

Time-resolved imaging-based CRISPRi screening

Daniel Camsund^{1,3}, Michael J. Lawson^{1,2,3*}, Jimmy Larsson¹, Daniel Jones¹, Spartak Zikrin¹, David Fange¹ and Johan Elf^{1*}

Our ability to connect genotypic variation to biologically important phenotypes has been seriously limited by the gap between live-cell microscopy and library-scale genomic engineering. Here, we show how in situ genotyping of a library of strains after time-lapse imaging in a microfluidic device overcomes this problem. We determine how 235 different CRISPR interference knockdowns impact the coordination of the replication and division cycles of *Escherichia coli* by monitoring the location of replication forks throughout on average >500 cell cycles per knockdown. Subsequent in situ genotyping allows us to map each phenotype distribution to a specific genetic perturbation to determine which genes are important for cell cycle control. The single-cell time-resolved assay allows us to determine the distribution of single-cell growth rates, cell division sizes and replication initiation volumes. The technology presented in this study enables genome-scale screens of most live-cell microscopy assays.

The last decade has shown remarkable development in genome engineering, fronted by applications of Cas9-mediated gene targeting^{1,2}. In combination with inexpensive large-scale DNA oligonucleotide synthesis, these techniques make it possible to generate pool-synthesized cell strain libraries with specific perturbations genome-wide^{3,4}. More recently, methods have been developed for screening Cas9 genome-edited libraries by sorting the library members based on the expression of a fluorescent protein and then sequencing cells with similar expression level⁵ or using single-cell RNA sequencing on libraries of CRISPR interference (CRISPRi) perturbations to determine the state of the transcriptome for individual cells^{6–8}. These methods are, however, blind to cellular dynamics and intracellular localization of relevant molecules and are thus greatly limited in the types of phenotypes that can be explored.

The progress in genome-scale engineering and expression perturbation has been accompanied by equally impressive developments in microscopy and microfabrication, which enable characterization of complex phenotypes at high temporal and spatial resolution in living cells under well-controlled conditions^{9–15}. While the power of these methods enables deep insight into cellular biophysics, the limitation of working with one strain at a time prohibits studying the impact of genes whose function is not already, at least to some degree, known. Given the rapid progress within the previously separate fields of imaging and genomics, the lack of efficient techniques for time-resolved single-cell phenotyping of pool-synthesized genetic strain libraries constitutes a severe bottleneck in biological research. We have recently proposed a tentative solution to the problem using a microfluidic method¹⁶ that we now demonstrate scaled to hundreds of genes for tens of thousands of single cells, converting the concept into a practical tool. A related method that allows imaging of a large library of bacterial cells adhered to a coverslip was published¹⁷ at the same time as our recent study¹⁶. Here, we use the microfluidic method to perform a large-scale screen of complex phenotypes to identify the regulatory elements of replication–division coordination in bacterial cells.

Results

Overview of the method. The heart of our method is a microfluidic device that enables both high-resolution dynamic phenotyping and

subsequent in situ genotyping of the individual strains. The microfluidics approach allows us to keep the bacteria in a constant state of exponential growth over hundreds of generations while imaging them at high resolution. The fluidic device¹³ is an adaption of the ‘mother machine chip’¹⁵, where each cell trap has a 300 nm constriction at the end that enables fast loading, as well as media and probe exchange (see the left side of Fig. 1 and Supplementary Fig. 1 for a schematic of the chip). Each strain occupies a defined space in the fluidic chip, but the genotypes of the different strains are unknown at the time of phenotyping. After the phenotypes have been determined, the cells are chemically fixed in the chip, and the genotypes are optically inferred by sequential fluorescence in situ hybridization (FISH) to a barcode (Fig. 1). We refer to the technique, which we developed previously in a small proof-of-principle study¹⁶, as DuMPLING (dynamic μ -fluidic microscopy phenotyping of a library before in situ genotyping).

We used the DuMPLING technique to characterize the coordination of the replication and division cycle in *E. coli* by tracking the chromosome replication forks throughout the cell cycle in a CRISPRi library. Replication initiation in *E. coli* is triggered at a fixed volume per chromosome^{18,19} independent of growth rate, but the underlying molecular mechanism is largely unknown. In this work, replication initiation was studied directly by observing a strain with a chromosomally integrated *seqA-yfp* fusion. SeqA binds hemimethylated DNA in the wake of the replication machinery and can thus be used to track replication foci (Fig. 2). In addition, the cell size at division and replication initiation was determined using phase contrast imaging. By imaging hundreds of cell cycles for each CRISPRi perturbation, we aimed to identify which genes are involved in setting the accuracy of the initiation volume. This screen could not have been performed without monitoring the dynamics of replication initiation directly in individual cells.

Implementation and analysis. We constructed the library in a host strain of MG1655 a *seqA-yfp* fusion for tracking the replication forks, dCas9 expressed from an anhydrotetracycline (aTc)-inducible promoter for CRISPRi and T7 RNA polymerase (T7 RNAPol) expressed from an arabinose-inducible promoter for barcode RNA expression. Each of the 235 library members expressed a unique

¹Department of Cell and Molecular Biology, SciLifeLab, Uppsala University, Uppsala, Sweden. ²Present address: Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. ³These authors contributed equally: Daniel Camsund, Michael J. Lawson.

*e-mail: lawsonjmichael@gmail.com; johan.elf@icm.uu.se

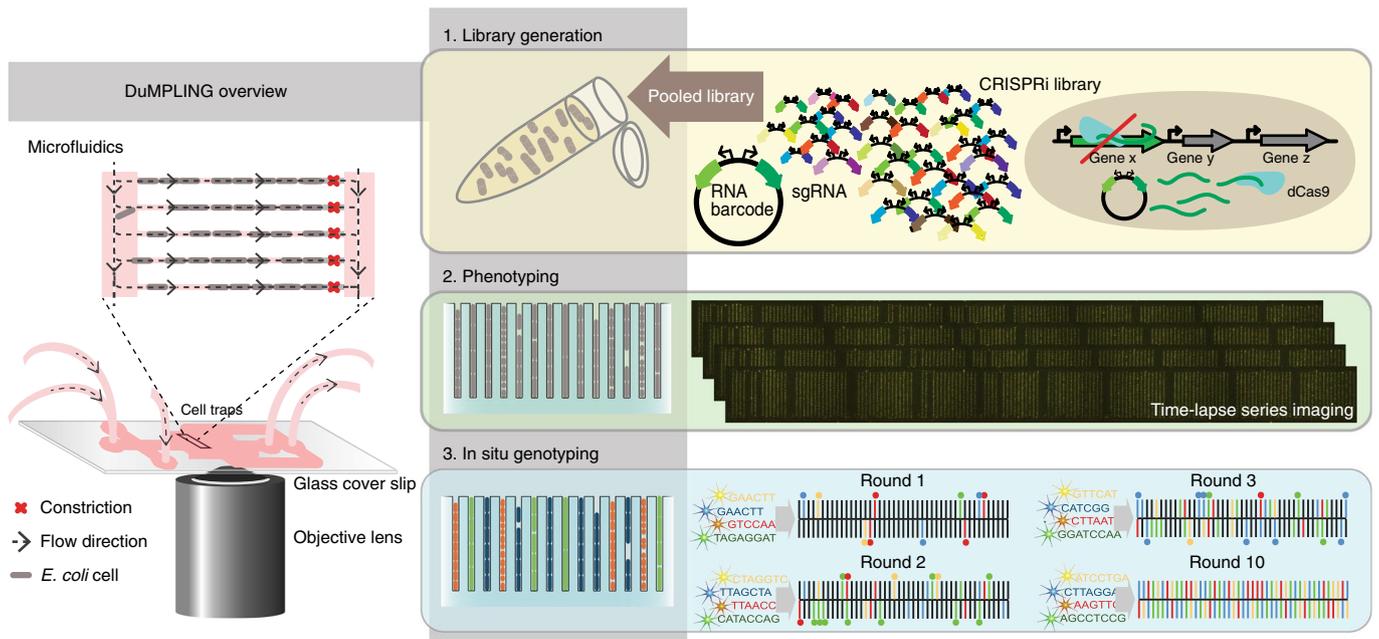


Fig. 1 | Assay workflow. Step 1, the bacteria in the CRISPRi library contain pool-synthesized plasmids each expressing a barcode and a corresponding sgRNA for repressing a specific gene in the *E. coli* chromosome. Step 2, the library of cells is loaded into a fluidic device where each strain occupies a spatially separated trap and where the cells can be monitored with highly sensitive time-lapse fluorescence microscopy for hours or days. Step 3, after phenotyping the cells are fixed, and the identities of the strains are revealed by sequential FISH probing for the barcodes.

