



Article

A systematic study of the determinants of protein abundance memory in cell lineage

Zongmao Gao^a, Haoyuan Sun^b, Shanshan Qin^a, Xiaojing Yang^{a,*}, Chao Tang^{a,b,c,*}^a Center for Quantitative Biology, Peking University, Beijing 100871, China^b Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China^c School of Physics, Peking University, Beijing 100871, China

ARTICLE INFO

Article history:

Received 19 May 2018

Received in revised form 12 June 2018

Accepted 2 July 2018

Available online 19 July 2018

Keywords:

Protein abundance memory

Intrinsic noise

Extrinsic noise

Epigenetic inheritance

Cell lineage

ABSTRACT

Proteins are essential players of life activities. Intracellular protein levels directly affect cellular functions and cell fate. Upon cell division, the proteins in the mother cell are inherited by the daughters. However, what factors and by how much they affect this epigenetic inheritance of protein abundance remains unclear. Using both computational and experimental approaches, we systematically investigated this problem. We derived an analytical expression for the dependence of protein inheritance on various factors and showed that it agreed with numerical simulations of protein production and experimental results. Our work provides a framework for quantitative studies of protein inheritance and for the potential application of protein memory manipulation.

© 2018 Science China Press. Published by Elsevier B.V. and Science China Press. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Proteins are important functional players in cells, whose abundances directly affect physiological functions and the fate of the cells [1–4]. Since proteins are directly passed on from parent to progeny cells in each cell division, there are correlations in the abundance of each type of proteins within a cell lineage [5]. Such correlations may further lead to similar cell behavior within the lineage, thereby playing an important role in inheritance and environmental adaptation. The progeny cell inherits more than just the specific protein molecules from the parent cell; it also inherits a large number of cellular components, including the transcription and translation machineries, which complicates the protein abundance correlation in a cell lineage. Both the strength and the time scale of protein abundance memory in a cell lineage significantly depend on individual cases [6,7]: while the abundance memory of cyclins in budding yeast lasts less than one cell cycle [8], the memory of the expression of the lac gene in *E. coli* cell lineages lasts for many generations [1].

Recently, an increasing number of studies have suggested that the abundance memory of certain proteins in cell lineages was likely an active and flexible strategy for cells to adapt to changing environments, rather than a simple by-product of cell division [1,5,7,9]. For example, it was suggested that a long memory of

protein abundance may help cell adapt to a fluctuating environment [1]. In addition, some proteins inherited from a parental cell can directly affect the apoptotic probability of the progeny cells [10,11].

Given the importance of protein inheritance in cell lineages, increasing theoretical and experimental efforts have been devoted to study the effects of different factors on protein inheritance and its potential impact on cellular behavior [1,5,12,13]. Such investigations undoubtedly contributed to our understanding of many issues related to protein inheritance. However, most studies thus far have focused on the inheritance of a particular protein or the contribution of one or a few specific factors to protein inheritance [1,10]. In reality, protein inheritance in the cell lineage is often affected by multiple factors at the same time. Do different factors have different effects on protein inheritance? Are the influences of these factors independent or synergetic? What is the quantitative relationship between the protein inheritance and the various influencing factors? The lack of answers to these questions has hindered our understanding of the general phenomenon of protein inheritance.

In this study, by combining theoretical approach and quantitative experiments, we systematically investigated the contributions of various factors to protein memory along the cell lineage. We constructed a simple model and, for the first time, gave an analytic formula for the dependence of protein memory on various factors: protein synthesis, protein degradation, volume ratio for uneven division, doubling time, intrinsic noise, extrinsic noise, partition

* Corresponding authors.

E-mail addresses: xiaojing_yang@pku.edu.cn (X. Yang), tangc@pku.edu.cn (C. Tang).

noise during cell division, and the time scales of different noises. We validated the theory by performing both computer simulations and quantitative experiments.

2. Materials and methods

2.1. Stochastic simulations of the model

All simulations were performed using MATLAB (MathWorks). Stochastic simulation was performed using the Gillespie Algorithm [14–16]. To incorporate the extrinsic noise, the fluctuating production rate of protein was modeled as one “species”. The system had two variables: the production rate number $K_p(t)$ and the protein number $S(t)$. The numbers of the protein and the production rate were updated according to the propensity based on Gillespie Algorithm.

The propensity was derived from the deterministic equations by incorporating a system size Ω through the conventional Van Kampen expansion [17] and a rescaled constant for the protein production rate number. For the model simulation with the extrinsic noise incorporated into the production rate, the propensity function at time t was $[k_{pp}\Omega, 1/\tau, a_1K_p(t), S(t)k_d]$, where $k_{pp}\Omega$ and $1/\tau$ are the birth and death rates of protein production rate number $K_p(t)$, $a_1K_p(t)$ and k_d are the birth and death rates of protein number $S(t)$. In the steady-state, we had $\langle K_p(t) \rangle = k_{pp}\Omega\tau$ for average production rate number and $\langle S(t) \rangle = a_1k_{pp}\Omega\tau/k_d$ for the average protein copy number. We also had $k_p(t) = K_p(t)/\Omega$ for the protein production rate and $X(t) = S(t)/\Omega$ for protein concentration.

After sufficient time simulation, protein copy number reached a steady state, and the generated protein level data were recorded per cell doubling time T . The process was simulated 2000 times. Then, the Pearson correlation was calculated based on the recorded proteins.

2.2. Single-cell measurements using time-lapse microscopy

Standard methods were used throughout the study. To prepare the cells for time-lapse microscopy, we inoculated congenic W303 (*MATa his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) cells from a colony into liquid SD, grew the cells for 12 h, and then diluted and cultured them for 12 h. Next, the cells that grew exponentially in SD liquid medium were seeded into a microfluidic chip in the same medium. For each experiment, stacks of 9 images were acquired every 5 min with 30 ms exposure for the bright-field channel,

and 50 ms for the red channel and green channels. Microcolonies were tracked throughout the time series by identifying overlapping areas. Cell segmentation and tracing were performed based on bright field images and automatically obtained using the MATLAB customized software cellseg, which we previously developed [18,19]. Fluorescence quantification was performed using cellseg and ImageJ with Image5D plugin. The maximum intensity projection of z-stacks was reported for experiments to obtain the protein intensity.

