LETTERS
https://doi.org/10.1038/s41564-019-0553-z
nature microbiology

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

Mengyu Wang[1,2,3,7], Jing Zhang[2,7], Heng Xu[4,5] and Ido Golding[1,2,6*]

Single-cell measurements of mRNA copy numbers inform our understanding of stochastic gene expression[1–3], but these measurements coarse-grain over the individual copies of the gene, where transcription and its regulation take place stochastically[4,5]. 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 bacteria[2] and eukaryotes[6] and showed how gene expression noise is modulated by physiological parameters[1,3,5,7]. However, whole-cell mRNA measurements represent 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 full lifetime of mRNA molecules and cannot distinguish actively transcribed mRNA from completed transcripts. Combined, these factors limit our ability to reliably map measured whole-cell mRNA numbers to the underlying stochastic kinetics of a single gene of interest[4,5,8].

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[9]. 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)[10], 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 molecules[5,11]. 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 ($\lesssim$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 gene[12] (Fig. 1c). Proximal mRNA was also enriched for the 5′ region of the gene, as would be expected from the presence of incomplete transcripts[12] (Fig. 1d). Finally, inhibiting transcription initiation using rifampicin led to the gradual disappearance, within a few minutes, of proximal mRNA, as would be expected for the completion and release of already-initiated transcripts (Fig. 1e). We found similar behaviour when examining two additional promoters (P_{lacI} and phage lambda P_R) under multiple growth conditions and in live cells, where RNA was labelled using MS2 tagging[13] (Supplementary Figs. 3–7). The evidence thus indicates that gene-proximal mRNA 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 medium[14]. 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_{el}, to a final length, L (refs. [15,16]). Once the transcript is complete, mature mRNA is released from the gene into the cytoplasm[15,16]. The degradation of both nascent

[1]Department of Physics, University of Illinois at Urbana–Champaign, Urbana, IL, USA. [2]Verna and Marrs McLean Department of Biochemistry and Molecular Biology, Baylor College of Medicine, Houston, TX, USA. [3]Graduate Program in Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX, USA. [4]School of Physics and Astronomy, Shanghai Jiao Tong University, Shanghai, China. [5]Institute of Natural Sciences, Shanghai Jiao Tong University, Shanghai, China. [6]Department of Microbiology, University of Illinois at Urbana–Champaign, Urbana, IL, USA. [7]These authors contributed equally: Mengyu Wang, Jing Zhang. *e-mail: igolding@illinois.edu

