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

Prathitha Kar[1,2], Sriram Tiruvadi-Krishnan[3], Jaana Männik[3], Jaan Männik[*3], and Ariel Amir[†1,4]

[1]*School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02134, USA*

[2]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA*

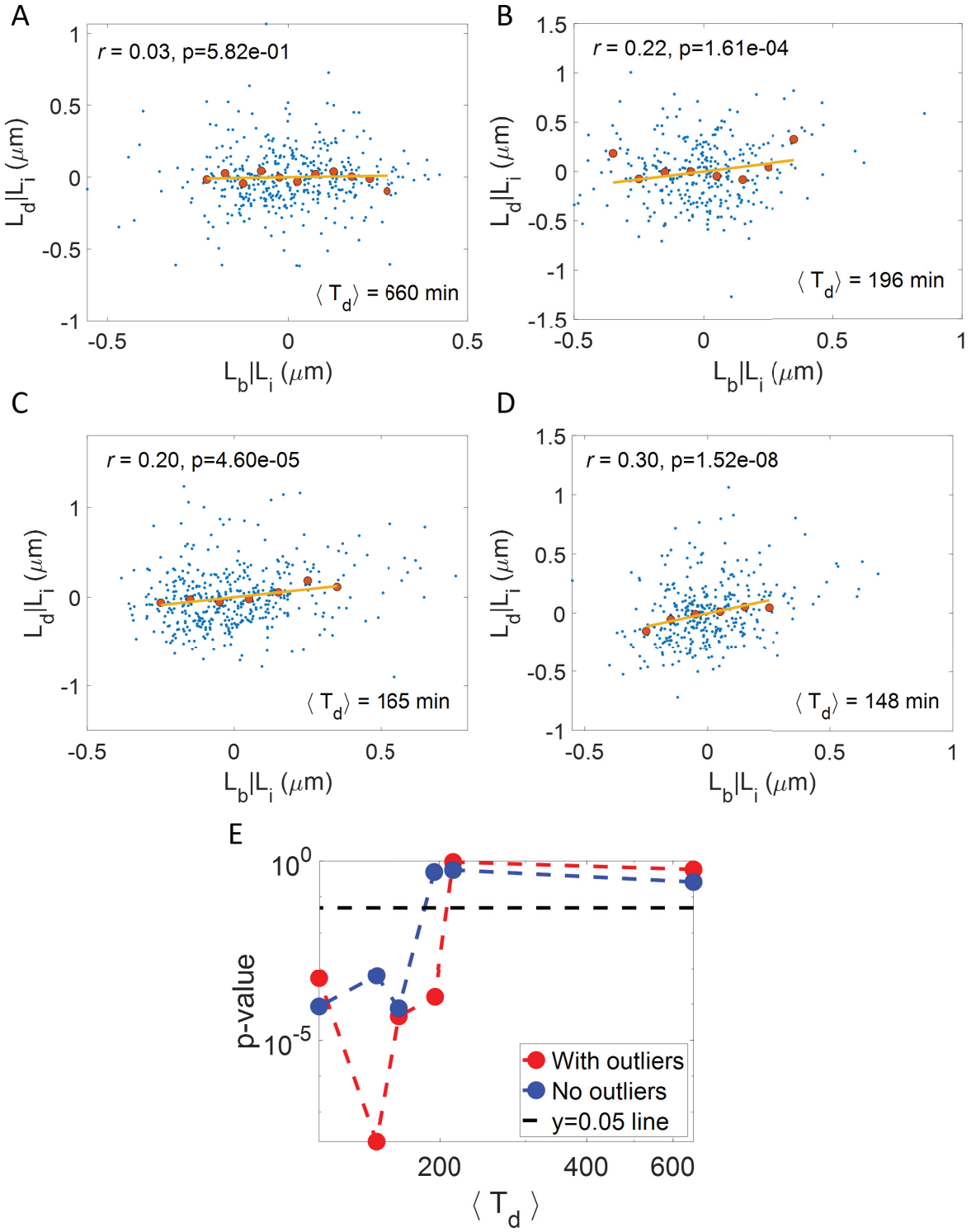
[3]*Department of Physics and Astronomy, University of Tennessee, Knoxville, TN 37996, USA*

[4]*Department of Complex Systems, Weizmann Institute of Science, Rehovot 7610001, Israel*

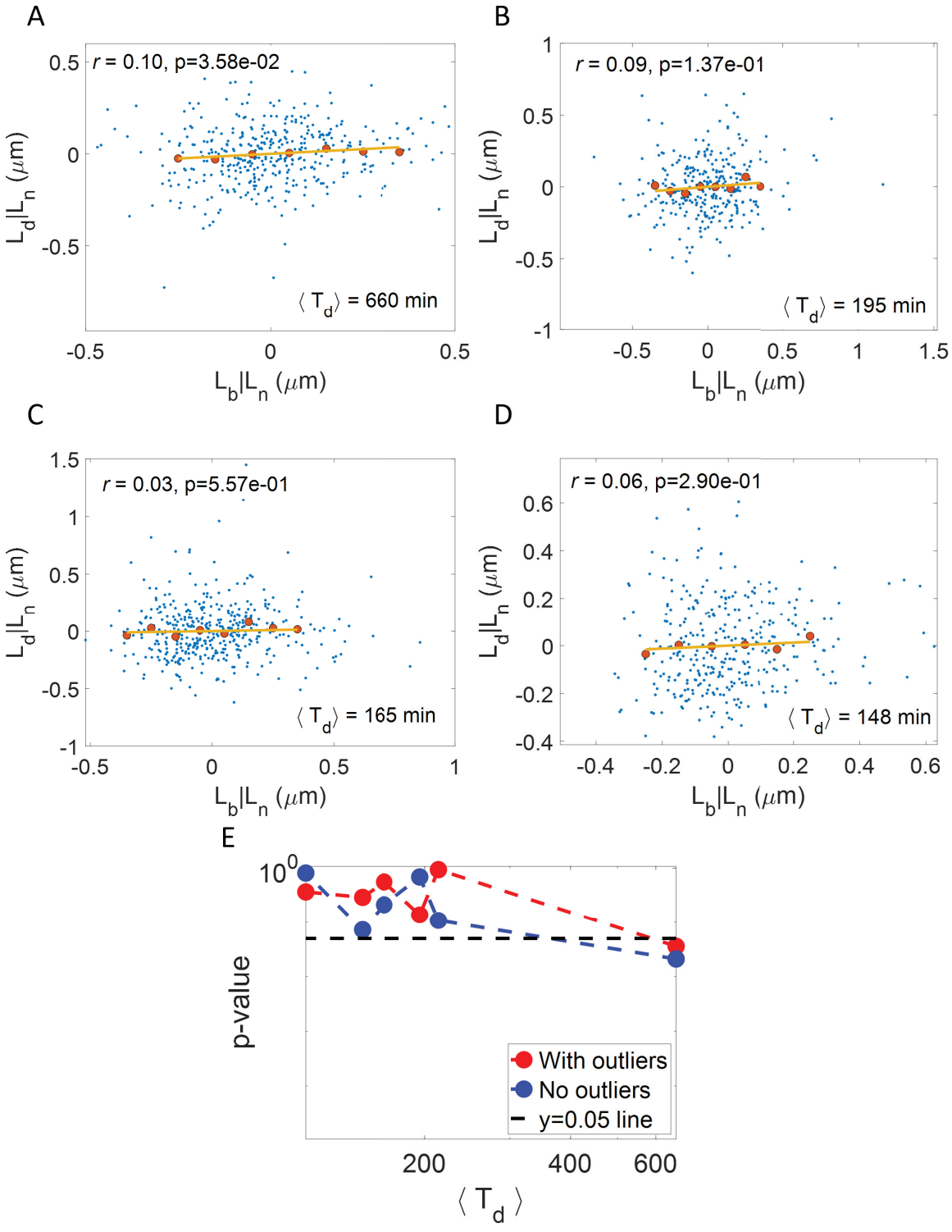
[*]Corresponding author- email: jmannik@utk.edu; phone: +1 (865) 974 6018

