Measuring transcription at a single gene copy reveals hidden drivers of bacterial individuality

Mengyu Wang1,2,3,7, Jing Zhang2,7, Heng Xu4,5 and Ido Golding1,2,6*

Single-cell measurements of mRNA copy numbers inform our understanding of stochastic gene expression1−3, but these measurements coarse-grain over the individual copies of the gene, where transcription and its regulation take place stochastically4−6. Here, we combine single-molecule quantification of mRNA and gene loci to measure the transcriptional activity of an endogenous gene in individual Escherichia coli bacteria. When interpreted using a theoretical model for mRNA dynamics, the single-cell data allow us to obtain the probabilistic rates of promoter switching, transcription initiation and elongation, mRNA release and degradation. Unexpectedly, we find that gene activity can be strongly coupled to the transcriptional state of another copy of the same gene present in the cell, and to the event of gene replication during the bacterial cell cycle. These gene-copy and cell-cycle correlations demonstrate the limits of mapping whole-cell mRNA numbers to the underlying stochastic gene activity and highlight the contribution of previously hidden variables to the observed population heterogeneity.

Counting RNA in individual cells revealed the bursty nature of transcription in bacteria7 and eukaryotes8 and showed how gene expression noise is modulated by physiological parameters9,10. However, whole-cell mRNA measurements report the summed contribution from multiple (sister) copies of the same gene, whose number doubles during the cell cycle and whose activity may be coupled in unknown ways. Cellular levels also integrate over the number of an endogenous gene in individual E. coli bacteria. When interpreted using a theoretical model for mRNA dynamics, the single-cell data allow us to obtain the probabilistic rates of promoter switching, transcription initiation and elongation, mRNA release and degradation. Unexpectedly, we find that gene activity can be strongly coupled to the transcriptional state of another copy of the same gene present in the cell, and to the event of gene replication during the bacterial cell cycle. These gene-copy and cell-cycle correlations demonstrate the limits of mapping whole-cell mRNA numbers to the underlying stochastic gene activity and highlight the contribution of previously hidden variables to the observed population heterogeneity.

Here we set out to measure the transcriptional activity of an individual gene copy within a single E. coli cell. We hypothesized that active transcription can be quantified by measuring the amount of mRNA that is localized to the transcribed gene. We therefore used two-colour labelling to simultaneously mark the gene locus and the mRNA produced from the gene within the same cell. The gene locus was labelled using the fluorescent repressor operator system (FROS)11, where a tet operator array, inserted near the gene, is bound by the cognate fluorescently tagged repressor (TetR-YFP). mRNA from the endogenous gene was detected using single-molecule fluorescence in situ hybridization (smFISH)11 (Fig. 1a; Supplementary Table 1 lists the promoters and genomic loci examined in this work). We used automated image analysis to identify the fluorescent foci in each channel (Supplementary Fig. 1) and measure the copy number of gene loci and mRNA molecules12−16. The FROS system allowed the reliable counting of gene copies in both live and fixed cells and did not affect the cell growth rate or mRNA expression from the gene (Supplementary Fig. 2).

Focusing first on the lactose promoter, P_{lac}, we measured the spatial distance between each lacZ mRNA molecule and the nearest lac locus in the cell. This revealed two distinct mRNA populations, one close to the gene locus (<300 nm) and the other further away (Fig. 1b and Supplementary Fig. 3). In accordance with the hypothesis that the gene-proximal mRNA signal corresponds to actively transcribed molecules, the proximal population was almost absent if the labelled locus did not correspond to the transcribed gene (Fig. 1b). Under conditions of high expression, the signal from gene-proximal mRNA was stronger than that of mRNA further away, consistent with the simultaneous presence of multiple nascent mRNA at the gene12 (Fig. 1c). Proximal mRNA was also enriched for the gene-proximal mRNA signal corresponds to actively transcribed molecules. Applying an mRNA-to-gene distance criterion (and correcting for the probability of random colocalization; Supplementary Fig. 8) allows us to classify cellular mRNA into the nascent (actively transcribed) and mature (complete) species. We can likewise determine whether a given gene copy is currently being transcribed and measure the amount of nascent mRNA at the gene at different expression levels (Fig. 1f and Supplementary Figs. 1, 7 and 9).

We next sought to use single-cell measurements of nascent mRNA to characterize the kinetics of mRNA processes taking place at the gene: transcription initiation, elongation, decay and release. To that end, we induced P_{lac} expression by adding isopropyl-β-D-thiogalactoside (IPTG) to the growth medium14. We measured the amount of nascent, mature and total lacZ mRNA per cell, at different times after induction (Fig. 2). To interpret the observed kinetics, we formulated a mathematical model for nascent mRNA dynamics (Fig. 2a). In the model, transcription initiation is followed by mRNA synthesis (elongation) at a speed, v_{syn}, to a final length, L (refs. 15,16). Once the transcript is complete, mature mRNA is released from the gene into the cytoplasm15,16. The degradation of both nascent...
and mature mRNA is assumed to initiate at rate $k_d$ (ref. 13). mRNA degradation is limited by the competition between the degradation machinery and translating ribosomes, leading to a degradation speed that is equal to $v_d$ (ref. 13).

Induction kinetics in glucose medium (where $P_{lac}$ exhibits a large dynamic range; Fig. 1f and ref. 14) showed good agreement between the theoretical model and experimental data (Fig. 2b and Supplementary Fig. 10). As predicted, mature lacZ mRNA appears only once the first transcript is completed and released, at time $L/v_d \approx 130$ s. This is also the time at which nascent mRNA levels reach a steady state, reflecting the balance of transcript initiation and release. These discontinuous features in nascent and mature mRNA kinetics are absent from the kinetics of total cellular lacZ mRNA (Fig. 2b). Fitting the data to the theoretical model allowed us to estimate the mRNA elongation speed ($v_d = 42 \pm 2$ nt s$^{-1}$; s.e.m. from two experiments) and degradation rate ($k_d = 0.0078 \pm 0.0003$ s$^{-1}$). Both estimates were consistent with independent measurements using mRNA counting alone and with previously reported values (Supplementary Figs. 11 and 12 and refs. 7,17,19,20). Measuring the steady-state amount of nascent mRNA at different $P_{lac}$ induction levels suggested that $v_d$ i is positively correlated with the rate of initiation (Supplementary Fig. 13), consistent with earlier reports20,21. The assumptions that nascent mRNA is degraded and that degradation proceeds at speed $v_d$ are further supported by simultaneous analysis of the mRNA signals from the 5′ and 3′ regions of the gene during induction (Supplementary Fig. 14).

