Statistical properties and dynamics of phenotype components in individual bacteria

Lee Susman\textsuperscript{1,2}, Maryam Kohram\textsuperscript{3}, Harsh Vashistha\textsuperscript{3}, Jeffrey T. Nechleba\textsuperscript{3}, Hanna Salman\textsuperscript{3,4}, and Naama Brenner\textsuperscript{1,5,*}

\textsuperscript{1}Network Biology Research Lab, Technion - Israel Institute of Technology, Haifa 32000, Israel
\textsuperscript{2}Interdisciplinary Program of Applied Math, Technion - Israel Institute of Technology, Haifa 32000, Israel
\textsuperscript{3}Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA
\textsuperscript{4}Department of Computational and Systems Biology, University of Pittsburgh, Pittsburgh, PA 15260, USA
\textsuperscript{5}Department of Chemical Engineering, Technion - Israel Institute of Technology, Haifa 32000, Israel

*Corresponding Author: nbrenner@technion.ac.il

Abstract

Cellular phenotype is characterized by different components such as cell size, protein content and cell cycle time. These are global variables that are the outcome of multiple internal microscopic processes. Accordingly, they display some universal statistical properties and scaling relations, such as distribution collapse and relation between moments. Cell size statistics and its relation to growth and division has been mostly studied separately from proteins and other cellular variables. Here we present experimental and theoretical analyses of these phenotype components in a unified framework that reveals their correlations and interactions inside the cell. We measure these components simultaneously in single cells over dozens of generations, quantify their correlations, and compare population to temporal statistics. We find that cell size and highly expressed proteins have very similar dynamics over growth and division cycles, which result in parallel statistical properties, both universal and individual. In particular, while distribution shapes of fluctuations along time are common to all cells and components, other properties are variable and remain distinct in individual cells for a surprisingly large number of generations. These include temporal averages of cell size and protein content, and the structure of their auto-correlation functions. We explore possible roles of the different components in controlling cell growth and division. We find that in order to stabilize exponential accumulation and division of all components across generations, coupled dynamics among them is required. Finally, we incorporate effective coupling within the cell cycle with a phenomenological mapping across consecutive cycles, and show that this model reproduces the entire array of experimental observations.
Significance. A population of microorganisms grown from a single ancestor share identical genetic material and common history. Nevertheless they exhibit broad variability in all measured properties: size, shape, molecular content etc. How this variability is produced despite the strong inheritance in a population is a central question in biophysics. In this work we utilize a special experimental system that enables us to monitor two properties simultaneously: cell size and protein copy number. We measure them in individual bacteria as they grow, divide and transfer them across many generations. Based on these measurements we develop a dynamical model of a single cell that takes into account the coupling between different cellular properties, their exponential accumulation within the cell cycle, and the control of cell division. The model reproduces the statistical properties of phenotypic variability observed in populations.

Introduction

Phenotypic variability is a hallmark of microbial cell populations, even when clonal and grown under uniform conditions [1, 2, 3]. Variability appears in many measured cellular properties, including cell size [4, 5, 6, 7, 8], molecular content [9, 10], organelle copy number [11], cell-cycle time [12, 13, 14], response timescale [15] and more. A clonal population is produced from a single ancestor by growth and division. Many cellular properties are, to some degree, inherited along generations. Thus phenotypic variability is tightly connected to growth and division dynamics [16, 17, 18]. These dynamics are regulated by internal cellular processes; the problem of growth and division regulation has been studied for decades, with recent renewed interest following single-cell technologies [19].

Of special interest and biological relevance among phenotypic properties is cellular protein content, which can determine functional characteristics of the cell. In microorganisms, the copy number of all proteins fluctuates considerably among clonal cells [20, 9]. Large variability is expected for low copy-number proteins, where number fluctuations and molecular noise dominate. However, highly expressed proteins – where number fluctuations are negligible – also exhibit large variability among cells [21, 22, 23, 24]. These fluctuations have been attributed to multiple sources, such as noise in the protein production process [25, 26], cell division [27, 28, 29], and growth rate [30]. Recent studies have focused on the transcription process, shedding light on variability in copy number of mRNA [31, 32, 33]. We note however, that there is a timescale separation between mRNA and protein lifetimes; as a result, there is little correlation between the two [24]. Moreover, possibly due to their buffering from molecular processes, highly expressed protein variability was found to display some simple and universal features: First, the distributions of such proteins exhibit a collapse under two-parameter scaling, for both bacteria and yeast under a broad range of conditions [34]. Qualitatively similar distribution shapes had been reported in various experimen-
tal works [35, 23, 22, 24, 36]. Second, for highly expressed proteins, the variance $\sigma^2$ is a quadratic function of their mean $\mu$ ($\sigma^2 = A\mu^2 + B\mu + C$; see [34]). This relation holds over large datasets, including genome-wide measurements (24; see also [34] supplementary material). Such universal features under a broad range of conditions suggests that the plethora of noisy molecular processes affecting protein content, are integrated to ultimately buffer cellular statistics from the molecular realm.

Universal distributions and scaling relations between moments have been observed also for cell size and, more generally, body mass in an ecological context. For cell size it was found however that distributions collapse under a one-parameter scaling, implying a more stringent relation ($\sigma^2 \propto \mu^2$) between mean and variance [37, 38, 39, 7, 8]. Since cell-size is linked to the control of cell division, statistical properties of cell-size have been considered in this context [5, 39, 8, 7]. Although cell mass is composed largely of proteins, the statistical properties of highly expressed specific proteins have been mostly studied separately; a quantitative understanding of the correlations between cell size and various highly expressed proteins, is not well developed.

Recently, measurements of cellular properties such as protein content and cell size in individual bacteria over extended times have become available [19]. These measurements provide new insight into the statistical properties of populations by revealing dynamical aspects of cell growth and protein expression [40, 39, 41, 8, 42]. Our previous work has shown that, for E. coli, the single-cell protein traces exhibit the same universal statistical properties as large populations, showing the same distribution shape and scaling between moments [43, 41]. In addition, the single-cell traces of cell-size and abundant protein are not only qualitatively similar but also statistically correlated on a cycle-by-cycle basis [41]. However, a general and fundamental description of the relation between single-cell dynamics and population statistics is still lacking. Statistical and scaling properties of cell-size and highly expressed proteins in microorganisms suggest that they are governed by similar or related coarse-grained mechanisms. These relations between cell-size, protein content and cell-division control are at the focus of our current study.

Here, we present a comprehensive experimental and theoretical analysis of single-cell dynamics along multiple generations, offering a unified framework for “phenotype components” - cell-size and highly expressed protein content. Stable measurements over dozens of generations allow us to clearly identify individual statistical features of single cells that are maintained distinct throughout their lifetime. These features are contrasted and compared to universal ones summarized above, that are common to all single-cell traces and population snapshots. Both universal and individual statistical properties are practically identical for cell-size and protein traces.

A stochastic mapping model relating consecutive generations is applied first to each phenotype component separately. Cycles of exponential growth and division converge to a stable process
only with some restraining force on the exponents; the resulting mapping model is equivalent to a particle under the influence of an effective potential (from which an effective restraining force is derived), and a noise source [5, 44, 45]. This mapping accounts quantitatively for both universal and individual statistical properties of each component. In particular we find that the effective potential is an individual property distinct among cells; this in turn results in a distinct long-time average of phenotype components, and distinct long-time oscillatory structure in their auto-correlation function.