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

# Supplementary Figures

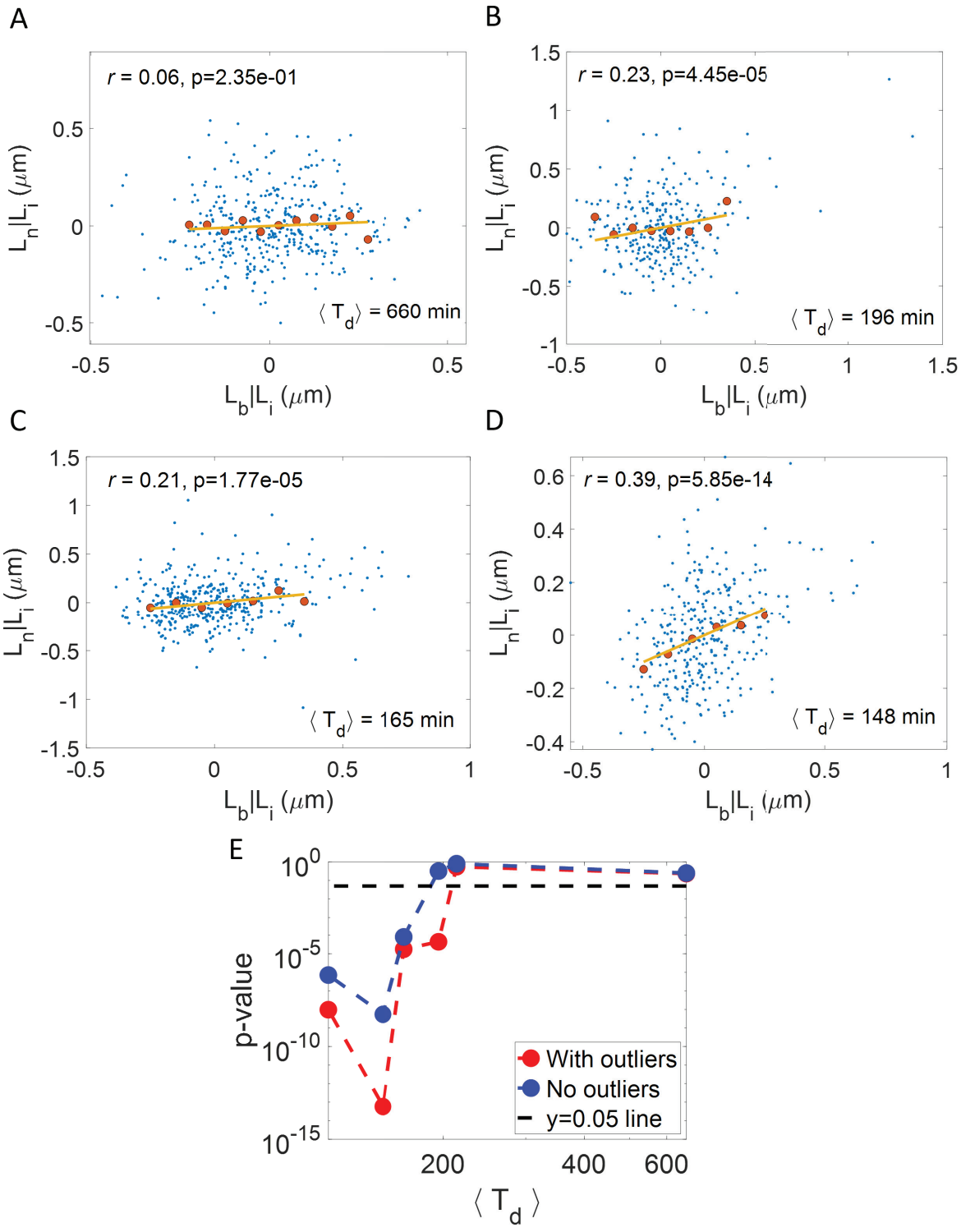


Supplementary Figure 1: **Linking birth, initiation and division: A-D.** Residuals obtained on linear regression of $L_d$ on $L_i$ ($L_d|L_i$) and $L_b$ on $L_i$ ($L_b|L_i$) are plotted for **A.** Acetate medium (generation time = 660 min, N = 401 cells, average number of origins at birth, $\langle n_{ori} \rangle = 1$). **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.


A
r = 0.10, p=3.58e-02
⟨ T_d ⟩ = 660 min
L_d|L_n (μm)
L_b|L_n (μm)
B
r = 0.09, p=1.37e-01
⟨ T_d ⟩ = 195 min
C
r = 0.03, p=5.57e-01
⟨ T_d ⟩ = 165 min
D
r = 0.06, p=2.90e-01
⟨ T_d ⟩ = 148 min
E
p-value
With outliers
No outliers
y=0.05 line
⟨ T_d ⟩

