Article

Synchronized oscillations in growing cell populations are explained by demographic noise

Enrico Gavagnin,[1,*] Sean T. Vittadello,[2] Gency Gunasingh,[3] Nikolas K. Haass,[3] Matthew J. Simpson,[4] Tim Rogers,[5] and Christian A. Yates[5]

[1]School of Biological Sciences, University of Bristol, Bristol, United Kingdom; [2]School of BioSciences, University of Melbourne, Melbourne, Victoria, Australia; [3]The University of Queensland, The University of Queensland Diamantina Institute, Brisbane, Queensland, Australia; [4]School of Mathematical Sciences, Queensland University of Technology, Brisbane, Queensland, Australia; and [5]Department of Mathematical Sciences, University of Bath, Bath, United Kingdom

ABSTRACT   Understanding synchrony in growing populations is important for applications as diverse as epidemiology and cancer treatment. Recent experiments employing fluorescent reporters in melanoma cell lines have uncovered growing subpopulations exhibiting sustained oscillations, with nearby cells appearing to synchronize their cycles. In this study, we demonstrate that the behavior observed is consistent with long-lasting transient phenomenon initiated and amplified by the finite-sample effects and demographic noise. We present a novel mathematical analysis of a multistage model of cell growth, which accurately reproduces the synchronized oscillations. As part of the analysis, we elucidate the transient and asymptotic phases of the dynamics and derive an analytical formula to quantify the effect of demographic noise in the appearance of the oscillations. The implications of these findings are broad, such as providing insight into experimental protocols that are used to study the growth of asynchronous populations and, in particular, those investigations relating to anticancer drug discovery.

SIGNIFICANCE   Recent experiments have reported strong evidence of periodic oscillations in the proportion of young and old melanoma cells. The biological mechanism generating this synchronization and the potential impact that can have on commonly used experimental protocols is still unclear. Here, we studied a population of melanoma cells for which we found oscillations in the proportions of cells in each phase of the cell cycle. We demonstrate that these observations may be triggered by intrinsic demographic noise alone rather than any active synchronization mechanism requiring cell-cell communication. Our findings may have implications for typical experimental protocols, which aim to produce asynchronous cell populations.

INTRODUCTION

Growing populations are a crucial feature of many biological phenomena, from the clonal expansion of cancer cell lines to the increase in numbers of infected individuals during a disease outbreak. A deeper understanding of cell proliferation sheds light on a vast range of biological processes, from morphogenesis to tumor growth (1,2); understanding and predicting the time evolution of these growing populations is, therefore, of fundamental biological relevance (3,4). The initial stages of growth in both these scenarios are typically considered to be exponential as cells replicate without restriction or disease spreads into an otherwise susceptible population.

Standard mathematical modeling approaches assume that cell divisions are independent events with exponentially distributed waiting times. This gives rise to exponential growth in unstructured populations (5). In cell biology, this approach has been supported by classic experimental studies for large populations under favorable growth conditions (6,7). However, when smaller populations are considered—for example, clones of a single progenitor cell—the classical model of exponential growth fails to capture the variable per capita growth rates caused by nonexponentially distributed cell cycle times, and more sophisticated models are necessary (8–12).

Because of recent technological advances, we are now able to access accurate data revealing the structure of dynamic cell populations (13,14), using, among other tools, proliferation assays: an in vitro experimental protocol in which the growth of cell populations is monitored over time (15). In particular, in recent work (16), we assayed

---

the proliferation of melanoma cells labeled with Fluorescent Ubiquitination-based Cell Cycle Indicator (FUCCI ([17],[18]); see Fig. 1 and Materials and methods), which allowed us to track the number of cells in particular phases of the cell cycle over a time span of 48 h. Strikingly, the proportion of cells in the first phase of the cell cycle, gap 1 (G1), demonstrates clear and unexplained fluctuations during the entire duration of the experiment. This synchrony, reproducible over multiple cell lines and experimental replicates, has potentially serious implications for studies of rapidly dividing cell populations and demonstrates that classical and widely adopted exponential models of population growth are insufficient to capture either individual-level or population-scale growth dynamics.

Inspired by our experimental findings, we build a multistage mathematical model for cell proliferation that represents the cell cycle as a series of discrete stages ([16]). The waiting time distribution between consecutive stages is exponential, meaning that the cell cycle time (CCT) follows a more general class of distributions, known as hypoexponential in Yates et al. ([8]). This family of distributions has been shown to provide good agreement with the experimental cell cycle time distribution data ([8],[13],[14],[19]). By deriving a deterministic representation of the population dynamics under the multistage approach, we reproduce the cell-cycle fluctuations observed in the experiments. This suggests that multistage models are a suitable framework for investigating

the phenomenon of cell-cycle synchronization. However, because in ([16]), the parametrization of the model is carried out individually for each experimental trajectory, our previous study did not explain the origin of such oscillatory phenomena nor their asymptotic behavior.

In a growing cell population, cell division acts as a natural source of synchronicity by increasing the number of phase-synchronized cells. This progeny form of synchronicity plays a fundamental role when the population size is small (for example, in the case of single-cell lineages). As the population grows, the effect of initially synchronous progeny gradually decreases, and it eventually becomes negligible for a large population, at which point the phase distribution—the distribution of cells in each phase of the cell cycle—reaches an invariant state, and the total population grows exponentially (asymptotic regime). When a small population is sampled from a larger one, for example during a typical cell proliferation assay, the intrinsic randomness of the process plays a dual role. Stochastic finite-size effects can lead to the sampled population being far from the invariant distribution. That sampled population will have a synchronized oscillatory state. However, the individual variations in the cell cycle time tend to break the correlation of synchronized cells, resulting in a gradual desynchronization the population (transient regime). The presence of these two regimes (a transient-oscillatory regime and an asymptotic-exponential regime) is a common feature of many structured growing populations ([9],[10],[20]). It is not surprising, therefore, that these two phases play distinct but critically important roles in the dynamics of a growing cell population.

The extent to which this natural transient phase is responsible for the observed oscillatory phenomenon remains unclear. Our goal is to design a suitable theoretical framework based on the multistage model (Multistage agent-based model), which allows us to quantify for a given population the magnitude of synchronicity triggered and sustained by demographic noise—a general term that describes to the randomness emerging at the level of a finite-sized population when individual birth and death events occur stochastically. The focus of our analysis is on quantifying the expected amplitude of the oscillations as this quantity can be easily measured and directly linked to experimental observations, even when the data are limited to short time windows.

In this study, we establish that strong oscillations observed in growing populations of cells can be triggered by demographic noise alone. This finding demonstrates that it may not be necessary to appeal to an external synchronization mechanism requiring cell-cell communication to explain synchronization observed in experiments. To do this, we first analyze the multistage model with a particular focus on characterizing the transient and asymptotic phases. By deriving a stochastic mesoscopic model, we study the effect of stochasticity in the system and obtain an analytical
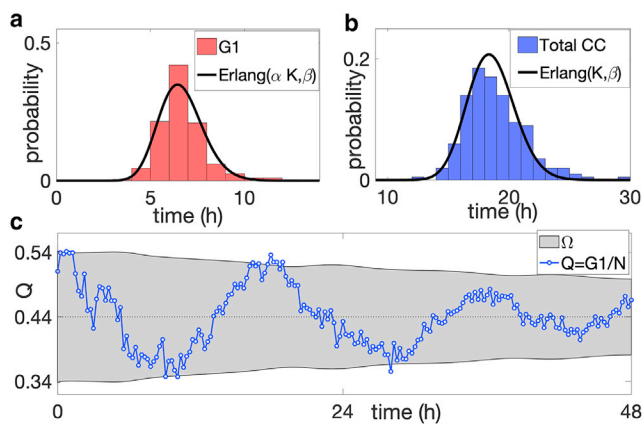

FIGURE 1 Cell cycle and phase time distributions and synchronized oscillations. (*a*) shows a comparison between the distribution of the duration of the G1 phase obtained by tracking 200 randomly chosen cells (*red histogram*) and the Erlang distribution with the same mean and variance (*black curve*). (*b*) shows the comparison between the distribution of the full cell cycle time of the same tracked cells (*blue histogram*) together with the corresponding Erlang distribution. (*c*) shows one time series trajectory of the proportion of G1 cells, Q, obtained from the experiments (*blue line*), together with the envelope of two SDs from the mean (*light gray region*) predicted using the multi-stage model. The parameters of the multi-stage models are obtained by fitting the distribution of the total cell cycle time and G1 duration (see Supporting materials and methods, section S4): $K = 92$, $\alpha K = 33$, $\beta = 4.96 \text{ h}^{-1}$, $\nu = 94.3$, and $N_0 = 381$. To see this figure in color, go online.

formula that can be used to quantify the amplitude of the fluctuations because of finite-sample effects. Finally, we parametrize the multistage model by fitting the G1 and total cell cycle time distributions, obtained from single-cell tracking data, and compare our predictions with the time series obtained in the experiments.

Our central finding is that the fluctuations in the subpopulation of G1 cells in the proliferation assay are of the same magnitude as those induced by demographic noise alone, which suggests that finite-sample effects provide a straightforward, yet often overlooked, explanation of the observed synchronization. Our study examines the specific impact of demographic noise on the dynamics of the amplitude of the oscillations, predicting that the observed synchronicity is a transient phenomenon for which we can predict the corresponding characteristic decay time.

The fact that the observed synchrony is generated by demographic noise and not a feature peculiar to the cell line we studied means that we expect the same phenomenon to be observed in a wide range of other populations undergoing stochastic growth, as exemplified in a number of studies (4,15,21–23). Moreover, the generality of the mathematical models adopted throughout allows this study to serve as a showcase of the potential applications of multistage models. The possible implications of our study therefore range from revised experimental protocols to altered cancer treatment schedules and from new ways of understanding early infection progression within the body to strategies for prevention of the spread and dissemination of disease in the early stages of an outbreak. The characterization of the transient nature of synchronized behavior may lead the way to new experimental designs for a broad range of experimental protocols in which cell cycle synchronization is of vital importance (24).

