repressed regardless of the glucose levels. (**D**) A simplified network diagram representing the interactions and regulations within the inositol utilization pathway.

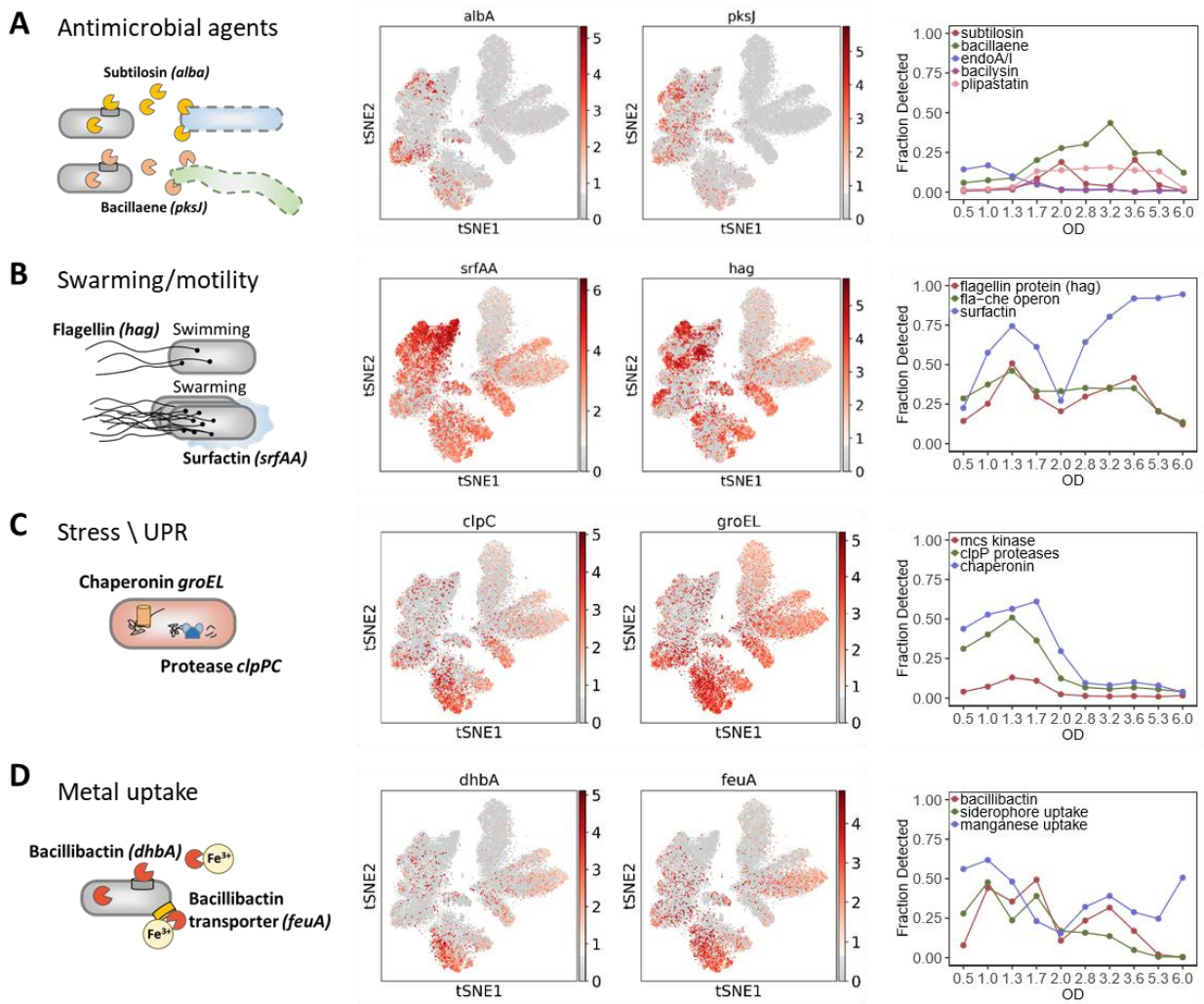

Fig. 5. Intrinsic stress responses and developmental gene expression. (A) Antimicrobial agents and endoA toxin-antitoxin secretion. Subtilosin and bacillaene action diagram. Overlay of representative subtilosin (*albA*) and bacillaene (*pksJ*) production pathway gene expression on the t-SNE. Fraction of cells expressing the indicated antimicrobials and endotoxin-antitoxin as a function of OD. **(B)** Swarming and motility genes expression. Flagellin and surfactin action diagram. Overlay of surfactin (*srfAA*) and flagellin (*hag*) gene expression on the t-SNE. Fraction of cells expressing the indicated motility operons as a function of OD. **(C)** Intrinsic stress and unfolded protein response (UPR) genes expression. GroEL chaperonin and ClpCP protease action diagram. Overlay of *clpC* and *groEL* gene expression on the t-SNE. Fraction of cells expressing the indicated genes as a function of OD. **(D)** Metal (iron and manganese) uptake genes expression. Bacillibactin siderophore and transporter action diagram. Overlay of bacillibactin (*dhbA*) and siderophore transporter (*feuA*) production pathway gene expression on the t-SNE. Fraction of cells expressing the indicated metal uptake genes as a function of OD. The genes used for fractional plots are listed in Supplementary Table S2.