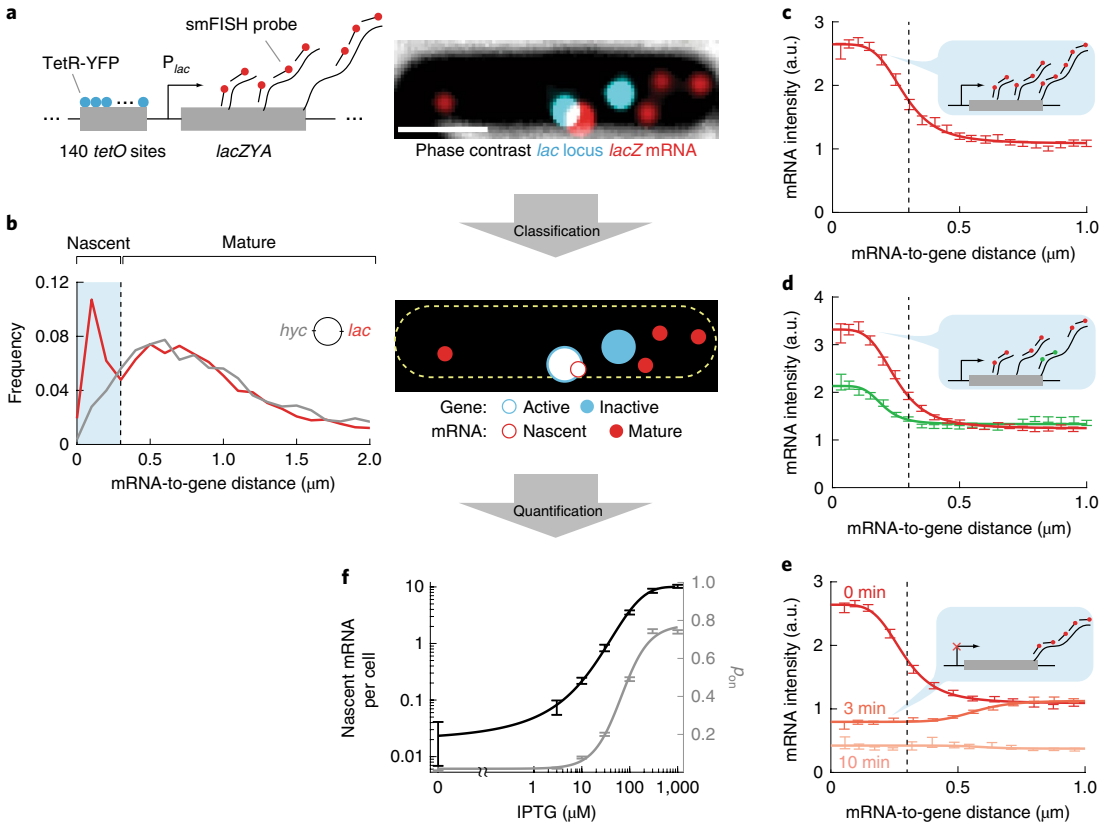

Fig. 1 | Detecting active transcription at a single gene copy. a, Left, the *lac* locus is detected through the binding of TetR-YFP to an array of 140 *tetO* sites inserted nearby in the *E. coli* chromosome. Endogenous *lacZ* mRNA, transcribed from P_{lac}, is simultaneously detected using smFISH. Right, 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-gene 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 *lac* 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.) and the fraction of transcriptionally active P_{lac} copies (p_{on}, grey, mean ± s.d.) as a function of inducer (IPTG) concentration. The data were fitted to a Hill function. See Supplementary Note for further details.

and mature mRNA is assumed to initiate at rate k_i (ref. [17]). mRNA degradation is limited by the competition between the degradation machinery and translating ribosomes, leading to a degradation speed that is equal to v_{el} (ref. [17]).

Induction kinetics in glucose medium (where P_{lac} exhibits a large dynamic range; Fig. 1f and ref. [18]) 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 $L/v_{el} \approx 130$ 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 ($v_{el} = 42 \pm 2$ 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_{el} is positively correlated with the rate of initiation (Supplementary Fig. 13), consistent with earlier reports[20,21]. The assumptions that nascent mRNA is degraded and that degradation proceeds at speed v_{el} 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 $g \approx 150$ min at 30 °C, compared to $g \approx 50$ 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