Whereas mRNA kinetics in glucose agreed with the theoretical expectation, this was not the case when we repeated the induction experiment in glycerol, a slow-growth medium (doubling time $g = 150$ min at 30 °C, compared to $g = 50$ min in glucose at 37 °C; Supplementary Fig. 15). As seen in Fig. 2c, the appearance of cytoplasmic mRNA, upon the completion of the first transcript, was not immediately accompanied by the stabilization of the gene-proximal (presumably, nascent) mRNA level. Instead, gene-proximal mRNA continued to accumulate, eventually reaching a steady-state level that was higher than that predicted by the model. To explain these observations, we hypothesized that, under these growth conditions, complete (fully transcribed) mRNA molecules are not all immediately

---

**Fig. 1 | Detecting active transcription at a single gene copy.**

a. Left, the lac locus is detected through the binding of TetR-YFP to an array of 140 tetO sites inserted nearby in the E. coli chromosome. Endogenous lacZ mRNA, transcribed from $P_{lac}$, is simultaneously detected using smFISH. Right, in the imaged cell, two sister lac loci are present. One locus is colocalized with a strong smFISH signal, indicating active $P_{lac}$ transcription (see subsequent panels). A number of mature (cytoplasmic) lacZ mRNA are also seen. Cells were grown at 37 °C in glucose medium supplemented with cyclic AMP (cAMP) and IPTG. Scale bar, 1 μm. b. Left, the distribution of distances between each lacZ mRNA spot and the labelled gene locus closest to it in the cell. Data are shown for the lac locus (red) and the $hyc$ locus (located opposite $lac$ on the other arm of the chromosome, grey). The distributions were used to define a distance threshold for the gene-proximal mRNA population (here, 300 nm, cyan shading and dashed line). Right, by applying the mRNA-to-distance threshold, cellular mRNA can be classified into nascent (actively transcribed) and mature; likewise, each gene copy is classified as transcriptionally active or inactive. c. The intensity of the lacZ smFISH signal (mean ± s.e.m.) as a function of distance from the lacZ locus. The data were binned and fitted to a Hill function. The vertical dashed line indicates the distance threshold defined in b. The diagrams (blue bubbles) represent our interpretation of the data. a.u., arbitrary units. d. The same as c, measured for the 5′ (red) and 3′ (green) regions of lacZ mRNA, labelled using different sets of smFISH probes. e. The same as c, measured at different times after the addition of rifampicin to inhibit transcription initiation. f. The amount of nascent lacZ mRNA per cell (black, mean ± s.e.m.) as a function of inducer (IPTG) concentration. The data were fitted to a Hill function. See Supplementary Note for further details.
released into the cytoplasm; instead, about half of them (55 ± 5%) remain in the vicinity of the transcribed gene for the full lifetime of the mRNA. Incorporating this feature into our theoretical model yielded good agreement with the experimental data (Fig. 2c and Supplementary Fig. 15). In further support of the hypothesis of mature mRNA retention, the ratio of 5′ to 3′ signals in gene-proximal mRNA in glycerol was lower than expected for nascent mRNA, consistent with the presence of complete-but-unreleased mRNA molecules (Supplementary Fig. 16). Interestingly, we found that vigorous centrifugation of the cells (4,500g for 5 min) lowered the level of gene-proximal mRNA back to the expected level for the nascent species only (Fig. 2c and Supplementary Fig. 17) and restored the 5′ enrichment of proximal mRNA (Supplementary Fig. 16). The effect of centrifugation is thus consistent with the presence of two mRNA populations at the gene, with only the actively transcribed molecules strongly tethered (and therefore irremovable by centrifugation). We observed a similar behaviour for the lambda PR promoter (Supplementary Fig. 18). Two additional slow-growth media, succinate and acetate (g ≈ 120 and 240 min at 37°C, respectively) also showed evidence of mature RNA retention (Supplementary Fig. 18). The ability to remove retained mature RNA by centrifugation is used in subsequent experiments to allow the quantification of transcriptional activity without the confounding effects of mRNA retention.

![Diagram](image)

**Fig. 2 | Analysing nascent mRNA reveals the stochastic kinetics of transcript initiation, elongation, release and degradation.** a. A model for mRNA kinetics. The promoter stochastically switches (with rates $k_{on}$, $k_{off}$) between active and inactive states. In the active state, stochastic transcription initiation (with rate $k_{ini}$) is followed by mRNA synthesis (elongation) at a constant speed ($v_m$). Once the transcript is complete, mature mRNA is released from the gene into the cytoplasm. Degradation of both nascent and mature mRNA is initiated at the same rate ($k_{d}$). b. The levels of total, nascent and mature lacZ mRNA per cell at different times after adding IPTG. Cells were grown in glucose (37°C, 1mM cAMP). The data points and error bars show the experimental data (mean ± s.e.m. from two experiments). The solid lines show the fit to the theoretical model. c. The same as b, for cells grown in glycerol (30°C). The experimental data (mean ± s.e.m. from two experiments) were fitted using a revised model where a fraction of mature mRNA remains at the gene after centrifugation (empty circle). d. Following $P_{lac}$ activity in live cells. Each lac locus was detected through Tet-mCherry binding to the nearby tetO array. Endogenous lacZYA was replaced by 48 MS2 binding sites (48MS2bs), and $P_{lac}$-48MS2bs transcripts were detected using MS2-GFP. The activity state of each sister $P_{lac}$ copy (on/off) was determined based on the presence/absence of an RNA signal at the gene. Cells were grown at 30°C in LB media supplemented with 1mM IPTG. The yellow dashed lines indicate the cell boundaries. Scale bar, 1μm (all frames). e. The distributions of ‘on’ and ‘off’ durations for individual $P_{lac}$ copies measured in live cells. The data points and error bars show the experimental data (normalized counts; error bars indicate the standard deviation). The solid lines show exponential fits, allowing an estimation of the probabilistic rates of promoter switching. f. The distributions of nascent (per gene copy) and total (per cell) lacZ mRNA at different times following induction. The data (grey bars) are from one of the experiments included in b. The solid lines are fits to the stochastic model. All histograms were truncated along the y axis for visibility. g. The same as f, for the experiment in c. The data were fitted to the revised model that incorporates mRNA retention at the gene. h. The estimated rate of $P_{lac}$ switching to the active state ($k_{ini}$) and the transcriptional burst size ($b=k_{on}/k_{off}$), as a function of IPTG concentration, for cells grown in glucose. Steady-state lacZ expression data from exponentially growing cells were fitted to the stochastic model. Markers indicate the best-fit parameters. Error bars represent the range of estimated parameters from the top 0.2% likelihood fitting results. The solid lines show the fit to a Hill function. See Supplementary Note for further details.

2120
In the analysis above, we used the population-averaged measurements to interrogate mRNA kinetics. Next we aimed to use the full single-cell dataset for inferring the stochastic kinetics of a single promoter. Following the transcription from individual $P_{\text{lac}}$ copies in live cells (Fig. 2d) revealed that the durations of promoter activity (defined by the presence of nascent RNA) and inactivity periods both follow exponential distributions (Fig. 2e and Supplementary Fig. 19). This indicated that, despite the complex dynamics of transcriptional regulation, promoter activity can be phenomenologically described using stochastic two-state kinetics, as previously concluded from whole-cell mRNA measurements. We thus extended our mathematical model (Fig. 2a) to include stochastic promoter kinetics and used the model to fit the copy-number distributions of nascent and total cellular $\text{lacZ}$ mRNA during $P_{\text{lac}}$ induction, in both glucose and glycerol (Fig. 2f and Supplementary Figs. 20 and 21). The agreement between theory and experiment suggests that we can reliably capture the stochastic kinetics of nascent mRNA at the single-cell level. Applying the same procedure to cells expressing $\text{lacZ}$ at different steady-state levels (Supplementary Fig. 22) allowed us to identify that, upon $P_{\text{lac}}$ induction, the probabilistic rate of promoter activation, $k_{\text{on}}$, is the main parameter being modulated to vary expression (Fig. 2h and Supplementary Fig. 23), in a similar way to that reported in eukaryotes.

