Article

Transcription–replication interactions reveal bacterial genome regulation

https://doi.org/10.1038/s41586-023-06974-w

Received: 22 March 2023

Accepted: 14 December 2023

Published online: 24 January 2024

Check for updates

Andrew W. Pountain[1], Peien Jiang[1,2], Tianyou Yao[3], Ehsan Homaee[3,4,10], Yichao Guan[3,10], Kevin J. C. McDonald[3,10], Magdalena Podkowik[5], Bo Shopsin[5,6], Victor J. Torres[6,7], Ido Golding[3,8] & Itai Yanai[1,9] ✉

Organisms determine the transcription rates of thousands of genes through a few modes of regulation that recur across the genome[1]. In bacteria, the relationship between the regulatory architecture of a gene and its expression is well understood for individual model gene circuits[2,3]. However, a broader perspective of these dynamics at the genome scale is lacking, in part because bacterial transcriptomics has hitherto captured only a static snapshot of expression averaged across millions of cells[4]. As a result, the full diversity of gene expression dynamics and their relation to regulatory architecture remains unknown. Here we present a novel genome-wide classification of regulatory modes based on the transcriptional response of each gene to its own replication, which we term the transcription–replication interaction profile (TRIP). Analysing single-bacterium RNA-sequencing data, we found that the response to the universal perturbation of chromosomal replication integrates biological regulatory factors with biophysical molecular events on the chromosome to reveal the local regulatory context of a gene. Whereas the TRIPs of many genes conform to a gene dosage-dependent pattern, others diverge in distinct ways, and this is shaped by factors such as intra-operon position and repression state. By revealing the underlying mechanistic drivers of gene expression heterogeneity, this work provides a quantitative, biophysical framework for modelling replication-dependent expression dynamics.

Bacterial gene regulation occurs primarily at the level of transcription[5], and decades of research have produced a wealth of knowledge about RNA polymerase and its interactions with promoters, repressors and activators of transcription. However, this work has been based primarily on measurements averaged across a population of millions of cells, thus limiting our resolution of gene circuits. Unlike in eukaryotic cells, transcription in rapidly proliferating bacteria occurs on a chromosome that is under continuous replication[6,7]. Although there has been some exploration of the effects of replication on individual genes[8,9], the transcriptome-wide consequences of this perturbation are largely unknown[10,11]. Traditionally, measuring global gene expression during the replication cycle has been hampered by the requirement for analysis of synchronized populations at a bulk level, limiting this analysis to organisms such as *Caulobacter crescentus*[12–14], where natural biological features facilitate synchronization, or to populations synchronized by batch synchronization methods such as starvation[15] or temperature shift[16] that may be both of questionable efficacy and liable to introduce artefacts[17]. Bacterial single-cell RNA sequencing[18–21] (scRNA-seq) has recently emerged as a tool to understand variation in gene expression in unperturbed, unsynchronized bacterial populations. By applying this approach to two distant species under rapid growth conditions, *Staphylococcus aureus* and *Escherichia coli*, we uncovered unexpected drivers of gene expression throughout the cell cycle in prokaryotes.

Global gene covariance in bacteria

We first aimed to study gene expression at the single-cell level in proliferating bacterial populations by applying the recently described prokaryotic expression profiling by tagging RNA in situ and sequencing[18] (PETRI-seq) method for scRNA-seq to *S. aureus* cells in exponential phase (Fig. 1a). We detected on average 135 transcripts across 73,053 individual cells (Extended Data Fig. 1a and Supplementary Table 1). As the transcriptome measurements were highly sparse, we denoised them using the single-cell variational inference (scVI) method[22]. Studying gene–gene correlations on a local scale, we observed the expected high correlations between the expression profiles of genes residing in the same operon (Fig. 1b). When we investigated gene–gene correlations on a genomic scale, we discovered a notable X-shaped pattern of gene expression correlations (Fig. 1c). We can break this pattern into three major elements (Fig. 1d): (1) correlation between genes near to each other on the chromosome; (2) correlation between genes equidistant from the origin of replication; and (3) correlation between origin-proximal and

[1]Institute for Systems Genetics, NYU Grossman School of Medicine, New York, NY, USA. [2]Department of Biology, New York University, New York, NY, USA. [3]Department of Physics, University of Illinois at Urbana Champaign, Urbana, IL, USA. [4]Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [5]Department of Medicine, Division of Infectious Diseases, NYU Grossman School of Medicine, New York, NY, USA. [6]Department of Microbiology, NYU Grossman School of Medicine, New York, NY, USA. [7] Department of Host-Microbe Interactions, St. Jude Children's Research Hospital, Memphis, TN, USA. [8]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL, USA. [9]Department of Biochemistry and Molecular Pharmacology, NYU Grossman School of Medicine, New York, NY, USA. [10]These authors contributed equally: Ehsan Homaee, Yichao Guan, Kevin J. C. McDonald. ✉e-mail: Itai.Yanai@nyulangone.org

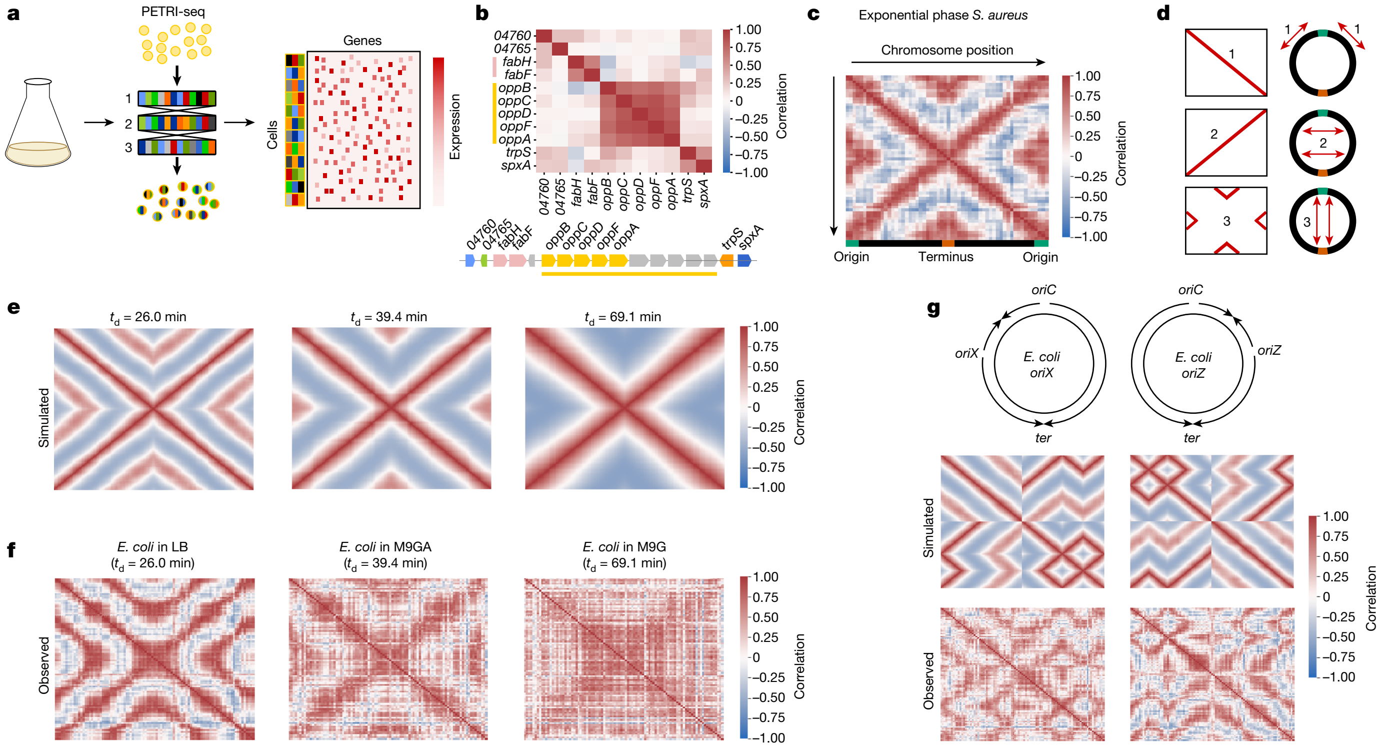

Fig. 1 | scRNA-seq reveals a global pattern of replication-associated gene covariance. a, The PETRI-seq workflow[18]. **b**, Local operon structure is captured by gene–gene correlations (Spearman's *r*) in *S. aureus* strain USA300 LAC. Operons are indicated by shared colours of genes. Grey genes indicate those removed by low-count filtering. **c**, Heat map of the global gene–gene correlations according to chromosomal position. Spearman correlations were calculated based on scVI-smoothed expression averaged in 50-kb bins by chromosome position. **d**, Schematic depicting the individual elements of positive gene–gene correlations in **c** according to their chromosomal locations. **e**, Simulated correlation patterns in unsynchronized *E. coli* populations at three different growth rates. **f**, Spearman correlations between scaled data averaged into 50-kb bins, as for **c**, but for *E. coli* grown at indicated growth rates (see Supplementary Table 1 and Methods). **g**, Top, schematic of predicted replication patterns in two *E. coli* strains with ectopic origins. Middle, predicted correlation patterns based on the copy number simulation. Bottom, real correlation patterns in *oriX* and *oriZ* mutant strains, as in **c**.

terminus-proximal genes. This pattern was also evident in a second independent dataset of 21,257 cells (Extended Data Fig. 1e), however we did not observe it for cells in stationary phase (Extended Data Fig. 1g), suggesting that it is a property of proliferating cells.

This proliferation dependence and correlation of genes equidistant from the origin of replication (Fig. 1d) led us to hypothesize that the X-shaped pattern reflects the effect of DNA replication on gene expression. In the model organism *E. coli*, when the cell doubling time is less than the approximately constant 40–50 min period for one complete round of DNA replication from the origin to the terminus (the C-period), simultaneous overlapping cycles of replication occur[6,23,24]. This leads to growth rate dependence in replication patterns and suggests that any effects of replication on global gene correlations should also be growth rate-dependent. To test this, we developed a simulation to predict growth rate-dependent gene expression correlations arising from replication-dependent changes in gene dosage (Fig. 1e and Extended Data Fig. 1h–k). Of three growth rates simulated, the intermediate growth rate (doubling time (t_d) = 39.4 min) led to a pattern most similar to *S. aureus* (Fig. 1c). 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. 1e). When we measured expression patterns in *E. coli* grown at these three rates, we found that each pattern closely corresponded to its respective simulation (Fig. 1f and Extended Data Fig. 1j). This and further evidence (Extended Data Fig. 1l–p and Methods) demonstrate that replication-driven gene dosage changes are a plausible mechanism driving chromosome-wide expression correlation patterns.