The effective restraining force acting on exponential growth has a biological meaning in terms of the control of cell division [5, 46, 47]. The similarity between the effective description of proteins and cell size raises the question of the relation between them and how they coordinate control of cell division. We therefore next consider the phenotype components as parts of a single stochastic dynamical system. New data are presented on phenotype components measured simultaneously in single cells. These data reveal strong correlations between phenotype components but also variability among generations and among individual cells. We find that a multi-component dynamical system with cycles of exponential growth and division, can only be stable if components interact dynamically during the cell-cycle. We develop an approach which describes the cellular phenotype components as part of a multi-dimensional system of interactions, representing the internal complexity of the cell in an abstract manner. The model reproduces both universal and individual properties of single cell traces. Our results point to limits of current experimental methods in determining mechanisms of cellular control, and offer several directions for future research.

Results

Balance between noise and effective potential shapes fluctuations in phenotype components

Several recent studies have focused on cell-size dynamics in single cells, revealing exponential-like accumulation throughout the cell cycle with abrupt drops at division [40, 39, 46, 45, 7].

Here we use an experimental method which traps a cell at the bottom of a microfluidic channel and follow its properties over multiple generations [40, 41]. A portion of a measured trace, displaying cell length over growth and division cycles, is shown in Fig. 1A (blue lines, top and bottom panels, different growth media). Since cells maintain approximately constant width in fixed conditions, we use cell length as a proxy for cell size. Highly expressed proteins in E-coli show a qualitatively similar single-cell dynamics, as demonstrated in our recent work [41]. A sample measurement of GFP expressed from the $\lambda$ PR promoter (red line, top panel) and from the LAC promoter (green line, bottom panel) are shown in the same figure. This similarity is not surprising
for the LAC promoter, since it is metabolically important in the medium used in these experiments (see Methods for details). Nevertheless, protein expression from the λ PR promoter, an alien promoter derived from a bacteriophage and which does not play any metabolic role in the cell, still reveals similar dynamics highly correlated with cell length. These results suggest that highly expressed proteins - similarly to cell size - are global cellular variables, whose dynamics reflect the integrations of multiple microscopic processes. We therefore treat them on equal footing as “phenotype components”, and adopt a phenomenological, coarse-grained approach to describe their dynamics [34, 41, 49].

While there are several important events in the cell cycle, such as DNA replication and transitions between cycle stages, the most dramatic one is cell division. In contrast to a smooth accumulation of proteins, biomass and other molecular content throughout the cycle, at division cellular content is abruptly divided approximately in half. Thus, we include cell division as an explicit event connecting consecutive cell cycles. Following previous coarse-grained models of cell growth over multiple generations [5, 46, 48, 7, 49, 8], we use a simple parametrization given by:

\[ x_n(t) = x_n(0) e^{\alpha_n t}, \quad 0 \leq t \leq T_n \]

\[ x_{n+1}(0) = f_n x_n(T_n) \]

where \( x_n(t) \) denotes any one of the phenotype components during the \( n \)-th cell-cycle, of duration \( T_n \); \( \alpha_n \) is the exponential accumulation rate; and \( f_n \) the fraction at division, connecting the end of the current cell-cycle to the initial condition of the next cycle. The measured trace can be described to high accuracy by fitting the three parameters: \( \alpha_n, f_n \) and \( T_n \), for each cell-cycle (black lines in Fig. 1A). These parameters vary over consecutive cell cycles: the exponential accumulation rates, \( \alpha_n \), are normally distributed across cell-cycles with average and variance depending on medium conditions and bacterial strain [39, 41, 8, 50, 51]. The division fraction \( f_n \) are also normally distributed around 1/2, with a relatively narrow distribution (typical standard deviation \( \sim 0.15 \) for *E. coli* [41]). The distribution of cell-cycle times has a broad tail, and has been well documented and studied in many cell types [52, 40, 46, 39, 13]. These variations in cell-cycle parameters are sufficient to account for the statistical properties of cell length and protein, suggesting that additional fluctuations over time (i.e. rapid fluctuations around the exponential approximation within a cell-cycle) can be neglected [41].

Within this approximation, one may combine the two equations above to relate the values of any phenotype component at the start of cell-cycle, \( x_n = x_n(0) \), across generations:

\[ x_{n+1} = f_n x_n e^{\alpha_n T_n}. \]
This defines a discrete-time stochastic process - a mapping from one cell-cycle to the next, where three random variables characterize each cycle. The exponential rates $\alpha_n$ show a strong negative correlation with the cell-cycle time $T_n$; the total accumulation exponent $\phi_n = \alpha_n T_n$ is approximately independent across generations, but is negatively correlated with the phenotype component at the beginning of the cell-cycle $x_n$ [49]. Figs. 1B,C show this correlation for cell length and protein copy-number respectively, which is well described as:

$$\phi_n = \tilde{\phi} - \beta \ln x_n + \xi_n$$

where $\beta$ and $\tilde{\phi}$ are the empirical slope and intercept of the best linear fit to each of the correlations ($\beta_L$ for cell length, $\beta_P$ for protein etc.) and $\xi_n$ accounts for fluctuations around this fit (Fig. 1B,C insets).

Taking the logarithm of the mapping Eq. (2) and using the correlation Eq. (3), one finds the discrete equation of motion:

$$\ln x_{n+1} - \ln x_n = -\beta \ln x_n + \ln f_n + \tilde{\phi} + \xi_n = -\beta \ln x_n + \eta.$$  

(4)

This shows that the dynamics of the exponential accumulation of phenotype components ($\ln x_n$) follows a discrete-time Langevin equation, which describes a particle under the influence of a deterministic linear force with noise. Here $\beta$ plays the role of an effective restraining force. The other terms (marked together as $\eta$) introduce noise, but may also have a nonzero average, as discussed below. In biological terms, the accumulating exponentials jitter due to various fluctuations in the system. On the other hand they are restrained by cellular mechanisms that regulate cell division such that the phenotype components do not diverge. Recent theoretical work has studied a generalized nonlinear form of this map [45]. The correlations in Figs. 1B,C, however, show that a linear force provides an excellent approximation to the data. Thus one may imagine the particle fluctuating near the bottom of a potential well, from which the restraining force is derived.

For cell length, Fig. 1B shows that in both experimental conditions $\beta \approx 0.5$. This corresponds to the "adder" model of cell division control, where a constant biomass is added per cell-cycle [5, 42, 7]. The equivalence of the above mapping for $\beta = 0.5$ and the adder model was discussed in our previous work [49]. In contrast, for proteins, different values of $\beta$ are found; in the data shown here, $\beta_P = 0.12$ for fluorescent protein expressed from the LAC promoter and $\beta_P = 0.26$ for protein expressed from the $\lambda$-PR promoter. The significance of the effective restraining force for proteins is discussed at length below.

