A quantitative model for the transcriptional landscape of the bacterial cell cycle

Andrew Pountain[1], Peien Jiang[1,2], Magdalena Podkowik[3], Bo Shopsin[3,4], Victor J. Torres[4], Itai Yanai[1,5]

[1]Institute for Computational Medicine, NYU Grossman School of Medicine, New York, NY USA
[2]Department of Biology, New York University, New York, NY USA
[3]Department of Medicine, Division of Infectious Diseases, NYU Grossman School of Medicine, New York, NY, USA
[4]Department of Microbiology, NYU Grossman School of Medicine, New York, NY USA
[5]Department of Biochemistry and Molecular Pharmacology, NYU Langone Health, New York, NY, USA

Abstract

Regulation of gene activity during the cell cycle is fundamental to bacterial replication but is challenging to study in unperturbed, asynchronous bacterial populations. Using single cell RNA-sequencing of heterogeneous *Staphylococcus aureus* populations, we uncovered a global gene expression pattern dominated by chromosomal position. We show that this pattern results from the effect of DNA replication on gene expression, and in *Escherichia coli*, changes under different growth rates and modes of replication. By constructing a quantitative model in each species that links replication to cell cycle gene expression, we identified divergent genes that may be instead subject to distinct regulation, and applied this cell cycle framework to characterize heterogeneity in responses to antibiotic stress. Our approach reveals a highly dynamic cell cycle transcriptional landscape and may be broadly applicable across species.

Introduction

Growth and division in bacteria requires coordinated regulation of gene function at the level of cytoplasm synthesis, cell wall production and division, and genome replication and segregation. Understanding how these processes are controlled and intersect with other cellular functions will yield fundamental insights into the physiology of these organisms. While control of transcript abundance is a fundamental regulatory mechanism, analysis of cell cycle transcript dynamics for most bacteria has not been possible. A central reason for this has been the difficulty of synchronizing inherently noisy bacterial cell cycles to enable measurement of cell cycle processes as a function of time. Hence global analysis of cell cycle gene expression has been limited to organisms such as *Caulobacter crescentus (1–3)* where natural biological features facilitate synchronization, or to populations synchronized by bulk treatments such as starvation (*4*) or temperature shift (*5*), although the efficacy of such batch synchronization methods has been called into question (*6*).

Single-cell RNA sequencing (scRNA-seq) has transformed the analysis of heterogeneous populations of eukaryotic cells by identifying a diversity of hitherto unanticipated cell types and states (*7–9*). The recent development of methods for scRNA-seq in prokaryotes (*10–12*)

represents a major technological advance, but the analysis of the resulting data is challenged by the low mRNA content of bacteria that leads to very few transcripts detected per cell. Here we apply one method, PETRI-seq (*10*), to acquire the transcriptomes of large numbers of individual bacteria and develop a new analysis approach to reveal a global pattern of gene covariance dependent on genes' chromosomal position. We show that this pattern is driven by the effect of DNA replication on gene expression and that modeling this effect allows, for the first time, resolution of the cell cycle transcriptome from unsynchronized bacterial populations. Applying this approach to two highly medically important but unrelated bacterial pathogens, the Gram-negative rod *Eschericha coli* (*E. coli*) and the Gram-positive coccus *Staphylococcus aureus* (*S. aureus*), we identify genes with expression patterns that diverge from this replication-driven effect, with roles from core bacterial physiological processes to stress responses. We further show how our framework can be used to understand cell cycle variation and proliferation state as a source of heterogeneity in responses to antibiotic treatment.

This study provides a powerful new analysis approach to apply methods such as PETRI-seq to interrogate bacterial expression dynamics. The global replication-driven expression dynamics that we uncover may not only act as a regulatory mechanism in some cases, but also form a useful framework and an essential consideration for interpretation of studies of bacterial gene expression heterogeneity. Moreover, our work raises questions about the regulatory function of these transcriptional dynamics, as well as their contribution to population heterogeneity in key phenotypes from stress tolerance to virulence.

Results

Global gene expression in bacterial populations is shaped by chromosome position.
To investigate transcriptional heterogeneity in proliferating bacterial populations, we applied PETRI-seq (*10*) to 73,053 *S. aureus* cells in exponential phase (Fig. 1A). By including further methodological optimizations (see Materials & Methods), we detected on average 135 transcripts per cell (measured as unique molecular identifiers (UMI)) (Fig. 1B). To denoise the data, we applied the scVI method, an unsupervised deep learning approach (*13*). This allowed us to recover previously reported gene-gene correlations, including a general covariance of genes within operons (Fig. 1C). When we expanded this analysis to chromosome-wide gene-gene correlations, we discovered a striking 'X-shaped' pattern of gene expression covariance (Fig. 1D). Beyond the expected diagonal reflecting coordinated gene expression at the level of operons, the anti-diagonal reflected correlations between genes at a similar distance from the origin of replication, between the "arms" of the circular chromosome, as well as a correlation between genes at the origin and terminus (Fig. 1E). This pattern was strengthened by averaging expression into 50 kb bins by chromosome position (Fig. 1F), and was reproducible in a second independent dataset under the same conditions of 21,257 cells (Figure S1, Table S1, Dataset D4), and also detectable without scVI, although the signal was substantially noisier (Fig. 1G). Interestingly, the pattern was abolished when we applied this analysis to 55,894 cells in stationary phase, suggesting that this pattern is a property of actively proliferating cells (Fig. 1H).

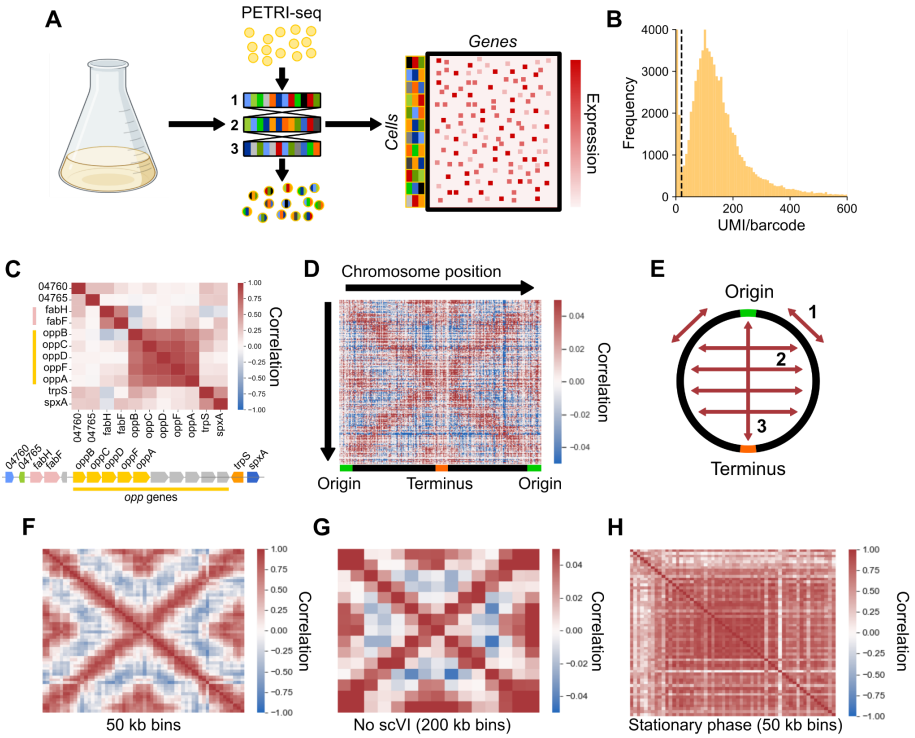

Figure 1: scRNA-seq of proliferating *S. aureus* reveals a global pattern of gene expression covariance. A) PETRI-seq workflow. *S. aureus* cells were fixed and permeabilized, then subjected to three rounds of cDNA barcoding to give transcripts of each cell a unique barcode combination. **B)** Histogram of total UMI per cell barcode in exponential *S. aureus*. The dotted line indicates the minimum threshold of 20 UMI/cell. **C)** Local operon structure is captured by gene-gene correlations (Spearman's *r*). Operons are indicated by shared colors of genes. Gray genes indicate those removed by low-count filtering. Names of *SAUSA300_RS04760* and *SAUSA300_RS04765* are truncated. **D)** Global gene-gene correlations reflect chromosomal position. Spearman correlations were calculated based on scVI-modeled expression. Genes are ordered by chromosomal position, with the replication origin at zero. **E)** Schematic of positive correlation patterns in **(D)**. Correlations are between 1) neighboring genes, 2) genes equidistant from the origin of replication, and 3) direct origin-terminus correlations. **F)** Correlations as in **(D)** but with scVI-modeled expression scaled to *z*-scores then averaged in 50 kb bins. **G)** Correlations as in **(D)** but without modeling with scVI and averaging expression in 200 kb bins. **H)** Correlations as in **(F)** but in stationary phase *S. aureus*.

***The rate of replication initiation determines transcript heterogeneity patterns*.**

As we observed correlations among genes that are equidistant from the origin of replication and cells in stationary phase did not show such correlations, we hypothesized that the 'X-shaped pattern' is related to DNA replication. To test this, we developed a simulation of gene copy number effects resulting from DNA replication patterns at different growth rates, to investigate their influence on gene-gene correlations (Fig. S2, Materials & Methods). First, we simulated a scenario in which the time between replication initiation events (i.e. the doubling time t_d) is on average equal to the time to replicate the whole chromosome (known as the "C-period") (Fig. 2A). This closely recapitulated the X-shaped pattern, supporting a role for replication.

A replication-driven pattern could in theory arise from a contaminating effect of genomic DNA contributing to the expression matrix. However, samples are DNase-treated and barcodes are introduced using reverse transcriptase. Furthermore, we observed that 1) gene expression varied over ~5 orders of magnitude, inconsistent with genomic DNA (Fig. S3A); 2) the X-shaped pattern is stronger with more highly expressed genes (Fig. S3B & C, the opposite of what would be predicted with background genomic DNA contamination); 3) we also identified the X-shaped pattern in a previously published dataset of bulk-measured, synchronized *C. crescentus* (*2*) (Fig. S3D). Together, this evidence demonstrates that it is very unlikely that the pattern could result purely from genomic DNA contamination.

Since DNA replication patterns are sensitive to replication rate, we predicted that gene-gene correlations will vary with population growth rate. In *E. coli*, it has long been established that overlapping cycles of replication occur simultaneously at high rates of proliferation, whereas at slower proliferation rates one round of replication finishes before the next one starts (*14*). We therefore measured the doubling times (t_d) of *E. coli* grown in three medium conditions (Fig. 2B): LB (26.0 min), M9 minimal medium with 0.4% glucose and 0.2% amino acids (M9 + Glc + AA, 39.4 min), and M9 medium with 0.4% glucose only (M9 + Glc, 69.1 min). A C-period of 42 minutes has also been reported for the K-12 MG1655 strain used here (*15*, *16*). These parameters allowed us to simulate each growth rate and predict the effect on gene-gene correlations (Fig. 2C). At an intermediate rate of growth (t_d = 39.4 min), we predicted a correlation pattern similar to that observed for *S. aureus* (Fig. 1D). However, simulating faster growth produced a nested “multi-X” pattern resulting from overlapping cycles of replication, and slower growth greatly reduced origin-terminus correlations (Fig. 2C).

When we compared these predictions to the observed data for *E. coli* grown under the three conditions, we observed a strikingly close correspondence between simulated and observed expression patterns (Fig. 2D). Correlations became less defined at slower growth rates, although this may reflect noise due to lower transcript counts (Fig. S1C), reflecting lower RNA content at slower growth rates (*17*). The M9 + Glc condition further resembled bulk RNA-seq of synchronized *C. crescentus* (Fig. S3D) (*2*), which undergoes a single round of replication prior to asymmetric division (*1*), similar to the situation for slower-growing *E. coli*. Finally, if this pattern is indeed driven by the effect of gene copy number on expression levels (as assumed in our simulation), this should create a relationship between origin distance and expression levels. Despite high variation in intrinsic promoter activity, we found that on average expression decreased with distance from the origin, and this effect was stronger at faster growth rates (Fig. S4).

To further test our ability to predict global correlations from expected replication patterns, we examined strains in which normal replication is perturbed. We compared wild-type *E. coli* in LB to two strains that had ectopic origins of replication at either 9 o’clock (*oriX*) or 3 o’clock (*oriZ*) positions in addition to *oriC* (*18*–*20*). In these strains, replication initiates simultaneously at both origins, and uses the same terminus, *ter* (*18*). Our simulation predicted highly perturbed correlation patterns that mirrored each other, as the ectopic origins were nearly equidistant from *oriC* on each side of the chromosome (Fig. 2E). Again, observed patterns matched closely with

our predictions (Fig. 2E). These results support the notion that DNA replication produces complex effects on transcriptional heterogeneity and these effects are sensitive to proliferation rate and other perturbations.

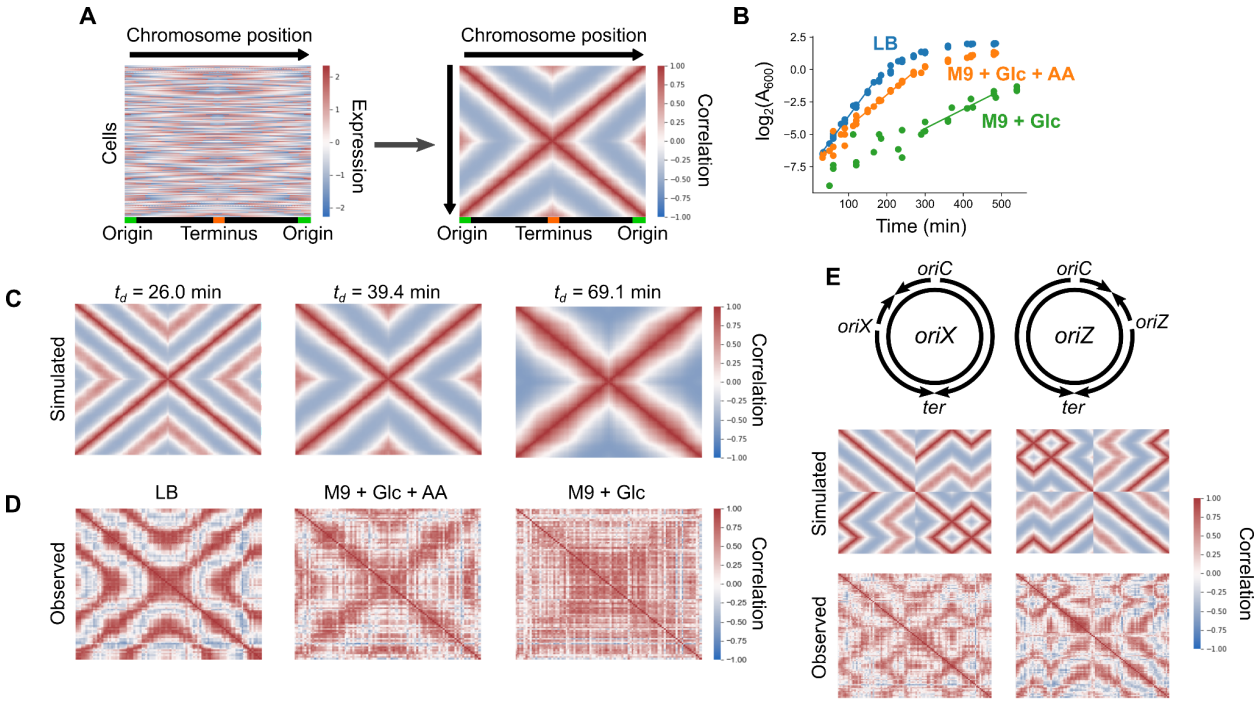

Figure 2: Global gene expression patterns are determined by doubling time-dependent DNA replication patterns. A) Simulation of DNA copy number effects predicts the global gene covariance pattern. For 1,000 simulated, unsynchronized cells where the doubling time t_d is equal to the C-period, the normalized, scaled gene expression matrix (*left*) is used to calculate gene-gene correlations (*right*). **B)** Growth of *E. coli* in three conditions. Doubling times were calculated based on the linear portions of growth (marked as fitted lines). **C)** Simulated correlation patterns in unsynchronized cells at three different growth rates. **D)** Spearman correlations between scaled data averaged into 50 kb bins, as for Fig. 1F but for *E. coli* grown at three growth rates. **E)** Introducing ectopic origins of replication leads to predictable perturbations in gene expression heterogeneity. *Top*: schematic of predicted replication patterns based on previous studies (*18–20*). *Middle*: Predicted correlation patterns based on the copy number simulation model. *Bottom*: Real correlation patterns in *oriX* and *oriZ* mutant strains, as in Fig. 2D.