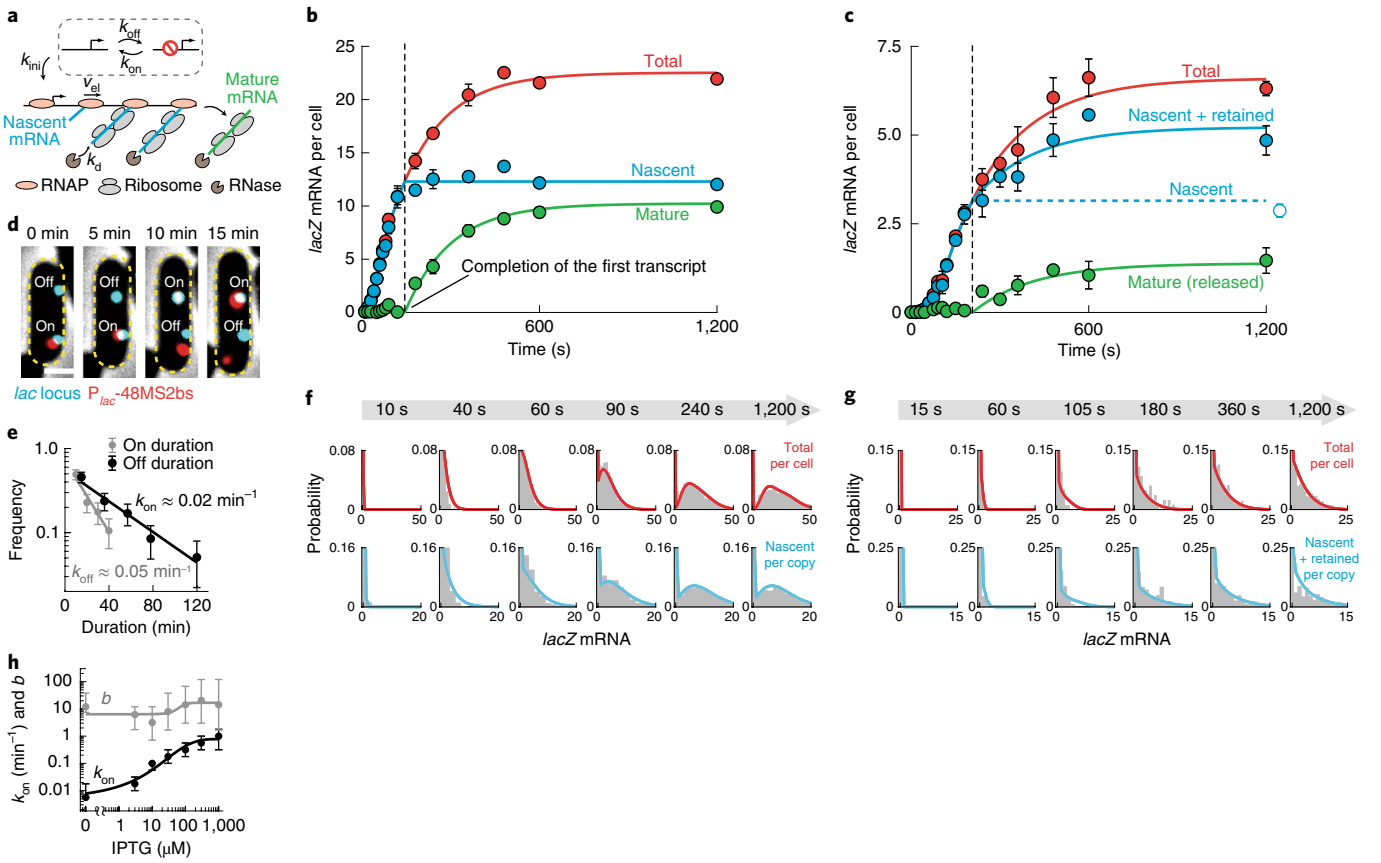

Fig. 2 | Analysing nascent mRNA reveals the stochastic kinetics of transcript initiation, elongation, release and degradation. a, A model for mRNA kinetics. The promoter stochastically switches (with rates k_{on}, k_{off}) between active and inactive states. In the active state, stochastic transcription initiation (with rate k_{ini}) is followed by mRNA synthesis (elongation) at a constant speed (v_{el}). 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, 1 mM 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 completion. The predicted steady-state level of nascent *lacZ* mRNA (cyan dashed line) is consistent with measurements using a protocol that includes cell centrifugation (empty circle). **d**, Following P_{lac} activity in live cells. Each *lac* locus was detected through TetR-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 1 mM IPTG. The yellow dashed lines indicate the cell boundaries. Scale bar, 1 μm (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_{on}) and the transcriptional burst size ($b = k_{ini}/k_{off}$), 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.

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 (4,500*g* 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 P_R promoter (Supplementary Fig. 18). Two additional slow-growth media, succinate and acetate ($g \approx 120$ 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.