2.3. Quantification of protein half-lives by FACS

W303 yeast strain with *Adh1Pr-GFP* (expressing GFP protein) or *Adh1Pr-GFP-PEST* (expressing GFP protein with a PEST tag) were grown in 5 mL of synthetic medium with 2% (w/v) glucose overnight at 30 °C and rotating. The overnight culture was diluted to an OD₆₀₀ value of 0.1 in 20 mL of fresh medium and incubated until the cells reached the mid-logarithmic growth phase. Cycloheximide (translation inhibitor [20,21]) was added to a final concentration of 200 µg/mL, which is sufficiently high to inhibit protein synthesis without inducing a critical growth defect during the experiment. Next, 0.5 ml of yeast cells were quickly obtained from the culture every 10 min and 4.5 mL PBS buffer was added. FACS was performed to determine the protein fluorescence.

3. Results

3.1. Memory of protein abundance along the cell lineage with intrinsic noise

To investigate the memory of protein abundance along the cell lineage, we first constructed a simple model in which the rate of protein deposition is determined by the rates of protein synthesis (k_p) and degradation/dilution (k_d). The degradation/dilution rate k_d includes two parts (i.e., $k_d = k_{dil} + k_{deg}$): regulated degradation (k_{deg}) and dilution rate (k_{dil}) due to cell growth and division [22]. We considered both symmetrical and asymmetrical division by introducing the division size ratio α , which is defined as the ratio of corresponding volume to the total volume (i.e., $\frac{V_M}{V_M+V_D}$ for mother/larger cells, and $\frac{V_D}{V_M+V_D}$ for daughter/smaller cells). As a result, $k_{dil} = -\ln(\alpha)/T$, where T is the cell doubling time [5]. We assumed that the system was in steady state, and only the mother and daughter lineages were discussed for simplicity (Fig. 1).

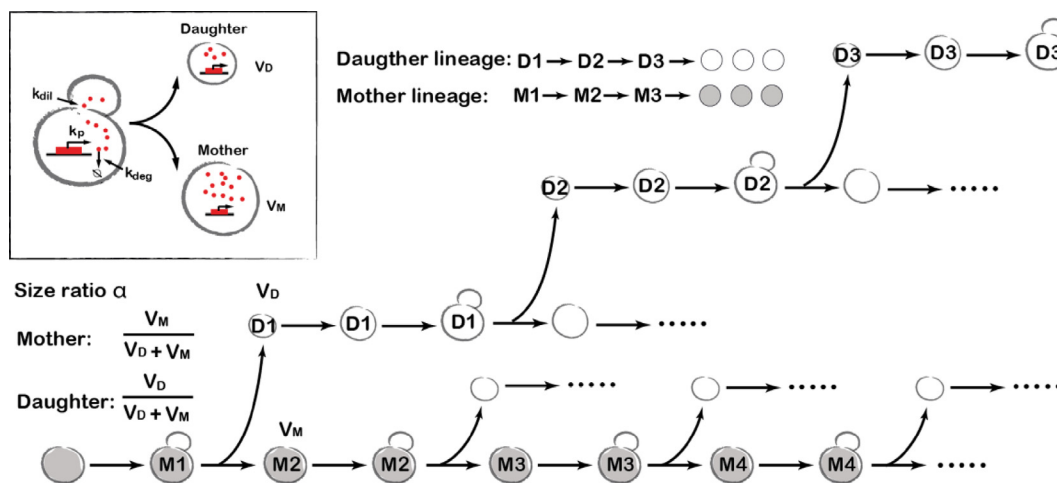


Fig. 1. Schematic view of the protein inheritance model in a cell lineage. The parameter k_p represents the protein production rate, while k_{deg} and k_{dil} represent the degradation rate and the dilution rate, respectively. The cell size ratio is the ratio of corresponding volume to the total volume (i.e., $\frac{V_M}{V_M+V_D}$ for mother cells, and $\frac{V_D}{V_M+V_D}$ for daughter cells).

Because intrinsic noise and extrinsic noise may have different effects on protein memory in the cell lineage [5], we first investigated a system with intrinsic noise only. The dynamics of protein concentration follows the stochastic differential equation (SDE) [23,24]

$$\frac{dX(t)}{dt} = k_p - k_d X(t) + \sqrt{D} \zeta(t), \quad (1)$$

where $X(t)$ is the protein concentration, k_p the production rate, k_d the degradation/dilution rate. $\zeta(t)$ represents the intrinsic noise, which is a rapidly fluctuating random variable with zero mean $\langle \zeta(t) \rangle = 0$ and $\langle \zeta(t) \zeta(t') \rangle = \delta(t - t')$ [25,26]. D is the noise strength. We assumed that the system was at steady state and that the fluctuation of protein abundance around its steady-state value $x(t) = X(t) - \langle X(t) \rangle$ was small. Note that our model was constructed simply based on the production rate and degradation/dilution rate without considering other kinds of regulations. Thus, the protein expression only had one steady state.

Pearson correlation coefficients between different generations were used to measure the strength of protein memory. The Pearson correlation of protein abundance between generation 1 and generation n in the mother/daughter lineages can be calculated as the following equation (see Supplementary Materials for details):

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.scib.2018.07.010>.

$$\rho(x^1, x^n) = \alpha^{(n-1)} e^{-(n-1)k_{deg}T}, \quad (2)$$

where $x^i (i = 1, 2, \dots, n)$ is the protein abundance fluctuation of the i th generation cell in the mother/daughter lineage, α is the cell size division ratio, and T is the corresponding mother/daughter cell doubling time.

Thus, with only the intrinsic noise, the protein correlation increases with the cell size division ratio α and decreases exponentially with the degradation rate k_{deg} and the doubling time T .

3.2. Memory of protein abundance along the cell lineage with both intrinsic noise and extrinsic noise

Next, we added extrinsic noise into the system and investigated the memory of protein abundance in the cell lineage with both intrinsic and extrinsic noises. Extrinsic noise may derive from either or both of production and degradation/dilution processes due to the fluctuation of environment and/or cell heterogeneity. We studied separately the impact of extrinsic noise on production and degradation/dilution.

We first incorporated extrinsic noise into the production rate [27]. We assumed that the fluctuation of the production rate k_p was described by the following equation [15,24,28,29]:

$$\frac{dk_p(t)}{dt} = k_{pp} - \frac{1}{\tau} k_p(t) + \sqrt{\frac{2\sigma_{ex}^2}{\tau}} \zeta(t), \quad (3)$$

where $\zeta(t)$ represents a Gaussian white noise with $\langle \zeta(t) \zeta(t') \rangle = \delta(t - t')$ and zero mean $\langle \zeta(t) \rangle = 0$. σ_{ex}^2 is the variance of the production rate and τ is the correlation timescale of the extrinsic noise.