Quantitative modeling of the replication effect reveals the dynamics of gene expression throughout the E. coli *cell cycle.*

Since DNA replication exerts a strong influence over gene expression, we hypothesized that this effect can be used to resolve a cell's position within the replication cycle given only its transcriptome. To study the distribution of cellular states in a population of cells, we projected LB-grown *E. coli* cells in two dimensions by uniform manifold approximation and projection (UMAP). When we performed UMAP on expression averaged by chromosomal position (which we saw strengthened global correlation patterns, Fig. 1F), cells arranged into a "wheel" shape (Fig. 3A). To determine the order of cells along this wheel, we calculated cells' angle θ_c between UMAP coordinates (Fig. 3A). Examining gene expression as a function of θ_c, we observed

waves of gene expression progressing from the origin to the terminus (Fig. 3B), suggesting that cells' positions on this wheel reveal their replication state.

As we observed that expression for most genes is strongly influenced by a cell's replication state, we reasoned that we should also be able to order genes by their cell cycle expression and that this would generally reflect their order of replication. To do this, we projected the genes themselves into two dimensions to derive a gene angle, θ_g (Fig. 3C). Ordering genes by θ_g, we observed a close relationship of this metric and the distance from the origin of replication (Fig. 3D), suggesting that θ_g does indeed capture the order of replication. However, we also detected "wrapping" such that θ_g started back again at zero after passing through a full cycle. While *E. coli* under all growth rates exhibited an origin-angle relationship, the gradient of change of θ_g with respect to origin distance decreased with increasing doubling time (Fig. 3D). This gradient is convertible to an "overlap fraction", the fraction of one round of replication happening before the previous one has finished. These overlapping rounds mean that distant genes are simultaneously co-replicated, resulting in the observed gene-gene correlation patterns.

The two parameters – the cell angle θ_c and the gene angle θ_g – led us to construct a predictive model for expression of a given gene (by θ_g) at a given point in the cell cycle (by θ_c) (Fig. 3F). Within this simplified model, as cell state rotates around θ_c, the genes maximally expressed simultaneously rotate by θ_g (Fig. 3G). Thus based on a given pattern of gene expression, the model infers the state of the cell along the cell cycle; conversely, for a particular cell cycle state, the model infers an expected gene expression pattern.

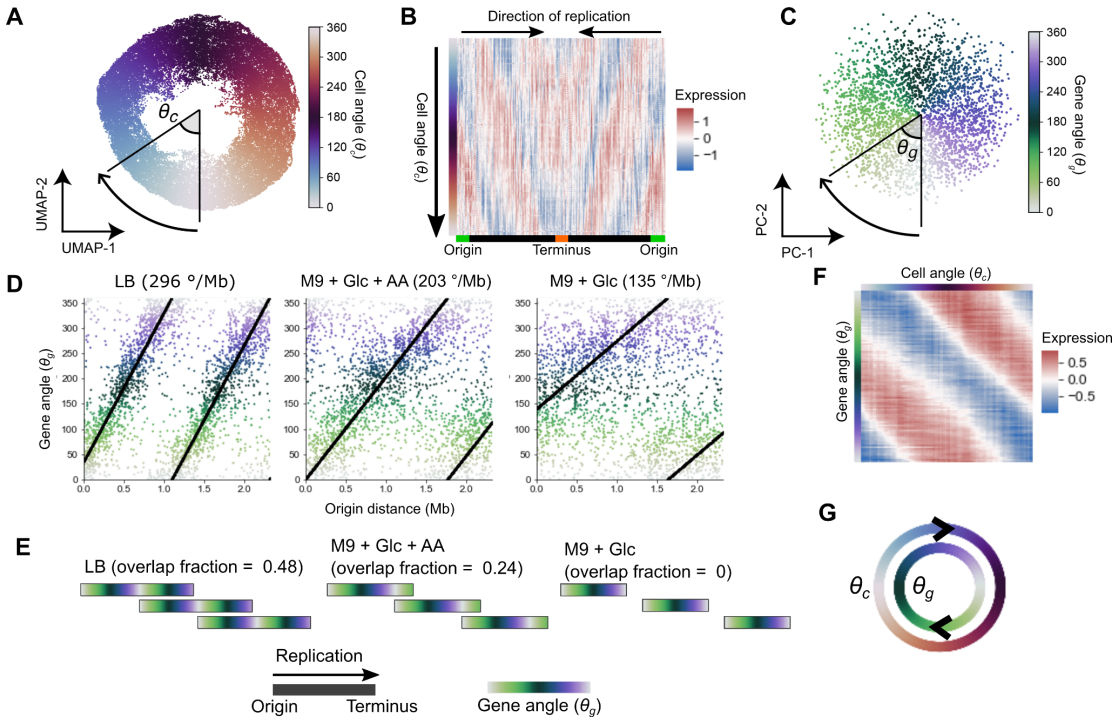

Figure 3: Ordering expression by cell angle and gene angle establishes a quantitative model of cell cycle gene expression. A) UMAP of LB-grown *E. coli* with expression averaged in 100 kb bins by chromosome position. Cell angle θ_c is the angle between UMAP dimensions relative to the center. For UMAP without averaging, see Fig. S5A. **B)** Heatmap of scaled gene expression averaged in 100 bins by θ_c. **C)** Derivation of gene angle θ_g. Principal component analysis was performed on the transpose of the matrix in **(B)**, and θ_g was defined as the angle between principal components (PCs) 1 and 2. Genes form a wheel in UMAP (Fig. S5B). **D)** The relationship between θ_g and origin distance for each *E. coli* growth condition. The gradient (see Materials & Methods) is indicated. **E)** Predicted replication patterns. Slower growth leads to a reduced fraction of overlapping expression. Each bar represents a single round of replication. Overlapping rounds lead to shared θ_g in simultaneously-replicated chromosomal regions. **F)** Expression in LB-grown *E. coli* is averaged in 100 bins first by θ_c then by θ_g. **G)** A model of cell cycle dynamics parameterized by cycling of cells by θ_c with concurrent cycling of gene expression by θ_g.

Correcting for replication patterns reveals evidence of cell cycle gene regulation in E. coli.

While most genes appear to follow a predictable pattern of cell cycle gene expression, identifying genes that display divergent behavior could reveal cell cycle regulation by other factors. We achieved this by developing a model to predict gene expression dynamics resulting from DNA replication alone based on a gene's distance from the replication origin and the relationship between cell angle θ_c, gene angle θ_g, and gene expression (Fig. 4A, see Materials & Methods). Overall, we found a moderately strong correlation of this prediction with the observed data (Pearson's r = 0.59, Fig. S5C). Crucially, this prediction allows us to "correct" for the effect of replication on gene expression by subtracting replication-predicted expression from the observed expression, largely eradicating the global chromosome position-dependent correlation pattern (Fig. 4B). Next, we computed the divergence of expression dynamics from our global model as the standard deviation of residual cell cycle expression after correcting for replication-predicted expression ($\sigma_{corrected}$). This value is highly concordant across replicates for highly variable genes (Pearson's r = 0.80, Fig. 4C, Fig. S5E). For genes with low $\sigma_{corrected}$ values, the observed expression closely tracked the replication-predicted expression (Fig. S5F). On the other hand, 70 genes that showed reproducible high $\sigma_{corrected}$ values were classified as "replication-divergent".

To understand the dynamics of the replication-divergent genes, we compared the observed cell cycle expression pattern to that expected based on replication alone (Fig. 4D). Overall, genes in the same operon showed highly similar dynamics (black lines in Fig. 4D). To aid interpretation, we shifted θ_c so that the zero angle was the most likely point for a replication initiation event (see Materials & Methods). Several replication-related and origin-proximal genes show a decrease in transcript abundance early in the cycle. Among these, the key regulator of replication initiation *dnaA*, and several other of these genes, have been shown to transiently drop in transcript abundance around the time of initiation (*21–24*). In contrast, the DNA polymerase subunit DnaE is required for replication elongation (*25*), explaining its increase concurrent with the *dnaA* drop. Similarly, initiation of a new replication cycle increases the DNA synthesis rate, leading to increased demand and transcription of the ribonucleotide reductase (*nrd*) genes to produce dNTP monomers, as described previously (*22*, *23*, *26*). We also observed a transient drop in β (*rpoB*) and β' (*rpoC*) subunits of RNA polymerase. Previously, it

was reported that the α (*rpoA*) subunit was not cell cycle-regulated (*21*). In keeping with this, *rpoA*, which is in a separate operon from *rpoB-rpoC*, was not in our replication-divergent gene list and its expression closely tracks the pattern expected from replication.

Besides genes involved in replication and transcription, two major functional groups dominated the replication-divergent gene list: those involved in cell surface synthesis and division (23 genes, Fig. 4E) and energy production and conversion (22 genes). The large operon of cell wall biosynthesis and division genes including *ftsZ* increased early and dropped towards the end of the cycle. Cell cycle fluctuation of the chromosome partition protein *mukB* has been reported previously (*23*). However, we also saw dynamic regulation of numerous operons involved in lipopolysaccharide biosynthesis and outer membrane function (*skp-lpxD*, *pqiB*, *ugd-wbbK-wbbJ-wbbI-wbbH-glf*, *waaU-waaY-waaJ*). Unexpectedly, we observed dynamic transcription of numerous genes involved in energy metabolism, particularly pyruvate dehydrogenase (*aceE-aceF-lpd*), α-ketoglutarate dehydrogenase (*sucB*), and succinyl-CoA synthetase (*sucC-sucD*) in the tricarboxylic acid cycle and electron transport chain complexes cytochrome bo_3 oxidase (*cyo*), NADH:ubiquinone oxidoreductase (*nuo*), and the F_0F1-ATP synthase (*atp*). We also identified large, replication-divergent fluctuations in the *nemRA* operon, encoding an electrophile-regulated transcriptional repressor, NemR, and a reactive electrophile-detoxifying enzyme, NemA (*27–29*). Typically, cell cycle variation in expression of these divergent genes amounted to around 3-4-fold changes in expression (Fig. S5G), although some genes such as *dnaA* showed particularly strong silencing at specific points. These dynamics were in general highly reproducible in a second dataset (Fig. S6, Dataset D2).

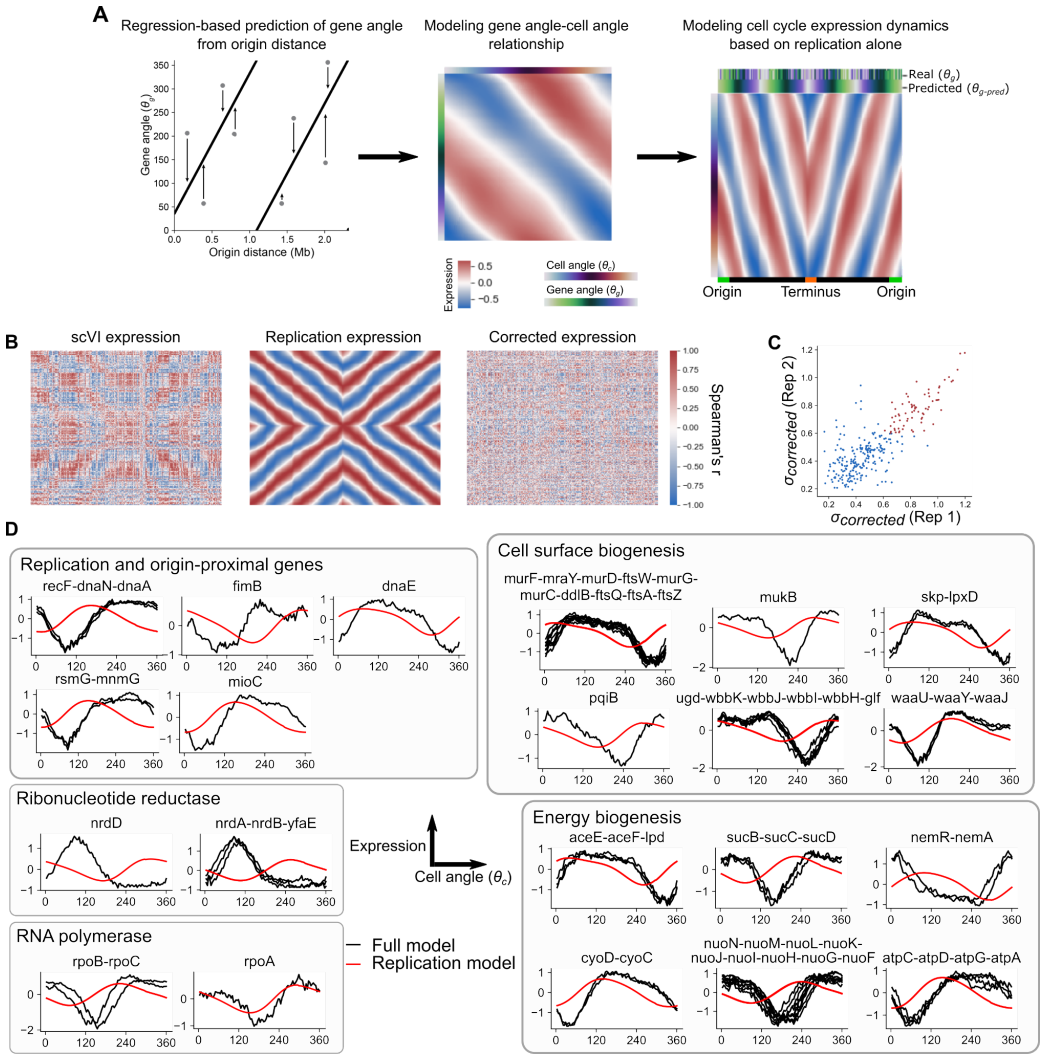

Figure 4: Modeling the replication effect allows identification of replication-divergent genes. A) Pipeline for predicting replication-only transcriptional dynamics. A regression model predicts gene angle $\theta_{g\text{-}pred}$ based on origin distance alone (*left*) and this is converted into expression by cell angle θ_c using a second regression model (*middle*). Ordering genes by chromosome position (*right*) shows a smoothed version of the expression pattern in Fig. 3B. **B)** Gene-gene correlations across θ_c-binned expression data (100 bins) for the full scVI model (*left*), the replication-only model (*middle*), and the corrected model that is the difference of the two expression matrices (*right*). **C)** Comparison of $\sigma_{corrected}$ between LB-grown *E. coli* in Datasets D1 & D2 of genes classed as highly variable in both datasets (287 genes). Red indicates replication-divergent genes ($\sigma_{corrected} > 0.6$). **D)** Expression across 100 bins averaged by θ_c. Expression is *z*-scores derived from scVI (Full model, black) or predicted as a replication effect (Replication model, red). The zero value for θ_c is set as described in Materials & Methods. Genes that are adjacent and tandemly arrayed are plotted together (usually genes in the same operon).

Cell cycle analysis reveals novel replication-divergent operons in S. aureus.

Having established our approach for the analysis of cell cycle transcriptional dynamics in *E. coli*, we then applied this approach to the Gram-positive pathogen *S. aureus*. To ensure that cells

were in balanced growth, we sampled *S. aureus* at a lower growth density than the previous experiment (see Materials & Methods), and achieved a faster doubling time (t_d = 24.4 min vs 30.1 min in the previous experiment, Fig. S7A & B).

As before, UMAP of position-averaged gene expression arranged cells into a wheel (Fig. 5B), and progression around this wheel as indicated by θ_c reflected movement of gene expression from the origin to the terminus (Fig. 5C). Ordering genes by gene angle θ_g (Fig. S7C) also captured a relationship between θ_g and distance from the origin (Fig. 5D). The gradient of this relationship indicated overlap in rounds of replication (Fig. 5E), although less than *E. coli* with a similar doubling time (LB, Fig. 3E), presumably because the much smaller genome (2.9 Mb compared to 4.6 Mb in *E. coli* K-12) reduces the need for overlap in replication. In contrast, another Gram-positive pathogen with an even smaller genome (2.1 Mb), *Streptococcus pneumoniae*, did not exhibit multiple-fork replication under normal conditions (*30*). Thus, overall, despite the large evolutionary distance between *E. coli* and *S. aureus*, our cell cycle analysis revealed a global chromosome position-dependent pattern in both species suggesting that these replication effects are widely conserved.

