

Using conditional independence tests to elucidate causal links in cell cycle regulation in *Escherichia coli*

Prathitha Kar^{a,b}, Sriram Tiruvadi-Krishnan^c, Jaana Männik^c, Jaan Männik^{c,1}, and Ariel Amir^{a,2}

^aSchool of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02134, USA; ^bDepartment of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA; ^cDepartment of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996, USA

This manuscript was compiled on January 21, 2023

How cells regulate their cell cycles is a central question for cell biology. Models of cell size homeostasis have been proposed for bacteria, archaea, yeast, plant and mammalian cells. New experiments bring forth high volumes of data suitable for testing existing models of cell size regulation and proposing new mechanisms. In this paper, we use conditional independence tests in conjunction with data of cell size at key cell cycle events (birth, initiation of DNA replication and constriction) in the model bacterium *Escherichia coli* to select between the competing cell cycle models. We find that in all growth conditions that we study the division event is controlled by the onset of constriction at mid-cell. In slow growth, we corroborate a model where replication-related processes control the onset of constriction at mid-cell. In faster growth, we find that the onset of constriction is affected by additional cues beyond DNA replication. Finally, we also find evidence for the presence of additional cues triggering initiations apart from the conventional notion where the mother cells solely determine the initiation event in the daughter cells via an adder per origin model. The use of conditional independence tests are a novel approach in the context of understanding cell cycle regulation and it can be used in future studies to further explore the causal links between cell events.

Conditional independence | Cell cycle | *Escherichia coli*

Cell size is regulated across all forms of life. The advent of single-cell experiments has advanced our understanding of these regulatory mechanisms over the past decade (1–3). Single cells growing in microfluidic channels when combined with fluorescence microscopy can be used to track the size and the timing of cell-cycle events such as birth, DNA replication initiation, termination, septum formation and division (4–10). Existing models of cell-cycle regulation can be tested against the high-throughput data obtained from these experiments and the data can be used to hypothesize new models.

Previous studies have proposed cell cycle models where cells are assumed to initiate a round of DNA replication upon adding a constant size per origin, on average, from the previous initiation (8, 9, 11–13). This model of replication initiation control, referred to as “adder per origin”, predicts that the size added per origin between successive initiations of DNA replication is uncorrelated with the size at initiation for single-cells which has been observed experimentally (8, 9).

However, the proposed cell cycle models differ in how the division event is controlled by the DNA replication process (8, 9, 11, 12, 14). Cooper and Helmstetter proposed that cell division follows the initiation of DNA replication after a constant time has elapsed (14) (we will refer to this as the CH model). Within this model, this constant time is the sum of the time taken for DNA replication (the C period) and the

time from termination of DNA replication to division (the D period), see Figure 1. In the parallel adder model proposed for *Mycobacterium smegmatis*, cell division occurs after the cell has increased by a constant size per origin from replication initiation (15). This model was later proposed for *E. coli*, where it was referred to as “double adder” (9). We will use the term, parallel adder (PA), in this paper to describe two adders working in parallel (initiation to initiation and initiation to division). In both CH and PA models, division is controlled solely by the replication initiation event. A competing model suggests that division happens independently of the DNA replication process (8). In this model, the division is controlled by accumulation of a key protein to a threshold level, starting from cell birth. A middle ground is the concurrent processes model where division is controlled by a combination of cues, some of which originate from cell birth and others from the initiation of DNA replication (13, 16, 17). Identifying the correct statistical analysis method and model has been contentious.

Much of the support for the models hypothesized above comes from absence or presence of correlations between two variables describing cell cycle. Recent studies have found different models described above to be consistent with the same data using the same analysis method (18–20). There is a lack of consensus on the use of a statistical method to study cell cycle regulation which leads to a lack of consensus on the underlying cell cycle model. To address this, we will go beyond two-variable correlations in this paper and use concepts of

Significance Statement

Cells across all domains of life grow and divide such that their sizes are tightly regulated, yet the coordination of these processes remains poorly understood. Previous works proposed different models for this coupling in bacteria, in some of which division is controlled by DNA replication processes while in others it is uncoupled from it. We combine experimental data on single-cell *E. coli* growth with the powerful methodology of causal inference to show that constriction is a cell cycle checkpoint controlled exclusively by DNA replication processes in slow-growth conditions, while additional cues are at play in faster growth conditions. We also show that control of the DNA replication cycles is more complex than previously thought, paving the way for future studies.

The authors declare that they have no competing interests.

¹ E-mail: jmannik@utk.edu; phone: +1 (865) 974 6018; ² Email: arielamir@seas.harvard.edu; phone: +1 (617) 495 5818

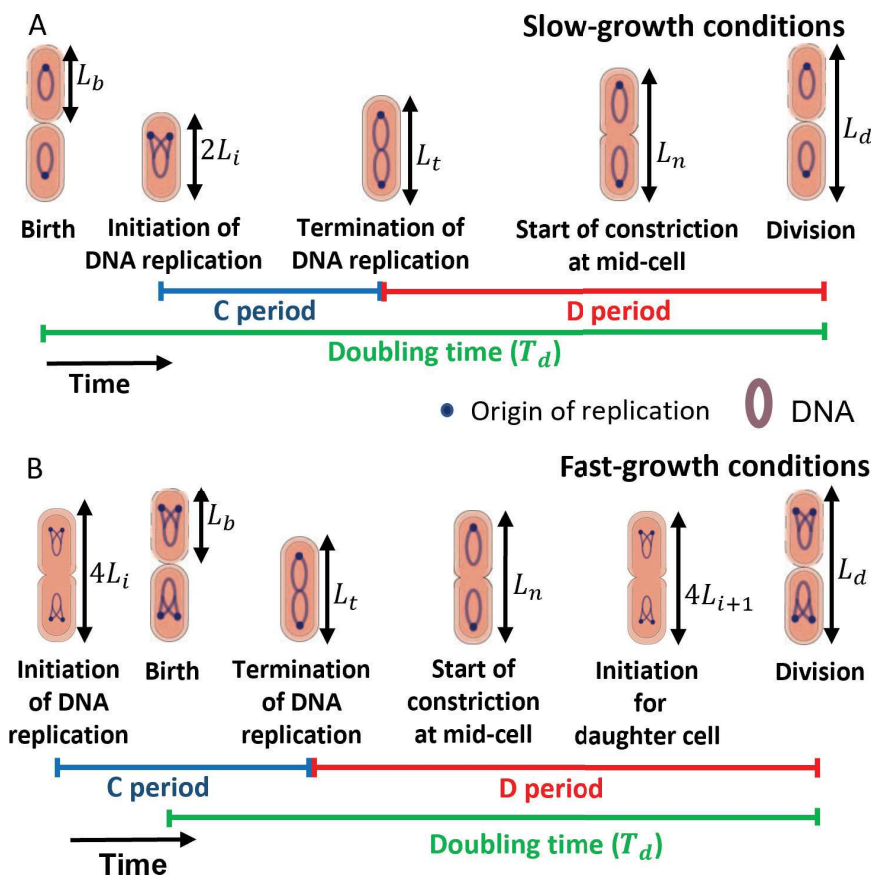


Fig. 1. Key cell cycle events analyzed. **A.** At slower growth rates, cells are born with a single origin of replication. **B.** In faster growth conditions, cells are born with multiple origins. In the case of cells with two origins at birth, the initiation starts in the mother cell. The C period is the time taken for DNA replication and the D period is the time between termination of DNA replication and cell division. The lengths at these events considered in the data analysis are also shown.

causal inference relying on conditional correlations involving more than two variables to study causal statements.

While causal inference is widely used in epidemiology, sociology, economics, and computer science (21), it has not been utilized previously in testing cell cycle models. Specifically, we study the relation between replication and the onset of constriction and find that in the slowest growth conditions, replication is the limiting factor controlling the onset of constriction, but in faster growth conditions additional regulatory cues need to be invoked to explain the data. Furthermore, we find that the onset of constriction directly leads to division without the involvement of any additional regulatory mechanisms that retain the memory of birth size. Finally, the data suggest that replication initiation in the mother cell is not the sole factor controlling the initiation in the daughter cell, as was suggested previously (8, 9, 11–13). While the casual inference methodology we are using is agnostic to the details of the underlying molecular mechanisms, it allows us to gain important insights on the possible regulatory network architecture and narrow down the potential biological pathways.

Results

Replication control on division is growth rate dependent. To investigate how the cell cycle events are controlled in *E. coli*, we used data from recent experiments in 6 different growth media (10). These data have been collected at slow (average number of origins at birth ~ 1) and moderately fast growth rates (average number of origins at birth > 1). The experiments were conducted at 28°C where the growth rates were about twice as slow as that in 37°C (22). The data contains

the timings of cell cycle events such as birth, initiation of DNA replication, termination, start of septum formation and division for hundreds of cell cycles and the corresponding cell lengths for those events.

Previous works have considered correlations between cell cycle variables such as the size at birth (L_b) and size at division (L_d) to infer cell cycle models (2). Using linear regression, we show the best linear fit between L_b and L_d for a fast growth condition in Figure 2A and a slow growth condition in Figure 2B. For cells growing in fast growth conditions (Figure 2A), the underlying equation is close to $L_d = L_b + \Delta L$ and cells are assumed to be following an adder model where cells divide on addition of constant size from birth (5, 8, 9). In slower growth conditions, the cells have been shown to follow a near-adder (Figure 2B and Ref (6)). Ref (23) provided a general framework to infer the cell cycle regulation strategy from L_d vs L_b plots. In this model, a cell born at size L_b divides at size L_d by employing a regulatory mechanism $f(L_b)$ (a deterministic element), to which noise is added. Mathematically, for the case of size-additive noise this corresponds to the equation $L_d = f(L_b) + \eta$, where η is the noise in division size independent of L_b . This is an example of a structural causal model (SCM), widely used in causal inference (24). The SCMs can be visualized using directed acyclic graphs. The nodes in the graph are connected via directed edges with the direction of the arrows going from cause (variables on right side of the SCM) to effect (variable on left side of the SCM). Each node in the graph represents a variable which may either correspond to an observable quantity obtained in the experiments or to an unobserved variable. In the graphs that we will study in the

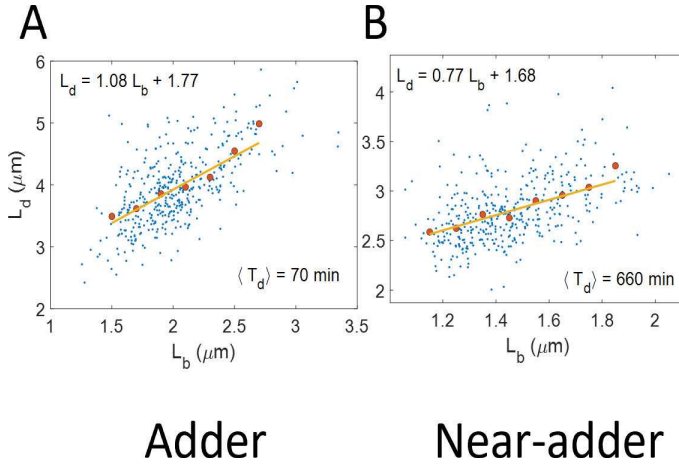


Fig. 2. Correlation between birth and division: A-B. L_d vs L_b graphs are plotted for data from Ref (10) in **A**. Glucose-cas medium (generation time ($\langle T_d \rangle$) = 70 min, $N = 409$ cells). The best linear fit has a slope = 1.08 (0.97, 1.19). **B**. Acetate medium (generation time ($\langle T_d \rangle$) = 660 min, $N = 401$ cells). The best linear fit has a slope = 0.77 (0.65, 0.89). The numbers in the parenthesis here represent the 95% confidence interval. In Figures 2A-2B, the blue dots represent the raw data, the red dots the binned data and the yellow line the best linear fit.

paper, the nodes will correspond to cell lengths at cell cycle events (see Section S2 in SI text for an explanation as to why using lengths is advantageous compared to using the timing of the events). In the graphs, the absence of an edge between two nodes shows that there is no direct causal effect between the two variables. In the case of cell cycle regulation models, the SCMs and the graphs are independent of the nature of noise, and so will be our conclusions. The latter can be either size or time additive (25).

The linear relation between L_b and L_d can also be explained by other cell cycle models such as the CH and PA models. Next, we construct causal graphs for these *E. coli* cell-cycle models. In the CH and PA models, initiation of DNA replication controls when division happens. In slow-growth conditions where the number of origins at birth is 1, initiation and division occur in the same cell cycle (Figure 1A). However, in the faster-growth conditions used in the experiments, replication initiation could start in the mother cell and the number of origins at birth is 2 (Figure 1B). Mathematically, the size at division is determined by $L_d = 2(L_i + \Delta_{id}) + \eta$ for the PA model, where L_i is the initiation size per origin number taken right after initiation, Δ_{id} is the size per origin added between initiation and division and η is a size additive noise. In the CH model where cells are undergoing exponential growth with growth rate λ , $L_d = 2L_i e^{\lambda(C+D)} + \eta_t$, where η_t is a size additive noise. This is shown as an arrow from L_i to L_d (Figure 3A). In these models, initiation size $L_i = \frac{L_{i-1} + \Delta_{ii}}{2} + \xi$, where L_{i-1} is the previous initiation size per origin, Δ_{ii} is the size per origin added between the consecutive initiations and ξ is a size additive noise. This is shown as an arrow from L_{i-1} to L_i in Figure 3A. The previous initiation event (L_{i-1}) also controls the division event in the mother cell (L_{d-1}) or equivalently the birth event of the current cell cycle (arrow from L_{i-1} to L_b). L_{i-1} is a confounder which means that it is a common cause of two events, in this case, L_b and L_i . In a second class of models referred to as “concurrent processes” (13, 16, 17), the division size is determined by the slowest of two processes- 1. constant size addition from birth (adder) at fixed growth rate 2. a time C+D from initiation of DNA replication (where each of the two processes is also subject to noise). The corresponding SCM for exponentially growing cells with growth rate λ is $L_d = \max(L_b + \Delta'_{bd} + \delta'_{bd}, L_i e^{\lambda(C'+D') + \delta'_{C+D}})$ and it is represented by arrows from L_b to L_d and L_i to L_d , in the graph shown in Figure 3B. L_d is a common effect of

L_b and L_i and in this case, L_d is said to be a collider. Note that the measured average C+D period will be determined by the competition between the two processes, and therefore could be different than $C' + D'$. Similarly, the measured average size added between birth and division will be different than Δ'_{bd} . In the concurrent processes model’s SCM, δ'_{bd} and δ'_{C+D} are the noise terms in Δ'_{bd} and $C' + D'$, respectively. The noise terms are independent of each other and are also uncorrelated with L_b and L_i . Similar to CH and PA models, birth (L_b) and initiation (L_i) are associated by a common cause, the initiation in the previous cell cycle (L_{i-1}) (Figure 3B). Note that the birth event in previous cell cycle (L_{b-1}) also controls the division event in mother cell (L_{d-1}) according to the concurrent process model and hence, it controls birth in current cell cycle (L_b). We do not show the L_{b-1} to L_b causal link here as the omission of the link will have no effect on our analysis. For complete causal diagrams, see Section S3 in the SI text. A third model, the independent adder (IA) model is also shown in Figure 3C where the division length is solely controlled by the birth length (arrow from L_b to L_d) independently of the initiation length (7, 8, 26). The initiation is controlled by the previous initiation as in the CH, PA and concurrent processes models (arrow from L_{i-1} to L_i). Importantly, the links between L_i and L_b , and L_i and L_d are absent as initiation is independently controlled from division. Directed acyclic graphs (DAGs) such as the ones shown in Figures 3A-3C can be used to determine correlations and conditional correlations.

Correlations and conditional correlations are determined from the DAGs using a set of rules known as d-separation (21). These rules will be briefly explained below. In graph 3A, since L_i controls L_d , they will be correlated. L_b and L_i are correlated via the confounder, L_{i-1} . Only under specific conditions where the effects of the two links cancel each other, L_b and L_i will be uncorrelated. Directed acyclic graphs encode information beyond two-variable correlations, namely, conditional independencies (CI). Conditional correlation $r(L_b, L_d | L_i)$ means finding the correlation between two variables, L_b and L_d upon fixing the value of a third variable, L_i . In graph 3A, L_b and L_d are uncorrelated upon fixing the value of L_i and the path between L_b and L_d is then closed (in contrast, a path connecting two variables and leading to their correlations is defined as open, for example, the path between L_b and L_d without conditioning in graph 3A). In graph 3B, the collider

L_d blocks the path between L_b and L_i and the path between L_b and L_i via L_d is closed. The path opens upon conditioning on a collider or any descendant of a collider: for instance, in graph 3B upon conditioning on the variable L_d , the path between L_b and L_i via L_d will be open. To summarize, a path is closed if a non-collider in the path is conditioned upon or if a collider and its descendants are not conditioned on. In the case of multiple paths between two variables, the variables are uncorrelated if all the paths between those variables are closed (Section S1 in SI text). In this paper, we will go beyond the previously used methodology of using two variable correlations (Figures 2A-2B) and use CI tests to select cell cycle models.

The model corresponding to graph 3C (IA model) predicts that L_i will be uncorrelated with L_b and L_d (prediction shown below the graph in panel 3C), as initiation is not linked to either birth or division. We find using experimental data that the Pearson correlation coefficients between L_b and L_i ($r(L_b, L_i)$), and L_i and L_d ($r(L_i, L_d)$) are non-zero in all six measured growth conditions (Table S1). Note that we have excluded the two fastest growth conditions from (10) because of incomplete tracking of DNA replication initiation in these data sets. The result rules out IA model as a viable model for cell cycle regulation, as was previously argued in Ref (9). In contrast, both models shown in graphs 3A and 3B predict that L_i will be correlated with L_b and L_d . Thus, we have to go beyond two variable correlations and use CI tests to distinguish between the two graphs.

To distinguish the models in graph 3A and 3B, we will condition on the initiation length, L_i , and calculate the conditional correlation ($r(L_b, L_d|L_i)$) between L_b and L_d . We predict using d-separation that L_b and L_d are uncorrelated on fixing L_i in graph 3A. However, they are predicted to be correlated in graph 3B as there is a direct causal link between L_b and L_d . We validated the method using synthetic data generated by existing models following the methodology outlined in Ref (25) (Section S4 in SI text).

The simplest way of calculating $r(L_b, L_d|L_i)$ using experimental data is by calculating the correlation between L_b and L_d in the small interval ($L_i - dL, L_i + dL$). We do not use this method because the number of data points of L_b and L_d corresponding to each interval in the available datasets is too small making the conditional correlations hard to interpret (see Supplementary Figure 5). In order to obtain the conditional correlation, we will instead remove the influence of L_i from L_b and L_d using linear regression. To that end, we assume linear dependence of L_b and L_d on L_i . The linear relations can be rationalized as Taylor expansions around the mean of the non-linear relations between L_b and L_i , and L_d and L_i . The residuals obtained upon carrying out the linear regression of L_b on L_i ($L_b|L_i$) and L_d on L_i ($L_d|L_i$) represent the effect of sources other than L_i on L_b and L_d , respectively. The correlation $r(L_b, L_d|L_i)$ is calculated by obtaining the Pearson correlation coefficient between the residuals $L_b|L_i$ and $L_d|L_i$ (see Materials and Methods, (27)). In this method of calculating the conditional correlation, we use the complete dataset available for each growth medium. Note that when we refer to conditional correlations as vanishing throughout the paper we mean that the Pearson correlation coefficient is not statistically significant when using a p-value as the metric at a significance level of 0.05.