Then, the protein correlation in the mother/daughter lineage can be calculated as follows (see Supplementary Materials for details):

$$\rho(x^1, x^n) = \alpha^{(n-1)} e^{-k_{deg}(n-1)T} + \frac{2\sigma_{ex}^2(k_{dil} + k_{deg})(\alpha^{(n-1)} e^{-k_{deg}(n-1)T} - e^{-\frac{(n-1)T}{\tau}})}{(D(\frac{1}{\tau} + k_{dil} + k_{deg}) + 2\sigma_{ex}^2)(\frac{1}{\tau} - k_{dil} - k_{deg})}. \quad (4)$$

Eq. (4) was plotted in Fig. 2 with various factors as variables. Note that the dependence of protein memory on the division size ratio α , the degradation rate k_{deg} , and the doubling time T are similar as in case of the intrinsic noise only.

Eq. (4) can be simplified in two important limits. In the white noise limit, $\tau \rightarrow 0$ ($1/\tau \rightarrow \infty$, i.e. the extrinsic noise fluctuates

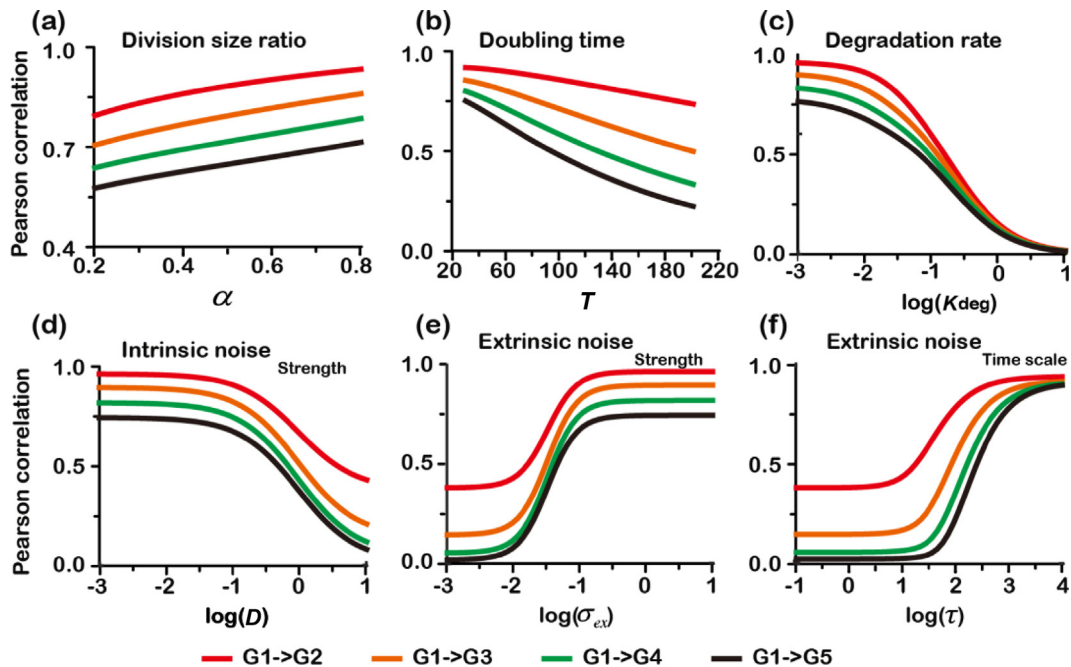


Fig. 2. Dependence of protein memory along the cell lineage on different factors based on Eq. (4). Each subpanel shows the relationship between one factor and the Pearson correlation. The parameter of the factor is varied while other parameters fixed: (a) cell size division ratio α , (b) doubling time T , (c) protein degradation rate k_{deg} , (d) strength of intrinsic noise D , (e) amplitude of extrinsic noise σ_{ex} , and (f) the time scale of extrinsic noise τ . Different colors represent the correlation between 1st generation and different generations along the cell lineage. Parameters values are $T = 50$, $k_{deg} = 0.01$, $\alpha = 0.63$, $\tau = 500$, $D = 0.1$, $\sigma_{ex} = 0.1$.

rapidly), the above Eq. (4) is degenerated into Eq. (2), the case with intrinsic noise only. Hence, fast fluctuating white extrinsic noise does not change the protein memory behavior. In the adiabatic limit, $\tau \rightarrow \infty$ ($1/\tau \rightarrow 0$, i.e., the extrinsic noise is almost a constant in a long time), then the protein correlation in the mother/daughter lineage is simplified as (see Supplementary Materials for details):

$$\rho(x^1, x^n) = 1 + \frac{1}{\eta} (\alpha^{(n-1)} e^{-(n-1)k_{\text{deg}}T} - 1), \quad (5)$$

where η ($\eta = \sigma_{\text{Total}}^2 / \sigma_{\text{Intrinsic}}^2$) is the variance ratio between the total variance of protein abundance and the variance caused by intrinsic noise.

Note that there can be two different sources of extrinsic noise. One is the external environment fluctuation; the other originates from cell to cell variability. Thus, the adiabatic limit case ($\tau \rightarrow \infty$) can be further subdivided into two different cases. In case A, the population of the cells under consideration was reproduced from one single cell. In this case, intrinsic noise is dominant ($\eta \rightarrow 1$), and Eq. (5) is degenerated into Eq. (2). In case B, the population of the cells under consideration was reproduced from

multiple cells that had some initial cell-to-cell variability. When the cell-to-cell variability is small, this case is similar to case A. When the cell-to-cell variability is dominant in the system ($\eta \gg 1$), the protein abundance in the cell lineage becomes highly correlated ($\rho \rightarrow 1$). These results show that intrinsic and extrinsic noises play different roles in protein memory.

Next, we tried to incorporate extrinsic noise into the protein degradation/dilution process, and then into both the protein synthesis and degradation/dilution processes. We found that the contribution of different factors to the correlation were generally similar (Figs. S1, S2, S3 and see Supplementary Materials for details). There is, however, an interesting non-monotonic dependence of correlation on the time scale of extrinsic noise (Fig. S2c and d), due to the interaction between the noise from production rate and that from degradation/dilution rate.