Correcting for the effects of replication to eliminate global chromosome position-dependent correlations as we introduced for *E. coli* (Fig. S7D & E), we identified replication-divergent genes. First, we examined the *dnaA*-containing operon, as well as the ribonucleotide reductase (*nrd*) genes (Fig. 5F, reproduced in a second replicate in Fig. S8). Both showed similar dynamics to *E. coli*, reflecting the shared demands of DNA replication. The nucleoid occlusion factor *noc*, which in *S. aureus* regulates replication initiation (*31*), showed similar dynamics to *dnaA*. In comparison to *E. coli*, however, fewer genes diverged from the replication pattern, with only 23 identified as replication-divergent across two replicates (Fig. S7F). These included the chaperone genes, *groL* and *groES*, and a series of genes within the regulon of the transcription factor GbaA (*31*–*34*). GbaA is a transcriptional repressor responsive to reactive electrophilic species that regulates two divergent operons that include itself, as well as several enzymes involved in electrophile detoxification (Fig. 5G) (*33*, *34*). Since the GbaA regulon is responsive to similar electrophilic species to the *nemRA* operon identified as divergent in *E. coli* (Fig. 4D), and genes within this regulon showed relatively large fluctuations in expression over the cell cycle (5.2-fold peak-to-trough ratio, compared to a median 1.9-fold across all genes) (Fig. 5H, Fig. S7I), we further investigated whether this was driven by fluctuations in GbaA activity.

Performing PETRI-seq on a *gbaA* transposon insertion mutant, *gbaA*$^-$, we found that removal of GbaA repression increased the expression of the left-directed operon (GbaA-L) 137-fold compared to the wild-type JE2 strain (Fig. 5I). No increase was seen in the right-directed operon (GbaA-R), which contains *gbaA* itself, presumably due to disruption by the insertion. However, GbaA-L genes fluctuated less in *gbaA*$^-$ than the wild-type JE2 strain (1.6-fold compared to 3.2-fold on average), and followed the replication-expected pattern much more closely (Fig. 5J) (mean $\sigma_{corrected}$ values 0.68 and 0.33 for JE2 and *gbaA*$^-$, respectively). In contrast, deletion of the stress-responsive sigma factor *sigB* decreased absolute expression but did not impact the divergent pattern (Fig. S7J & K). This suggests that GbaA activity fluctuates throughout the cell cycle.

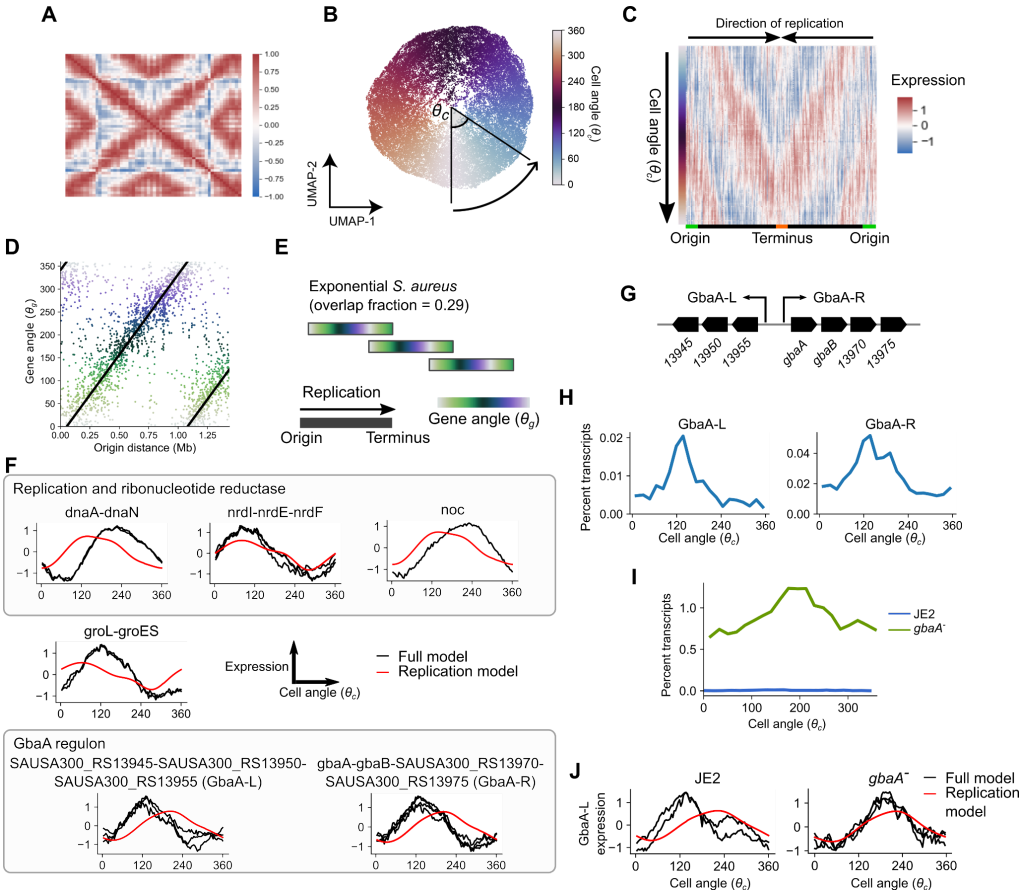

Figure 5: Cell cycle analysis reveals distinct regulatory dynamics of replication, stress, and virulence genes in *S. aureus*. A) Spearman correlations between scaled data averaged into 50 kb bins, as for Fig. 1F but with cells at lower density. **B)** UMAP of *S. aureus* with gene expression averaged in 50 kb bins by chromosome position. Cells are colored by the cell angle θ_c between UMAP dimensions relative to the center of the projection. **C)** Heatmap of expression after averaging by θ_c. **D)** The relationship between gene angle θ_g (Fig. S7B) and origin distance in *S. aureus*. **E)** The expected pattern of replication in *S. aureus*, as in Fig. 3E. **F)** Expression across 100 bins averaged by θ_c comparing observed and replication-predicted expression as in Fig. 4D. **G)** Structure of the GbaA regulon containing divergent left and right operons. **H)** Percentage of total mRNA transcripts derived from UMI counts, averaged in 20 bins by θ_c for the left (GbaA-L) and right (GbaA-R) operons of the GbaA regulon. **I)** Percent transcripts for operon Gba-L displayed as in **(H)** but comparing JE2 and *gbaA*$^-$ strains. **J)** Observed and replication-predicted expression for Gba-L as in **(K)** but comparing JE2 and *gbaA*$^-$ strains.

Cell cycle analysis provides insight into response heterogeneity and growth rate changes during antibiotic treatment.

The cell cycle fluctuation in transcript abundance described above creates population heterogeneity among exponentially growing cells. This heterogeneity is a concern if it influences bacterial responses to antibiotics (*35*, *36*). To determine whether the cell cycle influences antibiotic response, we treated exponentially-growing *S. aureus* (Fig. 6A) with a high dose (10 µg/ml) of vancomycin, inhibitor of cell wall synthesis (*37*). Vancomycin treatment produced a rapid response, with 137 genes induced at least twofold after six minutes (Fig. 6B). As expected, some of the most strongly-induced genes were in the *vraR* operon, which encodes

the VraSR two-component system that is part of the cell wall damage response (*38*). However, we also observed the strong induction of a group of contiguous genes involved in pyrimidine biosynthesis (labeled *pyr* genes). We considered two possible scenarios of cell cycle heterogeneity in treatment response: firstly, variation could be produced because higher gene copy numbers allow greater induction; secondly, cell cycle state could influence the regulatory response. Therefore, we examined the correlation between cell cycle expression dynamics before and after treatment. In general, there was a positive correlation in cell cycle expression pattern before and after treatment, suggesting that baseline cell cycle expression state informs the magnitude of post-treatment expression (Fig. 6C & D). This may be because when a gene is in a two-copy state, it can reach higher expression levels faster. In contrast, a few induced genes, including *argH*, exhibited a negative correlation in cell cycle dynamics before and after treatment (Fig. 6D). These perturbations may imply that the regulatory response is cell cycle-dependent in this case. The *argG-argH* operon also showed particularly high cell cycle variation after treatment, with a 4.6-fold peak-to-trough ratio, further suggesting differential induction by cell cycle state (Fig. S9A).

Examining the vancomycin response over time, principal component analysis (PCA) recovered the time order along principal component 1 (PC-1) (Fig. 6E). Gene set enrichment analysis (GSEA) (*39*, *40*) indicated that higher PC-1 scores were associated with induction of amino acid biosynthesis genes and reduced oxidative phosphorylation, transcription, and translation genes (Fig. S9F). This is consistent with induction of the stringent response, a growth-arrest response associated with vancomycin tolerance in *S. aureus* and other Gram-positive pathogens (*41–44*). To test whether PC-1 was correlated with slowed proliferation state, as expected from a stringent response, we grouped cells into eight equally sized bins by PC-1 (Fig. 6F, Fig. S9C) and calculated our “overlap fraction” metric of cell replication state (Fig. 3E) on each bin. There was a gradual decline in this metric followed by a rapid shift in the final PC-1 bin (Fig. 6G). Consistent with this, at the level of individual samples there was little perturbation in global gene covariance patterns until 25 min (Fig. S9D & E), and at a population level growth rate transitioned at ~15 min post treatment (Fig. 6A). During this response, *pyr* genes were strongly induced at 6 min (Fig. 6B) before rapidly declining by 25 min (Fig. 6H). However, in a subset of 25 min-treated cells, PC-1 remained lower and *pyr* gene expression remained high (Fig. 6H), indicating reduced responsiveness. Strikingly, these cells also displayed a reduced ratio of mRNA to rRNA (Fig. 6I) than the more responsive cells at this time point or cells at earlier time points. While most rRNA is packaged in ribosomes and relatively stable, mRNA is unstable and hence the mRNA to rRNA ratio serves as a metric for global transcriptional activity while controlling for variation in the efficiency of cell permeabilization (which should affect all RNA equally). Therefore, reduced progression into stringent responses-related growth arrest is associated with reduced transcriptional activity. Overall, these results demonstrate that our analysis framework can reveal both cell cycle heterogeneity in initial vancomycin responses, but also analyze heterogeneity in growth arrest responses that may determine antibiotic survival.

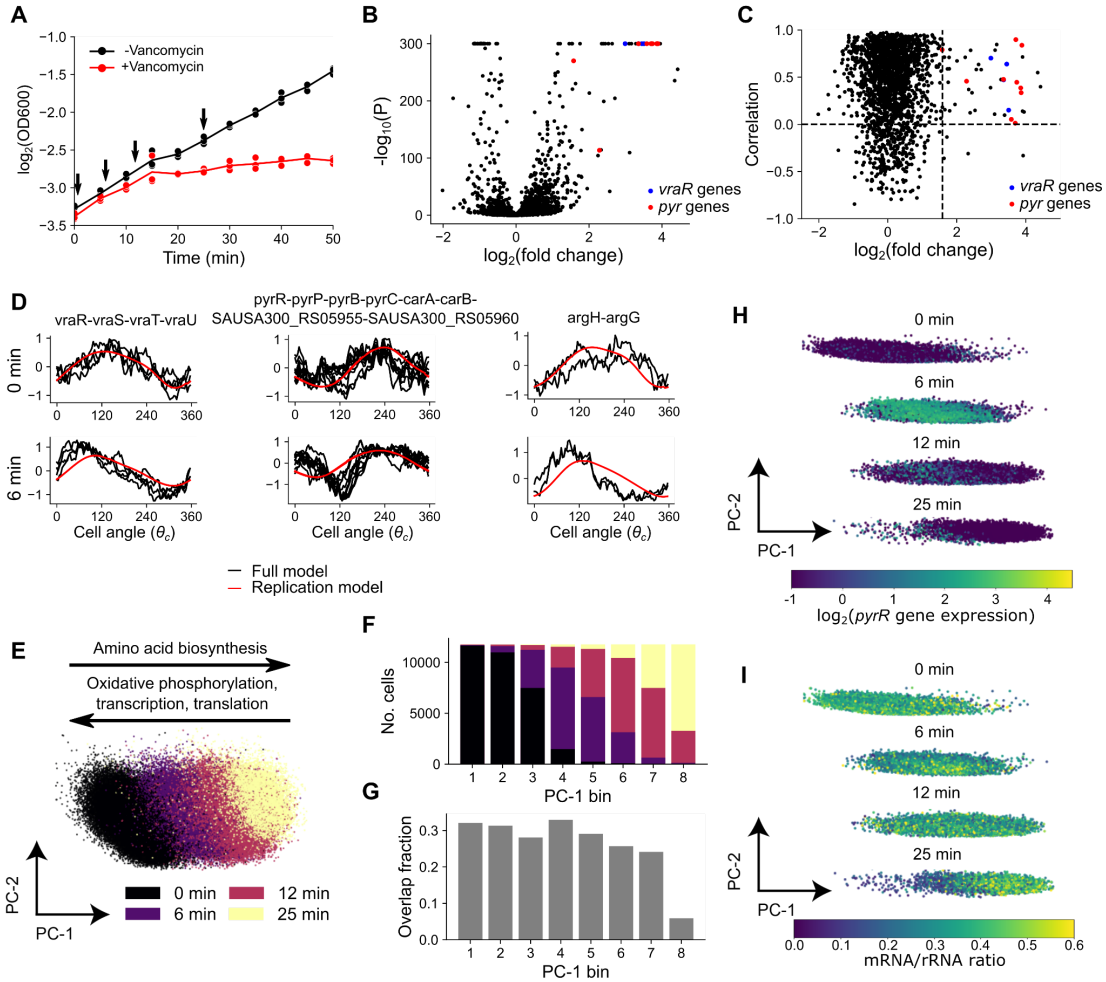

Figure 6: The *S. aureus* response to vancomycin exhibits initial cell cycle dependence and heterogeneous shifts in proliferation state. A) Growth response to vancomycin. Sampling points for PETRI-seq are indicated by arrows. **B)** Volcano plot of $\log_2$ fold change against $-\log_{10}$ *P*-values (Mann-Whitney-U test with Benjamini-Hochberg adjustment) of 6 min vancomycin compared to untreated cells. *P*-values >300 are set to 300. **C)** Comparing $\log_2$ fold change and correlation of cell cycle expression before and after 6 min vancomycin treatment. Dashed lines indicate genes with positive correlations and >threefold induction. **D)** Expression before and after 6 min vancomycin, comparing observed and replication-predicted expression as in Fig. 4D. **E)** PCA on cells at all time points. Annotated arrows indicate significant KEGG pathways (*45*, *46*) by GSEA based on the sign of normalized enrichment score (Fig. S9F). **F)** Sample distribution across eight bins by PC-1. Colors are as in **(E)**. **G)** Overlap fraction for each PC-1 bin. **H & I)** PCA split by sample showing log-expression of the *pyr* genes **(H)** and the ratio between total mRNA and rRNA counts **(I)**. mRNA/rRNA ratios are higher than expected but may reflect greater accessibility of mRNA to *in situ* reverse transcription in fixed cells (*10*).

Discussion

Transcriptome technologies have greatly contributed to our understanding of bacterial gene regulation (*47*, *48*), but measurement averaged across bulk populations has obscured their true transcriptional dynamics. Here, we demonstrate that DNA replication is a major driver of

expression heterogeneity in bacteria at the level of individual cells. While gene copy number effects on expression have been noted for individual genes (*49–52*) and have been theorized to influence global expression heterogeneity (*53*), we demonstrate for the first time a global influence of replication on gene expression. Our scRNA-seq analysis not only allowed us to model cell cycle expression in unsynchronized populations, but also to consider the cell cycle as a continuous process, revealing transcriptional dynamics in unprecedented resolution. The view that emerges from the model is that replication leads to a continually shifting transcriptome. In part, this results from the generally short half-life of bacterial mRNAs (*54*), which render RNA abundance highly sensitive to transcription rate due to copy number changes. However, there are likely myriad other factors influencing these dynamics, including competition for RNA polymerase (*55*) and the direct influence of DNA replication on local promoter environments (*56*, *57*). Building on our modeling approach may yield great insights into the factors influencing these processes.

How does the global replication effect on the transcriptome influence cell physiology? Conservation of position on the origin-terminus axis has been noted for some genes (*58*), but the precise replication pattern depends on growth rate (*59*) so the relationship between cell cycle function and evolutionary selection pressures on chromosomal position is unclear. Nevertheless, specific instances of phenotypic effects of copy number-mediated expression regulation have been noted, with relevance to sporulation, competence, and virulence (*30*, *60–62*). These considerations suggest that further exploration of the relationship between chromosomal function and cell cycle function will lead to more insight into bacterial cell physiology.