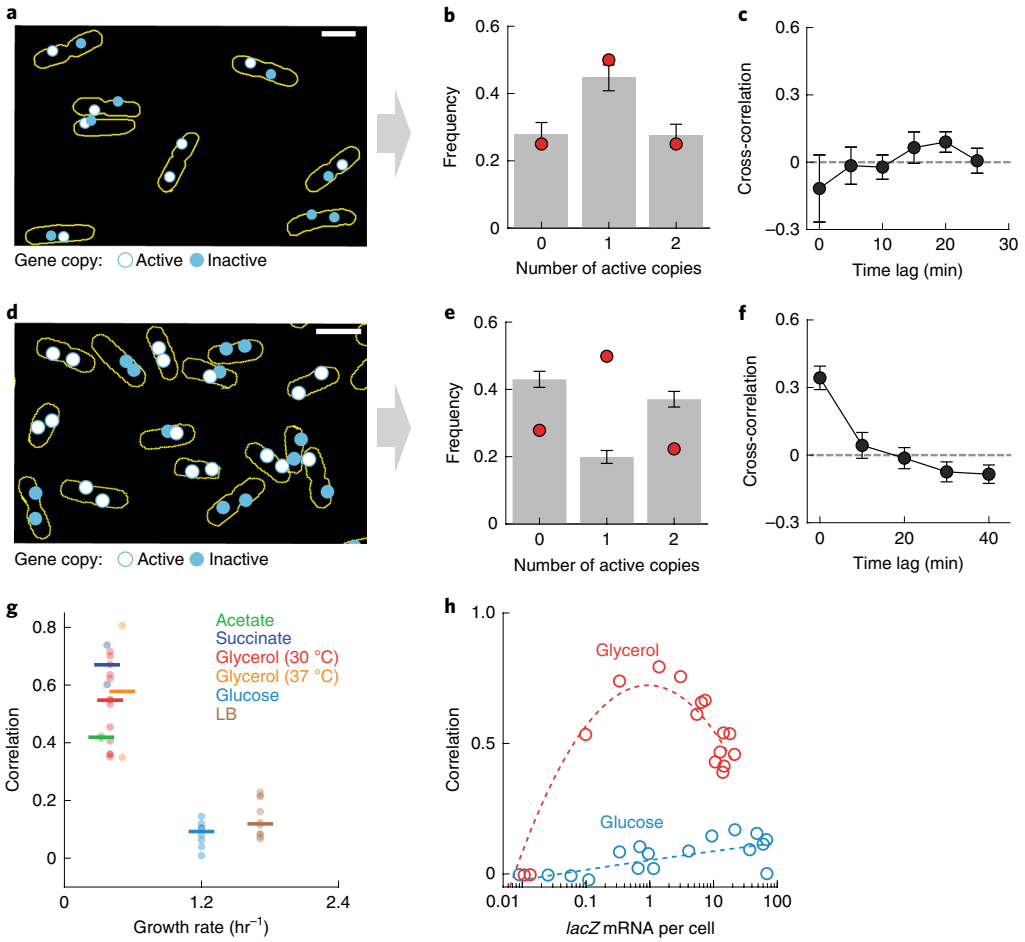

Fig. 3 | Promoter activity is coupled to the activity of additional gene copies in the cell. a, The transcriptional activity of individual copies of P_{lac} in cells grown in glucose (37 °C, 1 mM cAMP and 100 μM IPTG). For each cell that had two endogenous sister *lac* copies, the position and activity of each copy is overlaid on the cell boundary (yellow), as in Fig. 1b above (*lacZ* mRNA is not shown). Scale bar, 2 μm. **b**, The distribution of the number of active P_{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_{lac}–48MS2bs 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_{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 glycerol, 30 °C in succinate and 37 °C in acetate. **h**, The correlation between sister copies of P_{lac} as a function of expression level (total *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.

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_{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[4], promoter activity can be phenomenologically described using stochastic two-state kinetics, as previously concluded from whole-cell mRNA measurements[2]. 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 *lacZ* mRNA during P_{lac} induction, in both glucose and glycerol (Fig. 2f,g 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 *lacZ* at different steady-state levels (Supplementary Fig. 22) allowed us to identify that, upon P_{lac} induction, the probabilistic rate of promoter activation, k_{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[16,22,23].

As part of the replication cycle of the bacterial chromosome, multiple copies of the same gene are often present in the same cell[24,25]. The stochastic activity of these individual copies may be correlated due, for example, to fluctuations in an upstream regulator[26]. This effect was inferred from the presence of so-called 'extrinsic noise' in mRNA and protein expression[3,26], 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_{lac} activity in cells grown in glucose, under induction conditions where the fraction of transcriptionally active copies (p_{on}) was approximately half. Specifically, we examined the subpopulation of cells with exactly two sister copies of the *lac* locus. We found that the fractions of cells that had 0, 1 and 2 transcriptionally active