MATERIALS AND METHODS

Cells and cell culture

The human melanoma cell line C8161 (kindly provided by Mary Hendrix, Chicago, IL) was genotypically characterized (25,26), grown as described by Spoerri et al. (27), and authenticated by short tandem repeat fingerprinting (Queensland Institute of Medical Research Berghofer Medical Research Institute, Herston, Australia).

We maintain the cell cultures to prevent any induced synchronization from cell cycle arrest in the G1 phase. In general, such induced synchronization can occur through various experimental conditions, namely contact inhibition of proliferation at relatively high population densities (28), decreased pH of the growth medium because of the concentration of acidic cell metabolites such as lactic acid (29), and reduced availability of nutrients such as serum (21). We prevent induced synchronization by passaging the cells every 3 days and on the day before setting up an experiment to maintain a subconfluent cell density and a fresh growth medium so that the cell culture conditions are never such that they cause G1 arrest.

We note that there are other factors that can induce cell synchronization. For example, during the suspension, before seeding, some cells may die because of detachment—in particular, those close to mitosis. Clearly, this phenomenon might lead to further deviation from the invariant distribution and, hence, to amplification of the appearance of the oscillations. Because the aim of our study is to quantify the oscillations arising only from finite-sample effects, we do not account for this phenomenon in our model.

FUCCI

To generate stable melanoma cell lines expressing the FUCCI constructs, mKO2-hCdt1 (30–120) and mAG-hGem (1–110) (17) were subcloned into a replication-defective, self-inactivating lentiviral expression vector system as previously described (30). The lentivirus was produced by cotransfection of human embryonic kidney 293T cells with four plasmids, including a packaging defective helper construct (pMDLg/pRRE), a Rev plasmid (pRSV-Rev), a plasmid coding for a heterologous (pCMV-VSV-G) envelope protein, and the vector construct harboring the FUCCI constructs mKO2-hCdt1 (30–120) and mAG-hGem (1–110). High-titer viral solutions for mKO2-hCdt1 (30/120) and mAG-hGem (1/110) were prepared and used for co-transduction into eight biologically and genetically well-characterized melanoma cell lines (see above), and subclones were generated by single-cell sorting (18,27,31).

Image processing and analysis

The microscopy data consist of multichannel time series stacks that are processed and analyzed automatically with Fiji/ImageJ and MATLAB (The MathWorks, Natick, MA), as described in Vittadello et al. (16).
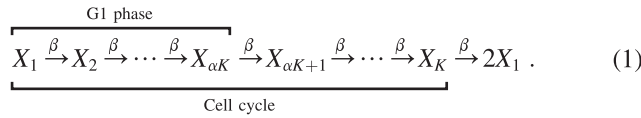
To obtain the time distribution of the G1 phase cells and of the cell cycle, we selected 200 cells toward the beginning of the experiment. To do this, we first labeled all the automatically detected cells on the first frame of the processed merged image of the red and green channels (using the routine Analyze particle of Fiji/ImageJ); we then selected 100 labeled mother cells uniformly, without replacement. For each selected mother cell, we manually recorded the time intervals corresponding to the G1 phase (i.e., between the mitosis event of the mother cell and the first appearance of the cell in the green channel) and to the cell cycle (i.e., between the mitosis event of the mother cell and the last appearance of the cell in the green channel) of its two daughter cells. We ignored cells that did not reach mitosis before the end of the experiment or move out of the microscopy window (1% of the selected cells).

Multistage agent-based model

We adopt an agent-based model for the growth and division of cells, following (8,13,19). In this formulation, the cell cycle is represented as a series of K stages, through which a cell progresses before it divides. We choose the waiting time to progress from one stage to the next to be exponentially distributed with rate β, independent from all other events. When a cell passes through the final stage, it divides into two new daughter cells, both initialized at stage one. This is a simplified model of the cell cycle; however, it is sufficient for the purposes of this study, and (as we will show later) it gives a good fit to experimentally observed distributions of cell cycle time.

The K stages of our model are grouped into sections corresponding to the known phases of the cell cycle. In particular, we say that a cell is in the G1 phase if it is in one of the first αK stages, where $\alpha \in [1/K, 2/K, \ldots, 1]$ is a constant to be determined by comparison with data. Expressed as a sum of exponential random variables, the duration of both the G1 and the entire cell cycle are Erlang distributed with parameters (K, β) and $(\alpha K, \beta)$, respectively. Fig. 1, a and b show the maximal likelihood fit of the model simultaneously to both the duration of the G1 phase and the total cell cycle time for the melanoma cell line C8161. In this example, we find parameters $K = 92$, $\alpha K = 33$, and $\beta = 4.96 \text{ h}^{-1}$. The measured cell cycle time has an average of 18.5 h with SD around 2 h (see Supporting materials and methods, section S4).

We define the state vector $X(t) = (X_1(t), X_1(t), \ldots, X_K(t))$, where $X_k(t)$ denotes the number of cells in stage k at time t. Our model can be represented as a series of chemical reactions, namely

$$
\overbrace{X_1 \xrightarrow{\beta} X_2 \xrightarrow{\beta} \cdots \xrightarrow{\beta} X_{\alpha K}}^{\text{G1 phase}} \xrightarrow{\beta} X_{\alpha K+1} \xrightarrow{\beta} \cdots \xrightarrow{\beta} X_K \xrightarrow{\beta} 2X_1 . \qquad (1)
$$

$$
\underbrace{\phantom{X_1 \xrightarrow{\beta} X_2 \xrightarrow{\beta} \cdots \xrightarrow{\beta} X_{\alpha K} \xrightarrow{\beta} X_{\alpha K+1} \xrightarrow{\beta} \cdots \xrightarrow{\beta} X_K}}_{\text{Cell cycle}}
$$

Note that one can write down a system of master equations for the set of chemical reactions above; however, this system is analytically solvable only for the simple case of $K = 1$ (8). We write $N = \sum_{k=1}^{K} X_k$ for the total number of cells and $G = \sum_{k=1}^{\alpha K} X_k$ for the number of cells in G1 phase. As the population grows, the proportion of cells in each stage will eventually converge to a fixed value $\lim_{t \to \infty} X_i/N = u_i$, for $I = 1, \ldots, N$. Together, these proportions are known as the invariant stage distribution, $\boldsymbol{u}$. In Understanding the transient and asymptotic dynamics, we prove this and derive an exact expression for the limit $\boldsymbol{u}$. On shorter time horizons, the system exhibits transitory oscillations about the invariant distribution (32,33).

RESULTS

Multistage model recapitulates experimental observations

To assess the amplitude of oscillations, in what follows, we develop a mathematical theory for the behavior of the proportion, $Q = G/N$, of G1-phase cells (see Material and methods). The first part of our analysis reveals long-lived damped oscillations in the expected value of Q in a growing population, whereas the second shows how this effect is initiated and sustained by demographic noise.

Our experimental data are 30 time series of images taken from proliferation assays, as previously reported in Vittadello et al. (16) (Cells and cell culture and Fig. S3 for three snapshots of the microscopy images). Each time series captures a 48-h time window after an incubation period of 24 h. In Fig. 1 c, we report an example of an experimental trajectory of Q (*blue line*) with marked oscillations and the envelope of two SDs about the mean, Ω (*light gray region*), obtained from the multistage model. The trajectory shows clear oscillations about the mean (about three complete cycles in the 48-h window), and 94% of the data points lie inside Ω (see Supporting materials and methods, section S6 for all the 30 time series of the experiments). The period of the oscillations is approximately one CCT, i.e., 18.5 h, which is confirmed by the power spectrum analysis (reported in Figure S5). In Fig. 2 a, we compare the amplitude of the oscillations; we present the time series of the sample SD of Q (the proportion of G1-phase cells), $\sigma_Q(t)$, obtained from the 30 experimental trajectories (*blue line*), together with the corresponding theoretical value (*gray line*) predicted from the multistage model (see Supporting materials and methods, section S3). The plot shows that, despite the fluctuation in the experimental data, there is a good agreement between the two time series for the entire duration of the experiment.
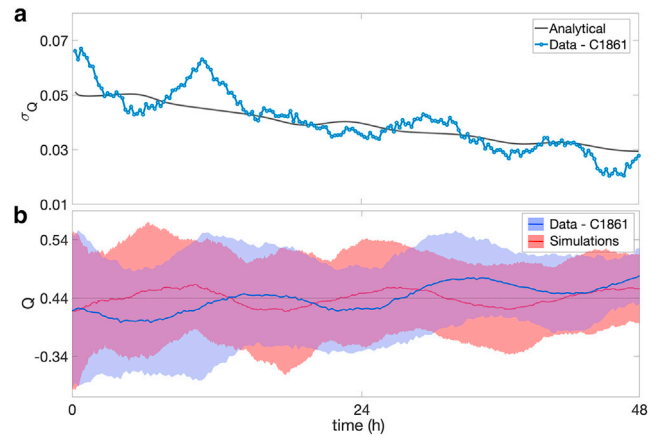

FIGURE 2 Comparison amplitude of oscillations between experimental data and model prediction. (*a*) Shown is a time series of the SD of Q, σ_Q, obtained from the experiments (*blue line*) and the analytical prediction of the model (*gray line*). (*b*) Shown is an overlay of envelopes of two sample SDs (*shaded regions*) around the sample means (*full lines*). The blue envelope is obtained from the 30 experimental time series. The red envelope is obtained from 30 independent simulated trajectories. The initial state of the simulations is chosen to match the mean and sample SD of the data. The parameters of the model are the same as in Fig. 1. To see this figure in color, go online.

In Fig. 2 b, we display the envelopes of two sample SDs around the sample mean of $Q(t)$, obtained from the 30 experimental time series (*blue*) and from 30 independent simulations of the multistage model (Langevin model; see Finite-sample effects trigger and amplify oscillations) initialized to match the experimental sample mean and SD at time 0. The plot highlights some qualitative similarities between the two envelopes; in particular, the rate of decay of the sample SD obtained from simulations is consistent with the experiments. Interestingly, a periodicity of approximately one CCT is visible in the time series of the sample mean, which is qualitatively captured in the trajectory obtained from independent simulations, suggesting that this phenomenon may be due to the relatively small size of the sample (30 trajectories).