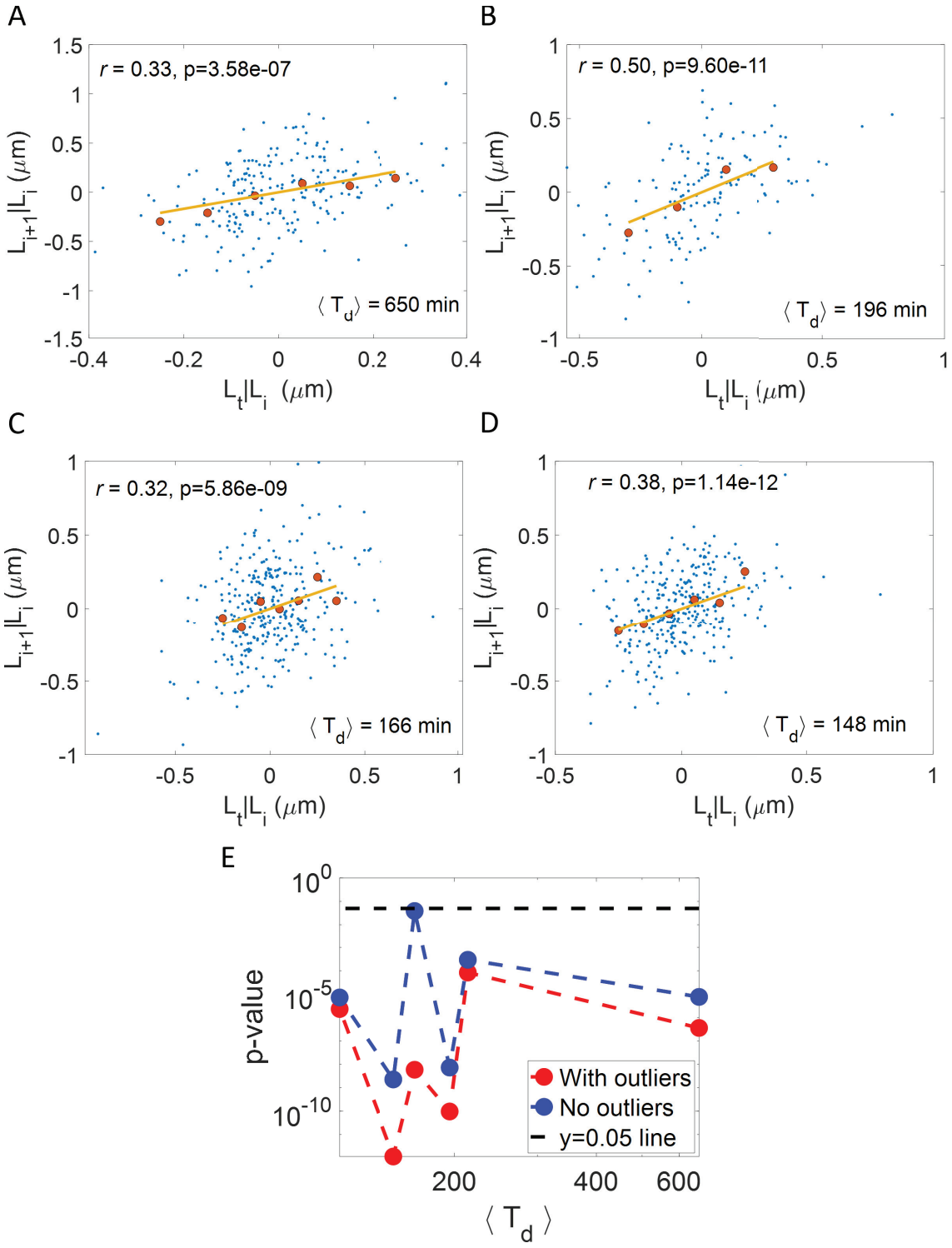
16

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.


A
r = 0.06, p=2.35e-01
⟨ T_d ⟩ = 660 min
L_n|L_i (μm)
L_b|L_i (μm)
B
r = 0.23, p=4.45e-05
⟨ T_d ⟩ = 196 min
C
r = 0.21, p=1.77e-05
⟨ T_d ⟩ = 165 min
D
r = 0.39, p=5.85e-14
⟨ T_d ⟩ = 148 min
E
p-value
With outliers
No outliers
y=0.05 line
⟨ T_d ⟩

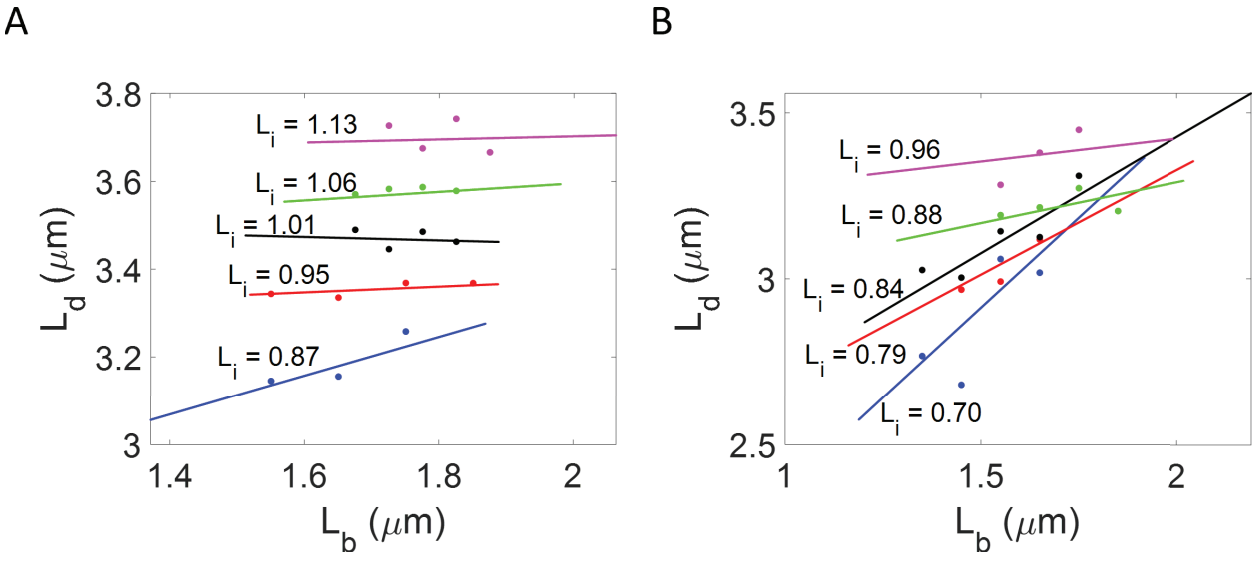
17

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.


A
r = 0.33, p=3.58e-07
⟨ T_d ⟩ = 650 min
L_{i+1}|L_i (μm)
L_t|L_i (μm)
B
r = 0.50, p=9.60e-11
⟨ T_d ⟩ = 196 min
L_{i+1}|L_i (μm)
L_t|L_i (μm)
C
r = 0.32, p=5.86e-09
⟨ T_d ⟩ = 166 min
L_{i+1}|L_i (μm)
L_t|L_i (μm)
D
r = 0.38, p=1.14e-12
⟨ T_d ⟩ = 148 min
L_{i+1}|L_i (μm)
L_t|L_i (μm)
E
p-value
With outliers
No outliers
y=0.05 line
⟨ T_d ⟩

18

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.




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 $\mu m$. 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 $\mu m$. We find that all subsets have a non-zero correlation between $L_b$ and $L_d$.