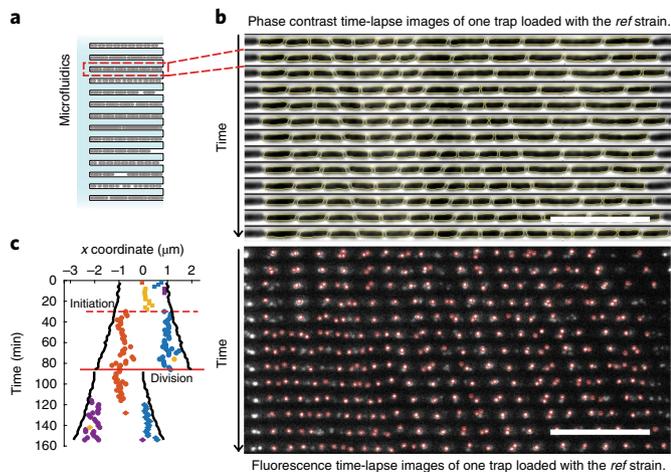


Fig. 2 | Analysis. **a**, Cartoon of cells growing in a microfluidic chip. **b**, Example kymographs in phase contrast (top) and fluorescence (bottom), with automated cell segmentation (yellow) or detection of SeqA-YFP clusters (red circles), respectively, overlaid. Each row corresponds to the same cell trap. The time increment is 2 min per row. Scale bar, 10 μm . Longer example of fluorescence imaging with cell and cluster detection is shown in Supplementary Video 1. **c**, Example of combined cell segmentation (cell poles in solid black lines) and SeqA-YFP cluster detection (all markers of different shapes and colors) for one cell and one of its daughters. SeqA-YFP clusters with the same marker color and shape belong to trajectories connected by u-track.

single-guide RNA (sgRNA), directing dCas9 to bind and repress a specific gene (see Methods section ‘Design of CRISPRi spacers’). The library included all known nonlethal cell cycle-related targets, as well as 38 *y*-genes²⁰, 28 of which are largely uncharacterized or have an unknown function. By inducing the dCas9 expression

with 1 $\mu\text{g}\mu\text{l}^{-1}$ aTc, the target genes were downregulated (see Supplementary Fig. 2 for burden on growth of dCas9 and sgRNA expression). In control experiments, 14 of 15 individually cloned sgRNA plasmids resulted in 4–100-fold repression (Supplementary Note 1). The remaining target, the gene for the transcription factor Rob, was hardly repressed at all, possibly due to its transcriptional autoregulation²¹. Each plasmid also encoded a 20-base pair (bp) barcode sequence that could be expressed as RNA by a T7 promoter. The barcode was uniquely coordinated with the sgRNA (see Methods section ‘Design and cloning of the DuMPLING library’) sequence to identify which sgRNA was expressed in which strain.

We loaded 40 pool-synthesized strains at a time into the microfluidic device to phenotype on average 562 generations per strain in an 8 h experiment. Before phenotyping, the bacteria were grown with dCas9 induced for 9 h to establish steady-state phenotypes. This time also allowed the mother cell at the end of the cell trap to divide enough times to set the genotype of the whole trap. To achieve sufficient time resolution, we imaged each position (30 traps per position) for 20 ms every minute in the phase contrast channel (example of time-lapse imaging in Supplementary Video 2) and for 300 ms every 2 min in the fluorescence channel (514 nm at 5.3 W cm^{-2}) (see example of time-lapse imaging in Supplementary Video 3). This limited us to 90 positions, or 2,700 cell traps per experiment. This also gives enough traps per strain to account for uneven representation of individual genotypes in the library (Supplementary Fig. 3). The limitation for how many strains we can analyze simultaneously is currently set by the number of traps we can image during phenotyping (for example, how fast we can image and move the stage) but not by how many strains can be made or genotyped in parallel.

After phenotyping, the barcode RNAs were expressed by T7 RNAPol induction and the cells were subsequently fixed and permeabilized. The strains were identified in situ by sequential FISH probing in four colors; each round of probing identifies the positions of four unique strains (see Supplementary Fig. 4 and the image analysis subsection on ‘Genotyping’ in Methods for details on quantifying genotyping fluorescence). Each round was completed in

less than 30 min, since no stripping of probes was required, and all strains could be identified within 6 h following fixation. In complementary experiments where we loaded all strains simultaneously, we used combinatorial FISH probing which can identify N^R genotypes in R rounds of N colors. This combinatorial approach requires stripping and rehybridizing probes to the same target, as opposed to the sequential FISH approach used here, in which each barcode RNA is targeted only once with no stripping step. The combination of the slow and incomplete stripping in the chip with the inability to use primary probes makes the combinatorial approach less practical (see Supplementary Note 2 for additional information about the advantages and drawbacks of different genotyping methods in the polydimethylsiloxane (PDMS) chip).

After connecting the genotypes to the phenotypes through their spatial positions, we analyzed how the repression of individual genes affected the cell size and growth rate (Fig. 2b, top) as well as coordination of replication with the division cycle (Fig. 2b, bottom). The kymograph for one of the 2,700 analyzed cell traps is displayed in Fig. 2c. The bacteria were segmented using the per object ellipse fit (POE) method²² and tracked by the Baxter algorithm²³. The SeqA-YFP foci were detected by a wavelet-based method²⁴ and the replication forks were tracked simultaneously through the generations using the u-track algorithm²⁵ (Fig. 2c). Each experiment, including 2,700 cell traps that were imaged every minute for 8 h, resulted in 220 Gb of image data and took 2 h to analyze on 45 cores using customized parallelized image analysis routines.

The impact of perturbations on the *E. coli* cell cycle. In Fig. 3a–c we show comparisons of the average growth rates and cell sizes at division and replication initiation, respectively, for the 215 strains for which we obtain data from a minimum of five independent cell traps and 40 complete cell cycles (statistics are available in Supplementary Table 1). The genotypes that are substantially different from the reference control strain (*ref*, see Methods section ‘Design and cloning of the DuMPLING library’ for strain details) in replicate experiments are indicated by the name of the sgRNA-targeted gene. The sizes at birth (or division) and initiation can get both bigger and smaller than *ref*, whereas growth rates typically only get smaller. The birth size is the area of the segmented cell (Fig. 2b) in the first frame after the division event, which in turn is defined by the Baxter algorithm. The initiation size is defined in the Methods (see Supplementary Fig. 5). The deviations from *ref* are mostly uncorrelated between the properties, with two notable exceptions: the *tol-pal* cluster is smaller and the *fis, diaA* cluster is larger in both initiation and birth sizes (Fig. 3a).

As a control for correct genotyping and cell segmentation, the average growth rates obtained from the single-cell time-lapse imaging compare well with the corresponding bulk experiment (Fig. 3d). To make bulk estimates of the growth rates, we performed a competition assay of the whole library in liquid culture and determined the time-dependent relative abundance of each genotype by next generation sequencing (NGS) (Fig. 3d, assay details in Supplementary Note 3). Additional controls are described in Supplementary Note 4. We compared the phenotypes of selected strains from the DuMPLING screen with those obtained with the corresponding knockouts or specific sgRNA knockdowns constructed and measured one strain at a time (Supplementary Note 4 and Supplementary Table 3). The good agreement in phenotypes implies that the genotyping is robust, although there were notable exceptions with off-target effects or where synthesis errors in the sgRNA coding sequence were selected for (Supplementary Note 4).

An information-rich way to simultaneously characterize replication and cell cycle processes in individual strains is the fork distribution plot, which shows the probability of finding a replication fork in a specific position in the cell (horizontal axis) for cells of

a given size (vertical axis). The distribution for the *ref* strain is shown in Fig. 3e. The data in the fork distribution plots can be used to estimate the average initiation volume (see Supplementary Fig. 5, the image analysis subsection on ‘Phenotyping’ in the Methods and ref.¹⁸). The good agreement between this bulk estimate and the average of estimates obtained from individual cells (see Fig. 3f for comparison and Supplementary Fig. 5 and the image analysis subsection on ‘Phenotyping’ in the Methods for estimation of a single-cell initiation size) acts as a control for our ability to accurately estimate initiation size for individual cell trajectories.

Figure 4 shows the fork distribution plots for all genotypes from one experiment that met the criteria described in the figure legend (for results from repeat experiments see Supplementary Figs. 6–11). Most of these are similar to the unperturbed *ref* (Fig. 4a), but a number of variant classes can be identified: Fig. 4b (*fis, dedD* and so on) represents strains with large division size and Fig. 4c (*clpP, nth, pal* and so on) represents strains with small division size. Correspondingly, Fig. 4d (*fis, diaA* and so on) represents strains with large initiation size and Fig. 4e (*tol, pal* and so on) represents strains with small initiation size. Figure 4f (*hda, ihf* and so on) represents strains for which the replication initiation size is undefined, suggesting that there may be important cell-to-cell variability in the initiation size and, consequently, that these genes are important for regulation of DNA replication. In this category, the *recF* and *skp* phenotypes may be due to collateral repression of *dnaA* and *dnaE*, respectively, which are located in the same operons (Supplementary Table 4). The strains *seqA, dam* and *damX* have few SeqA-YFP foci, which makes sense since *dam* encodes the DNA methylase and SeqA only binds hemimethylated DNA. *damX* is located upstream of *dam* in the same operon and its repression will also downregulate *dam*.