In contrast, only a minority of genes show strong evidence of dynamics that diverge from the replication effect. While bulk studies have reported ~15-20% of genes exhibiting cell cycle variation (*1*, *3*, *22*), our work suggests that many of these genes may fluctuate due to DNA replication effects, or in the case of *C. crescentus (1, 3)*, reflect unique requirements of its asymmetric division process. Our observations support the notion that cell cycle regulation of most cytoplasmic processes (e.g. translation, metabolism) is unnecessary if the cytoplasm is synthesized at a constant, exponential rate (*59*). DNA replication, however, undergoes discrete changes of rate (*14*). Timing of replication initiation is influenced by accumulation of DnaA, which is autoregulated at the level of transcription in *E. coli (63–65)*. Similarly, the concentration of dNTPs in the cell must be carefully controlled by regulation of ribonucleotide reductase activity, with lack or excess resulting in disrupted replication or increased mutation, respectively (*66*, *67*). Increased demand due to initiation of a new round of replication induces *nrd* gene transcription, through regulators including NrdR (*68–70*). The striking conservation of *dnaA* and *nrd* dynamics in *E. coli* and *S. aureus* reflect the shared physiological demands of replication.

E. coli, a Gram negative rod-shaped bacterium, regulates the transcription of cell wall and outer membrane synthesis and cell division genes (such as *ftsZ*, reported previously (*71*, *72*)). The rate of lipid synthesis is discontinuous during elongation and septation in the *E. coli* cell cycle (*73*) and cell membrane and wall biosynthesis must be tightly coordinated (*74*). Our data suggest that transcriptional regulation could play a role in this process in *E. coli*. By contrast, *S.*

aureus, which lacks an outer membrane, shows less evidence of regulation of cell surface biogenesis genes. *E. coli* (but not *S. aureus*) also dynamically regulates several energy biogenesis and electron transport chain transcripts. Membrane potential is important for cell division protein localization (*75*), while changes in metabolic flux could place changing demands on cellular bioenergetic processes. In line with this, in both *E. coli* and *S. aureus*, we observed large fluctuations in reactive electrophile-responsive regulons (NemR and GbaA regulons, respectively). In the case of *S. aureus*, this was driven by fluctuating GbaA activity since deletion of this repressor caused genes to revert to the replication-expected pattern, while deletion of other regulators did not. The physiological ligands to which GbaA is responding are not clear but may include reactive electrophiles such as aldehydes and quinones that emerge during metabolic stress (*33*, *34*). Treatment of cells with N-acetylcysteine, which scavenges some reactive electrophilic species (*76*), was not able to perturb the pattern (Fig. S7L). However, endogenous levels of methylglyoxal (an inducer of the GbaA regulon (*33*, *34*)) can influence cell length in *B. subtilis* (*77*), while in *C. crescentus* cell cycle changes in metabolism and redox potential influence cell division and chromosome segregation (*78*, *79*). Therefore, further investigation into the fluctuations driving GbaA and NemR activity may illuminate new ways in which bacterial metabolism connects to cell cycle progression.

While the cell cycle is fundamental to microbial physiology, it is also a source of heterogeneity that may influence how bacteria respond to external threats. Hence we applied our analysis framework to cells treated with vancomycin, the antibiotic of choice for methicillin-resistant *S. aureus*. We observed rapid transcriptional changes within six minutes, including a specific cell wall damage response. In general, cell cycle expression patterns of induced genes were correlated before and after treatment, implying that the mode of induction of these genes is cell cycle-independent but that cells with higher baseline expression reach higher levels of induction faster. For other genes, such as *argH*, induction led to a different cell cycle expression pattern, implying that cell cycle state (e.g. whether a cell is undergoing septum formation or division) influenced response signaling. Over time, a stringent-like response emerged that correlated with a shift to a reduced growth state, as inferred from the replication pattern. Strikingly, not all cells undergo this transition, and failure to do so was associated with reduced total mRNA content. While future work is required, this may imply that cells that do not reduce their growth in time undergo dysfunction that disrupts transcriptional activity. Since vancomycin tolerance in clinical isolates is prevalent and associated with treatment failure (*37*, *80*), understanding what drives these response kinetics is crucial. Overall, our framework provides multiple new ways of understanding heterogeneity in antibiotic responses.

Our work raises new questions about the regulators of cell cycle dynamics and their potential consequences for cell cycle function and heterogeneity. Moreover, while our modeling approach assumes a “steady state” of growth at a constant rate in a homogeneous environment, we believe that this approach can be generalized to many bacterial species and may be extended to more complex scenarios, such as changing environmental conditions (as in the case of vancomycin treatment here), as well as to identify and separate populations with mixed replication rates (with potential consequences for antibiotic tolerance (*81*)). Finally, scRNA-seq in bacteria has not yet widely been applied. While it may seem to be impeded by the extremely

low mRNA content per cell, our work conclusively demonstrates that underlying this noise are biological processes that can be quantitatively modeled and understood. Therefore, while new approaches may be required, we believe that the emergence of scRNA-seq for bacteria will illuminate bacterial physiology in unprecedented detail.

Materials and Methods

Bacterial strains and growth conditions

Strains used are listed in Table S1. All *E. coli* strains (a gift from Dr. Christian Rudolph) were routinely grown in modified Luria Broth (LB) (1% tryptone (Sigma-Aldrich), 0.5% yeast extract (Sigma-Aldrich), 0.05% NaCl, pH adjusted to 7.4 (*19*)). For growth in minimal media, an M9 base (1X M9 minimal salts (Gibco), 2 mM $MgSO_4$, 0.2 mM $CaCl_2$) was supplemented with 0.4% glucose (M9 + Glucose) or with both 0.4% glucose and 0.2% acid casein peptone (Acros Organics) (M9 + Glucose + AA). All *S. aureus* strains were routinely grown in Bacto tryptic soy broth (TSB) (BD Biosciences). The *gbaA* transposon mutant was provided by the Network on Antimicrobial Resistance in *Staphylococcus aureus* (cat. # NR-46898) and the *ΔsigB* mutant was a gift from Dr. Alexander Horswill).

Growth curves

Strains were grown overnight in LB (*E. coli*) or TSB (*S. aureus*) at 37°C, shaking at 225 rpm. For initial experiments with *S. aureus* (Datasets D3 & D4), strains were diluted to an A_{600} value of 0.05 in prewarmed TSB, after which A_{600} was measured at the times specified. A_{600} was measured on a BioMate 3S spectrophotometer (Thermo Scientific). For experiments with *S. aureus* in balanced growth (Datasets D5-D7), overnight cultures were diluted in TSB first to 0.005, then after 3 hr diluted again to 0.005 before measuring A_{600} at the time intervals specified. For *E. coli* growth curves, strains were incubated for 2 hr in the desired medium then diluted again in the same prewarmed medium to an A_{600} value of 0.005, after which A_{600} was measured at the time intervals specified. Where *E. coli* cells were diluted into a different medium, cells were washed once with PBS prior to dilution. To measure growth rate, a linear model $\log_2(A_{600}) \sim mT + c$ was calculated for the linear portion of this relationship (where T is the time in minutes) using the LINEST function in Microsoft Excel and the doubling time in minutes t_d was calculated as $1/m$.

Vancomycin treatment

For experiments measuring responses to vancomycin, cells were brought into balanced growth as above by diluting overnight cultures to 0.005, diluting again to 0.005 after 3 hrs, and then incubating for a further 90 min (for an A_{600} of ~0.1). Cells were then exposed to 10 µg/ml vancomycin (~5-10X minimum inhibitory concentration (*82*, *83*)). For growth curves, A_{600} was measured every 5 min for 50 min. For PETRI-seq, cells were taken at 6, 12, and 25 min post treatment and rapidly chilled with a dry ice-ethanol bath to <10°C before transferring to a new, pre-chilled tube on ice and centrifugation at 3,220 x g, 5 min, 4°C. This was to ensure that the transcriptome at these fast time points did not continue to change during harvesting.

PETRI-seq analysis

Cells were grown as described for the growth curves except that after specific time intervals (for *S. aureus*, 2 hr 20 min in initial experiments, 1 hr 30 min in balanced growth experiments; for *E. coli*, 2 hr, 3 hr, and 7 hr in LB, M9 + Glucose + AA, and M9 + Glucose, respectively, when growth rates appeared constant (Fig. 2C, Fig. S3A)) cells were harvested by centrifugation and resuspension in 4% formaldehyde in PBS. For *S. aureus* initial experiments, centrifugation was at 10,000 x g, 1 min at room temperature and for *E. coli* and balanced growth *S. aureus* experiments, centrifugation was at 3,220 x g, 5 min, 4°C. PETRI-seq was carried out as described previously (*10*) with the following modifications. Initial fixing, permeabilization, and DNase treatment were carried out as described but with cell wall permeabilization using 100 µg/ml lysostaphin (Sigma-Aldrich) for *S. aureus* and 100 µg/ml lysozyme (Thermo Scientific) for *E. coli*. For Dataset D4, samples were split into processing with or without DNase treatment and subsequent wash steps, to test whether this would affect correlation patterns (suggesting contaminating genomic DNA could play a role). However, no difference was observed in the presence or absence of DNase treatment, although UMI/barcode was slightly higher after DNase treatment (Table S1). For barcoding, the number of cells included was reduced from 3 x 10^7 to a maximum of 1 x 10^7, since preliminary experiments indicated lower input at this stage was associated with a higher UMI/barcode for *S. aureus*. Tagmentation was performed using the EZ-Tn5 transposase (Lucigen) as described in the latest version of the PETRI-seq protocol (available at https://tavazoielab.c2b2.columbia.edu/PETRI-seq/updates_April2021/PETRI_Seq_Protocol.pdf). Briefly, the transposase was loaded by incubating EZ-Tn5 with pre-annealed oligonucleotides (/5Phos/CTGTCTCTTATACACATCT and GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG) at 4 µM and 40% glycerol at room temperature for 30 min. Tagmentation was then performed incubating samples with loaded EZ-Tn5 (at a final further dilution of 400x) and 2x Tagment DNA buffer (either using the Nextera 2x Tagment DNA (TD) buffer or 20 mM Tris(hydroxymethyl)aminomethane; 10 mM $MgCl_2$; 20% (vol/vol) dimethylformamide, pH adjusted to 7.6 with acetic acid (*84*)). After incubating for 5 min at 55°C and decreasing the temperature to 10°C, either Nextera NT buffer (Illumina) or 0.2% sodium dodecyl sulfate was added, allowing neutralization to proceed for 5 min. Final amplification was performed with Q5 polymerase (New England Biolabs) using the NEBNext Universal i5 primer (New England Biolabs) and the N7 indices from the Nextera XT Index Kit v2 Set A (Illumina) as also described in the updated PETRI-seq protocol. Sequencing was performed on an Illumina NextSeq 500 to obtain 58 x 26 base paired-end reads. For each barcoding experiment, multiple libraries of ~20,000 cells were prepared and sequenced, and no batch effects were noted across libraries.

Pre-processing and scVI analysis

Initial demultiplexing of barcodes, alignment, and feature quantification was performed using the analysis pipeline described in (*10*) except that feature quantification was performed at the gene level rather than operon level. Reference sequences and annotations were obtained from Genbank (https://www.ncbi.nlm.nih.gov/genbank/). *E. coli* reads were aligned to the K-12 MG1655 reference assembly (GCA_000005845.2) and *S. aureus* to the USA300_FPR3757 reference assembly (GCF_000013465.1). To quantify rRNA, all rRNA genes were treated as a single feature (since high sequence similarity impedes unique assignment of rRNA reads to individual genes). After initial processing, counts by cell barcode were pooled across different

libraries and initial filtering was performed using Scanpy v1.7.1 (*85*). Barcodes with fewer than 15 UMI (20 for Dataset D3, D5-7) were removed, as well as any genes with fewer than 50 UMI across all included barcodes (100 for Dataset D3). To generate the denoised representation of the data, scVI v0.9.0 (*13*) was applied with the following hyperparameters, chosen through grid search to distinguish between closely related *S. aureus* strains in a pilot dataset: two hidden layers, 64 nodes per layer, five latent variables, a dropout rate of 0.1, and with a zero-inflated negative binomial gene likelihood (other hyperparameters maintained as defaults). Denoised expression values based on the scVI model were obtained using the scVI function "get_normalized_expression".

Cell cycle analysis

Cells were assigned to cell cycle phases by calculating the angle θ_c relative to the origin between *x* and *y* coordinates in a two-dimensional UMAP embedding of the data as $\tan^{-1}(x / y)$, similar to the ZAVIT method our lab has described previously (*86*, *87*). Embeddings were computed by averaging *z*-scores derived from denoised expression values within bins according to chromosomal location (50-400 kb bins, depending on the dataset), and then performing two-dimensional UMAP analysis using the umap-learn v0.5.1 library in Python (https://umap-learn.readthedocs.io/en/latest/) with the 'correlation' distance metric. These embeddings were then mean-centered (Fig. 3A & 5A). To get the expression by cell angle matrix used in Fig. 3B, gene expression was then averaged within 100 equally spaced bins of θ_c to produce a cell angle-binned expression matrix. To order genes based on their cell cycle expression, gene angle, θ_g, was calculated as follows. PCA was performed on the transpose of the cell angle-binned expression matrix and θ_g was calculated as the angle between PCs 1 and 2 relative to the origin. Together, θ_c and θ_g are metrics for ordering of cells and genes, respectively, within the model of cell cycle gene expression described here.

Modeling the gene angle-origin distance relationship

While there was a strong relationship between origin distance *D* and gene angle θ_g, modeling this relationship is challenged by the fact that the relationship is "wrapped" with an unknown periodicity with respect to *D* (Fig. 3D & 5C) (i.e. after a period of increased θ_g with *D*, θ_g starts again at zero). To fit this relationship, a custom Bayesian regression analysis was developed according to the following model partially adapted from (*88*), with both θ_g and *D* standardized to the range -π to π:

$$\theta_g \sim von\ Mises(A,\ \kappa)$$
$$cos(A) = \beta_1 cos(\gamma D) - \beta_2 sin(\gamma D)$$
$$sin(A) = \beta_2 cos(\gamma D) + \beta_1 sin(\gamma D)$$

Where:

$$log(\kappa) \sim Gaussian(0,\ 1)$$
$$\beta_1 \sim Gaussian(0,\ 0.5)$$
$$\beta_2 \sim Gaussian(0,\ 0.5)$$
$$log(\gamma) \sim Gaussian(0,\ 0.5)$$

The von Mises probability distribution is a circular probability distribution here parameterized by A, the predicted mean angle, and κ, the concentration parameter (higher κ implies greater concentration of the distribution around A). The parameter γ can be interpreted as the gradient of D with respect to θ_g after standardizing both variables to to the range -π to π. To give the gradient in °/Mb (as in Fig. 3D), this value is multiplied by 360 divided by origin-terminus distance in Mb. The inverse of γ, $1/\gamma$, is the gradient of θ_g with respect to D (after range standardization) and therefore is the fraction of the origin-terminus distance covered within a single span of θ_g. Therefore, $1 - 1/\gamma$ is the fraction of D during which the next round of replication has already initiated, referred to as the "overlap fraction" in Fig. 3E & 5D. Here, γ is constrained to be positive by the lognormal prior distribution (Fig. S10), which is appropriate since the ordering of angles θ_g are reversed (i.e. 360 - θ_g when θ_g is in degrees) if during analysis this relationship shows a negative trend. This can occur because the directionality of PCs used to calculate θ_g is arbitrary. Posterior distributions for the parameters were obtained by Hamiltonian Monte-Carlo sampling using Rstan v2.21.3 *(89)*. Fitted values for θ_g based on D ($\theta_{g\text{-}pred}$) were calculated by determining $\theta_{g\text{-}pred}$ for all sampled parameter values and then calculating the mean value of $\theta_{g\text{-}pred}$ as $\tan^{-1}(\text{mean}(\sin(\theta_{g\text{-}pred})) / \text{mean}(\cos(\theta_{g\text{-}pred})))$.