Understanding the transient and asymptotic dynamics

To understand the interplay between the transient-oscillatory dynamics and asymptotic-exponential growth, we begin by writing down the equations governing the dynamics of the expected number of cells in each stage $\overline{\boldsymbol{x}} = \mathbb{E}[\boldsymbol{X}]$. Here, expected should be interpreted as the average over many experiments with precisely the same initial condition; we will later see that the variability of the initial condition is a different feature that can also lead to the emergence of oscillations. From the model formulation, we directly obtain

$$
\frac{d\overline{\boldsymbol{x}}}{dt} = \beta \mathcal{S} \overline{\boldsymbol{x}} , \qquad (2)
$$

where β is the rate of progression through the model stages, and S is the corresponding stoichiometry matrix. This matrix has nonzero entries $S_{k,k} = -1$ and $S_{k,k+1} = 1$ for $k = 1, \ldots, K - 1$ describing progression between stages and $S_{K,K} = -1$ and $S_{1,K} = 2$ describing cell division.

For the purpose of the analysis, we assume $\beta = K$ throughout so that the average cell cycle time is normalized to unity. The characteristic polynomial of the matrix S is $P(y) = (y + 1)^K - 2$, from which the eigenvalues of S are $\lambda_k = \xi_k 2^{\frac{1}{k}} - 1$ for $k = 1, \ldots, K$, where $\xi_k = e^{2\pi i k/K}$ is a k-th root of unity. By solving a series of recursive equations, one can write down the left and right eigenvectors associated with the k-th eigenvalue of S, which we denote $\boldsymbol{u}^k$ and $\boldsymbol{v}^k$, respectively. Specifically, we have

$$u_j^k = \frac{2\lambda_K}{(1 + \lambda_k)^j} \quad \text{and} \quad v_j^k = \frac{1}{K} \frac{(1 + \lambda_k)^j}{2\lambda_K} . \tag{3}$$

We drop the index k whenever we refer to the eigenvalue with maximal real part and the corresponding eigenvectors, i.e., $\lambda = \lambda_K = 2^{\frac{1}{k}} - 1$, $\boldsymbol{u} = \boldsymbol{u}^K$, and $\boldsymbol{v} = \boldsymbol{v}^K$.

Notice that from the system of Eq. 2, we can write $\overline{\boldsymbol{x}}(t) = e^{KtS}\overline{\boldsymbol{x}}_0$, where $\overline{\boldsymbol{x}}_0$ denotes the initial number of cells per stage. To study the matrix exponential e^{KtS}, we first notice that we can write down the (i, j) element in terms of the eigenvalues and eigenvectors of S as

$$\left[e^{KtS}\right]_{i,j} = \sum_{k=1}^{K} u_i^k v_j^k e^{K\lambda_k t} . \tag{4}$$

Notice that as $t \to \infty$, the leading term of Eq. 4 is $u_i v_j e^{K\lambda t}$, and hence, $\overline{\boldsymbol{x}}(t) \sim \boldsymbol{u} e^{K\lambda t}$ determines the long-time behavior of the system of Eq. 2. We can use this fact to study the limiting behavior of Q: we write $\overline{Q}(t) = \sum_{i=1}^{\alpha K}[e^{KSt}\overline{\boldsymbol{x}}_0]_i / \sum_{i=1}^{K}[e^{KSt}\overline{\boldsymbol{x}}_0]_i$, and by looking at the first two terms of 4, we obtain $\lim_{t \to \infty} Q(t) = \sum_{i=1}^{\alpha K} u_i = 2(1 - 2^{-\alpha}) = Q_*$. Notice that convergence to Q_* occurs with an exponential decay rate given by the spectral gap of the stoichiometry matrix, $\Re[\lambda_{K-1}] - \lambda_K$ (see Fig. 3).

We now focus on the transient behavior of the system of Eq. 2. We substitute the Eq. 3 into the Eq. 4, and by exploiting a remarkable identity of the Mittag-Leffler function (34), we transform the finite sum over eigenvalues on the right-hand side of 4 into an infinite sum over the cycles of the oscillatory solutions. Precisely, we write

$$\left[e^{KtS}\right]_{i,j} = \frac{1}{K} \sum_{k=1}^{K} (1 + \lambda_k)^{j-i} e^{K\lambda_k t}, \tag{5}$$

$$= \sum_{n=0}^{+\infty} \frac{1}{2\pi i} \oint 2^n (1 + z)^{-1 - Kn - (i-j)} e^{Kzt} dz, \tag{6}$$

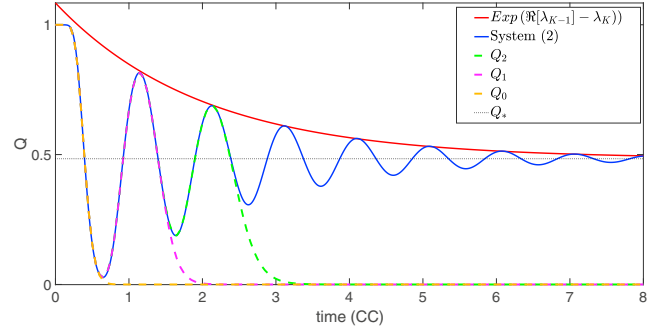$$= \sum_{n=0}^{+\infty} \varphi_n(t, i, j) , \tag{7}$$

FIGURE 3 The transient-oscillatory dynamics. The figure shows the plot of the ratio $Q(t)$ obtained by solving the deterministic system of Eq. 2 numerically (*blue solid*) initialized with $X(0) = N_0 \boldsymbol{e}_1$ and parameters $K = 40$, $N_0 = 100$, and $\alpha = 0.4$. The dashed lines represent the short time approximation obtained by truncating the expression of Eq. 7 up to $\overline{n} = 0$ (*dashed yellow*), $\overline{n} = 1$ (*dashed pink*), and $\overline{n} = 2$ (*dashed green*). The red solid line shows the exponential decay of the oscillations. To see this figure in color, go online.

where $\varphi_n(t, i, j) = 2^n e^{-Kt} (Kt)^{Kn+i-j} / (Kn + i - j)!$.

We can now use Eq. 7 to approximate e^{KtS} for short times by truncating the sum over n to a finite index, $\overline{n}$. For example, let us consider an initial population of N_0 cells perfectly synchronized at the beginning of the cell cycle, i.e., $\overline{\boldsymbol{x}}_0 = N_0 \boldsymbol{e}_1$. Then, we define $G_{\overline{n}} = N_0 \sum_{k=1}^{\alpha K} \sum_{n=0}^{\overline{n}} \varphi(n, i, 1)$ and $N_{\overline{n}} = N_0 \sum_{k=1}^{K} \sum_{n=0}^{\overline{n}} \varphi(n, i, 1)$. In Fig. 3, we plot $Q_{\overline{n}} = G_{\overline{n}}/N_{\overline{n}}$, for $\overline{n} = 0, 1$ and 3, together with Q obtained by solving the system of Eq. 2 numerically. The plot illustrates how each term of the sum in Eq. 7 contributes one additional oscillation to the transient dynamics of the proportion of G1-phase cells. We now have a complete picture of how oscillations propagate on average in the growing population. It remains for us to show how those oscillations are created and sustained.

Finite-sample effects trigger and amplify oscillations

There are two sources of randomness that are relevant to our model of cell population growth: the finite-size effects involved in the initial sampling and the individual cell variation of CCT. To take into account the stochasticity in the initial population of cells, we sample the initial stage distribution, $\boldsymbol{x}(0)$, by drawing from the invariant distribution u. Precisely, the proportion of cells on stage i at time 0 is modeled as $x_i(0) = v Y_i$ for $i = 1, \ldots, K$, where $Y_i \sim \text{Po}(u_i/v)$ are independent Poisson random variables and v is a parameter modulating the level of the stochastic involved in the sampling.

Next, we aim to quantify the effects of inherent stochasticity in the agent-based model. Performing a finite-size expansion of the master equation associated with the model (35,36), we derive a system of stochastic differential equations for the density of cells relative to the initial population
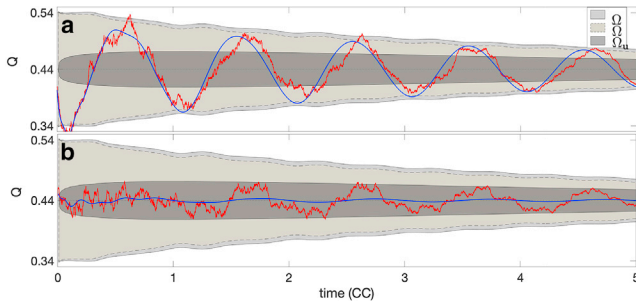
FIGURE 4 Finite-sample effects amplify the oscillations of Q. The two panels show the overlay of the three envelopes Ω (*light gray region with solid line*), $\overline{\Omega}$ (*medium gray region with dashed line*), and Ω_u (*dark gray region with solid line*) together with two trajectories of $Q(t)$ obtained by solving numerically (by using the Euler-Maruyama method with time step $\Delta t = 10^{-3}$) the Langevin model (*red line*) and the deterministic system of Eq. 2 (*blue line*) with the same, random initial condition. The two panels show two independent realizations of the models with different stochastic, initial conditions. The plots provide an example of two possible scenarios in which the intrinsic stochasticity of the Langevin model amplifies the oscillations of the deterministic system (*a*) or it triggers the emergence of new oscillations (*b*). To see this figure in color, go online.

size. Let N_0 denote the average initial number of cells and $x = X/N_0$; then, for large but finite N_0, we obtain the Langevin equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = K\mathcal{S}\,x + \sqrt{\frac{K}{N_0}}\mathcal{S}\,\boldsymbol{\eta}(t)\,, \tag{8}$$

where $\eta(t)$ is a K-dimensional white noise vector with correlator $\mathbb{E}[\eta_i(t)\eta_j(t')] = x_i\delta_{ij}\delta(t - t')$. The first term on the right describes the average behavior of the model and is the same as in Eq. 2. The second term captures the stochastic contributions arising from the finiteness of the population.