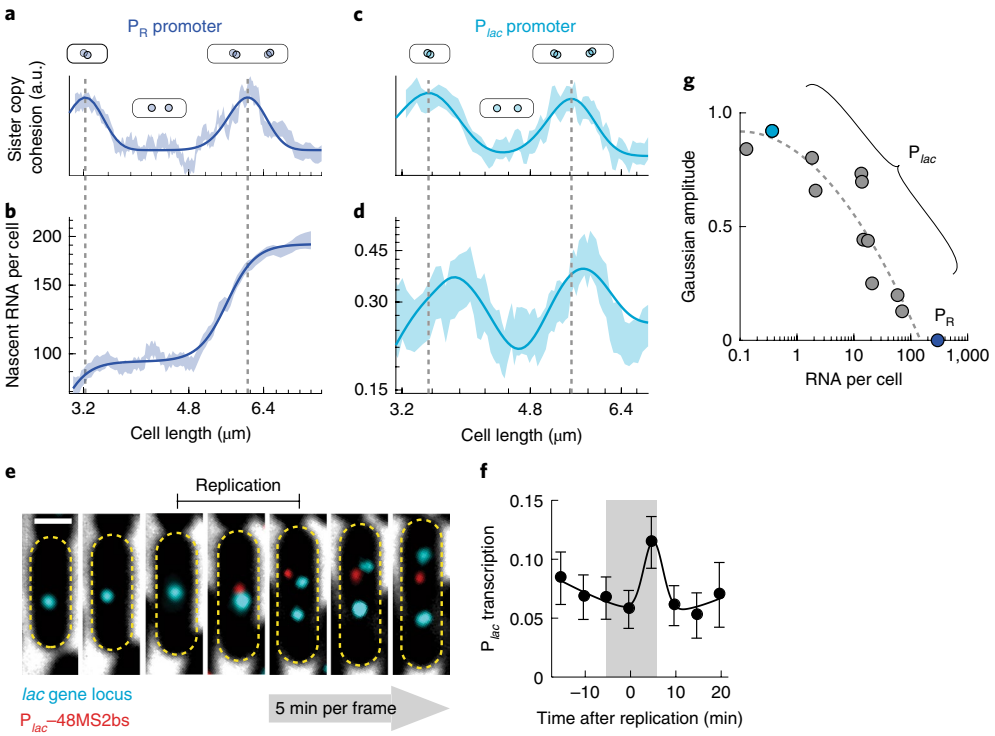

Fig. 4 | Promoter activity is coupled to the cell-cycle event of gene replication. a, The intensity of individual TetR-YFP (*hyc* gene locus) foci, which serves as a proxy for sister-copy cohesion (see schematics above the plot), as a function of cell length. The shading represents binned data (mean ± s.e.m., cells grown in LB, 30 °C). The solid line shows the fit to a sum of two Gaussians corresponding to consecutive gene-replication events. The vertical dashed lines are the Gaussian centres, indicating the estimated timing of gene replication. **b**, Nascent RNA per cell from the P_R promoter as a function of cell length. The data are from the same cells as in **a**. A P_R-24MS2bs reporter was placed near the *hyc* locus and nascent RNA levels were measured using smFISH against the MS2bs sequence. The shading represents binned data and the solid line shows the fit to a sum of two Hill functions corresponding to the consecutive doublings of gene dosage. **c**, The same as **a**, but for the *lac* locus (cells grown in LB with 10 μM IPTG). **d**, Nascent *lacZ* mRNA per cell from the P_{lac} promoter as a function of cell length. Data are from the same cells as in **c**. The shading represents binned data and the solid line shows the fit to a sum of two Hill functions and two Gaussians located at the Hill transition points, corresponding to a pulse of transcription around the time of gene replication. **e**, Time-lapse images demonstrating the occurrence of P_{lac} transcription close to the time of gene replication. Cells were grown in LB with no IPTG. The horizontal black bar indicates the uncertainty in estimating replication time due to the frame rate and sister cohesion. Yellow dashed lines indicate the cell boundaries. Scale bar, 1 μm (all frames). **f**, The probability of P_{lac} transcription (appearance of P_{lac}–48MS2bs signal) as a function of time relative to 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. **g**, The effect of gene replication on P_{lac} transcription as a function of the expression level. 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.

copies followed a binomial distribution, as would be expected if each P_{lac} 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_{lac} MS2 reporter also revealed a very low level of temporal correlation between two promoter copies within the same cell (Fig. 3c).

Repeating the same analysis for cells grown in glycerol (again at $p_{on} \approx 0.5$) yielded dramatically different results: two copies of P_{lac} within the same cell were highly correlated in their activity, as indicated by the distribution of number of active copies (Fig. 3d,e and Supplementary Fig. 24; $r = 0.58 \pm 0.11$), the correlation in nascent mRNA levels between the two copies (Supplementary Fig. 25) and the temporal correlation in live cells (Fig. 3f). Cells in other slow-growth media also exhibited high correlation between sister P_{lac} copies (Fig. 3g and Supplementary Fig. 26). The lambda P_R promoter showed similar dependence on the growth conditions, with two promoter copies having higher correlation in a slower growth medium (Supplementary Fig. 27). 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 Fig. 28) and genomic location (Supplementary Fig. 26). 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 untranslated 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 two unseparated sister loci[10] (Fig. 4a and Supplementary Fig. 30). In fixed samples, we used cell length as a proxy for the progression of the cell cycle[25]. Gene replication took place at well-defined cell-length intervals[25] (Fig. 4a) and the cell length at gene replication exhibited the expected dependence on genomic position (Supplementary Fig. 31). In the same length-sorted cells, we then measured the total amount of nascent RNA from the promoter, reflecting the transcriptional activity at a given cell-cycle phase. The results for P_R, considered a strong 'constitutive' promoter[26], revealed that transcription closely follows the gene dosage (Fig. 4b). Thus,

the transcriptional activity of each P_R 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[27]. Replication-induced transcription could stem, for example, from the transient displacement of a repressor by the replication fork[28]. To test this intriguing hypothesis, we examined the cell-cycle dependence of P_{lac} 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 P_{lac} 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 P_{lac} 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"[1,29,30].

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 gene locus of interest was labelled using FROS[31]. 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 (140*tetO*) were placed inside the *mhpA* gene, ~3.5 kb from *lac* (ref. [10]). Using this construct, we then built a series of strains where the 140*tetO* array is placed at different genomic loci.

We first built the plasmid pJZ087, which was used to place the 140*tetO* array at the gene locus of interest. The plasmid carries the 140*tetO* array and a single Flp recognition target (*FRT*) site that allows the integration of the whole plasmid into strains from the Keio collection of single gene deletions[32] using *FRT*-Flp recombination[33]. To construct pJZ087, we amplified a fragment containing a single *FRT* site, the R6K replication origin (*pir* dependent replication, for eliminating the plasmid after Flp-*FRT* recombination) and a kanamycin resistance cassette (Kan^R). This fragment was amplified from pKG137, a derivative of pCE37 (ref. [33]), using the primers FRT-FROS-FP-H1 and FRT-FROS-RP-H2, and then recombineered into plasmid pBH23 (ref. [10]) using standard protocols[34]. pBH23 carries the 140*tetO* array, with a gentamycin resistance cassette (Gen^R) 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 pCP20 (ref. [35]). The 140*tetO* array was then moved to a clean genetic background using P1 transduction[34]. 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.

To detect promoter activity in live cells, an array of MS2bs was placed under the control of the promoter of interest. The transcribed MS2bs formed stem-loops, which were specifically bound by the MS2 coat protein, fused to a fluorescent protein[2,13].