Next, we use the experimental data to test whether

$r(L_b, L_d|L_i)$ is zero or not. We plot the residuals obtained using linear regression of L_d on L_i ($L_d|L_i$) and L_b on L_i ($L_b|L_i$). We find the correlation coefficients between the residuals to be negligible for the two slowest growth media (Figure 3D and Supplementary Figure 1A) and non-zero for the other growth conditions (Figure 3E and Supplementary Figures 1B-1D). Thus, graph 3A is consistent with the data in the two slowest growth conditions while the model in graph 3B is consistent with data in the faster growth conditions. The correlations are tabulated for each growth medium in Table S2. Accounting for possible outliers in the data (keeping the middle 95% percentile data of both axes), we find the p-value to be above significance level of 0.05 in the three slowest growth condition (Supplementary Figure 1E). This finding is still in agreement with the hypothesis of the replication process becoming more limiting for determining division in slower growth conditions. We also checked whether growth rate $\lambda = \frac{1}{T_d} \ln(\frac{L_d}{L_b})$ affected the correlations between the residuals. The correlation coefficients between the residuals obtained using linear regression of L_d on L_i and λ ($L_d|(L_i, \lambda)$) and L_b on L_i and λ ($L_b|(L_i, \lambda)$) are shown in Table S2. We still find the correlations to be close to zero for the two slowest growth conditions and non-zero for the others. We also analyzed previously published datasets (8, 9, 13) and found that they were consistent with a model where both birth and replication processes limit division in fast growth, and replication becomes more limiting in slower growth conditions (Section S5 in SI text).

To conclude, in two slowest experimental growth conditions, division is solely controlled by replication (consistent with CH/PA models). However, in faster four growth conditions, additional processes starting from cell birth also control division (consistent with concurrent processes model).

The onset of constriction solely controls the division size.

Previous studies propose the start of septum formation at mid-cell as an important checkpoint involved in length control (7, 10). However, most of the previous cell cycle models, including the aforementioned ones, did not explicitly contain this checkpoint, but only considered the division event. In this section, we show that cells exert size control at the start of constriction at mid-cell and the constriction process ultimately culminates in division, without additional regulation on the division timing. We will use cell lengths at birth, the onset of constriction (L_n) and division as a proxy to denote the events. The onset of constriction can be determined by labeling FtsN with a fluorescent fusion protein; FtsN is the last known essential component of the *E. coli* divisome to assemble at the mid-cell before constriction starts (28–33). The accumulation of FtsN at the mid-cell thus indicates the start of septum formation, as was validated in Ref. (10).

We hypothesize a causal graph based on our prior knowledge about the start of septum formation at mid-cell. Previous works suggest that an accumulation of a threshold amount of cell division proteins such as FtsZ (8), or cell wall precursors (7) starting from birth is responsible for triggering constriction at mid-cell. For both scenarios and assuming also a balanced growth, we expect $L_n = L_b + \Delta_{bn} + \xi$, where Δ_{bn} is the size added between birth and the onset of constriction and ξ is a size additive noise. This relation is depicted by an arrow from L_b to L_n in the graph of Figure 4A, where the arrow from L_n to L_d represents commitment to division upon the onset of constriction. A competing model is shown in Figure 4B, where

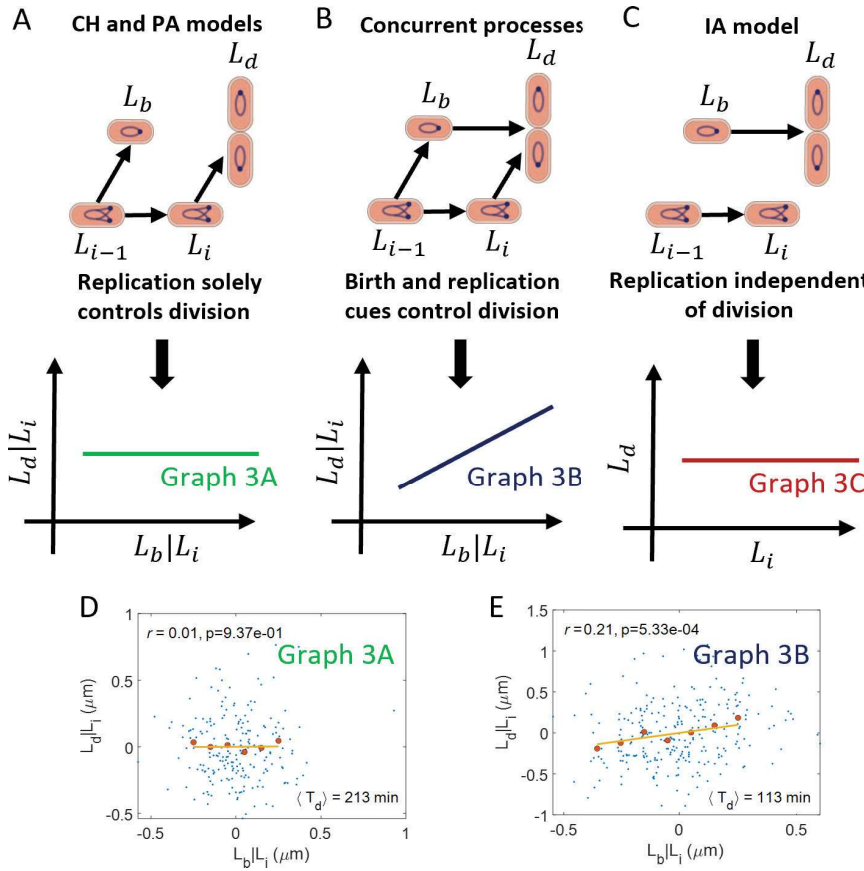


Fig. 3. Causal graphs and CI tests linking birth, initiation and division: **A.** A causal graph linking lengths at birth (L_b), initiation (L_i) and division (L_d). In this graph, the replication initiation controls the division event. We predict $r(L_b, L_d|L_i) = 0$ for the graph. **B.** Both birth and initiation simultaneously control the division event. We predict $r(L_b, L_d|L_i) \neq 0$ for the graph. **C.** A graph corresponding to the independent adder (IA) model. Here, the division process is independent of the replication process. L_b and L_d are uncorrelated from L_i . **D-E.** Residuals obtained on linear regression of L_d on L_i ($L_d|L_i$) and L_b on L_i ($L_b|L_i$) are plotted for data from Ref. (10) in **D.** Alanine medium (generation time = 213 min, $N = 215$ cells, average number of origins at birth, $\langle n_{ori} \rangle = 1.07$). **E.** Glucose medium (generation time = 113 min, $N = 259$ cells, $\langle n_{ori} \rangle = 1.98$). The conditional correlation, $r(L_b, L_d|L_i)$ is negligible for the alanine medium (consistent with graph 3A) while it is non-zero for the glucose medium (consistent with graph 3B). In Figures 3D-3E, the blue dots represent the raw data, the red dots the binned data and the yellow line the best linear fit. r is the Pearson correlation coefficient between the variables in the x and y axes. p in the plots are the p-values. We reject the null hypothesis that the correlation is zero if the p-value is less the significance level set at 0.05.

in addition to the onset of constriction, another biochemical process starting at cell birth is limiting for the division event (for example, the accumulation of another key protein).

We expect that the variables L_b , L_n and L_d will be correlated with each other for both graphs 4A and 4B. This is because L_n shares a cause and effect relationship with L_d and L_b , respectively. This is indeed what we observe in the experimental data for all six growth media as shown in Table S3. Note that the relation between birth and the onset of constriction deviated from an adder model in all growth conditions.

Next, we test the predictions of conditional independence obtained by applying d-separation on the graphs in Figure 4A and 4B. For the graph in Figure 4A, we predict $r(L_b, L_d|L_n) = 0$ using d-separation while for Figure 4B, $r(L_b, L_d|L_n)$ is non-zero. To test these predictions, we find the correlation between the residuals obtained on linear regression of L_b on L_n (denoted as $L_b|L_n$) and L_d on L_n ($L_d|L_n$). The plots of the residuals are shown in Figure 4C and Figure 4D for cells growing in a slow growth medium (alanine, generation time = 213 min) and a fast growth medium (glucose, generation time = 113 min), respectively. In Figures 4C-4D, we show the correlation between the residuals to be close to zero. Similar negligible correlations are also obtained for four other growth media as shown in Supplementary Figures 2A-2D and Table S3 with the corresponding p-values (Supplementary Figure 2E) above the significance level. Thus, the graph in Figure 4A is consistent with the experimental data.

These results show that the onset of constriction can be regarded as a cell cycle checkpoint that solely controls the cell

size at division without any additional cues from cell birth.

Cell cycle model involving the onset of constriction. In the previous section, we verified that the onset of constriction can be regarded as a cell cycle checkpoint. Previously, we showed that replication controls division in slow-growth conditions and is one of the factors controlling division in fast growth conditions. In this section, we combine these two results into a single, coherent model and discuss models where replication is coupled to constriction (which, in turn, is coupled to division).

To this end, we adapt the cell cycle models of graphs 3A and 3B by hypothesizing that birth size and replication initiation size control the size at onset of *constriction* instead of division size. The graph in Figure 5A corresponds to a model where initiation controls constriction (arrow from L_i to L_n). Such a control may be exerted by nucleoid occlusion, whereby a chromosome blocks the formation of FtsZ ring via DNA binding proteins(34) or sterically (35). Within this model, constriction may start when the chromosome segregation is underway, lowering the DNA density at the mid-cell and relieving the effects of nucleoid occlusion (10). Since termination of DNA replication follows causally from initiation, within the graph we may depict this mechanism by an arrow from initiation of DNA replication to constriction. Thus, a limiting factor that controls the start of constriction may be the start of DNA replication (Figure 5A). A competing model is shown in graph 5B where the size at onset of constriction is simultaneously controlled by birth size (arrow from L_b to L_n) and initiation size (arrow from L_i to L_n). In this model, accumulation of division proteins and nucleoid occlusion may

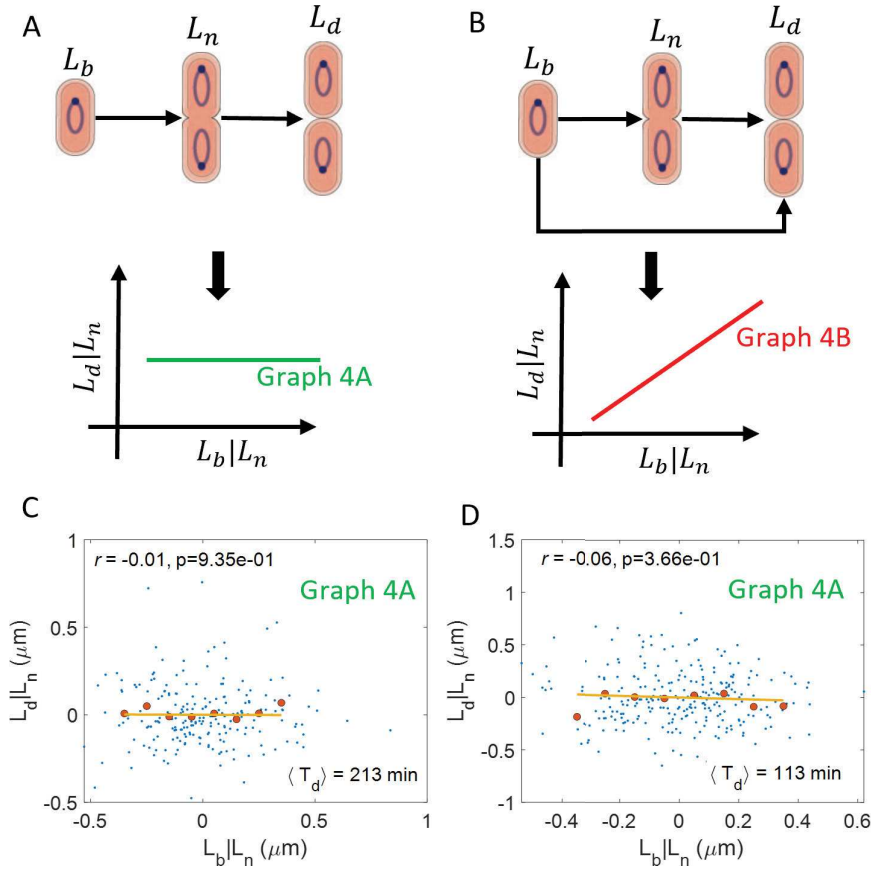


Fig. 4. Linking birth, onset of constriction and division:
A. A causal graph linking the length at birth (L_b), onset of constriction (L_n) and division (L_d). In this graph, division is solely controlled by the onset of constriction. We predict that $r(L_b, L_d|L_n) = 0$ for the graph. **B.** A causal graph where multiple processes - from birth and from the onset of constriction control division. We predict a non-zero $r(L_b, L_d|L_n)$ for the graph. **C-D.** Residuals obtained on linear regression of L_d on L_n ($L_d|L_n$) and L_b on L_n ($L_b|L_n$) are plotted for **C**. Alanine medium (generation time = 213 min, $N = 215$ cells, $\langle n_{ori} \rangle = 1.07$). **D.** Glucose medium (generation time = 113 min, $N = 259$ cells, $\langle n_{ori} \rangle = 1.98$). The conditional correlations, $r(L_b, L_d|L_n)$ are close to zero for both of the growth conditions (consistent with graph 4A).

both play a limiting role on start of constriction (10). Based on the results in the previous section, constriction culminates in division. This is shown as arrows from L_n to L_d in Figures 5A and 5B.

Both models predict that L_i and L_n will generally be correlated (in contrast to models where protein accumulation from birth triggers constriction, independently of DNA replication processes). We indeed find them to be correlated in experimental data in all 6 growth conditions (Table S4). Next, we use d-separation to predict correlations and conditional correlations between the cell cycle variables in graphs 5A and 5B. Graph 5A predicts L_b and L_n to be uncorrelated when conditioned upon L_i , while graph 5B predicts them to be correlated. To test these predictions, we plot the residuals $L_n|L_i$ and $L_b|L_i$ in Figures 5C-5D, Supplementary Figures 3A-3D. We find the correlations between the residuals to be zero for the two slowest growth conditions while it is non-zero for other growth conditions (see p-values in Supplementary Figure 3E). We also considered the correlations $r(L_b, L_n|(L_i, \lambda))$ to control for the effects of growth rate. The results obtained are similar to that shown in Figures 5C-5D, Supplementary Figures 3A-3D. Thus, we find graph 5A to be consistent with data in the two slowest growth conditions while graph 5B to be consistent with data in faster growth conditions.

Next, we show that our predictions of correlations and conditional correlations using graphs 5A-5B are in agreement with the conditional correlations discussed in the previous sections. Graph 5A predicts $r(L_b, L_d|L_i)$ to be zero, while graph 5B predicts a non-zero correlation. These predictions are identical to those of graphs 3A and 3B, respectively. As

previously discussed, $r(L_b, L_d|L_i)$ is non-zero in the four faster growth conditions while it is zero in the two slowest growth conditions. Thus, we again find graph 5A to be consistent with the two slower growth conditions while graph 5B is consistent with the other four growth conditions. We also showed that $r(L_b, L_d|L_n) = 0$ in the experiments for all growth conditions. This is consistent with our predictions obtained using d-separation for both graphs 5A and 5B.

To probe the molecular mechanisms that might be involved in coupling of the replication cycle to the division cycle, we used mutants that lack proteins which link the replication and division processes. The $\Delta zapA$, $\Delta zapB$, $\Delta matP$, $\Delta slmA$, FtsK K997A and $\Delta minC$ mutants were grown in M9 glycerol+trace elements medium ($T_d = 148$ min in wildtype cells (WT)) (10). In this growth condition, our analysis indicated the onset of constriction is controlled by two concurrent pathways (graph 5B). If these proteins were to mediate the coupling between the replication processes and the onset of constriction then on removing these proteins in the mutants, we expect the correlation between initiation and the onset of constriction upon conditioning on birth to be zero. However, we find that the correlation $r(L_i, L_n|L_b)$ in both the WT and mutants is non-zero except in the Min mutants which undergo polar divisions (see Section S6 in SI text). One possible explanation for the difference in the correlation $r(L_i, L_n|L_b)$ between cells undergoing mid-cell and polar divisions in the Min mutants is nucleoid occlusion as proposed previously in this section and in Ref (10). According to this idea, nucleoid density at mid-cell blocks the formation of the Z-ring until the later stages of the replication process, thus, coupling replication and the onset

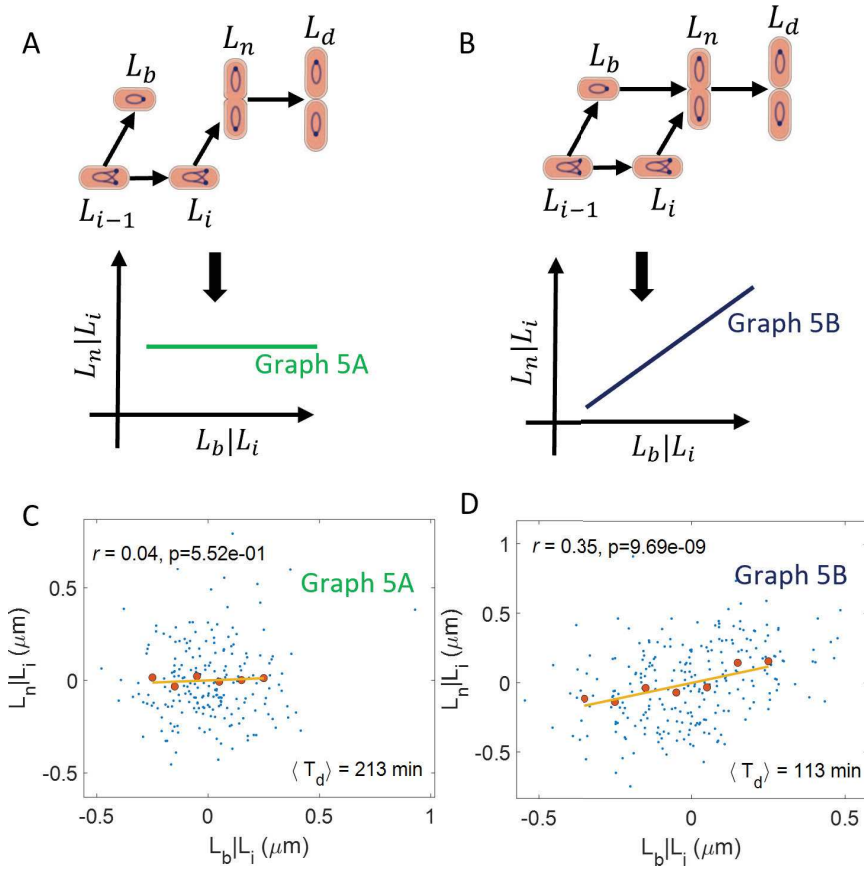


Fig. 5. Cell cycle regulation model: **A.** A causal graph linking the lengths at birth (L_b), initiation (L_i), onset of constriction (L_n) and division (L_d). In this graph, constriction is controlled by the DNA replication process. We predict $r(L_b, L_n|L_i) = 0$ for the graph. **B.** A causal graph in which the size at the onset of constriction is simultaneously controlled by the birth size and the initiation size. We predict a non-zero $r(L_b, L_n|L_i)$ for the graph. **C-D.** Residuals obtained on linear regression of L_n on L_i ($L_n|L_i$) and L_b on L_i ($L_b|L_i$) are plotted for **C.** Alanine medium (generation time = 213 min, $N = 215$ cells, $\langle n_{ori} \rangle = 1.07$). **D.** Glucose medium (generation time = 113 min, $N = 259$ cells, $\langle n_{ori} \rangle = 1.98$). The conditional correlation, $r(L_b, L_n|L_i)$ is close to zero for the alanine medium (consistent with graph 5A) while it is non-zero for the glucose medium (consistent with graph 5B).

of constriction while polar divisions are not inhibited by such factors and they can happen independently of replication, thus, leading to a lack of causal link between replication and the onset of constriction.