Modeling the cell angle-gene angle relationship

To predict expression based on cell angle θ_c and gene angle θ_g, a polynomial regression model was constructed using scikit-learn v0.24.1(*90*). Specifically, both angles were converted to radians and then transformed into $\cos(\theta_c)$, $\sin(\theta_c)$, $\cos(\theta_g)$, and $\sin(\theta_g)$. All combinations and interactions up to a fourth degree polynomial were constructed using the scikit-learn PolynomialFeatures function. These features were then used to fit a Ridge regression model ($\alpha = 10$). The model was trained on scVI expression z scores averaged first in 100 bins by θ_c then in 100 bins by θ_g (i.e. the expression matrix used for Fig. 3F). An alternative approach considered was a non-linear approach using the scikit-learn implementation of kernel ridge regression with kernel "rbf". However, the fourth degree polynomial model performed similarly and was computationally far more efficient so was chosen (increasing the polynomial degree further made little difference to performance).

Predicting expression dynamics based on DNA replication alone

To derive a prediction of cell cycle gene expression dynamics based on the expected effect of replication alone, the two regression models above were combined to yield the pipeline in Fig. 4A. Firstly, the gene angle-origin distance model was used to predict the expected value $\theta_{g\text{-}pred}$ from origin distance D. Next, cell cycle expression was predicted using the cell angle-gene angle regression model using $\theta_{g\text{-}pred}$ values. For cell angle θ_c, values used were the average θ_c values of cells binned into 100 equally spaced bins by θ_c. This gives a replication-predicted gene expression matrix of 100 bins x number of genes. The success of this model fit was evaluated based on the correlation with the θ_c-binned expression z-scores derived from scVI (Fig. S5C & S7D), as well as the loss of global chromosome position-dependent gene-gene correlations upon correction of scVI expression with replication-predicted expression (Fig. 4B, Fig. S7E). Additionally, we used this modeling approach to set the zero angle for gene expression plots. Using this approach, we predicted the gene expression profile by θ_c for an

imaginary gene at D = 0 (i.e. at the origin of replication). We then determined the value of θ_c giving the minimum predicted expression, reasoning that if increased expression in this model is responsive to a doubling of copy number, the doubling event should occur at the expression minimum. Therefore, we determined this angle, θ_0 to be the most likely value of θ_c at which replication initiation occurs, rotating the angles by the operation (θ_c - θ_0) *mod* 360 to set this point as 0°. While we have not experimentally verified that θ_0 represents the true point of replication initiation, nevertheless it is a useful and theoretically principled point by which to standardize and compare different datasets, where otherwise the point of 0° is arbitrary.

Identifying replication-divergent genes

We identified replication-divergent genes based on two criteria: absolute variability by cell angle θ_c and divergence from the replication model. First, we identified highly variable genes as follows (based on the method implemented in Seurat v3 (*91*)). We normalized raw counts for library size (so that the total sum of UMI for each barcode was the median UMI/barcode), then to reduce sparsity while retaining cycle information, we averaged counts across 20 bins by θ_c. Next, we log2-transformed the data (eliminating any genes with zero values after binning to allow log-transformation). We observed a negative overall relationship between the mean and variance of genes in log-transformed data (Fig. S5D), to which we fitted a regression line with locally weighted scatterplot smoothing (LOWESS) using the Python package statsmodels v0.12.2 (*92*). We used this fit to develop a mean-dependent variance threshold. In all cases, genes were considered highly variable if they had a ratio of observed to LOWESS-predicted variance > 1.3 as well as a $\log_2$ mean normalized expression > -10. These thresholds typically classified ~25% of genes as highly variable. Next, to quantify divergence from the replication model, we subtracted the replication-predicted expression from the scVI-derived expression *z*-scores (both averaged in 100 bins by θ_c) to "correct" for the effect of replication, and then calculated the standard deviation of this replication-corrected value, $\sigma_{corrected}$. A high $\sigma_{corrected}$ indicates that the dynamics behave differently from that expected based on replication alone. Thresholds for $\sigma_{corrected}$ (0.6 for *E. coli*, 0.5 for *S. aureus*) were determined manually based on inspection of the relationship between $\sigma_{corrected}$ across two datasets and choosing a value above which the correlation between datasets was stronger (Fig. S5E & Fig. S7G) (below the threshold, lack of reproducibility of $\sigma_{corrected}$ suggests divergences are small and dominated by noise). To calculate peak/trough fold changes in expression, normalized gene expression derived from scVI was averaged into 100 bins by θ_c and then the ratio between the fourth highest and fourth lowest values were calculated (this was chosen instead of maximum/minimum values to increase robustness to noise).

Simulating the effect of DNA replication on gene expression

We predicted the effect of DNA copy number on gene-gene correlations using a simulation written in Python (see Fig. S2) as follows. Cells were represented by genomes with 200 genes, each represented as a single integer and divided into individual replication units. In the simplest case, genomes were divided into two units of 100 genes (i.e. the two "arms" of the chromosome). In each cell, replication initiation events were simulated at intervals determined by a Poisson distribution with expected value μ. After an initiation event, replication proceeds in stepwise fashion along the length of each replication unit, doubling the copy number at each

point until the end of that replication unit has been reached. We also simulate "cell division" events in which all copy numbers are halved. These are timed independently from replication initiation but in the same way (at Poisson-distributed intervals with rate μ), with an additional offset from the first replication initiation event. In practice, we found that this offset did not affect correlations, since all genes are scaled equally. We used an initial offset of 150 steps (i.e. 1.5x the time to replicate a 100 gene replication unit, equivalent to the 40 min C-period + 20 min D-period originally proposed for *E. coli* B/r (*14*)). For each simulation, we generated 1,000 cells. Cells were initiated one at a time to yield an unsynchronized population, then the simulation was run for a further 1,000 steps with the whole population. We then normalized expression by total counts and calculated Spearman correlations across all genes. In order to simulate specific doubling times, the rate μ was calculated as $\mu = \left(n \times t_d\right) / t_c$ where *n* is the number of genes in the longest replication unit (here, 100 genes), t_d is the doubling time, and t_c is the C-period (here a value of 42 min was chosen for *E. coli* MG1655 based on (*15*)). The t_d/t_c ratio represents the fraction of one round of chromosomal replication that can take place in one cell cycle. Finally, for simulation of cells with additional origins of replication, genes were split into replication units according to the following assumptions: a) all origins initiate replication simultaneously; b) replication stops at the termination site *ter*, which is halfway along the chromosome; c) genes are replicated by the nearest origin (unless the replication fork must pass through *ter* to reach that gene).

Bulk RNA-seq analysis

For the analysis of bulk RNA-seq from (*2*) (Fig. S3D), we accessed data from the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession ID GSE46915. Counts were size factor-normalized with DESeq2 v1.32.0 (*93*), then data were standardized to *z*-scores and averaged into 100 kb bins by chromosomal position. Spearman correlations of binned values across all time points and replicates are shown.

Acknowledgements

We thank Yitzhak Pilpel, Jonas Schluter, Ido Golding, and Timothée Lionnet for critical discussions on the project and the manuscript, and Saeed Tavazoie, Sydney Blattman, and Wenyan Jiang for initial advice on implementing PETRI-seq. We thank Christian Rudolph and his lab for providing the *E. coli* strains. We also further thank members of the Yanai lab for advice and suggestions. This work was funded by grants R21AI169350 (IY) R01AI143290 (IY), R01AI137336 (BS, VJT, and IY) from the NIH.

References

1. M. T. Laub, H. H. McAdams, T. Feldblyum, C. M. Fraser, L. Shapiro, Global analysis of the genetic network controlling a bacterial cell cycle. *Science*. **290**, 2144–2148 (2000).

2. G. Fang, K. D. Passalacqua, J. Hocking, P. M. Llopis, M. Gerstein, N. H. Bergman, C. Jacobs-Wagner, Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co-expression and evolution. *BMC Genomics*. **14**, 450 (2013).

3. B. Zhou, J. M. Schrader, V. S. Kalogeraki, E. Abeliuk, C. B. Dinh, J. Q. Pham, Z. Z. Cui, D. L. Dill, H. H. McAdams, L. Shapiro, The global regulatory architecture of transcription during the Caulobacter cell cycle. *PLoS Genet.* **11**, e1004831 (2015).

4. N. J. De Nisco, R. P. Abo, C. M. Wu, J. Penterman, G. C. Walker, Global analysis of cell cycle gene expression of the legume symbiont Sinorhizobium meliloti. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3217–3224 (2014).

5. A. C. Bandekar, S. Subedi, T. R. Ioerger, C. M. Sassetti, Cell-Cycle-Associated Expression Patterns Predict Gene Function in Mycobacteria. *Curr. Biol.* **30**, 3961–3971.e6 (2020).

6. S. Cooper, The synchronization manifesto: a critique of whole-culture synchronization. *FEBS J.* **286**, 4650–4656 (2019).

7. B. Xia, I. Yanai, A periodic table of cell types. *Development*. **146** (2019), doi:10.1242/dev.169854.

8. R. Nayak, Y. Hasija, A hitchhiker's guide to single-cell transcriptomics and data analysis pipelines. *Genomics*. **113**, 606–619 (2021).

9. E. Papalexi, R. Satija, Single-cell RNA sequencing to explore immune cell heterogeneity. *Nat. Rev. Immunol.* **18**, 35–45 (2018).

10. S. B. Blattman, W. Jiang, P. Oikonomou, S. Tavazoie, Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat Microbiol*. **5**, 1192–1201 (2020).

11. 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*. **371** (2021), doi:10.1126/science.aba5257.

12. 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).

13. R. Lopez, J. Regier, M. B. Cole, M. I. Jordan, N. Yosef, Deep generative modeling for single-cell transcriptomics. *Nat. Methods*. **15**, 1053–1058 (2018).

14. S. Cooper, C. E. Helmstetter, Chromosome replication and the division cycle of Escherichia coli B/r. *J. Mol. Biol.* **31**, 519–540 (1968).

15. O. Michelsen, M. J. Teixeira de Mattos, P. R. Jensen, F. G. Hansen, Precise determinations of C and D periods by flow cytometry in Escherichia coli K-12 and B/r. *Microbiology*. **149**, 1001–1010 (2003).

16. R. Allman, T. Schjerven, E. Boye, Cell cycle parameters of Escherichia coli K-12. *J. Bacteriol.* **173**, 7970–7974 (1991).

17. M. Schaechter, O. Maaloe, N. O. Kjeldgaard, Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium. *J. Gen. Microbiol.* **19**, 592–606 (1958).

18. X. Wang, C. Lesterlin, R. Reyes-Lamothe, G. Ball, D. J. Sherratt, Replication and segregation of an Escherichia coli chromosome with two replication origins. *Proc. Natl.*

Acad. Sci. U. S. A. **108**, E243–50 (2011).

19. J. U. Dimude, M. Stein, E. E. Andrzejewska, M. S. Khalifa, A. Gajdosova, R. Retkute, O. Skovgaard, C. J. Rudolph, Origins Left, Right, and Centre: Increasing the Number of Initiation Sites in the Chromosome. *Genes* . **9** (2018), doi:10.3390/genes9080376.

20. D. Ivanova, T. Taylor, S. L. Smith, J. U. Dimude, A. L. Upton, M. M. Mehrjouy, O. Skovgaard, D. J. Sherratt, R. Retkute, C. J. Rudolph, Shaping the landscape of the Escherichia coli chromosome: replication-transcription encounters in cells with an ectopic replication origin. *Nucleic Acids Res.* **43**, 7865–7877 (2015).

21. P. W. Theisen, J. E. Grimwade, A. C. Leonard, J. A. Bogan, C. E. Helmstetter, Correlation of gene transcription with the time of initiation of chromosome replication in Escherichia coli. *Mol. Microbiol.* **10**, 575–584 (1993).

22. P. L. Echtenkamp, D. B. Wilson, M. L. Shuler, Cell cycle progression in Escherichia coli B/r affects transcription of certain genes: Implications for synthetic genome design. *Biotechnol. Bioeng.* **102**, 902–909 (2009).

23. P. Zhou, J. A. Bogan, K. Welch, S. R. Pickett, H. J. Wang, A. Zaritsky, C. E. Helmstetter, Gene transcription and chromosome replication in Escherichia coli. *J. Bacteriol.* **179**, 163–169 (1997).

24. Y. Hirota, A. Ryter, F. Jacob, Thermosensitive mutants of E. coli affected in the processes of DNA synthesis and cellular division. *Cold Spring Harb. Symp. Quant. Biol.* **33**, 677–693 (1968).

25. M. L. Gefter, Y. Hirota, T. Kornberg, J. A. Wechsler, C. Barnoux, Analysis of DNA polymerases II and 3 in mutants of Escherichia coli thermosensitive for DNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **68**, 3150–3153 (1971).

26. L. Sun, J. A. Fuchs, Escherichia coli ribonucleotide reductase expression is cell cycle regulated. *Mol. Biol. Cell*. **3**, 1095–1105 (1992).

27. E. Ozyamak, C. de Almeida, A. P. S. de Moura, S. Miller, I. R. Booth, Integrated stress response of Escherichia coli to methylglyoxal: transcriptional readthrough from the nemRA operon enhances protection through increased expression of glyoxalase I. *Mol. Microbiol.* **88**, 936–950 (2013).

28. M. J. Gray, W.-Y. Wholey, B. W. Parker, M. Kim, U. Jakob, NemR is a bleach-sensing transcription factor. *J. Biol. Chem.* **288**, 13789–13798 (2013).

29. C. Lee, J. Shin, C. Park, Novel regulatory system nemRA-gloA for electrophile reduction in Escherichia coli K-12. *Mol. Microbiol.* **88**, 395–412 (2013).

30. J. Slager, M. Kjos, L. Attaiech, J.-W. Veening, Antibiotic-induced replication stress triggers bacterial competence by increasing gene dosage near the origin. *Cell*. **157**, 395–406 (2014).

31. T. Pang, X. Wang, H. C. Lim, T. G. Bernhardt, D. Z. Rudner, The nucleoid occlusion factor Noc controls DNA replication initiation in Staphylococcus aureus. *PLoS Genet.* **13**, e1006908 (2017).

32. Y. You, T. Xue, L. Cao, L. Zhao, H. Sun, B. Sun, Staphylococcus aureus glucose-induced biofilm accessory proteins, GbaAB, influence biofilm formation in a PIA-dependent manner. *Int. J. Med. Microbiol.* **304**, 603–612 (2014).

33. A. Ray, K. A. Edmonds, L. D. Palmer, E. P. Skaar, D. P. Giedroc, Glucose-Induced Biofilm Accessory Protein A (GbaA) Is a Monothiol-Dependent Electrophile Sensor. *Biochemistry*. **59**, 2882–2895 (2020).

34. V. Van Loi, T. Busche, V. N. Fritsch, C. Weise, M. C. H. Gruhlke, A. J. Slusarenko, J. Kalinowski, H. Antelmann, The two-Cys-type TetR repressor GbaA confers resistance under disulfide and electrophile stress in Staphylococcus aureus. *Free Radic. Biol. Med.* **177**, 120–131 (2021).

35. W. A. Weigel, P. Dersch, Phenotypic heterogeneity: a bacterial virulence strategy. *Microbes Infect.* **20**, 570–577 (2018).

36. M. Huemer, S. Mairpady Shambat, J. Bergada-Pijuan, S. Söderholm, M. Boumasmoud, C. Vulin, A. Gómez-Mejia, M. Antelo Varela, V. Tripathi, S. Götschi, E. Marques Maggio, B. Hasse, S. D. Brugger, D. Bumann, R. A. Schuepbach, A. S. Zinkernagel, Molecular reprogramming and phenotype switching in lead to high antibiotic persistence and affect therapy success. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2014920118.

37. J. C. Barna, D. H. Williams, The structure and mode of action of glycopeptide antibiotics of the vancomycin group. *Annu. Rev. Microbiol.* **38**, 339–357 (1984).

38. A. Belcheva, D. Golemi-Kotra, A close-up view of the VraSR two-component system. A mediator of Staphylococcus aureus response to cell wall damage. *J. Biol. Chem.* **283**, 12354–12364 (2008).

39. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).

40. V. K. Mootha, C. M. Lindgren, K.-F. Eriksson, A. Subramanian, S. Sihag, J. Lehar, P. Puigserver, E. Carlsson, M. Ridderstråle, E. Laurila, N. Houstis, M. J. Daly, N. Patterson, J. P. Mesirov, T. R. Golub, P. Tamayo, B. Spiegelman, E. S. Lander, J. N. Hirschhorn, D. Altshuler, L. C. Groop, PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).

41. M. Moscoso, M. Domenech, E. García, Vancomycin tolerance in Gram-positive cocci. *Environ. Microbiol. Rep.* **3**, 640–650 (2011).