By pooling data from different cells of the same genotype, we lose information about cell-to-cell variation for the different genotypes. As a first-order description of the cell-to-cell variation, we plotted the coefficient of variation (CV) against the corresponding average in growth rate, birth size and initiation size for each genotype in Fig. 5a–c. Corresponding illustrative examples of the full distributions are also given for selected genotypes in Fig. 5d–f. For birth size, most perturbations give larger average as well as CV (Fig. 5b). Interestingly, *fis* and *dedD* do, however, increase the average size without an increase in CV. In a few cases, the CV increases much more than the change in the average. For example, *hda* repression gives an 80% increase in variation of the replication initiation size with only a 10% reduction in the average (Fig. 5c), suggesting an important regulatory role. With respect to cell-to-cell variation in growth rate, we observe a striking trend that gene knockdowns that reduce growth rate also give rise to large cell-to-cell variation (Fig. 5a). A plausible explanation is that the repression limits cell growth in only one factor and that stochastic fluctuations in this factor alone directly impact the growth rate. In general, the increase in variation due to CRISPRi may indicate that the repressed gene product has an active role in limiting the phenotypic fluctuations. Alternatively, the phenotype is highly sensitive to the gene product and thus increased relative fluctuations in the repressed gene product cause increased fluctuations in the phenotype.

Discussion

Circling back to the original question of which genes are important for triggering replication at a fixed cell size, a number of candidates stand out in Figs. 4g and 5c. If we exclude genes with well-documented and clearly unrelated functions, we end up with *hda, diaA, ihfA, ihfB, fis, yjgH* and *yeeS*. *diaA* is previously known for its direct interaction with DnaA at replication initiation²⁶. *hda, ihfA, ihfB* and *fis* are all associated with DnaA-ATP to ADP cycling²⁷. Further characterization of *yjgH* shows that the observed phenotype is due to an off-target effect of the sgRNA (Supplementary Note 4).

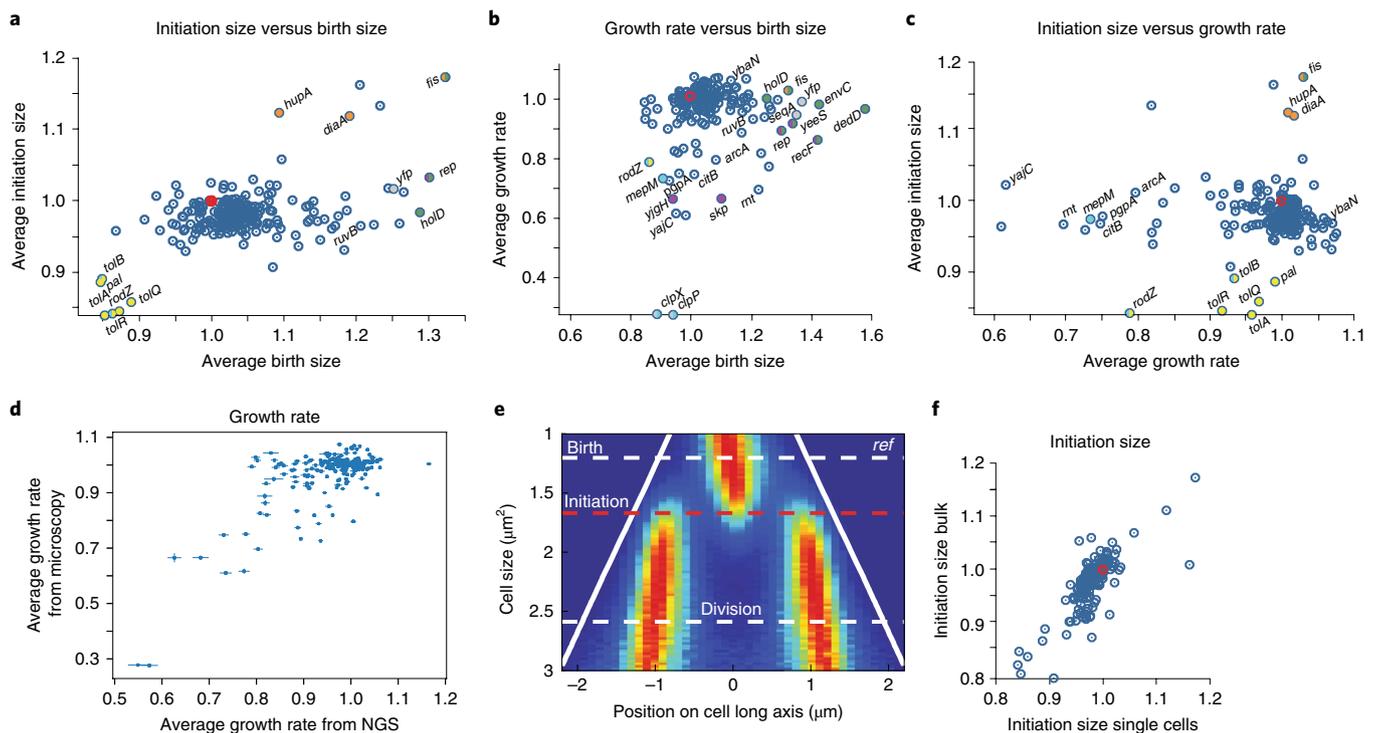


Fig. 3 | Phenotypic data averaged for each genotype. a–c, Two-dimensional plots in phenotype space, where CRISPRi knockdowns with notable deviation from the reference control strain (*ref*, red dot in **a–c** and **f**) are labeled by the name of the targeted gene. Outlier dots have been classified and labeled as follows: green, large birth size; cyan, small birth size; orange, large initiation size; yellow, small initiation size; gray, few detected foci; purple, undefinable initiation size. As these properties are not all mutually exclusive, dots may be multicolored. The numbers of data points used to estimate the averages are given in Supplementary Table 1. Correlations to replica experiments are described in Methods section ‘Reproducibility of the DuMPLING assay’. Horizontal, average normalized birth size; vertical, average normalized initiation size. **b**, Horizontal, average normalized birth size; vertical, average normalized growth rate. **c**, Horizontal, average normalized growth rate; vertical, average normalized initiation size. **d**, Control experiment showing the correlation between the relative growth rate derived from the DuMPLING experiments with that from a pooled competition assay (Pearson correlation = 0.7). The DuMPLING error bars correspond to the s.e.m.s normalized to the *ref* mean growth rate. For the NGS data, error bars denote the s.e.m.s as estimated from 1,000 bootstrap resamplings of the quality-controlled NGS sequence data. The number of sequence reads corresponding to each library member before and after competition is given in Supplementary Table 2. **e**, Fork distribution plot. Horizontal, SeqA-YFP cluster location along the long axis of the cell (from old cell pole to new); vertical, cell size, color indicates the probability of finding a replication fork at a given position along the cell axis and at a given cell size. Initiation size corresponds to the average of individually tracked replication forks. **f**, Average normalized initiation size fit to bulk replication forks (vertical) (Supplementary Fig. 5a) compared with the average of single-cell initiation size estimates (Supplementary Fig. 5b) (Pearson correlation = 0.78, $n = 197$ different genotypes, in which average initiation events were detected (image analysis subsection on ‘Phenotyping’ in the Methods)). The number of data points used to estimate each average is given in Supplementary Table 1.

The remaining uncharacterized gene that shows an evident perturbation in the accuracy of replication is *yeeS*. This gene deserves more specific study. We also note that downregulation of *pgsA* does not show a perturbation in replication initiation although the gene product is key in the synthesis of fatty acids, which some studies have implicated in DnaA-ADP to ATP conversion at the membrane²⁸. Thus, overall, our results support models for replication initiation control based on (1) DnaA-ATP to ADP cycling through the regulatory inactivation of DnaA (RIDA) mechanism mediated by Hda²⁹, (2) Fis and Ihf binding the DnaA reactivating sequences (DARS)³⁰ and (3) *datA* locus dependent DnaA-ATP hydrolysis (DDAH)³¹ regulated by Ihf.

We can envisage two types of mechanistic models that would trigger replication at a nearly fixed volume per chromosome and that, at least partially, can be mapped to genes identified in this study. In the first type of model, DnaA is converted from an ADP- to ATP-bound state at a rate proportional to the number of chromosomes through RIDA and DDAH and from ATP to ADP at a rate proportional to the cell volume; for example, by a saturated enzyme present at a constant concentration. This enzyme remains

to be discovered. Such a modification–demodification model can be made ultra-sensitive³² to the ratio of chromosomes and cell volume, but it is hard to rationalize the role of the DARS sites in this context. In the second type of model, the concentration of DnaA-ATP is kept constant such that the number of DnaA-ATP molecules is proportional to the volume of the cell. If, in addition, DnaA molecules are bound to high-affinity titration sites on the chromosome, that is, DnaA-boxes, the free DnaA-ATP concentration shoots up and replication initiates when the titration sites are filled; that is, at a fixed volume-to-titration site ratio. The importance of the DARS sites is equally mysterious in this model. Additional work is needed to precisely define the mechanistic details of replication initiation.