To conclude, we showed that in slow-growth conditions replication initiation controls the onset of constriction and hence, division, while in fast growth conditions there are additional limiting factors.

Initiation is not solely controlled by initiation in the previous cell cycle. So far, we have discussed the control of the division cycle and the link between the replication and division cycle. A question that arises is what controls the replication cycle. The main events in the DNA replication cycle are the initiation and termination of replication. As we discussed earlier, previous works suggested that the initiation happens via an adder per origin model (8, 9, 11, 16, 36). In the model, the initiation size per origin of the daughter cell (L_{i+1}) is related to the initiation size per origin of the current cell cycle (L_i) as $L_{i+1} = \frac{L_i + \Delta_{ii}}{2} + \xi$, and $r(L_i, L_{i+1})$ is expected to be 0.5. The experimental data analyzed show the correlation to be close to 0.5 (Table S7).

We also include replication termination in our model. Previous analysis suggests termination occurs after a constant time from initiation (the C period) (11, 14), consistent with a constant speed of the replication forks as observed in single-molecule experiments (37, 38). We include this prior knowledge in graph 6A as a causal link between initiation and termination, where we denote the length at termination of DNA replication as L_t . L_i , L_t and L_{i+1} are correlated with

each other in graph 6A. These predictions are consistent with the correlations in experimental data for all six growth conditions (Table S7). Furthermore, we predict that L_t and L_{i+1} will be uncorrelated upon conditioning on L_i in graph 6A. However, we find that $r(L_t, L_{i+1}|L_i)$ is non-zero in all growth conditions (Figures 6C-6D, Supplementary Figures 4A-4E, Table S7). In fact, this result is consistent with a model proposed in graph 6B which assumes that initiation in the daughter cell is also controlled by termination along with initiation in the current cell cycle. We predict using d-separation on graph 6B that $r(L_i, L_{i+1}|L_t)$ is non-zero which is consistent with our experiments. Graph 6B was also consistent with the data published in Ref (13) (Section S5 in SI text).

To further test the model proposed in graph 6B, we use data from cells whose C period was longer as compared to the WT cells (10). This was achieved by deleting *thyA* and controlling the amount of thymine in the growth medium (39). $\Delta thyA$ cells grown in thymine concentrations of 500 $\mu g/mL$ at 28°C in glycerol + trace elements medium had identical replication period as WT cells. However, on decreasing the concentration to 15 $\mu g/mL$, the C period showed a step-wise increase by approximately 40% (10). An increase in the C period may lead to termination in the current cell cycle happening after the initiation for the next cell cycle has started. Such a temporal order will violate the model presented in graph 6B where termination is a cause of initiation in daughter cells. The variation in timings at termination (T_{rt}), division (T_d) and initiation for the next cell cycle (T_{i+1}) are shown in Figure 6E for the $\Delta thyA$ strain. Time $t=0$ on the x-axis corresponds to the time when cells were shifted to

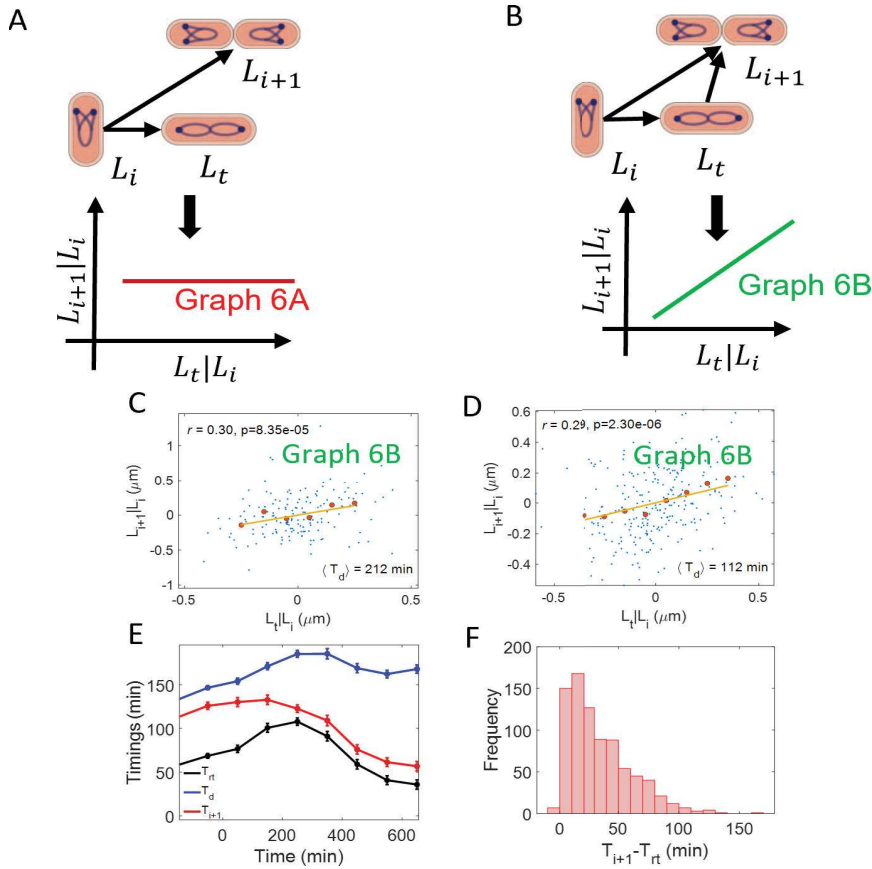


Fig. 6. Control of replication initiation: **A.** A causal graph linking the lengths at initiation per origin (L_i), termination (L_t) and the lengths at initiation per origin in the daughter cells (L_{i+1}). In this graph, the initiation in the daughter cells is solely controlled by initiation in the current cell cycle. We predict $r(L_t, L_{i+1}|L_i) = 0$ for the graph. **B.** A causal graph in which the initiation in the daughter cells is controlled simultaneously by initiation and termination in the current cell cycle. We predict a non-zero $r(L_t, L_{i+1}|L_i)$ for the graph. **C-D.** Residuals obtained on linear regression of L_{i+1} on L_i ($L_{i+1}|L_i$) and L_t on L_i ($L_t|L_i$) are plotted for **C.** Alanine medium (generation time = 212 min, $N = 167$ cells, $\langle n_{ori} \rangle = 1.08$). **D.** Glucose medium (generation time = 112 min, $N = 255$ cells, $\langle n_{ori} \rangle = 1.98$). The conditional correlation, $r(L_t, L_{i+1}|L_i)$ is non-zero for both alanine and glucose media (consistent with graph 6B). **E-F.** *ΔthyA* cells are grown in a thymine concentration of 500 $\mu\text{g/mL}$ in M9 glycerol+trace elements medium at 28°C. The cells are then shifted to a thymine concentration of 15 $\mu\text{g/mL}$. Upon shifting to a lower thymine concentration, the C period of the cells increases. We measure the timings at the termination of DNA replication (T_{rt}) and the DNA replication initiation for the next cell cycle (T_{i+1}) in multiple cells throughout the experiment (both before and after the shift). **E.** Variation of timing (relative to cell birth) of termination, initiation for next cell cycle and division is plotted. We show the binned data where the cell events' timings are averaged in each bin based on the time when cell divides (x-axis). Time $t=0$ in x axis represents the time when cells are shifted to the lower thymine concentration. **F.** We plot the distribution of $T_{i+1} - T_{rt}$ timings for all cells measured in the experiment.

15 $\mu\text{g/mL}$ thymine concentration. Strikingly, we find at the single-cell level that only few cells have the time $T_{i+1} - T_{rt} \leq 0$ and it is always greater than -6 min (Figure 6F). Since the measurement interval is 4 min, an error in the measurement of the initiation and termination events by one time frame can lead to a minimum time difference $T_{i+1} - T_{rt} = -8$ min even though the events coincide. Thus, the data is consistent with the temporal ordering of events in graph 6B even when the replication timings are perturbed. We note that graph 6B is unlikely to apply to faster growth conditions where overlapping rounds of replication have been reported (14, 40).

To conclude, we rule out the model in which initiation in the next cycle is controlled solely by initiation in the current cycle, showing that control over replication initiation is more complex than previously thought.

Discussion

In the paper, we make use of causal inference i.e., conditional independence tests, to interrogate cell cycle models. An ideal cell-cycle model should be able to reproduce the joint probability density of all cell cycle variables measured. Since the amount of data collected is finite, previous cell cycle modeling studies have relied on using certain correlations (or lack of correlations) between cell cycle variables to hypothesize models (6, 8, 9, 13, 17, 23, 41). The model simulations are then compared to experiments using specific correlations. The model which agrees the most with these chosen correlations is accepted as the underlying model. However, multiple models having different causal structures can agree with these limited correlations making it difficult to choose a particular

causal model (20). Conditional independence tests allow us to reject models in a robust manner that do not depend on the fine-tuned details of the models but instead only relies on the structure of the causal network (i.e., which variables control which other variables). The framework relies on testing whether conditional correlations are zero or not - without resorting to their precise numerical values.

Our goal was to test several models previously proposed for the bacterial cell cycle ranging from models in which DNA replication was assumed to control cell division to models where DNA replication cycles are independent of the cell division cycles (and a class of models interpolating the two, in which division couples not only to DNA replication but also to additional cues). Note that, generally, this framework of causal inference cannot determine the model structure *de novo* but rather allows us (in certain cases) to rule out particular models.

After validating our method on synthetic data, we used causal inference methods on recently obtained data measuring key cell cycle variables (length and time of cell birth and division, initiation and termination of DNA replication and constriction of the division ring) (10). We found that our data agreed with replication being the sole limiting factor for division in the two slowest growth conditions (Figure 3D, Supplementary Figure 1A). In faster growth conditions, the data agreed with a model in which birth size and replication initiation size both controlled division size (Figure 3E, Supplementary Figures 1B-1D).

Although the onset of constriction has not been included in previous cell cycle models, it can be expected to be an essential

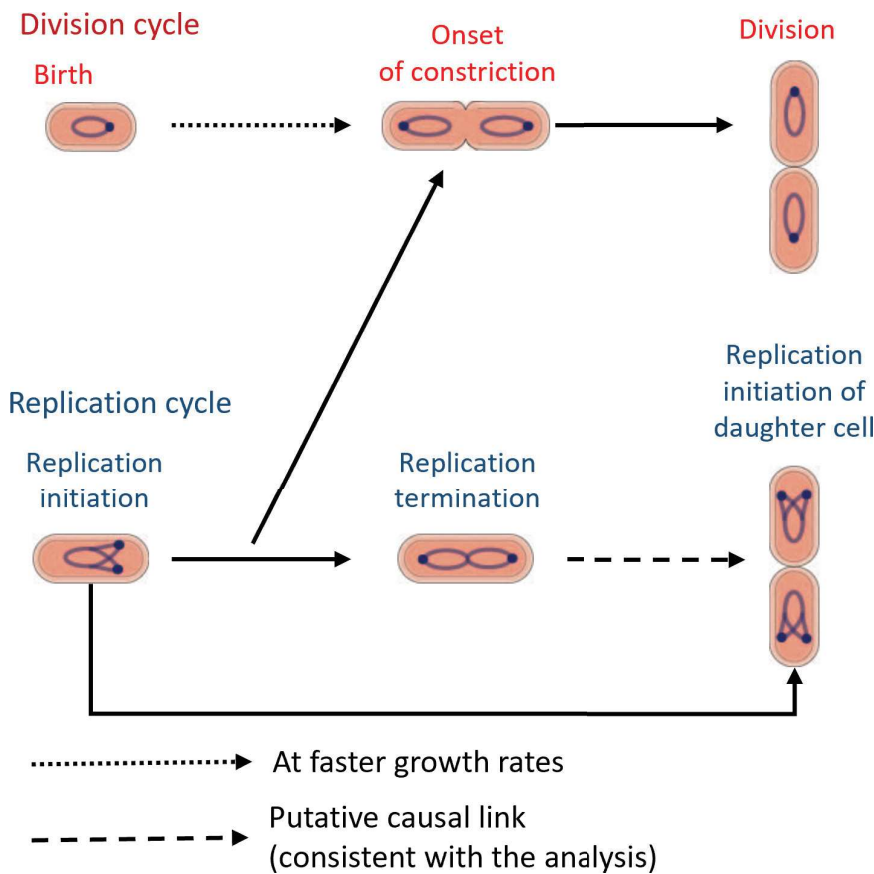


Fig. 7. Schematics of the proposed cell cycle model: In slow growth conditions, replication related processes, possibly nucleoid occlusion, limits the onset of constriction. In fast growth conditions, additional processes such as accumulation of cell division proteins, FtsZ or cell wall precursors, might also control the onset of constriction independent of replication. Finally, the constriction process culminates in the division event. The start of the replication cycle is controlled by the replication initiation event in the previous cell cycle as well as an additional cue which is linked with the termination event.

cell cycle checkpoint in *E. coli*. We tested this idea using conditional correlations. We find the conditional correlations between birth and division lengths to be zero when conditioned on constriction length (Figures 4C-4D, Supplementary Figures 2A-2D). Note that we condition on the length at the start of constriction because of availability of data at that time point. However, a biochemical reaction leading to onset of constriction may occur at some prior time close to the start of constriction (with lengths highly correlated with L_n) which may result in zero $r(L_b, L_d|L_n)$. For example, the incorporation of one of the many proteins in the Z ring can be the limiting step. Once the protein is incorporated to form the Z ring, the constriction starts after a small time delay. The existing data is not sufficient to distinguish these molecular steps yet. Regardless of the nature of the biochemical processes, our analysis confirms that onset of constriction controls cell cycle progression from birth to division. Thus, including the constriction event into the cell cycle is important for theoretical and experimental studies involving cell cycle regulation.

Combining these two results led us to envision a coarse-grained model for cell size regulation in which the constriction event is controlled by the DNA replication process alone in slow growth. In fast growth, the onset of constriction must be controlled by additional regulatory processes linked to cell birth and not controlling DNA replication initiation. In all growth conditions division is downstream of the onset of constriction (Figure 7). We corroborated the model predictions for the conditional correlation $r(L_b, L_n|L_i)$, predicted to be zero in slow growth conditions (Figure 5C, Supplementary Figure 3A) and non-zero in fast growth (Figure 5D, Supple-

mentary Figures 3B-3D). An appealing molecular mechanism that explains the causal control of replication initiation over the onset of constriction is that of nucleoid occlusion, in which septum formation is blocked by a replicating nucleoid (34). The nucleoid occlusion or absence thereof at the cell poles explains the lack of correlations between replication and constriction in Min mutants undergoing polar divisions. Previous work also showed that in slow growth conditions increasing the DNA replication time, using mutants where external thymine levels determine the C period, delays the start of constriction (10). In both the wild-type and the thymine mutants, the constriction process does not start until the DNA density at the mid-cell has decreased. In fast growth conditions, replication is not the sole limiting process, as evidenced by the non-vanishing conditional correlations. One possible additional mechanism is the accumulation of division proteins such as FtsZ (8), or cell wall precursors (7) that controls the trigger for constriction. The cell cycle regulation model discussed here are in agreement with the models proposed using correlations between the timings of different cell cycle events in Ref (10). Our analysis of data in Ref (6, 8, 9, 13) as well the analysis in Ref (13) itself are consistent with a model where the replication process becomes more limiting for determining division in the slower growth conditions.

We also studied the DNA replication cycle using the CI methodology. It has been suggested that accumulation of a threshold amount of the initiator protein DnaA in its ATP bound form is needed to initiate DNA replication (42, 43). The accumulation starts from the previous initiation and the initiation size of the previous replication cycle controls the

initiation size in the current replication cycle via an adder per origin model (9, 11, 12, 44). Furthermore, the termination of DNA replication happens after a C period elapses since the initiation (14). The adder per origin model predicts that $r(L_i, L_{i+1}) = 0.5$. The correlations $r(L_i, L_{i+1})$ reported in previous studies (8, 9) and observed in the experiments analyzed in this paper are close to 0.5, thus, lending support to an adder per origin model. However, such a model (Figure 6A) would also predict that the correlation between initiation in the daughter cell and termination event when conditioned upon the initiation event in the current cell cycle is zero. We find the conditional correlations to be non-zero in all six growth conditions (Figures 6C-6D, Supplementary Figures 4A-4D). This agrees with the graph shown in Figure 6B which suggests a more complicated model than previously thought. The current assumption is that DnaA accumulation triggers initiation. However, the availability of DnaA to initiate the next round of replication is not dependent on termination. One possibility is that replisome components other than DnaA are limiting for replication initiation. Note that, in our paper, the validity of these results are tested at growth rates that do not necessitate overlapping rounds of replication forks (doubling times less than the C period).

The termination event was used for the conditional correlation analysis because of its availability from the experiments (10). However, we cannot rule out the possibility that other events (correlated with termination) instead of termination in graph 6B could also predict a non-zero correlation between L_t and L_{i+1} upon conditioning on L_i . However, such an event cannot be cell division. The data from almost all available growth conditions studied in the paper show that at least some initiation events can precede cell division. Such time ordering violates the causality principle. Furthermore, replication initiation can start without any division in filamentous *E. coli* cells (45). The presence of more than a single initiation event per cell cycle was also the basis for rejecting a cell cycle model called the sequential adder, containing an adder from birth to initiation and another from initiation to division (15).

A possible alternative event for termination controlling the next initiation can be related to some replication-dependent conformational change within the nucleoid. It has been hypothesized that nucleoid tethered to the midcell (called the progression control complex or the PCC) inhibits both the onset of constriction and the next initiation (46). Once the cell has completed certain growth requirements, the PCC undergoes conformational changes permitting the next initiation and constriction formation to occur. These conformational changes could potentially happen at termination or close to it. If this hypothesis is correct, termination and the next initiation would be correlated upon conditioning on the initiation of the current cell cycle and as such this scenario will be able to explain the data.

It remains for future studies to determine at which growth rates the next initiation becomes uncorrelated from the previous termination event. The future studies can also identify if some conformational change in the nucleoid precedes the initiation or if there is some rate-limiting component beyond DnaA that controls the initiation. In the latter experiments, up regulation of the limiting component could shift initiation earlier and lead to disappearance of correlations.

To conclude, our analysis leads to a new cell cycle model in

E. coli linking division and replication cycles, which extends the previously developed concurrent processes model (Figure 7). To come to this result, we used a versatile method of inference involving conditional independence tests. The technique may prove useful in analyzing and critically testing cell cycle models also in other organisms.

Materials and Methods

Obtaining conditional correlations. The method used to calculate conditional correlations throughout the paper was introduced in Results section. In this section, we discuss the method from a mathematical perspective.