42. T. Geiger, B. Kästle, F. L. Gratani, C. Goerke, C. Wolz, Two small (p)ppGpp synthases in Staphylococcus aureus mediate tolerance against cell envelope stress conditions. *J. Bacteriol.* **196**, 894–902 (2014).

43. M. Singh, M. Matsuo, T. Sasaki, T. Hishinuma, N. Yamamoto, Y. Morimoto, T. Kirikae, K. Hiramatsu, RNA Sequencing Identifies a Common Physiology in Vancomycin- and Ciprofloxacin-Tolerant Staphylococcus aureus Induced by Mutations. *Antimicrob. Agents Chemother.* **64** (2020), doi:10.1128/AAC.00827-20.

44. J. Abranches, A. R. Martinez, J. K. Kajfasz, V. Chávez, D. A. Garsin, J. A. Lemos, The molecular alarmone (p)ppGpp mediates stress responses, vancomycin tolerance, and virulence in Enterococcus faecalis. *J. Bacteriol.* **191**, 2248–2256 (2009).

45. M. Kanehisa, S. Goto, KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30 (2000).

46. M. Kanehisa, M. Furumichi, Y. Sato, M. Ishiguro-Watanabe, M. Tanabe, KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.* **49**, D545–D551 (2021).

47. R. Sorek, P. Cossart, Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat. Rev. Genet.* **11**, 9–16 (2010).

48. M. J. Filiatrault, Progress in prokaryotic transcriptomics. *Curr. Opin. Microbiol.* **14**, 579–586 (2011).

49. M. G. Chandler, R. H. Pritchard, The effect of gene concentration and relative gene dosage on gene output in Escherichia coli. *Mol. Gen. Genet.* **138**, 127–141 (1975).

50. M. B. Schmid, J. R. Roth, Gene location affects expression level in Salmonella typhimurium. *J. Bacteriol.* **169**, 2872–2875 (1987).

51. C. Sousa, V. de Lorenzo, A. Cebolla, Modulation of gene expression through chromosomal positioning in Escherichia coli. *Microbiology*. **143 (Pt 6)**, 2071–2078 (1997).

52. E. N. Trip, J.-W. Veening, E. J. Stewart, J. Errington, D.-J. Scheffers, Balanced transcription of cell division genes in Bacillus subtilis as revealed by single cell analysis. *Environ. Microbiol.* **15**, 3196–3209 (2013).

53. J. R. Peterson, J. A. Cole, J. Fei, T. Ha, Z. A. Luthey-Schulten, Effects of DNA replication on mRNA noise. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 15886–15891 (2015).

54. D. W. Selinger, R. M. Saxena, K. J. Cheung, G. M. Church, C. Rosenow, Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. *Genome Res.* **13**, 216–223 (2003).

55. D. J. Jin, C. Cagliero, Y. N. Zhou, Growth rate regulation in Escherichia coli. *FEMS Microbiol. Rev.* **36**, 269–287 (2012).

56. C. J. Dorman, M. J. Dorman, DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophys. Rev.* **8**, 89–100 (2016).

57. C. St Germain, H. Zhao, J. H. Barlow, Transcription-Replication Collisions-A Series of Unfortunate Events. *Biomolecules*. **11** (2021), doi:10.3390/biom11081249.

58. P. Sobetzko, A. Travers, G. Muskhelishvili, Gene order and chromosome dynamics coordinate spatiotemporal gene expression during the bacterial growth cycle. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E42–50 (2012).

59. S. Cooper, "4 - Cytoplasm Synthesis during the Division Cycle" in *Bacterial Growth and Division*, S. Cooper, Ed. (Academic Press, San Diego, 1991), pp. 63–93.

60. J. Slager, J.-W. Veening, Hard-Wired Control of Bacterial Processes by Chromosomal Gene

Location. *Trends Microbiol.* **24**, 788–800 (2016).

61. J. Narula, A. Kuchina, D.-Y. D. Lee, M. Fujita, G. M. Süel, O. A. Igoshin, Chromosomal Arrangement of Phosphorelay Genes Couples Sporulation and DNA Replication. *Cell*. **162**, 328–337 (2015).

62. A. Soler-Bistué, J. A. Mondotte, M. J. Bland, M.-E. Val, M.-C. Saleh, D. Mazel, Genomic location of the major ribosomal protein gene locus determines Vibrio cholerae global growth and infectivity. *PLoS Genet.* **11**, e1005156 (2015).

63. T. Atlung, E. S. Clausen, F. G. Hansen, Autoregulation of the dnaA gene of Escherichia coli K12. *Mol. Gen. Genet.* **200**, 442–450 (1985).

64. R. E. Braun, K. O'Day, A. Wright, Autoregulation of the DNA replication gene dnaA in E. coli K-12. *Cell*. **40**, 159–169 (1985).

65. I. P. Menikpurage, K. Woo, P. E. Mera, Transcriptional Activity of the Bacterial Replication Initiator DnaA. *Front. Microbiol.* **12**, 662317 (2021).

66. E. Torrents, Ribonucleotide reductases: essential enzymes for bacterial life. *Front. Cell. Infect. Microbiol.* **4**, 52 (2014).

67. L. J. Wheeler, I. Rajagopal, C. K. Mathews, Stimulation of mutagenesis by proportional deoxyribonucleoside triphosphate accumulation in Escherichia coli. *DNA Repair* . **4**, 1450–1456 (2005).

68. I. Grinberg, T. Shteinberg, B. Gorovitz, Y. Aharonowitz, G. Cohen, I. Borovok, The Streptomyces NrdR transcriptional regulator is a Zn ribbon/ATP cone protein that binds to the promoter regions of class Ia and class II ribonucleotide reductase operons. *J. Bacteriol.* **188**, 7635–7644 (2006).

69. E. Torrents, I. Grinberg, B. Gorovitz-Harris, H. Lundström, I. Borovok, Y. Aharonowitz, B.-M. Sjöberg, G. Cohen, NrdR controls differential expression of the Escherichia coli ribonucleotide reductase genes. *J. Bacteriol.* **189**, 5012–5021 (2007).

70. B. L. McKethan, S. Spiro, Cooperative and allosterically controlled nucleotide binding regulates the DNA binding activity of NrdR. *Mol. Microbiol.* **90**, 278–289 (2013).

71. P. Zhou, C. E. Helmstetter, Relationship between ftsZ gene expression and chromosome replication in Escherichia coli. *J. Bacteriol.* **176**, 6100–6106 (1994).

72. T. Garrido, M. Sánchez, P. Palacios, M. Aldea, M. Vicente, Transcription of ftsZ oscillates during the cell cycle of Escherichia coli. *EMBO J.* **12**, 3957–3965 (1993).

73. C. E. Carty, L. O. Ingram, Lipid synthesis during the Escherichia coli cell cycle. *J. Bacteriol.* **145**, 472–478 (1981).

74. A. N. Gray, A. J. F. Egan, I. L. Van't Veer, J. Verheul, A. Colavin, A. Koumoutsi, J. Biboy, A. F. M. Altelaar, M. J. Damen, K. C. Huang, J.-P. Simorre, E. Breukink, T. den Blaauwen, A. Typas, C. A. Gross, W. Vollmer, Coordination of peptidoglycan synthesis and outer membrane constriction during Escherichia coli cell division. *Elife*. **4** (2015), doi:10.7554/eLife.07118.

75. H. Strahl, L. W. Hamoen, Membrane potential is important for bacterial cell division. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 12281–12286 (2010).

76. P. Moldéus, I. A. Cotgreave, M. Berggren, Lung protection by a thiol-containing antioxidant: N-acetylcysteine. *Respiration*. **50 Suppl 1**, 31–42 (1986).

77. S.-M. Shin, S.-H. Song, J.-W. Lee, M.-K. Kwak, S.-O. Kang, Methylglyoxal synthase regulates cell elongation via alterations of cellular methylglyoxal and spermidine content in Bacillus subtilis. *Int. J. Biochem. Cell Biol.* **91**, 14–28 (2017).

78. J. Hartl, P. Kiefer, A. Kaczmarczyk, M. Mittelviefhaus, F. Meyer, T. Vonderach, B. Hattendorf, U. Jenal, J. A. Vorholt, Untargeted metabolomics links glutathione to bacterial cell cycle progression. *Nat Metab*. **2**, 153–166 (2020).

79. S. Narayanan, B. Janakiraman, L. Kumar, S. K. Radhakrishnan, A cell cycle-controlled redox switch regulates the topoisomerase IV activity. *Genes Dev.* **29**, 1175–1187 (2015).

80. N. S. Britt, N. Patel, T. I. Shireman, W. I. El Atrouni, R. T. Horvat, M. E. Steed, Relationship between vancomycin tolerance and clinical outcomes in Staphylococcus aureus bacteraemia. *J. Antimicrob. Chemother.* **72**, 535–542 (2017).

81. M. R. Brown, P. J. Collier, P. Gilbert, Influence of growth rate on susceptibility to antimicrobial agents: modification of the cell envelope and batch and continuous culture studies. *Antimicrob. Agents Chemother.* **34**, 1623–1628 (1990).

82. W. Abdelhady, A. S. Bayer, K. Seidl, D. E. Moormeier, K. W. Bayles, A. Cheung, M. R. Yeaman, Y. Q. Xiong, Impact of vancomycin on sarA-mediated biofilm formation: role in persistent endovascular infections due to methicillin-resistant Staphylococcus aureus. *J. Infect. Dis.* **209**, 1231–1240 (2014).

83. H. Kaneko, H. Nakaminami, K. Ozawa, T. Wajima, N. Noguchi, In vitro anti-biofilm effect of anti-methicillin-resistant Staphylococcus aureus (anti-MRSA) agents against the USA300 clone. *J Glob Antimicrob Resist*. **24**, 63–71 (2021).

84. Q. Wang, L. Gu, A. Adey, B. Radlwimmer, W. Wang, V. Hovestadt, M. Bähr, S. Wolf, J. Shendure, R. Eils, C. Plass, D. Weichenhan, Tagmentation-based whole-genome bisulfite sequencing. *Nat. Protoc.* **8**, 2022–2032 (2013).

85. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

86. M. Levin, L. Anavy, A. G. Cole, E. Winter, N. Mostov, S. Khair, N. Senderovich, E. Kovalev, D. H. Silver, M. Feder, S. L. Fernandez-Valverde, N. Nakanishi, D. Simmons, O. Simakov, T. Larsson, S.-Y. Liu, A. Jerafi-Vider, K. Yaniv, J. F. Ryan, M. Q. Martindale, J. C. Rink, D. Arendt, S. M. Degnan, B. M. Degnan, T. Hashimshony, I. Yanai, The mid-developmental transition and the evolution of animal body plans. *Nature*. **531**, 637–641 (2016).

87. H. Zalts, I. Yanai, Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nat Ecol Evol*. **1**, 113 (2017).

88. S. R. Jammalamadaka, S. Rao Jammalamadaka, A. SenGupta, Topics in Circular Statistics. *Series on Multivariate Analysis* (2001), , doi:10.1142/4031.

89. Stan Development Team, *RStan: the R interface to Stan* (2021; https://mc-stan.org/).

90. L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. Vanderplas, A. Joly, B. Holt, G. Varoquaux, API design for machine learning software: experiences from the scikit-learn project. *arXiv [cs.LG]* (2013), , doi:10.48550/ARXIV.1309.0238.

91. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck 3rd, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive Integration of Single-Cell Data. *Cell*. **177**, 1888–1902.e21 (2019).

92. S. Seabold, J. Perktold, "Statsmodels: Econometric and statistical modeling with python" in *Proceedings of the 9th Python in Science Conference* (SciPy, 2010; https://conference.scipy.org/proceedings/scipy2010/seabold.html).

93. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

94. M. R. Kiedrowski, J. S. Kavanaugh, C. L. Malone, J. M. Mootz, J. M. Voyich, M. S. Smeltzer, K. W. Bayles, A. R. Horswill, Nuclease modulates biofilm formation in community-associated methicillin-resistant Staphylococcus aureus. *PLoS One*. **6**, e26714 (2011).

95. M. A. Benson, S. Lilo, T. Nygaard, J. M. Voyich, V. J. Torres, Rot and SaeRS cooperate to activate expression of the staphylococcal superantigen-like exoproteins. *J. Bacteriol.* **194**, 4355–4365 (2012).

96. T. Bae, E. M. Glass, O. Schneewind, D. Missiakas, Generating a collection of insertion mutations in the Staphylococcus aureus genome using bursa aurealis. *Methods Mol. Biol.* **416**, 103–116 (2008).

97. P. D. Fey, J. L. Endres, V. K. Yajjala, T. J. Widhelm, R. J. Boissy, J. L. Bose, K. W. Bayles, A genetic resource for rapid and comprehensive phenotype screening of nonessential Staphylococcus aureus genes. *MBio*. **4**, e00537–12 (2013).

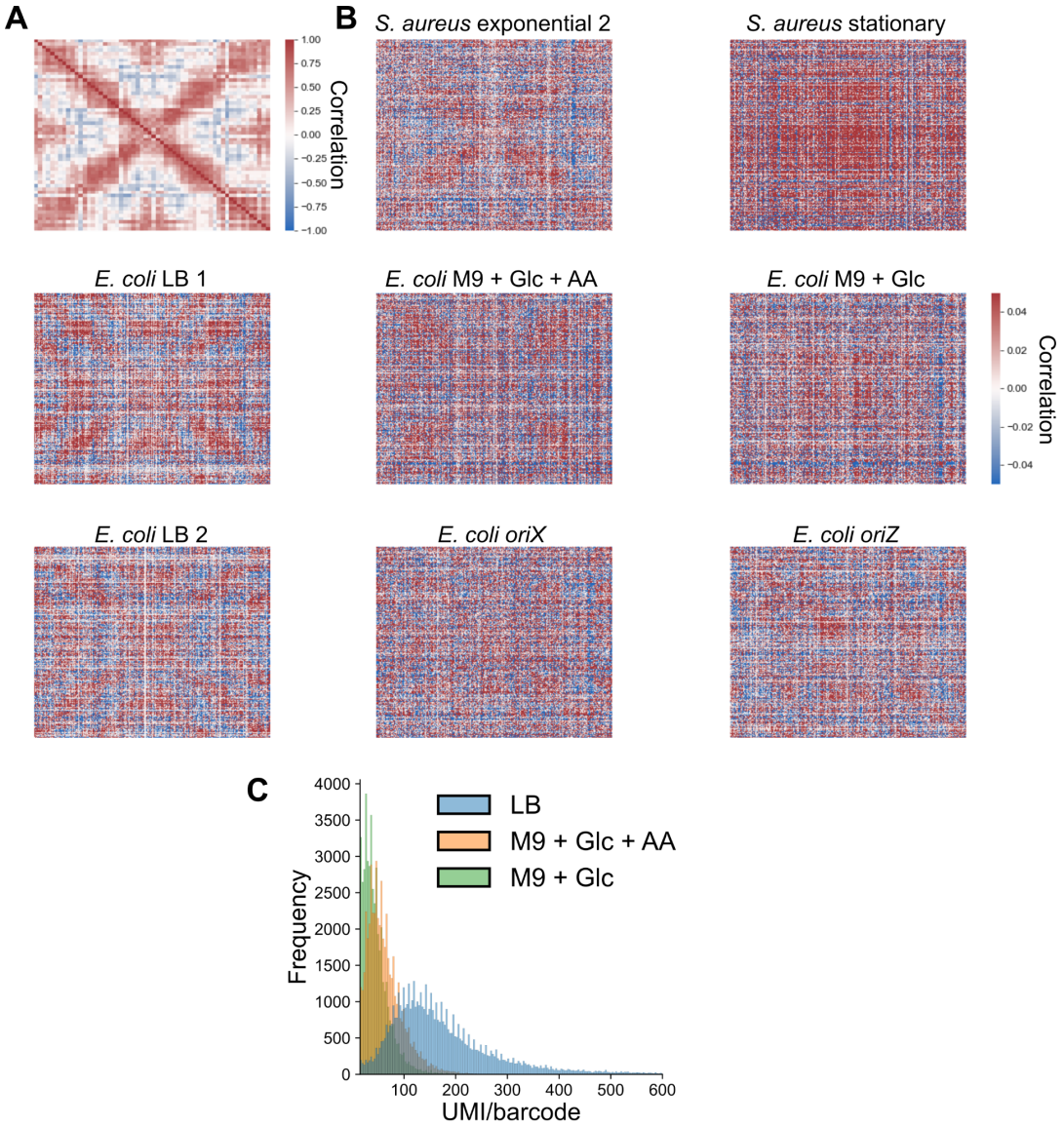

Figure S1: Global Spearman correlation patterns in PETRI-seq datasets. A) Gene-gene correlations in exponential *S. aureus* data from Dataset D4, averaged in 50 kb bins, as for Dataset D3 in Fig. 1F. **B)** Initial correlations from unbinned, scVI-predicted gene expression data. Sample "*S. aureus* exponential 2" is from Dataset D4, whereas *E. coli* LB replicates 1 and 2 are from Dataset D1 and Dataset D2, respectively. **C)** Histogram of total UMI/barcode under each growth condition. Median UMI/barcode was 152 (LB), 56 (M9 + Glc + AA), and 40 (M9 + Glc) after filtering barcodes with less than 15 or greater than 2,000 UMI. This loss of UMI may explain the weaker correlation signals in slower-growing cell populations.