As part of the replication cycle of the bacterial chromosome, multiple copies of the same gene are often present in the same cell. The stochastic activity of these individual copies may be correlated due, for example, to fluctuations in an upstream regulator. This effect was inferred from the presence of so-called ‘extrinsic noise’ in mRNA and protein expression, but it was not previously possible to directly measure these gene-copy correlations. To ask whether the activity of individual gene copies is coupled, we again focused first on $P_{\text{lac}}$ activity in cells grown in glucose, under induction conditions where the fraction of transcriptionally active copies ($p_{\text{on}}$) was approximately half. Specifically, we examined the subpopulation of cells with exactly two sister copies of the $\text{lac}$ locus. We found that the fractions of cells that had 0, 1 and 2 transcriptionally active

---

**Fig. 3** | Promoter activity is coupled to the activity of additional gene copies in the cell. a. The transcriptional activity of individual copies of $P_{\text{lac}}$ in cells grown in glucose (37 °C, 1 mM cAMP and 100 µM IPTG). For each cell that had two endogenous sister lac copies, the position and activity of each copy is overlaid on the cell boundary (yellow), as in Fig. 1b above ($\text{lacZ}$ mRNA is not shown). Scale bar, 2 µm. b. The distribution of the number of active $P_{\text{lac}}$ copies in cells that had two copies of the gene. The grey bars show the data (normalized counts) for cells grown in glucose; error bars indicate the standard deviation. The red dots indicate the fit to a model assuming independent activity of the two copies (binomial distribution). c. The cross-correlation (mean ± s.e.m.) between two copies of $P_{\text{lac}}$ in the same cell measured in live cells grown in glucose. d–f. The same as a–c, respectively, for cells grown in glycerol (30 °C, 10 µM IPTG). g. The correlation between sister copies of $P_{\text{lac}}$ as a function of growth rate. The dots represent the data and the horizontal bars show the median across samples. Cells were grown in the following conditions: 30 °C in LB, 37 °C in glucose, 37 °C in glycerol, 30 °C in succinate, 37 °C in acetate. h. The correlation between sister copies of $P_{\text{lac}}$ as a function of expression level (total $\text{lacZ}$ mRNA per cell) for cells grown in glucose and glycerol. The circles represent the data and the dashed lines are polynomial fits serving as a guide for the eye. See Supplementary Note for more details.
copies followed a binomial distribution, as would be expected if each \( P_{nr} \) copy acted independently (Fig. 3a,b and Supplementary Fig. 24). Consistent with this observation, the measured copy–copy correlation in activity \( (r) \) was very low \( (r=0.12 \pm 0.05) \), as was the correlation between the nascent mRNA levels of the two copies (Supplementary Fig. 25). Live-cell measurements in the \( P_{nr} \) promoter showed similar dependence on the growth conditions, with two promoter copies having higher correlation in a slower growth medium (Supplementary Fig. 26). In addition to the dependence on growth rate, the degree of correlation (and the corresponding extrinsic noise value) also varied with expression level (Fig. 3h and Supplementary Figs. 28 and 29) and genomic location (Supplementary Fig. 28). On the other hand, the correlation between two sister copies did not depend on their physical distance or cell length (Supplementary Fig. 29) and was observed in both translated and untranscribed RNA (Fig. 3 and Supplementary Figs. 24–28).

Beyond the ability to count the gene copies in a given cell, our reporter system also allowed us to identify the time within the cell cycle at which gene replication took place, as indicated by the appearance of \( P_{nr} \) transcription close to the time of gene replication. The dots and error bars show the experimental data (mean ± s.d.), the solid line is a guide for the eye and the grey shading indicates the uncertainty in estimating the replication time. The procedure in d was repeated for cells at different IPTG concentrations. The amplitude of the fitted Gaussian is plotted versus the number of RNA per cell. The values corresponding to b and d are highlighted in dark and light blue, respectively. The dashed line shows the fit to a second degree polynomial. See Supplementary Note for further details.
the transcriptional activity of each P₀ copy is constant throughout the cell cycle. The same trend was observed when the promoter was placed at different genomic loci (Supplementary Fig. 31).

It has long been speculated that, rather than being uniformly probable throughout the cell cycle, transcription from low-expression promoters takes place only briefly, following gene replication. Replication-induced transcription could stem, for example, from the transient displacement of a repressor by the replication fork. To test this intriguing hypothesis, we examined the cell-cycle dependence of P₀ activity under low-IPTG conditions, where it is repressed by LacI. We found that, rather than simply following gene dosage, the amount of nascent lacZ mRNA exhibited a strong transient increase around the time of gene replication (Fig. 4c,d). A similar pattern could be seen in other growth conditions (Supplementary Fig. 32). We also observed the transient increase in transcription around gene replication by following P₀ activity in live cells (Fig. 4e,f and Supplementary Fig. 33). We further verified that the coupling between transcription and gene replication was not an artefact of the gene labelling scheme, by measuring lacZ mRNA numbers in genetically unmodified cells (strain MG1655), and similarly observed a higher probability of finding mRNA in cells whose length corresponded to the timing of gene replication (Supplementary Fig. 34). These results all indicate that the replication of a strongly repressed P₀ copy is accompanied by a transient increase in its activity. Consistent with the idea that the increased activity reflects a transient relief of LacI repression, we found that the relative effect of replication gradually diminished as lacZ expression increased, that is, as repression was relieved (Fig. 4g and Supplementary Fig. 32).

Measuring mRNA at the resolution of a single gene, rather than the whole cell, dramatically improved our ability to characterize the life history of mRNA during and after transcription. It also revealed how the stochastic activity of a single gene copy depends on the presence of additional copies and on the event of gene replication. Additional work will be required to elucidate the origins of these dependencies. In any event, their presence highlights the need to continue removing the hidden variables that drive cellular heterogeneity, rather than simply attributing this heterogeneity to unknowable “noise”. To detect promoter activity in live cells, an array of MS2b2s was placed under the control of the promoter of interest. The transcribed MS2b2s formed stem-loops, which were specifically bound by the MS2 coat protein, fused to a fluorescent protein.