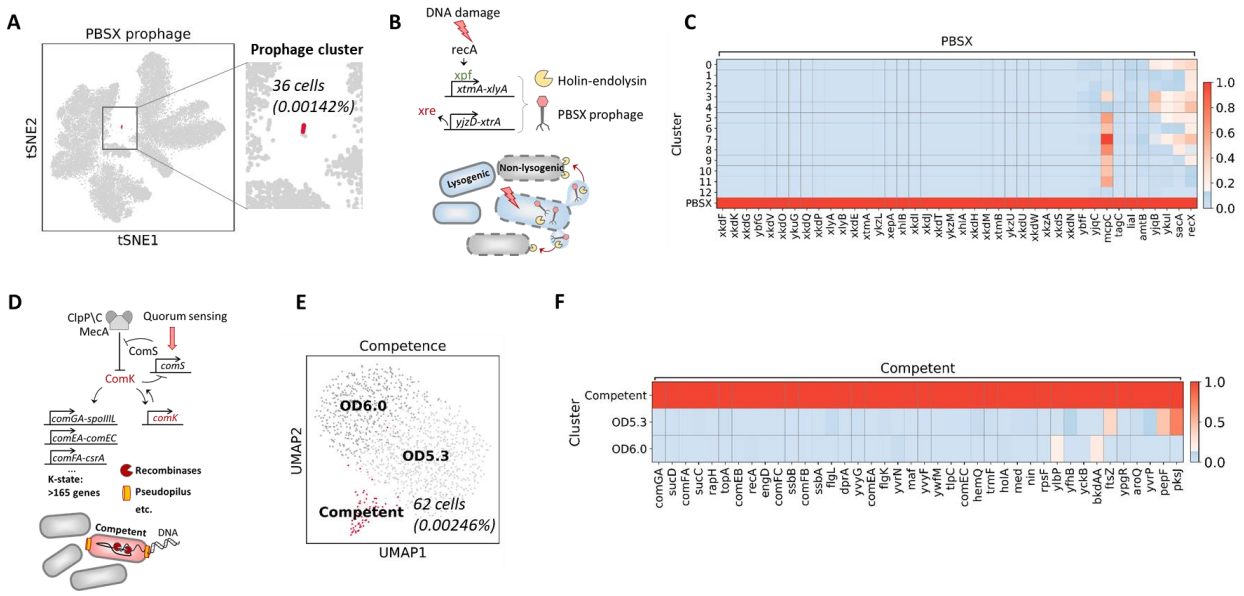

Fig. 6. Rare developmental states induced by cellular stress. (**A**) PBSX prophage induction and killing of the non-lysogenic cells. Upon DNA damage, two large prophage operons are induced and non-infectious phage particles along with endolysins are released. The latter degrade cell walls of other cells containing heterologous prophage elements. (**B**) PBSX prophage cluster (36 cells) shown on the t-SNE plot. (**C**) Genes enriched in the PBSX prophage cluster, including both prophage and host genes. (**D**) Overview of competence development. Activated by stress and quorum-sensing, ComS binds to and redirects MecA complex from degrading master competence regulator ComK. Amplified via positive autoregulation, ComK induces >165 genes involved in extracellular DNA uptake and cellular physiology. (**E**) UMAP embedding of the subclustered OD5.3 and 6.0 samples, showing the competence cluster (62 cells). (**F**) Genes enriched in the discovered competence cluster relative to the rest of the cells in OD5.3 and 6.0 samples.

References and Notes:

1. A. Raj, A. van Oudenaarden, Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*. **135**, 216–226 (2008).

2. A. Eldar, M. B. Elowitz, Functional roles for noise in genetic circuits. *Nature*. **467**, 167–173 (2010).

3. A. Kuchina, L. Espinar, T. Cagatay, A. O. Balbin, F. Zhang, A. Alvarado, J. Garcia-Ojalvo, G. M. Suel, Temporal competition between differentiation programs determines cell fate choice. *Molecular Systems Biology*. **7**, 557–557 (2014).

4. M. B. Elowitz, Stochastic Gene Expression in a Single Cell. *Science*. **297**, 1183–1186 (2002).

5. J. R. Russell, M. T. Cabeen, P. A. Wiggins, J. Paulsson, R. Losick, Noise in a phosphorelay drives stochastic entry into sporulation in Bacillus subtilis. *EMBO J.* **36**, 2856–2869 (2017).

6. J. C. W. Locke, M. B. Elowitz, Using movies to analyse gene circuit dynamics in single cells. *Nat. Rev. Microbiol.* **7**, 383–392 (2009).