A particle with the above dynamics, over a large number of time steps exhibits a Gaussian
Figure 1: Universal statistical properties of the different phenotype components. (A) example traces of cell-length (blue) and total fluorescence (red and green) measured in different media (LB top panel, and M9CL lower panel). The fluorescent protein in LB medium was expressed from the $\lambda$-phage PR promoter, while in M9CL it was expressed from the LAC promoter. (B,C) correlations between the accumulated exponents ($\phi_n = \ln \left( \frac{x_n(T_n)}{x_n(0)} \right)$, with all symbols defined in the main text) and value at the start of the cell-cycle ($\ln x_n(0)$) for each phenotype component presented in (A). The data was collected from different traces each over 50 generations long (59 traces for measurements in LB and 17 traces in M9CL). Data points are color-coded by density of the points. Black lines: best linear fits, $\phi_n = \bar{\phi} - \beta \ln x_n$. Insets: noise distribution around the linear fit. (D) Population distributions in scaled variables $(x - \langle x \rangle) / \sigma_x$, highlighting the distribution shapes of cell length (blue crosses), $\lambda$ PR expression (red circles), and LAC promoter expression (green circles). Solid lines: prediction of mapping model with shape factor estimated from the entire collection of traces.
Figure 2: Illustration of discrete-time Langevin dynamics and the parameters affecting the distribution shape. (A) Discrete-time Langevin dynamics of $y = \ln x$ in a potential well corresponding to a linear restraining force ($F(y) = -\beta y$ with $\beta = 0.5$), and noise of strength $\sigma_\eta$. For simplicity of illustration we take $\langle \eta \rangle = 0$. The motion forms stochastic trajectories along time (A, right) that fluctuate with larger amplitude for larger noise (orange vs. blue trajectory). These fluctuations form stationary distributions (corresponding colors in the potential well), with variance $S^2 = \sigma_\eta^2/(2\beta - \beta^2)$. (B) In the original coordinates, the phenotype component forms a log-normal distribution with shape factor $S$. Plotting it in scaled variables highlights this shape and allows comparison of distributions with different scales. In general, different Gaussian variances correspond to differently shaped log-normal distributions.

distribution of its coordinate with a variance $S^2 = \sigma_\eta^2/(2\beta - \beta^2)$. Transforming this back to the original phenotype component $x_n$, one finds a log-normal distribution [5, 49]:

$$P(x) = \frac{1}{xS\sqrt{2\pi}} \exp \left[ -\frac{(\ln x - \mathcal{M})^2}{2S^2} \right].$$

with

$$\mathcal{M} = \frac{\langle \eta \rangle}{\beta}, \quad S^2 = \frac{\sigma_\eta^2}{2\beta - \beta^2},$$

(5)

corresponding to the mean and variance of $\ln x$. Here $\langle \eta \rangle$ and $\sigma_\eta$ are the mean and standard deviation of the variable $\eta = \ln f_n + \bar{\phi} + \xi_n$, respectively. Fig. 2 illustrates Langevin dynamics in a potential-well, comparing low (blue) and higher (orange) levels of noise. For the higher noise level, the trajectory shows larger fluctuations (Fig. 2A, right, orange trajectory) and the Gaussian distribution is wider (Fig. 2A, left, orange trajectory).

The shape of a log-normal distribution is determined solely by the factor $S$. Stated differently, a family of log-normal distributions sharing the same $S$ collapse on one another under two parameter scaling: $(x - \langle x \rangle)/\sigma_x$ (see SI-1). We therefore term this parameter the "shape factor" in what follows. Fig. 2B shows how the different variances of the Gaussian distributions in the potential well transform to a different shape factors for the phenotype components $x_n$, which is revealed when plotting their log-normal distributions in appropriately scaled axes.

The distribution shape of any phenotype component fluctuations can now be predicted from its dynamical trace parameters, namely the noise variance $\sigma_\eta^2$ and the effective restraining force $\beta$. Fig. 1D shows that this prediction fits well with the empirical distribution for all measured phenotype
components - cell length and both highly expressed proteins. The estimated shape parameters are in the range $0.25 - 0.5$, resulting in similarly shaped log-normal distributions. The broader spectrum of shapes potentially attained by the log-normal distribution is only revealed when $S$ spans a much larger range, see SI-1. Further discussion of this robustness of distribution shape is presented in the next section, in the context of individual traces.

**Statistical properties of individual traces along time - individuality and universality**

Long enough traces of phenotype components can reveal the statistical properties of single cells along time. It is of interest to compare these properties to those of the population, or collection of traces. Such a comparison can shed light on the internal structure of the population which gives rise to the statistical properties as a whole.

Individual traces have both distinct and universal statistical properties. Fig. 3A depicts analysis of the cell length averaged over increasing time-windows one trace, where the last time-point is the average over the entire trace. It is clearly seen that each cell converges over its lifetime to a distinct average cell length. Similar results were obtained for highly expressed proteins, as well as in a different growth medium (SI-3). Temporal averages over the entire trace span a range of up to five-fold for protein content and up to two-fold for cell length. Analysis of single-cell traces measured in another lab [40] also show individuality in temporal averages (see SI-3).

In contrast, the distribution shape of scaled fluctuations is common to all measured traces, which collapse on the same curve under two-parameter scaling. Our previous work has demonstrated this universality for protein distributions [41]; Fig. 3B shows the same scaling holds also for cell length. The population-averaged curve is depicted by a black line. This scaling results in a more accurate collapse of the data than one-parameter scaling (see SI-2).

We have seen that the effective mapping model presented above predicts the distribution shape. Additionally it predicts the absolute moments and in particular the average, as a function of dynamic trace parameters (Eq. 5). We use this result to shed light on both the individual and the universal properties of the traces. To this end we extract the relevant dynamical parameters - noise variance and effective restraining force - for each trace separately. Fig. 3C shows the best linear fit to the correlation between $\phi_n$ and $\ln x_n$ (Eq. 3), for a collection of individual cell-length traces. Each line was obtained as a fit to a figure similar to 1B,C where data points were taken from a single individual trace along time. It can be clearly seen that each trace is characterized by a different slope $\beta$ and intercept $\bar{\phi}$. However, these are not independent parameters; Fig. 3D reveals a tight linear
Figure 3: Individual (left panels) vs. universal (right) statistical properties of cell-length. Data collected from cells grown in LB medium at 32C. (A) Cumulative average cell-length of example traces, computed over an increasing time window. The last point is the average over the entire trace, spanning a range 2.5-3 µm. (B) Collapse of scaled fluctuations from individual traces, showing the distribution shape is universal, and coincides with the scaled population distribution (solid black line). (C) Linear fits to effective-force plots (similar to Fig. 1B), only here for data collected from different traces separately. Clearly individual traces have different dynamic parameters: slope ($\beta$) and intercept ($\phi$). Different dynamic parameters are not independent, as seen by all of them crossing at a pivot point (blue dot). (D) They obey $\phi = \phi_0 + \beta \ln x_0$, where $\phi_0 = 0.78, \ln x_0 = 0.83$ are the equilibrium population-level values. (E) Individually distinct temporal averages estimated from traces, as a function of the value predicted by the model with independently estimated dynamic trace parameters. (F) Measured data points (blue: cell length; red: protein expressed from the λ PR promoter) are presented on a contour map of the shape factor $S$, showing that they are located mostly along a region of constant $S$. Inset: data points for cell-length (blue dots) and theoretical shape factor (line) projected on the $\beta$ axis, reveal a knee-like structure with data points located in the flat part of the graph. (additional data are presented in SI-3).
relationship between the two:

\[ \bar{\phi} = \phi_0 + \beta \ln x_0 \]  

(6)

The discrete-time equation of motion now takes the form:

\[ \ln x_{n+1} - \ln x_n = -\beta (\ln x_n - \ln x_0) + \phi_0 + \ln f_n + \xi_n. \]  

(7)

The identification of \(\phi_0\) and \(\ln x_0\) as population-level parameters, whereas \(\beta\) varies among traces, suggests a clear interpretation of the dynamics for individual traces. \(\ln x_0\) is the bottom of the effective potential well, the equilibrium value to which each phenotype component is attracted by cell-cycle control. For cell length it corresponds to \(L_0 \approx 2.3\mu m\), the average cell length over the population of traces (blue circle, intersection of all lines in Fig. 3C. A similar value is found also in the parallel experiment in a different medium (SI-3). For proteins, this equilibrium value depends on the particular protein (SI-3). In contrast, \(\phi_0\) is similar for all phenotype components and is found empirically \(\approx \ln 2\). This corresponds to the (logarithmic) average fold-increase of each phenotype component per cell-cycle. In principle this should cancel on average with \(\ln f\) for symmetric division.