# Supplementary Tables

| **Media** | **No. of cells** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_b}, \boldsymbol{L_d})$ | $(\boldsymbol{L_b}, \boldsymbol{L_i})$ | $(\boldsymbol{L_i}, \boldsymbol{L_d})$ |
|---|---|---|---|---|---|---|
| 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** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_b}, \boldsymbol{L_d} \mid \boldsymbol{L_i})$ | $(\boldsymbol{L_b}, \boldsymbol{L_d} \mid (\boldsymbol{L_i}, \boldsymbol{\lambda}))$ | $(\boldsymbol{L_i}, \boldsymbol{L_d} \mid \boldsymbol{L_b})$ |
|---|---|---|---|---|---|---|
| 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 ($\lambda$), cell birth ($L_b$), initiation ($L_i$) and cell division ($L_d$) events.

| **Media** | **No. of cells** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_b}, \boldsymbol{L_d})$ | $(\boldsymbol{L_b}, \boldsymbol{L_n})$ | $(\boldsymbol{L_n}, \boldsymbol{L_d})$ | $(\boldsymbol{L_b}, \boldsymbol{L_d} \mid \boldsymbol{L_n})$ | $(\boldsymbol{L_n}, \boldsymbol{L_d} \mid \boldsymbol{L_b})$ |
|---|---|---|---|---|---|---|---|---|
| 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** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_i}, \boldsymbol{L_n})$ | $(\boldsymbol{L_b}, \boldsymbol{L_n} \mid \boldsymbol{L_i})$ | $(\boldsymbol{L_b}, \boldsymbol{L_n} \mid (\boldsymbol{L_i}, \boldsymbol{\lambda}))$ | $(\boldsymbol{L_i}, \boldsymbol{L_n} \mid \boldsymbol{L_b})$ |
|---|---|---|---|---|---|---|---|
| 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 ($\lambda$), cell birth ($L_b$), initiation ($L_i$), and onset of constriction ($L_n$) events.

| **Media** | **No. of cells** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_i}, \boldsymbol{L_d} \mid \boldsymbol{L_n})$ | $(\boldsymbol{L_n}, \boldsymbol{L_d} \mid \boldsymbol{L_b})$ | $(\boldsymbol{L_n}, \boldsymbol{L_d} \mid \boldsymbol{L_i})$ |
|---|---|---|---|---|---|---|
| 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** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_b}, \boldsymbol{L_d} \mid (\boldsymbol{L_i}, \boldsymbol{L_n}))$ | $(\boldsymbol{L_i}, \boldsymbol{L_d} \mid (\boldsymbol{L_b}, \boldsymbol{L_n}))$ | $(\boldsymbol{L_n}, \boldsymbol{L_d} \mid (\boldsymbol{L_b}, \boldsymbol{L_i}))$ |
|---|---|---|---|---|---|---|
| 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** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_i}, \boldsymbol{L_{i+1}})$ | $(\boldsymbol{L_i}, \boldsymbol{L_t})$ | $(\boldsymbol{L_t}, \boldsymbol{L_{i+1}})$ | $(\boldsymbol{L_t}, \boldsymbol{L_{i+1}} \mid \boldsymbol{L_i})$ | $(\boldsymbol{L_i}, \boldsymbol{L_{i+1}} \mid \boldsymbol{L_t})$ |
|---|---|---|---|---|---|---|---|---|
| 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 elements | 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.

| **Media** | **No. of cells** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_{d-1}}, \boldsymbol{L_d} \mid \boldsymbol{L_i})$ | $(\boldsymbol{L_{d-1}}, \boldsymbol{L_n} \mid \boldsymbol{L_i})$ |
|---|---|---|---|---|---|
| Acetate | 160 | 664 | 1 | -0.10 (-0.25, 0.06) | 0.11 (-0.04, 0.26) |
| Alanine | 145 | 215 | 1.06 | -0.06 (-0.22, 0.11) | 0.01 (-0.15, 0.18) |
| Mannose | 101 | 230 | 1.22 | 0.07 (-0.13, 0.26) | 0.09 (-0.11, 0.28) |
| Glycerol | 288 | 159 | 1.33 | 0.21 (0.10, 0.32) | 0.20 (0.09, 0.31) |
| Glycerol +trace elements | 162 | 139 | 1.62 | 0.33 (0.18, 0.46) | 0.32 (0.18, 0.46) |
| Glucose | 190 | 112 | 1.99 | 0.24 (0.10, 0.37) | 0.34 (0.21, 0.46) |

Table S8: Pearson correlation coefficients along with their 95% CI are shown for six different growth media.

| **Media** | **No. of cells** | $\langle \mathbf{T_d} \rangle$ **(min)** | $\langle n_{ori} \rangle$ | $(\boldsymbol{L_t}, \boldsymbol{L_{i+1}} \mid (\boldsymbol{L_i}, \frac{\boldsymbol{L_b}}{\boldsymbol{L_{d-1}}}))$ |
|---|---|---|---|---|
| Acetate | 72 | 642 | 1 | 0.28 (0.05, 0.48) |
| Alanine | 109 | 213 | 1.06 | 0.27 (0.08, 0.43) |
| Mannose | 46 | 216 | 1.39 | 0.60 (0.38, 0.76) |
| Glycerol | 220 | 160 | 1.36 | 0.35 (0.23, 0.46) |
| Glycerol +trace elements | 146 | 139 | 1.64 | 0.40 (0.25, 0.53) |
| Glucose | 187 | 111 | 1.99 | 0.26 (0.12, 0.39) |

Table S9: Pearson correlation coefficients along with their 95% CI are shown for six different growth media.

# 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 - $L_b \to L_n \to L_d$, path 2 - $L_b \leftarrow L_{i-1} \to L_i \to L_n \to L_d$..
  - 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

example, and then discuss its generalization.

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

$$L_y = L_x + L_{xy} + \eta_{xy}, \tag{S1}$$

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

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

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

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

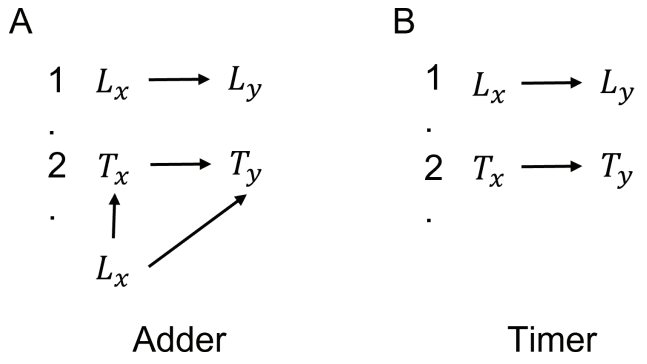



Figure S1: **DAGs for the adder and timer model: A-B.** We show DAGs involving the sizes (in graph 1) and the timings (graph 2) at cell events X, and Y for **A.** An adder model. **B.** A timer model.

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

$$T_y = T_x + T_{xy} + \eta_{t,xy}, \tag{S3}$$

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,

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

$L_x$ is independent of $\eta_{t,xy}$ in the timer model. If $L_x$ is independent of growth rate ($\lambda$), 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,

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

$\lambda(t)$ 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.