7. J. Wang, L. Chen, Z. Chen, W. Zhang, RNA-seq based transcriptomic analysis of single bacterial cells. *Integr Biol (Camb)*. **7**, 1466–1476 (2015).

8. 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).

9. Y. Liu, P. Jeraldo, J. S. Jang, B. Eckloff, J. Jen, M. Walther-Antonio, Bacterial Single Cell Whole Transcriptome Amplification in Microfluidic Platform Shows Putative Gene Expression Heterogeneity. *Anal. Chem.* **91**, 8036–8044 (2019).

10. A. B. Rosenberg, C. M. Roco, R. A. Muscat, A. Kuchina, P. Sample, Z. Yao, L. Gray, D. J. Peeler, S. Mukherjee, W. Chen, S. H. Pun, D. L. Sellers, B. Tasic, G. Seelig, Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding. *Science*, eaam8999 (2018).

11. V. F. Wendisch, D. P. Zimmer, A. Khodursky, B. Peter, N. Cozzarelli, S. Kustu, Isolation of Escherichia coli mRNA and comparison of expression using mRNA and total RNA on DNA microarrays. *Anal. Biochem.* **290**, 205–213 (2001).

12. 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** (2016), doi:10.1098/rsta.2015.0069.

13. F. Arsène, T. Tomoyasu, B. Bukau, The heat shock response of Escherichia coli. *International Journal of Food Microbiology*. **55**, 3–9 (2000).

14. W. Schumann, The Bacillus subtilis heat shock stimulon. *Cell Stress Chaperones*. **8**, 207–217 (2003).

15. H. A. Thieringer, P. G. Jones, M. Inouye, Cold shock and adaptation. *BioEssays*. **20**, 49–57 (1998).

16. A. Chastanet, D. Vitkup, G.-C. Yuan, T. M. Norman, J. S. Liu, R. M. Losick, Broadly heterogeneous activation of the master regulator for sporulation in Bacillus subtilis. *Proceedings of the National Academy of Sciences*. **107**, 8486–8491 (2010).

17. Y. Luo, K. Asai, Y. Sadaie, J. D. Helmann, Transcriptomic and Phenotypic Characterization of a Bacillus subtilis Strain without Extracytoplasmic Function Factors. *Journal of Bacteriology*. **192**, 5736–5745 (2010).