We first constructed a series of plasmids carrying 24 or 48 MS2b2s under the control of different promoters. We started with pIG-48bs and pIG-24bs, which carry 48 and 24 MS2b2s, respectively, under the control of P₀ (ref. 36). Next, pBS48bs-Cm² and pBS24bs-Cm² were constructed by replacing the ampicillin resistance cassette (Amp⁰) in pIG-48bs and pIG-24bs with a chloramphenicol resistance cassette (Cm²), using recombineering. Cm² was amplified from pKΔ3 (ref. 37) using primers OC201-2 and OC202. pZO54 was constructed by replacing P₀ in pBS48bs-Cm² with the phage lambda promoter Pₐ. The Pₐ sequence was amplified from wild-type phage lambda using the primers PR-short-PP and PR-short-RP, cut with NotI and PciI and ligated into BS24bs-Cm², which was isolated by the same enzyme.

To place each MS2 reporter cassette (promoter-MS2b2s-Cm²) in the chromosome, each cassette was amplified from the corresponding plasmid and then recombineered into the lac locus. From there, it was PCR amplified and recombineered into other loci as needed. Transcription of these arrays was detected using MS2-GFP from plasmid pJZ107, whose construction is described below. The recombineering primers, promoters and genomic loci are listed in Supplementary Table 4.

The dual-reporter strains (FROS and MS2) were built by moving the MS2 reporter cassette to the strains carrying the 140tetO array in the chromosome, using transduction: Plasmids pJZ133 was transformed into E. coli DH5α cells, which were subsequently used to transform competent Escherichia coli strains to express TetR-YFP (used for smFISH experiments). Plasmid pJZ152 was transformed into these dual-reporter strains to simultaneously express TetR-mCherry and MS2-GFP (used for live-cell imaging). These plasmids are described below.

In the original configurations of both the FROS and MS2 systems, the fluorescently labelled binding proteins are expressed from inducible promoters. This often results in non-uniform expression across cells (data not shown). To optimize the expression level and achieve improved uniformity among cells, we placed our binding proteins (TetR-YFP, TetR-mCherry and MS2-GFP) under the control of constitutive synthetic promoters. Specifically, we used the pSR67 series of plasmids (pSR67.1–5) in which the protein of interest is expressed from one of five Anderson collection promoters of different strengths. To construct plasmid pJZ133, we amplified the tetYFP fragment from pDM21 using the primers TetR-YFP-GG-FP and TetR-YFP-GG-RP. The backbone of pSR67.1 (containing Pₐ, Bba_J23117 and 162 arbitrary Anderson promoter units) was amplified using primers pSR67-GG-FP and pSR67-GG-RP. The two fragments were digested using Bsal and ligated using Golden Gate assembly. pJZ102 was constructed in a similar way, except that tetR-mCherry was amplified from pKG110 using the primers TetR-mCherry-GG-FP and TetR-mCherry-GG-RP.

To construct plasmid pJZ107, we used the m2-gfp fragment from pJZ102 by using the primers TetR-mCherry-GG-FP and TetR-mCherry-GG-RP. The two fragments were digested using Bsal and ligated using Golden Gate assembly. We next combined tetR-mCherry and m2-gfp in a single plasmid. pJZ152 was constructed in the following way. The P₀-m2-gfp fragment was amplified from pJZ107 using the primers MS2-TetR-yfp-FP and MS2-TetR-yfp-TR. The backbone of pSR67.3 (containing Pₐ, Bba_J23105 and 623 arbitrary Anderson promoter units) was amplified using the primers pSR67-GG-FP and pSR67-GG-RP. The two fragments were ligated with using Bsal and ligated using Golden Gate assembly. The backbone of pJZ152 was constructed in a similar way, but with the two proteins placed in the reverse order. P₀-tetR-mCherry was amplified from pJZ102 by using the primers MS2-TetR-FFP and TetR-mCherry-GG-RP and ligated to the backbone of pJZ107 (containing Pₐ-tetR-yfp-GG-FP).

pJZ186 was constructed to lower the expression level of TetR-mCherry. This was achieved by replacing the ribosomal binding site Bba_0034 (Registry of Standard Biological Parts, http://parts.igem.org/Ribosome_Binding_Sites/Prokaryotic/Constitutive/Community_Collection) with Bba_0031 (0.7 relative strength to Bba_0034). A fragment containing Pₐ and Bba_0031 was synthesized as double-stranded DNA (gBlocks, IDT) with flanking Bsal cut sites. This fragment was then digested with Bsal and ligated with the backbone of pJZ156, which was amplified using the primers P₀ and P₀ and digested with the same enzyme.

pJZ16 was constructed in a similar manner to pJZ186, except that tetR-mCherry was placed upstream of m2-gfp. The same fragment (used above) containing Pₐ and Bba_0031 was digested with Bsal and ligated with the backbone of pJZ152, which was amplified using the primers P₀ and P₀ and digested with the same enzyme.

Methods

Bacterial strains and plasmids. All bacterial strains are listed in Supplementary Table 1, plasmids are listed in Supplementary Table 2 and primers are listed in Supplementary Tables 3 and 4. The construction of strains and plasmids is described below.

The locus of interest was labelled using FROS⁰. An array of tet operators (tetO) was placed near the gene of interest in the chromosome. Visualization of the locus under the microscope was accomplished through the cognate binding protein TetR, which is fused to a fluorescent protein. We started with a strain where 140 tetO copies (140tetO) were placed inside the mhpA gene, ~3.5 kb from lac (ref. 30). Using this construct, we then built a series of strains where the 140tetO array is placed at different genomic loci.

We first built the plasmid pJZ087, which was used to place the 140tetO array at the gene locus of interest. The plasmid carries the 140tetO array and a single FRT recognition site (FRT) site that allows the integration of the whole plasmid into strains from the Keio collection of single gene deletions using FRT-FRT recombination. To construct pJZ087, we amplified a fragment containing a single FRT site, the R6K replication origin (pir dependent replication, for eliminating the plasmid after FRT-FRT recombination) and a kanamycin resistance cassette (Kan³). This fragment was amplified from pKG137, a derivative of pC357 (ref. 31), using the primers FRT-FROS-FP-H1 and FRT-FROS-PR-H2, and then recombineered into plasmid pB123 (ref. 32) using standard protocols. pB123 carries the 140tetO array, with a gentamicin resistance cassette (Gen³) inside the array. The primers used to construct the plasmids are provided in Supplementary Table 3.

pJZ087 was next inserted into several Keio strains using FLP-FRT recombination. FLP recombinase was expressed from plasmid pC200 (ref. 30). The 140tetO array was moved to a clean genetic background using pC1 transduction. The array was detected by expressing either TetR-YFP from plasmid pJZ133, or TetR-mCherry, from plasmid pJZ102. The construction of these plasmids is described below.
(Minimal M9 broth minus carbon (Teknova), supplemented with 0.4% glycerol (Fisher Scientific)); (4) succinate (Minimal M9 broth minus carbon (Teknova), supplemented with 0.4% succinate (Sigma-Aldrich)); (5) acetate (Minimal M9 broth minus carbon (Teknova), supplemented with 0.4% Pyruvate (Sigma-Aldrich)). Cultures from fresh colonies were grown overnight (14–16 h) with antibiotics when appropriate: 100 μg/ml ampicillin (Fisher Scientific), 50 μg/ml kanamycin (Fisher Scientific), 17 μg/ml chloramphenicol (Fisher Scientific), 5.5 μg/ml gentamicin (Sigma-Aldrich), 100 μg/ml spectinomycin (Fisher Scientific).