Ideally, then, all cells would have the same average length over time, equal to the population average \(x_0 = L_0\). However, if fluctuations in division ratio and in accumulation factor do not compensate one another over the lifetime of the cell, this deviation will contribute to the temporal average of cell length:

\[ \langle \ln x \rangle = \ln x_0 + \frac{\langle \ln f \rangle + \phi_0 + \langle \xi \rangle}{\beta}. \]  

(8)

Computing these fluctuations directly from the traces, Fig. 3E confirms that this prediction captures the source of their distinct average cell length. Indeed, a direct calculation from the data shows that over any given trace, the average accumulation is not exactly 2 while the average division is not exactly 1/2. Similar arguments hold also for protein content, with an equilibrium value depending on protein type (SI-3).

Notably, the strength of the effective potential \(\beta\) emerges as a distinct property of individual cells over time that can vary over a significant range. In the data presented in Fig. 3D, the range of data-points on the \(x\)-axis is 0.4-1. Nevertheless, we have seen that the scaled distributions collapse to an excellent approximation. This is somewhat surprising since the shape factor also depends on \(\beta\) (Eq. 5). To understand how such robustness of the distribution shape is achieved, we plot the data points in the \((\beta, \sigma)\) plane for many traces (Fig. 3F, overlaid on a two-dimensional contour map of the shape factor predicted by Eq. 5). It is seen that the data points are scattered in a region
where contours are almost parallel to the $\beta$-axis, implying that significant changes in $\beta$ can be made without strongly affecting $S$. This can also be seen by plotting $S$ directly as a function of $\beta$, revealing a shallow dependence in the relevant region (Fig. 3F inset). Moreover, although there is still a range of shape factors in the data, the shape of the log-normal distribution does not vary considerably in this range, as already discussed above for the population analysis (see SI-1).

These results show that all traces can be described by a log-normal distribution with approximately the same shape. It follows from this property that the variance is a quadratic function of the mean among all traces (see SI-1). This property has been established for highly expressed proteins [43, 24, 34], but is more difficult to verify for cell length because of the relatively narrow range it spans.

Another statistical property characterizing the traces separately along time is the autocorrelation function (ACF). To disentangle the long-term dynamics from the short accumulation and division cycles, we consider the phenotype components at the beginning of consecutive cell cycles, as in the mapping above. Their ACFs reveal a temporal structure spanning many generations, seen in Fig. 4 for cell-length in LB medium (additional data for protein content and in another growth medium are presented in SI-4). For some of the traces, damped oscillations can be seen. The typical periods and amplitudes of these oscillations are distinct for each individual trace [6]. As a consequence of this incoherency, these structures are washed out by averaging traces (black line in Fig. 4A).

How can these oscillations be explained in terms of the mapping model? When averaging over the ensemble from which noise is drawn, the ACF of mappings such as Eq. 7 decreases exponentially with a time-constant of $\approx \ln(1-\beta)$. This is in line with the smooth form that appears after averaging over all traces. In a single realization, one cannot directly calculate the ACF. However, a recent theory estimates the probability of oscillatory patterns being generated at random across
The period of oscillations was approximated as the mean peak-to-peak distance, which was computed as a function of the restraining force $\beta$. These results have established that the discreteness of the Langevin mapping is sufficient to produce oscillations over time in individual trajectories, which disappear after averaging over the ensemble.

Fig. 4B shows this theoretical prediction (black solid line) together with the corresponding quantities computed from our experimental traces for cell length (blue Xs; binned data are shown as large blue circles) and protein content (similar to length, in red). Although the individual traces show a large scatter, binning them by $\beta$ gives a good agreement with the theory. This is expected since the theory is probabilistic and predicts an average over realizations for any given $\beta$. The theory becomes a better predictor of model simulations for longer traces (see SI-5). We conclude that the data are consistent with a simple interpretation: the long-term properties of the ACF arise from purely stochastic effects, in combination with the inherent discreteness of cell-division. Taken together, the results of this section point to the effective force strength $\beta$ as determining both individual statistical properties of the traces: the temporal average and the structure of the ACF.

**Relationship between phenotype components**

Close inspection of the traces in Fig. 1A reveals that the exponential rates characterizing the accumulation of different proteins are correlated with cell-length across cell-cycles. To generalize these results further towards multiple phenotype components, we constructed a bacterial strain that expresses two different color fluorescent proteins from two distinct promoters (LAC promoter, and $\lambda$-phage PR promoter), and measured both proteins and cell length simultaneously in each cell (see Methods). Our results reveal a strong correlation between all three exponential rates measured in the same cell-cycle, as can be seen in Fig. 5A depicting a 3-dimensional view of the exponential accumulation rates $\alpha_i$. Examples of three simultaneously measured traces can be seen in SI-5. One possibility to explain these strong correlations is that the copy-number of highly expressed proteins is simply proportional to cell size. In this case, protein dynamics are enslaved to those of cell size; such dynamics were recently proposed for membrane proteins making up the cell surface [54]. In terms of statistical properties, this would imply a constant (or narrowly distributed) protein density. Fig. 5B shows that, rather than being a narrow distribution, the density spans a broad range, approximately 5-fold in concentration. Moreover, plotting directly the instantaneous total fluorescence intensity vs. cell length shows that, although they are correlated ($C \approx 0.5$), there is a large spread both within single-cell traces and among individual cells (Fig. 5C). These results suggest that the relationship between protein and cell size is not a simple proportion; they seem proportional within a cell-cycle, but their ratio varies among cell-cycles and among individual cells. This nontrivial relation raises the question about the origin of exponential growth observed in all
Figure 5: Relationship between the different phenotype components. (A) 3-way correlation between exponential growth rates $\alpha_n$ for cell-length, LAC promoter expression, and $\lambda$ PR expression all measured simultaneously in the same cells. Each point represents one cell cycle. (B) Distribution of GFP density (fluorescence per unit cell area), representing the density of the LAC proteins, computed from 3 different traces (colors). Note the broad distribution spanning a 5-fold range. (C) Scatter-plot of instantaneous GFP (representing the total amount of LAC proteins) vs. cell-length. Linear relations hold within cell cycles, but their slope changes from cycle to cycle and centers on different averages for individual cells (colors).
measured components, particularly proteins that do not auto-regulate their own expression.

Another question that arises when considering all phenotype components as a whole, concerns the correlations described previously as effective force plots, such as 1B,C. These are usually thought to reflect a mechanism of cell-division control; why are such correlations found also for proteins? Do they imply a role for proteins in controlling division, or is it merely an indirect induced correlation?

To test for possible relationships between phenotype components that are consistent with the multiple-component aspects of the data, we explored several models with coupling between them. First we consider the simplest model where one component (e.g. cell size) controls the division time by Eq. 3, while other components independently accumulate over the cell cycle (see Methods). Such a model is found to be unstable: a realistic level of division noise causes instability in those components that do not directly control division time. This is true even for the extreme case where accumulation exponential rates are identical in all components (results are presented in SI-6).

How, then, do multiple components continue to accumulate exponentially and divide over many generations without them being involved in controlling division time directly? To answer this question we go beyond the mapping model, which relates consecutive generations, to include also dynamics of phenotype components within the cell cycle. Our guiding hypothesis is that the cell constitutes a complex system composed of many interacting variables. When we measure any specific phenotype component, such as protein content or cell size, we are probing the outcome of this complex system; when we measure several components simultaneously, an effective interaction between them will be observed. This should not be interpreted as a simplistic causal relation, but rather as an effective description of their participation in the large dynamical system that makes up the cell, and which may also contain many other hidden variables.