We first constructed a series of plasmids carrying 24 or 48 MS2bs under the control of different promoters. We started with pIG-48bs and pIG-24bs, which carry 48 and 24 MS2bs, respectively, under the control of P_{lac} (ref. [13]). Next, pBS48bs-Cm^R and pBS24bs-Cm^R were constructed by replacing the ampicillin resistance cassette (Amp^R) in pIG-48bs and pIG-24bs with a chloramphenicol resistance cassette (Cm^R), using recombineering. Cm^R was amplified from pKD3 (ref. [36]) using primers OC201-2 and OC202.

pJZ054 was constructed by replacing P_{lac} in pBS24bs-Cm^R with the phage lambda promoter P_R. The P_R sequence was amplified from wild-type phage lambda using the primers PR-short-FP and PR-short-RP, cut by NotI and PciI and ligated into pBS24bs-Cm^R, which was cleaved by the same enzymes.

To place each MS2 reporter cassette (promoter-MS2bs-Cm^R) 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 by expressing 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 140*tetO* array in the chromosome, using P1 transduction. Plasmid pJZ133 was transformed into the dual-reporter 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[10,13]. 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[37].

To construct plasmid pJZ133, we amplified the *tetR-yfp* fragment from pDM21 using the primers TetR-YFP-GG-FP and TetR-YFP-GG-RP. The backbone of pSR67.1 (containing P_{J23117}, 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 BsaI and ligated using Golden Gate assembly[38]. 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 amplified the *ms2-gfp* fragment from pIG-K133 using the primers MS2-GFP-GG-FP3 and MS2-GFP-GG-FP3. The backbone of pSR67.3 (containing P_{J23105}, BBa_J23105 and 623 arbitrary Anderson promoter units) was amplified using the primers pSR67-GG-FP2 and pSR67-GG-RP3. The two fragments were digested using BsmBI and ligated using Golden Gate assembly.

We next combined *tetR-mCherry* and *ms2-gfp* in a single plasmid. pJZ152 was constructed in the following way. The P_{J23105}-*ms2-gfp* fragment was amplified from pJZ107 using the primers MS2-TetR-GG-FP2 and MS2-GFP-GG-RP3. The backbone of pJZ102 (containing P_{J23117}-*tetR-mCherry*) was amplified using the primers pSR67-D-GG-FP and pSR67-D-GG-RP. The two fragments were digested using BsmBI and ligated using Golden Gate assembly. pJZ156 was constructed in a similar way, but with the two proteins placed in the reverse order. P_{J23117}-*tetR-mCherry* was amplified from pJZ102 using the primers MS2-TetR-GG-FP2 and TetR-mCherry-GG-RP2 and ligated to the backbone of pJZ107 (containing P_{J23105}-*ms2-gfp*).

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 (0.07 relative strength to BBa_0034). A fragment containing P_{J23117} and BBa_0031 was synthesized as double-stranded DNA (gBlocks, IDT) with flanking BsmBI cut sites. This fragment was then digested with BsmBI and ligated with the backbone of pJZ156, which was amplified using the primers P9 and P10 and digested with the same enzyme.

pJZ416 was constructed in a similar manner to pJZ186, except that *tetR-mCherry* was placed upstream of *ms2-gfp*. The same fragment (used above) containing P_{J23117} and BBa_0031 was digested with BsmBI and ligated with the backbone of pJZ152, which was amplified using the primers P9 and P11 and digested with the same enzyme.

Growth media and conditions. Five different growth media were used in this study: (1) LB (1 l of medium contained 10 g tryptone (BD Biosciences), 5 g yeast extract (BD Biosciences), 5 g NaCl (Fisher Scientific) and 1 μM NaOH (Fisher Scientific)); (2) glucose (Minimal M9CA broth (Teknova)); (3) glycerol

(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% acetate (Sigma-Aldrich)).

Cultures from fresh colonies were grown overnight (14–16 h) with antibiotics when appropriate: 100 μg ml^{-1} ampicillin (Fisher Scientific), 50 μg ml^{-1} kanamycin (Fisher Scientific), 17 μg ml^{-1} chloramphenicol (Fisher Scientific), 5.5 μg ml^{-1} gentamicin (Sigma-Aldrich), 100 μg ml^{-1} spectinomycin (Fisher Scientific). Depending on the growth conditions of the overday cultures, we used two different overnight culture set-ups. If the overday condition was LB, the overnight cultures were grown in LB as well; otherwise, the overnight cultures were grown in glucose. If the overday condition was glycerol, succinate or acetate, the overnight (glucose) culture was diluted at least 1:800. The overday 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 overday 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 $(OD)_{600} \approx 0.2$. Each sample was then treated according to the procedures 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 $OD_{600} \approx 0.2$. 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 smFISH 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 $OD_{600} \approx 0.2$. Rifampicin (Fisher Scientific) was added to a final concentration of 500 μg ml^{-1} (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 $OD_{600} \approx 0.2$–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 $OD_{600} \approx 0.2$–0.4. Cells were then prepared for imaging according to the procedures described below.