To further test our ability to predict global gene expression correlations from expected replication patterns, we examined strains in which normal replication is perturbed. We studied two *E. coli* strains with ectopic origins of replication at either the 9 o'clock (*oriX*) or 3 o'clock (*oriZ*) positions in addition to *oriC*[25–27]. In these strains, replication was shown to initiate simultaneously at both native and inserted origins, while ending at the same terminus[25]. Again, our simulation effectively predicted the perturbed correlation patterns in these strains (Fig. 1g). Collectively, these results support the notion that DNA replication produces a predictable effect on transcriptional heterogeneity within a population of proliferating bacteria, and that this effect is sensitive to growth rate and genetic perturbations.

Cell cycle state inference

As our gene-level analysis revealed the effect of replication on gene expression, we next carried out a cell-level analysis that exploited this insight to resolve individual cells on the basis of their replication state. To observe cell–cell relationships, we projected the transcriptomes of LB-grown *E. coli* cells into a two-dimensional space after collapsing the expression of individual genes into 100-kb regions to strengthen the chromosome position-dependent signal (as in Fig. 1c). We found a distinctive wheel-shaped arrangement of the cells (Fig. 2a), indicating the capturing of a cyclical process occurring within the population. Hypothesizing that this wheel reflects the cell cycle, we computed a cell angle index (θ_c), which simply orders the cells according to their geometric angle from the centre in this space. Examining gene expression as a function of θ_c, we observed waves of gene expression progressing from the origin to the terminus (Fig. 2b), suggesting that the positions of cells on this wheel indeed reveal their replication state. Similar periodical patterns were observed in simulated data (Extended Data Fig. 1k)

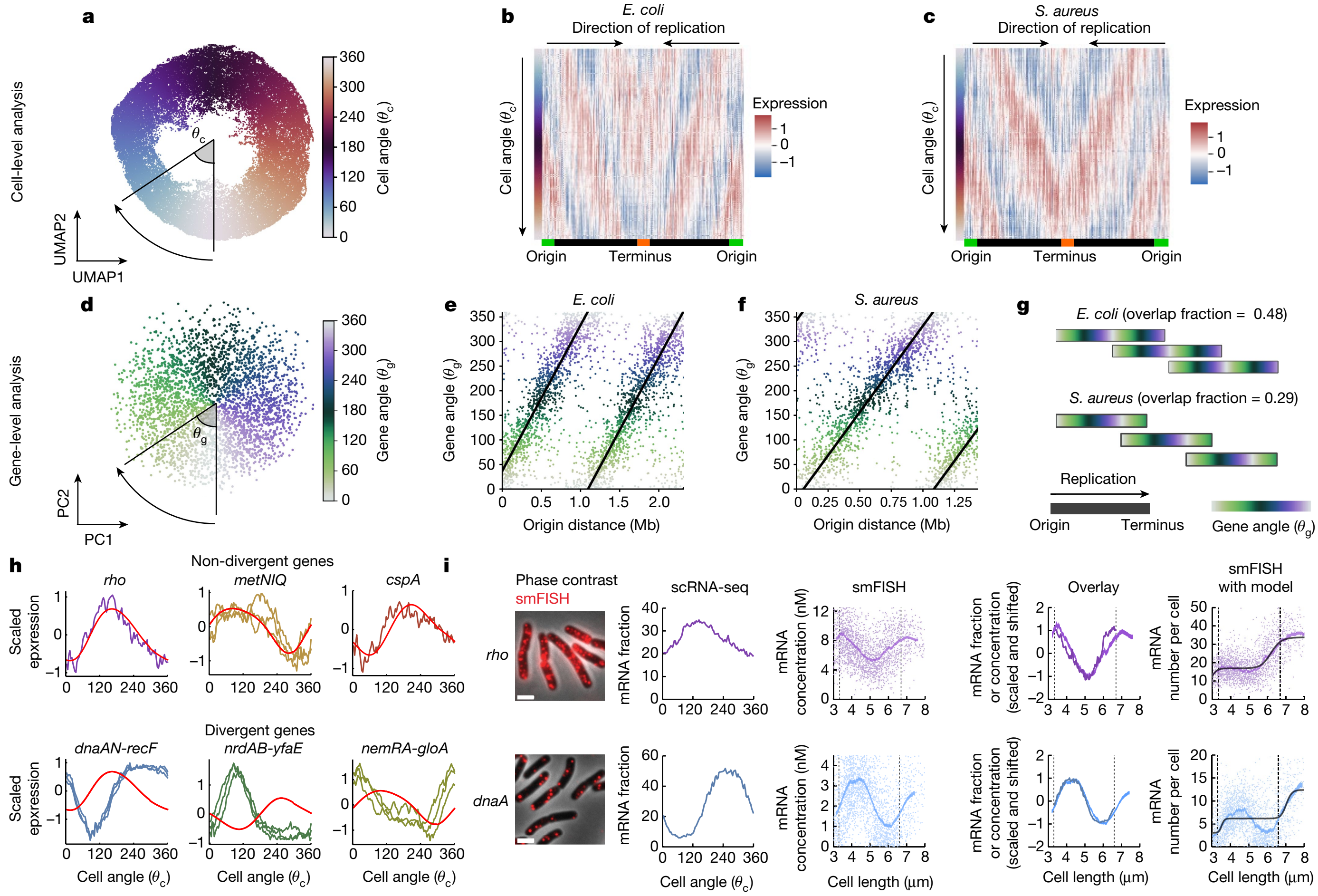

Fig. 2 | Cell cycle analysis of bacterial gene expression. a, Two-dimensional projection by uniform manifold approximation and projection (UMAP) of LB-grown *E. coli* with expression averaged in 100-kb bins by chromosome position. Cell angle θ_c is the angle between UMAP dimensions relative to the centre. **b**,**c**, Heat maps of scaled gene expression in *E. coli* (**b**) and *S. aureus* (**c**) averaged in 100 bins by θ_c. **d**, Principal component analysis (PCA) at the gene level. θ_g is defined as the angle between principal components PC1 and PC2. **e**,**f**, The relationship between θ_g and origin distance for *E. coli* grown in LB (**e**) and *S. aureus* grown in TSB (**f**). **g**, Predicted replication patterns in LB-grown *E. coli* and TSB-grown *S. aureus*. Overlapping replication rounds lead to shared θ_g in simultaneously replicated chromosomal regions. **h**, Expression of genes in operons across 100 bins averaged by θ_c. Expression is shown as *z*-scores derived from scVI (jagged lines) or predicted as a replication effect (smooth, red lines). **i**, Comparison of scRNA-seq and smFISH data for two genes (see Extended Data Fig. 6a for more genes and further details). Left to right: microscopy images of *E. coli* cells labelled using gene-specific smFISH probes; scRNA-seq expression shown as fraction of total cellular mRNA; mRNA concentration, measured using smFISH, as a function of cell length (single-cell measurements are indicated alongside the moving average; dashed lines indicate the inferred mean values at birth and division); alignment of scaled data from smFISH and scRNA-seq measurements; and absolute mRNA copy number, measured using smFISH, as a function of cell length. Data are representative of two independent experiments.

and in *S. aureus* cells (Fig. 2c and Extended Data Fig. 2b). These data suggest that the transcriptome alone can be used to infer the replication state of a cell, and that this holds across different bacterial species.

As we observed that gene expression moved in waves during progression along the cell angle trajectory, we reasoned that we should also be able to order genes according to their expression profiles. We thus developed a similar ordering metric, which we denoted the gene angle (θ_g) (Fig. 2d). Consistent with a role of replication in driving expression patterns, we observed a linear relationship between the θ_g of a gene and its genomic distance from the origin of replication in both species (Fig. 2e,f). This suggested that θ_g may be ordering genes according to their order of replication. The relationship however is not a simple ordering of genes from origin to terminus. For *E. coli*, we observed that the period of θ_g (the chromosomal distance associated with a 360° rotation) was much less than the full origin–terminus distance, meaning that genes at multiple positions on the origin–terminus axis had the same θ_g value (Fig. 2e). This intuitively relates to the fact that at high growth rates, multiple overlapping rounds of replication lead to simultaneous replication of genes at different distances from the origin. To quantify this further, we used the gradient between the θ_g and the distance from the origin to estimate an 'overlap fraction' (Fig. 2g), meaning the fraction of one round of replication happening before the previous one has finished. When we compared *E. coli* at different growth rates, we observed that, in line with expectations[6,23], decreasing proliferation speed in *E. coli* is associated with reduced overlap in rounds of replication (Extended Data Fig. 2d,e). In *S. aureus*, too, we observed that genes close to the origin and terminus had similar θ_g values, in line with our observation of a direct correlation between genes in these regions (Fig. 1c,d). This implies that *S. aureus* can also exhibit multiple rounds of simultaneous replication at fast growth rates.

If the gene angle captures the order of replication of a gene then it should be possible to compute the average speed of DNA polymerase using the doubling time and the θ_g–origin distance gradient. For *E. coli* in LB, this estimate was 780 bp s^{-1} (Extended Data Fig. 2f), which is very close to previously reported values[28,29] of around 800 bp s^{-1}. This would correspond to a C-period of around 50 min to replicate the full 2.3 Mb distance from origin to terminus. In *S. aureus*, we predict a slightly slower replication speed of 687 bp s^{-1}. However, its smaller genome (1.4 Mb from origin to terminus) means that a shorter C-period (around 35 min) is inferred, leading to less overlap in rounds of replication than *E. coli* (Fig. 2g) despite very similar doubling times. Therefore, the gene angle θ_g provides a quantitative and interpretable description of the relationship between gene expression and global replication patterns. Given the relationship between θ_g and the cell angle θ_c (Extended Data Fig. 2g,h), we could therefore devise an inference model that predicts expression of a given gene (by θ_g) at a given point in the cell cycle (by θ_c), purely on the basis of its distance from the origin of replication (Extended Data Fig. 2l). This model effectively captured the global chromosome position-dependent expression pattern (Extended Data Fig. 4a,b) and crucially, gives us a baseline prediction for determining whether or not individual genes behave according to global, replication-driven trends.