To gain more insight into the behavior of this model, we start by writing down an Ornstein-Uhlenbeck (OU) model, which approximates the behavior of the Langevin equation as a stationary Gaussian process (see Supporting materials and methods, section S1). By employing standard results for stationary processes (36,37), we obtain an expression for the time correlation matrix of x (see Supporting materials and methods, section S2). We then use the information gained from the OU process to calculate the SD of $Q(t)$, denoted by $\sigma_Q(t)$. Notice that by using the OU approximation, $Q(t)$ is defined as a ratio of Gaussian-distributed random variables, which under certain conditions (discussed in Supporting materials and methods, section S3) can be approximated by a Gaussian. Finally, we compute the envelope of two SDs about the mean of Q, defined as

$$\Omega(t) = \left[Q_* - 2\sigma_Q(t), Q_* + 2\sigma_Q(t)\right]. \tag{9}$$

We denote with Ω the envelope obtained from the Langevin model with initial random sampling, with $\overline{\Omega}$ the envelope obtained from the system of Eq. 2 with initial random

sampling and with Ω_u the envelope obtained from the Langevin model with the deterministic initial condition $x_0 = u$. In the two panels of Fig. 4, we overlay $\Omega, \overline{\Omega}$, and Ω_u, together with two numerical trajectories of Q: one (*red*) obtained by solving Eq. 8 and one (*blue*) by solving the system of Eq. 2, both initialized with the same random initial condition.

The results in Fig. 4 show that in all three cases considered, accounting for the finite-sample stochasticity can lead to a persistent departure of Q from the equilibrium value. In the two cases that account for the initial random sampling, Ω and $\overline{\Omega}$, the envelopes present an evident initial departure from equilibrium, which is sustained for several cell cycle times, halving after approximately four periods. The inherent dynamical stochasticity of the Langevin model tends to amplify the departure from the equilibrium, as evident in Ω. Interestingly, both these envelopes have slightly fluctuating edges. In contrast, the envelope initialized at the invariant distribution Ω_u shows an initial, fast expansion, followed by a phase of slower decay. Notice that Ω_u lies well inside Ω for all time. This suggests that the initial random sampling plays a role for the entire duration of the experiment. The numerical trajectories overlaid show good agreement with these findings. In particular, the solution of system of Eq. 2 (*blue line*) lies well inside $\overline{\Omega}$, whereas the simulation of the Langevin equation (*red line*) shows a larger departure, but it remains almost entirely inside the envelope Ω.

Notice that both trajectories in Fig. 4 *a* show clear oscillations about the origin with similar phases. The Langevin solution has increased oscillation amplitude in comparison to the solution of the deterministic system of Eq. 2. Although this amplification phenomenon, because of the stochasticity of the Langevin model, is common, we also find that, for some initializations, the oscillations appear only in the Langevin model and not in the deterministic model, as shown in Fig. 4 *b*.

To quantify the appearance of the oscillations, we look at the time-autocorrelation function of $G(t)$, which we define as

$$A(t, t') = \rho[G(t), G(t')]\,, \tag{10}$$

where ρ denotes the correlation coefficient (defined in Eq. S9) and can be computed using the formula for the correlation matrix (see Supporting materials and methods, section S2). Fig. 5 shows the evolution of the autocorrelation function, $A(t, t')$ as function of t', for $t = 0, 2$, and 4, respectively. In each panel, we plot $A(t, t')$ calculated analytically using the correlation matrix (*black solid line*) and the simulated value obtained by averaging 50 independent trajectories of the Langevin model (*orange line*). All three panels show a good agreement between the analytical formula and the simulated counterpart. Moreover, the results confirm the presence of strong fluctuations on the time autocorrelation
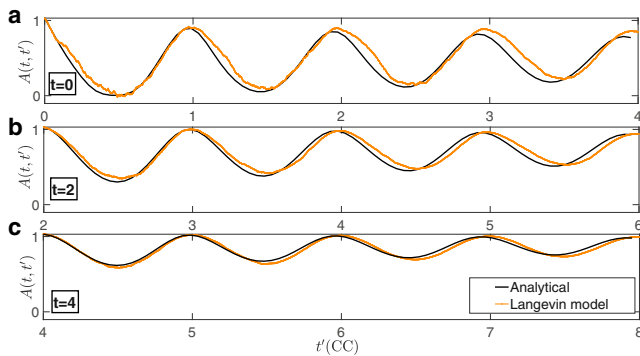
FIGURE 5   The time-autocorrelation function. The (a–c) show the time-autocorrelation function $A(t, t')$ at time $t = 0$, 2 and 4, respectively, obtained analytically from the full stochastic model (*back line*) and by averaging over 50 independent simulations of the Langevin model (*orange line*). The parameters of the model are the same as in Fig. 1. Time is normalized with respect to the average cell cycle time. To see this figure in color, go online.

of G, i.e., the number of cells in the G1-phase, with a period of exactly one cell cycle. As Q converges to the equilibrium, the amplitude of the oscillations decreases, and the autocorrelation function $A(t, t')$ tends to unity.

DISCUSSION

In this work, we highlight the importance of demographic noise to the early dynamics of growing populations with nonexponentially distributed generation times. We demonstrate that finite-sample effects can recapitulate the synchronization in the cell-cycle phase that we previously observed (16). To provide insight in these observations, we adopt a multistage approach to model both the total cell cycle time distribution and the distribution of the G1 duration, and we derive both a deterministic and a stochastic representation for the time evolution of the ratio Q. We find that the stochasticity in the initial sampling of cells can lead to a departure from the invariant distribution, which triggers a transient-oscillatory phase. The presence of intrinsic stochasticity in the dynamics tends to amplify these oscillations and delay their exponential decay.

We characterize the transient and asymptotic phases of the multistage model by deriving an analytical formula for the variance of the amplitude of the oscillations. By comparing our results with the experimental data from a proliferation assay of C8161 melanoma cells, we find that the frequency of the amplitude of the observed oscillations are consistent with those generated by demographic noise alone. Our findings suggest that finite-sample stochasticity plays a crucial role in the early stage dynamics of growing populations and that it can provide an explanation for observed synchronization in the subpopulation of the cell cycle phases.

From a theoretical point of view, our study provides a further understanding of the relation between cell-cycle distribution and global population dynamics. Although our

analysis employed a multistage model, the results of our analysis are amenable to extension to a more general type of cell-cycle distributions. In fact, for certain choices of the model parameters, the Erlang distribution adopted in this article is an excellent approximation of a Gaussian distribution. In Supporting materials and methods, section S5, we compute the relative entropy (Kullback-Leibler divergence) between Erlang and Gaussian distributions to show that the Erlang distribution tends to a Gaussian distribution as $K \to \infty$. In principle, one could use this fact to study the applicability of our findings to Gaussian cell cycle time and, hence, compare our results with other relevant studies that rely on a Gaussian approach (9,10). For example, Jafarpour (9) used a Gaussian model to study the connection between mother-daughter size regulation in bacteria and the decay of transient fluctuations. Our study focuses on synchrony emerging even in the absence of correlation of CCTs. We expect that accounting for such mechanisms will tend to amplify the amplitude of the oscillations predicted by the model; however, the analysis of this phenomenon is left for future study.

From an experimental point of view, our results highlight the routinely overlooked importance of the sample size when performing experiments that involve small populations. In particular, any data interpretation should be carried out with the role played by finite-sample stochasticity in mind. Employing larger initial populations, for example by increasing the size of microscopy images, would diminish the amplitude of the synchronization. However, because the oscillations described in this article are an intrinsic phenomenon because of the finiteness of the population, the aim of any intervention would be to mitigate their effects because they cannot be completely eradicated. Our analysis provides a novel platform to quantify the extent of finite-sample effects, which can then be used to assess their relevance in experimental contexts.

One of the central implications of this study is the need for further experiments to understand and control the fluctuating phenomena observed in proliferating cell populations. A major aspect that remains unclear and that could serve as a guideline when designing new experimental protocols is the spatial extent of the oscillatory phenomenon. Accounting for cell motility and spatial correlations of mother-daughter cells is likely to play an important role in the context of synchronizing subpopulations (38). In particular, we should expect cell motility to diminish spatial correlation, leading to faster decay of the oscillations of cells in a given window of space. To progress in this direction, we advocate that new experiments need to be undertaken with increased microscopy field of view and larger cell populations. Similarly, the accompanying theoretical framework could be adapted by studying spatial extensions of the multistage models presented in this article. Ultimately, a clearer understanding of the interplay of correlation of mother-daughter CCTs, cell motility, and population size would

help to identify the spatial scales at which the synchronization phenomenon shows up and, in turn, those at which it is irrelevant.

Although we primarily focus on explaining the synchronized oscillatory behavior observed in subpopulations of melanoma cells in a proliferation assay, we stress that the general protocol developed here will be useful to determine the extent of effects because of finiteness of the cell population in a broad class of applications. Our proliferation assays are typical experimental protocols used to investigate the efficiency of cell cycle-inhibiting drugs (4,21); hence, our findings may impact upon the reproducibility of such experiments, the efficacy of treatment protocols (22,23), and the findings of mathematical models of these experiments (39–41). Our work suggests that inherent synchronization will also occur in bacterial populations and, consequently, that studies of bacterial pathogen growth may be impacted. Many experimental protocols rely on the synchronization of cell populations to study the structural and molecular events that occur throughout the cell cycle, providing information about gene expression patterns, posttranscriptional modification, and contributing to drug discovery (24). The improved understanding of the impacts of demographic noise on the evolution of synchronous populations showcased here will shed light on the potential impact that desynchronization has on the results of these studies.

SUPPORTING MATERIAL

Supporting material can be found online at https://doi.org/10.1016/j.bpj.2021.02.017.

AUTHOR CONTRIBUTIONS

E.G. designed the analytical tools, analyzed the data, and wrote the manuscript. S.T.V. performed the experiments. N.K.H. generated the FUCCI-C8161 cells. All authors provided feedback and gave approval for final publication.