Depending on the growth conditions of the overnight cultures, we used two different overnight culture set-ups. If the overnight condition was LB, the overnight cultures were grown in LB as well; otherwise, the overnight cultures were grown in glucose. If the overnight condition was glycerol, succinate or acetate, the overnight (glucose) culture was diluted at least 1:800. The overnight cultures for each type of experiment were grown as described below. Detailed information regarding strains and growth conditions is provided in the Supplementary Note.

For smFISH steady-state experiments, the overnight cultures were used to prepare overnight cultures at dilutions ranging from 1:200 to 1:2,000 in 30 ml medium with the appropriate supplements and grown in 250 ml baffled flasks to an optical density (OD)\(_{600}\) ≈ 0.2. Each sample was then treated according to the procedure described below. For smFISH induction experiments, the overnight cultures were diluted (1:250 to 1:1,000) in 200 ml medium with the appropriate supplements and grown in 1,000 ml baffled flasks to OD\(_{600}\) ≈ 0.2. IPTG (Sigma–Aldrich) was added to a final concentration of 100 μM (glycerol) or 1,000 μM (glucose) at t = 0. 10 ml samples were collected at different time points and treated according to the procedures described below. For FISH rifampicin experiments, the overnight cultures were diluted (1:250 to 1:1,000) into 120 ml medium with the appropriate supplements and grown in 1,000 ml baffled flasks to OD\(_{600}\) ≈ 0.2. Rifampicin (Fisher Scientific) was added to a final concentration of 500 μg/ml (ref. 39) at t = 0. 10 ml samples were collected at different time points and treated according to the procedures described below.

For live-cell snapshots, the overnight cultures were diluted (1:500 to 1:2,000) in 10 ml medium with appropriate supplements and grown in 125 ml baffled flasks to OD\(_{600}\) ≈ 0.2–0.4. Cells were then prepared for imaging according to the procedures described below. For live-cell time-lapse videos, the overnight culture was diluted (1:500 to 1:2,000) in 10 ml medium with appropriate supplements and grown in 125 ml baffled flasks to OD\(_{600}\) ≈ 0.2–0.4. Cells were then prepared for imaging according to the procedures described below.

**smFISH.** The smFISH protocol was described in detail previously40. Briefly, a set of antisense DNA oligo probes was designed against the gene of interest and synthesized with a 3' amine modification (LGC Biosearch Technologies). The oligos were pooled, covalently linked to fluorescent dyes (Invitrogen) and purified through ethanol precipitation. Probe sequences and fluorescent dyes are listed in Supplementary Table 5. Cells were grown as described above, then harvested, fixed and permeabilized. Cells were incubated with fluorescently labelled probes, washed and then imaged as described below. We made the following modifications relative to the original protocol from ref. 40: (1) A final concentration of 1% formaldehyde was used for fixation. For steady-state experiments, we added an additional step between cell harvesting and fixation. Following harvesting (centrifuging at 4,500 g for 5 min), the cell pellets were resuspended in 1 ml PBS, then centrifuged at 4,500 g for 1 min. This washing step is meant to ensure the proper pH, since YFP is pH sensitive41. (2) For non-steady-state experiments, at each time point, the culture was taken out and directly mixed 1:1 (equal volume) with 2% formaldehyde solution in 2× PBS.

**Preparation of MS2 reporter cells for imaging.** For the snapshots, cells were grown as described above. Unless otherwise noted, 1 ml of each sample was transferred to a 1.5 ml microcentrifuge tube and centrifuged at 15,000 rpm for 30 s and the cell pellet was resuspended in 50 μl of the same medium. Cells were then imaged as described below. For the tests on the effect of centrifugation on RNA retention, cells were prepared for imaging in two different ways. Without centrifugation: cells were directly taken from the culture and placed under the microscope for imaging. With centrifugation: to mimic the procedure of cell preparation in smFISH steady-state experiments, cells were harvested by centrifuging at 4,500 g for 5 min, then washed in 1 ml PBS at 4,500 g for 1 min. The cell pellet was resuspended in 50 μl 1× PBS for imaging.

For non-steady-state experiments (for example, drug treatment), our protocol was adapted from the corresponding smFISH experiments. Cells were fixed and prepared for taking snapshots in two different ways. Without centrifugation: cells were directly mixed 1:1 (equal volume) with 2% formaldehyde solution in 2× PBS for fixing; cells were then washed twice in 1× PBS and prepared for imaging as described below. With centrifugation: cells were harvested by centrifuging at 4,500 g for 5 min and then washed in 1 ml 1× PBS at 4,500 g for 1 min; cells were resuspended in 1 ml 1% formaldehyde solution in 1× PBS for fixation, washed twice in 1 ml 1× PBS and prepared for imaging.

For the time-lapse videos, cells were grown as described above. Videos were acquired using the CellASIC ONIX microfluidic system (Millipore) placed in a temperature-controlled enclosure (Okolab), following the manufacturer’s protocol. In brief, cells and media were first pipetted into the appropriate wells in the microfluidic plate. The plate was then sealed to the ONIX manifold and placed under the microscope. Cells were loaded and trapped in the imaging area. Both temperature and flow speed were maintained for at least 30 min before imaging, to achieve stable cell growth. Medium (2% glucose) was added to the ONIX (0.2 µg/ml) at 4,500 g for 5 min and washed in 1 ml 1× PBS at 4,500 g for 1 min; cells were resuspended in 1 ml 1% formaldehyde solution in 1× PBS for fixation, washed twice in 1 ml 1× PBS and prepared for imaging.

**Microscopy.** We used an inverted epifluorescence microscope (Eclipse Ti, Nikon), equipped with motorized stage control (ProScan III, Prior Scientific), a universal specimen holder, a mercury lamp (Intensilight C-HGFI, Nikon), filter sets (YFP, GFP, TxRed, Cy5, Nikon) and either an EMCCD camera (Cascade II: 1024 Photometrics) or a CMOS camera (Prime 95B, Photometrics). A ×100, NA 1.40, oil-immersion phase-contrast objective (Plan Apo, Nikon) was used, as well as a ×2.5 magnification lens (Nikon) in front of the camera.

To acquire the snapshots, cells were prepared for imaging as described above for live-cell analysis. The samples were then placed onto the microscope’s slide holder and the cells were visually located using the phase-contrast channel. In all of the experiments, we used 100 ms exposure for phase-contrast images. For the fluorescence channels (YFP, GFP, TxRed, Cy5), we used exposure times between 0.2 and 1 s, with a gain of 2,000–3,500 (when using the EMCCD camera). Fixed-cell snapshots were taken at 9 z positions (focal planes) with steps of 200 nm or 5 z positions with steps of 300 nm. Live-cell snapshots were taken at 5–7 z positions with steps of 300 nm. A set of images with multiple z positions is denoted as an ‘image stack’ and the image of each z position as ‘a slice’. Images were acquired at multiple slice positions, to image a total of 400–4,000 cells per sample (typically 10–30 positions).