Finally, we tested whether we could use this framework to infer replication dynamics under changing growth rates. We considered two scenarios: a growth rate 'shift up' upon transferring *E. coli* into a richer growth medium[30] (Extended Data Fig. 3a–d) and a 'shift down' upon exposing *S. aureus* to high concentrations of the antibiotic vancomycin (Extended Data Fig. 3e–h). Consistent with previous observations[30], we found that in both cases while the transcriptional response to these stimuli happened very rapidly, the shift in the replication pattern (more or less overlapping in growth acceleration or deceleration, respectively), occurred only after a delay. Therefore, our analysis of replication dynamics from scRNA-seq data is not only robust to external perturbations to the global transcriptome, but also provides information on relative changes to replication patterns independent of other transcriptional changes.

Canonical and divergent gene expression

To test whether the wheel-shaped distribution of cells indeed reflected cell cycle-dependent gene expression, we turned to single-molecule fluorescence in situ hybridization[8,31] (smFISH), a scRNA-seq-independent approach that uses microscopy to detect individual transcripts in single cells. We first identified operons whose genes' expression did or did not fit the pattern predicted by our inference model (Fig. 2h). We then compared our measurements for genes within the selected operons to cell cycle-dependent gene expression measurements obtained using smFISH[8,31]. Between scRNA-seq and smFISH, the overall expression levels of the genes were in close quantitative agreement (Extended Data Fig. 5d). The smFISH approach resolves *E. coli* cell cycle states by using cell length to infer cell age, thus defining the cell cycle relative to division timing[8] (that is, the time since cell birth). By contrast, we defined cell angle $\theta_c = 0$ to be the assumed time of replication initiation (Methods). As expected given these differing 'start' points, we observed a phase shift in expression profiles between the two methods that was consistent across genes (Extended Data Fig. 5e). Modelling of total DNA content as a function of cell length supported that this phase shift was roughly consistent with our choice of $\theta_c = 0$ as the point of replication initiation (Extended Data Fig. 5f), albeit with some discrepancy (Methods). By correcting for this phase shift between methods, we aligned the scRNA-seq profile to that of the smFISH data (Fig. 2i and Extended Data Fig. 6a). In doing so, we observed that expression dynamics inferred by the two methods were highly correlated, confirming that our scRNA-seq approach indeed captures cell cycle-dependent expression.

When we analysed cell cycle expression patterns of genes that did or did not diverge from the expected pattern (Fig. 2h), we noted a number of key differences. Both scRNA-seq and smFISH showed that the amplitude of cell cycle expression (that is, the relative change between cell cycle minimum and maximum expression) was higher for these divergent genes than for the non-divergent ones (Extended Data Fig. 5g). Moreover, whereas the scRNA-seq measurements capture only relative expression of a gene as a fraction of total cellular mRNA, the smFISH experiments additionally provide absolute abundance (that is, mRNA copy number). This revealed that in non-divergent genes, there was a discrete twofold stepwise increase in expression (Fig. 2i and Extended Data Fig. 6a,c), consistent with genes that are sensitive to gene dosage but otherwise exhibit constant transcription rates[8]. Divergent genes, however, did not conform to this simple step function (although there was variation between replicates in the low-expressed *nemA*; Extended Data Fig. 6a,c). These observations support an interpretation that the pattern predicted by the inference model corresponds to a canonical cell cycle expression pattern driven by gene dosage: genes that fit this pattern increase in expression only upon their replication, whereas divergent genes are governed by additional factors, leading to a higher amplitude in cell cycle expression than be explained by copy number effects alone.

Expression timing and promoter distance

We next investigated those genes whose phase in cell cycle expression differed from predictions. Divergent genes can vary in phase of expression by peaking either earlier or later in the cell cycle than predicted (Fig. 3a). In *E. coli*, we observed a systemic bias whereby the majority of divergent genes showed delayed expression, meaning that the peak of expression was later than expected based upon chromosomal location (Fig. 3a). Many of these genes were encoded in large operons, such as those involved in energy biogenesis (for example, *nuo* and *atp* operons) and cell surface synthesis (for example, the *mraZ*–*ftsZ* operon). We found that genes with a more distal position within these operons exhibited a greater delay (Fig. 3b and Extended Data Fig. 7a). Globally, this correlation between the delay—measured as 'angle difference' (see Fig. 3a and Supplementary Table 5)—and distance from the transcriptional start site (TSS) pattern was highly significant (Fig. 3c). Moreover, this delay was clearly relative to the timing of replication: in genes whose replication-predicted pattern changed in the *oriZ* mutant, expression also shifted in this strain such that the delay was relative to their new replication time (Fig. 3b).

We hypothesized that this delayed phenotype arises owing to the time for RNA polymerase (RNAP) to reach genes after replication by DNA polymerase (DNAP) has occurred. The speed of RNAP has previously been estimated[8,32] as approximately 40 nucleotides (nt) per second in *E. coli*, much slower than the approximately 800 nt s^{-1} speed for DNAP[28,29] (see also Extended Data Fig. 2f). By performing linear regression to measure the angle difference–transcriptional distance relationship (Fig. 3c) and converting θ_g into time by assuming that 360° is equivalent to one doubling time of 26 min, we inferred that distance from the TSS is associated with a delay that is consistent with an average RNAP speed of 35 nt s^{-1} (33 nt s^{-1} and 38 nt s^{-1} in two replicates; Extended Data Fig. 7c). Therefore, our data support the hypothesis that when a gene is replicated, the time for its expression to increase to the higher-expressed state (due to higher gene dosage) correlates with the time for RNAP to reach that same gene after transcription from the replicated locus restarts.

In addition to the delay, however, we also observed that when examining how expression changes after an operon is replicated, genes close to the TSS immediately increase to a new higher state, as expected from the increase in gene dosage, whereas genes far from the TSS initially drop before then recovering to the new state (Fig. 3d). This manifests as an increasing cell cycle expression amplitude (peak expression versus trough expression) of genes far from their TSS, a trend that was present as a weak but highly significant correlation across the genome (Extended Data Fig. 7d). We can interpret this effect as follows: the replication of an operon produces local disruption of ongoing transcription[33]. For genes close to the TSS, expression can immediately resume at a higher rate from the duplicated locus. However, for genes far from the TSS, the new rounds of transcription may take several minutes to

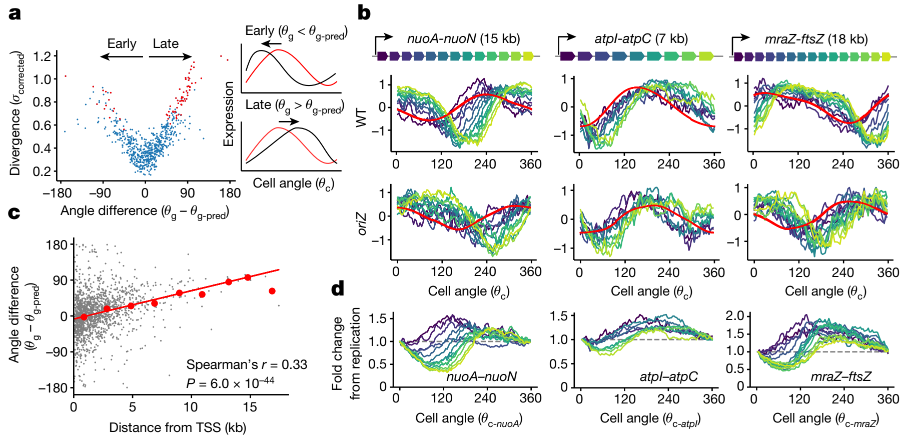

Fig. 3 | Intra-operon position produces a characteristic delay in expression dynamics in *E. coli*. a, Left, plot of divergence from predictions ($\sigma_{corrected}$) against the difference between predicted and observed angles for all genes with high cell cycle variance (Methods). Divergent genes are shown in red. Angle difference (θ_g – predicted θ_g ($\theta_{g\text{-pred}}$)) represents whether a gene is expressed earlier ($\theta_g < \theta_{g\text{-pred}}$) or later ($\theta_g > \theta_{g\text{-pred}}$) than expected (black arrows). Right, schematic of angle difference effects. Black and red lines represent observed and predicted cell cycle expression, respectively. **b**, Cell cycle expression of operons containing 'delayed' genes for wild type (WT) and the *oriZ* mutant, coloured by position within the operon. Model-predicted expression is represented in red. **c**, Plot of distance from the TSS against angle difference. The red line indicates the linear model fit and red points indicate averages of 2-kb bins. The *P* value was calculated based on a two-sided hypothesis test (Methods). **d**, Fold change in expression of genes within an operon relative to expression at the operon's predicted time of replication. Genes are coloured by their position in the operon and cell angle is adjusted so that the zero point is the predicted replication time of the first gene in each operon (*nuoA*, *atpI* and *mraZ*). Dotted grey lines indicate the level of expression upon replication.

reach them, during which time a drop in expression is observed due to mRNA degradation. An implication of this is that internal promoters should buffer the effects of replication-associated loss of transcription. While the *nuo* operon contains no well-evidenced internal promoters, a substantial amount of transcription at the distal end of the *mraZ*–*ftsZ* operon is driven from internal promoters[34,35]. Expression amplitude rises linearly with TSS distance in the *nuo* operon whereas it tails off at the location of the internal promoters within the *mraZ*–*ftsZ* operon (Extended Data Fig. 7b), supporting our inference. Thus internal promoters may prevent excessive fluctuations in key genes such as *ftsZ*, which determines the timing of cell division[36], and in principle it may even be possible to use these signatures to infer the presence of internal promoters within operons.