Our aim is to calculate the correlation between variables A and B when conditioned upon variables $\mathbf{X} = \{X_1, X_2, X_3, \dots, X_n\}$. Here, $\mathbf{X}$ is a set of n variables which are being conditioned upon. Conditional correlation when conditioned upon $\mathbf{X}$ means finding the correlation on fixing the values of *all* variables in the set $\mathbf{X}$. Fixing $\mathbf{X}$ would remove the effects of variability in $\mathbf{X}$ on other variables.

We use a method based on partial regression to calculate the conditional correlation (47). To achieve this, we try to find the effect of $\mathbf{X}$ on variables A and B . The random variables A , B and $\mathbf{X}$ will correspond to cell lengths at various events in the manuscript. Since cell lengths are narrowly distributed about their means, we need to know the dependence/effects of $\mathbf{X}$ on A and B around their means. Hence, we can Taylor expand the non-linear dependence of A , and B on $\mathbf{X}$ around the means and consider terms to first order. We represent it as,

$$A = \sum_{i=1}^n a_i X_i + \eta, \quad [1]$$

$$B = \sum_{i=1}^n b_i X_i + \xi. \quad [2]$$

a_i s and b_i s are calculated by multiple linear regression of A on $\mathbf{X}$ and B on $\mathbf{X}$, respectively. η and ξ capture the effects on A and B , respectively, from sources other than $\mathbf{X}$ i.e., they represent variability in A and B on removing the effects of $\mathbf{X}$. η and ξ are therefore the residuals obtained from the multiple linear regression of A on $\mathbf{X}$ and B on $\mathbf{X}$, respectively. The conditional correlation between A and B when conditioned upon $\mathbf{X}$ (denoted as $r(A, B|\mathbf{X})$) is obtained by finding the Pearson correlation coefficient between the residuals η and ξ .

ACKNOWLEDGMENTS. The authors thank Daniel Needleman, Jane Kondev, Hanna Salman, Aleksandra Walczak, Marco Cosentino Lagomarsino, and Sven van Teeffelen for the useful discussions and feedback on the manuscript. This work has been supported by the US-Israel BSF research grant 2017004 (JM), the National Institutes of Health award under R01GM127413 (JM), NSF CAREER 1752024 (AA), NIH grant 103346 (AA) and NSF award 1806818 (PK).

1. L Willis, KC Huang, Sizing up the bacterial cell cycle. *Nat. Rev. Microbiol.* **15**, 606–620 (2017).
2. PY Ho, J Lin, A Amir, Modeling cell size regulation: From single-cell-level statistics to molecular mechanisms and population-level effects. *Annu. Rev. Biophys.* **47**, 251–271 (2018).
3. S Jun, F Si, R Pugatch, M Scott, Fundamental principles in bacterial physiology—history, recent progress, and the future with focus on cell size control: a review. *Reports on Prog. Phys.* **81**, 056601 (2018).
4. P Wang, et al., Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103 (2010).
5. M Campos, et al., A constant size extension drives bacterial cell size homeostasis. *Cell* **159**, 1433–1446 (2014).
6. M Wallden, D Fange, EG Lundius, Ö Baltekin, J Elf, The synchronization of replication and division cycles in individual *E. coli* cells. *Cell* **166**, 729–739 (2016).
7. LK Harris, JA Theriot, Relative rates of surface and volume synthesis set bacterial cell size. *Cell* **165**, 1479–1492 (2016).
8. F Si, et al., Mechanistic origin of cell-size control and homeostasis in bacteria. *Curr. Biol.* **29**, 1760–1770 (2019).
9. G Witz, E van Nimwegen, T Julou, Initiation of chromosome replication controls both division and replication cycles in *E. coli* through a double-adder mechanism. *eLife* **8**, e48063 (2019).
10. S Tiruvadi-Krishnan, et al., Coupling between DNA replication, segregation, and the onset of constriction in *Escherichia coli*. *Cell Reports* **38**, 110539 (2022).
11. PY Ho, A Amir, Simultaneous regulation of cell size and chromosome replication in bacteria. *Front. Microbiol.* **6**, 662 (2015).

12. M Berger, PRt Wolde, Replication initiation in *E. coli* is regulated via an origin-density sensor generating adder correlations. *arXiv preprint arXiv:2106.03674* (2021).
13. A Colin, G Micali, L Faure, MC Lagomarsino, S van Teeffelen, Two different cell-cycle processes determine the timing of cell division in *Escherichia coli*. *eLife* **10**, e67495 (2021).
14. S Cooper, CE Helmstetter, Chromosome replication and the division cycle of *Escherichia coli* br. *J. Mol. Biol.* **31**, 519–540 (1968).
15. MM Logsdon, et al., A parallel adder coordinates mycobacterial cell-cycle progression and cell-size homeostasis in the context of asymmetric growth and organization. *Curr. Biol.* **27**, 3367–3374 (2017).
16. G Micali, J Grilli, M Osella, MC Lagomarsino, Concurrent processes set *E. coli* cell division. *Sci. Adv.* **4**, eaau3324 (2018).
17. G Micali, J Grilli, J Marchi, M Osella, MC Lagomarsino, Dissecting the control mechanisms for DNA replication and cell division in *E. coli*. *Cell Reports* **25**, 761–771 (2018).
18. G Le Treut, F Si, D Li, S Jun, Comment on 'Initiation of chromosome replication controls both division and replication cycles in *E. coli* through a double-adder mechanism'. *bioRxiv* (2020).
19. G Witz, T Julou, E van Nimwegen, Response to comment on 'Initiation of chromosome replication controls both division and replication cycles in *E. coli* through a double-adder mechanism'. *bioRxiv* (2020).
20. G Le Treut, F Si, D Li, S Jun, Quantitative examination of five stochastic cell-cycle and cell-size control models for *Escherichia coli* and *Bacillus subtilis*. *Front. Microbiol.* p. 3278 (2021).
21. J Pearl, *Causality*. (Cambridge university press), (2009).
22. SL Herendeen, RA VanBogelen, FC Neidhardt, Levels of major proteins of *Escherichia coli* during growth at different temperatures. *J. Bacteriol.* **139**, 185–194 (1979).
23. A Amir, Cell size regulation in bacteria. *Phys. Rev. Lett.* **112**, 208102 (2014).
24. J Peters, D Janzing, B Schölkopf, *Elements of Causal Inference: Foundations and Learning Algorithms*. (The MIT Press), (2017).
25. P Kar, S Tiruvadi-Krishnan, J Männik, J Männik, A Amir, Distinguishing different modes of growth using single-cell data. *eLife* **10**, e72565 (2021).
26. S Taheri-Araghi, et al., Cell-size control and homeostasis in bacteria. *Curr. Biol.* **25**, 385–391 (2015).
27. JP Guilford, *Fundamental Statistics in Psychology and Education*. (McGraw-Hill), (1950).
28. DS Weiss, Last but not least: new insights into how FtsN triggers constriction during *Escherichia coli* cell division. *Mol. Microbiol.* **95**, 903–909 (2015).
29. B Liu, L Persons, L Lee, PA de Boer, Roles for both FtsA and the FtsBLQ subcomplex in FtsN-stimulated cell constriction in *Escherichia coli*. *Mol. Microbiol.* **95**, 945–970 (2015).
30. DP Haeussler, W Margolin, Splitsville: structural and functional insights into the dynamic bacterial Z ring. *Nat. Rev. Microbiol.* **14**, 305–319 (2016).
31. DO Daley, U Skoglund, B Söderström, FtsZ does not initiate membrane constriction at the onset of division. *Sci. Reports* **6**, 1–6 (2016).
32. S Du, J Lutkenhaus, At the heart of bacterial cytokinesis: The Z ring. *Trends Microbiol.* **27**, 781–791 (2019).
33. A Boes, S Olatunji, E Breukink, M Terrak, Regulation of the peptidoglycan polymerase activity of PBP1b by antagonist actions of the core divisome proteins FtsBLQ and FtsN. *mBio* **10**, e01912–18 (2019).
34. LJ Wu, J Errington, Nucleoid occlusion and bacterial cell division. *Nat. Rev. Microbiol.* **10**, 8–12 (2012).
35. CL Woldringh, The role of co-transcriptional translation and protein translocation (transertion) in bacterial chromosome segregation. *Mol. Microbiol.* **45**, 17–29 (2002).
36. H Zheng, et al., Interrogating the *Escherichia coli* cell cycle by cell dimension perturbations. *Proc. Natl. Acad. Sci.* **113**, 15000–15005 (2016).
37. NA Tanner, et al., Real-time single-molecule observation of rolling-circle DNA replication. *Nucleic Acids Res.* **37**, e27–e27 (2009).
38. TM Pham, et al., A single-molecule approach to DNA replication in *Escherichia coli* cells demonstrated that DNA polymerase III is a major determinant of fork speed. *Mol. Microbiol.* **90**, 584–596 (2013).
39. R Pritchard, A Zaritsky, Effect of thymine concentration on the replication velocity of DNA in a thymineless mutant of *Escherichia coli*. *Nature* **226**, 126–131 (1970).
40. M Chandler, R Bird, L Caro, The replication time of the *Escherichia coli* K12 chromosome as a function of cell doubling time. *J. Mol. Biol.* **94**, 127–132 (1975).
41. I Soifer, L Robert, A Amir, Single-cell analysis of growth in budding yeast and bacteria reveals a common size regulation strategy. *Curr. Biol.* **26**, 356–361 (2016).
42. T Katayama, K Kasho, H Kawakami, The dnaa cycle in *Escherichia coli*: activation, function and inactivation of the initiator protein. *Front. Microbiol.* **8**, 2496 (2017).
43. R Reyes-Lamothe, DJ Sherratt, The bacterial cell cycle, chromosome inheritance and cell growth. *Nat. Rev. Microbiol.* **17**, 467–478 (2019).
44. F Barber, PY Ho, AW Murray, A Amir, Details matter: noise and model structure set the relationship between cell size and cell cycle timing. *Front. Cell Dev. Biol.* **5**, 92 (2017).
45. I Gelber, A Aranovich, M Feingold, I Fishov, Stochastic nucleoid segregation dynamics as a source of the phenotypic variability in *E. coli*. *Biophys. J.* **120**, 5107–5123 (2021).
46. NE Kleckner, K Chatzi, MA White, JK Fisher, M Stouf, Coordination of growth, chromosome replication/segregation, and cell division in *E. coli*. *Front. Microbiol.* **9**, 1469 (2018).
47. MP Allen, *Understanding regression analysis*. (Springer), (1997).

1 Using conditional independence tests to
2 elucidate causal links in cell cycle regulation in
3 *Escherichia coli* - Supplementary Information

4 Prathitha Kar^{1,2}, Sriram Tiruvadi-Krishnan³, Jaana Männik³, Jaan Männik^{*3},
5 and Ariel Amir^{†1}

6 ¹*School of Engineering and Applied Sciences, Harvard University, Cambridge,*
7 *MA 02134, USA*

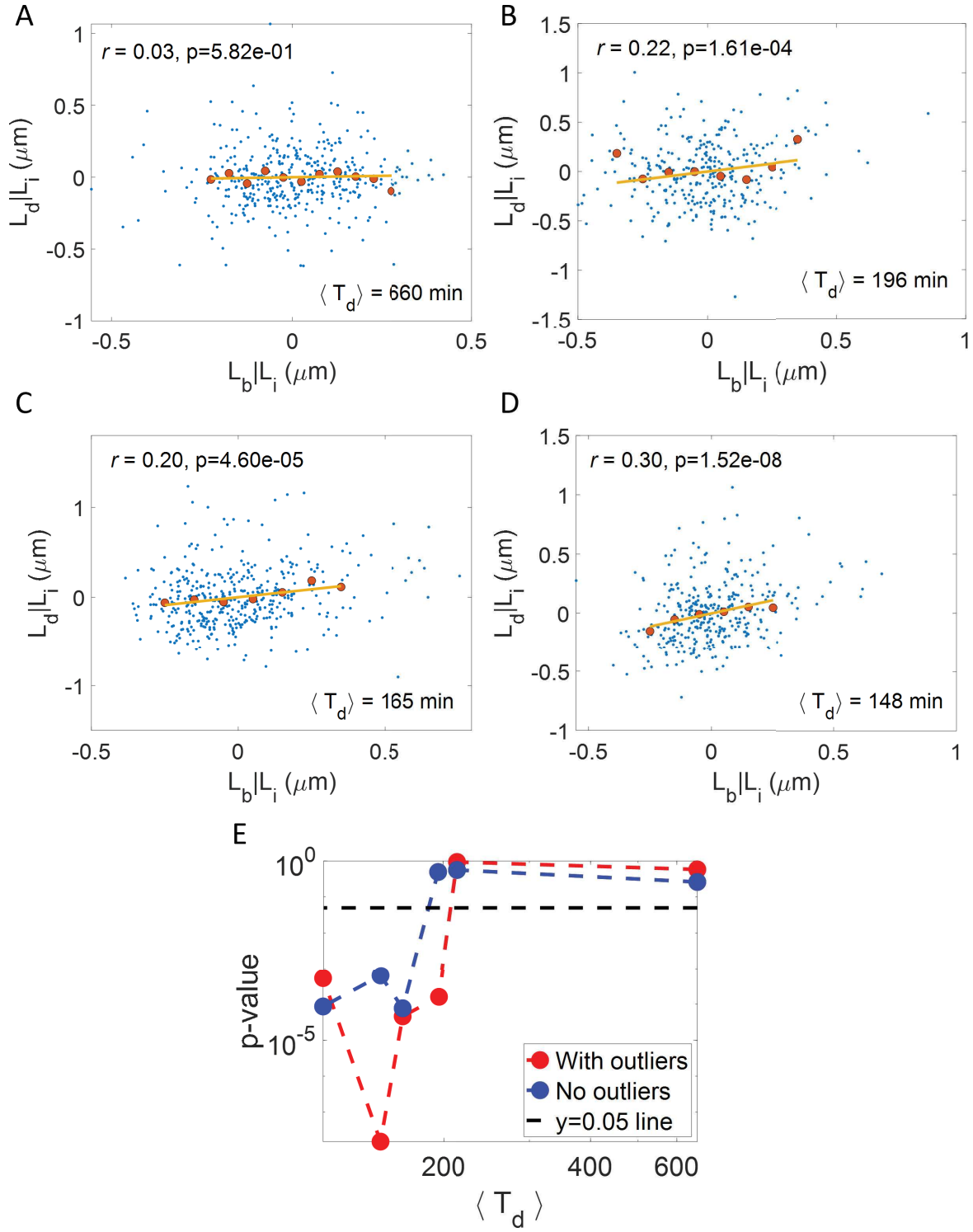
8 ²*Department of Chemistry and Chemical Biology, Harvard University,*
9 *Cambridge, MA 02138, USA*

10 ³*Department of Physics and Astronomy, University of Tennessee, Knoxville,*
11 *TN 37996, USA*

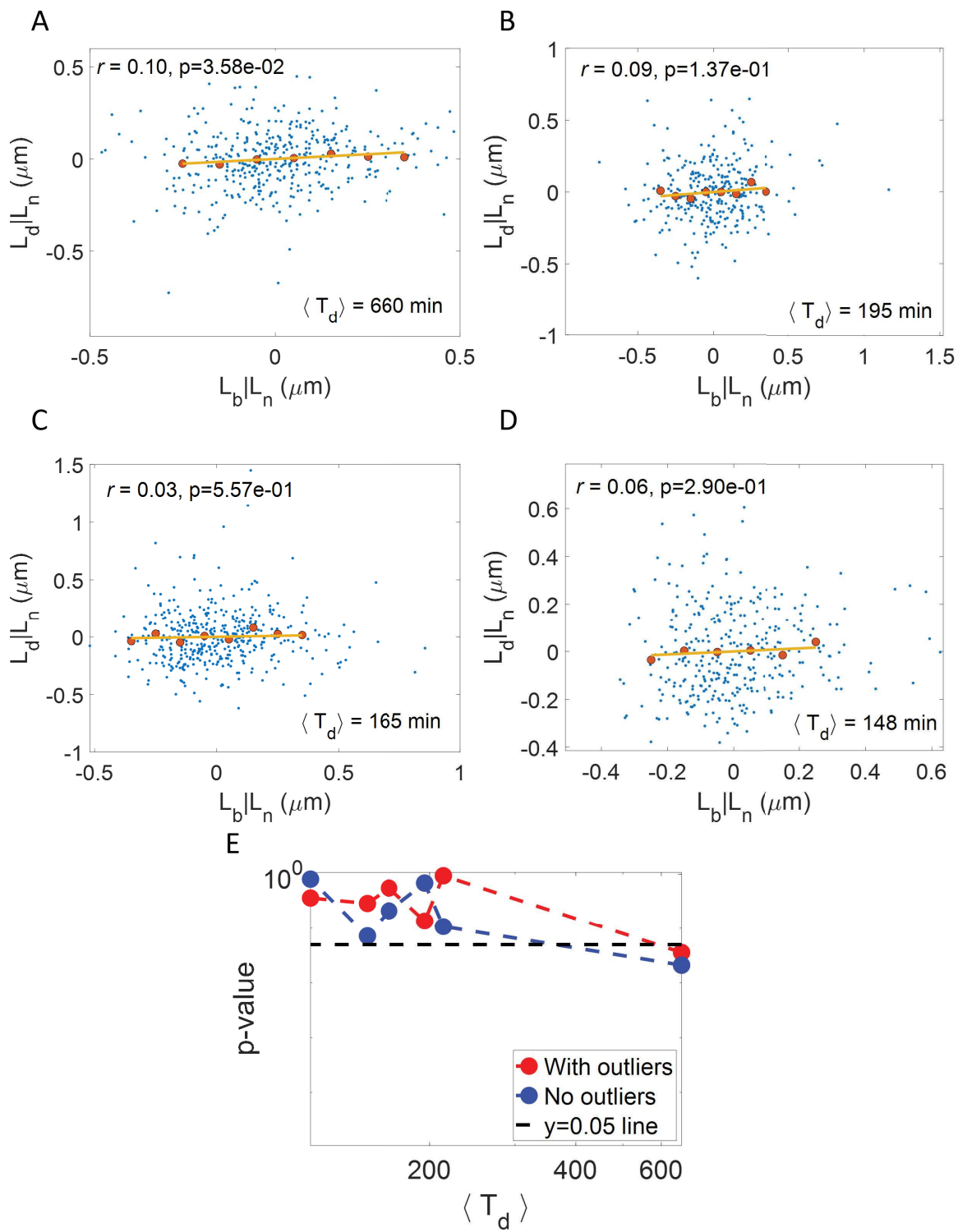
*Corresponding author- email: jmannik@utk.edu; phone: +1 (865) 974 6018

†Corresponding author- email: arielamir@seas.harvard.edu; phone: +1 (617) 495 5818