ACKNOWLEDGMENTS

REFERENCES

1. Gilbert, S. 2000. Developmental Biology. Sinauer Associates, Inc, Sunderland, MA.
2. Evan, G. I., and K. H. Vousden. 2001. Proliferation, cell cycle and apoptosis in cancer. *Nature.* 411:342–348.
3. Mort, R. L., R. J. H. Ross, …, C. A. Yates. 2016. Reconciling diverse mammalian pigmentation patterns with a fundamental mathematical model. *Nat. Commun.* 7:10288.
4. Haass, N. K., and B. Gabrielli. 2017. Cell cycle-tailored targeting of metastatic melanoma: challenges and opportunities. *Exp. Dermatol.* 26:649–655.
5. Murray, J. 2007. Mathematical biology: I. An introduction. Springer Science and Business Media, New York.
6. Monod, J. 1949. The growth of bacterial cultures. *Annu. Rev. Microbiol.* 3:371–394.
7. Laird, A. K. 1965. Dynamics of tumour growth: comparison of growth rates and extrapolation of growth curve to one cell. *Br. J. Cancer.* 19:278–291.
8. Yates, C. A., M. J. Ford, and R. L. Mort. 2017. A multi-stage representation of cell proliferation as a markov process. *Bull. Math. Biol.* 79:2905–2928.
9. Jafarpour, F. 2019. Cell size regulation induces sustained oscillations in the population growth rate. *Phys. Rev. Lett.* 122:118101.
10. Pirjol, D., F. Jafarpour, and S. Iyer-Biswas. 2017. Phenomenology of stochastic exponential growth. *Phys. Rev. E.* 95:062406.
11. Kuritz, K., W. Halter, and F. Allgöwer. 2018. Passivity-based ensemble control for cell cycle synchronization. *In* Emerging Applications of Control and Systems Theory. R. Tempo, S. Yurkovich, and P. Misra, eds. Springer, pp. 1–13.
12. Mosheiff, N., B. Martins, …, N. Balaban. 2018. Inheritance of cell-cycle duration in the presence of periodic forcing. *Phys. Rev. X.* 8:021035.
13. Vittadello, S. T., S. W. McCue, …, M. J. Simpson. 2018. Mathematical models for cell migration with real-time cell cycle dynamics. *Biophys. J.* 114:1241–1253.
14. Simpson, M., W. Jin, …, S. McCue. 2018. Stochastic models of cell invasion with fluorescent cell cycle indicators. *Physica A.* 510:375–386.
15. Riss, T., R. Moravec, …, L. Minor. 2016. Cell viability assays. *In* Assay Guidance Manual [Internet]. S. Markossian, G. S. Sittampalam, A. Grossman, K. Brimacombe, M. Arkin, D. Auld, C. P. Austin, J. Baell, J. M. M. Caaveiro, T. D. Y. Chung, N. P. Coussens, J. L. Dahlin, V. Devanaryan, T. L. Foley, M. Glicksman, M. D. Hall, J. V. Haas, S. R. J. Hoare, J. Inglese, P. W. Iversen, S. D. Kahl, S. C. Kales, S. Kirshner, M. Lal-Nag, Z. Li, J. McGee, O. McManus, T. Riss, P. Saradjian, O. J. Trask, Jr., J. R. Weidner, M. J. Wildey, M. Xia, and X. Xu, eds. Eli Lilly and Company and the National Center for Advancing Translational Sciences.
16. Vittadello, S., S. McCue, …, M. Simpson. 2019. Mathematical models incorporating a multi-stage cell cycle replicate normally-hidden inherent synchronization in cell proliferation. *J. R. Soc. Interface.* 16:20190382.
17. Sakaue-Sawano, A., H. Kurokawa, …, A. Miyawaki. 2008. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell.* 132:487–498.
18. Haass, N. K., K. A. Beaumont, …, W. Weninger. 2014. Real-time cell cycle imaging during melanoma growth, invasion, and drug response. *Pigment Cell Melanoma Res.* 27:764–776.
19. Gavagnin, E., M. J. Ford, …, C. A. Yates. 2019. The invasion speed of cell migration models with realistic cell cycle time distributions. *J. Theor. Biol.* 481:91–99.
20. Baker, R., and G. Röst. 2019. Global dynamics of a novel delayed logistic equation arising from cell biology. *J. Nonlin. Sci.* 30:397–418.
21. Beaumont, K. A., D. S. Hill, …, N. K. Haass. 2016. Cell cycle phase-specific drug resistance as an escape mechanism of melanoma cells. *J. Invest. Dermatol.* 136:1479–1489.
22. Welsh, S. J., H. Rizos, …, G. V. Long. 2016. Resistance to combination BRAF and MEK inhibition in metastatic melanoma: where to next? *Eur. J. Cancer.* 62:76–85.
23. Hill, D. S., S. Martin, …, P. E. Lovat. 2009. Combining the endoplasmic reticulum stress-inducing agents bortezomib and fenretinide as a novel therapeutic strategy for metastatic melanoma. *Clin. Cancer Res.* 15:1192–1198.
24. Banfalvi, G. 2017. Cell Cycle Synchronization. Springer, New York.

25. Davies, M. A., K. Stemke-Hale, ..., J. E. Gershenwald. 2009. Integrated molecular and clinical analysis of AKT activation in metastatic melanoma. *Clin. Cancer Res.* 15:7538–7546.

26. Smalley, K. S., R. Contractor, ..., M. Herlyn. 2007. Ki67 expression levels are a better marker of reduced melanoma growth following MEK inhibitor treatment than phospho-ERK levels. *Br. J. Cancer.* 96:445–449.

27. Spoerri, L., K. Beaumont, ..., N. K. Haass. 2017. Real-time cell cycle imaging in a 3D cell culture model of melanoma. *Methods Mol. Biol.* 1612:401–416.

28. Davis, P. K., A. Ho, and S. F. Dowdy. 2001. Biological methods for cell-cycle synchronization of mammalian cells. *Biotechniques.* 30:1322–1326, 1328, 1330–1331.

29. Taylor, I. W., and P. J. Hodson. 1984. Cell cycle regulation by environmental pH. *J. Cell. Physiol.* 121:517–525.

30. Smalley, K. S., P. Brafford, ..., M. Herlyn. 2005. Up-regulated expression of zonula occludens protein-1 in human melanoma associates with N-cadherin and contributes to invasion and adhesion. *Am. J. Pathol.* 166:1541–1554.

31. Beaumont, K. A., A. Anfosso, ..., N. K. Haass. 2015. Imaging- and flow cytometry-based analysis of cell position and the cell cycle in 3D melanoma spheroids. *J. Vis. Exp.* 106:e53486.

32. Strässle, C., B. Sonnleitner, and A. Fiechter. 1988. A predictive model for the spontaneous synchronization of *Saccharomyces cerevisiae* grown in continuous culture. I. Concept. *J. Biotechnol.* 7:299–317.

33. Strässle, C., B. Sonnleitner, and A. Fiechter. 1989. A predictive model for the spontaneous synchronization of *Saccharomyces cerevisiae* grown in continuous culture. II. Experimental verification. *J. Biotechnol.* 9:191–208.

34. Paris, R. B. 2002. Exponential asymptotics of the Mittag–Leffler function. *Proc. R. Soc., Ser. B, London.* 458:3041–3052.

35. Morris, R. G., and T. Rogers. 2014. Growth-induced breaking and unbreaking of ergodicity in fully-connected spin systems. *J. Phys. A Math. Theor.* 47:342003.

36. Gardiner, C. 2009. Stochastic Methods. Springer Berlin, New York.

37. Van Kampen, N. G. 1992. Stochastic Processes in Physics and Chemistry. Elsevier, Amsterdam.

38. Baker, R. E., and M. J. Simpson. 2010. Correcting mean-field approximations for birth-death-movement processes. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 82:041905.

39. Altinok, A., F. Lévi, and A. Goldbeter. 2009. Identifying mechanisms of chronotolerance and chronoefficacy for the anticancer drugs 5-fluorouracil and oxaliplatin by computational modeling. *Eur. J. Pharm. Sci.* 36:20–38.

40. Clairambault, J. 2011. Optimizing cancer pharmacotherapeutics using mathematical modeling and a systems biology approach. *Per. Med.* 8:271–286.

41. Lévi, F. 2006. Chronotherapeutics: the relevance of timing in cancer therapy. *Cancer Causes Control.* 17:611–621.

Supplemental information

Synchronized oscillations in growing cell populations are explained by demographic noise

Enrico Gavagnin, Sean T. Vittadello, Gency Gunasingh, Nikolas K. Haass, Matthew J. Simpson, Tim Rogers, and Christian A. Yates

This document contains the supplementary materials which accompany the paper Gavagnin et al. [1]. Sections S.1, S.2 and S.3 contain some details of the mathematical derivation of the analytical formula for the envelope of two standard deviation Q. In Section S.4 we explain the method adopted to parametrise our multi-stage model from the experimental images. Section S.5 contains the computation of the relative entropy between an Erlang and a Gaussian distribution. Section S.6 contains supplementary figures. All accompanying codes of the publication (MATLAB 2019b) can be found at https://github.com/EnricoGavagnin/sync_multi_stage.git