Finally, we tested whether similar trends could be observed in *S. aureus*. In contrast to *E. coli*, we observed neither an excess of delayed genes among the divergent genes (Extended Data Fig. 7f), nor an effect of distance from the TSS on expression amplitude (Extended Data Fig. 7d). We did however measure a delay as a function of distance from the TSS (Extended Data Fig. 7c), which we found to be consistent with an RNAP elongation speed of 71 nt s^{-1} (59, 64 and 92 nt s^{-1} across three replicates). The difference between the two species persisted even when operons were redefined according to unified criteria[37] (Extended Data Fig. 7e). The RNAP speed of *S. aureus* has not been measured, but in *B. subtilis*, which is—similarly to *S. aureus*—a firmicute of the order Bacillales, experimental measurement of RNAP by a reporter system suggested that it was substantially faster (75–80 nt s^{-1}) than its counterpart in *E. coli* measured by the same method[38,39] (around 48 nt s^{-1}). Therefore, we tested whether we could use our method to detect this faster RNAP speed in *B. subtilis* (Extended Data Fig. 8). Despite multiple additional sources of heterogeneity in this species—including multicellular chain growth, prophage activation and sporulation programs[19]—we nevertheless resolved a highly overlapping replication pattern (Extended Data Fig. 8d,e) from which we inferred a DNAP speed of 632 bp s^{-1} (631 and 634 bp s^{-1} across two replicates), which is close to that of *S. aureus* and, as previously reported[40], around 80% that of the *E. coli* DNAP. However, as predicted, we estimated a faster RNAP speed for *B. subtilis* of 96 nt s^{-1} (Extended Data Fig. 8g) (95 and 98 nt s^{-1} in two replicates). Similar trends could be observed across species for individual long operons (Extended Data Fig. 8i). Therefore, the intra-operon effect that we observe in *E. coli* is not conserved across species, and probably follows from variation in RNAP elongation speed.

Repression-driven expression pulses

Our analysis above revealed recurrent, replication-coupled patterns of cell cycle expression. To enable comparison of replication-coupled expression dynamics across different chromosomal loci, we aligned their mean-standardized expression profiles so that zero on the *x* axis represents the point of a gene's replication and $\theta_{c\text{-rep}}$ represents the cell cycle progression from this point (Fig. 4a). We refer to this as the transcription–replication interaction profile (TRIP) and propose that it enables an explicit focus on the transcriptional response of each gene to the perturbation caused by its replication, independent of when in the cell cycle that gene is replicated. Among genes whose cell cycle expression was reproducible across replicates (Supplementary Table 6), the profile for most genes was similar, rising rapidly owing to a doubling of gene dosage before declining as a relative fraction of the transcriptome as other genes are replicated and thus increase their own fractional abundance. Many genes, however, exhibited patterns that could not be explained by gene dosage effects alone.

To identify the range of behaviours, we partitioned *E. coli* genes into 20 clusters based on their TRIPs (Fig. 4b). Of these, several exhibited particularly divergent expression, differing from the expected pattern in both the timing of their dynamics (for example, when their peak or trough expression occurs) and the amplitude (that is, the relative difference between maximum and minimum cell cycle expression). Cluster 12 in *E. coli* (Ec12) comprised the *nrdAB*–*yfaE* operon and cluster Ec5 contained the *dnaAN*–*recF* operon and other delayed expression genes, including some *nuo* genes. Cluster Ec17 showed an early-peaking pulse in expression with greater amplitude than most genes (Fig. 4c). Many genes in these clusters were in operons that encode repressors, at least some of which have autorepressive activity (including *nemA*, which is co-transcribed with the autorepressor *nemR*) (Supplementary Table 4). Cluster Ec9, whose members peak at the expected time but show increased amplitude (Fig. 4d), also included several repressed genes (Supplementary Table 4), such as the glyoxylate shunt operon *aceBAK*, which is repressed by IclR. Other clusters comprising genes expressed at low levels showed similar trends (Extended Data

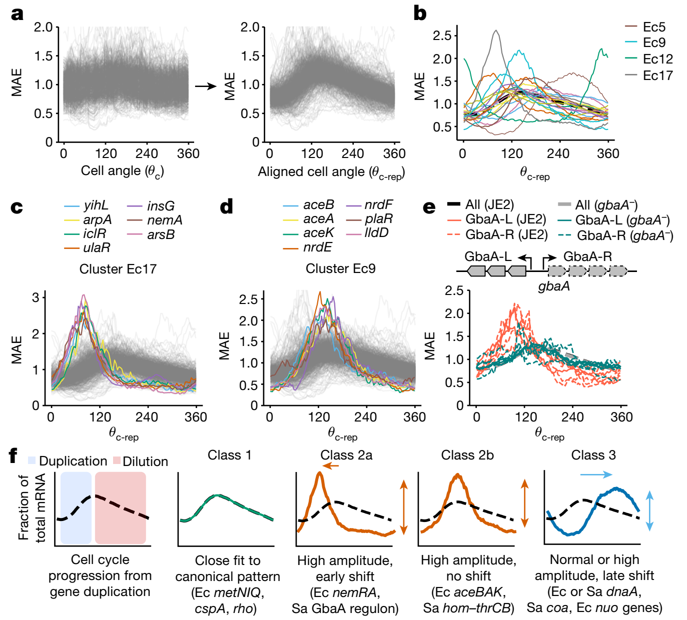

Fig. 4 | TRIPs. a, Procedure to generate TRIPs. Expression is mean-adjusted by division of each gene by its mean (left), then aligned by rotating cell angle so the predicted replication time expression is at zero (see Methods). MAE, mean-adjusted expression. **b**, Average aligned expression profiles for 20 *k*-means clusters in *E. coli*. The dotted black line represents average expression across all reproducible genes. **c**,**d**, Plots of individual genes from cluster Ec17 (**c**) and Ec9 (**d**). **e**, Bottom, TRIPs for GbaA regulon genes in JE2 and a *gbaA*⁻ (*SAUSA300_RS13960*) transposon mutant. Thick black and grey lines represent average expression across all reproducible genes. Top, the operon structure. **f**, TRIP classes. Left, canonical TRIP driven by gene dosage. The other graphs show archetypal classes of TRIPs that adhere to (class 1) or diverge (classes 2 and 3) from this pattern. Genes in *E. coli* and *S. aureus* are preceded by Ec and Sa, respectively.

Fig. 9a), and clusters with low expression that showed high amplitude were enriched for repressor targets (false discovery rate (FDR) = 0.1; Extended Data Fig. 9b). This suggested a broad association between repression state and replication-associated pulses in gene TRIPs.

When we extended this analysis to *S. aureus*, we noted extreme divergence in TRIP clusters within the core genes of genome-integrated mobile genetic elements (MGEs) (Extended Data Fig. 9e,f). Beyond MGE genes, however, a range of behaviours was evident, similar to those observed in *E. coli* (Extended Data Fig. 9c). For example, we observed high amplitude and delayed dynamics in cluster *S. aureus* (Sa) 9, comprising *dnaAN*, as well as several high-amplitude clusters (Extended Data Fig. 9d). Sa18 was almost exclusively composed of genes in the GbaA regulon (Extended Data Fig. 9d,g). By contrast, cluster Sa12 showed delayed dynamics (Extended Data Fig. 9d). Notably, this included several genes involved in stress and virulence.

We reasoned that transcriptional repression could be driving the high-amplitude pulses observed for TRIPs of genes in certain clusters (Ec9, Ec17, Sa11 and Sa18), because of the enrichment of repressors and the low expression levels among high-amplitude clusters (Extended Data Fig. 9a,b), and based on previous observations[8,41,42]. Therefore, we focused on genes of the *S. aureus* GbaA regulon (Extended Data Fig. 9g), which showed a particularly strong early pulse in expression. This regulon consists of two oppositely oriented operons (referred to here as GbaA-L and GbaA-R; Fig. 4e) that are repressed by GbaA. GbaA is an electrophile-sensitive transcriptional repressor encoded by *gbaA* within the GbaA-R operon[43,44]. To test whether GbaA repression was responsible for the divergent dynamics of its regulon, we compared wild-type TRIPs to those of a *gbaA* transposon mutant, in which GbaA-mediated repression should be relieved. Since transposon insertion happens within the GbaA-R operon, transcription of this locus was disrupted. However, in the GbaA-L operon we observed a more than 100-fold increase in expression (Extended Data Fig. 9h) due to loss of repression. As predicted, this loss of repression was accompanied by a clear reversion of GbaA-L TRIPs to the expected pattern in the transposon mutant, as well as reduced expression amplitude (Fig. 4e). To further verify that relief of GbaA repression at the promoter was directly responsible for this change, we measured transcription of a reporter gene from the GbaA-L promoter at an alternative chromosomal locus. Although repression by GbaA was less efficient at this locus than for native GbaA-L (Extended Data Fig. 9i), we nonetheless observed a spike in reporter expression on a wild-type JE2 background that was GbaA-dependent (Extended Data Fig. 9j). These observations suggest that repression drives the high-amplitude pulses in expression seen for low-expressed genes.

By comparing TRIPs across the genome, we identify several archetypal classes (Fig. 4f). Class 1 TRIPs reflect the canonical dosage-driven response. For genes outside this category, we observe divergence of TRIPs along two main axes: heterochrony, or differential expression phase (that is, timing of expression changes), and heterometry, or differential amplitude (or peak/trough ratio). Many repressed operons exhibit heterometry (class 2a and 2b), whereas a subset of these peak earlier than expected (heterochrony) (class 2a). Genes can also exhibit heterochrony as a delayed expression profile (class 3). Each class of TRIP may therefore reveal distinct features of local gene regulatory contexts.

Biophysical modelling of TRIPs

The presence of shared classes among TRIPs may suggest recurrent mechanistic drivers. To explore these hypothetical mechanisms in a quantitative manner, we interpreted the expression patterns using several biophysical models for cell cycle-dependent transcription. We first leveraged the procedure developed above of aligning scRNA-seq and smFISH data (Fig. 2i) to convert our *E. coli* scRNA-seq measurements (fraction mRNA as a function of θ_c) to the estimated absolute mRNA copy number as a function of time since the last cell division (Extended Data Fig. 10a). This produced traces analogous to those measured by smFISH (Fig. 2i and Extended Data Fig. 6a,c). We next fitted those traces to a biophysical model[11] in which mRNA production rate is proportional to gene dosage. In this model, gene replication at time t_r results in the doubling of mRNA level over a time inversely proportional to the mRNA degradation rate k_d (Fig. 5). We found that 39% of reproducible genes were well fitted by this model (Extended Data Fig. 10b,c,i), with good consistency between replicates (Extended Data Fig. 10b). Moreover, the fitted values of k_d were significantly correlated with published values (Extended Data Fig. 10d), and the inferred replication times, t_r, mirrored the expectations based on each gene's chromosomal location (Extended Data Fig. 10e). Thus, a large proportion of genes evaluated show cell cycle-dependent expression that is consistent with a simple biophysical model of dosage-dependent transcription.