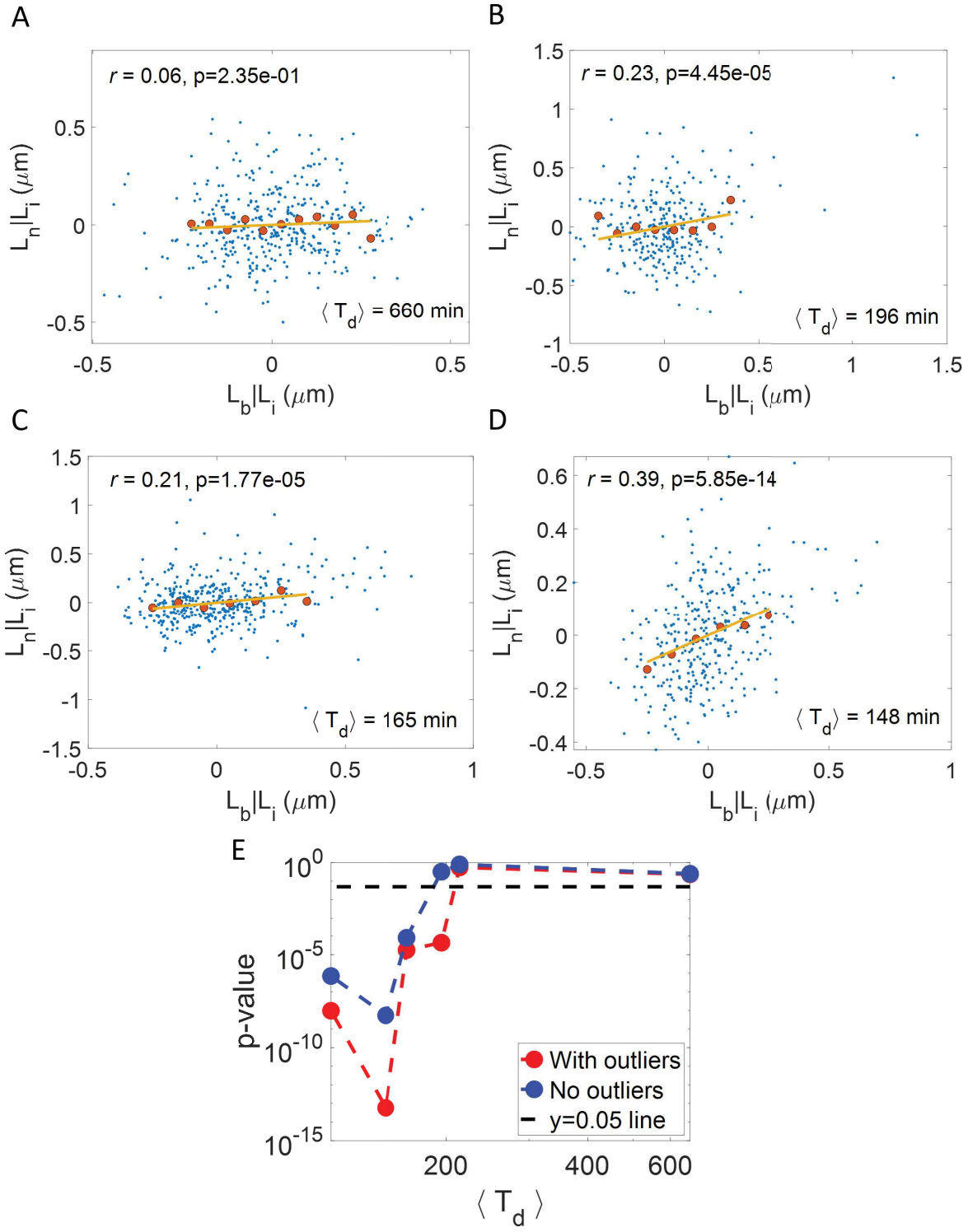
12 Supplementary Figures



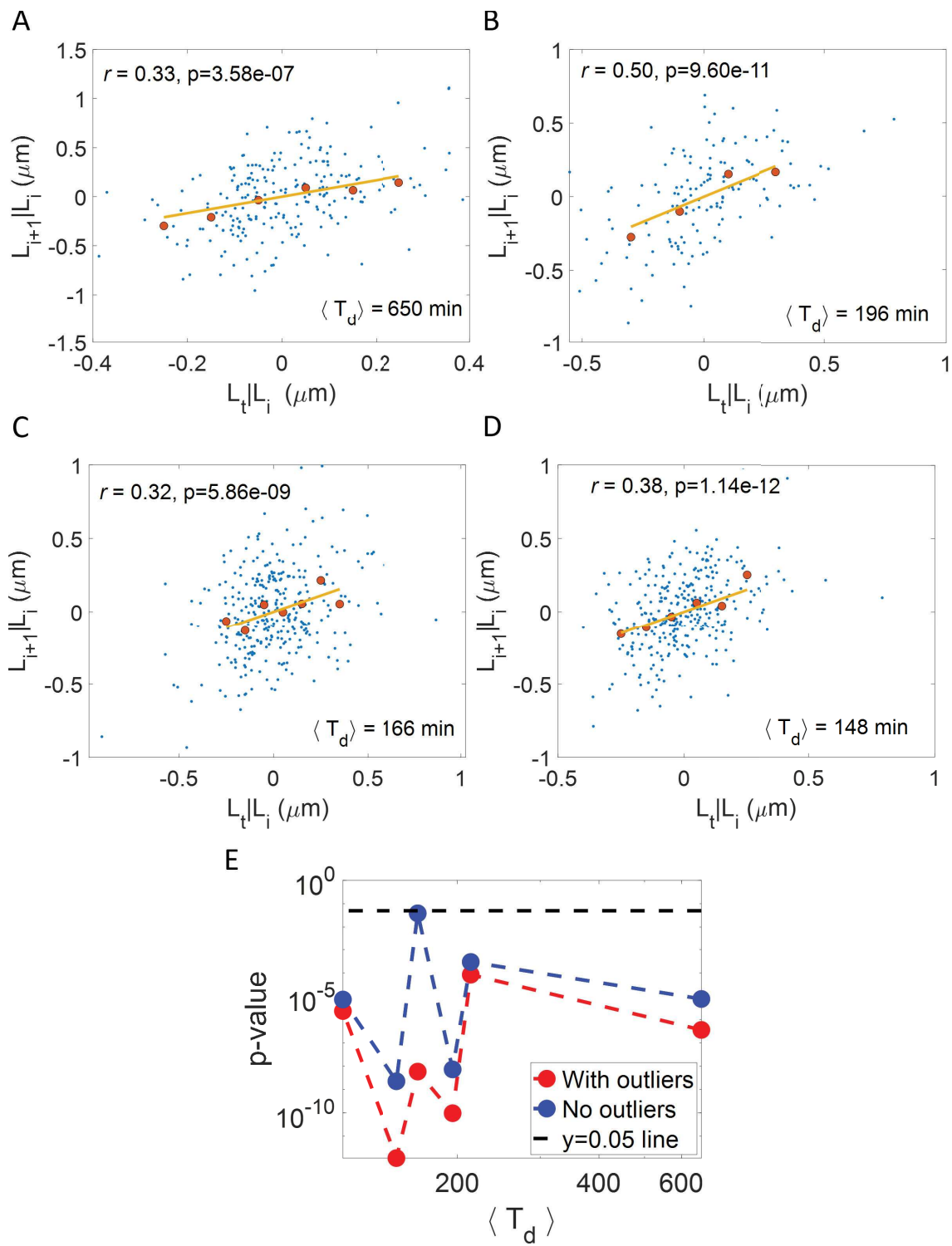
Supplementary Figure 1: **Linking birth, initiation and division: A-D.** Residuals obtained on linear regression of L_d on L_i ($L_d|L_i$) and L_b on L_i ($L_b|L_i$) are plotted for **A.** Acetate medium (generation time = 660 min, $N = 401$ cells, average number of origins at birth, $\langle n_{ori} \rangle = 1$). **B.** Mannose medium (generation time = 196 min, $N = 298$ cells, $\langle n_{ori} \rangle = 1.30$). **C.** Glycerol medium (generation time = 165 min, $N = 419$ cells, $\langle n_{ori} \rangle = 1.33$). **D.** Glycerol+trace elements medium (generation time = 148 min, $N = 344$ cells, $\langle n_{ori} \rangle = 1.60$). The conditional correlation, $r(L_b, L_d|L_i)$ is close to zero for the slowest growth condition (consistent with graph 3A of main text) while the correlations are non-zero for the other conditions shown here (consistent with graph 3B of main text). **E.** Using the data in Ref [S1], we obtain p-values as a function of average doubling time ($\langle T_d \rangle$) for the null hypothesis that the correlation $r(L_b, L_d|L_i)$ is zero and an alternate hypothesis that the correlations are non-zero. Red dots represent the p-values obtained without removing any data points. Blue represents the p-values obtained after the outliers are removed and the data points which are in the middle 95% percentiles of both axes are kept. Dotted line represents the significance level which is set at 0.05.



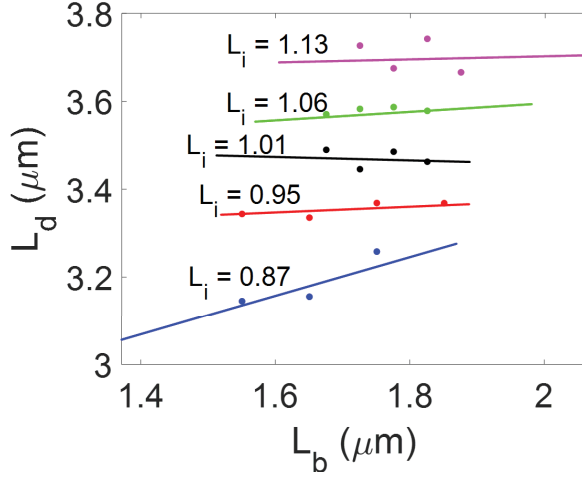
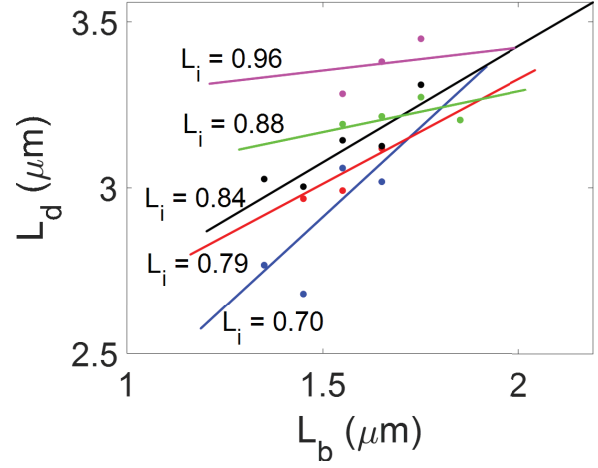
Supplementary Figure 2: **Linking birth, onset of constriction and division: A-D.** Residuals obtained on linear regression of L_d on L_n ($L_d|L_n$) and L_b on L_n ($L_b|L_n$) are plotted for **A.** Acetate medium (generation time = 660 min, $N = 401$ cells, $\langle n_{ori} \rangle = 1$). **B.** Mannose medium (generation time = 195 min, $N = 302$ cells, $\langle n_{ori} \rangle = 1.30$). **C.** Glycerol medium (generation time = 165 min, $N = 420$ cells, $\langle n_{ori} \rangle = 1.33$). **D.** Glycerol+trace elements medium (generation time = 148 min, $N = 344$ cells, $\langle n_{ori} \rangle = 1.60$). The conditional correlations, $r(L_b, L_d|L_n)$ are close to zero for all growth conditions (consistent with graph 4A of main text). **E.** Using the data in Ref [S1], we obtain p-values as a function of average doubling time ($\langle T_d \rangle$) for the null hypothesis that the correlation $r(L_b, L_d|L_n)$ is zero and an alternate hypothesis that the correlations are non-zero. Red dots represent the p-values obtained without removing any data points. Blue represents the p-values obtained after the outliers are removed and the data points which are in the middle 95% percentiles of both axes are kept. Dotted line represents the significance level which is set at 0.05.



Supplementary Figure 3: **Cell cycle regulation model: A-D.** Residuals obtained on linear regression of L_n on L_i ($L_n|L_i$) and L_b on L_i ($L_b|L_i$) are plotted for **A.** Acetate medium (generation time = 660 min, $N = 401$ cells, $\langle n_{ori} \rangle = 1$). **B.** Mannose medium (generation time = 196 min, $N = 298$ cells, $\langle n_{ori} \rangle = 1.30$). **C.** Glycerol medium (generation time = 165 min, $N = 419$ cells, $\langle n_{ori} \rangle = 1.33$). **D.** Glycerol+trace elements medium (generation time = 148 min, $N = 344$ cells, $\langle n_{ori} \rangle = 1.60$). The conditional correlation, $r(L_b, L_n|L_i)$ is close to zero for the slowest growth condition (consistent with graph 5A of main text) while the correlations are non-zero for the other conditions shown here (consistent with graph 5B of main text). **E.** Using the data in Ref [S1], we obtain p-values as a function of average doubling time ($\langle T_d \rangle$) for the null hypothesis that the correlation $r(L_b, L_n|L_i)$ is zero and an alternate hypothesis that the correlations are non-zero. Red dots represent the p-values obtained without removing any data points. Blue represents the p-values obtained after the outliers are removed and the data points which are in the middle 95% percentiles of both axes are kept. Dotted line represents the significance level which is set at 0.05.



Supplementary Figure 4: **Control of replication initiation: A-D.** Residuals obtained on linear regression of L_{i+1} on L_i ($L_{i+1}|L_i$) and L_t on L_i ($L_t|L_i$) are plotted for **A.** Acetate medium (generation time = 650 min, N = 233 cells, $\langle n_{ori} \rangle = 1$). **B.** Mannose medium (generation time = 196 min, N = 146 cells, $\langle n_{ori} \rangle = 1.38$). **C.** Glycerol medium (generation time = 166 min, N = 324 cells, $\langle n_{ori} \rangle = 1.35$). **D.** Glycerol+trace elements medium (generation time = 148 min, N = 320 cells, $\langle n_{ori} \rangle = 1.62$). The conditional correlations, $r(L_t, L_{i+1}|L_i)$ are non-zero for all growth conditions shown here (consistent with graph 6B of main text). **E.** Using the data in Ref [S1], we obtain p-values as a function of average doubling time ($\langle T_d \rangle$) for the null hypothesis that the correlation $r(L_t, L_{i+1}|L_i)$ is zero and an alternate hypothesis that the correlations are non-zero. Red dots represent the p-values obtained without removing any data points. Blue represents the p-values obtained after the outliers are removed and the data points which are in the middle 95% percentiles of both axes are kept. Dotted line represents the significance level which is set at 0.05.

A**B**

Supplementary Figure 5: **A.** Parallel adder model was simulated using parameters obtained from alanine growth medium in Ref [S1]. The simulations were carried out for a single lineage of 400 generations. The data was arranged in ascending order based on the initiation length per origin and divided into 5 subsets with equal number of points in each subset. Each subset has initiation lengths per origin in a small interval centered around L_i . We show the average initiation length per origin, L_i , in each subset in μm . The binned data and best linear fit for each of the subset is plotted. We find that, while most subsets have nearly zero slope (horizontal best linear fit) in agreement with the parallel adder model ($r(L_b, L_d|L_i) = 0$), the smallest initiation length per origin subset deviates from a horizontal line. Such discrepancies make it difficult to narrow down on the model by dividing the datasets into small subsets and using binning. **B.** Concurrent processes model was simulated using parameters obtained from glycerol growth medium in Ref [S1]. The simulations were carried out for a single lineage of 419 generations. The data was again divided into 5 subsets based on the initiation length per origin. The binned data and best linear fit for each of the subset is plotted. We show the average initiation length per origin in each subset in μm . We find that all subsets have a non-zero correlation between L_b and L_d .

17 **Supplementary Tables**

Media	No. of cells	$\langle T_d \rangle$ (min)	$\langle n_{ori} \rangle$	(L_b, L_d)	(L_b, L_i)	(L_i, L_d)
Acetate	401	660	1	0.48 (0.40, 0.55)	0.63 (0.57, 0.68)	0.73 (0.68, 0.78)
Alanine	215	213	1.07	0.55 (0.45, 0.63)	0.76 (0.69, 0.81)	0.72 (0.65, 0.78)
Mannose	298	196	1.30	0.41 (0.31, 0.50)	0.54 (0.46, 0.62)	0.46 (0.37, 0.55)
Glycerol	419	165	1.33	0.37 (0.29, 0.45)	0.49 (0.41, 0.56)	0.44 (0.36, 0.52)
Glycerol + trace elements	344	148	1.60	0.25 (0.15, 0.35)	-0.06 (-0.17, 0.04)	0.37 (0.28, 0.46)
Glucose	259	113	1.98	0.30 (0.18, 0.40)	0.46 (0.36, 0.55)	0.24 (0.12, 0.35)

Table S1: Pearson correlation coefficients along with their 95% confidence intervals (CI) are shown for six different growth media with generation times, $\langle T_d \rangle$. Correlations are found for cell length variables corresponding to cell birth (L_b), initiation of DNA replication (L_i) and cell division (L_d) events.

Media	No. of cells	$\langle T_d \rangle$ (min)	$\langle n_{ori} \rangle$	$(L_b, L_d L_i)$	$(L_b, L_d (L_i, \lambda))$	$(L_i, L_d L_b)$
Acetate	401	660	1	0.03 (-0.07, 0.13)	0.03 (-0.07, 0.13)	0.63 (0.57, 0.69)
Alanine	215	213	1.07	0.01 (-0.13, 0.14)	0.03 (-0.11, 0.16)	0.56 (0.46, 0.64)
Mannose	298	196	1.30	0.22 (0.11, 0.32)	0.37 (0.26, 0.46)	0.31 (0.20, 0.41)
Glycerol	419	165	1.33	0.20 (0.10, 0.29)	0.14 (0.05, 0.24)	0.32 (0.24, 0.41)
Glycerol + trace elements	344	148	1.60	0.30 (0.20, 0.39)	0.29 (0.19, 0.39)	0.40 (0.31, 0.49)
Glucose	259	113	1.98	0.21 (0.09, 0.33)	0.17 (0.05, 0.29)	0.13 (0, 0.24)

Table S2: Pearson correlation coefficients along with their 95% CI are shown for six different growth media. Conditional correlations are found for growth rate (λ), cell birth (L_b), initiation (L_i) and cell division (L_d) events.

Media	No. of cells	$\langle T_d \rangle$ (min)	$\langle n_{ori} \rangle$	(L_b, L_d)	(L_b, L_n)	(L_n, L_d)	$(L_b, L_d L_n)$	$(L_n, L_d L_b)$
Acetate	401	660	1	0.48 (0.40, 0.55)	0.50 (0.43, 0.57)	0.85 (0.82, 0.88)	0.10 (0.01, 0.2)	0.81 (0.77, 0.84)
Alanine	215	213	1.07	0.55 (0.45, 0.63)	0.62 (0.53, 0.69)	0.89 (0.86, 0.91)	-0.01 (-0.14, 0.13)	0.84 (0.79, 0.87)
Mannose	302	195	1.30	0.41 (0.31, 0.50)	0.46 (0.37, 0.54)	0.79 (0.74, 0.83)	0.09 (-0.03, 0.2)	0.74 (0.69, 0.79)
Glycerol	420	165	1.33	0.37 (0.28, 0.45)	0.45 (0.37, 0.52)	0.79 (0.75, 0.83)	0.03 (-0.07, 0.12)	0.75 (0.71, 0.79)
Glycerol + trace elements	344	148	1.60	0.25 (0.15, 0.35)	0.29 (0.19, 0.38)	0.75 (0.70, 0.79)	0.06 (-0.05, 0.16)	0.73 (0.68, 0.78)
Glucose	259	113	1.98	0.30 (0.18, 0.40)	0.47 (0.37, 0.56)	0.70 (0.63, 0.76)	-0.06 (-0.18, 0.07)	0.67 (0.59, 0.73)

Table S3: Pearson correlation coefficients along with their 95% CI are shown for six different growth media. Correlations and conditional correlations are found for cell length variables corresponding to cell birth (L_b), onset of constriction (L_n) and cell division (L_d) events.