3.3. Small copy numbers of protein

In order to make the protein correlation solvable analytically, we described the stochastic behavior of protein abundance using the Langevin equation (Eq. (1)), where the intrinsic noise amplitude was determined by a free parameter D . However, the

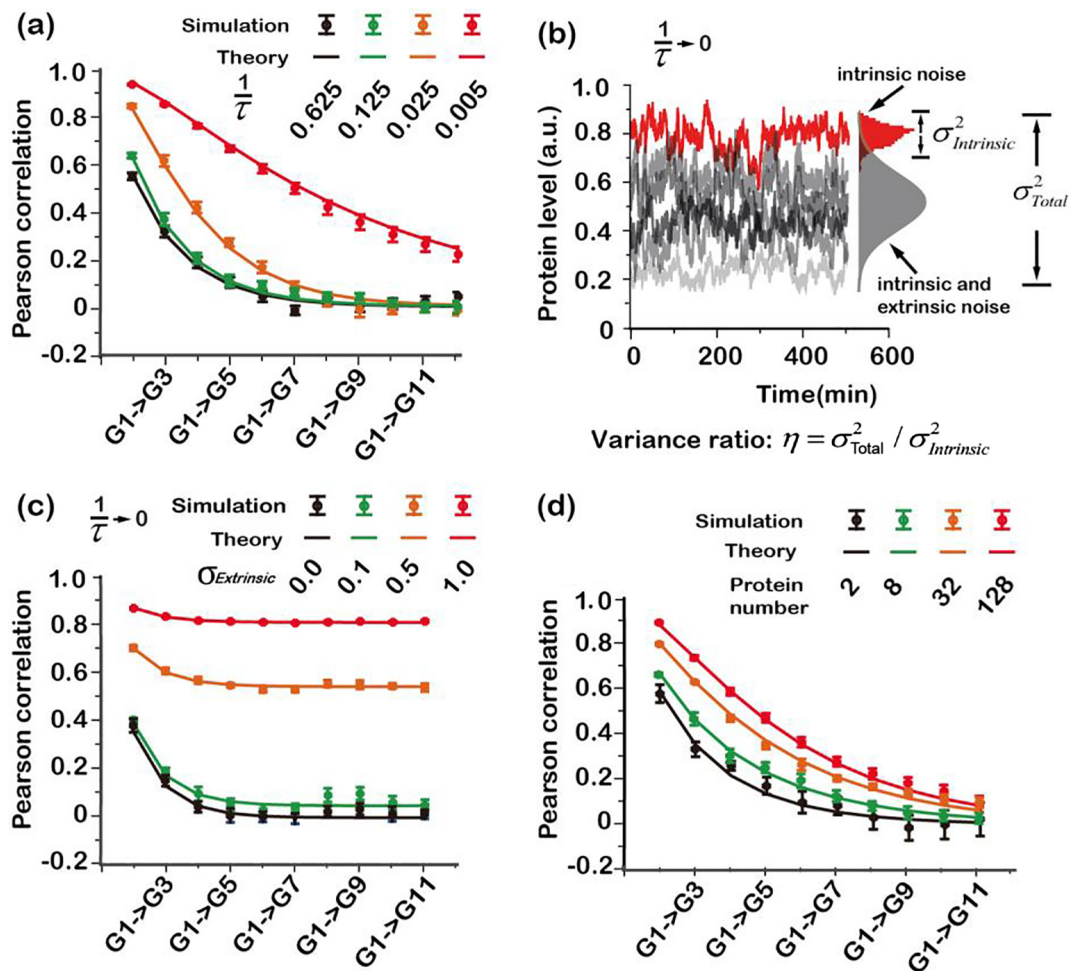


Fig. 3. Comparison of the protein inheritance between theory (lines) and simulation (symbols) with extrinsic noise incorporated into the production rate. (a) The results with different time scales of extrinsic noise. The parameter values were $\Omega = 500$, $k_{pp}\tau = 0.02$, $k_d = 0.02$, $a_1 = 1$ and $T = 30$, where Ω is the system volume size, and a_1 is a rescale constant for production rate in the simulation. (b) The schema of variance caused by intrinsic noise, total variance of protein abundance, and the variance ratio for the adiabatic limit case B. (c) The results with different strength of extrinsic noise for the adiabatic limit case B. The parameter values were $\Omega = 500$, $k_p\Omega \sim N(\mu, \sigma_{\text{extrinsic}})$, $\mu = 10$, $k_{pp} = 0$, $1/\tau = 0$, $k_d = 0.02$, $a_1 = 1$ and $T = 30$. For each simulation, the value of production rate was sampled from the normal distribution. (d) The results with different average protein copy number. The parameters value were $\Omega = 500$, $k_{pp}\Omega = 0.1$, $1/\tau = 0.01$, $k_d = 0.02$, and $T = 30$. The value of a_1 was calculated based on $\langle S(t) \rangle = a_1 k_{pp}\Omega\tau/k_d$, when the average of protein number $\langle S(t) \rangle$ was set to a specific value. The Pearson correlation coefficients of protein levels were estimated from an ensemble of 2000 simulated cells, error bars show the standard deviation of 10 independent simulations (see Materials and Methods for more details).

parameter D is related to the production and degradation/dilution rates in the real chemical reaction processes. In addition, our SDE of protein abundance rely on assumption of small noise, which could breakdown when the copy number of proteins are very small [30], as is the case for many proteins in *E. coli* and budding yeast [31,32].

To address these concerns, we used the Gillespie algorithm to simulate the gene expression and degradation/dilution processes with both intrinsic and extrinsic noises [14–16]. In order to compare the simulation results with the theoretical prediction (Eq. (4)), we estimated the noise strength D and the variance of production rate σ_{ex}^2 based on the chemical Langevin equation [33]. In this case, we had $D = 2k_p$, where parameter k_p is the production rate, and $\sigma_{ex}^2 = k_{pp}\tau$, where k_{pp} and $1/\tau$ is the birth and death terms of production rate (Eq. (3)) [24,33] (see Supplementary Materials for details).

Our stochastic simulation using the Gillespie algorithm agrees perfectly with the theory prediction (Fig. 3). It is worth noting that, in the adiabatic limit case B (i.e. $1/\tau \rightarrow 0$), the production rate for each single cells is a constant, but drawn from a distribution for

a population of cells due to cell-to-cell heterogeneity. Consequently, the temporal protein abundance fluctuation of single cell is controlled by the intrinsic noise, and $\sigma_{intrinsic}^2$ and σ_{Total}^2 can be estimated by measuring the variance of the protein level using single cell time series and among the population, respectively (Fig. 3b). In simulation, we directly investigated the influence of the variance ratio $\eta = \sigma_{Total}^2/\sigma_{intrinsic}^2$ by verifying the variance of protein production rate for the cell population as $1/\tau \rightarrow 0$ (Fig. 3c).