18. B. Zhu, J. Stülke, SubtiWiki in 2018: from genes and proteins to functional network annotation of the model organism Bacillus subtilis. *Nucleic Acids Res*. **46**, D743–D748 (2018).

19. H. Ludwig, G. Homuth, M. Schmalisch, F. M. Dyka, M. Hecker, J. Stülke, Transcription of glycolytic genes and operons in Bacillus subtilis: evidence for the presence of multiple levels of control of the gapA operon: Regulation of glycolysis in Bacillus subtilis. *Molecular Microbiology*. **41**, 409–422 (2001).

20. J. T. Larsson, Coordinated patterns of cytochrome bd and lactate dehydrogenase expression in Bacillus subtilis. *Microbiology*. **151**, 3323–3335 (2005).

21. Y. Wei, A. A. Guffanti, M. Ito, T. A. Krulwich, *Bacillus subtilis* YqkI Is a Novel Malic/Na $^{+}$ -Lactate Antiporter That Enhances Growth on Malate at Low Protonmotive Force. *Journal of Biological Chemistry*. **275**, 30287–30292 (2000).

22. P. N. Lowe, J. A. Hodgson, R. N. Perham, Dual role of a single multienzyme complex in the oxidative decarboxylation of pyruvate and branched-chain 2-oxo acids in Bacillus subtilis. *Biochemical Journal*. **215**, 133–140 (1983).

23. F. J. Grundy, D. A. Waters, S. H. Allen, T. M. Henkin, Regulation of the Bacillus subtilis acetate kinase gene by CcpA. *Journal of Bacteriology*. **175**, 7348–7355 (1993).

24. A. Z. Rosenthal, Y. Qi, S. Hormoz, J. Park, S. H.-J. Li, M. B. Elowitz, Metabolic interactions between dynamic bacterial subpopulations. *Elife*. **7** (2018), doi:10.7554/eLife.33099.

25. S. Fillinger, S. Boschi-Muller, S. Azza, E. Dervyn, G. Branlant, S. Aymerich, Two Glyceraldehyde-3-phosphate Dehydrogenases with Opposite Physiological Roles in a Nonphotosynthetic Bacterium. *Journal of Biological Chemistry*. **275**, 14031–14037 (2000).

26. K. B. Pechter, F. M. Meyer, A. W. Serio, J. Stulke, A. L. Sonenshein, Two Roles for Aconitase in the Regulation of Tricarboxylic Acid Branch Gene Expression in Bacillus subtilis. *Journal of Bacteriology*. **195**, 1525–1537 (2013).

27. Y. Fujita, Carbon Catabolite Control of the Metabolic Network in *Bacillus subtilis*. *Bioscience, Biotechnology, and Biochemistry*. **73**, 245–259 (2009).

28. R. J. Kleijn, J. M. Buescher, L. Le Chat, M. Jules, S. Aymerich, U. Sauer, Metabolic Fluxes during Strong Carbon Catabolite Repression by Malate in *Bacillus subtilis*. *Journal of Biological Chemistry*. **285**, 1587–1596 (2010).

29. K. D. Singh, M. H. Schmalisch, J. Stulke, B. Gorke, Carbon Catabolite Repression in Bacillus subtilis: Quantitative Analysis of Repression Exerted by Different Carbon Sources. *Journal of Bacteriology*. **190**, 7275–7284 (2008).

30. K. Yoshida, M. Yamaguchi, T. Morinaga, M. Kinehara, M. Ikeuchi, H. Ashida, Y. Fujita, myo-Inositol Catabolism in Bacillus subtilis. *J. Biol. Chem.* **283**, 10415–10424 (2008).

31. K. I. Yoshida, D. Aoyama, I. Ishio, T. Shibayama, Y. Fujita, Organization and transcription of the myo-inositol operon, iol, of Bacillus subtilis. *Journal of Bacteriology*. **179**, 4591–4598 (1997).