S.1. THE ORNSTEIN-UHLENBECK APPROXIMATION

We can simplify the Langevin model given by Eq. 8 of the main document by replacing the dependence on $\boldsymbol{x}$ in the correlator of $\boldsymbol{\eta}(t)$ with $\mathbb{E}\left[\boldsymbol{x}\right] = K\boldsymbol{u}e^{K\lambda t}$. The resulting equation consists of the high-dimensional non-autonomous Ornstein-Uhlenbeck (OU) process

$$\frac{\mathrm{d}\hat{\boldsymbol{x}}}{\mathrm{d}t} = K\mathcal{S}\hat{\boldsymbol{x}} + K\sqrt{\frac{e^{K\lambda t}}{N_0}}\mathcal{S}\boldsymbol{\psi}(t)\,, \tag{S.1}$$

where $\boldsymbol{\psi}(t)$ is a K-dimensional white noise vector with correlator $\mathbb{E}[\eta_i(t)\eta_j(t')] = u_i\delta_{ij}\delta(t - t')$. We explore the behaviour of the two models, the OU process given by Eq. S.1, and the Langevin equation 8 of the main document in Fig. S.1. The results suggest that the OU process is an accurate approximation of the Langevin equation, in particular the presence of the oscillations is evident in both the modelling regimes (Fig. S.1(a)). In Fig. S.1(b), we compare the distributions of $Q(t)$ at $t = 1, 3$ and 5 obtained by averaging over 1000 independent simulations, which show good agreement between the two models.
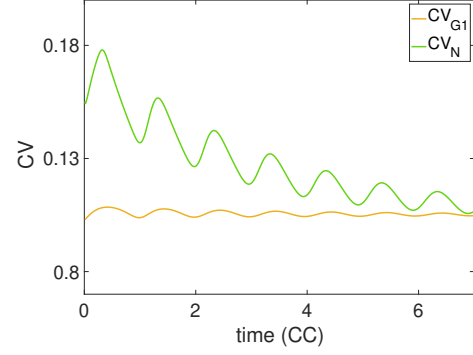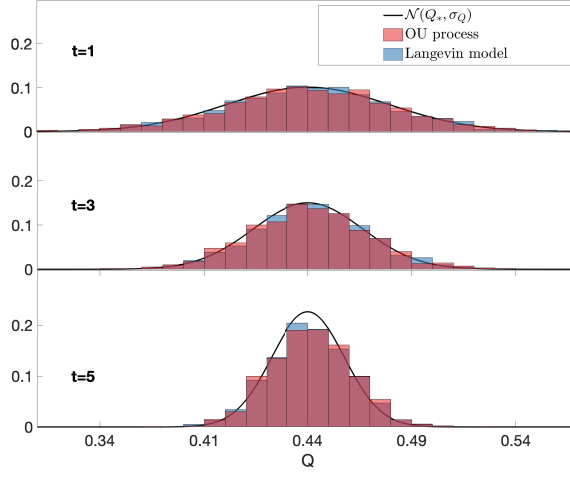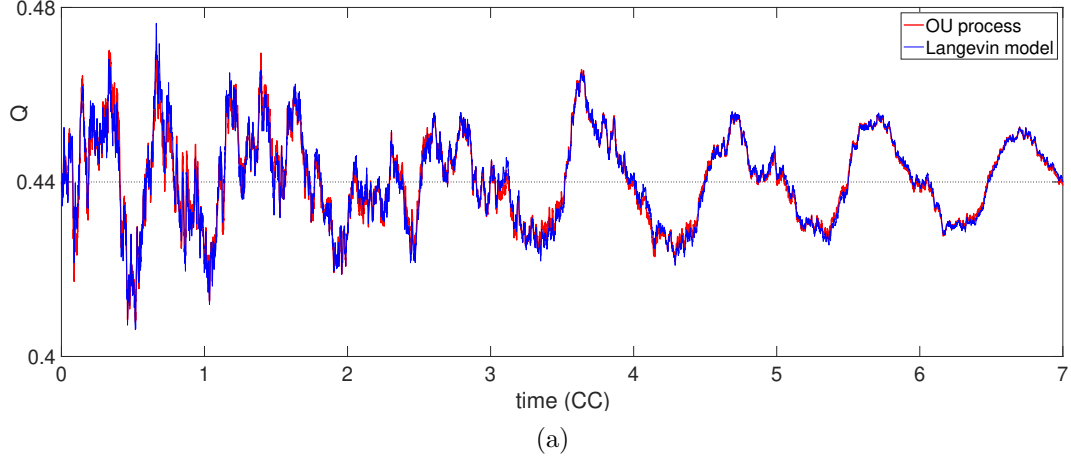
Fig. S.1. Comparison of OU process and the Langevin equation model. Panel (a) shows two evolutions of Q for the OU process (red) and the Langevin equation (blue). The two trajectories are realised using Euler-Maruyama method with time step $\Delta t = 10^{-3}$ and the same randomly generated numbers. In panel (b) we plot the distribution of Q at three time points ($t = 1$, $t = 3$ and $t = 5$). The two overlaid histograms represent the distributions of 1000 independent simulations of the OU process (red) and the Langevin Equation (blue). The black line represent the distribution $N(Q_*, \sigma_Q(t))$. All the parameters are the same as Fig. 2 of the paper. Panel (c) shows the values of $CV_{G1}(t)$ (green line) and $CV_N(t)$ (yellow line) indicating that Q will be Gaussian distributed for the chosen parameter values. The parameters are the same as Fig. 2 of the paper and time is normalised with reference to the average cell-cycle time.

S.2. THE CORRELATION MATRIX

For a stochastic initial condition, $\boldsymbol{x}_0$, as described in Section III C of the main document, we can compute the correlation matrix at time $t = 0$, as

$$C_0 = \mathbb{E}\left[x_i(0)x_j(0)\right] = \begin{cases} u_i u_j & \text{for } i \neq j \\ u_i^2 + \frac{u_i}{v} & \text{for } i = j \end{cases}. \tag{S.2}$$

We can rewrite this as $C_0 = \boldsymbol{u}\boldsymbol{u}^T + \frac{1}{v}M$, where $M = Diag(\boldsymbol{u})$.

We then focus on computing the correlation matrix $C(t, t') = \mathbb{E}\left[\hat{\boldsymbol{x}}(t)\,\hat{\boldsymbol{x}}^T(t')\right]$ for the OU process S.1, as an approximation for the correlation matrix of the Langevin model. By applying general results for OU processes (See Section 4.5 of [2]) we have:

$$C(t, t') = e^{Kt\mathcal{S}}C_0 e^{Kt'\mathcal{S}^T} + \frac{K}{N_0}\int_0^{\min(t,t')} e^{K(t-\tau)\mathcal{S}}\mathcal{S}\left(Me^{K\lambda\tau}\right)\mathcal{S}^T e^{K(t'-\tau)\mathcal{S}^T}\,d\tau. \tag{S.3}$$

We can use expression 5 of the main document and the fact that $e^{Kt\mathcal{S}}\boldsymbol{u}\boldsymbol{u}^T e^{Kt'\mathcal{S}^T} = \boldsymbol{u}\boldsymbol{u}^T e^{K(t+t')\lambda}$, to write down the (i,j) element of S.3 for $t < t'$ as

$$C_{i,j}(t, t') = u_i u_j e^{K(t+t')\lambda} + \frac{1}{vK^2}\sum_{k,l,m=1}^{K}\frac{1}{(1+\lambda_k)^{i-m}}\frac{1}{(1+\lambda_l)^{j-m}}\frac{2\lambda}{(1+\lambda)^m}e^{K(t\lambda_k + t'\lambda_l)}$$

$$+ \frac{1}{N_0 K}\sum_{k,l,m=1}^{K}\frac{\lambda_k}{(1+\lambda_k)^{i-m}}\frac{\lambda_l}{(1+\lambda_l)^{j-m}}\frac{2\lambda}{(1+\lambda)^m}\int_0^t e^{K[(t-\tau)\lambda_k + (t'-\tau)\lambda_l + \lambda\tau]}.$$

$$\tag{S.4}$$

Substituting the expressions 3 of the main text for $\boldsymbol{u}^k$ and $\boldsymbol{v}^k$, using the formula $\sum_{m=1}^{K}(1+\lambda_k)^m(1+\lambda_l)^m/(1+\lambda)^m = (1+\lambda_k)(1+\lambda_l)/[(1+\lambda_k)(1+\lambda_l)-(1+\lambda)]$ and

4

upon rearranging terms, we obtain

$$
\begin{aligned}
C_{i,j}(t,t') = & \frac{4\lambda^2}{(1+\lambda)^{i+j}} e^{K(t+t')\lambda} \\
& + \frac{2\lambda}{K^2} \sum_{k,l=1}^{K} \frac{(1+\lambda_l)^{1-j}(1+\lambda_k)^{1-i}}{(1+\lambda_k)(1+\lambda_l)-(1+\lambda)} \left[\frac{1}{v} e^{K(t\lambda_k+t'\lambda_l)} \right. \\
& \left. - \frac{\lambda_k \lambda_l}{N_0(\lambda-\lambda_k-\lambda_l)} \left(e^{K(t\lambda_k+t'\lambda_l)} - e^{K((t'-t)\lambda_l+t\lambda)} \right) \right].
\end{aligned}
\tag{S.5}
$$

It should be noted that alternative techniques could be employed to approximate the correlation matrix of the Langevin model. For instance, it is possible to write down the equation for the second and third moment of $\boldsymbol{x}$ directly from the Langevin equation (Eq. 8 of the main document) and to employ a moment closure approximation in order to truncate the dependency on the fourth moment. For reasonable choices of closure approximations, we expect such an approach would lead to similar formulae to those we obtained in this section by using the OU approximation.