We also examined the effect of average protein copy number in the steady state. We varied the average protein copy number from 128, 32, 8 to 2. The theory showed consistently very good agreement with the simulations for all cases, even when the average protein copy number decreased to 2 (Fig. 3d).

3.4. Comparison with experiments

We further tested our theory with quantitative experiments. We used budding yeast because it grows fast and is easy to manipulate genetically. More importantly, its asymmetrical division not only enabled us to examine the influence of the cell size division

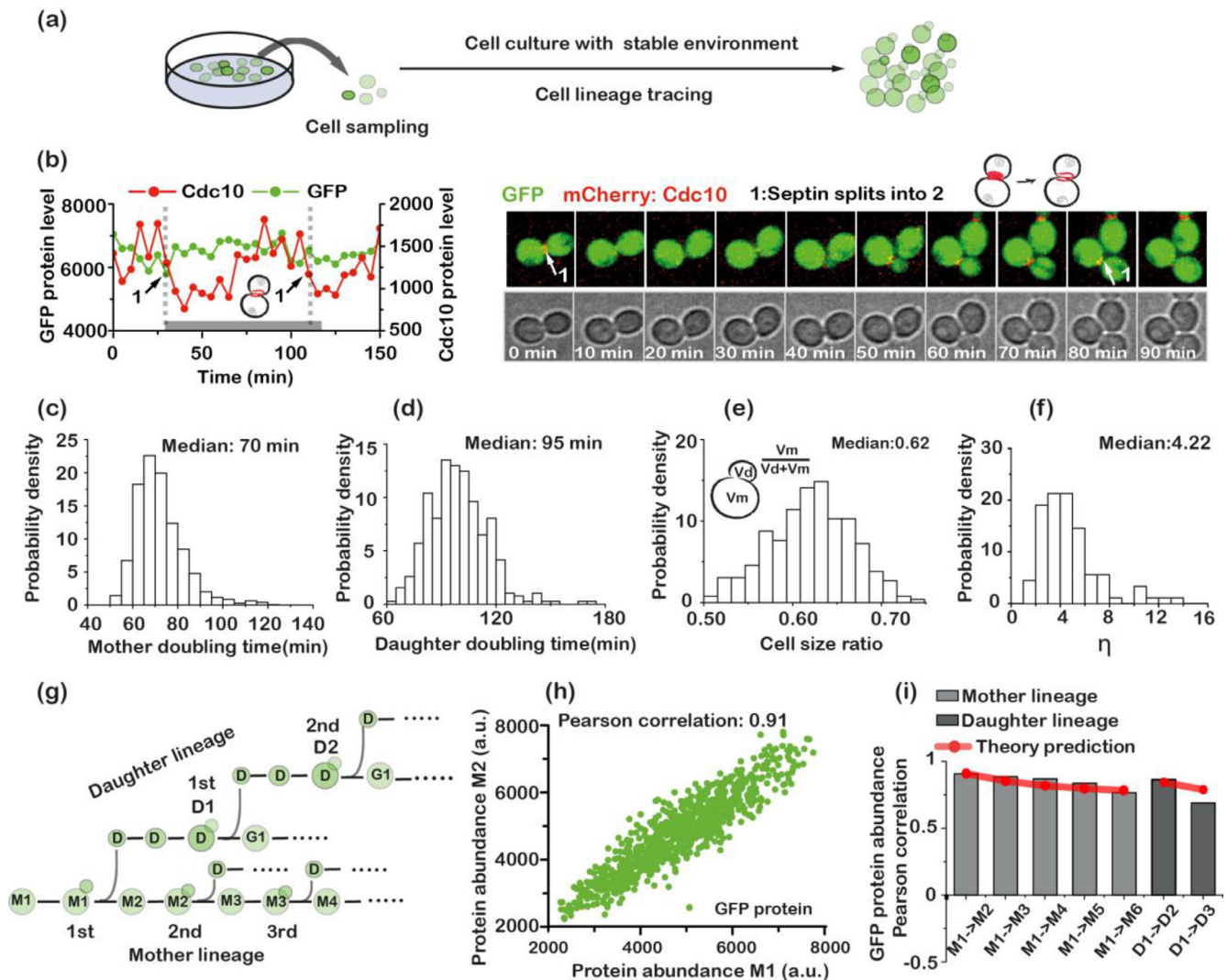


Fig. 4. Protein memory experiments with GFP. (a) Schematic diagram of experimental process. (b) Right panel: A typical phase/fluorescence time-course images (green, Adh1pr-GFP; red, Cdc10, cell cycle marker). Left panel: the corresponding profiles of protein levels. '1' marks the time when cytokinesis (septin splits into two) happens, and the time between two consecutive cytokinesis was defined as one generation. (c–f) The distributions of different parameters estimated from the experimental data. (g) Schematic diagram of the mother and daughter lineages, showing the time points of data analysis. (h) Scatter plot of the GFP protein abundance between M1 and M2. (i) Experimental results (bar) and theoretical predictions based on the parameters from the experimental data (red line) for GFP abundance along the mother lineage and daughter lineage.

ratio, but also provided two different protein dilution rates (doubling times) for mother and daughter cells simultaneously. We employed a microfluidics device, which enabled us to grow the cell population in a very stable environment. This setting would correspond to the case B with adiabatic limit for extrinsic noise ($1/\tau \rightarrow 0$) (Fig. 4a). Thus, we could directly estimate the variance ratio by measuring the variance of protein abundance in a single cell and in a population of cells, as shown in Fig. 3b.

First, we performed protein inheritance experiments with a GFP controlled by a constitutive promoter Adh1pr and not being actively degraded (i.e. $k_{deg} \approx 0$) (Fig. 4b) [34]. We monitored the concentration of GFP in individual cells in real time over generations using time-lapse fluorescence microscopy. To define generations, we fused an mCherry to the septin ring component Cdc10, whose assembly and disassembly mark the cell cycle entry and cytokinesis, respectively (Fig. 4b) [35]. Moreover, to avoid potential fluctuations due to different cell cycle phases, we always quantified protein abundance at the end of cell division (defined as when the septin ring splits into two) (Fig. 4b).

The protein abundance of hundreds of cells for each generation was obtained, and the Pearson correlations of the protein levels between the first generation and offspring generations were

calculated (Fig. 4g and h). Meanwhile, all parameters in the theory, including the cell size division ratio, doubling time, variance ratio η , were also obtained from the experimental data (Fig. 4c–f), and the median values of these parameters were then used to predict the lineage protein correlation based on Eq. (5) without parameter fitting.