We consider therefore a vector $\vec{x}$ of $N$ observable phenotype components, and seek the simplest model consistent with exponential accumulation over time. Previous work has proposed that autocatalytic-like dynamics can arise indirectly from linear interactions [55, 56]. Effective dynamics within the $n$-th cell cycle are then

$$\frac{d}{dt} \vec{x}_n(t) = K_n \vec{x}_n(t), \quad 0 < t < T_n,$$

with $K$ a matrix of interactions. In what follows this matrix will be arbitrary - i.e. drawn at random, since we are interested in properties that are insensitive to the exact details of the interactions it describes. Once $K$ has been specified the dynamics are deterministic, thus rapid intra-generation fluctuations are neglected in this model, in line with the above discussion of the data. Cell division distributes fractions $f_n^{(j)}$ and $1-f_n^{(j)}$ of phenotype component $j$ to each daughter cell, similarly to...
Figure 6: Multi-component model with coupled dynamics between phenotype components. In these simulations, phenotype components interact via a random interaction matrix. A single phenotype component, \( x_1 \), determines division events through a noisy threshold on its accumulated exponent, Eq. 3 with \( \beta = 0.5 \).

(A) Time evolution of three model phenotype component traces over 40 generations, in a simulation of a 10-dimensional system. Division events are determined by the first component only. (B) Top panel: Over the short cell-cycle time, model traces can be approximated by an exponential function with a single effective exponent (black lines). Bottom panel: normalizing cell-cycle traces to initial time and to value 1 at the cell-cycle start, emphasizes the variability in the effective exponents between generations. (C) 3-way correlations between effective exponential growth rates of different components similar to those in (A), from two different simulation runs (colors).

the one-component models discussed before:

\[
x_{n+1}(0)^{(j)} = f_{n}^{(j)} \cdot x_{n}^{(j)}(T_{n}).
\]

with \( f_{n}^{(j)} \sim N\left( \frac{1}{2}, \sigma^{2} \right) \) for each \( 1 \leq j \leq N \). We first consider the case where one component controls cell division through the relation in Eq. 3. In this case, we observe that dynamical interactions inside the cell cycle stabilize the system: accumulation and division persist over many generations in all components. Examples of the resulting traces for 3 out of 10 components in a simulation are shown in Fig. 6A. This picture is insensitive to the details of the interaction matrix \( K \) as long as it has at least one positive eigenvalue.

The dynamics of phenotype components inside a cell-cycle, however, are not exactly exponential: the trajectory within any cycle \( n \) is in fact given by:

\[
\hat{x}_{n}(t) = \sum_{i=1}^{N} c^{(i)} e^{\lambda_{i} t} v_{i} \quad 0 < t < T_{n}.
\]

Here \( \{\lambda_{i}\}_{i=1}^{N} \) are the eigenvalues of \( K \), \( \{v_{i}\}_{i=1}^{N} \) are the respective eigenvectors, and \( c \) are the projec-
Figure 7: Indirect induced correlations in a multi-component model with coupled dynamics. A single phenotype component, $x_1$, determines division events through a noisy threshold on its accumulated exponent, Eq. [5] with $\beta = 0.5$. Other phenotype components indirectly acquire a correlation between their accumulated exponent and initial value at cycle start, similar to Fig. 1B,C. This appears as a restraining force, although it is caused solely by their coupling with $x_1$ and their common division time. Black lines represent the best linear fit. Note that induced restraining force $\beta$ appears to be strongest for component 3 (0.56), while the controlling component has an effective restraining force of 0.48.

The resulting in-cycle trajectories are a linear combination of exponentials (Eq. [10]), where at least some of them are positive. Such a combination can be described, to a good approximation over a finite time, by a single exponential function with an effective exponential accumulation rate. Fig. 6B (upper panel) illustrates a model trajectory of one phenotype component, fitted to an effective exponential growth. Over long time-scales, a linear combination of exponential functions would be dominated by the leading exponent [56]. However, biological constraints limit the cell-cycle to relatively short times, over which the components increase by only a factor of $\sim 2$. Due to this limited time the effective exponent depends on all eigenvalues as well as on the prefactor of each exponent. As a result of this dependence, fluctuations emerge in the effective exponents from one generation to the next. These fluctuations are caused by the distribution of random fractions at the beginning of each cycle, which in turn reshuffles the prefactors of exponents in Eq. [10].

Examples of the effective exponentials for one phenotype component along consecutive gen-
erations, all normalized to 1 at the cell-cycle start, are presented in Fig. 6B (lower panel). Model simulation results are plotted in blue dots whereas exponential fits are shown with black lines. Our previous work has shown that variability in the exponential accumulation rates among generations is significant (CV ≈ 0.5), and crucial for obtaining the broad universal distribution of protein fluctuations from individual traces [41]. This variability arises here naturally as a result of the effective interactions and division noise, without the need to explicitly introduce a large stochastic element into the rate. Moreover, even though the exponential accumulation rate of each phenotype component exhibits wide variability across cell-cycles, a strong correlation is still observed between the exponential accumulation rates of all components on a cycle-by-cycle basis due to their common division time (Fig. 6C).

In addition to stabilizing the multi-component phenotype dynamics, the interactions within cell-cycles indirectly induce correlations that appear as effective restraining force plots (see Fig. 7). The points in the scatter-plots are collected from many traces produced by the model with the same set of parameters, one point for each cell-cycle, analogous to the experimental data presented in Fig. 1B,C. As expected, phenotype component 1 exhibits a strong correlation with cell-cycle time; it actually carries out the control of cell division. Perhaps more surprisingly, effective correlations emerge between accumulated exponents and initial condition for all other phenotype components. The slope of the induced correlation need not be identical for all components; if interactions are strong enough (large off-diagonal matrix elements of K), these induced correlations can be as strong or even stronger than those of the controlling component. Qualitatively similar results to Fig. 7 are obtained if two phenotype component control division, with generally two different restraining forces strengths $\beta_1$ and $\beta_2$ (see SI-7 for details of the two-component control mechanism and results). However, if the number of controllers approaches the total number of phenotype components, they behave statistically very different from one another, various distribution shapes appear, and generally the experimentally observed statistical properties are not reproduced.

The multi-component model with coupled dynamics, and with one or two components controlling division, reproduces practically all statistical properties described above in individual traces: universal distribution shape, individual temporal average and auto-correlations. A graphical display of all these properties is presented in Fig. 8 and SI-8.
Figure 8: Model results: Individual and universal statistical properties. The graphs in this figure are exactly as described in Fig. 3, only for simulations results. The statistical properties of $x_1$ are presented here, whereas those of another components can be seen in SI-8.
Discussion

The biological cell is a complex system with a myriad of interacting processes, whose integrated outcome influences its phenotypic properties: cell size, cell-cycle time, protein content and more. These are global variables, buffered from the microscopic realm as a result of integration over many processes. Therefore, such variables are best described by a mesoscopic, effective description. Indeed such descriptions have advanced tremendously in recent years. In particular the dynamics of cell-size growth and division has received much attention thanks to recent single-cell techniques that allow long-term measurement of these dynamics. However, being part of the same complex system, cell size and other global variables are not independent of one another. They constitute "components" of the phenotype of one and the same cell. In the current work our aim was to unify the mesoscopic description of two types of global variables, which have been studied separately until now: cell size and copy number of highly expressed proteins.