The DuMPLING approach, here scaled up from a proof of principle¹⁶ and deployed to identify the key regulators of the *E. coli* cell cycle, can be used to study all sorts of complex dynamic phenotypic traits that require sensitive or time-lapse microscopy for characterization. Plausible extensions include studies of gene expression dynamics in response to recoded promoter sequences, or other genetic regulatory elements; or live-cell enzymatic assays as a function of active site residue mutations. In eukaryotic

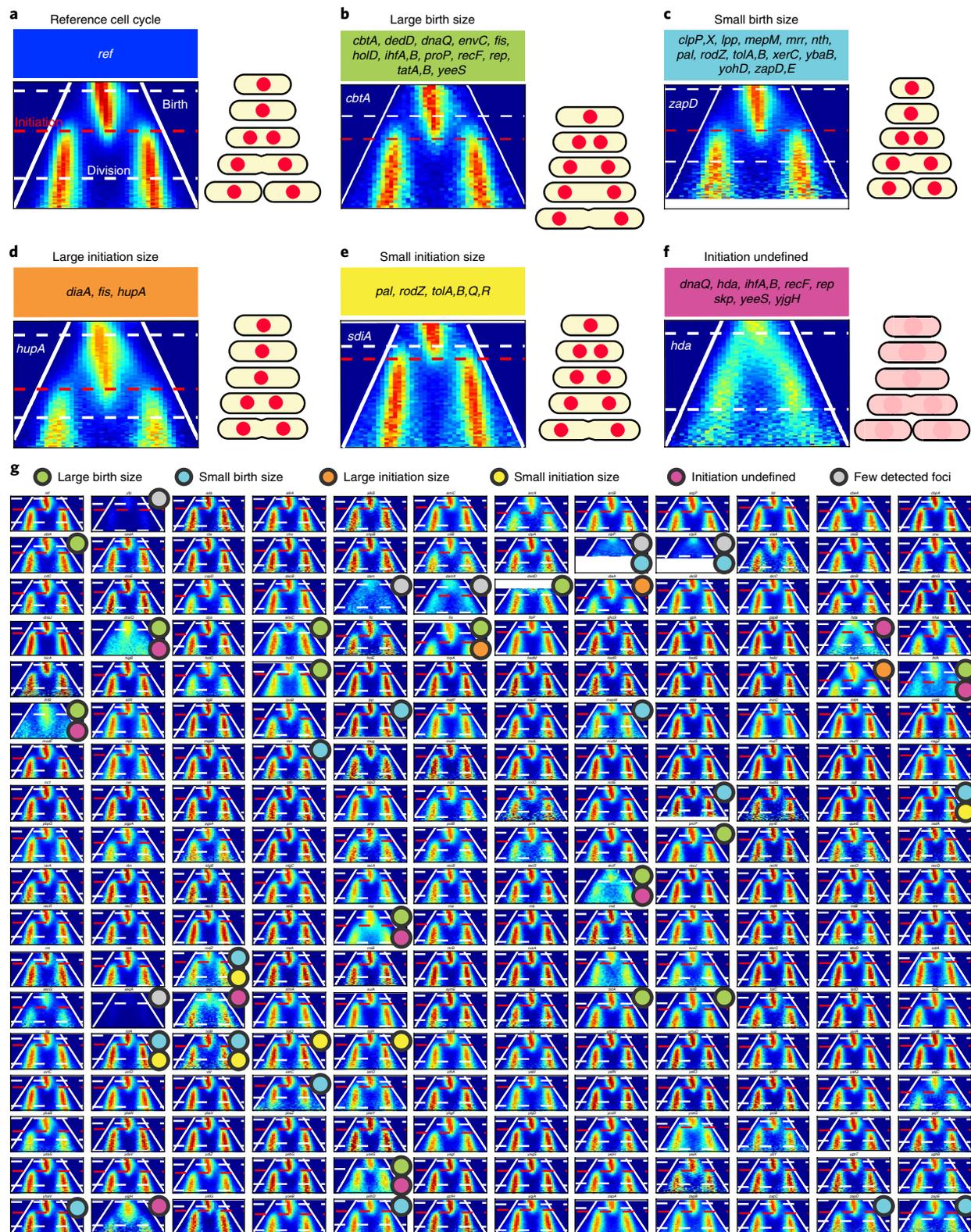


Fig. 4 | Fork distribution plots and time-resolved phenotypes. a, Example fork distribution plot for the unperturbed/control *ref* strain. White dashed lines, birth/division size; red dashed lines, replication initiation size, average of individually tracked replication forks; white solid lines, position of cell poles.

b–f, Examples of each classification in **g** and Figs. 3a–c and 5a–c (note, it is possible for a gene to belong to more than one classification). **b**, Large division size (only birth size in figure). **c**, Small division size (the plot is white at the bottom as cells with this phenotype do not reach the larger sizes of the figure and the area is left blank to keep all fork distribution plots on the same scale). **d**, Large initiation size. **e**, Small initiation size. **f**, Replication initiation size not definable. **g**, Plots for all knockdowns that had a minimum of five independent cell traps and 40 whole cell cycles. Differently colored dots indicate the classification of phenotypes where green, cyan, orange, yellow, purple and gray are large birth size (as in **b**), small birth size (as in **c**), large initiation size (as in **d**), small initiation size (as in **e**), initiation size undefined (as in **f**) and few detected foci, respectively. Fork distribution plots for a replica experiment are shown in Supplementary Figs. 6–11.

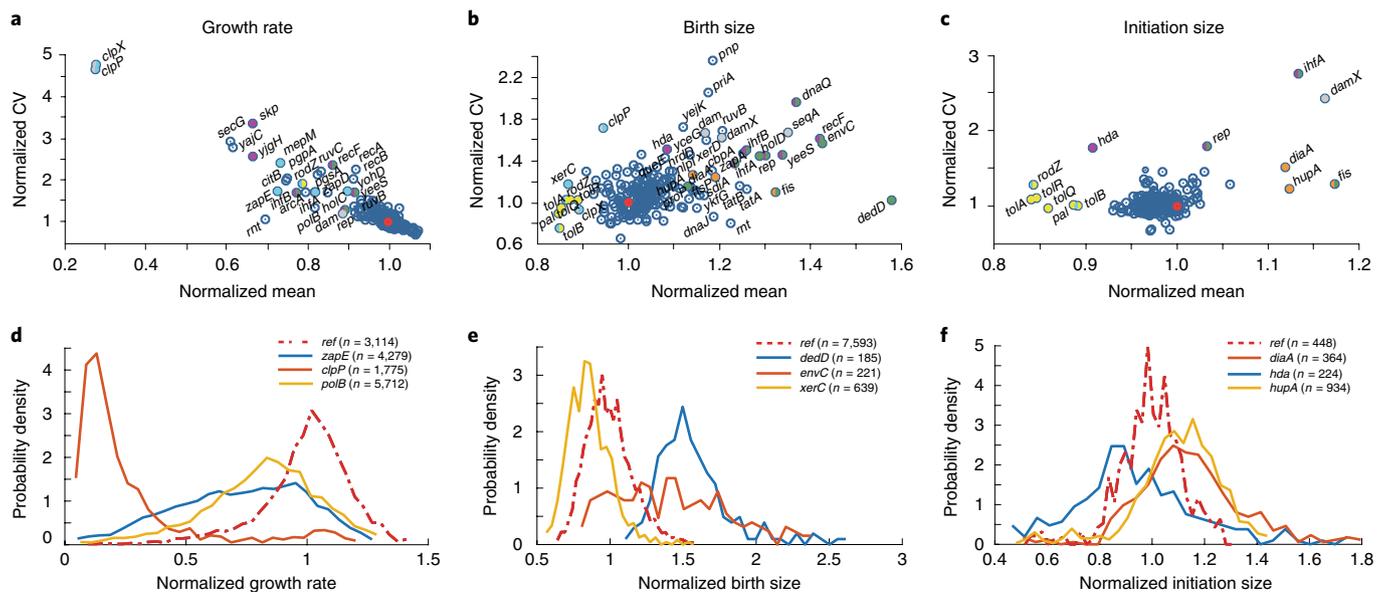


Fig. 5 | Cell-to-cell phenotypic variation for each genotype. **a–c**, Each plot has the CV on the vertical axis and mean on the horizontal axis, both normalized by *ref* (red dot at [1,1] in each plot), for growth rate (**a**), birth size (**b**) and initiation size (**c**). Outlier dots have been classified and color coded as in Fig. 4. The numbers of data points used to estimate the means and CVs in **a–c** are given in Supplementary Table 1. **d–f**, We show probability density (vertical axis) as a function of growth rate (**d**), birth size (**e**) and initiation size (**f**) for selected genotypes (*N*, number of cells). Examples of variation in single-cell phenotypes are shown in Supplementary Videos 4–7 for *ref*, *hda*, *dedD* and *clpP*, respectively. Comparative examples of cell-to-cell variation in specific, single-gene knockdown and knockout of *clpP* are shown in Supplementary Videos 8 and 9, respectively.

extensions, where spatial organization is critical, we also envisage interaction partner screens where the intracellular location of a labeled molecule of interest is altered in response to a library of gene knockdowns.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41592-019-0629-y>.