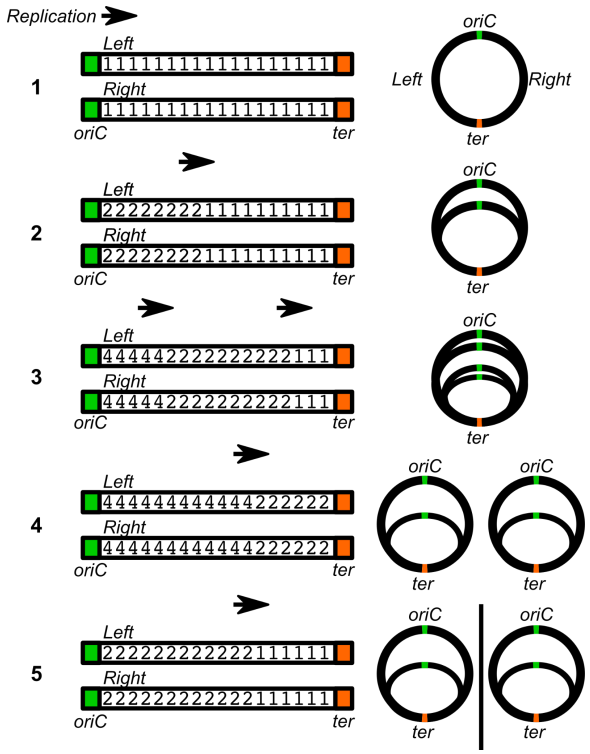

Figure S2: Schematic figure for the simulation of the effect of DNA replication on gene-gene correlations. Each “arm” of the circular chromosome is represented as an array of integers (initially ones), representing each gene. Replication proceeds stepwise from origin to terminus, doubling copy number as it does (steps 1 to 2). At high replication rates, a second round of replication will initiate before the first has finished (step 3). When one round of replication reaches the terminus, that round finishes and after a given time interval copy numbers are globally halved, reflecting cell division (steps 4 to 5). Figures on the right indicate the represented states on the circular chromosome.

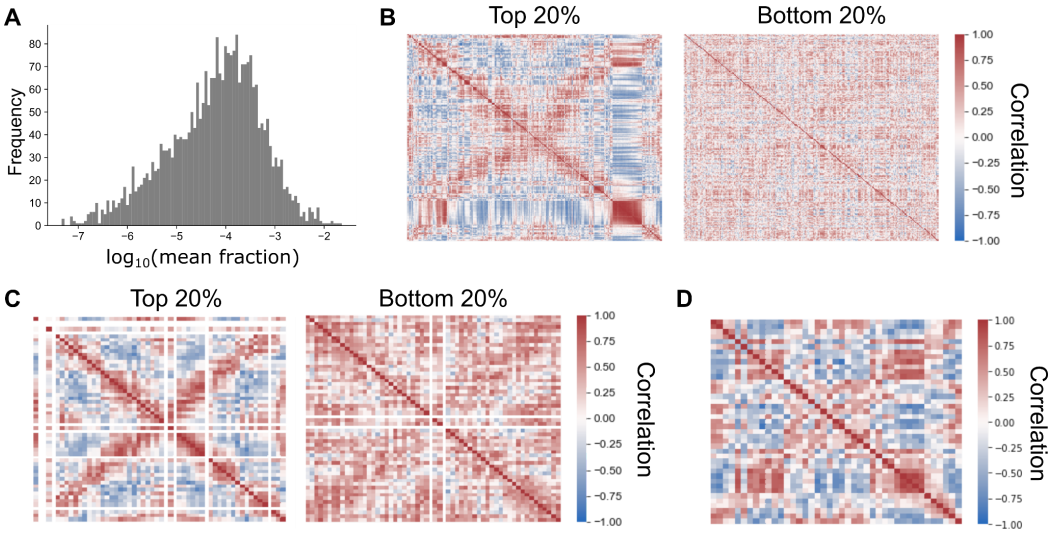

Figure S3: Evidence to demonstrate that the global gene covariance pattern results directly from gene expression. A) Histogram of average gene expression varies over several orders of magnitude. Raw expression counts were normalized by library size (to sum to 1 per barcode) and the average expression was calculated. **B)** Spearman correlations between genes in the top and bottom 20% of genes. Genes are arranged by chromosome order. **C)** Spearman correlations between top and bottom 20% of genes after averaging expression in 50 kb bins as in Fig. 1F. **D)** Gene expression correlations in synchronized *C. crescentus* bulk RNA-seq from (*2*). Scaled gene expression is averaged into 100 kb bins.

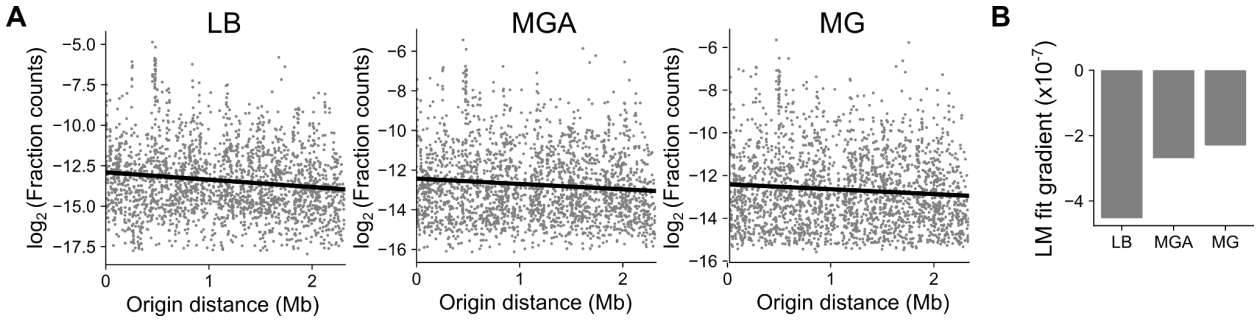

Figure S4: The relationship between origin distance and expression levels. A) For each *E. coli* growth condition, the average fraction of total mRNA UMI from each gene was calculated and $\log_2$-transformed. A linear regression model (black line) was fitted between log-fraction counts and origin distance. **B)** The gradient of the linear model fits in **(A)**. Note that in each case, there is a negative relationship, with a steeper gradient for faster growth rates. This is expected given that at fast growth rates, genes near the origin may attain higher copy number states (>2) than at slow growth rates. Spearman correlations are -0.13 (LB, $P = 3.8 \times 10^{-10}$), -0.09 (MGA, $P = 2.2 \times 10^{-5}$), and -0.07 (MG, $P = 6.0 \times 10^{-4}$).

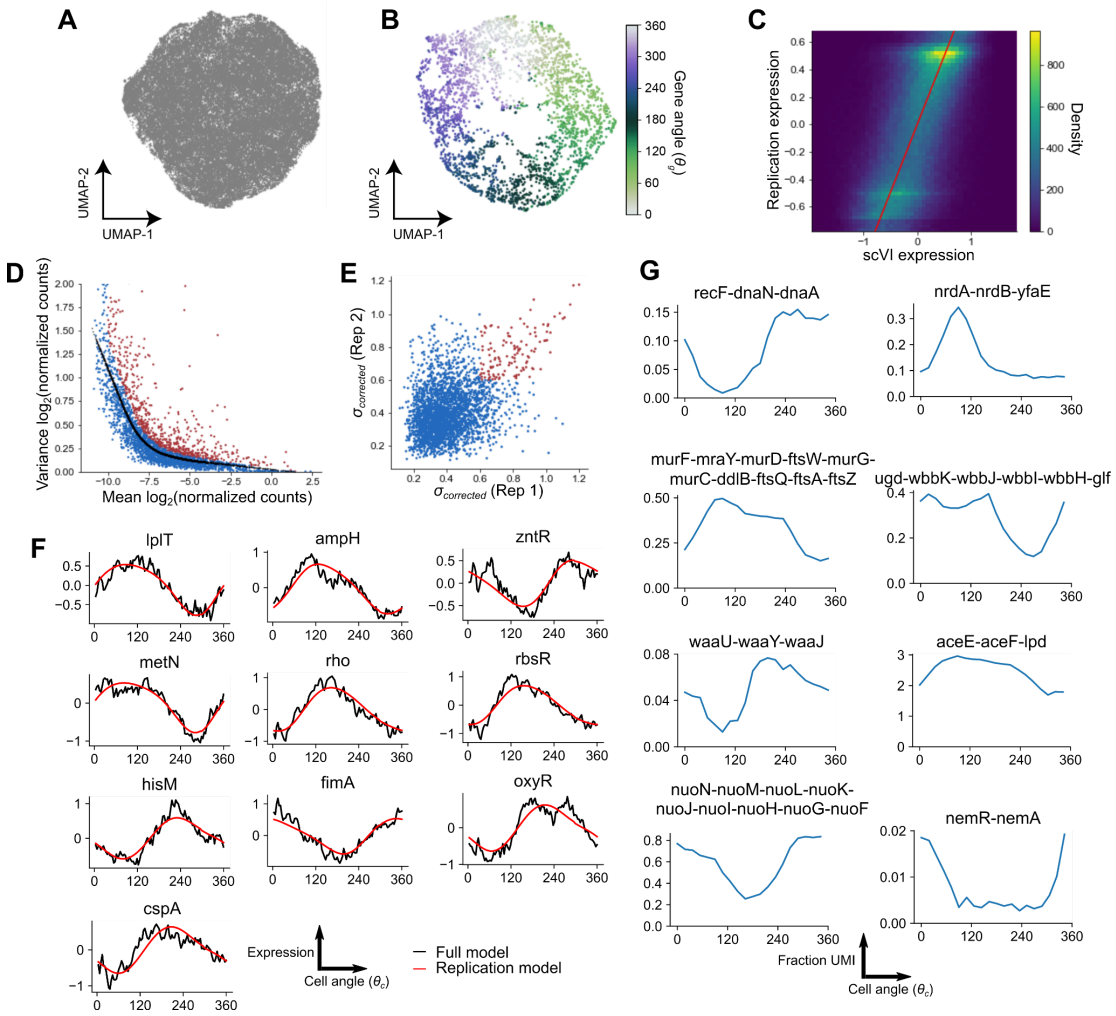

Figure S5: Cycle analysis of LB-grown *E. coli*. A) UMAP analysis of cells based on scVI-predicted expression. **B)** UMAP of genes, performed on the same data as the PCA in Fig. 3C. Gene angles shown are those derived from PCA. **C)** Two-dimensional histogram showing the relationship between observed expression from scVI and replication-predicted expression. Expression is averaged in 100 bins by cell angle θ_c and both matrices are flattened to one-dimensional vectors. The red line indicates $x = y$ i.e. the case where expression in both matrices is identical. **D)** The mean-variance relationship of log-transformed normalized counts. The black line indicates the locally weighted scatterplot smoothing (LOWESS)-fitted values and red points are genes classed as highly variable. See Materials & Methods for further details. **E)** Comparison of $\sigma_{corrected}$ (standard deviation of divergence from the replication model) between LB-grown *E. coli* in Dataset D1 and Dataset D2 of all genes present in both datasets. Red indicates $\sigma_{corrected} > 0.6$ in both datasets, meaning that they are considered replication-divergent. The Pearson correlation between replicates is 0.38. **F)** Cycle gene expression plots as in Fig. 4D of the ten highly variable genes with the lowest $\sigma_{corrected}$. **G)** Percent transcripts (calculated as the fraction of total mRNA UMI of a given transcript multiplied by 100) averaged across 20 bins by cell angle θ_c. Operons are selected from those displayed in Fig. 4D.

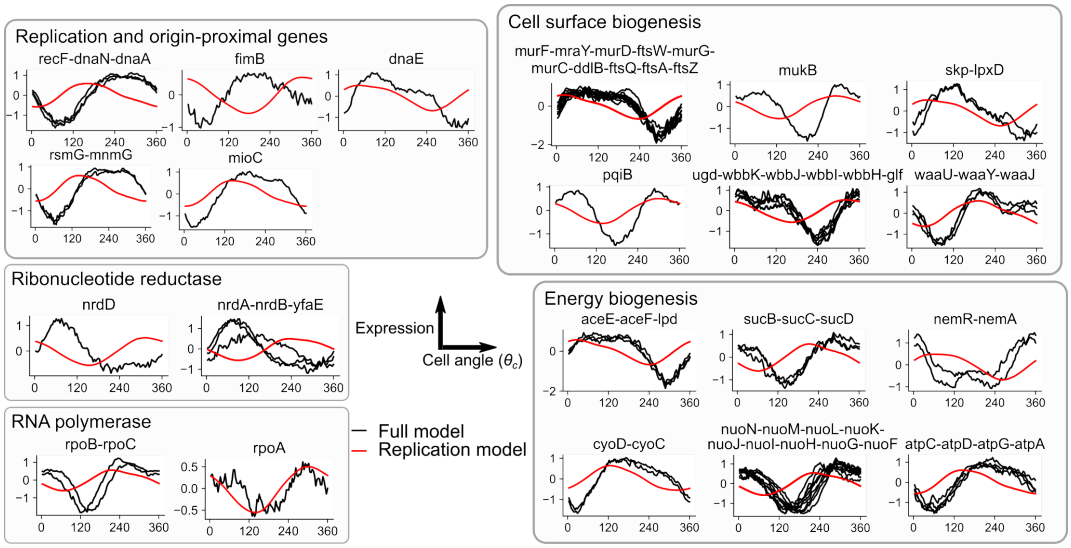

Figure S6: Cell cycle plots of divergent *E. coli* genes from another replicate. Plots are as in Fig. 4D, but taken from a second independent replicate (Dataset D2).