To acquire time-lapse videos, cells were prepared for imaging as described above. Time-lapse videos were taken at three z positions with steps of 500 nm. We used an exposure time of 100 ms, EM gain of 3,000 and Neutral Density (ND) filter 8–16 for the GFP channel; and an exposure time of 100–200 ms, EM gain of 3,500 and ND 1–2 for the TxRed channel. Videos were acquired at multiple slice positions (typically starting at 5 z positions and decreasing as the experiments progress). Depending on the growth conditions, videos were acquired at a frame rate of 5 or 10 min.

**Cell recognition and lineage tracking.** We used Schnitzcell41 to identify cells in the phase-contrast channel of snapshots of live and fixed cells, as well as time-lapse video images in an image stack, the slice with the highest variance of pixel values was identified as ‘in focus’ and used for cell segmentation. The segmentation results were visually inspected; poorly segmented cells were either discarded or manually corrected using the software’s graphical interface. For time-lapse videos, following segmentation, we used the built-in capability of Schnitzcell to track cell identity and lineage over time.

**Spot recognition and quantification.** We used Spätzcell11 to identify and quantify foci (‘spots’) in the fluorescence images of live and fixed samples. Briefly, Spätzcell first identifies the local maxima above a user-defined threshold, in every z slice in an image stack. It then connects the local maxima from different z slices to one another, and to the same spot. For each spot, the local maxima is defined as the one where the spot has the highest intensity. In that plane, the fluorescence intensity profile within a small region around each spot is fitted to one or more two-dimensional elliptical Gaussians, with the number of Gaussians equal to the number of local maxima within the region. The following properties of each spot are obtained from the fitting procedure and used in subsequent analysis: position, area (π times the major and minor axis of the fitted Gaussian), peak height (amplitude of the fitted Gaussian), spot intensity (volume underneath the fitted Gaussian).

Spätzcell was originally optimized for smFISH images, where there is almost no background fluorescence in the cell. In our FROS and MS2 images, where spots correspond to bound fluorescent proteins, there was often a high level of background fluorescence in the cell. To improve spot recognition in those images, we modified the first step in Spätzcell, namely identifying local maxima at different z slices, as follows. For each z slice, we performed an à trous wavelet three-plane decomposition and obtained the second wavelet plane28,42. We then calculated the Laplacian of the second wavelet plane and set a threshold to identify the local maxima. The subsequent steps (connecting spots in different z slices, identifying the focal z slice for each spot and fitting) were unchanged. The intensity profile used for fitting was obtained from the original (unprocessed) image.

To discard false positive spots in smFISH images, we followed the procedure described in ref. 11. Briefly, the distribution of spot peak heights in a given sample was compared to the results from a negative sample (a sample without the RNA of interest). A threshold was chosen such that ~99% of spots from the negative sample are below (dimmer than) the threshold. The same threshold value was then used in the experimental sample, with only spots brighter than the threshold considered real RNA spots and used in subsequent analysis.

For the time-lapse images, the two-dimensional scatter plot of peak height versus spot area was compared to the results from a negative sample (here, images of cells expressing the fluorescent protein but lacking the cognate binding sites in the chromosome, see Supplementary Fig. 30a). Manual gating was
then used to discard the spot population present in the negative sample and the choice of gating was confirmed by manual inspection of spots in a subset of images. RNA quantification was performed as described in ref. 1. Briefly, after discarding false positive spots, we first determined a low-expression sample, where individual RNA were spatially separated. We fitted the histogram of spot intensities to sums of Gaussians corresponding to one, two and three RNA molecules per spot. The centre of the first Gaussian was then used to estimate the fluorescence intensity corresponding to a single RNA. Subsequently, for each smFISH spot in any sample, we converted the measured spot intensity to an RNA number based on the above single-RNA intensity. Likewise, total RNA copy number per cell was calculated by summing the spot intensities of all spots within the same cell, converted to RNA number.

For gene-copy identification, we note that, after removing false positive spots, the majority of the remaining FROS spots corresponded to individual gene copies. Accordingly, the mean numbers of gene spots per cell were consistent with previously reported values and with the theoretically expected copy number at different genomic loci and growth rates (Supplementary Fig. 2a,b). Under most experimental conditions, cell fixation and the smFISH procedure resulted in only a minor loss (<10%) of FROS spots (Supplementary Fig. 2b).

Whereas most FROS spots represent individual gene copies, we also expect to observe a fraction of spots corresponding to replicated sister copies that are still in cohesion with each other and therefore optically inseparable14,15. Consistent with this expectation, under some experimental conditions we were able to observe two distinct populations of FROS spots, with the peak height of the brighter population approximately twofold that of the dimmer population (Supplementary Fig. 3a,b). The fraction of twice-brighter spots (≥400 300 nm) or mature (≥800 nm) was consistent with previously reported duration of sister-copy cohesion14,15. In time-lapse videos, we often observed the FROS spot intensity increasing before the spot split in two (Supplementary Fig. 3b), again consistent with the scenario of transient cohesion of replicated sister copies.

Measuring nascent RNA and identifying active gene copies in fixed cells. The identification of nascent RNA relies on accurately detecting colonized gene and RNA spots. We first corrected for the effect of chromatic aberration, which creates a shift in the relative positions of images acquired in different fluorescent channels. The correction for chromatic aberration was performed as described in ref. 16. Briefly, we imaged fluorescent beads (TetraSpeck, Fisher Scientific) using the same imaging parameters as the sample slides. In each imaged channel, the spots (individual RNA molecules) were identified and localized using SpatzCell. The theoretical offset between bead centres in two channels (Δx, Δy) was well described by a linear function of the bead position (x, y). Using this linear fit allowed us to correct the position of each bead. The same fit was then used to correct the images from the experimental samples.

After correcting the chromatic aberration, we calculated the distance from the centre of each RNA spot to the centre of its nearest gene locus in the same cell. Under multiple experimental conditions, the resulting histogram of RNA-to-gene distances (Fig. 1b and Supplementary Fig. 3) revealed two distinct populations of RNA residing, respectively, in close proximity (within <300 nm) to the gene and further away from it. This observation was rendered more quantitative by fitting the distance histogram to a sum of two Gaussian functions (Supplementary Fig. 8a). For wild-type PcrA at medium-to-high expression levels (where the two distinct populations were most clearly seen), the distance threshold was similar across different growth media (LB, glucose and glycerol; Supplementary Fig. 3). The threshold value was also similar in two additional promoters (Supplementary Fig. 6). The gene-proximal RNA population disappears when the RNA positions are numerically randomized (Supplementary Fig. 5). Using these observations, we classified each RNA spot on the basis of its distance to the nearest gene copy, as either nascent (<300 nm) or mature (≥300 nm). Similarly, each gene copy was classified as active (presence of RNA spot within 300 nm) or inactive (absence of RNA spot within 300 nm). We further corrected for the possible false identification of nascent RNA, as described in the Supplementary Note.