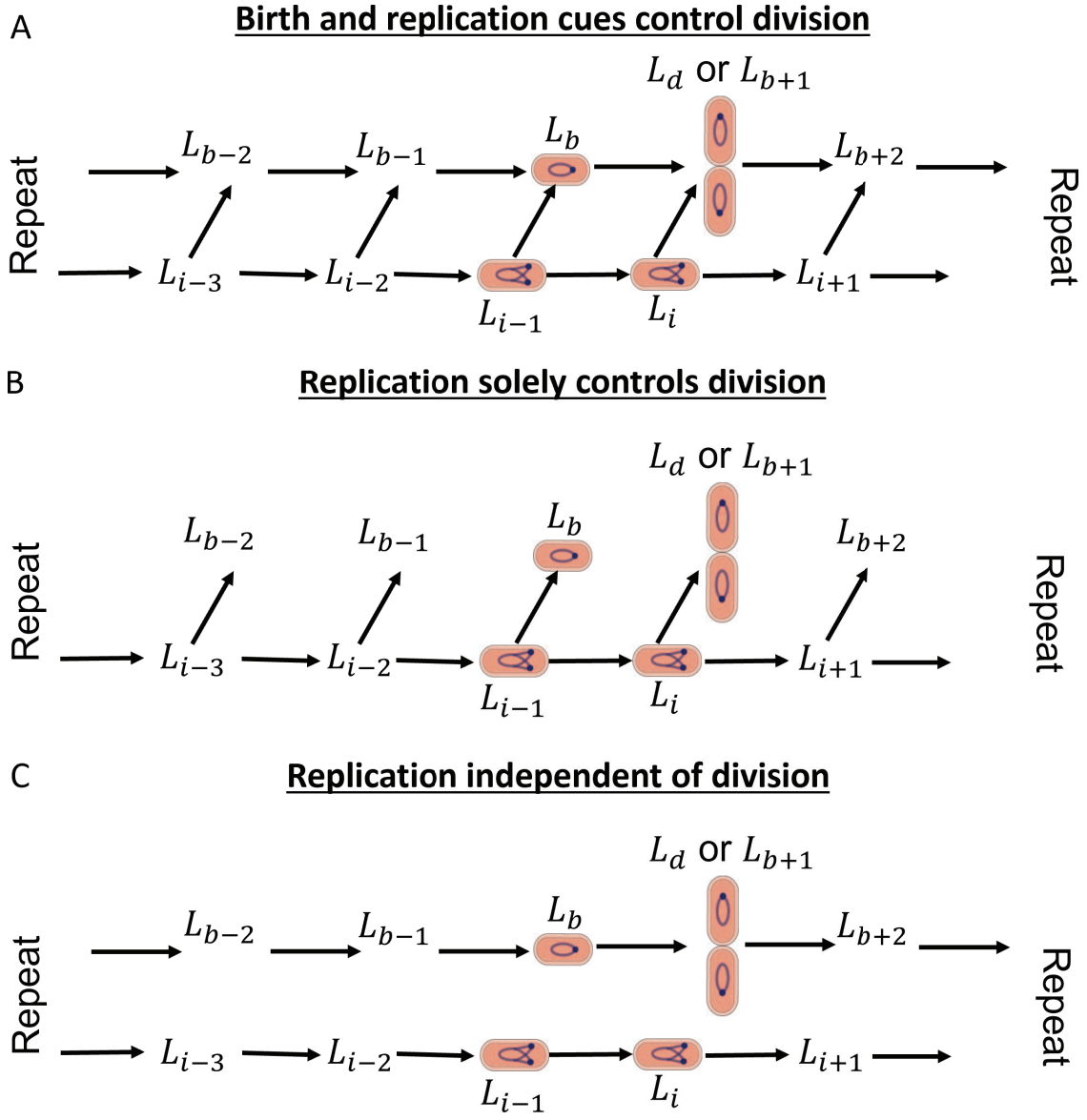



Figure S2: **A-C.** Causal graphs are shown spanning multiple generations. **A.** In this model, division is controlled by both birth and replication related processes. This is an extension of graph 3B in the main text. **B.** Division is solely controlled by replication. This is an extension of graph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