Received: 25 May 2019; Accepted: 4 October 2019;

Published online: 18 November 2019

References

- Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
- Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911 (2018).
- Garst, A. D. et al. Genome-wide mapping of mutations at single-nucleotide resolution for protein, metabolic and genome engineering. *Nat. Biotechnol.* **35**, 48–55 (2016).
- Wang, H. H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894–898 (2009).
- Rajagopal, N. et al. High-throughput mapping of regulatory DNA. *Nat. Biotechnol.* **34**, 167–174 (2016).
- Dixit, A. et al. Perturb-seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).
- Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882.e21 (2016).
- Jaitin, D. A. et al. Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896.e15 (2016).
- Liu, Z., Lavis, L. D. & Betzig, E. Imaging live-cell dynamics and structure at the single-molecule level. *Mol. Cell* **58**, 644–659 (2015).
- Li, N. et al. Single-molecule imaging and tracking of molecular dynamics in living cells. *Natl. Rev. Phys. Sci.* **4**, 739–760 (2017).
- Balzarotti, F. et al. Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes. *Science* **355**, 606–612 (2016).
- Liu, T.-L. et al. Observing the cell in its native state: imaging subcellular dynamics in multicellular organisms. *Science* **360**, eaaq1392 (2018).
- Balteski, Ö., Boucharin, A., Tano, E., Andersson, D. I. & Elf, J. Antibiotic susceptibility testing in less than 30 min using direct single-cell imaging. *Proc. Natl. Acad. Sci. USA* **114**, 9170–9175 (2017).
- 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).
- Wang, P. et al. Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103 (2010).
- Lawson, M. J. et al. In situ genotyping of a pooled strain library after characterizing complex phenotypes. *Mol. Syst. Biol.* **13**, 947 (2017).
- Emanuel, G., Moffitt, J. R. & Zhuang, X. High-throughput, image-based screening of pooled genetic-variant libraries. *Nat. Methods* **14**, 1159–1162 (2017).
- Wallden, M., Fange, D., Lundius, E. G., Balteski, Ö. & Elf, J. The synchronization of replication and division cycles in individual *E. coli* cells. *Cell* **166**, 729–739 (2016).
- Si, F. et al. Mechanistic origin of cell-size control and homeostasis in bacteria. *Curr. Biol.* **29**, 1760–1770 (2019).
- Ghatak, S., King, Z. A., Sastry, A. & Palsson, B. O. The γ -ome defines the 35% of *Escherichia coli* genes that lack experimental evidence of function. *Nucleic Acids Res.* **47**, 2446–2454 (2019).
- Schneiders, T. & Levy, S. B. MarA-mediated transcriptional repression of the *rob* promoter. *J. Biol. Chem.* **281**, 10049–10055 (2006).
- Ranefall, P., Sadanandan, S. K. & Wählby, C. Fast adaptive local thresholding based on ellipse fit. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)* 205–208 (2016).
- Magnusson, K. E. G., Jaldén, J., Gilbert, P. M. & Blau, H. M. Global linking of cell tracks using the Viterbi algorithm. *IEEE Trans. Med. Imaging* **34**, 911–929 (2015).
- Olivo-Marín, J.-C. Extraction of spots in biological images using multiscale products. *Pattern Recognit.* **35**, 1989–1996 (2002).
- Jaqaman, K. et al. Robust single-particle tracking in live-cell time-lapse sequences. *Nat. Methods* **5**, 695–702 (2008).
- Keyamura, K. et al. The interaction of DiaA and DnaA regulates the replication cycle in *E. coli* by directly promoting ATP DnaA-specific initiation complexes. *Genes Dev.* **21**, 2083–2099 (2007).
- Katayama, T., Kasho, K. & Kawakami, H. The DnaA cycle in *Escherichia coli*: activation, function and inactivation of the initiator protein. *Front. Microbiol.* **8**, 2496 (2017).

28. Saxena, R., Fingland, N., Patil, D., Sharma, A. K. & Crooke, E. Crosstalk between DnaA protein, the initiator of *Escherichia coli* chromosomal replication, and acidic phospholipids present in bacterial membranes. *Int. J. Mol. Sci.* **14**, 8517–8537 (2013).
 29. Camara, J. E. et al. Hda inactivation of DnaA is the predominant mechanism preventing hyperinitiation of *Escherichia coli* DNA replication. *EMBO Rep.* **6**, 736–741 (2005).
 30. Fujimitsu, K., Senriuchi, T. & Katayama, T. Specific genomic sequences of *E. coli* promote replicational initiation by directly reactivating ADP-DnaA. *Genes Dev.* **23**, 1221–1233 (2009).
 31. Kasho, K. & Katayama, T. DnaA binding locus *datA* promotes DnaA-ATP hydrolysis to enable cell cycle-coordinated replication initiation. *Proc. Natl Acad. Sci. USA* **110**, 936–941 (2013).
 32. Goldbeter, A. & Koshland, D. E. Jr. An amplified sensitivity arising from covalent modification in biological systems. *Proc. Natl Acad. Sci. USA* **78**, 6840–6844 (1981).
- 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 America, Inc. 2019

Methods

Strain construction and cloning. *Design and cloning of the DuMPLING test strain.* The DuMPLING recipient strain for studying CRISPRi perturbations of cell cycle-related genes (EL685, test strain) was made based on a previously published *E. coli* strain in which SeqA is expressed as a translational fusion to a yellow fluorescent protein (YFP)³⁵. To make the test strain, we used generalized P1 phage transduction to transfer an arabinose-inducible T7 RNAPol gene with a *tetA* selection marker in the *araB* locus from BL21-AI (Thermo Fisher Scientific) to the SeqA-YFP strain. Then, we repeated the transduction procedure to transfer a previously optimized low-leakiness aTc-inducible dCas9 gene with a spectinomycin resistance marker in *intC*¹⁶ to the combined SeqA-YFP and T7 RNAPol strain. This produced the *E. coli* MG1655 *INSseqA::yfp-cat INSaraB::T7 rnapol-tetA DELaraB INSintC::tetR-dcas9-aadA* test strain (EL685). The *seqA-yfp* locus and both transduced constructs were confirmed by colony PCR and by Sanger sequencing the flanks of the PCR products (*T7 rnapol-tetA*) or the complete constructs (*seqA-yfp* and *tetR-dcas9-aadA*).

Further, we tested the in vivo functionality of all three genetic constructs in EL685 as follows. First, the presence of SeqA-YFP foci in the test strain was microscopically confirmed using the setup described in the microscopy section below. Second, we transformed the pGuide-P1-*lacY* plasmid that carries the P1 barcode after the T7 promoter from ref.¹⁶ into our test strain (creating EL464) to confirm our ability to induce T7 RNAPol and express and detect RNA barcodes from T7 promoters. Induction of P1 barcode expression and genotyping using FISH with a truncated P1 probe complementary to the P1 barcode¹⁶ were performed in the same manner as in the DuMPLING screen. Finally, we tested the inducibility of our CRISPRi system. We used an sgRNA targeted against the lactose permease gene *lacY* and cultured cells in the presence of glucose or lactose, which should lead to inhibition of growth only in lactose media with aTc induction. To perform this assay, the growth of the test strain with the pGuide-P1-*lacY* plasmid (EL464), constitutively expressing an sgRNA against *lacY*, was compared with that of the test strain transformed with pGuide-empty (EL261), a negative control plasmid¹⁶. Three separate colonies each from the test strain with pGuide-P1-*lacY* or pGuide-empty were inoculated in lysogeny broth with 50 µg ml⁻¹ kanamycin (Km) and grown at 37 °C, shaking at 200 r.p.m. until in mid-exponential phase. Each replicate culture was diluted 1:400 into wells with 200 µl supplemented M9 (100 µM CaCl₂, 2 mM MgSO₄, 1× M9 salts, 1× RPMI amino acid mix (Sigma)) with 50 µg ml⁻¹ Km, either 0.4% (wt/vol) glucose or lactose, and either 0 or 100 µg µl⁻¹ aTc to induce CRISPRi, in a transparent 96-well plate with lid (Costar Assay Plate, REF 3370, Corning). The growth assay was carried out in an Infinite M200 plate reader (Tecan) at 37 °C, shaking for 60 s at 4.5 mm amplitude and measuring the absorbance at 600 nm (optical density 600) every 5 min. After medium background subtraction, the growth curves for each well were plotted (Supplementary Fig. 2) using a custom MATLAB script. There was an expected difference in growth rate between carbon sources; however, there was only a reduction of growth rate within a carbon source when cells were grown on lactose and aTc, with sgRNA expression from the pGuide-P1-*lacY* plasmid, confirming the function of the inducible dCas9 construct.