smFISH. The smFISH protocol was described in detail previously[11]. 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. [11]. (1) A final concentration of 1% formaldehyde was used for cell fixation. (2) For steady-state experiments, we added a washing step between cell harvesting and fixation. Following harvesting (centrifuging at 4,500*g* for 5 min), the cell pellets were resuspended in 1 ml 1 × PBS, then centrifuged at 4,500*g* for 1 min. This washing step is meant to ensure the proper pH, since YFP is pH sensitive[40]. (3) 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 r.p.m. 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 1 × 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 fixation; cells were then washed twice in 1 ml 1 × PBS and prepared for imaging as described below. With centrifugation: cells were harvested by centrifuging 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.

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 switching and flow rate settings were controlled using the CellASIC ONIX FG software. Cells were then imaged as described below.

Microscopy. We used an inverted epifluorescence microscope (Eclipse T*i*, Nikon), equipped with motorized stage control (ProScan III, Prior Scientific), a universal specimen holder, a mercury lamp (Intensilight C-HGFIE, 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 fixed and live cells. 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 '*z* slice'. Images were acquired at multiple slide 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 slide positions (typically starting with 50–100 cells per position, 5–10 positions). Depending on the growth conditions, videos were acquired at a frame rate of 5 or 10 min.

Cell recognition and lineage tracking. We used Schnitzcell[41] to identify cells in the phase-contrast channel of snapshots of live and fixed cells, as well as time-lapse videos. In every image stack, the *z* slice with the largest 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 Spätzcell[11] 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 that correspond to the same spot. For each spot, the in-focus plane 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 á trous wavelet three-plane decomposition and obtained the second wavelet plane[28,42]. 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 spots identified in FROS and MS2 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. [11]. Briefly, after discarding false positive spots, we first examined 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[5]. 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 inseparable[10,43]. 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 twice that of the dimmer population (Supplementary Fig. 30a). The fraction of twice-as-bright spots (10–20%) was consistent with the reported duration of sister-copy cohesion[10,44,45]. In time-lapse videos, we often observed the FROS spot intensity increasing before the spot split in two (Supplementary Fig. 30b), 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 colocalized 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. [46]. Briefly, we imaged fluorescent beads (TetraSpeck, Fisher Scientific) using the same imaging parameters as the sample slides. In each imaged channel, the spots (individual beads) were identified and localized using Spätzcell. The measured 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 P_{lac}, 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 similar 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 at two gene copies. To assess how this population affects the distribution of nascent RNA per gene copy, we performed the following calculation. $P_{nas}(m)$ 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: $P_{nas,mix}(m) = (1-c)P_{nas}(m) + c(P_{nas}(m) * P_{nas}(m))$ (where * represents the convolution operator). To evaluate the difference between $P_{nas}(m)$ and $P_{nas,mix}(m)$, 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 colocalization 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. Thus, each gene copy 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_{on} (data not shown).

The analysis of time-lapse videos is complicated by the need to keep track of 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. 30b). 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:

$$r = \frac{\langle ij \rangle - \langle i \rangle \langle j \rangle}{\sigma_i \sigma_j} \quad (1)$$

where i and j 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. [26] (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:

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

where $i(t)$ and $j(t)$ are the activities of the two gene copies in the same cell at time (t), N is the number of time points in the time series and τ is the lag time. In practice, we first used the built-in MATLAB cross-correlation function $xcorr(i(t) - \overline{i(t)}, j(t) - \overline{j(t)}$, 'biased') and then normalized the cross-correlation output by the standard deviations of both time series, to obtain $C(\tau)$. Because there is no natural 'order' between the two gene copies, we averaged the cross-correlation values of $C(\tau)$ 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. 4a,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 replication[10]. 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):

$$y = u_1 e^{-v_1(x-a)^2} + u_2 e^{-v_2(x-b)^2} + w_1 x^2 + w_2 x + w_3 \quad (3)$$

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 transcriptional 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

unrepressed 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:

$$y = c\left(1 + \frac{1}{1+\left(\frac{a}{x}\right)^{k_1}} + \frac{2}{1+\left(\frac{n_2 a}{x}\right)^{k_1}}\right) \quad (4)$$

The parameter n_2 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. [25]) 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_{lac}; 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:

$$y = c\left(1 + \frac{1}{1+\left(\frac{a}{x}\right)^{k_1}} + \frac{2}{1+\left(\frac{n_2 a}{x}\right)^{k_2}}\right) + u_1 e^{-v(x-a)^2} + u_2 e^{-v(x-n_2 a)^2} \quad (5)$$