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

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

# S4 Conditional independence tests on synthetic data

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

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

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

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

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

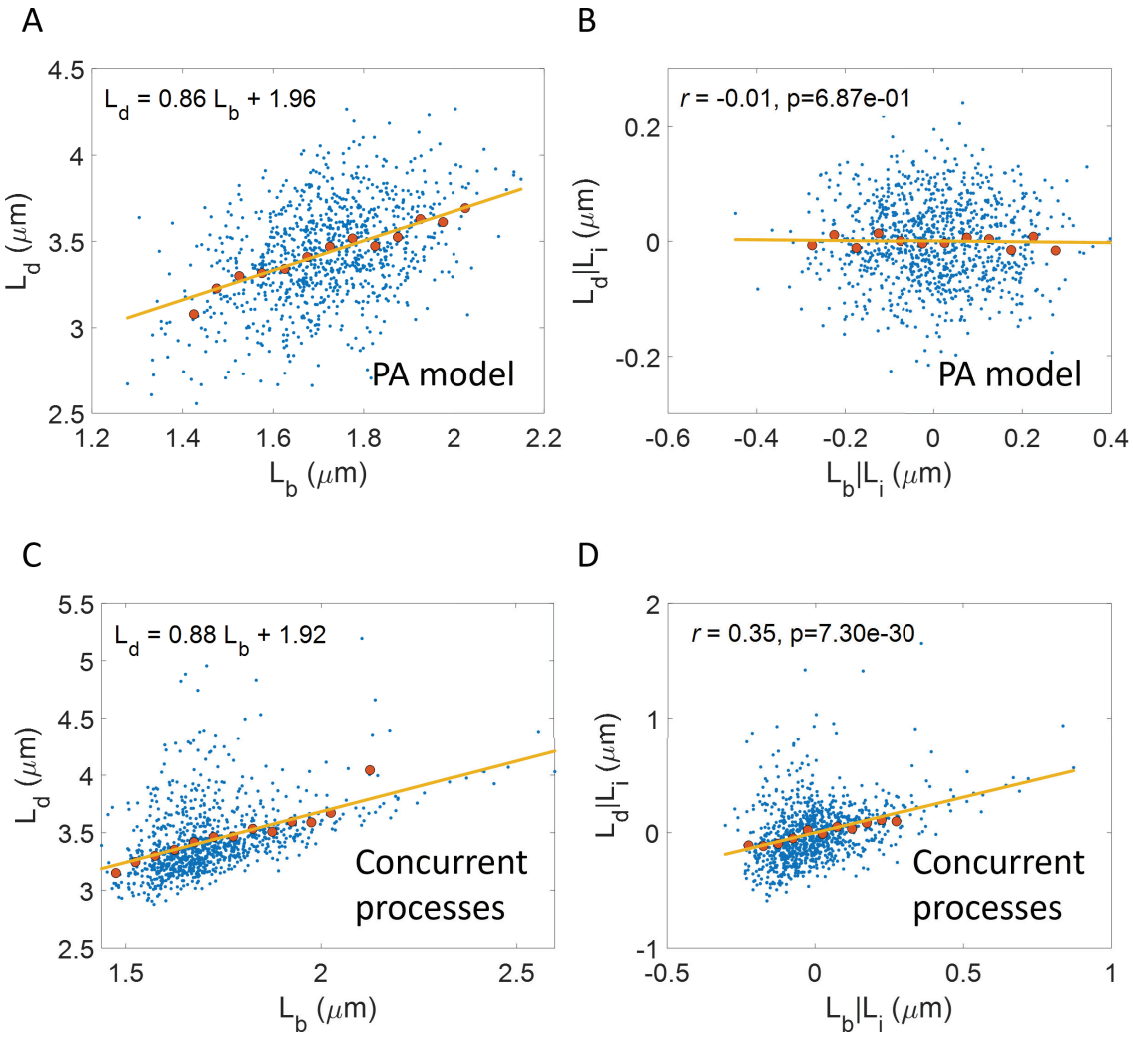
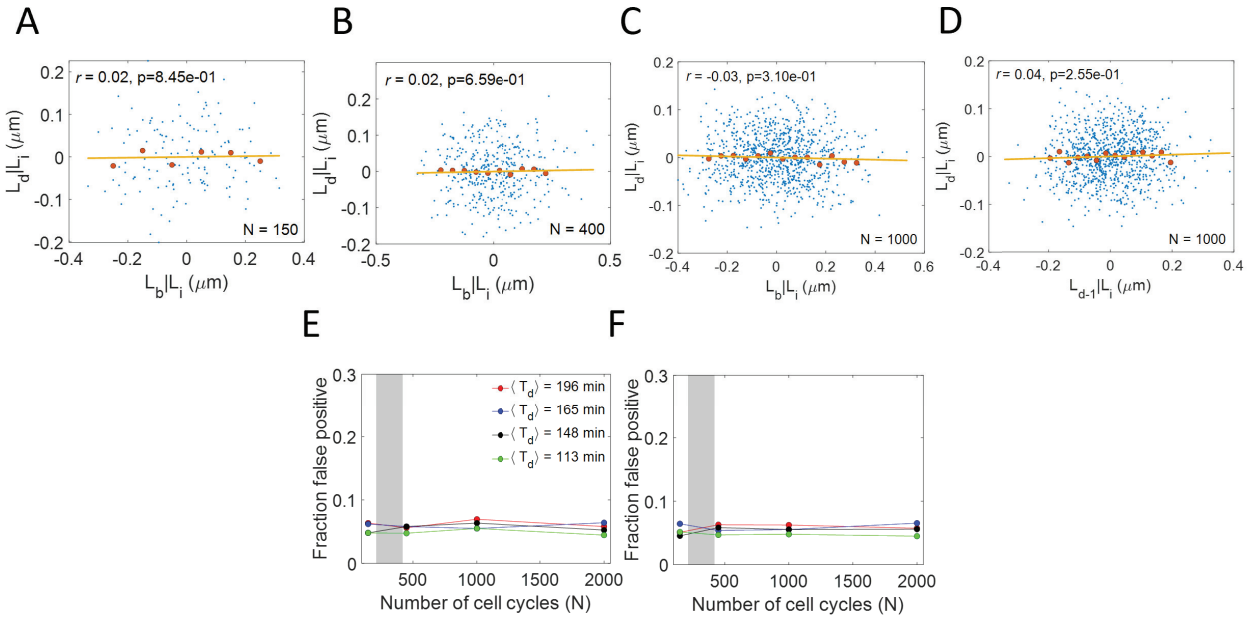
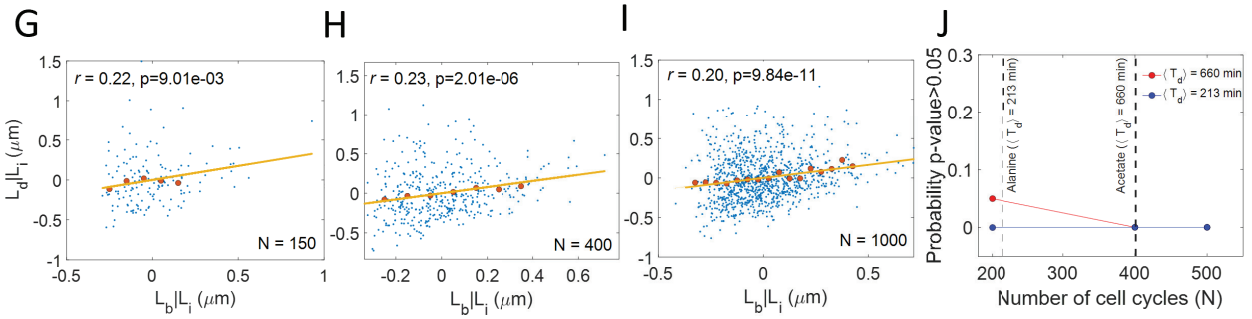


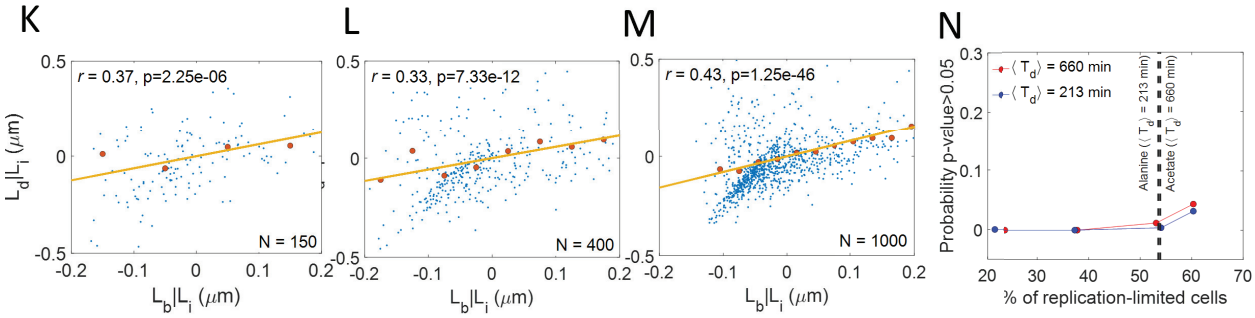

Figure S3: **Tests on synthetic data: A-B.** Simulations of cells undergoing exponential growth and following a parallel adder model are carried out and data is collected for 1000 cell cycles. For the synthetic data generated, we show **A.** $L_d$ vs $L_b$ plot. **B.** Residuals obtained on linear regression 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