32. K.-I. Yoshida, T. Shibayama, D. Aoyama, Y. Fujita, Interaction of a Repressor and its Binding Sites for Regulation of the Bacillussubtilis iol Divergon. *Journal of Molecular Biology*. **285**, 917–929 (1999).

33. B. C. Marciniak, M. Pabijaniak, A. de Jong, R. Dűhring, G. Seidel, W. Hillen, O. P. Kuipers, High- and low-affinity cre boxes for CcpA binding in Bacillus subtilis revealed by genome-wide analysis. *BMC Genomics*. **13**, 401 (2012).

34. D. Lopez, H. Vlamakis, R. Kolter, Generation of multiple cell types in *Bacillus subtilis*. *FEMS Microbiology Reviews*. **33**, 152–163 (2009).

35. E. Ben-Jacob, D. Schultz, Bacteria determine fate by playing dice with controlled odds. *Proceedings of the National Academy of Sciences*. **107**, 13197–13198 (2010).

36. E. T. Granato, T. A. Meiller-Legrand, K. R. Foster, The Evolution and Ecology of Bacterial Warfare. *Current Biology*. **29**, R521–R537 (2019).

37. C. E. Shelburne, F. Y. An, V. Dholpe, A. Ramamoorthy, D. E. Lopatin, M. S. Lantz, The spectrum of antimicrobial activity of the bacteriocin subtilosin A. *J Antimicrob Chemother*. **59**, 297–300 (2007).

38. P. S. Patel, S. Huang, S. Fisher, D. Pirnik, C. Aklonis, L. Dean, E. Meyers, P. Fernandes, F. Mayerl, Bacillaene, a novel inhibitor of procaryotic protein synthesis produced by Bacillus subtilis: production, taxonomy, isolation, physico-chemical characterization and biological activity. *J. Antibiot.* **48**, 997–1003 (1995).

39. L. Gao, J. Han, H. Liu, X. Qu, Z. Lu, X. Bie, Plipastatin and surfactin coproduction by Bacillus subtilis pB2-L and their effects on microorganisms. *Antonie van Leeuwenhoek*. **110**, 1007–1018 (2017).

40. J. van Gestel, M. Ackermann, A. Wagner, Microbial life cycles link global modularity in regulation to mosaic evolution. *Nature Ecology & Evolution*. **3**, 1184–1196 (2019).

41. D. B. Kearns, R. Losick, Cell population heterogeneity during growth of Bacillus subtilis. *Genes Dev.* **19**, 3083–3094 (2005).

42. D. López, H. Vlamakis, R. Losick, R. Kolter, Paracrine signaling in a bacterium. *Genes Dev.* **23**, 1631–1638 (2009).

43. D. R. Zeigler, Z. Prágai, S. Rodriguez, B. Chevreux, A. Muffler, T. Albert, R. Bai, M. Wyss, J. B. Perkins, The Origins of 168, W23, and Other Bacillus subtilis Legacy Strains. *J Bacteriol*. **190**, 6983–6995 (2008).

44. E. Krüger, E. Witt, S. Ohlmeier, R. Hanschke, M. Hecker, The clp proteases of Bacillus subtilis are directly involved in degradation of misfolded proteins. *J. Bacteriol.* **182**, 3259–3265 (2000).

45. E.-J. Blom, A. N. J. A. Ridder, A. T. Lulko, J. B. T. M. Roerdink, O. P. Kuipers, Time-Resolved Transcriptomics and Bioinformatic Analyses Reveal Intrinsic Stress Responses during Batch Culture of Bacillus subtilis. *PLOS ONE*. **6**, e27160 (2011).

46. D. Schultz, R. Schlüter, U. Gerth, M. Lalk, Metabolic Perturbations in a Bacillus subtilis clpP Mutant during Glucose Starvation. *Metabolites*. **7**, 63 (2017).

47. J. D. Helmann, Specificity of Metal Sensing: Iron and Manganese Homeostasis in Bacillus subtilis. *J. Biol. Chem.* **289**, 28112–28120 (2014).