To capture the dynamics of the TRIPs of genes with more complex expression patterns, we expanded our modelling approach. TRIP classes 2a and 2b exhibit 'pulses' of expression upon replication (Fig. 4f). We thus developed an 'activation model' in which mRNA production rate transiently increases upon replication before mRNA abundance decays back to a level driven by gene dosage (Fig. 5). Genes in high-amplitude expression clusters Ec17 (Fig. 4c; for example, *nemA* and *iclR*) and Ec9 (Fig. 4d; for example, *aceB*) had a good fit ($R^2 > 0.9$) and well-predicted gene replication times, t_r (Extended Data Fig. 10h), but were poorly fitted by the null (dosage-driven) model (Extended Data Fig. 10g). Genes fitting to the activation model were also enriched for repressors ($P < 10^{-4}$, hypergeometric test as in Extended Data Fig. 9b), whereas genes fitting to the null model were not ($P = 1.00$). Moreover, expression of *lacZ* was best-described by

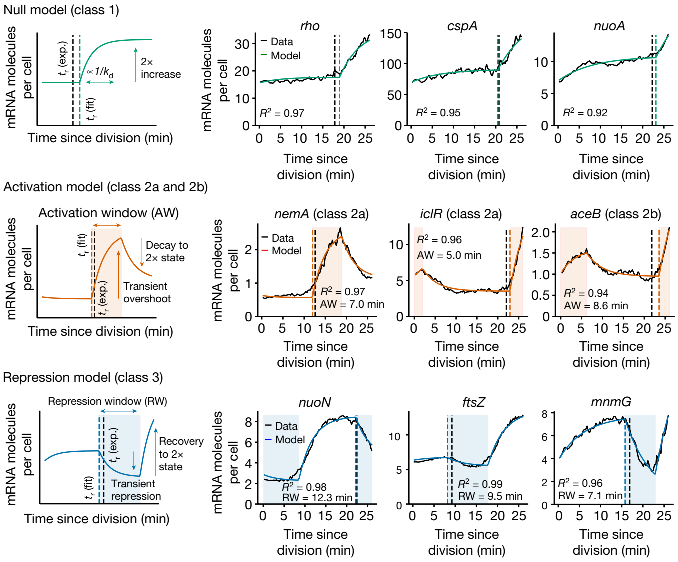

Fig. 5 | TRIP classes can be explained by simple biophysical models of expression. Fits of gene expression patterns to biophysical models of cell cycle-dependent transcription. Gene expression data from *E. coli* grown in LB were converted to absolute mRNA copy number as a function of time (Methods). Black traces show observed expression and coloured lines represent model fits. Vertical dashed lines indicate expected replication times based on predicted minimum gene expression (based on chromosome position) (t_r (exp.), black lines) or based on biophysical model fits (t_r (fit), coloured lines) (Methods). For each model, the left column is a schematic description of the fit and how it relates to model parameters (the null model parameters also apply to the expanded activation and repression models). The accompanying panels show these fits for individual genes (for two of these genes, *nemA* and *nuoN*, alternative plots of poorly fitting models are shown for comparison in Extended Data Fig. 10k).

the activation model (Extended Data Fig. 10l). Previously, Wang and colleagues[8] observed a replication-associated pulse in *lacZ* expression associated with the activity of its repressor LacI, further supporting our inference that many genes fitting to the activation model are in a repressed state. Finally, the activation model enabled us to interrogate what determines whether a gene displays early-peaking behaviour (class 2a) or not (class 2b). The IclR regulon encompasses both its own gene *iclR* and the neighbouring *aceBAK* operon. Whereas *iclR* peaks early (class 2a; Fig. 4c), *aceBAK* genes peak at the expected time (class 2b; Fig. 4d). In this case, model fits suggested that this difference is driven by a longer activation window for *aceBAK*, implying that it takes longer for IclR-dependent repression to be restored at *aceBAK* than at *iclR* (Extended Data Fig. 10n). Therefore, analysis of model fits can suggest useful hypotheses regarding specific regulatory circuits.

However, neither of these models can accurately describe genes with delayed expression timing ('class 3' TRIPs). Specifically, each fails to predict the replication timing both in the delayed cluster Ec5 (Fig. 4b and Extended Data Fig. 10h) and for genes far from their promoter (Extended Data Fig. 10j). Since many of these genes show a drop in transcript abundance upon replication (Fig. 3d), we introduced a repression model that features a transient window of reduced mRNA production rate upon gene replication. This model effectively captured genes distant from their TSS, including for those genes at the far end of *nuo* and *mraZ*–*ftsZ* operons (Fig. 5 and Extended Data Fig. 10j), whereas the promoter-proximal *nuoA* was sufficiently described by the null model (Fig. 5). Other genes with a delayed (class 3) TRIP that were well described by the repression model included genes immediately adjacent to the *oriC* locus (*mioC* and *mnmG*; Fig. 5 and Extended Data Fig. 10m), as well as *dnaA* (Extended Data Fig. 10m), all of which have previously been suggested to experience transient repression around the time of their replication[45,46]. Of the genes not captured by the null model, the repression model explained fewer genes than the activation model, and most of the former genes' repressed dynamics can be linked to their position within operons (Extended Data Fig. 10i). Thus, replication-dependent repression appears to be a relatively unusual phenomenon. Overall, the simple biophysical modelling provides gene-level estimation of testable gene regulatory parameters, including mRNA decay, transcription activation and transcription repression.

Discussion

Here we reveal the cell cycle transcriptional dynamics of rapidly dividing bacteria. Our work differs substantially not only from scRNA-seq analysis of cell cycle phase-specific genes in eukaryotic cells[47], but also from previous studies of cell cycle transcriptomes in bulk, synchronized bacterial populations. It was previously suggested that, at least for α-proteobacteria[12,15], transcription of cell cycle genes was temporally regulated according to function. Our observations, however, suggest that the more general situation in prokaryotes, at least under rapid growth conditions, is one in which cytoplasmic content is relatively invariant throughout the cell cycle[48], and rather that the major perturbation to gene expression is the local effect of chromosomal replication itself. For example, cell cycle fluctuation in mRNA of *ftsZ*, the major cell division regulator, has been described previously[49,50]. A direct link between *ftsZ* replication and transcriptional inhibition has been postulated[50], but the authors could not provide a satisfactory mechanistic explanation. Here, we provide a simple explanation of these augmented fluctuations in *ftsZ* abundance as the natural consequence of replication of a gene transcribed from a distant promoter (Fig. 3d), and not due to a cell division-coupled signalling event. Although global factors such as competition for RNA polymerase between genes[51] may still have a role, our work points to a central role of gene replication in driving cell cycle transcriptional dynamics.

We introduce the notion of the TRIP as a gene-level summary of the response of each gene to the perturbation of its own replication. This profile is analogous to the electrocardiogram, a time-resolved electrical pattern whose sophisticated, quantitative interpretation yields a wealth of information about cardiac function. Similarly, by continuing to refine the capture and analysis of TRIPs, we expect to gain an ever more detailed diagnostic picture of gene regulation. Presently, we can distinguish a number of broad classes of TRIPs (Fig. 4f). For many genes, expression changes upon replication are sensitive primarily to copy number increases, and their dynamics can be described simply by their mRNA production and decay rates, as well as their replication timing. Other classes exhibit heterometry (here, primarily high amplitude) and heterochrony (early or delayed expression) that reveal important features of their regulatory environment, from repression state to promoter usage. Our modelling analysis, however, could be extended to other regulatory motifs. For example, many operons with genes displaying particularly strong 'pulses' of expression upon replication are under autorepression (Supplementary Table 4), a network motif that facilitates rapid peaking of expression[52]. Biophysical models that can account for these specific motifs will enable us to both test and generate increasingly specific hypotheses about the regulatory context of a gene. Although a gene's fit to a specific biophysical model does not automatically entail a specific regulatory mechanism, it places constraints on the plausible molecular processes leading to that TRIP. Crucially, the ability to reveal regulatory motifs without genetic perturbations, as well as to infer relevant global parameters such as DNAP and RNAP speeds, will enable expansion of our approach across non-model organisms or strains in which the regulatory network is poorly characterized. We expect that the rapid improvement in scale and capture efficiency of bacterial scRNA-seq methodologies[4,53–55] will enable us to perform ever deeper and more quantitative analyses.

Single-cell analysis of eukaryotic cells has led to the development of a suite of analysis tools based principally on clustering and interpreting cell populations from specific marker genes[56]. Whereas these approaches may be applicable to prokaryotes in certain circumstances, a hallmark of bacterial systems biology has been the use of quantitative, model-driven analysis of gene regulation and physiology[2,57]. This arises from both the simplicity of bacteria and the reproducibility—with careful experimental design—of bacterial measurements. Physically inspired modelling has been applied to eukaryotic scRNA-seq, particularly in the context of RNA velocity[58], but in practice these approaches are typically used to infer cellular behaviours such as developmental processes[59]. Our work demonstrates that bacterial scRNA-seq can produce quantitative estimates of molecular-level dynamics on a genome-wide scale and in a wide range of organisms. Our cell cycle analysis and TRIP frameworks, however, are probably only the first examples of the type of quantitative analyses that are now enabled by single-cell technologies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-023-06974-w.