In previous studies we have focused on statistical properties of highly expressed protein copy number, with emphasis on their universal aspects. Our results showed that protein fluctuations along the lifetime of bacterial cells exhibit the same distribution shape as that of the entire population. In addition we have found a quadratic relation between the variance and mean, both among experimental conditions and among distinct individual traces [34, 41]. These results were analyzed using a discrete-time model that maps protein content from one cell-cycle to the next [49]; this mapping model is equivalent to a particle in a potential well with noise [5, 49, 44]. Here, these results were extended also to cell size. Our analysis in the current work reveals that cell-size (measured in our experiments as the length of bacteria trapped in a narrow channel), exhibits the same universal statistical properties reported previously for protein. Specifically, the distributions of cell-length in all traces are well described by a log-normal distribution with similar shape factor, resulting in their collapse onto the same curve under two-parameter scaling. In principle this would imply a quadratic dependence of variance on mean, however cell size spans a smaller dynamic range of averages relative to protein content, which renders the confirmation of such scaling difficult.

In addition to the universal properties, our data also reveals non-universal statistics of all phenotype components that remain distinct among individual bacteria over their lifetime. The long-time average of both cell size and copy-number of different proteins, as well as the ACF, were found to be distinct properties of single traces. In terms of the mapping model, the two dynamic parameters - noise strength and effective potential strength - vary significantly among cells. In addition, the average fold-growth and the average division ratio per cycle were also found to be different from cell to cell. Analysis of a large number of individual traces is consistent with a picture where population-level parameters define the equilibrium point for the effective potential well, but the steepness of the well is an individual property of each cell. As a result of this dis-
tinction, deviations from twofold increase and from half-fold division over the finite lifetime of the cell, accumulate differently in each individual cell. This provides a prediction of the individuality in temporal average of the different phenotype components.

The distinct nature of the ACF was also related to the different effective potentials in individual cells. The observed oscillations in ACFs are explained as a property of discrete-time stochastic mappings [53]. The oscillation period and the probability of observing oscillations, both depend the effective force strength. We thus observe in the data a direct dynamic consequence of the discreteness of cell cycles, which persists over multiple generations.

Beyond the similarity in statistical properties among phenotype components, they also exhibit strong correlations in individual cells on a cycle-by-cycle basis. On the other hand, the relation between them is inconsistent with a simple proportion, or a dominance of one phenotype (e.g. cell size) that determines all others. Therefore we examined multi-dimensional models of different coupling between the components, that are consistent with the observed dynamics and statistical properties. The minimal such model was found to require the dynamics of phenotype components to be coupled during the cell-cycle. We incorporated random linear interactions producing effective auto-catalytic dynamics during the cell-cycle [55, 56] with an effective restraining force controlling cell division, as in the mapping model. The finite duration of cell-cycle, the small dynamic range (≈ ×2) of exponential accumulation, and the imperfect nature of division, cause reshuffling of the different phenotype components. As a consequence, rather than a pure exponential accumulation which reflects only one timescale - the leading eigenvalue of the interaction matrix - several eigenvalues continue to contribute. This results in effective (rather than exact) exponents which vary over cycles and between components, still maintaining a positive correlation among them.

The proposed model highlights the notion of both protein content and cell size as global cellular variables. Such global variables result from the integrated action of many processes and therefore are statistically correlated, to various degrees, with one another. Most proteins in bacteria are highly expressed, with a relatively small effect of number fluctuations [24] and moreover their degradation is negligible over the timescale of a cell-cycle [57, 58]. These properties result in their smooth accumulation within the cell cycle, which buffers many microscopic noise sources. Their global nature is also reflected in the correlation of their copy-number with cell-cycle time, and in the robustness of their distribution shape, as discussed above. The model moreover questions our ability to deduce feedback mechanisms from statistical correlations. In particular, it shows that measurement of such correlations cannot distinguish between variables that control division and others that are indirectly correlated with it.

The identification of an effective parameter - the effective potential, or restraining force, strength - that links to the distinct features of individual traces immediately raises the question: what de-
terminates that parameter in individual cells and why are they different? Our model shows that it is difficult to coordinate multiple components to control division and maintain stability. One possibility is that feedback control is an emergent property of the cell which arises dynamically from complex interactions. Such dynamic feedback has been suggested as an organizing principle for mesoscopic-scale systems [59]. It is possible that a composite variable, not necessarily immediately measurable, influences division more than any one of the molecular constituents [60]. These are questions of broad implications which are beyond the scope of our current work. Our ability to answer them is limited by the number of long traces presently available and the time of observation. Moreover this kind of data are in principle limited by the cell lifetime. How the initial conditions come into play is also unknown since these are determined by previous generations. New experimental systems may be designed to address these challenges.

Methods

Experimental procedure and data processing. Wild type MG1655 E. coli bacteria were used in all experiments described. Protein content was measured through the fluorescence intensity of green fluorescent protein (GFP) or red fluorescent protein (tdTomato) inserted into the bacteria on a high or medium copy number plasmid and expressed under the control of the promoter of interest. For measuring the expression level of a metabolically relevant protein, GFP was expressed from the medium copy number plasmid pZA [61] under the control of the LAC promoter. For metabolically irrelevant protein, GFP was expressed from the same plasmid pZA but under the control of the viral λ-phage PR promoter. For simultaneous measurement of the expression of two proteins, GFP was expressed from the high copy number plasmid pUC19 under the control of the LAC promoter, while tdTomato was expressed from the pZA plasmid under the control of the λ-phage PR promoter. Cultures were grown overnight at 30°C, in M9 minimal medium supplemented with 1g/l casamino acids and 4g/l lactose (M9CL, for measuring the expression level from the LAC Promoter or both promoters together), or LB medium measuring the expression level from the λ-PR promoter only. The following day, cells were diluted in the same medium and regrown to early exponential phase, optical Density (OD) between 0.1 and 0.2. When cells reached the desired OD, they were concentrated 10X into fresh medium, and loaded into a microfluidic trapping device (Fig. 9). After cells were trapped in the device, fresh medium was flown continuously through it to supply nutrients. The cells were allowed to grow in this device for 10s of generations, while maintaining the temperature fixed, using a made-in-house incubator. Images of the channels were acquired every 3 to 6 minutes in phase contrast and fluorescence modes using a Zeiss Axio Observer microscope with a 100x objective. The size and protein content of the mother cell were measured from these images using the image analysis software microbeTracker [62]. These data were then used to gen-
Figure 9: Experimental Setup. Bacteria (depicted in green) were trapped in an array of long micro-channels (1 \( \mu \text{m} \) width 1 \( \mu \text{m} \) height), microfabricated in PDMS. The micro-channels were closed at one end and open at the other to large perpendicular channels (30 \( \mu \text{m} \) width 30 \( \mu \text{m} \) height), through which medium could be pumped in order to feed the trapped bacteria and to allow growth for many generations along the micro-channels.

erate traces such as those presented in Fig. 1A. Growth stability of cells in the microfluidic device was verified by comparing the average division time in the first and second halves of the trace. No trend was detected in all experiments (see SI-9).

Experimental data analysis Single-cell traces were analyzed using home-made MATLAB programs. Trace autocorrelation functions, Pearson correlation coefficients, trace distributions, linear curve fitting and Kolmogorov-Smirnov test, were all calculated by their implementations in MATLAB toolboxes. All trace ACFs calculations followed a 3-point-average smoothing of the measured basepoint values. Effective force parameters for length (\( \beta_L \)) and for protein (\( \beta_P \)) were estimated as the slope of the best linear fit to scatter plots such as Fig. 1B,C, namely, accumulated exponents as a function of log phenotype component at the beginning of each cycle. Using this fit, \( \xi \) was estimated as the difference between the data and the fit. The total mapping noise level \( \sigma_{\xi_{\text{fit}}} \) was estimated by computing the standard deviation of \( \xi_n + \ln f_n \), where \( f_n \) is the division ratio at generation \( n \). Data measured at the Jun lab were extracted from the webpage accompanying [40], and analyzed in the same way as our data.