S.3. THE ENVELOPE OF TWO STANDARD DEVIATIONS OF Q

We recall the definition of the envelope of two standard deviations of $Q(t)$ as $\Omega(t) = [Q_* - 2\sigma_Q(t), Q_* + 2\sigma_Q(t)]$, where $\sigma_Q(t)$ denotes the standard deviation of $Q(t)$. To compute σ_Q we employ the OU approximation (see Section S.1). From a fixed initial condition, the solutions of S.1 evolve as a Gaussian process with mean $\bar{\boldsymbol{x}}(t)/N_0$. We can write $G(t) \sim \mathcal{N}(\mu_G(t), \sigma_G(t))$ and $N(t) \sim \mathcal{N}(\mu_N(t), \sigma_N(t))$ where

$$
\mu_G(t) = Q_* e^{K\lambda t}, \qquad \sigma_G^2(t) = \sum_{i,j=1}^{\alpha K} C_{i,j}(t,t) - Q_*^2 e^{2K\lambda t}, \tag{S.6a}
$$

$$
\mu_N(t) = e^{K\lambda t}, \qquad \sigma_N^2(t) = \sum_{i,j=1}^{K} C_{i,j}(t,t) - e^{2K\lambda t}. \tag{S.6b}
$$

Notice that, Q is defined as a ratio between two Gaussian distribution and, in general, this does not imply that $Q(t)$ is Gaussian. However, Hayya et al. [3] showed that the ratio of two Gaussian can be well approximated as a Gaussian, under certain conditions on the *coefficient of variation* (CV) of the numerator and denominator. Precisely, provided that

$$CV_N = \frac{\sigma_N}{\mu_N} < 0.39 \qquad \text{and} \qquad CV_{G1} = \frac{\sigma_{G1}}{\mu_{G1}} > 0.005 \qquad (\text{S.7})$$

Hayya et al. [3] demonstrate that Q is close to a Gaussian distribution. Moreover, we can approximate the variance of Q by Taylor expanding to the second order which leads to

$$
\begin{aligned}
\sigma_Q^2 &\approx \sigma_N^2 \frac{\mu_G^2}{\mu_N^4} + \frac{\sigma_G^2}{\mu_N^2} - 2\rho\mu_G \frac{\sigma_N \sigma_G}{\mu_N^3} \\
&= \frac{1}{\mu_N^2} \left[\sigma_N^2 Q_*^2 + \sigma_G^2 - 2\sigma_N \sigma_G Q_* \, \rho\,[G, N] \right] ,
\end{aligned}
\qquad (\text{S.8})
$$

where ρ denotes the correlation coefficient, defined as

$$\rho\,[Y_1, Y_2] = \frac{\mathbb{E}\,[Y_1 Y_2] - \mathbb{E}\,[Y_1]\,\mathbb{E}\,[Y_2]}{\sqrt{\text{Var}\,[Y_1]\ \text{Var}\,[Y_2]}} . \qquad (\text{S.9})$$

Notice that we can compute $\mathbb{E}\,[G(t)N(t)]$ in Eq. S.9 in terms of the correlation matrix C as

$$\mathbb{E}\,[G(t)N(t)] = \sum_{i=1}^{K} \sum_{j=1}^{\alpha K} C_{i,j}(t,t) .$$

Finally we check that the conditions S.7 are satisfied for our parameter choices. In Fig. S.1 (c) we report the value of CV_{G1} (yellow line) and CV_N (green line) for the parameter values obtained in Section S.4. The two coefficients lie well inside the range indicated by Hayya et al. [3], suggesting that the ratio Q is close to a Gaussian distribution for the parameters considered. Notice that the plots in Fig. S.1(b) provide

6

further confirmation of this by showing good agreement between the distribution of Q and the Gaussian distribution $\mathcal{N}(Q_*, \sigma_Q(t))$.

S.4. PARAMETER INFERENCE

To infer the parameters of the multi-stage model, we simultaneously fit the distribution of the total cell-cycle time and of the G1 duration of 200 randomly selected cells.

Let $\boldsymbol{H}_T$ and $\boldsymbol{H}_{G1}$ denote the histogram representations of the probability density function (pdf) of the total cell-cycle time and the G1 duration, respectively, with a bin width of one hour. For example, $(\boldsymbol{H}_T)_i$ denotes the proportion of cells with a cell-cycle time in the interval $[ih, (i+1)h)$. We denote with $\boldsymbol{H}_{E(K,\beta)}$ the histogram obtained by discretising an Erlang distribution with parameters (K, β) with the same bin width, i.e. $(\boldsymbol{H}_{E(K,\beta)})_i = \frac{\beta^K}{(K-1)!} \int_i^{i+1} x^{K-1} e^{-\beta x} dx$.

For a given combination of parameters, (K, β, α), one can consider the distance measure

$$I(K, \beta, \alpha) = \|\boldsymbol{H}_T - \boldsymbol{H}_{E(K,\beta)}\|_1 + \|\boldsymbol{H}_{G1} - \boldsymbol{H}_{E(\alpha K, \beta)}\|_1 \,, \qquad \text{(S.10)}$$

where $\| - \|_1$ denotes the 1-norm. To determine the parameter combination which provides the best simultaneous fit of the two distribution, we evaluated the function I in the parameter range $K \in [10, 150]$ $\beta \in [1, 10]$ and $\alpha \in [0, 1]$. We find that the combination $K^* = 92$, $\beta^* = 4.96$ and $\alpha^* = 33/92$ minimises the statistic I in the parameter region considered and, hence, we select these parameters for the multi-stage model.

In order to determine the level of initial stochasticity of the model, we choose the value of the parameter v which provides the best fit between the experimental standard deviation and the one predicted from the model (see Figure 2 (b) of the main text).

7

This is done by minimising the 1-norm of the distance between the experimental time series of $\sigma_Q(t)$, obtained from the 30 experimental trajectories, and the corresponding theoretical time series obtained using formula S.5. After performing a direct parameter exploration for $v \in \{\delta_v k \,|\, \delta_v = 0.1, \; k = 1 \ldots 10^3\}$ we found that the distance between the two trajectories is minimised by choosing $v = 94.3$, hence we select this value as parameter for the model. The initial average population size, N_0 is chosen to match the experimental value, averaged over the 30 repeats, $N_0 \approx 381$.

S.5. THE KULLBACK LEIBLER DIVERGENCE BETWEEN ERLANG AND GAUSSIAN DISTRIBUTION

We compute the relative entropy (Kullback-Leibler divergence, D_{KL}) between an Erlang and a Gaussian distribution as a measure of the distance between the two distributions.

For two distributions, $p(x)$ and $q(x)$, the KL divergence is defined as:

$$D(p,q) = \int_{-\infty}^{\infty} p(x) \log \left[\frac{p(x)}{q(x)} \right] dx \,. \tag{S.11}$$

We set $p(x)$ to be the probability density function (pdf) of an $Erlang(K, \beta)$ and $q(x)$ to be the pdf of a Gaussian with same mean and variance, *i.e.* $\mathcal{N}\left(\frac{K}{\beta}, \frac{K}{\beta^2} \right)$. We then obtain

$$D(K,\beta) = D\left(p\left(K,\beta\right), q\left(\frac{K}{\beta}, \frac{K}{\beta^2}\right)\right)$$

$$= \frac{\beta^K}{(K-1)!} \int_0^\infty x^{K-1} e^{-\beta x} \log\left[\frac{\beta^{K-1}\sqrt{2\pi K}}{(K-1)!} x^{K-1} e^{\frac{\beta^2}{2K}\left(x - \frac{K}{\beta}\right)^2 - \beta x}\right] dx$$

$$= \log\left[\frac{\beta^{K-1}\sqrt{2\pi K}}{(K-1)!}\right] \frac{\beta^K}{(K-1)!} \int_0^\infty x^{K-1} e^{-\beta x} dx$$

$$+ (K-1)\frac{\beta^K}{(K-1)!} \int_0^\infty x^{K-1} \log(x) e^{-\beta x} dx \qquad \text{(S.12)}$$

$$+ \frac{\beta^K}{(K-1)!} \int_0^\infty \left[-\beta x + \frac{\beta^2}{2K}\left(x - \frac{K}{\beta}\right)^2\right] x^{K-1} e^{-\beta x} dx\,.$$

Notice that the first integral of Eq. S.12 is exactly the pdf of an Erlang distribution, which simplifies to unity. The second and third integral in S.12 require more work. By using integration by parts and upon simplification, we get to the final expression

$$D(K,\beta) = \log\left[\frac{\beta^{K-1}\sqrt{2\pi K}}{(K-1)!}\right] + (K-1)\left(\mathcal{H}_{K-1} - \log(\beta) - \gamma\right) - K + \frac{1}{2}, \qquad \text{(S.13)}$$

where $\mathcal{H}_{K-1} = \sum_{i=1}^{K-1} \frac{1}{i}$ is the $(K-1)$-th harmonic number and $\gamma = \lim_{n\to+\infty}\left(\sum_{i=1}^n \frac{1}{i} - \log(n)\right)$ denotes the Euler-Mascheroni constant.

Using the expression of Eq. S.13 it is possible to show that $D(K,\beta)$ is a decreasing function of K and $D(K,\beta) \sim \mathcal{O}(K^{-1})$ for $K \to +\infty$. This is not a surprise, since by central limit theorem we know that the Erlang distribution converges to a Gaussian with same mean and variance. Since the CV of an Erlang(K,β) is given by $K^{-\frac{1}{2}}$, we can rephrase by saying that the KL divergence scales proportionally to the square of the CV of the Erlang distribution.

Fig. S.2 shows the plot of $D(K,\beta)$ with $\beta = K$ for different values of K. In the overlaid panels the two distributions are compared for $K = 5, 10, 20, 30$ and 60. The results
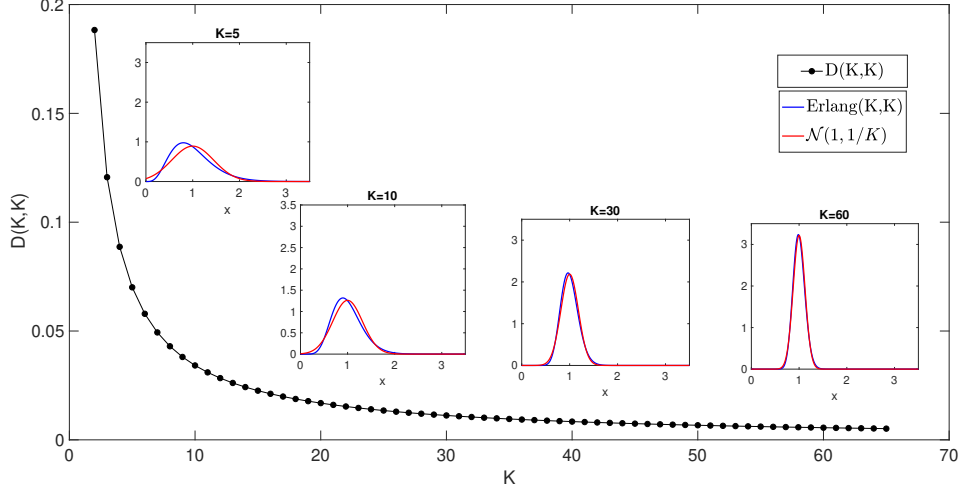
Fig. S.2. The Kullback Leibler (KL) divergence between an Erlang distribution and Gaussian distribution. The black dotted line in the main panel shows the KL divergence between an Erlang distribution of parameters (K, K) and a Gaussian distribution of parameters $(1, 1/K)$ as function of K. The four overlaid panels show the comparison of the two distributions, Erlang (blue) and Gaussian (Red), for $K = 5, 10, 30$ and 60 (from left to right).