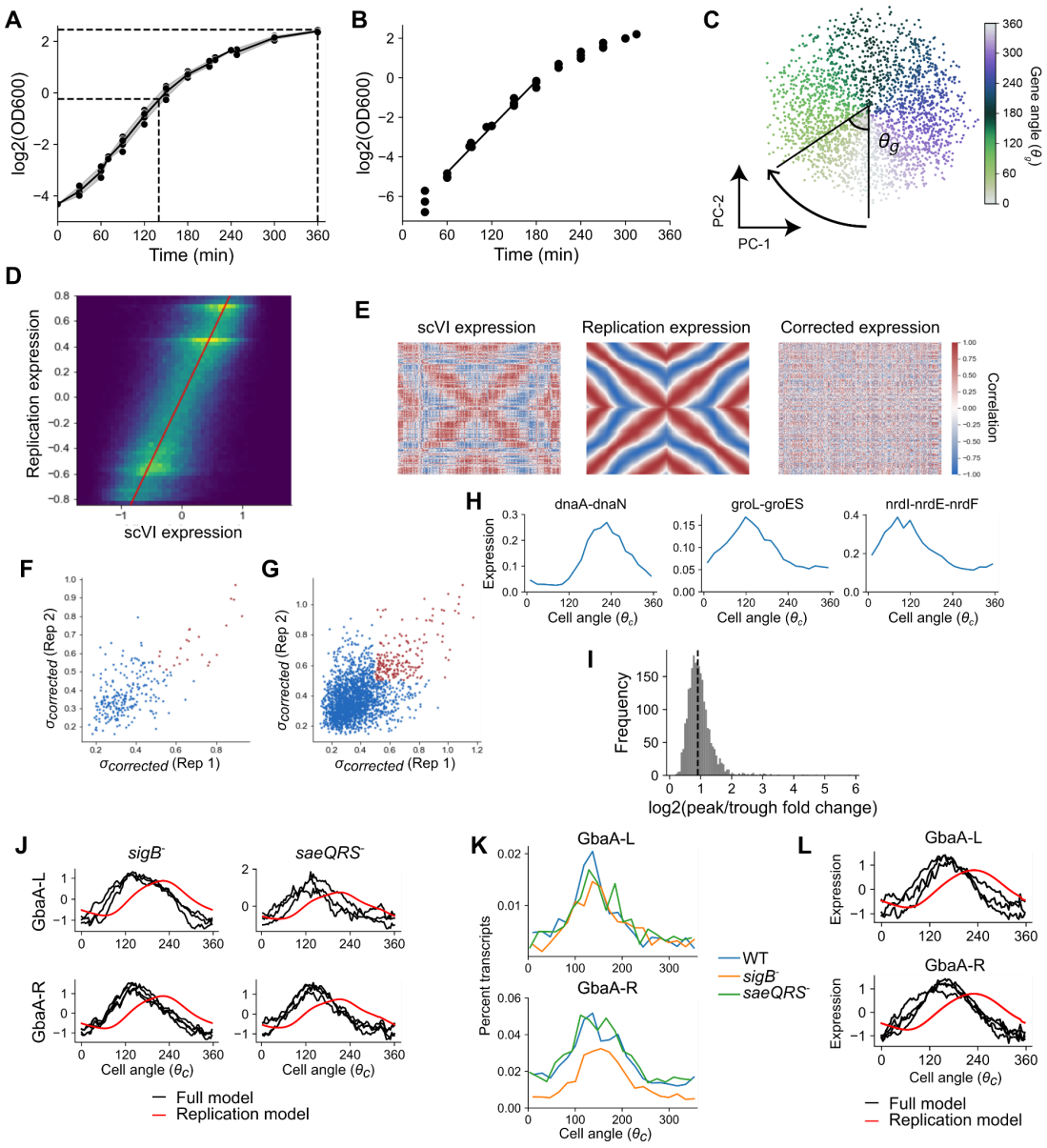

Figure S7: Cycle analysis of *Staphylococcus aureus*. A) Growth of *S. aureus* under standard growth conditions. The time and $\log_2(A_{600})$ values when exponential and stationary phase samples were taken are marked with dotted lines. The line is fitted to the mean at each time point, with the gray area representing standard deviation. Data are from five biological replicates. Doubling times for exponentially growing cells are estimated for the linear portion of the curve (~60-150 min). **B)** Growth of *S. aureus* under balanced growth conditions (see Materials & Methods). The black line indicates the linear portion from which doubling time was estimated. Data are from three biological replicates. **C)** Derivation of gene angle θ_g. Principal component analysis was performed on the transpose of the matrix in Fig. 5C to project genes into two dimensions by expression. Genes are colored by the angle θ_g between principal components 1 and 2 relative to the center of the projection. **D)** Two-dimensional histogram showing the relationship between observed expression from scVI and replication-predicted expression (Pearson's r = 0.67). Expression is averaged in 100 bins by cell angle θ_c and both matrices are flattened to one-dimensional vectors. The red line indicates $x = y$ i.e. the case where expression in both matrices is identical. **E)** Gene-gene correlations across θ_c-binned expression data (100 bins) for the full scVI model

(*left*), the replication-only model (*middle*), and the corrected model that is the difference of the two expression matrices (*right*). Note that correction largely eliminates the global correlation pattern. **F & G)** Comparison of $\sigma_{corrected}$ (standard deviation of divergence from the replication model) between *S. aureus* in Dataset D5 and Dataset D6 for highly variable genes in both datasets **(F)** (Pearson's r = 0.66) and all genes **(G)** (Pearson's r = 0.48). Red indicates $\sigma_{corrected} > 0.5$ in both datasets, meaning that they are considered replication-divergent. **H)** Percent transcripts (calculated as the fraction of total mRNA UMI of a given transcript multiplied by 100) averaged across 20 bins by cell angle θ_c, shown for select replication-divergent genes. **I)** Distribution of peak-trough fold changes calculated as described in Materials & Methods. The dotted black line indicates the position of the median (1.9-fold, shown here on a $\log_2$ scale). **J)** Cycle gene expression plots of left and right GbaA-responsive operons as in Fig. 5E, but shown for *S. aureus* strains with deletions of the *sigB* and *saeQRS* regulators. **K)** Percent transcripts for GbaA-L and GbaA-R operons displayed as in Fig. 5I but comparing wild-type USA300 to *sigB*$^-$ and *saeQRS*$^-$ mutants. Note that the pattern is the same but the *sigB*$^-$ mutant shows slightly lower overall expression. **L)** Cycle gene expression plots of left and right GbaA-responsive operons as in **(I)** but in wild-type cells treated with 10 mM N-acetylcysteine.

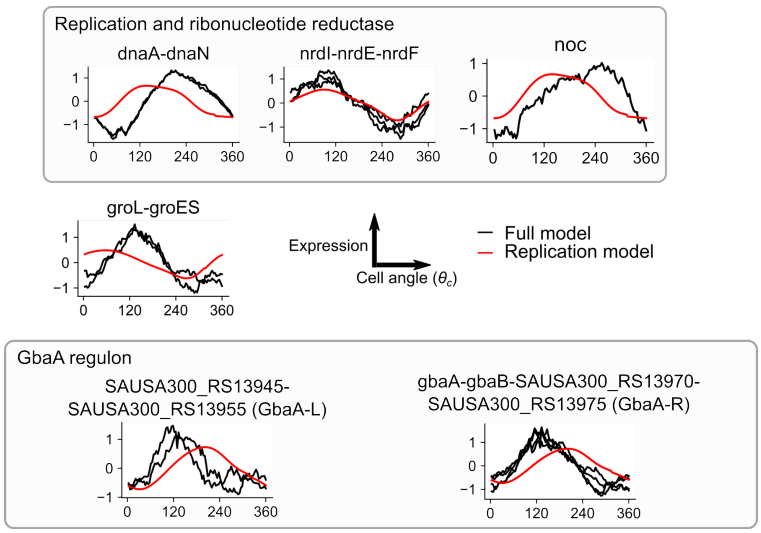

Figure S8: Cell cycle plots of divergent *S. aureus* genes from another replicate. Plots are as in Fig. 5F, but taken from a second independent replicate (Dataset D6).

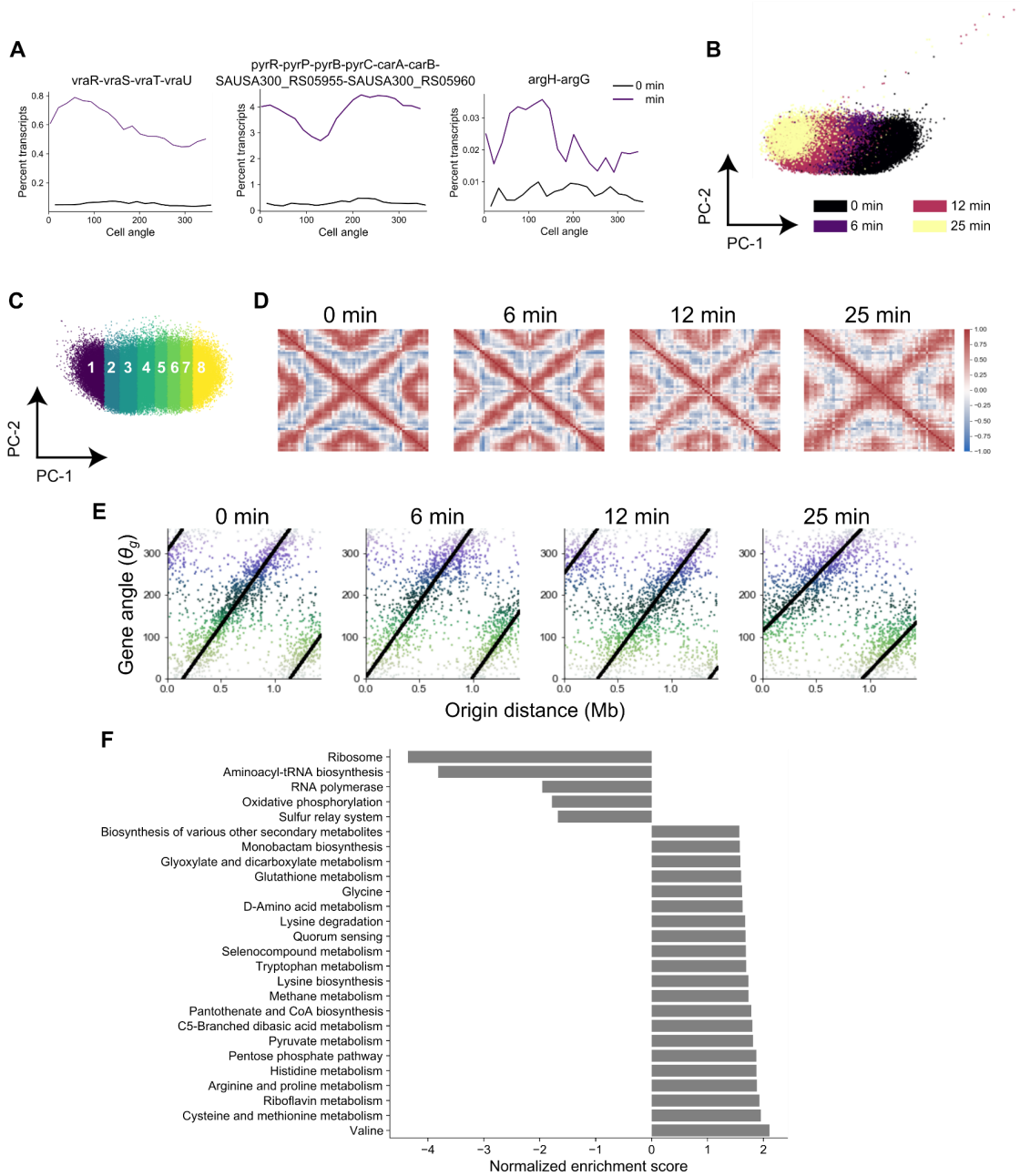

Figure S9: Analysis of *S. aureus* responses to vancomycin treatment. A) Comparison of cell cycle gene expression in untreated and 6 minute-treated cells for the genes in Fig. 6D. Expression is shown as percent transcripts (calculated as the fraction of total mRNA UMI of a given transcript multiplied by100) averaged across 50 bins by cell angle θ_c. **B)** Full PCA plot for time-dependent changes in vancomycin expression. For Fig. 6E, outlier cells are not shown and the orientation of PC-1 is reversed (so that PC-1 corresponds positively with time, since PC directions are arbitrary). **C)** PCA as in Fig. 6E colored by the eight PC1 bins as in Fig. 6F. **D)** Global gene-gene correlation plots as in Fig. 1F but for each time point of vancomycin treatment. **E)** Gene angle-origin distance plots as in Fig. 3D but for each time point. Note that the gradient at 25 min is substantially reduced compared to other time points. **F)** Results of GSEA (*39*, *40*) with KEGG pathways *(74, 75)* on PC-1 loadings (reversed as in Fig. 6E). Normalized enrichment scores (*39*) are shown for all pathways significantly enriched (false discovery rate = 0.1). GSEA was performed using the "gseapy" Python package v0.10.8.

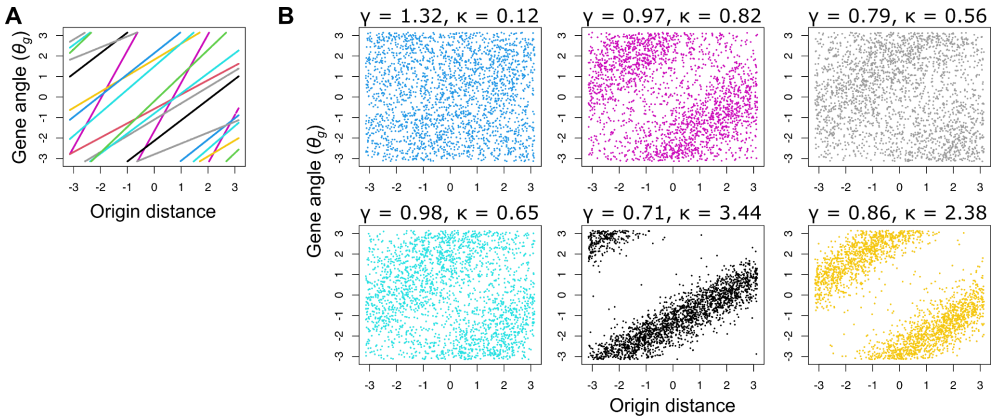

Figure S10: Sampling from the prior of the gene angle-origin distance regression model. Based on the model and priors specified in Materials & Methods, values were randomly sampled from the prior and used to predict either the expected gene angle *A* **(A)** or the predicted value of gene angle θ_g after von Mises sampling **(B)**. For each sampled set of parameters in **(B)** the gradient γ and concentration parameter κ are shown. Both θ_g and origin distance *D* are standardized to the range -π to π as per the model requirements. Overall, the prior assumptions of the model are that there is a positive, linear relationship between θ_g and *D*, but there is considerable flexibility regarding the gradient (and hence degree of wrapping), value of θ_g at $D = 0$, and noise.

Table S1: Information about datasets and samples used. A_{600} refers to the optical density at the time of harvesting. *Growth *E. coli* MG1655 in LB was measured in a separate series of experiments for each dataset.

Dataset	Sample	Strain	Medium	A_{600}	Doubling time (min)	# cells	Median mRNA UMI/barcode
D1	eco_lb_1	*E. coli* MG1655	LB	0.15	26.0 ± 1.3 (n = 4)*	57,627	152
	eco_mga_1	*E. coli* MG1655	M9 + Glc + AA	0.185	39.4 ± 2.3 (n = 4)	50,920	56
	eco_mg_1	*E. coli* MG1655	M9 + Glc	0.062	69.1 ± 9.8 (n = 3)	45,898	40
D2	eco_lb_2	*E. coli* MG1655	LB	0.152	27.0 ± 1.6 (n = 4)*	69,396	93
	eco_orix_1	*E. coli* MG1655 ΔlacIZYA oriX-<> (*18*, *20*)	LB	0.127	27.2 ± 2.4 (n = 4)	25,967	97
	eco_oriz_1	*E. coli* MG1655 ΔlacIZYA oriZ-<> (*19*)	LB	0.14	26.6 ± 2.1 (n = 4)	32,151	100
D3	sau_tsb_1	*S. aureus* USA300 LAC	TSB	0.97	30.1 ± 0.8 (n = 5)	73,053	135
D4	sau_exp_plus	*S. aureus* USA300 LAC	TSB	1.12	30.1 ± 0.8 (n = 5)	13,075	87
	sau_exp_minus	*S. aureus* USA300 LAC	TSB	1.12	30.1 ± 0.8 (n = 5)	8,182	57
	sau_stat_plus	*S. aureus* USA300 LAC	TSB	5.76	NA	40,772	27
	sau_stat_minus	*S. aureus* USA300 LAC	TSB	5.76	NA	15,122	24
D5	sau_wt_1	*S. aureus* USA300 LAC	TSB	0.088	24.9 ± 0.6 (n = 3)	49,307	159
	sau_sigb_1	*S. aureus* USA300 LAC (Erm^S) *ΔsigB (94)*	TSB	0.092	25.4 ± 0.4 (n = 3)	55,936	147
	sau_saeqrs_1	*S. aureus* USA300 LAC (Erm^S) *saeQRS::spec (95)*	TSB	0.076	25.8 ± 2.1 (n = 3)	39,662	174
D6	sau_wt_2	*S. aureus* USA300 LAC	TSB	0.112	24.9 ± 0.6 (n = 3)	38,426	136
	sau_je2_1	*S. aureus* JE2	TSB	0.107	NA	46,719	107
	sau_gbaa_1	*S. aureus* JE2 *SAUSA300_2515:: Tn(Erm)* (Nebraska library # NE355) (*96*, *97*)	TSB	0.103	NA	37,985	109
	sau_nac_1	*S. aureus* USA300 LAC	TSB	0.105	NA	52,071	117
D7	sau_wt_3	*S. aureus* USA300 LAC	TSB	NA	24.9 ± 0.6 (n = 3)	31,852	152
	sau_vanc6_1	*S. aureus* USA300 LAC	TSB	NA	NA	22,344	145
	sau_vanc12_1	*S. aureus* USA300 LAC	TSB	NA	NA	24,064	94
	sau_vanc25_1	*S. aureus* USA300 LAC	TSB	NA	NA	14,832	202