We observed excellent agreement between theoretical predictions and the experimental results for the mother lineage (Fig. 4i). For daughter lineage, the agreement was very good for the first to the second daughter. But for the first to the third daughter, the experimentally measured correlation was lower than the theoretic prediction. A possible reason may be that in contrast to the mother cell lineage, the protein abundance in the daughter cell lineage was obtained by monitoring different cells, which may introduce additional noise. Also, it was difficult to trace the daughter cell lineage more than 3 generations in our experimental system and we had only limited data for the 3rd generation daughters.

Next, we tested the theory with a degradable GFP ($k_{deg} \neq 0$). We fused a PEST degen to GFP (GFP-PEST), which would introduce an active degradation of the protein [36,37]. We first monitored the degradation of both GFP and GFP-PEST using cell cytometry after the addition of a protein synthesis inhibitor, and observed that

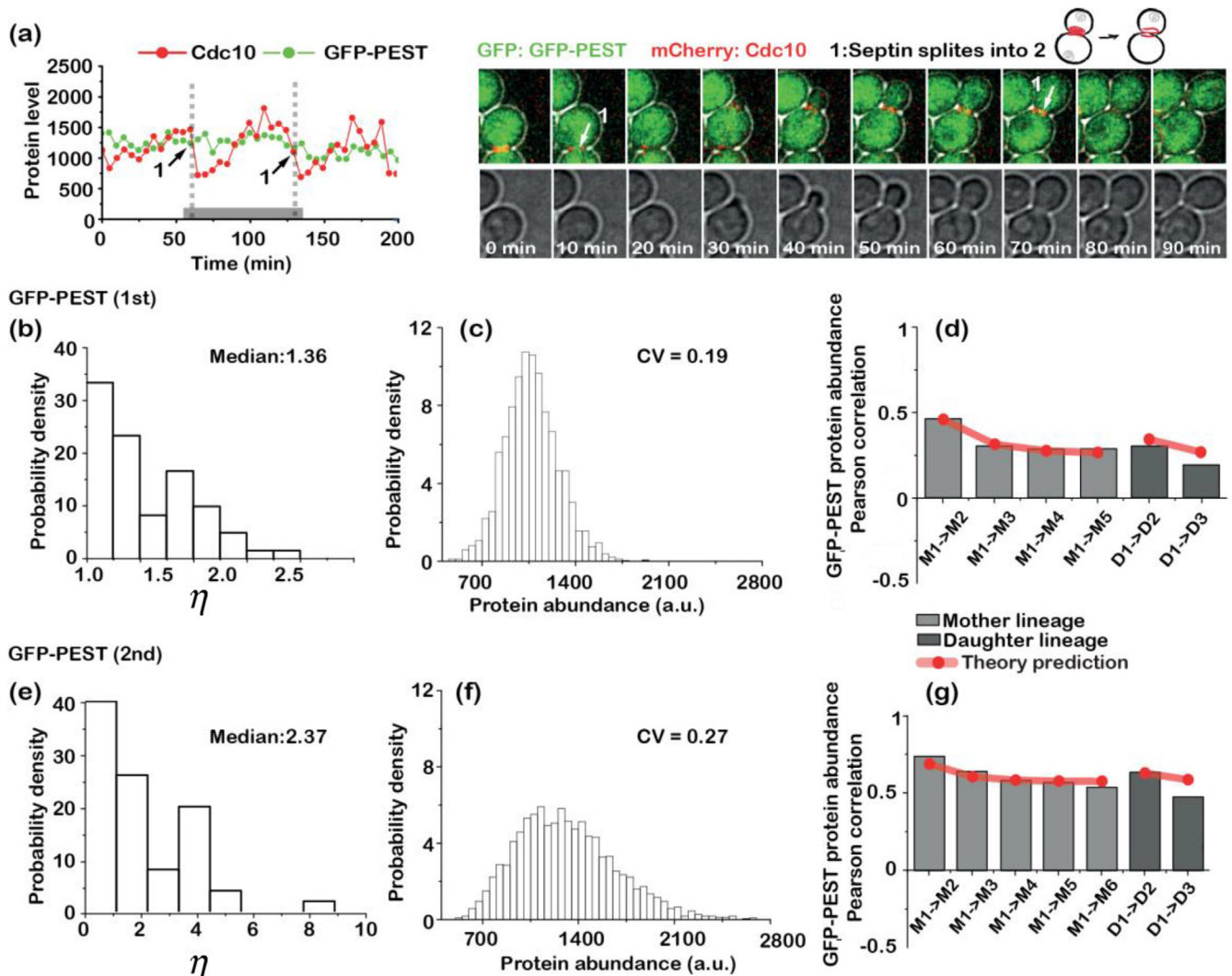


Fig. 5. Protein memory of GFP-PEST in cell lineage. (a) Right panel: A typical phase/fluorescence time-course images (green, Adh1pr-GFP-PEST; red, Cdc10, cell cycle marker). Left panel: the corresponding profiles of protein levels. "1" marks the time when cytokinesis (septin splits into two) happens, and the time between two consecutive cytokinesis was defined as one generation. (b) and (e) show the distributions of the protein variance ratio η obtained in the two experiments. (c) and (f) show the distributions of GFP-PEST protein abundance. (d) and (g) GFP-PEST memory along mother and daughter lineages estimated from experiments (bars) and predicted by the theory (red lines).