Design of CRISPRi spacers. Genes directly and indirectly connected to the *E. coli* cell cycle (that is, genes annotated with cell cycle-related gene ontology terms in the EcoCyc database³⁶, as well as genes connected to lipid, membrane, cytoskeleton or RNA modifications, or with unknown functions) were selected as targets for the CRISPRi library. However, genes annotated as essential in the EcoCyc database were excluded. One exception is the essential gene *dnaA*, the initiator of replication, which was added as a control. Further, the genes *holD*, *ihfA*, *rnt*, *ybeY* and *ydaS* were included in the library but later identified as essential³⁵. A spacer (*yfp*) against the specific *yfp* gene used for the SeqA-YFP fusion³⁵ was included as a control for CRISPRi of the SeqA-YFP expression. All selected genes are listed in Supplementary Table 5.

The spacers against the targeted genes were derived from an annotated FASTA file of all *E. coli* K12 MG1655 coding sequence (Genbank: U00096-3) using custom MATLAB functions. Non-template strand binding spacers with an 'NGG' PAM site, a GC-content ≤75% and ≥25%, and absence of restriction enzyme sequence motifs for use in subsequent cloning steps (BpiI 'GAAGAC' and its reverse complement, ApaI 'GGGCC' and SmaI 'CCCGG') were considered. SeqMap³⁶ was used to find off-target sites in the genome with up to five mismatches. Only spacers where all identified off-targets had ≥4 mismatches in total with ≥2 in the extended seed sequence of the spacer (the PAM-proximal 12 nucleotides (nt)) were included in the library. For this purpose, 'NAG' PAM sites were added to the count of extended seed mismatches. Finally, among all of the passed spacers, the spacers closest to the start codons of the genes were selected for the library. All selected spacers are listed in Supplementary Table 5.

Design of barcodes. The 20-nt barcode sequences were randomly generated and filtered from forbidden sequence motifs using custom MATLAB code. Specifically, sequences with SchI or BpiI restriction enzyme target sites, or partial target sites that could be recreated in later steps, were removed: SchI 'GAGTC', its reverse complement (rc) site and their partial 'AGTC' and 'ACTC' sites in the 5' end of the barcode. BpiI 'GAAGAC', its rc site and their partial 'GAAG' and 'GTCTT'

sites in the 3' end. Further, to limit the risk of G-quadruplexes and other stable structures, not more than two G's or C's in a row were allowed, and only one G in the 5' and 3' ends. Then, the barcodes were subjected to an initial screen for hairpins, dimers, GC-content and melting temperature using the MATLAB Bioinformatics toolbox with the following settings and boundaries: Hairpins 3 bp, dimers 6 bp, salt concentration 341 mM, oligonucleotide concentration 0.1 µM, 45% ≤ GC-content ≤ 70% and 55 °C ≤ melting temperature ≤ 65 °C. UNAFold³⁷ was used to more thoroughly screen a million such MATLAB-generated barcode sequences for hairpins and self-dimers using the settings temperature 25 °C, salt concentration 341 mM and oligonucleotide concentration 0.1 µM. Barcodes with melting temperatures of hairpins >28 °C and of self-dimers >10 °C were removed, leaving 425,567 barcodes. The remaining barcodes were used to generate a list of internally unique barcodes using custom MATLAB code and the Bioinformatics toolbox local alignment tool, producing 1,888 barcodes with ≤14 total matches and ≤8 consecutive matches against any other barcode in the list. In principle, these barcodes are usable for genotyping using direct FISH. However, to enable use of the barcodes for combinatorial FISH genotyping (Supplementary Note 2), additional sequence elements were added, including four mutually orthogonal elongation probe sequences using our previous design (see ref.¹⁶) and the third/fourth elongation probe sequences 'taggtgtccgaatccatc'/'ggatgtgatagccttgg', and the resulting sequences were further filtered. UNAFold was used to screen for hairpins and self-dimers using the same settings as above but the results were filtered with relaxed thresholds of >50 °C for hairpins and >30 °C for self-dimers for all barcode + elongation probe sequence combinations. This left 275 barcodes, of which 243 were considered for use in combinatorial FISH (Supplementary Note 2) and 40 were used for direct FISH. The 40 barcodes selected for use in the direct FISH DuMPLING cell cycle CRISPRi screen are listed in Supplementary Table 5.

Design and cloning of the DuMPLING library. The 234 spacers targeting *E. coli* genes for CRISPRi were divided into six subpools and paired up with barcodes 2–40 within each subpool. The internal library control sgRNA (*yfp*) was selected as the first library member of each subpool and paired with barcode 1 (Supplementary Table 5). The general design of the oligonucleotides in the pool is as in our previous work¹⁶. Briefly, the sgRNA is expressed by a constitutive sigma 70 promoter in the opposite direction of the T7 promoter for conditional expression of the RNA barcode, which is transcriptionally fused to a stable RNA that allows for accumulation of the RNA barcode in the cell. The barcode and spacer parts, with the divergent T7 and sigma 70 promoters in the center, are flanked by upstream 'GAAGACATCCCC' and downstream 'GTTTATGTCTTC' Golden Gate sites for cloning into the pGuide plasmid¹⁶. Finally, the Golden Gate sites are flanked by subpool-associated orthogonal primer sites, generated using custom MATLAB code, enabling specific amplification of individual subpools from the oligonucleotide pool that was acquired from a commercial manufacturer (CustomArray). The complete oligonucleotide library divided into the six subpools with the primer sites included is available in Supplementary Table 5.

The six subpools were amplified from the oligonucleotide pool using Emulsion PCR as in our previous study¹⁶ to minimize the risk for chimera formation and amplification bias. Pooled pGuide plasmid library assembly using BpiI Golden Gate and transformation of the test strain using electroporation was also as in our previous work, but each subpool was separately transformed. After recovery, the library cultures were split between (1) plating on lysogeny broth agar plates with 50 µg ml⁻¹ Km for colony-forming unit (c.f.u.) estimations and colony PCRs, and (2) dilution 1:100 for selective growth with 50 µg ml⁻¹ Km either in supplemented M9 medium (100 µM CaCl₂, 2 mM MgSO₄, 1× M9 salts, 1× RPMI amino acid mix (Sigma), 0.1 mM uracil, 0.4% succinate) or in lysogeny broth at 37 °C overnight to make glycerol cryostocks used for routine handling of the libraries. From the colony count of the plated portion we calculate that the post-recovery c.f.u.s for the electroporated subpool cultures used to inoculate overnight cultures for making library cryostocks were in the range of 34.8–100 million. The transformation of library subpools 1–6 into the test strain (EL685) created the library strains EL687/744/746/748/750/752. The c.f.u.s of all six subpool cryostocks ranged from ~0.4 to 2.7 million per µl and DuMPLING experiment inoculations were done with at least 1 million c.f.u.s to avoid library representation bottlenecks. Initial quality control of the plasmid libraries was done by Sanger sequencing of 48 colony PCR products (colony PCR and sequencing primers as in ref.¹⁶). No chimeras were found. For a more detailed analysis of library errors, see Supplementary Note 4.

Finally, a series of knockdown control plasmids based on pGuide were constructed, all harboring the barcode for the P1 probe¹⁶. First, different spacers were selected for quantitative PCR experiments (Supplementary Note 1) or for individual analysis on the chip (Supplementary Note 4) as controls for DuMPLING CRISPRi results: *clpP* (EL654), *dnaQ* (EL655), *pgsA* (EL1215), *recA* (EL656), *fis* (EL1089), *lpp* (EL653), *diaA* (EL1090), *ihfA* (EL1091), *ihfB* (EL1092), *uvrD* (EL1093), *dps* (EL1649), *hsdR* (EL1650), *mutS* (EL1651), *rob* (EL1652), *yafN* (EL1653), *yjgH* (EL1334), *dedD* (EL1331) and *yajC* (EL1345). Second, to account for potential lack of specificity in interesting but unexpected phenotypes from the CRISPRi screen, three off-target control plasmids with a second spacer lacking homology to the first one for the *dedD*, *yjgH* and *ybaN* genes were made (EL1333, EL1336 and EL1343, respectively) (Supplementary Note 4). Third, the negative control pGuide-P1-O1a plasmid expressing the *lacO* array-NT1 sgRNA³⁸, whose

target lacks a PAM site, was used to make a reference strain (*ref*) (EL702) for the DuMPLING CRISPRi screen (as described in the microscopy section). All control plasmids' sgRNA and barcode sequences were confirmed using Sanger sequencing.