The magnitudes of the Gaussian functions (u_1, u_2) 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 schnitz 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 alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* **15**, 1263–1271 (2008).
10. Joshi, M. C. et al. *Escherichia coli* sister chromosome separation includes an abrupt global transition with concomitant release of late-splitting intersister snaps. *Proc. Natl Acad. Sci. USA* **108**, 2765–2770 (2011).
11. Skinner, S. O. et al. Measuring mRNA copy number in individual *Escherichia coli* cells using single-molecule fluorescent in situ hybridization. *Nat. Protoc.* **8**, 1100–1113 (2013).
12. French, S. L. & Miller, O. L. Jr. Transcription mapping of the *Escherichia coli* chromosome by electron microscopy. *J. Bacteriol.* **171**, 4207–4216 (1989).
13. Golding, I. & Cox, E. C. RNA dynamics in live *Escherichia coli* cells. *Proc. Natl Acad. Sci. USA* **101**, 11310–11315 (2004).
14. Elf, J., Li, G. W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
15. Choubey, S., Kondev, J. & Sanchez, A. Deciphering transcriptional dynamics in vivo by counting nascent RNA molecules. *PLoS Comput. Biol.* **11**, e1004345 (2015).
16. Xu, H. et al. Stochastic kinetics of nascent RNA. *Phys. Rev. Lett.* **117**, 128101 (2016).
17. Chen, H. et al. Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*. *Mol. Syst. Biol.* **11**, 781 (2015).
18. Kuhlman, T. et al. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **104**, 6043–6048 (2007).
19. Moffitt, J. R. et al. Spatial organization shapes the turnover of a bacterial transcriptome. *eLife* **5**, e13065 (2016).
20. Proshkin, S. et al. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504–508 (2010).
21. Epshtein, V. & Nudler, E. Cooperation between RNA polymerase molecules in transcription elongation. *Science* **300**, 801–805 (2003).
22. Senecal, A. et al. Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep.* **8**, 75–83 (2014).
23. Skinner, S. O. et al. Single-cell analysis of transcription kinetics across the cell cycle. *eLife* **5**, e12175 (2016).
24. Neidhardt, F. C., Ingraham, J. L. & Schaechter, M. *Physiology of the Bacterial Cell: a Molecular Approach* (Sinauer Associates, 1990).
25. Wallden, M. et al. The synchronization of replication and division cycles in individual *E. coli* cells. *Cell* **166**, 729–739 (2016).
26. Elowitz, M. B. et al. Stochastic gene expression in a single cell. *Science* **297**, 1183–1186 (2002).
27. Guptasarma, P. Does replication-induced transcription regulate synthesis of the myriad low copy number proteins of *Escherichia coli*? *BioEssays* **17**, 987–997 (1995).
28. 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).
29. Golding, I. Infection by bacteriophage lambda: an evolving paradigm for cellular individuality. *Curr. Opin. Microbiol.* **43**, 9–13 (2017).
30. 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).
31. Lau, I. F. et al. Spatial and temporal organization of replicating *Escherichia coli* chromosomes. *Mol. Microbiol.* **49**, 731–743 (2003).
32. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
33. Ellermeier, C. D., Janakiraman, A. & Slauch, J. M. Construction of targeted single copy *lac* fusions using λ Red and FLP-mediated site-specific recombination in bacteria. *Gene* **290**, 153–161 (2002).
34. Thomason, L. et al. Recombineering: genetic engineering in bacteria using homologous recombination. *Curr. Protoc. Mol. Biol.* **78**, 1.16.1–1.16.24 (2007).
35. 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. *Gene* **158**, 9–14 (1995).
36. 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).
37. Schmidl, S. R. et al. Refactoring and optimization of light-switchable *Escherichia coli* two-component systems. *ACS Synth. Biol.* **3**, 820–831 (2014).
38. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PloS ONE* **3**, e3647 (2008).
39. Bernstein, 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).
40. 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).
41. Young, J. W. et al. Measuring single-cell gene expression dynamics in bacteria using fluorescence time-lapse microscopy. *Nat. Protoc.* **7**, 80–88 (2012).
42. Olivo-Marin, J.-C. Extraction of spots in biological images using multiscale products. *Pattern Recognit.* **35**, 1989–1996 (2002).
43. 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).
44. Nielsen, H. J. et al. Progressive segregation of the *Escherichia coli* chromosome. *Mol. Microbiol.* **61**, 383–393 (2006).
45. Nielsen, H. J. et al. Dynamics of *Escherichia coli* chromosome segregation during multifork replication. *J. Bacteriol.* **189**, 8660–8666 (2007).
46. 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. Elf, H. Garcia, M. Girard, J. Halliday, C. Herman, M. Joshi, D. Magnan, J. Moffitt, E. Nudler, R. Phillips, A. Sarrion-Perdigones, S. Sebastian, L. Sepúlveda, A. Singh, P. Sivaramakrishnan, S. Skinner, 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 1147498, 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.

nature research

Corresponding author(s): Ido Golding

Last updated by author(s): Jul 26, 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.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	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