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-D.** Parallel adder model was simulated over a single lineage of 1000 generations using parameters determined using experiments in acetate growth medium in Ref [S1]. The cells divided asymmetrically with the mean division ratio being 0.5 and the noise in the division ratio being determined using the experimental data in Ref [S1]. Using the synthetic data generated by the simulations, we plot the **C.** $L_d|L_i$ vs $L_b|L_i$ plot. **D.** $L_d|L_i$ vs $L_{d-1}|L_i$ plot. **E.** 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]. **F.** Parallel adder model was simulated over a single lineage of N generations for different growth media (colors represent the same growth rates as in Figure S4E). The cells divided asymmetrically with the mean division ratio being 0.5 and the noise in the division ratio being determined using the experimental data in Ref [S1]. For the 2000 iterations carried out, we found the fraction of cases where $r(L_{d-1}, L_d|L_i)$ 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]. **G-I, K-M.** 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 **G.** N = 150, x = 75%. **H.** N = 400, x = 75%. **I.** N = 1000, x = 75%. **K.** N = 150, x = 25%. **L.** N = 400, x = 25%. **M.** 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). **J, N.** For the concurrent processes model with parameters chosen using the acetate and alanine growth condition in Ref [S1], **J.** 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]. **N.** 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].

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

Next, we quantify the accuracy of rejecting the parallel adder model using the conditional independence tests. In our paper, we use p-values to classify a correlation as zero or non-zero. Under the null hypothesis that $r(L_b, L_d|L_i)$=0 with the alternate hypothesis being $r(L_b, L_d|L_i) \neq 0$, we reject the null hypothesis if the p-value is significant (less than the threshold, $\alpha = 0.05$). In such a case, we classify the data to follow a model where both birth 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 S4E). 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.

In the above simulations of the parallel adder model, the cells are assumed to divide symmetrically. However, one can relax this assumption by introducing noise in the division ratio which is defined as $\frac{L_b}{L_{d-1}}$, where $L_b$ is the size at birth and $L_{d-1}$ is the division size of the mother cell. The mean division ratio is still 0.5, i.e., the cells divide symmetrically on average. In such a model, $L_b$ and $L_d$ might be correlated via two paths- 1. via $L_i$ as discussed previously, 2. via the division ratio. While $L_b$ is clearly dependent on the division ratio, $L_d$ is also dependent on it in the case of the parallel adder model. Consider cells growing in a fast-growing media with initiations happening in the mother cell (2 origins at birth). According to the parallel adder model, $L_d$ happens upon the accumulation of size $\Delta_{id}$ per origin from

initiation. Mechanistically, this might correspond to the accumulation of a protein from initiation, and once it reaches a threshold amount division happens. For division in the current cell cycle in fast growth conditions, the accumulation of this protein starts from the mother cell. This protein is partitioned based on the division ratio at the start of the current cell cycle. The larger of the two daughter cells (larger $L_b$) has more of the protein and hence, needs to accumulate a lesser amount to undergo division. Thus, the division ratio affects the division size ($L_d$). $r(L_b, L_d|L_i)$ will be non-zero because $L_b$ and $L_d$ will be correlated via the division ratio. However, for the small division ratio noise observed in the experiments for *E. coli* ([S1, S11, S12]), we still expect $r(L_b, L_d|L_i)$ to be close to zero for the parallel adder model. This is indeed what we find as shown in Figure S4C for simulations of the parallel adder model with asymmetric divisions. We could also account for asymmetric divisions by using the conditional correlation $r(L_{d-1}, L_d|L_i)$. This correlation is unaffected by asymmetric divisions because the division length of the mother cell ($L_{d-1}$) is not influenced by the asymmetry in the division process unlike $L_b$ and $L_d$. We predict this correlation to be zero for the model where replication solely controls division (e.g., the parallel adder model) while it is non-zero for the concurrent processes model. For the simulations of the parallel adder model undergoing asymmetric divisions, we find $r(L_{d-1}, L_d|L_i)$ to be close to zero as shown in Figure S4D. Similar to the procedure followed before to generate Figure S4E, we find the false positive cases for varying N in the simulations of the parallel adder model undergoing asymmetrical divisions. We expect the correct model (parallel adder model) to be rejected in only 5% of all cases (set by significant threshold) when we use the p-values for the correlation $r(L_{d-1}, L_d|L_i)$. We indeed find it to be the case as shown in Figure S4F. Note that the correlation $r(L_{d-1}, L_d|L_i)$ will be theoretically non-zero in the simulations of the parallel adder model in growth conditions where initiations happen in the grandmother cells as $L_{d-1}$ and $L_d$ will be correlated via division ratio in the mother cell. However, in the experimental conditions that we analyze in this paper, we find initiations happening in the

same cell cycle or the mother cell. Thus, the conditional correlation $r(L_{d-1}, L_d|L_i)$ is zero for the parallel adder model as predicted and can account for division asymmetry.

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 S4G, S4K) and N=400 cells (Figures S4H, S4L). 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 S4J, 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 S4G-S4I). 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 [S13]. 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 S4N). 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 ($\approx$ 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 [S14]. 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 [S15] 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 [S15], 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 to the data on slow-growing cells in Ref [S16]. Both experiments were conducted at 37°C with similar *E. coli* strains. We find for the slower growth condition ($T_d$ = 223 min) in Figure S5I, the correlation $r(L_b, L_d|L_i)$ is lower in value than that in Figure S5G ($T_d$ = 197

min). Note that a bias in the data in Ref [S16] because the initiation is always in the same cell cycle as division (or C+D $< T_d$) might explain the non-zero $r(L_b, L_d|L_i)$ in slow-growth conditions. To conclude, birth related processes are less limiting for determining division in slower growth conditions in agreement with our results in the main text.