Construction of single gene knockout control strains. Keio collection knockout strains³⁹ were made as controls for the *dedD*, *yjgH*, *ybaN*, *yajC* and *clpP* DuMPLING knockdowns in the DuMPLING test strain (EL685). *dedD* (ECK2308), *yjgH* (ECK4242), *ybaN* (ECK0462), *yajC* (ECK0401) and *clpP* (ECK0431), corresponding to the Dan Andersson laboratory strains K1052, K1531, K583, K964 and K307, were phage transduced into the DuMPLING test strain using the knockout strains' Km resistance marker (*kanR*) for selection. The flippase plasmid pCP20 (ref.⁴⁰) was used to flip out the *kanR* cassettes and the *cat* gene (from *seqA-yfp-cat*) simultaneously from all knockout strains, but only the *cat* gene from the DuMPLING strain, to be used as a control. All gene deletions and flipping out of the resistance cassettes from both the knockouts and their control strain were confirmed by colony PCR using flanking locus-specific primers, Sanger sequencing of the PCR products and sensitivity towards the removed resistances. To confirm that the deleted genes had not survived through a genomic duplication event, additional colony PCRs were made using coding sequence-internal primers. Additionally, the DuMPLING test knockout strains were transformed with the negative control plasmid pGuide-P1-O1a to better mimic the conditions during the DuMPLING screen. This produced the final DuMPLING test knockout strains for *dedD* (EL1288), *yjgH* (EL1286), *ybaN* (EL1282), *yajC* (EL1283), *clpP* (EL1607) and the control strain with only the *cat* gene flipped out (EL1274). All strains are listed in Supplementary Table 6.

Microscopy experiments. Media. The medium used in the microscopy experiments, unless stated otherwise, is referred to as growth medium (GM) and is composed of 100 μM CaCl_2 , 2 mM MgSO_4 , 1 \times M9 salts, 0.4% (wt/vol) succinate (Sigma), 1 \times RPMI 1640 amino acid mix (Sigma), 0.1 mM uracil and 0.0425% (wt/vol) Pluronic F108 (Sigma) supplemented with 50 $\mu\text{g ml}^{-1}$ Km.

Optical setup. All microscopy experiments were carried out on a Ti2-E microscope (Nikon) equipped with a Plan Apo Lambda 100 \times objective (Nikon). All images were captured on a Zyla 4.2 (Andor) camera attached to the side port of the microscope eyepiece tube base. In both the phenotyping and genotyping experiments, the phase contrast imaging used a light-emitting diode light source (Nikon) together with an external phase module. For fluorescence imaging in the genotyping experiments, the sample was illuminated episcopically using a SOLA SE 365 FISH (Lumencor) attached to the microscope using an LG-N31 collimator (Sutter Instruments). The light from the SOLA is filtered by an excitation filter and mirrored on a dichroic before passing through the objective and illuminating the sample. The emitted light from each dye-labeled FISH probe passes through the same dichroic and is filtered by an emission filter. Each dye that is used has a specific combination of filters and dichroic: The filter combinations used are FF01-642/10, FF665-Di02 and FF02-684/24 for imaging of Cy5/TYE 665; FF01-585/11, Di02-R594 and FF01-625/15 for imaging of Texas Red; FF01-530/11, FF555-Di03 and FF01-575/19 for imaging of Cy3/TYE 563; and finally FF01-473/10, Di02-R488 and FF01-524/24 for imaging of Alexa 488. All filters are manufactured by Semrock. For fluorescence imaging in the phenotyping experiment, samples are episcopically illuminated using a 300 mW 515 nm Fandango laser (Cobolt). Before reaching the objective, the laser beam is shuttered using an AOTFnc together with MDPS (AA Opto-Electronic). The laser beam is then expanded ~ 21 times before being focused on the back aperture of the objective. The expanded laser light is reflected on a dichroic mirror, z405/514rpc (Chroma), before reaching the objective. The emitted light from the SeqA-YFP passes the same dichroic and is filtered by a RazorEdge Long Pass 514 (Semrock). During the phenotyping experiments, the microscope stage and the sample are kept at constant temperature using a temperature unit and lexan enclosure manufactured by Okolab.

Cell growth before loading the microfluidic chip. The *E. coli* cells were inoculated from glycerol stocks into lysogeny broth supplemented with Km at 50 $\mu\text{g ml}^{-1}$ (Sigma), and grown overnight in a 30 $^\circ\text{C}$ shaking incubator. The overnight cultures were diluted 1:500 in GM and grown for an additional 3 h in a 30 $^\circ\text{C}$ shaking incubator.

The microfluidic device. The microfluidic chip is made of PDMS (SYLGARD 184) bonded to a No. 1.5 coverslip (Menzel-Gläser). The process for making the mold, including 300 nm features, for the PDMS casing is described in ref.¹³. The microfluidic chip is mounted on the microscope and tubing (TYGON) is connected to the chip with metal tubing connectors. In some of the experiments we use a modified version of the chip, where the traps are slightly narrower, and the empty reference trap has constrictions in both ends.

Ports 2.0, 7.0, 8.0, 5.1, 5.2, 2.1, 2.2 and 6.0 (see Supplementary Fig. 1) are punched and connected to media containers. When pressures are listed, any port not included was left open and unpressurized. During most microfluidic procedures the GM containers are placed as follows: ports 2.0 (media in, front channel) and 6.0 (media in, back channel) containers are placed slightly higher than the chip and approximately 40 cm higher than the microscope table on which

the other containers are placed. However, during cell loading the back channel tubes (5.1, 5.2 and 6.0) are placed lower compared with the other media-containing tubes to increase flow from the front channel across the cell traps, thereby increasing the bacterial cell loading efficiency. After loading, the tube attached to port 6.0 (media in, back channel) is returned to 40 cm above the microscope table and the containers for ports 5.1 and 5.2 (back channel waste ports) are placed on the microscope table, thus using gravity to maintain a constant flow of media in the back channel.

Cell loading. The microfluidic chip is mounted on the microscope where the lexan enclosure has been preheated to 26 $^\circ\text{C}$ and wetted in GM supplemented with 1 $\mu\text{g l}^{-1}$ aTc using a pressure regulator (Elvesys). In ref.¹⁸, we show that the replication initiation triggering is consistent across many different growth conditions and we do not expect any difference with respect to the regulatory mechanism in the particular conditions used in this paper. Cells were loaded into the 1,000-nm traps from channels 2.1 and 2.2 (see Supplementary Fig. 1), with waste to ports 7.0 and 8.0 (pressures: 2.0: 60 mbar; 2.1: 180 mbar; and 2.2: 180 mbar; further details as described in ref.¹³). After loading, the port configuration and pressures were 2.0: 190 mbar; 7.0: closed; and 8.0: closed. After trapping the cells, the flow of fresh GM + aTc is adjusted to flow in both the front and back channels.

Phenotyping. After 10 h of aTc induction at 26 $^\circ\text{C}$, 90 positions on the chip are imaged in phase contrast (20 ms exposure) every minute and in fluorescence (515 nm, 5.3 W cm^{-2} , 300 ms exposure) every second minute for 8 h. Imaging is performed using in-house plugins to Micromanager.

Genotyping. After phenotyping, the RNA barcode expression is induced by adding GM containing 2% (wt/vol) arabinose (Sigma) to the chip for 3 h at 30 $^\circ\text{C}$. The medium was switched by opening and applying pressure from ports 7.0 and 8.0, reducing the pump-applied pressure to port 2 to zero, switching the medium reservoir containing GM with one containing GM + 2% arabinose, then reversing these steps to result in the configuration described for phenotyping above. All of the following steps are performed at room temperature. After arabinose induction, cells are fixed with 4% formaldehyde (Sigma) in PBS for 20 min followed by a PBS wash step. The fixed and washed cells are then permeabilized by 70% EtOH for 30 min (ports 7.0 and 8.0 open). Next, cells are rehydrated by sequential flowing in 40% EtOH, 20% EtOH and PBS for 20 min each. The rehydrated cells are probed with the reference strain (0.1 μM) probe (TYE 563) in 30% formamide (Sigma) and 2 \times SSC overnight. This allows the microscope to stabilize at room temperature before imaging. All positions are imaged in phase contrast (20-ms exposures) and in fluorescence (Cy3/TYE 563 channel, 500-ms exposures). Switching the fluids for genotyping of the 40 library members was performed manually or fully automated (using the M-Switch and the 2-Switch (Fluigent)). Pools of four probes (0.1 μM each) in 40% formamide and 2 \times SSC were added to the chip in ten separate rounds. Each round of probes was flowed into the chip for less than 7 min and incubated in the chip for 20 min before imaging all positions with phase contrast (20-ms exposure times) and with epifluorescence of the probes individually (TYE 665, TYE 563, Texas Red and Alexa Fluor 488) (500-ms exposure times). Imaging of all positions in phase contrast and four colors of fluorescence takes less than 7 min and the next round of probes is added immediately after finishing the imaging.