highlight the good level of similarity between the Erlang and Gaussian distributions for large K - small values of the CV. For example, for $K > 25$, *i.e.* CV< 0.2, we have $D(K, K) < 0.02$ which corresponds to good agreement between the two distributions.
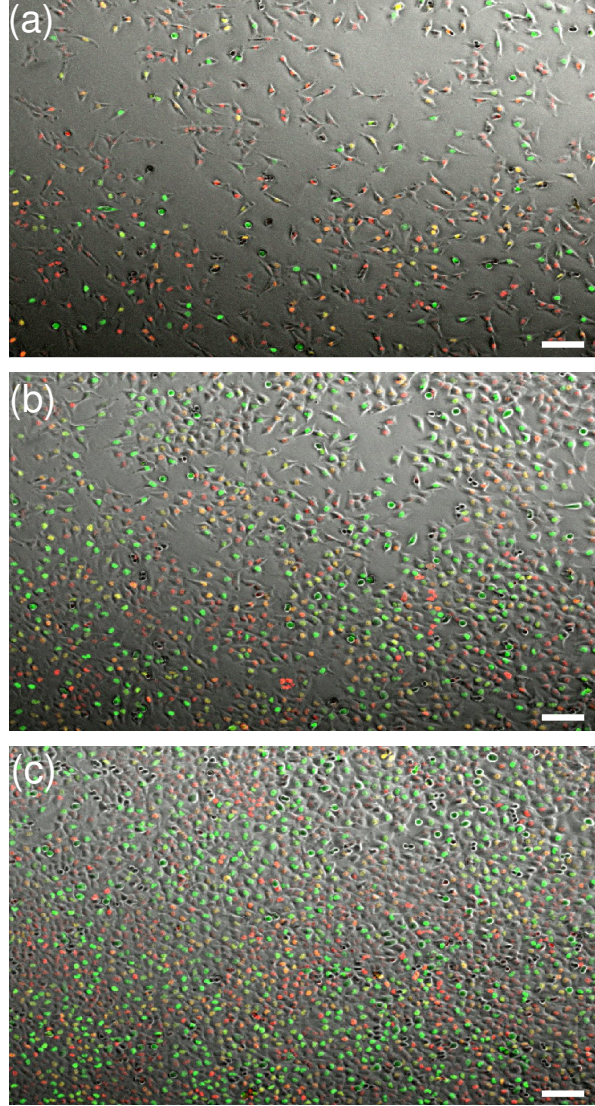
S.6. SUPPLEMENTARY FIGURES

Fig. S.3. Microscopy images of proliferation assay. Panels (a) to (c) show three snapshots of an epifluorescence microscopy image time series - merged images of bright-field and the red and green fluorescence channels. Panel (a) is taken at the beginning of the recording, $t = 0$ h, panel (b) is taken half-way through the recording, $t = 24$ h, and at panel (c) at the end of the recording, $t = 48$ h (scale bar $100\mu m$). The cells with red nuclei are in the G1 (gap 1) phase of the cell cycle, the ones with yellow nuclei are in the eS (early-synthesis) phase and those with green nuclei are in one of the remaining consecutive phases: S (synthesis), G2 (gap 2) or M (mitosis).
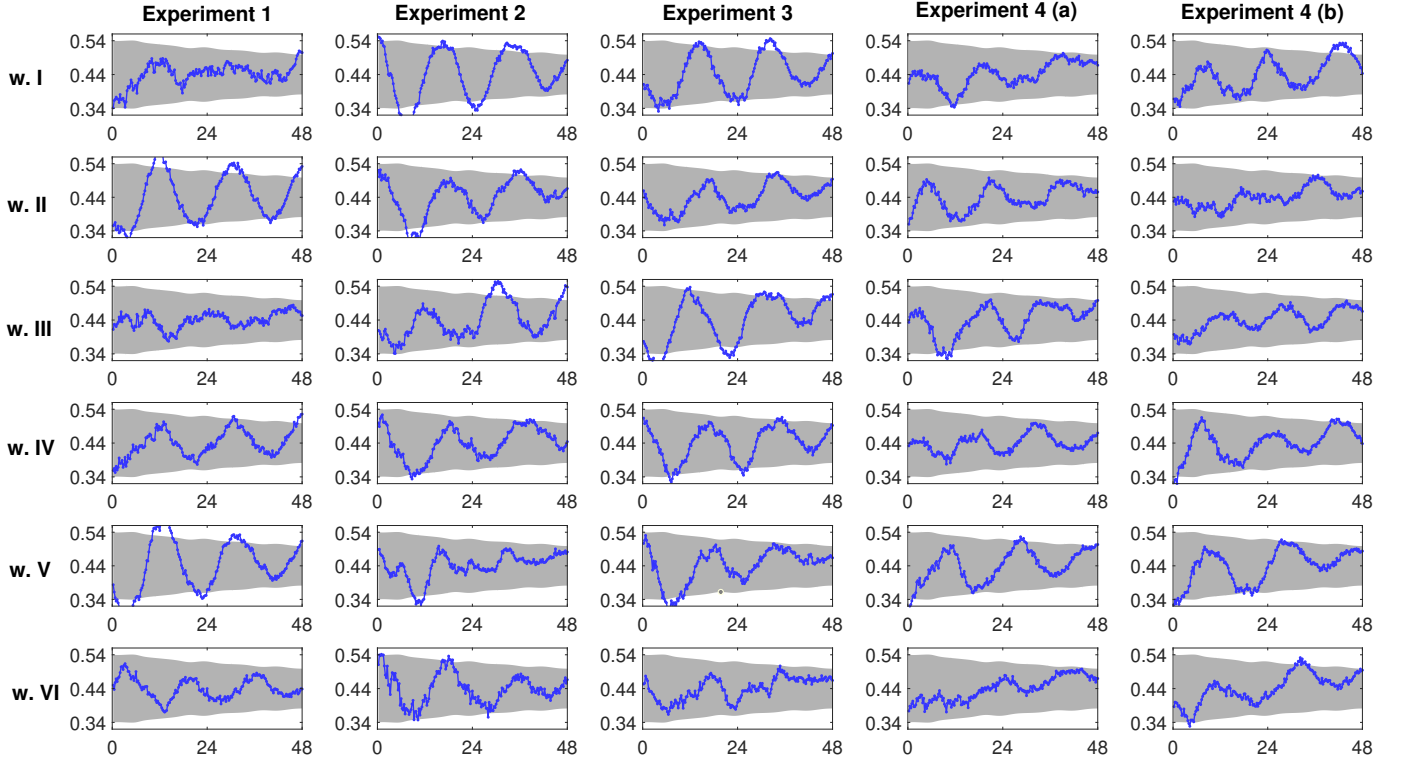
Fig. S.4. Comparison 30 time series obtained from the data (blue lines), together with the envelope of two standard deviations, Ω (light grey regions) predicted using the multi-stage model. The parameters of the multi-stage models are obtain by fitting the distribution of the total cell-cycle time and G1 duration (see Section S.4): $K = 92, \alpha K = 33, \beta = 4.96\text{h}^{-1}, v = 94.3$ and $N_0 = 381$.
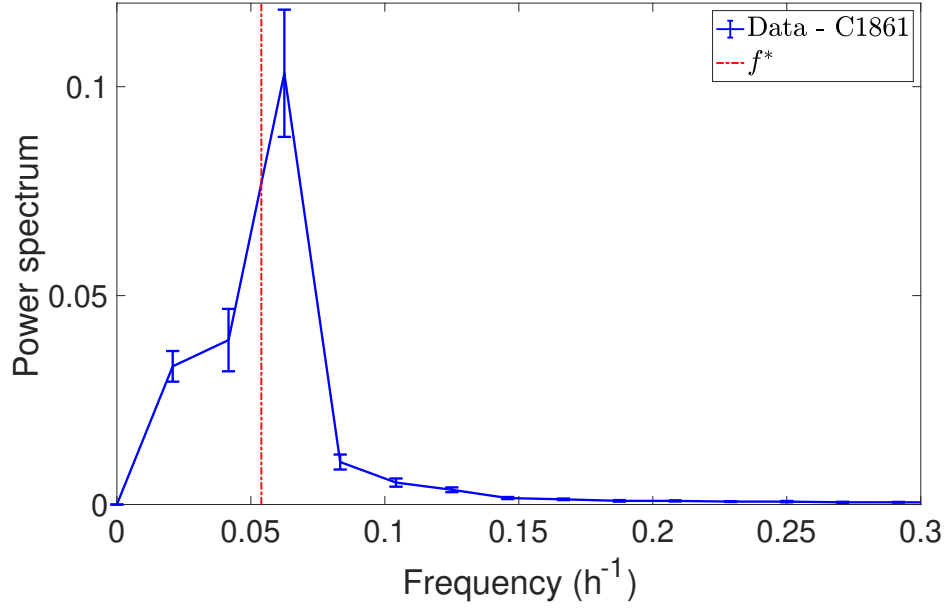
Fig. S.5. Average Fourier power spectrum analysis of the time series of $Q(t)$. The plot shows the experimental Fourier power spectrum of $Q(t)$ averaged over the 30 time series (blue lines - bars denote interval of one standard deviation). The red line highlights the dominant frequency predicted from the multi-stage model, corresponding to one CCT, $f^* = \beta/K = 0.054\,\mathrm{h}^{-1} = (18.5\ \mathrm{h})^{-1}$.

[1] E. Gavagnin, S.T Vittadello, G. Gunasingh, N.K Haass, M.J. Simpson, T. Rogers, and C.A. Yates. Synchronised oscillations in growing cell populations are explained by demographic noise. *t.b.a.*, 2020.

[2] C. Gardiner. *Stochastic methods*, volume 4. Springer Berlin, 2009.

[3] J. Hayya, D. Armstrong, and N. Gressis. A note on the ratio of two normally distributed variables. *Manag. Sci.*, 21(11):1338–1341, 1975.