For each gene spot, the amount of nascent RNA was measured by summing over the intensities of all RNA signals within the distance threshold and converting to RNA copy number using the single-RNA intensity. The resulting value is denoted as ‘nascent RNA per gene copy’. For each cell, the amount of nascent RNA was measured in a single way by summing over the intensities of all the nascent RNA signals in the cell and converting to the RNA copy number. The resulting value is denoted as ‘nascent RNA per cell’.

As mentioned above, a fraction of gene spots correspond to unseparated sister loci rather than individual gene copies. Therefore, nascent RNA measured at these loci corresponds to the total nascent RNA from two gene copies. How this population affects the distribution of nascent RNA per gene copy, we performed the following calculation. \( P_u(m) \) denotes the true distribution of nascent RNA per copy and \( c \) is the fraction of gene spots that are unseparated sister copies. The observed distribution of nascent RNA per gene spot (assuming independent transcription from the two sister copies) can then be written as:

\[
P_{\text{obs}}(m) = (1-c) \times P_u(m) + c \times P_u(m) \]

\( \ast \) represents the convolution operator. To evaluate the difference between \( P_u(m) \) and \( P_{\text{obs}}(m) \), we used our experimentally estimated kinetic parameters to calculate both distributions and found them to be statistically indistinguishable within our experimental accuracy (data not shown).

Identifying active gene copies in live cells. In live-cell snapshots, the activity state of each gene copy was determined as follows. First, cell segmentation and spot recognition were performed as described above, followed by correction for chromatic aberration and co-localization analysis of the gene and RNA signals as in fixed cells. After examining the RNA-to-gene distance histograms in multiple samples (Supplementary Fig. 7c), a value of 450 nm was chosen for the distance threshold between nascent and mature RNA spots, which was classified as active/inactive on the basis of the presence/absence of RNA within 450 nm of it. In a number of samples, image quality was insufficient to perform the automated analysis described above and instead we identified active gene copies manually by visually inspecting for the presence of RNA signal within 10 pixels (∼500 nm) of the gene. When applied to the same sample, manual and automated analysis yielded similar estimates of \( p_c \) (data not shown).

The analysis of time-lapse videos is complicated by the need to track cell and spot identity over time. For automated analysis, cell segmentation and lineage tracking were performed as described above. As part of the output of Schnitzcell, each branch of a cell lineage (called a ‘schnitz’) tracks a cell from birth to division. To identify spots and active gene copies, we first treated the time-lapse frames as snapshots and then incorporated the spot measurements into the original schnitz to keep track of cells, gene copies and RNA simultaneously. The intensity of each gene spot was used to estimate the gene-replication time (Supplementary Fig. 3b). In our analysis, we included only those schnitzs that fulfil the following criteria: the cell was successfully tracked through its full cell cycle and the cell doubling time was within 75–125% of the average doubling time of all cells in that video. When video quality was insufficient for automated analysis, we manually recorded the timing of cell birth and division, and of sister-copy separation, and the activity of each gene spot.

Calculating the correlation in activity between two gene copies. In snapshot experiments, the nascent RNA per gene copy was obtained as described above. Cells with two gene copies were gated as described below. The correlation in activity states of the two copies was calculated using:

\[
\tau = \frac{\langle i(t) \rangle - \langle i(t) \rangle}{\sigma_{ij}}
\]

\( \langle i(t) \rangle \) and \( \langle j(t) \rangle \) represent the activity (0/1) of the two copies in the same cell. In a similar way we calculated the correlation between nascent RNA levels of two gene copies (Supplementary Fig. 25) and the corresponding extrinsic noise, using the definition in ref. 16 (Supplementary Fig. 28).

In time-lapse videos, we tracked the activity (0/1) of individual gene copies in the cell over time, as described above. The cross-correlation between the two gene copies in the same cell was calculated as:

\[
C(\tau) = \frac{\sum_{t=1}^{N} \left( [i(t) - \langle i(t) \rangle] \times [j(t+\tau) - \langle j(t+\tau) \rangle] \right)}{\sqrt{\sum_{t=1}^{N} [i(t) - \langle i(t) \rangle]^2} \times \sum_{t=1}^{N} [j(t+\tau) - \langle j(t+\tau) \rangle]^2}
\]

where \( \langle i(t) \rangle \) and \( \langle j(t+\tau) \rangle \) are the activities of the two gene copies in the same cell at time \( t \) and \( t+\tau \), respectively. The standard deviation of the time series, to obtain \( C(\tau) \) and \( C(-\tau) \) to generate a symmetric function. As controls, we also calculated the corresponding cross-correlation for randomly shuffled data for each gene copy (data not shown).

Analysing cell-cycle data. Sister-copy cohesion, discussed above, was used to estimate the gene-replication time. In smFISH experiments, plotting the intensity of individual gene spots versus cell length (Fig. 3a,c and Supplementary Fig. 31) revealed peaks, corresponding to the cell-cycle phase with the highest occurrence of unseparated sister copies, that is, immediately following gene replication10. To estimate the cell-length position of the replication events, we fitted the binned data to a sum of two Gaussian functions (corresponding to replication events) and a second degree polynomial (capturing slower changes along the cell cycle):

\[
y = a_1 e^{-(x-s_1)^2} + a_2 e^{-(x-s_2)^2} + w_1 x^2 + w_2 x + w_3
\]

The centres of the two Gaussians \((a, b)\) were then used as the estimated cell lengths at which gene replication took place. As seen in Supplementary Fig. 31, these lengths exhibited the expected dependence on genomic locus.

To describe the transient response to gene replication, nascent RNA level per cell was plotted versus cell length, normalized to the sample mean and binned. In the case that promoter activity simply follows gene dosage (for example, for
unpressed P$_r$, Fig. 4b and Supplementary Fig. 31), we fitted the data to the sum of two Hill functions, corresponding to two rounds of gene replication:

$$y = c \left( 1 + \frac{1}{1 + \left( \frac{x}{y} \right)^a} + \frac{2}{1 + \left( \frac{x}{z} \right)^b} \right)$$

(4)

The parameter $r$ describes the fold change in cell length between the successive replication events. It is expected to be close to (but not necessarily equal to) two (ref. 15) and this is indeed what our analysis shows (Supplementary Figs. 31 and 32).

For datasets exhibiting a pulsatile response to the event of gene replication (for example, P$_r$-, Fig. 4d and Supplementary Figs. 31 and 32), we modified the fit by adding two Gaussians centred at the half-maximum points of the Hill functions:

$$y = c \left[ 1 + \frac{1}{1 + \left( \frac{x}{y} \right)^a} + \frac{2}{1 + \left( \frac{x}{z} \right)^b} \right] + u e^{-x-u} + v e^{-x-v}$$

(5)

The magnitudes of the Gaussian functions ($u$, $v$) were used to estimate the effect of gene replication on transcription (Fig. 4g and Supplementary Fig. 32).

To identify gene replication in time-lapse videos, we measured the total intensity of gene spots in the cell over time. For each smFISH fulfilling our gating criteria discussed above, we then manually identified the time point at which that intensity approximately doubled (Supplementary Fig. 30b). This point was estimated to be the gene replication time.

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

**Data availability**

The data that support the findings of this study are available from the corresponding author on request.