1. Bervoets, I. & Charlier, D. Diversity, versatility and complexity of bacterial gene regulation mechanisms: opportunities and drawbacks for applications in synthetic biology. *FEMS Microbiol. Rev.* **43**, 304–339 (2019).
2. Vilar, J. M. G., Guet, C. C. & Leibler, S. Modeling network dynamics: the lac operon, a case study. *J. Cell Biol.* **161**, 471–476 (2003).
3. Narula, J., Devi, S. N., Fujita, M. & Igoshin, O. A. Ultrasensitivity of the *Bacillus subtilis* sporulation decision. *Proc. Natl Acad. Sci. USA* **109**, E3513–E3522 (2012).
4. Homberger, C., Hayward, R. J., Barquist, L. & Vogel, J. Improved bacterial single-cell RNA-seq through automated MATQ-seq and Cas9-based removal of rRNA reads. *mBio* **14**, e0355722 (2023).
5. Balakrishnan, R. et al. Principles of gene regulation quantitatively connect DNA to RNA and proteins in bacteria. *Science* **378**, eabk2066 (2022).
6. Cooper, S. & Helmstetter, C. E. Chromosome replication and the division cycle of *Escherichia coli* B/r. *J. Mol. Biol.* **31**, 519–540 (1968).
7. Schaechter, M., Bentzon, M. W. & Maaloe, O. Synthesis of deoxyribonucleic acid during the division cycle of bacteria. *Nature* **183**, 1207–1208 (1959).
8. Wang, M., Zhang, J., Xu, H. & Golding, I. Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality. *Nat. Microbiol.* **4**, 2118–2127 (2019).
9. Narula, J. et al. Chromosomal arrangement of phosphorelay genes couples sporulation and DNA replication. *Cell* **162**, 328–337 (2015).
10. Slager, J. & Veening, J.-W. Hard-wired control of bacterial processes by chromosomal gene location. *Trends Microbiol.* **24**, 788–800 (2016).
11. Peterson, J. R., Cole, J. A., Fei, J., Ha, T. & Luthey-Schulten, Z. A. Effects of DNA replication on mRNA noise. *Proc. Natl Acad. Sci. USA* **112**, 15886–15891 (2015).
12. Laub, M. T., McAdams, H. H., Feldblyum, T., Fraser, C. M. & Shapiro, L. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* **290**, 2144–2148 (2000).
13. Fang, G. et al. Transcriptomic and phylogenetic analysis of a bacterial cell cycle reveals strong associations between gene co-expression and evolution. *BMC Genom.* **14**, 450 (2013).
14. Zhou, B. et al. The global regulatory architecture of transcription during the Caulobacter cell cycle. *PLoS Genet.* **11**, e1004831 (2015).
15. De Nisco, N. J., Abo, R. P., Wu, C. M., Penterman, J. & Walker, G. C. Global analysis of cell cycle gene expression of the legume symbiont *Sinorhizobium meliloti*. *Proc. Natl Acad. Sci. USA* **111**, 3217–3224 (2014).
16. Bandekar, A. C., Subedi, S., Ioerger, T. R. & Sassetti, C. M. Cell-cycle-associated expression patterns predict gene function in Mycobacteria. *Curr. Biol.* **30**, 3961–3971.e6 (2020).
17. Cooper, S. The synchronization manifesto: a critique of whole-culture synchronization. *FEBS J.* **286**, 4650–4656 (2019).
18. Blattman, S. B., Jiang, W., Oikonomou, P. & Tavazoie, S. Prokaryotic single-cell RNA sequencing by in situ combinatorial indexing. *Nat. Microbiol.* **5**, 1192–1201 (2020).
19. Kuchina, A. et al. Microbial single-cell RNA sequencing by split-pool barcoding. *Science* **371**, eaba5257 (2021).
20. Imdahl, F., Vafadarnejad, E., Homberger, C., Saliba, A.-E. & Vogel, J. Single-cell RNA-sequencing reports growth-condition-specific global transcriptomes of individual bacteria. *Nat. Microbiol.* **5**, 1202–1206 (2020).
21. Homberger, C., Barquist, L. & Vogel, J. Ushering in a new era of single-cell transcriptomics in bacteria. *microLife* **3**, uqac020 (2022).
22. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
23. Bremer, H. & Dennis, P. P. Modulation of chemical composition and other parameters of the cell at different exponential growth rates. *EcoSal Plus* **3**, https://doi.org/10.1128/ecosal.5.2.3 (2008).
24. Michelsen, O., Teixeira de Mattos, M. J., Jensen, P. R. & Hansen, F. G. Precise determinations of C and D periods by flow cytometry in *Escherichia coli* K-12 and B/r. *Microbiology* **149**, 1001–1010 (2003).
25. Wang, X., Lesterlin, C., Reyes-Lamothe, R., Ball, G. & Sherratt, D. J. Replication and segregation of an *Escherichia coli* chromosome with two replication origins. *Proc. Natl Acad. Sci. USA* **108**, E243–E250 (2011).
26. Dimude, J. U. et al. Origins left, right, and centre: increasing the number of initiation sites in the chromosome. *Genes* **9**, 376 (2018).
27. Ivanova, D. et al. 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).
28. Khodursky, A. B. et al. Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. *Proc. Natl Acad. Sci. USA* **97**, 9419–9424 (2000).
29. Pham, T. M. et al. A single-molecule approach to DNA replication in *Escherichia coli* cells demonstrated that DNA polymerase III is a major determinant of fork speed. *Mol. Microbiol.* **90**, 584–596 (2013).
30. Kjeldgaard, N. O., Maaloe, O. & Schaechter, M. The transition between different physiological states during balanced growth of *Salmonella typhimurium*. *J. Gen. Microbiol.* **19**, 607–616 (1958).
31. Skinner, S. O., Sepúlveda, L. A., Xu, H. & Golding, I. Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat. Protoc.* **8**, 1100–1113 (2013).
32. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504–508 (2010).
33. Pomerantz, R. T. & O'Donnell, M. The replisome uses mRNA as a primer after colliding with RNA polymerase. *Nature* **456**, 762–766 (2008).
34. de la Fuente, A., Palacios, P. & Vicente, M. Transcription of the *Escherichia coli* dcw cluster: evidence for distal upstream transcripts being involved in the expression of the downstream *ftsZ* gene. *Biochimie* **83**, 109–115 (2001).
35. Flärdh, K., Palacios, P. & Vicente, M. Cell division genes *ftsQAZ* in *Escherichia coli* require distant *cis*-acting signals upstream of *ddlB* for full expression. *Mol. Microbiol.* **30**, 305–315 (1998).
36. Lutkenhaus, J. F., Wolf-Watz, H. & Donachie, W. D. Organization of genes in the *ftsA–envA* region of the *Escherichia coli* genetic map and identification of a new fts locus (*ftsZ*). *J. Bacteriol.* **142**, 615–620 (1980).
37. Zaslaver, A., Mayo, A., Ronen, M. & Alon, U. Optimal gene partition into operons correlates with gene functional order. *Phys. Biol.* **3**, 183–189 (2006).
38. Zhu, M., Mu, H., Han, F., Wang, Q. & Dai, X. Quantitative analysis of asynchronous transcription–translation and transcription processivity in under various growth conditions. *iScience* **24**, 103333 (2021).
39. Zhu, M., Mori, M., Hwa, T. & Dai, X. Disruption of transcription–translation coordination in *Escherichia coli* leads to premature transcriptional termination. *Nat. Microbiol.* **4**, 2347–2356 (2019).
40. Sharpe, M. E., Hauser, P. M., Sharpe, R. G. & Errington, J. *Bacillus subtilis* cell cycle as studied by fluorescence microscopy: constancy of cell length at initiation of DNA replication and evidence for active nucleoid partitioning. *J. Bacteriol.* **180**, 547–555 (1998).
41. Golding, I. Revisiting replication-induced transcription in *Escherichia coli*. *Bioessays* **42**, e1900193 (2020).
42. Guptasarma, P. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *Bioessays* **17**, 987–997 (1995).
43. Ray, A., Edmonds, K. A., Palmer, L. D., Skaar, E. P. & Giedroc, D. P. Glucose-induced biofilm accessory protein A (GbaA) is a monothiol-dependent electrophile sensor. *Biochemistry* **59**, 2882–2895 (2020).
44. Van Loi, V. et al. 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).
45. Campbell, J. L. & Kleckner, N. E. coli *oriC* and the dnaA gene promoter are sequestered from dam methyltransferase following the passage of the chromosomal replication fork. *Cell* **62**, 967–979 (1990).
46. Theisen, P. W., Grimwade, J. E., Leonard, A. C., Bogan, J. A. & Helmstetter, C. E. Correlation of gene transcription with the time of initiation of chromosome replication in *Escherichia coli*. *Mol. Microbiol.* **10**, 575–584 (1993).
47. Buettner, F. et al. Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nat. Biotechnol.* **33**, 155–160 (2015).
48. Cooper, S. The *Escherichia coli* cell cycle. *Res. Microbiol.* **141**, 17–29 (1990).
49. Garrido, T., Sánchez, M., Palacios, P., Aldea, M. & Vicente, M. Transcription of *ftsZ* oscillates during the cell cycle of *Escherichia coli*. *EMBO J.* **12**, 3957–3965 (1993).
50. Zhou, P. & Helmstetter, C. E. Relationship between *ftsZ* gene expression and chromosome replication in *Escherichia coli*. *J. Bacteriol.* **176**, 6100–6106 (1994).
51. Lin, J. & Amir, A. Homeostasis of protein and mRNA concentrations in growing cells. *Nat. Commun.* **9**, 4496 (2018).
52. Rosenfeld, N., Elowitz, M. B. & Alon, U. Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323**, 785–793 (2002).
53. Ma, P. et al. Bacterial droplet-based single-cell RNA-seq reveals antibiotic-associated heterogeneous cellular states. *Cell* **186**, 877–891.e14 (2023).
54. Brennan, M. A. & Rosenthal, A. Z. Single-Cell RNA sequencing elucidates the structure and organization of microbial communities. *Front. Microbiol.* **12**, 713128 (2021).

55. Xu, Z. et al. Droplet-based high-throughput single microbe RNA sequencing by smRandom-seq. *Nat. Commun.* **14**, 5130 (2023).
56. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
57. Jun, S., Si, F., Pugatch, R. & Scott, M. Fundamental principles in bacterial physiology-history, recent progress, and the future with focus on cell size control: a review. *Rep. Prog. Phys.* **81**, 056601 (2018).
58. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
59. Bergen, V., Soldatov, R. A., Kharchenko, P. V. & Theis, F. J. RNA velocity-current challenges and future perspectives. *Mol. Syst. Biol.* **17**, e10282 (2021).

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

Methods

Bacterial strains and media

Strains used are listed in Supplementary Table 1. All *E. coli* strains (a gift from C. Rudolph) and *B. subtilis* (ATCC) 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)[26]. For growth in minimal media, an M9 base (1× M9 minimal salts (Gibco), 2 mM $MgSO_4$, 0.2 mM $CaCl_2$) was supplemented with 0.4% glucose (M9 + glucose: M9G) or with both 0.4% glucose and 0.2% acid casein peptone (Acros Organics) (M9 + glucose + amino acids: M9GA). All *S. aureus* strains were routinely grown in Bacto tryptic soy broth (TSB) (BD Biosciences). The *gbaA* transposon mutant was provided by the Biodefence and Emerging Infections (BEI) Resources Repository (NR-46898).

Growth curves and cell collection for PETRI-seq

All growth experiments were performed at 37° with shaking at 225 rpm.