Media	No. of cells	$\langle T_d \rangle$ (min)	$\langle n_{ori} \rangle$	(L_i, L_n)	$(L_b, L_n L_i)$	$(L_b, L_n (L_i, \lambda))$	$(L_i, L_n L_b)$
Acetate	401	660	1	0.75 (0.70, 0.79)	0.06 (-0.04, 0.16)	0.06 (-0.04, 0.16)	0.65 (0.59, 0.70)
Alanine	215	213	1.07	0.80 (0.74, 0.84)	0.04 (-0.09, 0.17)	0.05 (-0.09, 0.18)	0.64 (0.55, 0.71)
Mannose	298	196	1.30	0.54 (0.45, 0.61)	0.23 (0.12, 0.34)	0.31 (0.21, 0.41)	0.39 (0.28, 0.48)
Glycerol	419	165	1.33	0.61 (0.55, 0.67)	0.21 (0.11, 0.30)	0.17 (0.07, 0.26)	0.51 (0.43, 0.57)
Glycerol + trace elements	344	148	1.60	0.55 (0.47, 0.62)	0.39 (0.30, 0.48)	0.39 (0.29, 0.47)	0.60 (0.52, 0.66)
Glucose	259	113	1.98	0.42 (0.31, 0.51)	0.35 (0.24, 0.45)	0.32 (0.21, 0.43)	0.26 (0.14, 0.37)

Table S4: Pearson correlation coefficients along with their 95% CI are shown for six different growth media. Correlations and conditional correlations are found for growth rate (λ), cell birth (L_b), initiation (L_i), and onset of constriction (L_n) events.

Media	No. of cells	$\langle T_d \rangle$ (min)	$\langle n_{ori} \rangle$	$(L_i, L_d L_n)$	$(L_n, L_d L_b)$	$(L_n, L_d L_i)$
Acetate	401	660	1	0.27 (0.18, 0.36)	0.81 (0.77, 0.84)	0.67 (0.62, 0.72)
Alanine	215	213	1.07	0.05 (-0.09, 0.18)	0.84 (0.79, 0.87)	0.75 (0.69, 0.81)
Mannose	298	196	1.30	0.07 (-0.4, 0.18)	0.74 (0.68, 0.79)	0.72 (0.66, 0.77)
Glycerol	419	165	1.33	-0.09 (-0.18, 0.01)	0.76 (0.71, 0.79)	0.74 (0.69, 0.78)
Glycerol +trace elements	344	148	1.60	-0.07 (-0.18, 0.03)	0.73 (0.68, 0.78)	0.70 (0.64, 0.75)
Glucose	259	113	1.98	-0.08 (-0.2, 0.04)	0.67 (0.59, 0.73)	0.68 (0.61, 0.74)

Table S5: Pearson correlation coefficients along with their 95% CI are shown for six different growth media. Conditional correlations are found for initiation (L_i), onset of constriction (L_n) and cell division (L_d) events.

Media	No. of cells	$\langle T_d \rangle$ (min)	$\langle n_{ori} \rangle$	$(L_b, L_d (L_i, L_n))$	$(L_i, L_d (L_b, L_n))$	$(L_n, L_d (L_b, L_i))$
Acetate	401	660	1	-0.02 (-0.11, 0.08)	0.25 (0.16, 0.34)	0.67 (0.62, 0.72)
Alanine	215	213	1.07	-0.04 (-0.17, 0.10)	0.06 (-0.07, 0.19)	0.75 (0.69, 0.81)
Mannose	298	196	1.30	0.07 (-0.04, 0.18)	0.04 (-0.08, 0.15)	0.71 (0.65, 0.76)
Glycerol	419	165	1.33	0.07 (-0.03, 0.16)	-0.10 (-0.20, -0.01)	0.73 (0.68, 0.77)
Glycerol +trace elements	344	148	1.60	0.04 (-0.07, 0.14)	-0.06 (-0.16, 0.05)	0.67 (0.60, 0.72)
Glucose	259	113	1.98	-0.03 (-0.15, 0.09)	-0.06 (-0.18, 0.06)	0.66 (0.59, 0.73)

Table S6: Pearson correlation coefficients along with their 95% CI are shown for six different growth media. Conditional correlations when conditioned upon two variables are found for variables involving cell birth (L_b), initiation (L_i), onset of constriction (L_n), and cell division (L_d) events.

Media	No. of cells	$\langle T_d \rangle$ (min)	$\langle n_{ori} \rangle$	(L_i, L_{i+1})	(L_i, L_t)	(L_t, L_{i+1})	$(L_t, L_{i+1} L_i)$	$(L_i, L_{i+1} L_t)$
Acetate	233	650	1	0.58 (0.49, 0.66)	0.86 (0.82, 0.89)	0.64 (0.55, 0.71)	0.33 (0.21, 0.44)	0.09 (-0.04, 0.22)
Alanine	167	212	1.08	0.64 (0.54, 0.72)	0.88 (0.84, 0.91)	0.67 (0.58, 0.75)	0.30 (0.15, 0.43)	0.14 (-0.02, 0.28)
Mannose	146	196	1.38	0.49 (0.36, 0.61)	0.59 (0.48, 0.69)	0.64 (0.54, 0.73)	0.50 (0.37, 0.62)	0.18 (0.02, 0.33)
Glycerol	324	166	1.35	0.45 (0.36, 0.53)	0.71 (0.65, 0.76)	0.52 (0.43, 0.59)	0.32 (0.21, 0.41)	0.14 (0.03, 0.24)
Glycerol +trace ele- ments	320	148	1.62	0.56 (0.48, 0.63)	0.76 (0.71, 0.80)	0.63 (0.56, 0.70)	0.38 (0.29, 0.47)	0.16 (0.05, 0.27)
Glucose	255	112	1.98	0.55 (0.46, 0.63)	0.65 (0.58, 0.72)	0.55 (0.45, 0.63)	0.29 (0.17, 0.40)	0.31 (0.20, 0.42)

Table S7: Pearson correlation coefficients along with their 95% CI are shown for six different growth media. Correlations and conditional correlations are found for variables involving initiation (L_i), termination (L_t), and initiation in the next cell cycle (L_{i+1}) events.

S1 D-separation in the context of cell cycle

In the main text, we use directed acyclic graphs (DAG) to show causal relations. The edges are directed from cause to effect. Two vertices in the graph are connected by a path when there is a sequence of distinct vertices with an edge between them. We apply d-separation [S2, S3] to DAGs (Figures 3-6 of the main text) to predict correlations and conditional correlations. In this section, we will discuss in detail several examples of predicting the correlations and conditional correlations using d-separation.

Consider the graph in Figure 5B of the main text. We will choose two variables and check whether they are correlated or not when we condition upon other variables.

- L_b and L_d - There are two paths between L_b and L_d , path 1 - $L_b \rightarrow L_n \rightarrow L_d$, path 2 - $L_b \leftarrow L_{i-1} \rightarrow L_i \rightarrow L_n \rightarrow L_d$.
 - Without conditioning - Both paths 1 and 2 are open as there is no collider. So, L_b and L_d are d-connected and correlated.
 - Conditioning on L_i - Path 2 is blocked as we conditioned on a non-collider. However, path 1 is still open as we are not conditioning on any variables on the path. Hence, L_b and L_d are still d-connected and correlated.
 - Conditioning on L_n - Path 1 and 2 are both blocked as we are conditioning on the non-collider L_n . Hence, L_b and L_d are d-separated and uncorrelated.
 - Conditioning L_i and L_n - Path 1 and 2 are both blocked as we are conditioning on non-colliders L_i and L_n . Hence, L_b and L_d are d-separated and uncorrelated.

S2 Length is used to denote cell cycle events

In this section, we will discuss why cell lengths (L) and not the corresponding timings (T) are used as a proxy to denote the cell cycle events. We will illustrate this on a concrete

41 example, and then discuss its generalization.

42 Consider events X and Y in the cell cycle, assuming that Y occurs after a constant length
43 addition from event X (i.e., we are assuming an adder model). A possible mechanistic mech-
44 anism for this phenomenological model is the accumulation of an initiator protein starting
45 from X [S4]. We assume that the protein amount (P) when event X happens is zero and it
46 undergoes balanced biosynthesis i.e. $\frac{dP}{dL}$ is constant. The event Y happens when a threshold
47 amount of P has been reached. Mathematically, length at event Y (L_y) is related to length
48 at event X (L_x) by,

$$L_y = L_x + L_{xy} + \eta_{xy}, \quad (S1)$$

49 where L_{xy} is the average size added between X and Y and η_{xy} is a size additive noise
50 independent of L_x . The DAG for the structural causal model (SCM) in Eq. S1 is shown in
51 Figure S1A-1. Assuming exponential growth with rate λ and the adder model, the timing
52 of event Y (T_y) is related to T_x and L_x as,

$$T_y = T_x + \frac{1}{\lambda} \ln\left(1 + \frac{L_{xy} + \eta_{xy}}{L_x}\right). \quad (S2)$$

53 Therefore, we find the timing of the event Y is determined by the timing of the event X (T_x)
54 and the length at event X (L_x). The timing of the events X and Y have a relation as shown in
55 Figure S1A-2 where T_x and L_x both influence when Y happens. If X was also determined by
56 an adder, the timing of events (T_x and T_y) are associated with each other via a direct causal
57 link as well as through cell lengths (L_x). Thus, graphs involving the timing of events will
58 also need to include cell lengths. More generally, the DAGs in Figure S1A are identical when
59 the length at Y is determined by a general regulatory mechanism, $L_y = \alpha_x L_x + L_{xy} + \eta_{xy}$
60 (the adder model for Y corresponds to the particular case $\alpha_x = 1$ [S5]).

61 Next, we will consider the timer model where Y happens after an average time T_{xy} of
62 event X. A possible underlying mechanism is that a biochemical process starts at event X

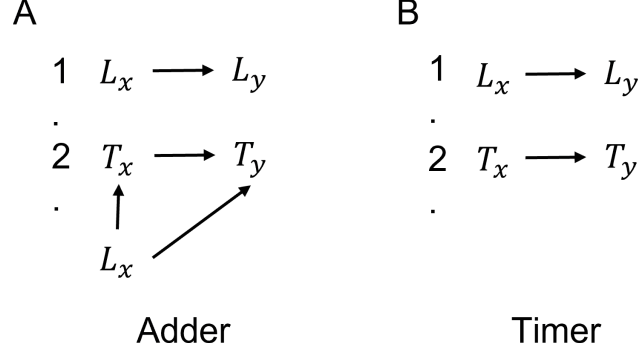


Figure S1: **DAGs for the adder and timer model: A-B.** We show DAGs involving the sizes (in graph 1) and the timings (graph 2) at cell events X, and Y for **A**. An adder model. **B**. A timer model.

and proceeds at a constant rate. In this case, the timing of event Y is,

$$T_y = T_x + T_{xy} + \eta_{t,xy}, \quad (\text{S3})$$

where T_x is the timing of event X and $\eta_{t,xy}$ is the time additive noise. The DAG for the SCM is an arrow from T_x to T_y (Figure S1B-2). Assuming exponential growth, the length at event Y (L_y) is related to length at event X (L_x) as,

$$L_y = L_x e^{\lambda(T_{xy} + \eta_{t,xy})}. \quad (\text{S4})$$

L_x is independent of $\eta_{t,xy}$ in the timer model. If L_x is independent of growth rate (λ), the DAG involving the lengths, L_x and L_y will be as shown in Figure S1B-1.

Therefore in both adder and timer like models, causal relations between events cannot be solely represented using their timings (Figure S1A-2) but they can be solely denoted by their lengths (Figures S1A-1, S1B-1).

Recent experiments on *E. coli* have shown that single cell lengths grows super-exponentially (faster than exponential growth) [S6, S7]. Next, we discuss whether using lengths to denote cell cycle events is appropriate in case of super-exponential growth.

Consider event Y was determined by a timer from X. Assuming super-exponential growth, the lengths at X and Y are related as,

$$L_y = L_x e^{\int_{T_x}^{T_y} \lambda(t') dt'}. \quad (\text{S5})$$

$\lambda(t)$ shows the variation of growth rate with time. The lengths at events X and Y following Eq. S5 cannot be represented by DAGs containing just the lengths of events X and Y. The causal diagrams might also have to include growth rate parameters which are not directly observed in the experiments.

However, for an adder between X and Y, the lengths at events X and Y will be related by Eq. S1 assuming balanced biosynthesis. The resulting DAG for the SCM is identical to that for exponential growth (Figure S1A-1). Thus, cell lengths seem to be the appropriate cell characteristic to represent the cell events in many biologically relevant cases.

S3 Representing cell cycles as causal graphs

In this section, we will show the complete causal graphs of various cell cycle models discussed in the main text.

Causal graphs discussed in the main text are assumed to follow the Causal Markov condition which states that, when conditioned upon all direct causes, the nodes of a causal graph are independent of its non-descendants. In causal graphs which follow the Causal Markov assumption, all variables which are the common causes of the variables in the graphs must also be in that graph [S3]. Note that all common causes for any pair of variables in graphs 3A-3C, 4A-4B, 5A-5B and 6A-6B of the main text are already included in the graphs.

Next, we will discuss the recursive nature of the causal graphs over multiple generations.

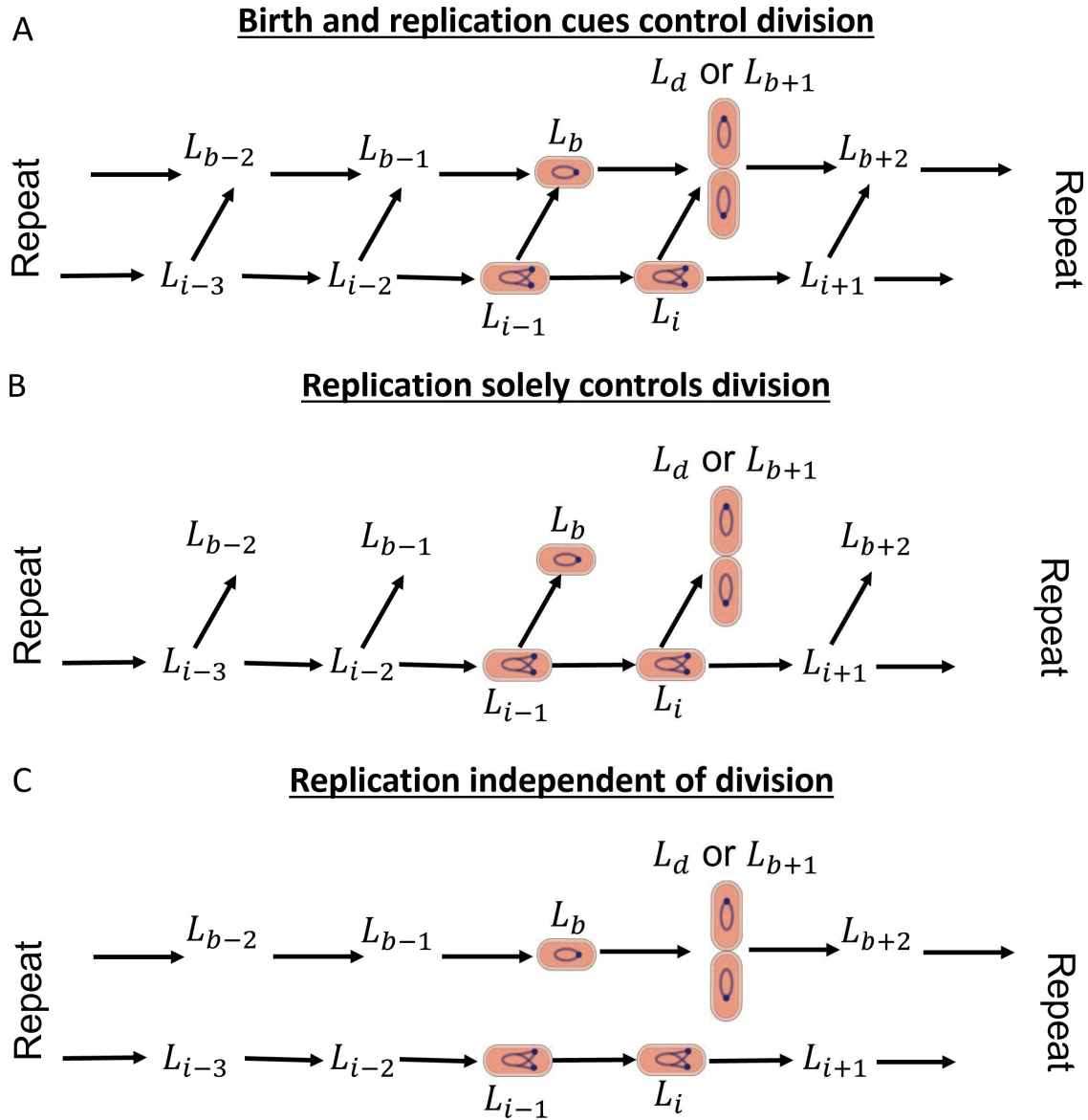


Figure S2: **A-C**. Causal graphs are shown spanning multiple generations. **A**. In this model, division is controlled by both birth and replication related processes. This is an extension of graph 3B in the main text. **B**. Division is solely controlled by replication. This is an extension of graph 3A in the main text. **C**. Division occurs independent of replication. This is an extension of graph 3C in the main text.

96 In the causal graph where birth and replication both control division or birth event of next
 97 cell cycle (Figures 3B and 5B of the main text), we do not show birth in the previous cell
 98 cycle (L_{b-1}) as a cause of L_b . Omitting L_{b-1} from graphs 3B and 5B does not change our
 99 predictions for the conditional correlation between the variables in those graphs because L_{b-1}
 100 is not a common cause of any pair of variables in the graph. Here, however, we will extend
 101 causal graphs 3B and 5B to include both causes of L_b i.e., L_{b-1} and L_{i-1} are included in the
 102 causal graphs.

103 We show the graph where both birth and replication control division in Figure S2A
 104 (extension of graph 3B in the main text). If we replace L_d by L_n which then causes L_d , we
 105 will get an extension for the graph 5B in the main text. In Figure S2A, we have a causal link
 106 from L_{b-1} to L_b in addition to $L_{i-1} \rightarrow L_b$ link. Upon including both L_{b-1} and L_{i-1} into the
 107 graph, we have to include its common cause - initiation previous to that of L_{i-1} i.e., L_{i-2} .
 108 L_{b-1} is also controlled by two events - L_{i-2} and previous birth event L_{b-2} . Thus, we obtain
 109 a recursive pattern which is shown in Figure S2A. For solely replication limited division, we
 110 show the causal graph in Figure S2B where birth size j generations before the current cell
 111 cycle (L_{b-j}) does not influence the birth size directly in the next cell cycle (L_{b-j+1}). Graph
 112 S2C shows a model where the division cycle is independent of the replication cycle.