**Code availability**

The custom MATLAB routines used for processing and analysing the fluorescence microscopy data are freely available from the corresponding author on request.

Received: 14 November 2018; Accepted: 2 August 2019; Published online: 16 September 2019

**References**

1. Sanchez, A. & Golding, I. Genetic determinants and cellular constraints in noisy gene expression. Science 342, 1188–1193 (2013).

2. Golding, I. et al. Real-time kinetics of gene activity in individual bacteria. Cell 123, 1025–1036 (2005).

3. Taniguchi, Y. et al. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538 (2010).

4. Jones, D. L., Brewster, R. C. & Phillips, R. Promoter architecture dictates cell-to-cell variability in gene expression. Science 346, 1533–1536 (2014).

5. Sepúlveda, L. A. et al. Measurement of gene regulation in individual cells reveals rapid switching between promoter states. Science 351, 1218–1222 (2016).

6. Raj, A. et al. Stochastic mRNA synthesis in mammalian cells. PLoS Biol. 4, e309 (2006).

7. So, L. H. et al. General properties of transcriptional time series in Escherichia coli. Nat. Genet. 43, 554–560 (2011).

8. Peterson, J. R. et al. Effects of DNA replication on mRNA noise. Proc. Natl Acad. Sci. USA 112, 15886–15891 (2015).

9. Zenklusen, D., Larson, D. R. & Singer, R. H. Single-RNA counting reveals noisy gene expression. Cell 166, 729–739 (2016).

10. Elowitz, M. B. et al. Stochastic gene expression in a single cell. Science 297, 1183–1186 (2002).

11. Gupta-Sarma, P. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of Escherichia coli? BioEssays 17, 987–997 (1995).

12. Hammar, P. et al. Direct measurement of transcription factor dissociation excludes a simple operator occupancy model for gene regulation. Nat. Genet. 46, 405–408 (2014).

13. Golding, I. Infection by bacteriophage lambda: an evolving paradigm for cellular individuality. Curr. Opin. Microbiol. 43, 9–13 (2017).

14. Symmons, O. & Raj, A. What’s luck got to do with it: single cells, multiple fates, and biological nondeterminism. Mol. Cell 62, 788–802 (2016).

15. Lai, I. F. et al. Spatial and temporal organization of replicating Escherichia coli chromosomes. Mol. Microbiol. 49, 731–743 (2003).

16. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 0006.0008 (2006).

17. Ellermeier, C. D., Janakiraman, A. & Slauch, J. M. Construction of targeted single copy lac fusions using 3′ and FLP-mediated site-specific recombination in bacteria. Gene 290, 153–161 (2002).

18. Thomason, L. et al. Recombining: genetic engineering in bacteria using homologous recombination. Curr. Protoc. Mol. Biol. 78, 1.6.1–1.6.24 (2007).

19. Cherepanov, P. P. & Wackernagel, W. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of FLP-catalyzed excision of the antibiotic-resistance determinant. Cell 158, 9–14 (1995).

20. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl Acad. Sci. USA 97, 6640–6645 (2000).

21. Schmidli, S. R. et al. Refactoring and optimization of light-switchable Escherichia coli two-component systems. ACS Synth. Biol. 3, 820–831 (2014).

22. Angler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. PLoS ONE 3, e3647 (2008).

23. Bornstein, J. A. et al. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl Acad. Sci. USA 99, 9697–9702 (2002).

24. Llopis, J. et al. Measurement of cytosolic, mitochondrial, and Golgi pH in single living cells with green fluorescent proteins. Proc. Natl Acad. Sci. USA 95, 6803–6808 (1998).

25. Voigt, J. W. et al. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. Nat. Protoc. 7, 80–88 (2012).

26. Olivo-Marín, J.-C. Extraction of spots in biological images using multiscale products. Pattern Recognit. 35, 1989–1996 (2002).

27. Bates, D. & Kleckner, N. Chromosome and replisome dynamics in E. coli: loss of sister cohesion triggers global chromosome movement and mediates chromosome segregation. Cell 121, 899–911 (2005).

28. Nielsen, H. J. et al. Progressive segregation of the Escherichia coli chromosome. Mol. Microbiol. 61, 383–393 (2006).

29. Nielsen, H. J. et al. Dynamics of Escherichia coli chromosome segregation during multifork replication. J. Bacteriol. 189, 8660–8666 (2007).

30. Xu, H. et al. Combining protein and mRNA quantification to decipher transcriptional regulation. Nat. Methods 12, 739–742 (2015).

**Acknowledgements**

We are grateful to the following people for their generous advice and for providing reagents: D. Bates, J. Elif, H. Garcia, M. Girard, J. Halliday, C. Herman, M. Joshi, D. Magnan, J. Moffit, E. Nuñez, R. Phillips, A. Sarrion-Perdigones, S. Sebastian, L. Skarstedt, A. Sokac, J. Tabor, K. Venken and all the members of the golding lab. Work in the golding lab is supported by grants from the National Institutes of Health (grant no. R01 GM082837), the National Science Foundation (grant nos. PHY 1147948, PHY 1430124 and PHY 1427654), the
Welch Foundation (grant no. Q-1759) and the John S. Dunn Foundation (Collaborative Research Award). H.X. was supported by the Burroughs Wellcome Fund Career Award at the Scientific Interface (grant no. 1013907), the Thousand Talents Plan of China (Programme for Young Professionals), the National Natural Science Foundation of China (grant no. 11774225), the National Key Research and Development Programme of China (grant no. 2018YFC0310800) and the National Science Foundation of Shanghai (grant no. 18ZR1419800). We gratefully acknowledge the computing resources provided by the CIBR Center of the Baylor College of Medicine.

Author contributions
M.W., J.Z., H.X. and I.G. designed the experiments. M.W. and J.Z. performed the experiments. M.W., J.Z. and H.X. analysed the data. M.W., J.Z. and I.G. wrote this Letter.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41564-019-0553-z.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to I.G.

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

© The Author(s), under exclusive licence to Springer Nature Limited 2019
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- Confirm that the statistical test(s) used AND whether they are one- or two-sided
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
- Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection All microscopy data was collected using Nikon Elements.

Data analysis Data analysis was performed using custom Matlab (2013b, 2015b, 2017b) code.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw images and analyzed data will be available upon request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf
Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

**Sample size**
The sample size for each figure is detailed in the Supplementary Note (Section 4, "DATA SUMMARY AND STATISTICS").

**Data exclusions**
No data was excluded.

**Replication**
There were no experiments that we failed to replicate or reproduce. The numbers of biological repeats for each figure are described in the figure captions and in the Supplementary Note (Section 4, "DATA SUMMARY AND STATISTICS").

**Randomization**
No randomization was performed, since all measurements and analysis are fully automated to prevent bias.

**Blinding**
No blinding was performed, since all measurements and analysis are fully automated to prevent bias.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

**Materials & experimental systems**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | Antibodies            |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology         |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |
| ☒   | Clinical data         |

**Methods**

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | ChIP-seq              |
| ☒   | Flow cytometry        |
| ☒   | MRI-based neuroimaging |