Model simulations All models explored were implemented as MATLAB programs. Simulations were carried out for model systems of various dimensions \( N \); no qualitative difference was found among high-dimensional simulations; results shown are for \( N = 10 \). To simulate the case of no interactions between components, \( K = 1.5I \), where \( I \) is the identity matrix on \( \mathbb{R}^N \). For the case of exponential rates that fluctuate across cycles, at each cycle an independent, zero-mean 0.2-variance Gaussian random variable was added to each diagonal element of \( K \). For dynamic interactions between components, each element in \( K \) was drawn at the beginning of a simulation from a Gaussian distribution centered at 0, such that the variance of off-diagonal elements was 20-fold smaller than that of diagonal elements. Except for the case of varying exponential without interactions,
the matrix was held fixed throughout all model cycles in a trace, as well as across different traces presented in each figure.

In each generation the linear dynamics were integrated over the interval \([0, T_n]\), where \(T_n\) is defined implicitly by the division control mechanism. For the case of one-component control, this was implemented by iterating the propagation loop as long as \(x^{(1)}(t) < x^{(1)}(0) e^{\phi_n}\), for a given cycle \(n\). Here \(\phi_n\) is defined by Eq. 3. For the case of two-component control, this was implemented by iterating the propagation loop as long as the following two conditions hold:

\[
x^{(1)}(t) < x^{(1)}(0) e^{\phi^{(1)}_n}; \quad x^{(2)}(t) < x^{(2)}(0) e^{\phi^{(2)}_n}.
\]

Here \(\phi^{(1)}_n\) is defined as before, and

\[
\phi^{(2)}_n = \phi^{(2)}_0 - \beta_2 (\ln x^{(2)}_n - \ln x^{(2)}_0) + \xi^{(2)}_n.
\]

At each division event, a ratio \(f_n\) was drawn independently (among generations and among components within a generation) from a Gaussian distribution \(\mathcal{N}(0.5, 0.15)\). The noise in division control was \(\xi \sim \mathcal{N}(0.5, 0.15)\). Remaining parameters: \(\beta = 0.5\), \(\phi_0 = 1\) and \(x_0 = 3\) in all simulations, and identical across components (in the case of two-parameter control).

Acknowledgements We thank Omri Barak, Erez Braun, Yoni Savir and Dan Stein for helpful comments. This work was supported by the US-Israel Binational Science Foundation (NB and HS), The Israel Science Foundation (grant No. 1566/11; NB), and the National Science Foundation (grant PHY-1401576; HS). HS also acknowledges his support by a Fulbright Scholar Award.

References

[1] S. V. Avery. Microbial cell individuality and the underlying sources of heterogeneity. Nature Reviews, 4:577–587, 2006.

[2] Steven J Altschuler and Lani F Wu. Cellular heterogeneity: do differences make a difference? Cell, 141(4):559–563, 2010.

[3] Martin Ackermann. A functional perspective on phenotypic heterogeneity in microorganisms. Nature Reviews Microbiology, 13(8):497–508, 2015.

[4] Stefano Di Talia, Jan M Skotheim, James M Bean, Eric D Siggia, and Frederick R Cross. The effects of molecular noise and size control on variability in the budding yeast cell cycle. Nature, 448(7156):947–951, 2007.

[5] A. Amir. Size regulation in bacteria. Phys. Rev. Lett., 112:208102, 2014.
[6] Y. Tanouchi, A. Pai, H. Park, S. Huang, R. Stamatov, N. E. Buchler, and L. You. A noisy linear map underlies oscillations in cell size and gene expression in bacteria. *Nature*, 523:357-360, 2015.

[7] S. Taheri-Araghi, S. Bradde, J. T. Sauls, N. S. Hill, P. A. Levin, J. Paulsson, M. Vergassola, and S. Jun. Cell-size control and homeostasis in bacteria. *Curr. Biol.*, 25(3):385–391, 2015.

[8] Andrew S Kennard, Matteo Osella, Avelino Javer, Jacopo Grilli, Philippe Nghe, Sander J Tans, Pietro Cicuta, and Marco Cosentino Lagomarsino. Individuality and universality in the growth-division laws of single e. coli cells. *Physical Review E*, 93(1):012408, 2016.

[9] Arjun Raj and Alexander van Oudenaarden. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell*, 135(2):216–226, 2008.

[10] Alvaro Sanchez and Ido Golding. Genetic determinants and cellular constraints in noisy gene expression. *Science*, 342(6163):1188–1193, 2013.

[11] Iain G Johnston, Bernadett Gaal, Ricardo Pires das Neves, Tariq Enver, Francisco J Iborra, and Nick S Jones. Mitochondrial variability as a source of extrinsic cellular noise. *PLoS Comput Biol*, 8(3):e1002416, 2012.

[12] O. Sandler, S. Pearl Mizrahi, N. Weiss, O. Agam, I. Simon, and N. Q. Balaban. Lineage correlations of single cell division time as a probe of cell-cycle dynamics. *Nature*, 519:468-471, 2015.

[13] Rami Pugatch. Greedy scheduling of cellular self-replication leads to optimal doubling times with a log-frechet distribution. *Proceedings of the National Academy of Sciences*, 112(8):2611–2616, 2015.

[14] P. G. Lord and A. E. Wheals. Variability in individual cell cycles of saccharomyces cerevisiae. *J. Cell Scie.*, 50:361–376, 1981.

[15] J. L. Spudich and D. E. Koshland. Non genetic individuality: chance and the single cell. *Nature*, 262:467–471, 1976.

[16] Vahid Shahrezaei and Samuel Marguerat. Connecting growth with gene expression: of noise and numbers. *Current opinion in microbiology*, 25:127–135, 2015.

[17] Noreen Walker, Philippe Nghe, and Sander J Tans. Generation and filtering of gene expression noise by the bacterial cell cycle. *BMC biology*, 14(1):1, 2016.

[18] Daniel J Kiviet, Philippe Nghe, Noreen Walker, Sarah Boulineau, Vanda Sunderlikova, and Sander J Tans. Stochasticity of metabolism and growth at the single-cell level. *Nature*, 514(7522):376–379, 2014.
[19] Fabai Wu and Cees Dekker. Nanofabricated structures and microfluidic devices for bacteria: from techniques to biology. *Chemical Society Reviews*, 45(2):268–280, 2016.

[20] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6:451–464, 2005.

[21] Arren Bar-Even, Johan Paulsson, Narendra Maheshri, Miri Carmi, Erin O’Shea, Yitzhak Pilpel, and Naama Barkai. Noise in protein expression scales with natural protein abundance. *Nature genetics*, 38(6):636–643, 2006.

[22] D. Volfson, J. Marciniak, W. J. Blake, N. Ostroff, L. S. Tsimring, and J. Hasty. Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439:861–864, 2006.

[23] N. Brenner, K. Farkash, and E. Braun. Dynamics of protein distributions in cell populations. *Physical Biology*, 3:172–182, 2006.

[24] Y. Taniguchi, P.J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X.S. Xie. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329:533, 2010.

[25] Mukund Thattai and Alexander Van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences*, 98(15):8614–8619, 2001.

[26] S. Krishna, B. Banerjee, T.V. Ramakrishnan, and G.V. Shivashankar. Stochastic simulations of the origins and implications of long tailed distributions in gene expression. *PNAS*, 102(13):4771–4776, 2005.

[27] Dann Huh and Johan Paulsson. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature genetics*, 43(2):95–100, 2011.

[28] T. Friedlander and N. Brenner. Cellular properties and population asymptotics in the population balance equation. 2008.