Constant growth conditions. Strains were grown overnight in LB (*E. coli* & *B. subtilis*) or TSB (*S. aureus*). 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 3 S spectrophotometer (Thermo Scientific). For experiments with *S. aureus* in balanced growth (Datasets D5-D8), overnight cultures were diluted in TSB first to 0.005, then after 3 h were diluted again to 0.005 before measuring A_{600} at the time intervals specified. Growth of *B. subtilis* was the same except with LB used as the growth medium. For *E. coli* growth curves, strains were diluted to an A_{600} value of 0.05 and incubated for 2 h 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, m is the gradient of the relationship between time and $\log_2(A_{600})$, and c is the y-intercept) using the LINEST function in Microsoft Excel and the doubling time in minutes t_d was calculated as $1/m$. All doubling times are provided in Supplementary Table 1.

For harvesting for PETRI-seq, cells were grown as described except that after specific time intervals (for *S. aureus*, 2 h 20 min in initial experiments, 1 h 30 min in balanced growth experiments; for *E. coli*, 2 h, 3 h and 7 h in LB, M9GA, and M9G, respectively, when growth rates appeared constant (Extended Data Fig. 1b); for *B. subtilis*, 1 h 30 min) cells were harvested by centrifugation and resuspension in 4% formaldehyde in PBS. For *S. aureus* initial experiments, centrifugation was at 10,000*g*, 1 min at room temperature and for *E. coli*, *B. subtilis*, and balanced growth *S. aureus* experiments, centrifugation was at 3,220*g*, 5 min, 4 °C. For *B. subtilis*, because sensitivity in the transcriptome to cold shock was previously noted upon gradual cooling during centrifugation at 4 °C (ref. 19), cells were initially cooled to <10 °C by rapid agitation in a dry ice ethanol bath followed by retention on ice to prevent further transcriptional changes after harvesting.

For growth under perturbed conditions, see Supplementary Information, 'Growth perturbation conditions'.

PETRI-seq

PETRI-seq was carried out as described previously[18] with modifications as described in the Supplementary Information, 'PETRI-seq modifications'.

Analysis of PETRI-seq data

Pre-processing of scRNA-seq data. Initial demultiplexing of barcodes, alignment, and feature quantification was performed using the analysis pipeline described[18], 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), *S. aureus* to the USA300_FPR3757 reference assembly (GCF_000013465.1), and *B. subtilis* to the 168 reference assembly (GCF_000009045.1). After initial processing, counts by cell barcode were pooled across different libraries (no batch effects were noted between libraries) and initial filtering was performed using Scanpy v1.7.1 (ref. 60). Barcodes with unique molecular identifiers (UMIs) below a threshold (10 for dataset D9 rep 2, 15 for dataset D1, D2, D4, and D10 (all samples but M9GA); 20 for dataset D3, D5–7, D9 rep 1, and D10 M9GA, 40 for dataset D8) were removed, as well as any genes with fewer than 50 UMIs across all included barcodes (100 for dataset D3 & D9). Note that within the main text, we refer to UMIs as simply 'transcripts' for the sake of clarity, although we acknowledge that with random priming more than one unique read could originate from a single mRNA molecule.

Data denoising and generation of gene–gene correlations. To generate the denoised representation of the data, scVI v0.9.0 (ref. 22) 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). The model was trained with default parameters. Denoised expression values based on the scVI model were obtained using the scVI function get_normalized_expression. Initial gene–gene correlations without binning (Extended Data Fig. 1c,f) were calculated from scVI-normalized counts using the get_feature_correlation_matrix function. For correlations with position-dependent binning, scVI-normalized expression matrices were first $\log_2$-transformed and then converted to z-scores by mean centering followed by division by the standard deviation. For Extended Data Fig. 1d, this was carried out using raw counts normalized by total UMI per cell, and transformation by $\log_2(x + 1)$ (to allow for zero values). After removing extrachromosomal genes, expression z-scores were averaged within bins (50 kb unless otherwise stated) and Spearman correlations between bins were calculated. For *B. subtilis* analysis, uniform manifold approximation and projection (UMAP) of initial scVI-smoothed data revealed two clusters in initial analysis, with one cluster arising due to PBSX mobilization (Extended Data Fig. 8c). To generate the final scVI models, we removed this cluster and repeated scVI on the remaining cells. Note that for PBSX annotation and for annotation in *S. aureus* MGEs (Extended Data Fig. 9f), the online tool Phaster[61] (phaster.ca) was used. For further discussion of evidence supporting our analysis of global gene correlations, see Supplementary Information, 'Quality assessment of global correlations in gene expression'.

Cell cycle analysis. Our quantitative framework describing gene expression as a function of replication cycle state (Fig. 2) is parameterized as the relationship between gene expression and two further parameters, cell angle θ_c (describing the position of cells within the cell cycle) and gene angle θ_g (describing the ordering of expression of genes within the cell cycle).

Derivation of cell angle. The position of cells within the cell cycle was determined as follows. First, scVI-derived z-scores (see 'Data denoising and generation of gene–gene correlations') were averaged into bins according to chromosomal location (50–400 kb bins, depending on the dataset). Binned data were then projected into two dimensions by UMAP analysis using the umap-learn v0.5.1 library in Python (https://umap-learn.readthedocs.io/en/latest/) with the 'correlation' distance metric, generating the wheel plots (Fig. 2a and Extended Data Fig. 2b). To assign cells to a particular replication cycle phase based on their position, the embeddings were first mean-centred and then the angle of each cell θ_c relative to the origin between x and y coordinates in a two-dimensional UMAP embedding was calculated as $\tan^{-1}(x/y)$, similar to the ZAVIT method described previously[62,63]. To obtain the expression × cell angle matrix used in Fig. 2b,c, scVI-denoised gene expression *z*-scores were then averaged within 100 equally spaced bins of θ_c to produce a cell angle-binned expression matrix. For UMAP without averaging, see Extended Data Fig. 2a.

Derivation of gene angle. After averaging gene expression within 100 equally spaced bins of θ_c (as in Fig. 2b,c), principal component analysis (PCA) was performed on the transpose of this matrix to generate a low-dimensional projection of genes based on their cell cycle expression (Fig. 2d). Analogous to the derivation of θ_c, gene angle θ_g was calculated as the angle between PCs 1 and 2 relative to the origin. As discussed, this parameter roughly relates to genes' order of expression within the cell cycle and indeed recapitulates the order in which genes are replicated (Fig. 2e,f). We chose to use PCA instead of UMAP to derive θ_g because while UMAP produces a 'wheel' similar to the cell-level analysis (Extended Data Fig. 2c), we reasoned that as a linear dimensionality reduction PCA would be more likely to give a consistent gene angle–origin distance relationship. However PCA-derived θ_g values still broadly capture the ordering when UMAP is performed (Extended Data Fig. 2c).

Predicting expression dynamics based on DNA replication alone. We developed two regression models to infer a gene's predicted gene angle $\theta_{\text{g-pred}}$ from its distance from the origin of replication (Supplementary Information, 'Modelling the gene angle–origin distance relationship') and then to predict cell cycle expression from this $\theta_{\text{g-pred}}$ value (Supplementary Information, 'Modelling the cell angle–gene angle relationship'). These models were combined to yield the pipeline in Extended Data Fig. 2i. Firstly, the gene angle–origin distance model (Supplementary Information, 'Modelling the gene angle–origin distance relationship') was used to predict the expected value $\theta_{\text{g-pred}}$ from origin distance *D*. Next, cell cycle expression was predicted using the cell angle–gene angle regression model (Supplementary Information, 'Modelling the cell angle–gene angle relationship') using $\theta_{\text{g-pred}}$ 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 × 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 (Extended Data Fig. 4a,f), as well as the loss of global chromosome position-dependent gene–gene correlations upon correction of scVI expression with replication-predicted expression (Extended Data Fig. 4b,g). Additionally, we used this modelling approach to set the zero angle for gene expression plots.

Setting the position of $\theta_c = 0$. Initially, the cell angle θ_c orders cells by their cell cycle position within a circle but the start point, when $\theta_c = 0$, is arbitrary. This is not only challenging to interpret but impedes comparing across replicates. Therefore, we standardized θ_c so that $\theta_c = 0$ was the predicted point of replication initiation. Using the inference approach described above, we predicted the gene expression profile by θ_c for an imaginary gene at $D = 0$ (that is, 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 approximately 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 ($\theta_c - \theta_0$) mod 360 to set this point as 0°. This interpretation is roughly in accordance with the estimated timing of replication initiation as determined directly from smFISH data (Extended Data Fig. 5f and Supplementary Information, 'Inferring cell cycle phase from the DAPI signal'). Crucially, however, it also provides a point of standardization that allows in-phase comparison of cell cycle expression profiles across independent replicates.

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. For details, see Supplementary Information, 'Identifying genes with high cell cycle variance' and 'Identifying genes with high divergence from predicted expression'.

Plotting expected and observed cell cycle expression patterns. To visualize the degree of divergence from predicted expression (for example, Figs. 2h and 3b), we take scVI-derived *z*-scores averaged in 100 bins by θ_c along with model predictions (see 'Predicting expression dynamics based on DNA replication alone'). We then adjust θ_c as described above (in 'Setting the position of $\theta_c = 0$'). Importantly, we are able to validate our analysis from the raw (non-scVI-derived) data by averaging normalized counts θ_c (Extended Data Fig. 4j). This demonstrates that cell cycle expression patterns are not an artefact of scVI denoising.

Analysing the effect of operon gene position on expression dynamics. We identified the excess of genes with a delayed expression profile by calculating the angle difference as $\tan^{-1}(\sin(\theta_g - \theta_{\text{g-pred}})/\cos(\theta_g - \theta_{\text{g-pred}}))$ where θ_g and $\theta_{\text{g-pred}}$ are the observed and predicted gene angles in radians, respectively. For operon annotations, *E. coli* and *B. subtilis* transcription units from Biocyc[64,65] (https://biocyc.org/) were used. To investigate the relationship between gene distance from TSSs and angle difference in *E. coli*, all genes in polycistrons (transcription units with more than one gene) were included. The distance was measured from the annotated transcription unit start site to the midpoint of each gene. Where genes were in multiple transcription units, the longest distance from a start site was taken. Angle difference was converted into time by dividing the angle by 360° then multiplying by the doubling time in seconds. For *S. aureus*, operon annotation was obtained from AureoWiki[66] (https://aureowiki.med.uni-greifswald.de/). Since this provided only the genes within an operon and not its start, the first base of the first gene was taken as the TSS.