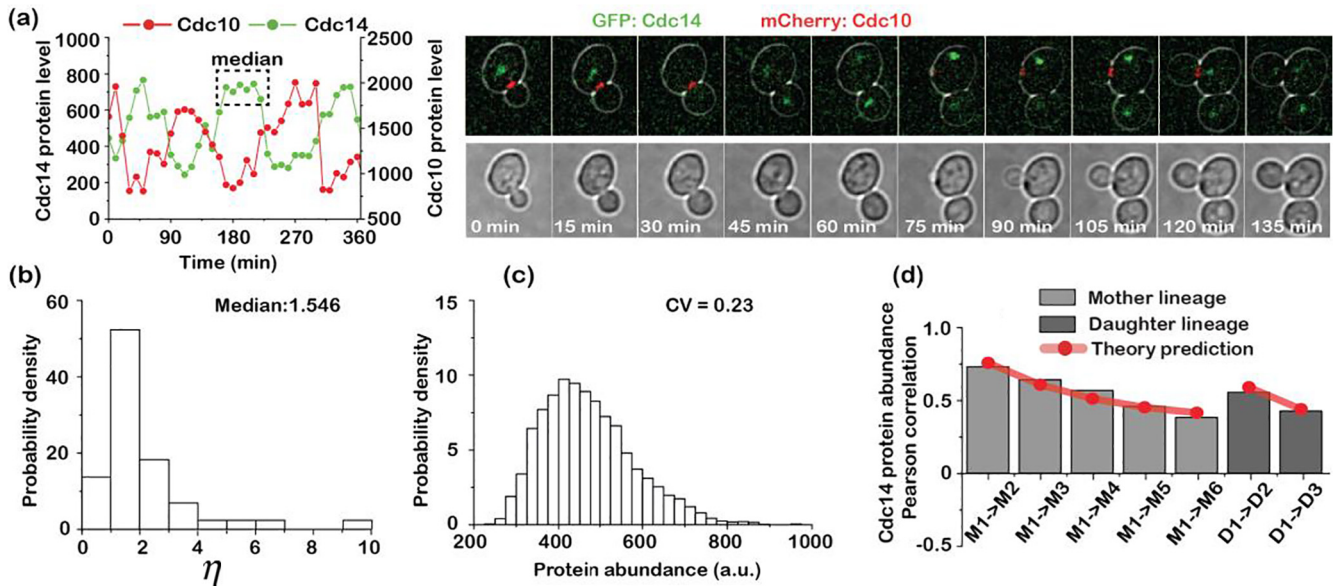


Fig. 6. Protein memory of the Cdc14 in cell lineages. (a) Right panel: A typical phase/fluorescence time-course images (green, Cdc14; red, Cdc10, cell cycle maker). Left panel: the corresponding profiles of protein levels. The median level of protein Cdc14 as it localized in the nucleolus was used. (b) The distribution of protein variance ratio η obtained in the experiments. (c) The distribution of Cdc14 protein abundance. (d) Cdc14 memory along the cell lineage from experiments (bars) and predicted by the theory (red lines).

GFP-PEST degraded much faster than GFP (Fig. S4). Note that the degradation of the PEST sequence depends on its phosphorylation state, and thus the degradation rate of GFP-PEST is cell-cycle dependent [38]. Indeed, we found that the fluctuation of protein abundance of GFP-PEST correlates with cell cycle progression (Fig. 5a). However, according to our theory the protein inheritance memory depends on the average degradation rate within a cell cycle. The measured average degradation rate of GFP-PEST was $k_{\text{deg}} = 0.0112$ which was used in Eq. (5) for prediction. We performed the lineage experiments twice using GFP-PEST with different initial cell-to-cell variability (Fig. 5c and f), which resulted in significantly different the protein abundance correlations along the lineages (Fig. 5d and g). Remarkably, the theoretic predictions agreed very well with both experiments (Fig. 5d and g), (except for the first to the third daughter for the same reason discussed above).

Furthermore, we tested the theory with an endogenous protein Cdc14, which is an essential cell cycle phosphatase. The localization of Cdc14 is regulated by multiple factors, although it does not seem to have regulated degradation [39,40]. During the early mitosis, Cdc14 is sequestered in the nucleolus by Net1 until M/A transition [39]. Since the localization of Cdc14 changes significantly during cell cycle, the median level of Cdc14 as it localized in the nucleolus was used in our correlation calculation to minimize the error (Fig. 6a). Again, the protein abundance correlation obtained from the experiment agreed very well with the theory (Fig. 6d).

4. Summary

Protein abundance memory plays an important role for cells to adapt to the environment [1]. However, multiple factors could affect the protein memory, such as protein production, degradation, random partition noise at division, extrinsic noise, intrinsic noise and the time scales of these noises. In this study, we combined theoretical, computational and experimental approaches, and systematically investigated the contributions of different factors to protein memory in cell lineages.

We found that these factors significantly and differently affected the protein memory in the cell lineage. First, the protein level correlation decreases with decreasing division size ratio (α). This would imply that the smaller daughter cell is naturally "fresher" (less correlated with the parent cell) than the mother cell, even without any specific asymmetric aging mechanisms [41]. Moreover, daughter cells commonly have a longer doubling time, making them even more "fresher" (lesser correlated with their parent cell). Second, the protein level correlation decreases exponentially with the protein degradation rate, suggesting that changing the degradation rate could be an easy way to adjust protein memory. Given that protein degradation is a common method of regulation and it can also be subject to environmental cues, this observation may provide insights into the regulation of protein degradation. Third, while intrinsic noise tends to diminish the protein abundance correlation in a cell lineage, extrinsic noise enhances the correlation (Fig. 2d and e).

Generally, there are two layers of protein inheritance, total protein amount and the protein with post-translational modification regulation. Here, we mainly focused on the memory of total protein amount, since the mechanism of protein post-translational modification varies case by case so that it is hard to have a general theoretical expression for this layer of inheritance. However, for the case in which protein with post-translational modification can be viewed as a continuous process (the post-translational modification rate can be viewed as the production rate), our theory could also be used to predict the protein memory.

We chose budding yeast to test our theory experimentally, and the theory agrees remarkably well with our quantitative experiments. Note that, although in principle all parameters described in our model could be tuned experimentally (such as cell size ratio by comparing the symmetric and asymmetric division cells, doubling time by culturing cells with different kinds of medium), in practice most parameters are correlated with each other. For example, in the cases of the protein memory of GFP and of GFP-PEST, adding PEST degron actually did not only change the protein degradation, but also the noise level. This is something to keep in mind, especially when trying to tune the protein memory.

Our theory was based on very general assumptions. The conclusions are not limited to yeast, but applicable to other cell types, such as bacterial and mammalian cells. Our work provides a framework and a platform to further explore protein memory in more complex situations, e.g. fluctuating environments with various time scales.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

We thank Qi Liu, Junyuan Gong and Lu Zhang for helpful discussions. This work was supported by the Chinese Ministry of Science and Technology (2015CB910300) and the National Natural Science Foundation of China (NSFC31700733).

Author contribution

Conceived and designed the projects: X.Y., Z.G. and C.T.; designed, performed experiments and analyzed data: Z.G., H.S., X. Y.; developed models and conducted the simulations: Z.G. and S. Q.; Wrote the manuscript X.Y., Z.G., S.Q. and C.T. All authors read and approved the final manuscript.