48. E. Guedon, C. M. Moore, Q. Que, T. Wang, R. W. Ye, J. D. Helmann, The global transcriptional response of Bacillus subtilis to manganese involves the MntR, Fur, TnrA and sigmaB regulons. *Mol. Microbiol.* **49**, 1477–1491 (2003).

49. N. Baichoo, T. Wang, R. Ye, J. D. Helmann, Global analysis of the Bacillus subtilis Fur regulon and the iron starvation stimulon: Bacillus subtilis Fur regulon. *Molecular Microbiology*. **45**, 1613–1629 (2002).

50. H. Pi, J. D. Helmann, Sequential induction of Fur-regulated genes in response to iron limitation in Bacillus subtilis. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12785–12790 (2017).

51. E. Mhatre, A. Troszok, R. Gallegos-Monterrosa, S. Lindstädt, T. Hölscher, O. P. Kuipers, Á. T. Kovács, The impact of manganese on biofilm development of Bacillus subtilis. *Microbiology (Reading, Engl.)*. **162**, 1468–1478 (2016).

52. H. E. Wood, M. T. Dawson, K. M. Devine, D. J. McConnell, Characterization of PBSX, a defective prophage of Bacillus subtilis. *J. Bacteriol.* **172**, 2667–2674 (1990).

53. A. I. Goranov, E. Kuester-Schoeck, J. D. Wang, A. D. Grossman, Characterization of the Global Transcriptional Responses to Different Types of DNA Damage and Disruption of Replication in Bacillus subtilis. *Journal of Bacteriology*. **188**, 5595–5605 (2006).

54. L. M. Anderson, K. F. Bott, DNA packaging by the Bacillus subtilis defective bacteriophage PBSX. *J. Virol.* **54**, 773–780 (1985).

55. R. Shingaki, Y. Kasahara, T. Inoue, S. Kokeguchi, K. Fukui, Chromosome DNA fragmentation and excretion caused by defective prophage gene expression in the early-exponential-phase culture of *Bacillus subtilis*. *Canadian Journal of Microbiology*. **49**, 313–325 (2003).

56. G. M. Süel, J. Garcia-Ojalvo, L. M. Liberman, M. B. Elowitz, An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*. **440**, 545–550 (2006).

57. R. M. Berka, J. Hahn, M. Albano, I. Draskovic, M. Persuh, X. Cui, A. Sloma, W. Widner, D. Dubnau, Microarray analysis of the Bacillus subtilis K-state: genome-wide expression changes dependent on ComK. *Molecular Microbiology*. **43**, 1331–1345 (2002).

58. M. Ogura, H. Yamaguchi, K. Kobayashi, N. Ogasawara, Y. Fujita, T. Tanaka, Whole-genome analysis of genes regulated by the Bacillus subtilis competence transcription factor ComK. *J. Bacteriol.* **184**, 2344–2351 (2002).

59. W. K. Smits, C. Bongiorni, J.-W. Veening, L. W. Hamoen, O. P. Kuipers, M. Perego, Temporal separation of distinct differentiation pathways by a dual specificity Rap-Phr system in *Bacillus subtilis*. *Molecular Microbiology*. **65**, 103–120 (2007).

60. T. Yadav, B. Carrasco, A. R. Myers, N. P. George, J. L. Keck, J. C. Alonso, Genetic recombination in Bacillus subtilis : a division of labor between two single-strand DNA-binding proteins. *Nucleic Acids Research*. **40**, 5546–5559 (2012).

Acknowledgments: This work was supported by NIH grants NIH R01HG009136 and R01HG009892 to G.S. All relevant sequencing files will be deposited to Sequence Read Archive (SRA). Processed data will be submitted to GEO. This work was supported by NIH A.K., L.B. W.D. and G.S. have filed a patent application for microSPLiT. C.R., A.B.R and G.S. are co-founders of Split Biosciences, a single-cell RNA sequencing company.

Supplementary Materials:

Materials and Methods

Figures S1-S16

Tables S1-S3

Tables (Auxiliary Excel Files) S4

Supplementary References (*1-28*)