Analysis of operon expression trends relative to timing of operon replication. For each operon shown (Fig. 3d), the predicted replication timing in degrees of θ_c was calculated as the predicted minimum in expression for the first gene in that operon (similar to calculation of TRIPs, 'Defining gene TRIPs'). The cell angle, θ_c, was then redefined such that $\theta_c = 0$ is the point of operon replication (denoted $\theta_{\text{c-}\mathit{nuoA}}$, $\theta_{\text{c-}\mathit{atpI}}$ and $\theta_{\text{c-}\mathit{mraZ}}$ for each operon). With an expected DNA polymerase speed[28,29] of around 800 bp s^{-1}, replication of the whole operon is expected to take less than 20 s, so it is assumed that genes within the operon are replicated simultaneously. Next, scVI-derived normalized expression of the *nuo*, *atp* and *mraZ*–*ftsZ* operons, averaged in 100 bins by θ_c, was converted to fold change relative to the replication point by dividing cell cycle expression by when $\theta_{\text{c-}\mathit{nuoA}}$, $\theta_{\text{c-}\mathit{atpI}}$ or $\theta_{\text{c-}\mathit{mraZ}}$, respectively, were equal to zero degrees. As shown in Fig. 3d, this reveals that genes far from, but not close to, the TSS display transient decreases in expression.

For analysis of correlations between a gene's position within its operon and its expression timing or amplitude, Spearman correlations were calculated using the spearmanr function in the Python package scipy v1.9.3 (ref. 67). This function was also used to calculate *P* values based on a two-sided test with the null hypothesis of no correlation. See documentation in https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.spearmanr.html for details.

Defining gene TRIPs. Our work uncovered evidence that cell cycle-dependent fluctuations in each gene's expression can arise from the diverse responses of that gene to perturbation by replication. Therefore, we define the TRIP as a gene's expression profile relative to its predicted timing of replication. We did this for all genes that showed good reproducibility between replicates (Spearman correlation > 0.7 between expression averaged in 100 bins by θ_c). To produce each gene's TRIP, we first take scVI-normalized expression and average in 100 positions by θ_c. To preserve information on the dynamics of each gene (since amplitude in cell cycle fluctuations is an important parameter in interpreting TRIPs), we do not log-transform or scale expression but instead divide by the mean expression across the cell cycle (to allow comparison of genes with different baseline expression levels). For each gene, we then determine the predicted timing of replication as the minimum in the predicted expression (see 'Predicting expression dynamics based on DNA replication alone'). We then rotate θ_c for that gene so that $\theta_c = 0$ corresponds to this predicted replication timing. We denote this as $\theta_{c\text{-}rep}$, calculated by the transformation $(\theta_c - \theta_{c\text{-}min})$ mod 360 where $\theta_{c\text{-}min}$ is the predicted expression minimum. Therefore, the TRIP preserves replication dynamics while allowing standardization across different replication timings and baseline expression values.

Clustering of TRIPs. Reproducible genes were clustered based on the TRIPs derived as above using *k*-means clustering. TRIPs for genes within each cluster were then averaged to give the cluster profiles in Fig. 4 and Extended Data Fig. 9. To analyse for enrichment of repressed genes in *E. coli*, repressor annotations were obtained from Biocyc[64,65] (https://biocyc.org/). We then assessed whether genes within each cluster were enriched for genes that had an annotated repressor, using the hypergeometric test to assess significance. Since we noticed that a handful of regulators had a much higher number of target genes than others (Extended Data Fig. 9b, right), we decided to exclude these 'global repressors' (20 or more repressive targets annotated), which decreased the background fraction of reproducible genes with an annotated repressor from 35% to 18%, thus improving the sensitivity and focusing the analysis on more specific repressor-target interactions. After performing the analysis on each cluster, we then adjusted for multiple comparisons using the Benjamini–Hochberg procedure, choosing a FDR of 0.1. Note that while most of the clusters showing enrichment were those early-peaking clusters with or without high amplitude, clusters Ec5 and Ec12 were also significant. Ec5 is dominated by *nuo* genes and *dnaA*, and Ec12 is the *nrdAB–yfaE* operon, all of which have annotated repressors.

Simulating the effect of DNA replication on gene expression

We predicted the gene–gene correlation patterns arising from DNA replication using a simulation written in Python (see Extended Data Fig. 1h–k) 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 (that is, 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 (that is, 1.5 times 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[6]). 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 $\mu = (n \times t_d)/t_c$ 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 ref. 24). 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: (1) all origins initiate replication simultaneously; (2) replication stops at the termination site *ter*, which is halfway along the chromosome; (3) 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[13] (Extended Data Fig. 1l), 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 (ref. 68), 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.

smFISH

For a full description of smFISH experimental procedures and analysis, see Supplementary Information, 'smFISH experiments and analysis'.

Biophysical modelling of TRIPs

For biophysical modelling of scRNA-seq data, expression profiles were first converted into inferred copy number as a function of time and then specific models were fitted. See Supplementary Information, 'Transformation of scRNA-seq data for biophysical modelling' and 'Biophysical modelling' for detailed descriptions.

Generation of chromosome-integrated reporter constructs in *S. aureus*

For generation of the reporter construct, we modified the pJC1111 vector[69], which integrates at the SaPI1 chromosomal attachment (att_C) site. The vector was linearized with restriction enzymes SphI and XbaI (New England Biolabs) and insertion fragments were amplified using Q5 polymerase (New England Biolabs). For the GbaA-L promoter, the intergenic region of the GbaA regulon (130 bp upstream of the *SAUSA300_RS13955* start codon) amplified from USA300 LAC genomic DNA using primers 5′-CCGTATTACCGCCTTTGAGTGAGCTGGCGGCCGCTGCATGGATTACACCTACTTAAAATTCTCTAAAATTGACAAACGG-3′ and 5′-AGTTCTTCTCCTTTGCTCATTATCAACACTCTTTTCTTTTATGATATTTAATAGTTATTGCAAATTCA-3′. *S. aureus* codon-optimized sGFP was amplified from the genomic DNA of *S. aureus* USA300 LAC previously transformed with the pOS1 plasmid (VJT67.63 (ref. 70)) using primers 5′-AAAAGAAAAGAGTGTTGATAATGAGCAAAGGAGAAGAACTTTTCACTG-3′ and 5′-ATAGGCGCGCCTGAATTCGAGCTCGGTACCCGGGGATCCTTTAGTGGTGGTGGTGGTGGTGGG-3′. Fragments were assembled using the NEBuilder HiFi assembly kit (New England Biolabs) and transformed into competent *E. coli* DH5α (New England Biolabs). The plasmid was purified and then electroporated into RN9011 (RN4220 with pRN7023, a CmR shuttle vector containing SaPI1 integrase), and positive chromosomal integrants were selected with 0.1 mM $CdCl_2$. Finally, this strain was lysed using bacteriophage 80α and the lysate was used to transduce JE2 and JE2 *gbaA*$^-$ strains, selecting for transduction on 0.3 mM $CdCl_2$.

Reporting summary

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

Data availability

All counts matrices and raw sequencing reads used to perform the scRNA-seq analysis are available in the Gene Expression Omnibus (GEO) under the accession number GSE217715. Previously published counts matrices used for bulk analysis of *C. crescentus* expression are available in the GEO under the accession number GSE46915.

Code availability

Example code used to generate the analyses and plots is available at https://github.com/yanailab/TRIPs.

60. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
61. Arndt, D. et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, W16–W21 (2016).
62. Levin, M. et al. The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–641 (2016).
63. Zalts, H. & Yanai, I. Developmental constraints shape the evolution of the nematode mid-developmental transition. *Nat. Ecol. Evol.* **1**, 113 (2017).
64. Keseler, I. M. et al. The EcoCyc database in 2021. *Front. Microbiol.* **12**, 711077 (2021).
65. Karp, P. D. et al. The BioCyc collection of microbial genomes and metabolic pathways. *Brief. Bioinform.* **20**, 1085–1093 (2019).
66. Fuchs, S. et al. AureoWiki—the repository of the *Staphylococcus aureus* research and annotation community. *Int. J. Med. Microbiol.* **308**, 558–568 (2018).
67. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
69. Chen, J., Yoong, P., Ram, G., Torres, V. J. & Novick, R. P. Single-copy vectors for integration at the SaPI1 attachment site for *Staphylococcus aureus*. *Plasmid* **76**, 1–7 (2014).
70. Benson, M. A. et al. *Staphylococcus aureus* regulates the expression and production of the staphylococcal superantigen-like secreted proteins in a Rot-dependent manner. *Mol. Microbiol.* **81**, 659–675 (2011).

Acknowledgements The authors thank Y. Pilpel, T. Lionnet and F. Matheis for critical discussions on the project and the manuscript; S. Tavazoie, S. Blattman and W. Jiang for initial advice on implementing PETRI-seq; C. Rudolph and his laboratory for providing the *E. coli* strains; P. Eichenberger for advice on *B. subtilis* growth; and M. Wang and D. Barkley, J. Wang and other members of the Yanai and Golding laboratories for advice and suggestions. The following funding was provided by the National Institutes of Health: R21AI169350 (I.Y.), R01AI143290 (I.Y.), R01AI137336 (B.S., V.J.T. and I.Y.), R35 GM140709 (I.G.). I.G. also received funding from the Alfred P. Sloan Foundation, and K.J.C.M. was supported by the Ralph O. Simmons Undergraduate Research Scholarship.

Author contributions A.W.P. and I.Y. conceived the project. A.W.P. generated and analysed the scRNA-seq data, with contributions from P.J. A.W.P., P.J. and M.P. produced the *S. aureus* strains. T.Y. and I.G. designed the smFISH approach with T.Y. and E.H. performing the experiments and Y.G., E.H. and T.Y. performing the analysis. Y.G., K.J.C.M. and T.Y. performed the biophysical modelling analysis. I.Y., I.G., B.S. and V.J.T. contributed funding and resources to the project. The original draft was written by A.W.P. with contributions from I.Y., I.G., T.Y., E.H., Y.G. and K.J.C.M., and additional reviewing and editing were provided by P.J., B.S., V.J.T. and M.P.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-023-06974-w.

Correspondence and requests for materials should be addressed to Itai Yanai.

Peer review information *Nature* thanks Lars Barquist and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.