113 **S4 Conditional independence tests on synthetic data**

114 Kar *et al.* showed that data analysis methods should be validated against synthetic data
 115 before being applied to experimental data [S6]. This prevents ambiguity and provides con-
 116 sensus about the use of the method. In this section, we validate the conditional independence
 117 tests using synthetic data generated by existing models.

118 We simulated a lineage of 1000 generations using the parallel adder (PA) model with
 119 exponentially growing single cells and perfectly symmetric division. According to the PA

120 model, the division event happens upon addition of constant size per origin from replication
 121 initiation. The DNA replication initiates upon adding a constant cell length per origin
 122 from the previous initiation [S8, S9]. Figure S3A shows the L_d vs L_b plot obtained from
 123 simulations of the PA model. The best linear fit is very close to the equation $L_d = L_b + \Delta L$.
 124 A similar equation is also obtained for simulations of concurrent processes model where single
 125 cells are undergoing exponential growth and perfectly symmetric division (Figure S3C). In
 126 this model, division is limited by slower of the two processes - 1. constant size addition on
 127 average from birth (adder) 2. a time C+D elapses from initiation of DNA replication (with
 128 both processes subject to noise). The replication initiation is controlled in the same manner
 129 as in the PA model.

130 In the main text, we showed that the conditional correlation $r(L_b, L_d|L_i)$ can be used to
 131 distinguish between two classes of model- 1. replication initiation solely controls division size
 132 shown in graph 3A (e.g. - PA model) and 2. birth and replication simultaneously control
 133 division as shown in graph 3B (e.g. - concurrent process model). Using d-separation, we
 134 predict L_b and L_d to be uncorrelated on fixing L_i in graph 3A. However, they are predicted
 135 to be correlated in graph 3B.

136 Next, we use the synthetic data to test the prediction that conditional correlation between
 137 L_b and L_d on fixing L_i ($r(L_b, L_d|L_i)$) is zero for the PA model and non-zero for the concurrent
 138 process model. We find $r(L_b, L_d|L_i)$ to be close to zero in the synthetic data generated
 139 using the PA model and the p-value to be not statistically significant at significance level
 140 of 0.05 (Figure S3B). This is consistent with our prediction of $r(L_b, L_d|L_i) = 0$ for the PA
 141 model. We also show the non-zero conditional correlation $r(L_b, L_d|L_i)$ for simulations of the
 142 concurrent process model (Figure S3D). The conditional correlations are in agreement with
 143 our predictions made using the directed acyclic graphs and d-separation. Hence, conditional
 144 independence tests can be used to differentiate between cell cycle models.

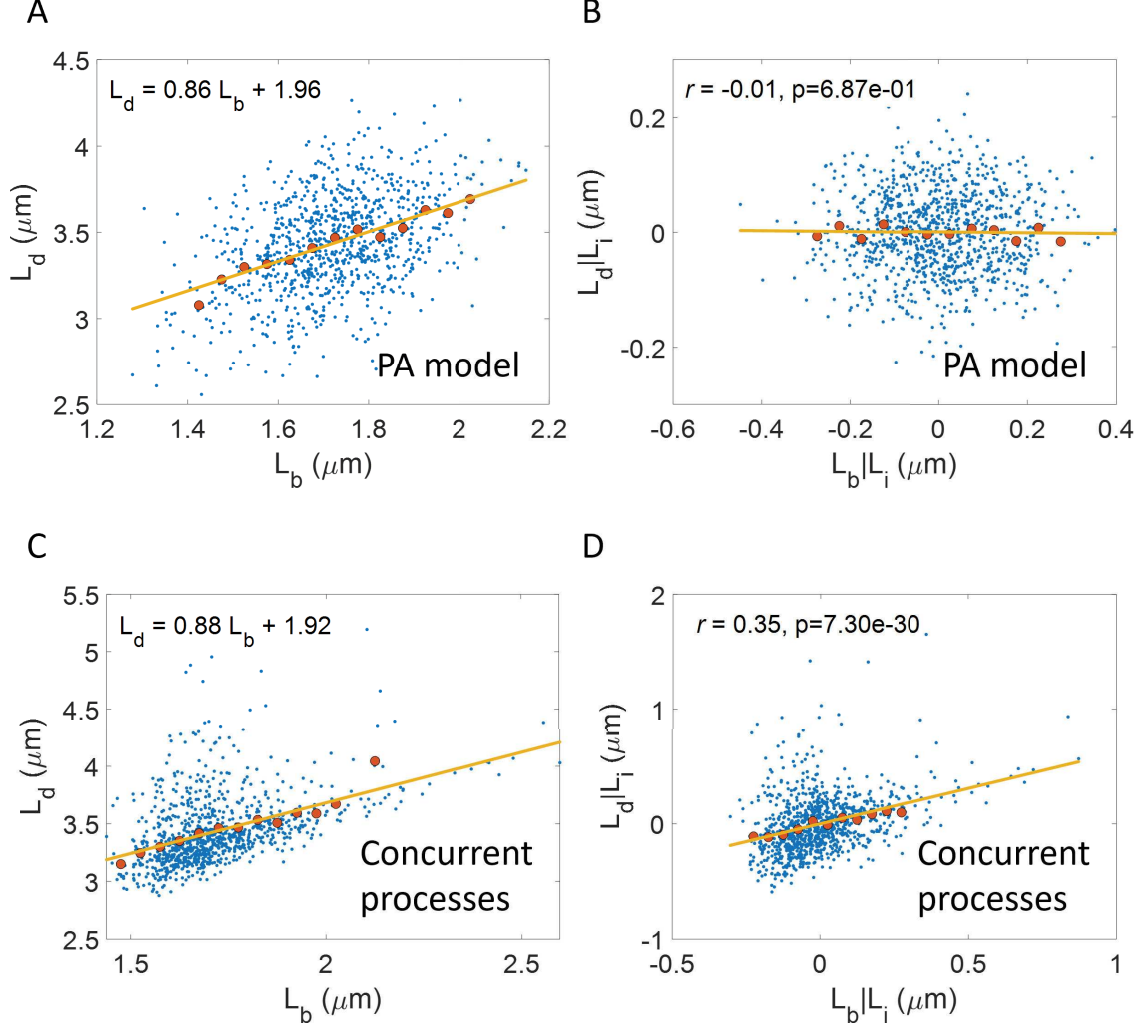
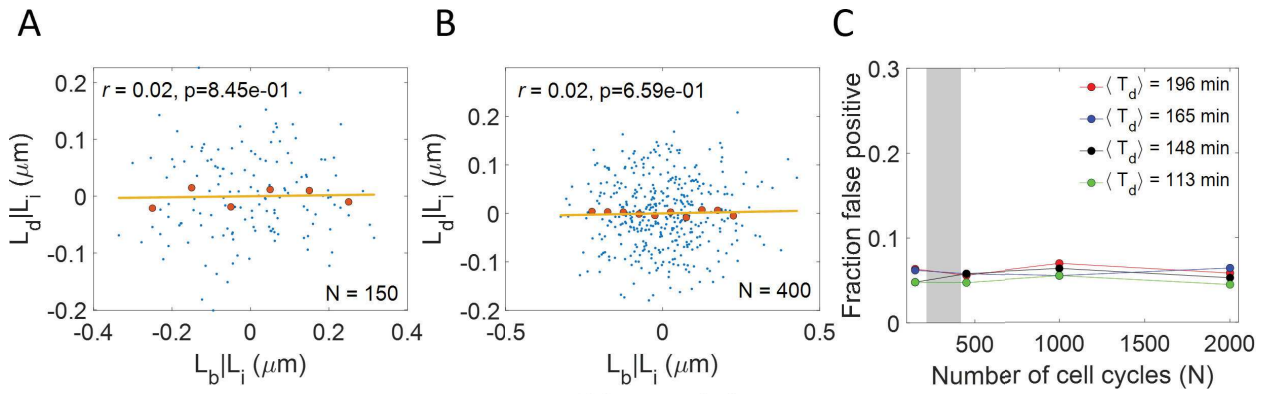
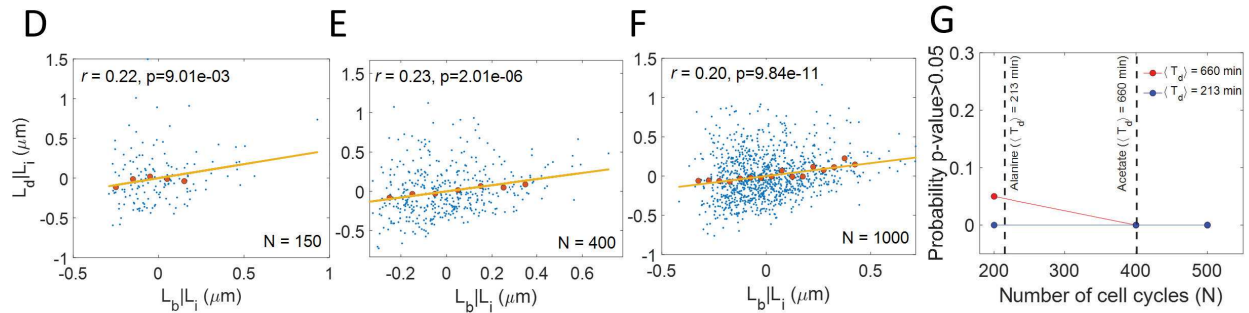


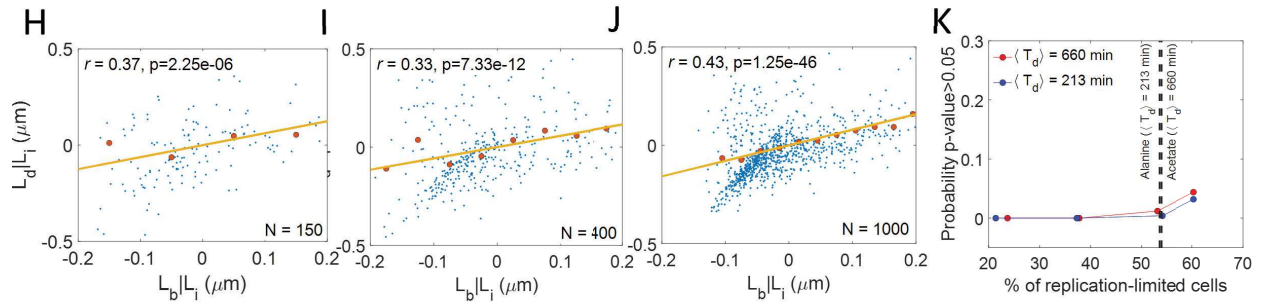
Figure S3: **Tests on synthetic data:** **A-B.** Simulations of cells undergoing exponential growth and following a parallel adder model are carried out and data is collected for 1000 cell cycles. For the synthetic data generated, we show **A.** L_d vs L_b plot. **B.** Residuals obtained on linear regression of L_d on L_i ($L_d|L_i$) and L_b on L_i ($L_b|L_i$) are plotted. The correlation, $r(L_b, L_d|L_i)$ is close to zero. **C-D.** Simulations of cells undergoing exponential growth and following a concurrent process model are carried out and data is collected for 1000 cell cycles. For the synthetic data generated, we show **C.** L_d vs L_b plot. **D.** Residuals obtained on linear regression of L_d on L_i ($L_d|L_i$) and L_b on L_i ($L_b|L_i$) are plotted. The correlation, $r(L_b, L_d|L_i)$ is non-zero. Here, the blue dots represent the raw data, the red dots represent the binned data and the yellow line represents the best linear fit.



PA model



Concurrent processes : Replication limiting in 75% cells



Concurrent processes : Replication limiting in 25% cells

145

146

Figure S4: **Testing conditional independence tests: A,B.** Simulations of cells undergoing exponential growth and following a parallel adder model are carried out and data is collected for **N = A. 150 cells. B. 400 cells.** We find that even for the smaller datasets, our method of calculating conditional correlations is consistent with the predictions obtained using d-separation (zero for PA model). **C.** Parallel adder was simulated for different growth medium (different colors) over a single lineage of **N** generations. For the 2000 iterations carried out, we found the fraction of cases where $r(L_b, L_d|L_i)$ was non-zero (p-value was less than 0.05). This fraction (fraction false positive) is plotted for varying values of **N**. The shaded region shows the range of **N** for the faster growth conditions in Ref [S1]. **D-F, H-J.** Simulations of cells undergoing exponential growth and following a concurrent process model are carried out for **N** cells. We simulate a concurrent processes model where the replication related processes are limiting for division in **x%** cells. We plot $L_d|L_i$ vs $L_b|L_i$ for **D. N = 150, x = 75%. E. N = 400, x = 75%. F. N = 1000, x = 75%. H. N = 150, x = 25%. I. N = 400, x = 25%. J. N = 1000, x = 25%.** In all of the plots, $r(L_b, L_d|L_i)$ are non-zero irrespective of the amount of data (**N**) and the strength of causal link between replication initiation and division (**x**). **G, K.** For the concurrent processes model with parameters chosen using the acetate and alanine growth condition in Ref [S1], **G.** Probability that the p-value is greater than 0.05 is plotted as a function of number of cell cycles (**N**). The dotted lines mark the values of **N** in acetate and alanine growth medium of Ref [S1]. **K.** Probability that the p-value is greater than 0.05 is plotted as a function of % of cells in which the replication process is limiting. The two dotted lines denote the quantity in case of experiments in acetate and alanine growth media [S1].

147 In the simulations in Figure S3, the correlation $r(L_b, L_d|L_i)$ was obtained for **N = 1000**
148 cells. However, in the experiments analyzed in the main text of the paper, the value of **N** is
149 between 150 and 400 cells. We plot $L_d|L_i$ vs $L_b|L_i$ for the simulations of PA model with **N**
150 = 150, and 400 cells in Figures S4A, and S4B, respectively. The correlations are negligible
151 and the p-values are not statistically significant (significance level $\alpha = 0.05$) in agreement
152 with the predictions of PA model.

153 Next, we quantify the accuracy of rejecting the parallel adder model using the conditional
154 independence tests. In our paper, we use p-values to classify a correlation as zero or non-
155 zero. Under the null hypothesis that $r(L_b, L_d|L_i)=0$ with the alternate hypothesis being
156 $r(L_b, L_d|L_i) \neq 0$, we reject the null hypothesis if the p-value is significant (less than the
157 threshold, $\alpha = 0.05$). In such a case, we classify the data to follow a model where both birth
158 and replication related processes are likely influencing the division event (e.g., the concurrent

process model). If the p-value is greater than 0.05, we cannot reject a model where replication solely limits division. Note that a p-value greater than 0.05 does not imply accepting the null hypothesis. In other words, a concurrent process model can also have a p-value greater than 0.05. Later, we will provide an estimate of the chances that a concurrent process model has p-value greater than 0.05. However, first, we will use simulations of the parallel adder model over a single lineage of N generations. We repeat the simulations over 2000 iterations and find the number of cases where we reject the parallel adder model (p-value < 0.05). The fraction of cases where the p-value is less than 0.05 for the parallel adder model is our error metric (false positive error). If the fraction of false positive is high, then there are greater chances of rejecting the parallel adder model and incorrectly classifying it as a concurrent process model.

We carried out the simulation of the parallel adder model for a varying number of cell cycles N for all growth rates in Ref [S1] where we find a non-zero $r(L_b, L_d|L_i)$. For the 2000 iterations of the parallel adder model, we calculate the p-value and compare it to the significance threshold. We explain the calculation of p-value briefly. The p-value is the probability that the test statistic has a value as extreme as the one we find using the data. The test statistic in our case is $r\sqrt{\frac{N-2}{1-r^2}}$, where r is the sample Pearson correlation coefficient which has a variance of $\sqrt{\frac{1-r^2}{N-2}}$. The test statistic is assumed to follow a t-distribution with $N-2$ degrees of freedom under the null hypothesis that the actual underlying correlation is 0. From the definition of significance threshold (set at 0.05 in our case) which is the probability of rejecting models when the null hypothesis is true (in this case, $r(L_b, L_d|L_i) = 0$), we expect the false positive error to be 5%. The calculation of a p-value assumes that the correlation is found for two normally distributed variables [S10]. However, if the normality assumption of the data does not hold one might expect an error different from the expected 5%. Such deviations from normality in cell cycle variables might arise when simulating a model of exponentially growing cells with time additive noise. Therefore, we use simulations to show

that the error is still close to 5% even in case of time additive noise (low noise regime). Using the p-values in the 2000 iterations, we find the fraction of false positive cases for different N to be $\approx 5\%$ in all growth conditions (Figure S4C). So, to conclude, the significance threshold sets the error rate of rejecting a replication controlled division model (e.g., parallel adder) even if it is the actual underlying model.

Similar to the parallel adder model, we checked that the correlations $r(L_b, L_d|L_i)$ are non-zero for different N in the case of concurrent processes model. For simulations of the concurrent processes model, we find $r(L_b, L_d|L_i)$ to be non-zero when N = 150 cells (Figures S4D, S4H) and N=400 cells (Figures S4E, S4I). We also estimated the fraction of cases in which the p-value is greater than 0.05 when the underlying model is concurrent processes model. Note that the null and alternate hypothesis is the same as before. We simulated 200 iterations of the concurrent processes model with parameters chosen using the experimental data in the alanine ($T_d = 213$ min) and acetate ($T_d = 660$ min) growth media of Ref [S1]. We chose these slow-growth conditions because we find p-values > 0.05 (null hypothesis: $r(L_b, L_d|L_i)=0$, alternate: $r(L_b, L_d|L_i) \neq 0$) in these growth conditions. In Figure S4G, we show for a varying N that there are nearly zero cases where the p-value >0.05 . The values of N in the slower growth conditions of Ref [S1] are marked as dotted lines. While p-values greater than 0.05 does not imply that the underlying model is replication solely controls division, we show using simulations that it is unlikely to be a model where both birth and replication related processes control division.

In the case of concurrent processes model, $r(L_b, L_d|L_i)$ is non-zero because there is a direct causal link between L_b and L_d (see Figure 3B in main text). The value of $r(L_b, L_d|L_i)$ will also depend on the strength of this causal link: making replication related processes more limiting for the division event compared with the birth-related processes (i.e., they limit division in a larger fraction of cells) will lead to a smaller value of $r(L_b, L_d|L_i)$. We wanted to test that our method of calculating conditional correlation behaves as expected

on changing the strength of the causal links. To make the replication process more limiting for division, we change the parameters of the concurrent processes model i.e., decrease the length added between birth and division. We find that $r(L_b, L_d|L_i)$ is still non-zero even if replication is the limiting process in 75% of cells (Figures S4D-S4F). As carried out previously for varying N , we calculate the probability that the p-values are greater than 0.05 for varying strengths of causal links in the concurrent processes model. We simulate 250 iterations of the concurrent processes model with parameters chosen using the experimental data in the alanine ($T_d = 213$ min) and acetate ($T_d = 660$ min) growth media of Ref [S1]. We control the % of cells where replication limits division by varying the size added between birth and division as explained previously. Assuming that the cells growing in slow-growth conditions in Ref [S1] follow the concurrent processes model, we can also estimate the % of cells where replication controls division in the case of experiments [S11]. The experimental values are shown as dotted lines for the alanine and acetate growth medium. We find that the fraction of cases where the p-value is greater than 0.05 is small for a wide range of values (Figure S4K). Thus, the underlying model is unlikely to be a concurrent processes model for a p-value greater than 0.05 (null hypothesis: $r(L_b, L_d|L_i)=0$, alternate: $r(L_b, L_d|L_i) \neq 0$).