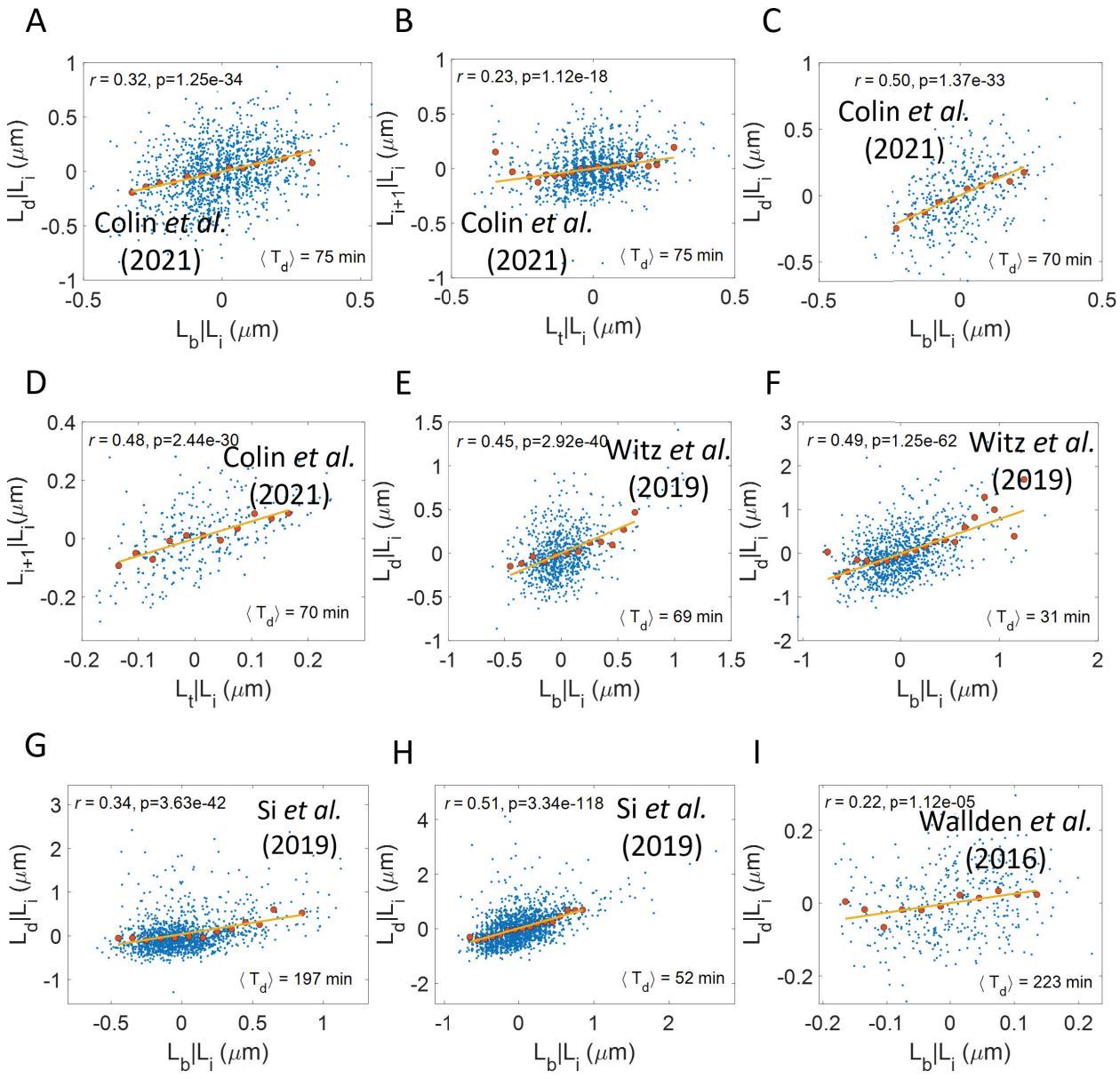


Figure S5: **Consistency with published results:** We analyze conditional correlations using previously published datasets on *E. coli*. **A,B.** Using data from Ref [S12] for N = 1380 cells ($\langle n_{ori} \rangle = 1.30$) growing in minimal media (M9+NH4Cl+glycerol), we plot **A.** $L_d|L_i$ vs $L_b|L_i$. 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 [S15] 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 [S16] 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, $\langle n_{ori} \rangle = 1$).

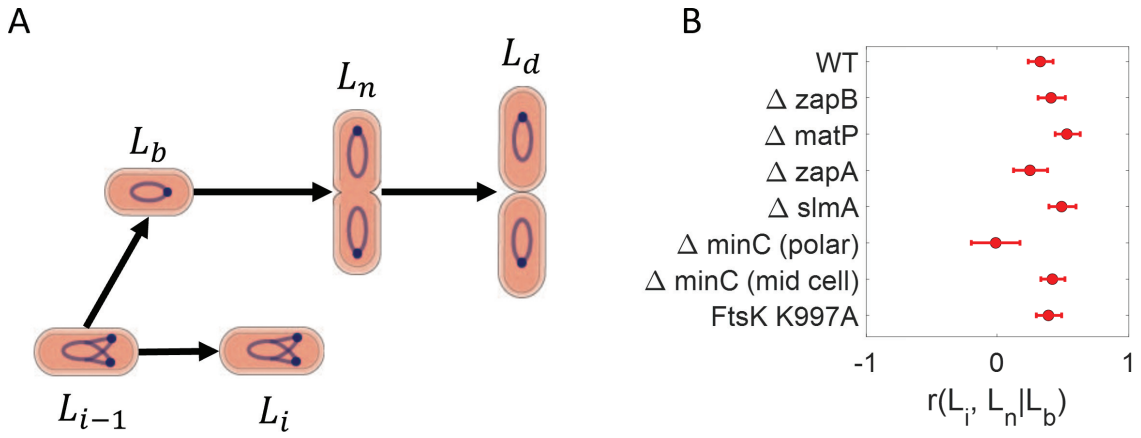



Figure S6: **A** A possible causal graph depicting the cell cycle in the Min mutants which undergo polar divisions. The mutants are hypothesized to lack mechanisms which couple the replication process to the onset of constriction. This is shown as a lack of arrow from $L_i$ to $L_n$. Since the mutants 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 [S17].

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 [S17]. 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., Sassetti, 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. Eun, Y.J., Ho, P.Y., Kim, M., LaRussa, S., Robert, L., Renner, L.D., Schmid, A., Garner, E., and Amir, A. (2018). Archaeal cells share common size control with bacteria despite noisier growth and division. Nature Microbiology *3*, 148–154.

S12. Colin, A., Micali, G., Faure, L., Lagomarsino, M.C., and van Teeffelen, S. (2021). Two different cell-cycle processes determine the timing of cell division in *Escherichia coli*. eLife *10*, e67495.

S13. Micali, G., Grilli, J., Osella, M., and Lagomarsino, M.C. (2018). Concurrent processes set *E. coli* cell division. Science Advances *4*, eaau3324.

S14. Zheng, H., Ho, P.Y., Jiang, M., Tang, B., Liu, W., Li, D., Yu, X., Kleckner, N.E., Amir, A., and Liu, C. (2016). Interrogating the *Escherichia coli* cell cycle by cell dimension perturbations. Proceedings of the National Academy of Sciences *113*, 15000–15005.

S15. Si, F., Le Treut, G., Sauls, J.T., Vadia, S., Levin, P.A., and Jun, S. (2019). Mechanistic origin of cell-size control and homeostasis in bacteria. Current Biology *29*, 1760–1770.

S16. Wallden, M., Fange, D., Lundius, E.G., Baltekin, Ö., and Elf, J. (2016). The synchronization of replication and division cycles in individual *E. coli* cells. Cell *166*, 729–739.

S17. Männik, J. and Bailey, M.W. (2015). Spatial coordination between chromosomes and cell division proteins in *Escherichia coli*. Frontiers in Microbiology *6*, 306.