References

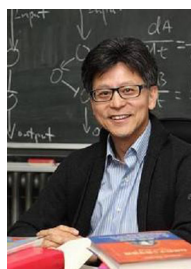
- [1] Lambert G, Kussel E. Memory and fitness optimization of bacteria under fluctuating environments. *PLoS Genet* 2014;10.
- [2] Levy SF, Ziv N, Siegal ML. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol* 2012;10.
- [3] Cerulus B, New AM, Pougach K, et al. Noise and epigenetic inheritance of single-cell division times influence population fitness. *Curr Biol* 2016;26:1138–47.
- [4] Brickner DG, Cajigas I, Fondufe-Mittendorf Y, et al. H2A.Z-mediated localization of genes at the nuclear periphery confers epigenetic memory of previous transcriptional state. *PLoS Biol* 2007;5:704–16.
- [5] Sigal A, Milo R, Cohen A, et al. Variability and memory of protein levels in human cells. *Nature* 2006;444:643–6.
- [6] Rando OJ, Verstrepen KJ. Timescales of genetic and epigenetic inheritance. *Cell* 2007;128:655–68.
- [7] Veening J-W, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* 2008;62:193–210.
- [8] Bloom J, Cross FR. Multiple levels of cyclin specificity in cell-cycle control. *Nat Rev Mol Cell Biol* 2007;8:149–60.
- [9] Snijder B, Pelkmans L. Origins of regulated cell-to-cell variability. *Nat Rev Mol Cell Biol* 2011;12:119–25.
- [10] Spencer SL, Gaudet S, Albeck JG, et al. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 2009;459:428–32.
- [11] Paek AL, Liu JC, Loewer A, et al. Cell-to-cell variation in p53 dynamics leads to fractional killing. *Cell* 2016;165:631–42.
- [12] Rosenfeld N, Young JW, Alon U, et al. Gene regulation at the single-cell level. *Science* 2005;307:1962–5.
- [13] Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature* 2010;467:167–73.
- [14] Gillespie DT. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem* 2007;58:35–55.
- [15] Gillespie D. Exact numerical simulation of the Ornstein-Uhlenbeck process and its integral. *Phys Rev E* 1996;54:2084–91.
- [16] Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 1977;81:2340–61.
- [17] Wallace EWJ, Petzold LR, Gillespie DT, et al. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET Syst Biol* 2012;6:102–15.
- [18] Lau K. Elucidating the systems design principles of the yeast cell cycle network [doctoral dissertation]. 2009.
- [19] Yang X, Lau KY, Sevim V, et al. Design principles of the yeast G1/S switch. *PLoS Biol* 2013;11.
- [20] Ennis HL, Lubin M. Cycloheximide: aspects of inhibition of protein synthesis in mammalian cells. *Science* 1964;146:1474–6.
- [21] Schneider-Poetsch T, Ju J, Eyler DE, et al. Inhibition of eukaryotic translation elongation by cycloheximide and lactimidomycin. *Nat Chem Biol* 2010;6:209–17.
- [22] Alon U. An introduction to systems biology: design principles of biological circuits. 2008;215.
- [23] Van Kampen NG. Stochastic processes in physics and chemistry. 1981;2.
- [24] Qin S, Tang C. Early-warning signals of critical transition: effect of extrinsic noise. *Phys Rev E* 2018;97:32406.
- [25] Elowitz M, Levine AJ, Siggia ED, et al. Stochastic gene expression in a single cell. *Science* 2002;297:1183–6.
- [26] Kærn M, Elston TC, Blake WJ, et al. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 2005;6:451–64.
- [27] Hilfinger A, Paulsson J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc Natl Acad Sci U S A* 2011;108:12167–72.
- [28] Shahrezaei V, Swain PS. The stochastic nature of biochemical networks. *Curr Opin Biotechnol* 2008;19:369–74.
- [29] Dunlop MJ, Cox RS, Levine JH, et al. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nat Genet* 2008;40:1493–8.
- [30] Paulsson J. Models of stochastic gene expression. *Phys Life Rev* 2005;2:157–75.
- [31] Choi PJ, Cai L, Frieda K, et al. A stochastic single molecule event triggers phenotype switching of a bacterial cell. *Science* 2008;322:442.
- [32] Cai L, Friedman N, Xie XS. Stochastic protein expression in individual cells at the single molecule level. *Nature* 2006;440:358–62.
- [33] Gillespie DT. Chemical Langevin equation. *J Chem Phys* 2000;113:297–306.
- [34] Corish P, Tylersmith C. Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Eng* 1999;12:1035.
- [35] Madden K, Snyder M. Cell polarity and morphogenesis in budding yeast. *Annu Rev Microbiol* 1998;52:687–744.
- [36] Rechsteiner M, Rogers SW. PEST sequences and regulation by proteolysis. *Trends Biochem Sci* 1996;21:267–71.
- [37] Rogers S, Wells R, Rechsteiner M, et al. Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. *Science* 1986;234:364–8.
- [38] García-Alai MM, Gallo M, Salame M, et al. Molecular basis for phosphorylation-dependent, PEST-mediated protein turnover. *Structure* 2006;14:309–19.
- [39] Stegmeier F, Amon A. Closing mitosis: the functions of the Cdc14 phosphatase and its regulation. *Annu Rev Genet* 2004;38:203–32.
- [40] Sullivan M, Morgan DO. Finishing mitosis, one step at a time. *Nat Rev Mol Cell Biol* 2007;8:894–903.
- [41] Eldakak A, Rancati G, Rubinstein B, et al. Asymmetrically inherited multidrug resistance transporters are recessive determinants in cellular replicative ageing. *Nat Cell Biol* 2010;12:799–805.



Zongmao Gao is a Ph.D. candidate in the Center for Quantitative Biology, Peking University. His undergraduate background is in physics. His graduate study is interdisciplinary between physics and biology, with current research interest in the robustness of cell cycle and how the cell responds to fluctuating environment.



Xiaojing Yang received her Ph.D. degree from the School of Physics, Peking University. She was a postdoctoral researcher at University of California, San Francisco, USA. She currently is a research associate in the Center for Quantitative Biology at Peking University, and her research interest focuses on the cell cycle regulation and optogenetics.



Chao Tang received his Ph.D. degree in physics from the University of Chicago. He was a Full Professor at the University of California San Francisco before returning to China full time in 2011. He is now a Chair Professor of Physics and Systems Biology, and the director of the Center for Quantitative Biology at Peking University. His current research interest is at the interface between physics and biology.