To conclude, we show that the conditional independence tests can be applied to experimental data even if the number of cells, N is relatively small in the dataset (≈ 150). The conditional correlations obtained were found to be consistent with our predictions from d-separation even when the causal link between two cell cycle events was weak.

S5 Consistency with published results

In this section, we apply conditional independence tests to already published datasets and compare the results to that obtained in the main text. A significant difference between the datasets analyzed here and that in the main text is that the onset of constriction is

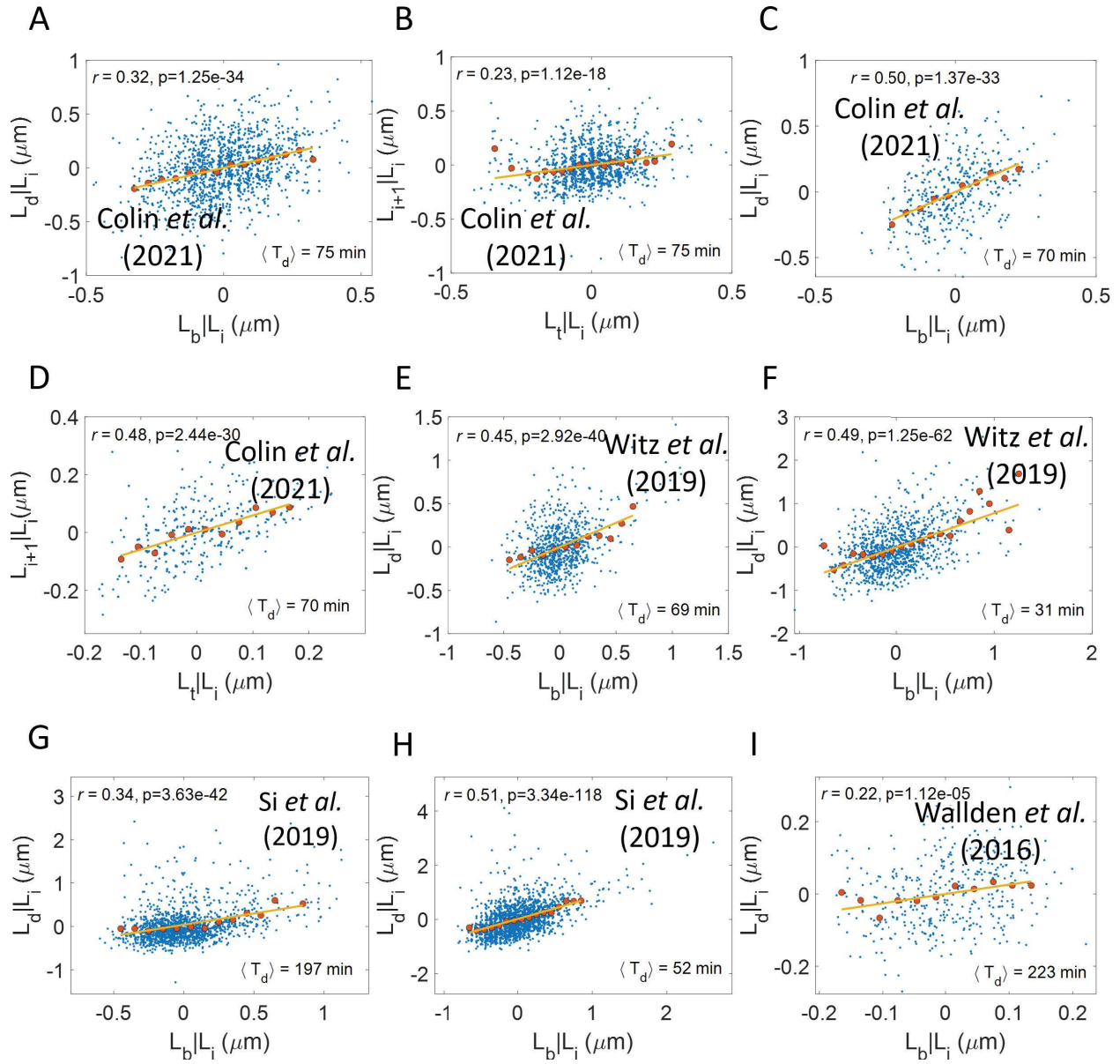
not measured in these datasets. Thus, we cannot examine cell cycle models with the onset of constriction as a checkpoint. However, we can still test the predictions of PA and Cooper-Helmstetter (CH) model (Figure 3A), the concurrent processes model (Figure 3B) and the adder per origin between initiations (Figures 6A-6B in main text). We use the datasets published in Ref [S12] because it contains cell length data at replication initiation and termination events, which we use in Figure 6 of the main text.

In Ref [S12], *E. coli* cells were grown in microfluidic devices and the single-cell characteristics at cell replication and division were measured for multiple cells. The cells were grown in minimal media (M9+NH₄Cl+glycerol) with an average doubling time, $\langle T_d \rangle = 75$ min. This growth condition is comparable to the faster growth conditions in the main text (note that the doubling time for experiments in the main text is roughly twice that of presented here as those were conducted at 28°C). Using the data in this growth condition, we will test whether the replication process is the sole limiting process controlling division. In the main text, this class of models is represented in Figure 3A. A competing model is the concurrent processes model where multiple processes from birth, and replication initiation control division (Figure 3B). We predict $r(L_b, L_d|L_i)$ to be zero for Figure 3A and it is non-zero for the class of models represented by Figure 3B. Using experimental data, we find a non-zero $r(L_b, L_d|L_i)$ as shown in Figure S5A. This is in agreement with the model proposed in the main text as well as Ref [S12]. Next, we test if the initiation in the next cell cycle is controlled solely by initiation in the current cell cycle. The two competing models proposed are presented in Figures 6A and 6B. For the model with adder per origin between initiations as the sole control for initiations, we expect $r(L_t, L_{i+1}|L_i)$ to be zero (Figure 6A). Using the experimental data in Ref [S12], we find that $r(L_t, L_{i+1}|L_i)$ is non-zero (Figure S5B) which is in agreement with our results in the main text. We obtain same qualitative results for experimental replicates. Thus, DNA replication initiation is controlled by additional processes apart from replication initiation in the previous cell cycle.

To test the concurrent processes model, the birth related processes were made more limiting by increasing the D period (time between replication termination and division) in Ref [S12]. Cells were treated with sub-inhibitory concentrations of MreB-polymerization inhibitor A22 which led to an increase in the width of cells and also an increase in D period [S13]. We tested the replication control over division and the adder per origin control between initiations for A22 treated cells (concentration = 50 $\mu g/mL$). We found $r(L_b, L_d|L_i)$ to be non-zero (Figure S5C) and greater than that of untreated cells (Figure S5A), thus, favoring a concurrent processes model where birth related processes are more limiting for division. We also found $r(L_t, L_{i+1}|L_i)$ to be non-zero (Figure S5D), in agreement with the results in the main text and also for untreated cells. The concurrent processes model for division and additional processes apart from DNA replication start controlling the next initiation were also consistent with the data obtained from different concentrations of A22 treated cells.

We also analyzed datasets published in Ref [S9] and Ref [S14] where *E. coli* cells were grown in microfluidic devices. In these datasets, DNA replication termination was not marked but the length at birth, the length at replication initiation and the length at division were collected. Using these data, we could test the replication control over division. For the experiments in Ref [S9], $r(L_b, L_d|L_i)$ was found to be non-zero (Figures S5E-S5F). The growth conditions in these datasets were comparable to the faster growth conditions in the main text which also showed a non-zero $r(L_b, L_d|L_i)$ and was consistent with the concurrent processes model. Upon analyzing the datasets in Ref [S14], we find a non-zero $r(L_b, L_d|L_i)$ for both faster growth condition (Figure S5H) and slower growth condition (Figure S5G). The non-zero $r(L_b, L_d|L_i)$ in fast growth conditions is consistent with our results in main text and the experiments analyzed in this section. For the slow growth condition shown in Figure S5G, we find that $r(L_b, L_d|L_i)$ is lower in value than that of the faster growth condition (Figure S5H), with the binned relation showing a nearly flat region in the regime where the data is most abundant. We also compared the correlations in Figure S5G

287 to the data on slow-growing cells in Ref [S15]. Both experiments were conducted at 37°C
288 with similar *E. coli* strains. We find for the slower growth condition ($T_d = 223$ min) in
289 Figure S5I, the correlation $r(L_b, L_d|L_i)$ is lower in value than that in Figure S5G ($T_d = 197$
290 min). Note that a bias in the data in Ref [S15] because the initiation is always in the same
291 cell cycle as division (or $C+D < T_d$) might explain the non-zero $r(L_b, L_d|L_i)$ in slow-growth
292 conditions. To conclude, birth related processes are less limiting for determining division in
293 slower growth conditions in agreement with our results in the main text.



294

295

Figure S5: **Consistency with published results:** We analyze conditional correlations using previously published datasets on *E. coli*. **A,B.** Using data from Ref [S12] for $N = 1380$ cells ($\langle n_{ori} \rangle = 1.30$) growing in minimal media (M9+NH4Cl+glycerol), we plot **A.** $L_d|L_i$ vs $L_b|L_i$. We obtain a non-zero $r(L_b, L_d|L_i)$ consistent with the concurrent processes model. **B.** $L_{i+1}|L_i$ vs $L_t|L_i$. We obtain a non-zero $r(L_t, L_{i+1}|L_i)$ consistent with the predictions of Graph 6B in the main text. This rules out adder per origin between initiations as the sole control for DNA replication initiation. **C,D.** We use data from Ref [S12] where cells are treated with $50\mu g/mL$ of A22, a MreB polymerization inhibitor. These cells have a larger D period. We plot for $N = 506$ cells ($\langle n_{ori} \rangle = 1.98$), **C.** $L_d|L_i$ vs $L_b|L_i$. We obtain a non-zero $r(L_b, L_d|L_i)$ which is again consistent with the concurrent processes model. **D.** $L_{i+1}|L_i$ vs $L_t|L_i$. The non-zero $r(L_t, L_{i+1}|L_i)$ also rules out adder per origin between initiations being the sole control for initiation. **E-F.** Data was obtained from Ref [S9] and $L_d|L_i$ vs $L_b|L_i$ was plotted. Cells were grown in **E.** glycerol ($N = 777$ cells, $\langle n_{ori} \rangle = 1.7$). We obtain a non-zero $r(L_b, L_d|L_i)$ consistent with the concurrent processes model. **F.** glucose and eight amino acids ($N = 1039$ cells, $\langle n_{ori} \rangle = 2$). We also obtain a non-zero $r(L_b, L_d|L_i)$ consistent with the concurrent processes model. **G, H:** Data was obtained from Ref [S14] and $L_d|L_i$ vs $L_b|L_i$ was plotted. Cells were grown in **G.** M9 minimal medium with sodium acetate as the carbon source ($N=1554$ cells, $\langle n_{ori} \rangle = 1.2$). We obtain a non-zero $r(L_b, L_d|L_i)$ consistent with the concurrent processes model. **H.** MOPS medium with glucose as the carbon source ($N=1807$ cells, $\langle n_{ori} \rangle = 2$). We obtain a non-zero $r(L_b, L_d|L_i)$ consistent with the concurrent processes model. **I.** Data was obtained from Ref [S15] and $L_d|L_i$ vs $L_b|L_i$ was plotted. Cells were grown in M9 minimal medium and 0.4% acetate ($N=401$ cells, $\langle n_{ori} \rangle = 1$).

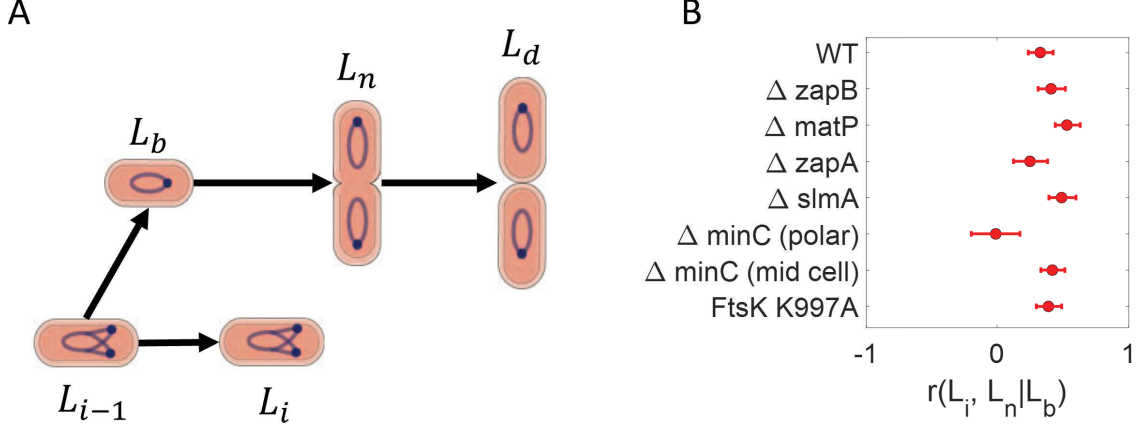


Figure S6: **A** A possible causal graph depicting the cell cycle in the Min mutants which undergo polar divisions. The mutants are hypothesized to lack mechanisms which couple the replication process to the onset of constriction. This is shown as a lack of arrow from L_i to L_n . Since the mutants are grown in glycerol+trace elements medium ($T_d \approx 148$ min in wildtype (WT)), birth related processes might still control the start of constriction. **B**. For WT, $\Delta zapB$, $\Delta matP$, $\Delta zapA$, $\Delta slmA$, $\Delta minC$ undergoing polar divisions, $\Delta minC$ undergoing midcell divisions, and the FtsK K997A strains, we show the conditional correlations $r(L_i, L_n | L_b)$.

S6 Analyzing mutants

In this section, we will probe the molecular mechanisms that might link the replication cycle and the onset of constriction using mutants studied in Ref [S1]. One such molecular system is the nucleoid occlusion factor, SlmA, which prevents the Z-ring formation until the Ter region of the chromosome moves to the mid-cell. Other proteins such as ZapA, ZapB and MatP are responsible for linking the Ter region of the chromosome to the Z-ring, thus, promoting Z-ring formation and constriction. The protein FtsK is part of the divisome and is involved in chromosome segregation at the mid-cell [S16].

If these proteins link the replication process to the onset of constriction, then removing them might start the constriction independent of the replication process. We expect L_i and L_n to be uncorrelated when L_b is conditioned upon. However, we find that the correlation $r(L_i, L_n | L_b)$ is non-zero for mutants obtained by removing SlmA ($\Delta slmA$), ZapA ($\Delta zapA$),

ZapB ($\Delta zapB$), MatP ($\Delta matP$) and on using a translocation defective FtsK K997A mutant (Figure S6B). This reiterates the conclusions reached in Ref [S1] that these molecular systems seem unlikely to be involved in the coupling between replication and the start of constriction.

We also analyzed Min mutants which have a defective Min system. Min proteins are responsible for the positioning of the Z-ring at the mid-cell [S16]. A defective Min system can lead to cell divisions occurring near the poles in addition to the symmetrical divisions at the mid-cell. We find that the conditional correlation $r(L_i, L_n|L_b)$ in Min mutant cells which undergo divisions at the mid-cell is also non-zero (Figure S6B). Next, we analyze only those Min mutant cells which undergo polar divisions. The proposed cell cycle for these mutants is shown in Figure S6A where the causal link between L_i and L_n is absent. Note that a link between L_b and L_{i-1} might still exist in these cells as their mother cells undergo divisions at mid-cell where we found replication and constriction (hence, division and birth in the next cell cycle) to be coupled. Cells which undergo polar divisions have a negligible $r(L_i, L_n|L_b)$ (Figure S6B) pointing to the lack of replication control over division (agreeing with the correlation in graph S6A). Ref [S1] proposed nucleoid occlusion as a possible mechanism for explaining the difference between cells undergoing polar and mid-cell divisions. Substantial nucleoid density at the mid-cell during the replication process hinders the formation of the Z-ring, thus, coupling replication and the start of constriction. However, the lower nuclear density at cell poles does not inhibit the Z-ring formation and constriction can start independently of the replication.

References

- S1. Tiruvadi-Krishnan, S., Männik, J., Kar, P., Lin, J., Amir, A., and Männik, J. (2022). Coupling between DNA replication, segregation, and the onset of constriction in *Escherichia coli*. Cell Reports 38, 110539.

- S2. Pearl, J. (2009). Causality (Cambridge university press).
- S3. Hernán, M.A. and Robins, J.M. (2020). Causal Inference: What If (Boca Raton: Chapman & Hall/CRC).
- S4. Barber, F., Ho, P.Y., Murray, A.W., and Amir, A. (2017). Details matter: noise and model structure set the relationship between cell size and cell cycle timing. *Frontiers in Cell and Developmental Biology* 5, 92.
- S5. Amir, A. (2014). Cell size regulation in bacteria. *Physical Review Letters* 112, 208102.
- S6. Kar, P., Tiruvadi-Krishnan, S., Männik, J., Männik, J., and Amir, A. (2021). Distinguishing different modes of growth using single-cell data. *eLife* 10, e72565.
- S7. Cylke, K.C. and Banerjee, S. (2022). Super-exponential growth and stochastic shape dynamics in rod-like bacteria. *bioRxiv* .
- S8. Logsdon, M.M., Ho, P.Y., Papavinasasundaram, K., Richardson, K., Cokol, M., Sasseti, C.M., Amir, A., and Aldridge, B.B. (2017). A parallel adder coordinates mycobacterial cell-cycle progression and cell-size homeostasis in the context of asymmetric growth and organization. *Current Biology* 27, 3367–3374.
- S9. Witz, G., van Nimwegen, E., and Julou, T. (2019). Initiation of chromosome replication controls both division and replication cycles in *E. coli* through a double-adder mechanism. *eLife* 8, e48063.
- S10. Student (1908). Probable error of a correlation coefficient. *Biometrika* , 302–310.
- S11. Micali, G., Grilli, J., Osella, M., and Lagomarsino, M.C. (2018). Concurrent processes set *E. coli* cell division. *Science Advances* 4, eaau3324.

- 353 S12. Colin, A., Micali, G., Faure, L., Lagomarsino, M.C., and van Teeffelen, S. (2021). Two
354 different cell-cycle processes determine the timing of cell division in *Escherichia coli*.
355 eLife *10*, e67495.
- 356 S13. Zheng, H., Ho, P.Y., Jiang, M., Tang, B., Liu, W., Li, D., Yu, X., Kleckner, N.E., Amir,
357 A., and Liu, C. (2016). Interrogating the *Escherichia coli* cell cycle by cell dimension
358 perturbations. Proceedings of the National Academy of Sciences *113*, 15000–15005.
- 359 S14. Si, F., Le Treut, G., Sauls, J.T., Vadia, S., Levin, P.A., and Jun, S. (2019). Mechanistic
360 origin of cell-size control and homeostasis in bacteria. Current Biology *29*, 1760–1770.
- 361 S15. Wallden, M., Fange, D., Lundius, E.G., Baltekin, Ö., and Elf, J. (2016). The syn-
362 chronization of replication and division cycles in individual *E. coli* cells. Cell *166*,
363 729–739.
- 364 S16. Männik, J. and Bailey, M.W. (2015). Spatial coordination between chromosomes and
365 cell division proteins in *Escherichia coli*. Frontiers in Microbiology *6*, 306.