[29] N. Brenner and Y. Shokef. Nonequilibrium statistical mechanics of dividing cell populations. *Phys. Rev. Lett.*, 99:138102, 2007.

[30] Saburo Tsuru, Junya Ichinose, Akiko Kashiwagi, Bei-Wen Ying, Kunihiko Kaneko, and Tetsuya Yomo. Noisy cell growth rate leads to fluctuating protein concentration in bacteria. *Physical biology*, 6(3):036015, 2009.

[31] Lok-hang So, Anandamohan Ghosh, Chenghang Zong, Leonardo A Sepúlveda, Ronen Segev, and Ido Golding. General properties of transcriptional time series in escherichia coli. *Nature genetics*, 43(6):554–560, 2011.
[32] Daniel L Jones, Robert C Brewster, and Rob Phillips. Promoter architecture dictates cell-to-cell variability in gene expression. *Science*, 346(6216):1533–1536, 2014.

[33] Kok Hao Chen, Alistair N Boettiger, Jeffrey R Moffitt, Siyuan Wang, and Xiaowei Zhuang. Spatially resolved, highly multiplexed rna profiling in single cells. *Science*, 348(6233):aaa6090, 2015.

[34] H. Salman, N. Brenner, C.-K. Tung, N. Elyahu, E.Stolovicki, L. Moore, A. Libchaber, and E. Braun. Universal protein fluctuations in populations of microorganisms. *Phys. Rev. Lett*, 108:238105, 2012.

[35] Sandeep Krishna, Bidisha Banerjee, TV Ramakrishnan, and GV Shivashankar. Stochastic simulations of the origins and implications of long-tailed distributions in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 102(13):4771–4776, 2005.

[36] Lindsay S Moore, Elad Stolovicki, and Erez Braun. Population dynamics of metastable growth-rate phenotypes. *PloS one*, 8(12):e81671, 2013.

[37] Kazufumi Hosoda, Tomoaki Matsuura, Hiroaki Suzuki, and Tetsuya Yomo. Origin of lognormal-like distributions with a common width in a growth and division process. *Physical Review E*, 83(3):031118, 2011.

[38] Andrea Giometto, Florian Altermatt, Francesco Carrara, Amos Maritan, and Andrea Rinaldo. Scaling body size fluctuations. *Proceedings of the National Academy of Sciences*, 110(12):4646–4650, 2013.

[39] Srividya Iyer-Biswas, Charles S Wright, Jonathan T Henry, Klevin Lo, Stanislav Burov, Yihan Lin, Gavin E Crooks, Sean Crosson, Aaron R Dinner, and Norbert F Scherer. Scaling laws governing stochastic growth and division of single bacterial cells. *Proceedings of the National Academy of Sciences*, 111(45):15912–15917, 2014.

[40] Ping Wang, Lydia Robert, James Pelletier, Wei Lien Dang, Francois Taddei, Andrew Wright, and Suckjoon Jun. Robust growth of escherichia coli. *Current biology*, 20(12):1099–1103, 2010.

[41] N. Brenner, E. Braun, A. Yoney, L. Susman, J. Rotella, and H. Salman. Single-cell protein dynamics reproduce universal fluctuations in cell populations. *Eur. Phys. J E*, 38:102, 2015.

[42] Denis Cottinet, Florence Condamine, Nicolas Bremond, Andrew D Griffiths, Paul B Rainey, J Arjan GM de Visser, Jean Baudry, and Jérôme Bibette. Lineage tracking for probing heritable phenotypes at single-cell resolution. *PloS one*, 11(4):e0152395, 2016.
[43] C. Furasawa, T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko. Ubiquity of log-normal distributions in intra-cellular reaction dynamics. *Biophysics*, 1:25–31, 2005.

[44] A. S. Kennard J. Grilli, M. Osella and M. C. Lagomarsino. Relevant parameters in models of cell division control. *arxiv* 1606.09284, 2016.

[45] David A Kessler and Stanislav Burov. Effective potential for cellular size control. *arXiv preprint arXiv:1701.01725*, 2017.

[46] M. Osella, E. Nugentc, and M. C. Lagomarsino. Concerted control of escherichia coli cell division. *PNAS*, 111(9):3431–3435, 2014.

[47] Manuel Campos, Ivan V Surovtsev, Setsu Kato, Ahmad Paintdakhi, Bruno Beltran, Sarah E Ebmeier, and Christine Jacobs-Wagner. A constant size extension drives bacterial cell size homeostasis. *Cell*, 159(6):1433–1446, 2014.

[48] Lydia Robert, Marc Hoffmann, Nathalie Krell, Stéphane Aymerich, Jérôme Robert, and Marie Doumic. Division in escherichia coli is triggered by a size-sensing rather than a timing mechanism. *BMC biology*, 12(1):1, 2014.

[49] N. Brenner, C. M. Newman, D. Osmanovic, Y. Rabin, H. Salman, and D. L. Stein. Universal protein distributions in a model of cell growth and division. *Phys. Rev. E*, 92:042713, 2015.

[50] Feiqiao Brian Yu, Lisa Willis, Rosanna Man Wah Chau, Alessandro Zambon, Mark Horowitz, Devaki Bhaya, Kerwyn Casey Huang, and Stephen R Quake. Long-term microfluidic tracking of coccoid cyanobacterial cells reveals robust control of division timing. *BMC biology*, 15(1):11, 2017.

[51] Mats Wallden, David Fange, Ebba Gregorsson Lundius, Özden Baltekin, and Johan Elf. The synchronization of replication and division cycles in individual e. coli cells. *Cell*, 166(3):729–739, 2016.

[52] HE Kubitschek. The distribution of cell generation times. *Cell proliferation*, 4(2):113–122, 1971.

[53] W. I. Newman. Emergence of patterns in random processes. ii. stochastic structure in random events. *Phys. Rev. E*, 89(062113), 2014.

[54]

[55] Cyril N Hinshelwood. On the chemical kinetics of autosynthetic systems. *Journal of the Chemical Society (Resumed)*, pages 745–755, 1952.

[56] S. Iyer-Biswas, G. E. Crooks, and N. F. Schererand A. R. Dinner. Universality in stochastic exponential growth. *Phys. Rev. Lett*, 113:028101, 2014.
[57] KL Larrabee, JO Phillips, GJ Williams, and AR Larrabee. The relative rates of protein synthesis and degradation in a growing culture of escherichia coli. *Journal of Biological Chemistry*, 255(9):4125–4130, 1980.

[58] Christian Trötschel, Stefan P Albaum, and Ansgar Poetsch. Proteome turnover in bacteria: current status for corynebacterium glutamicum and related bacteria. *Microbial biotechnology*, 6(6):708–719, 2013.

[59] Robert B Laughlin, David Pines, Joerg Schmalian, Branko P Stojković, and Peter Wolynes. The middle way. *Proceedings of the National Academy of Sciences*, 97(1):32–37, 2000.

[60] B. C. Daniels, Y-J Chen, J. P Sethna, R. N. Gutenkunst, and C. R. Myers. Sloppiness, robustness, and evolvability in systems biology. *Curr. Op. Biotech.*, 19:389395, 2008.

[61] Rolf Lutz and Hermann Bujard. Independent and tight regulation of transcriptional units in escherichia coli via the lacr/o, the tetr/o and arac/i1-i2 regulatory elements. *Nucleic acids research*, 25(6):1203–1210, 1997.

[62] Oleksii Slusarenko, Jennifer Heinritz, Thierry Emonet, and Christine Jacobs-Wagner. High-throughput, subpixel precision analysis of bacterial morphogenesis and intracellular spatio-temporal dynamics. *Molecular microbiology*, 80(3):612–627, 2011.