Image analysis. Phenotyping. The code for image preprocessing, cell segmentation and cell tracking is mainly based on the code described in ref.¹⁶. Since the same camera was used for the acquisition of phase contrast and fluorescence images, no landmark-based registration was required. Stacks of phase contrast images were aligned using normalized cross-correlation in the region of the dot barcodes imprinted next to empty traps.

Fluorescently labeled SeqA-YFP clusters were localized using the dot detection algorithm from ref.²⁴. To deal with overlapping spots, we measured the aspect ratios of ellipses with the same normalized second central moments as the spots. If the aspect ratio of an ellipse was > 2 , the corresponding spot was split into two parts along the minor axis, passing through the weighted spot centroid. For each of the parts, a new weighted centroid was calculated.

For estimation of a single-cell initiation size, we considered a joint set of dots detected in the cell and one of its daughter cells. The dots were connected into tracks using the single particle tracking algorithm u-track²⁵. We used the settings that allow tracks to split and merge. The obtained tracks were sorted by their initial appearance frames. The track with the earliest initial appearance frame that also contained dots both in the mother and the daughter cells were used to estimate the initiation size. The cell size in the track's appearance frame was used as an estimate of the cell initiation size. The average of single-cell initiation sizes (see Supplementary Fig. 5b) was computed by fitting a histogram of the data with a Gaussian. We used the MATLAB function *fit(gauss1)* and excluded estimates if the coefficient of determination was < 0.7 .

To estimate an average initiation size in bulk (see Supplementary Fig. 5a), we used an error function-based fitting¹⁸. We excluded estimates if the spread of initiation s_i was > 0.7 .

Genotyping. The phase contrast images were preprocessed the same way as in phenotyping above, and used for cell segmentation. Traps filled with cells by less than 20% were discarded. We computed an average cell pixel intensity $I(t,r,f)$ per each trap t , round r and fluorescence channel f . To deal with noise and unspecific aggregation of clusters in the Texas Red channel, we excluded in $I(t,r,f)$ the highest and the lowest 2.5% data values and the average intensity in the empty trap from the same position as trap t . For each fluorescence channel f , we computed a signal lower threshold by fitting the histogram of all trap signals in all rounds in the channel with a Gaussian. We used the MATLAB function $fit('gauss1')$ for the fitting. As shown in Supplementary Fig. 4, the threshold is $\delta = \mu(f) + 7 \times \sigma(f)$ where $\mu(f)$ is the mean and $\sigma(f)$ is the standard deviation of the fitted Gaussian. We said that cells in trap t show a genotype in round r and channel f , if it is the earliest round where $I(t,r,f) > \delta(f)$ and either $r=1$ or $I(t,r,f) > I(t,r-1,f) + 0.5 \times \delta(f)$.

Statistics and reproducibility. *Statistics.* No statistical tests were used in this work. Pearson's correlation coefficient was used to compare growth rates obtained using the DuMPLING method and the NGS method (Fig. 3d), to compare initiation sizes as determined by single-cell analysis and bulk analysis (Fig. 3f) and also to estimate the reproducibility between repeated DuMPLING experiments. Error bars are used in Fig. 3d to represent the standard error of the mean and generated as described in the figure legend. Please refer to the Reporting Summary for additional information.

Reproducibility of the DuMPLING assay. The entire DuMPLING assay was repeated at least two times for each library subpool. In Supplementary Figs. 6–11 we show multiple fork plots from several independent experiments for each genotype. Overall, the reproducibility is high, which is also substantiated by the correlation plots (Supplementary Fig. 12) of the main phenotypic traits between the two sets of experiments. In a few cases, such as *dam* and *clpP*, the fork plot distributions look different between the experiments. This variability is predominantly due to randomness in how many cell traps with erroneous sgRNA appear in each experiment. The case of *clpP* is discussed further in Supplementary Note 4, where selected genotypes of the DuMPLING screen are compared with strains with single-gene knockdowns. Please refer to the Life Sciences Reporting Summary for additional information.

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

Data availability

The microscopy images and image analysis output associated with the DuMPLING experiments are publicly available at the Image Data Resource (<https://idr.openmicroscopy.org/>) under the accession number [idr0065](https://doi.org/10.1038/s41592-019-0629-y). Other data from this study are available from the corresponding author upon request.

Code availability

The code used to analyze the DuMPLING microscopy images and generate figures associated with the DuMPLING experiments is provided as Supplementary Software.

References

- Babic, A., Lindner, A. B., Vulic, M., Stewart, E. J. & Radman, M. Direct visualization of horizontal gene transfer. *Science* **319**, 1533–1536 (2008).
- Keseler, I. M. et al. The EcoCyc database: reflecting new knowledge about *Escherichia coli* K-12. *Nucleic Acids Res.* **45**, D543–D550 (2017).
- Goodall, E. C. A. et al. The essential genome of *Escherichia coli* K-12. *mBio* **9**, e02096-17 (2018).
- Jiang, H. & Wong, W. H. SeqMap: mapping massive amount of oligonucleotides to the genome. *Bioinformatics* **24**, 2395–2396 (2008).
- Markham, N. R. & Zuker, M. in *Bioinformatics, Volume II: Structure, Function and Applications* (ed. Keith, J. M.) Ch. 1, 3–31 (Humana Press, 2008).
- Jones, D. L. et al. Kinetics of dCas9 target search in *Escherichia coli*. *Science* **357**, 1420–1424 (2017).
- 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).
- Cherepanov, P. P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of F₁-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9–14 (1995).

Acknowledgements

This work was supported by the Knut and Alice Wallenberg Foundation (2017.0291 and 2016.0077), the European Research Council (616047) and the Swedish Research Council (VR)(642-2013-7841 and 2016-06213). We are grateful to I. Barkefors for help with the manuscript and figures, to P. Karempudi for making microfluidic molds and to the Dan Andersson lab for kindly sharing *E. coli* strains.

Author contributions

J.E. conceived the DuMPLING method and SeqA application. D.C. developed cloning methods, and designed and made the strain library. J.E. and M.J.L. managed the project. D.F. and M.J.L. developed phenotyping methods. J.L. and M.J.L. developed genotyping methods. D.F. built the microscope. J.L. and M.J.L. performed microscopy experiments. S.Z. and D.F. developed the image analysis pipeline. S.Z. and D.F. analyzed the DuMPLING data. D.J. performed repression measurements and NGS growth rate experiments. J.E., M.J.L., D.C. and D.F. wrote the paper with input from all authors.

Competing interests

The DuMPLING technology is patented with European patent no. EP3167061 (B1).

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41592-019-0629-y>.

Correspondence and requests for materials should be addressed to M.J.L. or J.E.

Peer review information Rita Strack was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

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

Image preprocessing was made using custom MATLAB code and cell segmentation was made using the POE algorithm (version 2015) in Python, cell tracking was made using the Baxter algorithm (free software under MIT license, version 1.5.1) and u-track (free software under GNU General Public License, version 2.1.1)

Data analysis

Custom MATLAB code was used for image data processing and analysis of derived strain-specific genotypic and phenotypic data. NGS data was demultiplexed automatically on the iSeq machine (v100, Illumina), reads were joined using the open source software fastq-join (version 1.3.1), and data was analysed using custom Python 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 data will be publically deposited and accession codes will be provided before publication. The following figures will have associated raw data: Fig 3, Fig 5, Fig ST3, Fig ST5 and Fig S1.

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](https://www.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	No required sample size was determined in advance as we are not testing hypothesis
Data exclusions	For the DuMPLING microscopy imaging data, the following cell lineages were excluded from analysis: (i) Cell outlines where the size transiently dropped or increased by large amounts were deemed as missegmentation and not used in further analysis. (ii) Lineages from one cell generation with large shifts in size (non-transient) were excluded from further analysis. (iii) Lineages from one cell generation with large center of mass movements of the cell outlines were excluded from further analysis. Finally (iv), lineages from one cell generation with very short life span were excluded from further analysis unless they contained both a mother and two daughter cells. Further, to exclude strains present in a low number of traps making the derived data sensitive to strain mutations, we set the minimum traps threshold for specific strains at 5. Also, to ensure that we had sufficient strain-specific data for calculating reliable cell cycle-related characteristics, the minimum number of complete cell cycles threshold was set to 40. These limits were set empirically and were not pre-established. They are also reported in the manuscript. For the NGS data, individual reads were discarded that could not be reliably mapped to one of the DuMPLING library members or that displayed mutations (mutations in sgRNA or sgRNA promoter) which would deactivate CRISPRi.
Replication	Yes the DuMPLING experiments were replicated successfully. The new section "4 Statistics and Reproducibility" is added to the Online Methods.
Randomization	Randomization was not necessary as we are not testing hypothesis.
Blinding	Blinding was not necessary as we are not testing hypothesis

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
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging