Entropy, Ergodicity and Stem Cell Multipotency

Sonya J. Ridden,[1] Hannah H. Chang,[2] Konstantinos C. Zygalakis,[1] and Ben D. MacArthur[1,3,*]

[1] *Mathematical Sciences, University of Southampton, SO17 1BJ, UK*

[2] *5AM Ventures, Boston, MA 02109, USA*

[3] *Centre for Human Development, Stem Cells and Regeneration, University of Southampton, SO16 6YD, UK*

(Dated: October 20, 2015)

Populations of mammalian stem cells commonly exhibit considerable cell-cell variability. However, the functional role of this diversity is unclear. Here, we analyze expression fluctuations of the stem cell surface marker Sca1 in mouse hematopoietic progenitor cells using a simple stochastic model and find that the observed dynamics naturally lie close to a critical state, thereby producing a diverse population that is able to respond rapidly to environmental changes. We propose an information-theoretic interpretation of these results that views cellular multipotency as an instance of maximum entropy statistical inference.

PACS numbers: 87.10.Vg, 89.70.-a, 89.70.Cf, 87.18.Tt, 87.10.Mn, 87.17.Aa

arXiv:1504.07266v3 [q-bio.CB] 16 Oct 2015

Clonal populations of unicellular organisms often exhibit phenotypic diversity, which confers selective advantage under adverse environmental conditions. Well-known examples include antibiotic bacterial persistence, the lysis-lysogeny switch of λ-phage, competence development and sporulation of *B. subtilis*, and lactose uptake by *E. coli* [1]. The ubiquity of this phenomenon indicates that it is a generic, evolvable, mechanism that facilitates collective cellular dynamics by enabling robust, rapid responses to diverse environmental changes. Recently, stochastic fluctuations in expression of important marker proteins have been seen to generate functional diversity within multipotent mammalian stem cell populations, suggesting a similar role for cell-cell variability in higher organisms [2]. These observations have motivated speculation that functional multipotency (the ability to differentiate along a number of distinct cellular lineages) is a collective property of stem and progenitor cell populations, reflective of fitness constraints imposed at the population, rather than individual cell, level [3]. This perspective is appealing since such regulated cell-cell variability in principle allows a cellular population to remain primed to respond quickly to a range of different differentiation cues while remaining robust to cell loss. However, convincing demonstrations of the potency of individual stem cells appear to argue strongly against such a collective view (for example, single long-term repopulating hematopoietic stem cells are able to fully reconstitute the blood system of lethally irradiated adult mice and small numbers of pluripotent stem cells are able to rescue development of genetically compromised embryos [4]). Thus, it is still unclear how population-level and cell-intrinsic regulatory programs interact to control mammalian stem and progenitor cell dynamics.

Here we propose a theoretical framework that reconciles these disparate observations, which views cellular multipotency as an instance of maximum entropy statistical inference. In this view, individual cells satisfy any minimal regulatory constraints imposed upon them (such as basic metabolic requirements, etc.) yet, in the absence of defined instructions, are maximally noncommittal with respect to their remaining molecular identity, thereby generating a diverse population that is able to respond optimally to a range of unforeseen future environmental changes. Thus, rather than viewing the multipotent cell state as an attractor of the underlying molecular regulatory dynamics (i.e. associating cellular identities with well-defined, stable, patterns of gene expression – a common modeling assumption, that has received some experimental validation for differentiated cell types [5]), individual multipotent cells are characterized by fundamental *uncertainty* in their molecular state and their populations exhibit variability in accordance with this intrinsic uncertainty. However, since this model exchanges the attractor hypothesis at the single cell level for an ergodicity assumption for the underlying stochastic processes, each individual cell has the latent potential to assume every identity within the population, and thereby retains the regenerative capacity of the entire population. As this view is fundamentally stochastic, its corollary is that regulation of multipotency occurs at the level of probabilities (i.e. at the population level), rather than at the individual cell level.

In order to illustrate this perspective we consider here the expression dynamics of the stem cell surface marker Sca1 (stem cell antigen 1) in populations of multipotent EML mouse hematopoietic progenitor cells. It has previously been shown that Sca1 levels fluctuate stochastically in EML cells in culture, with extrinsic 'transcriptome-wide' noise driving transitions between Sca1 high and Sca1 low states, which transiently prime individual cells for erythroid and myeloid differentiation respectively and generate a characteristically bimodal Sca1 expression distribution within the population (see Fig. 1, bottom panel and Ref. [6]). However, the underlying mechanisms by which these stochastic fluctuations are regulated are not known. In the absence of this knowledge we assume here that the intracellular dynamics of the Sca1 expression

level $z(t)$ are described by a generic stochastic differential equation:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = a(z) + \sqrt{2d(z)}\xi(t),$$

where $\xi(t)$ is a standard one-dimensional white noise process $[\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(s)\rangle = \delta(t-s)]$ and $d(z)$ accounts for fluctuations in Sca1 levels due to both intrinsic sources (i.e. noise in the molecular processes involved in Sca1 production/decay, such as transcription, translation, translocation and degradation etc.) and extrinsic sources (i.e. fluctuations in upstream regulators and uncontrolled environmental noise). Rather than model Sca1 levels directly it is convenient to introduce a reaction coordinate $x(z)$ such that the Fokker-Planck equation for the probability density $\rho(x,t)$ has the form

$$\frac{\partial\rho}{\partial t} = L(\rho), \quad L(\rho) = \frac{\partial}{\partial x}\left(\frac{\mathrm{d}\psi}{\mathrm{d}x}\rho\right) + \sigma\frac{\partial^2\rho}{\partial x^2}, \qquad (1)$$

with scalar potential $\psi(x)$ and diffusion coefficient σ. Such a transformation, which maps the original dynamics to those of a Brownian particle in a one-dimensional potential field, may be achieved by application of Itō's lemma (see Supplemental Material for details). The stationary solution of Eq. (1) is the Boltzmann-Gibbs distribution

$$\rho_\infty(x) = Z^{-1}\exp(-\psi/\sigma), \quad Z = \int \exp(-\psi/\sigma)\,\mathrm{d}x. \quad (2)$$

This solution exists so long as $\psi(x)$ grows sufficiently rapidly as $|x| \to \infty$ that the partition function Z remains finite. In this case, the dynamics are ergodic and the free energy

$$\begin{aligned} F(\rho) &= \int \psi\rho\,\mathrm{d}x + \sigma\int \rho\log\rho\,\mathrm{d}x, \\ &= E(\rho) - \sigma S(\rho), \end{aligned}$$

where $E(\rho)$ and $S(\rho)$ are the energy and entropy functionals respectively, is a Lyapunov functional for the dynamics. Thus, in order to model Sca1 dynamics phenomenologically we need only chose an appropriate reaction coordinate x and form for the potential $\psi(x)$.

Since noise in protein expression often scales with abundance, a natural choice for the reaction coordinate is $x = \log z$, as has been taken elsewhere (see Supplemental Material for details) [1]. In the absence of detailed information on how Sca1 fluctuations are regulated, the potential $\psi(x)$ may be estimated numerically from the empirical Sca1 distribution by inverting Eq. (2). The model then has a single free parameter, the diffusion coefficient σ, which sets the timescale for the dynamics.

Estimates of σ and $\psi(x)$ were obtained by model fitting using maximum likelihood estimation to evolving Sca1 expression distributions obtained experimentally using flow cytometry starting from pre-selected populations of Sca1 low, mid, and high expressing cells as they equilibrate in culture over a period of 18 days (obtained in Ref. [6]). Despite the simplicity of this model, an excellent agreement with the experimental time-series data was observed from all three initial conditions, using the same numerically estimated potential and the same estimate of σ (Figs. 1–2).

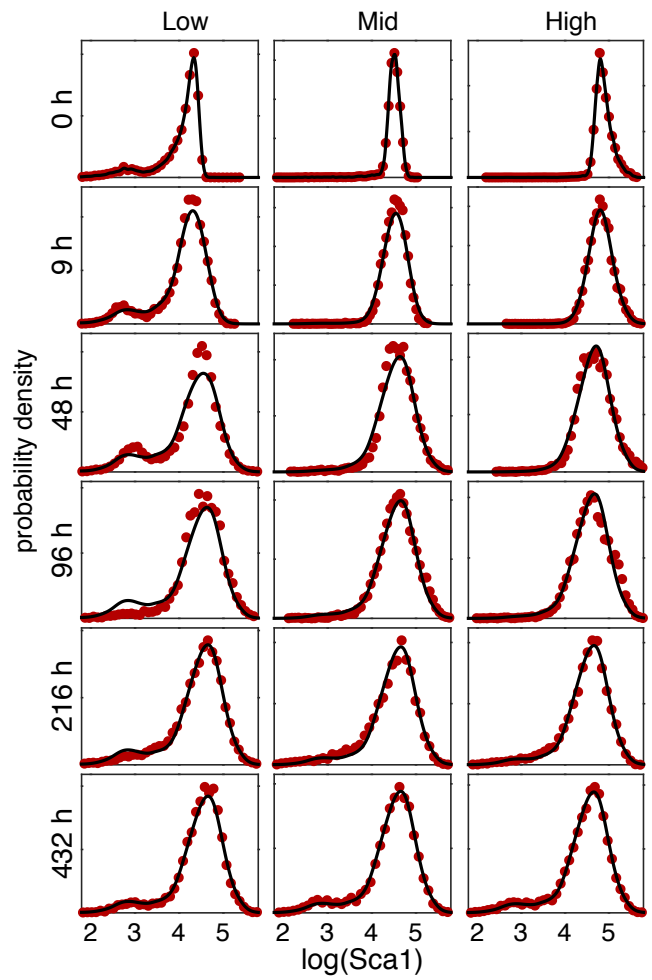

Figure 1. Model fit to experimental data. Model simulations using the same estimates of $\psi(x)$ and σ are shown against the three independent experimental time-series; simulations differ only in the experimentally prescribed initial conditions. Data is in dark red and the fitted model is in black. The potential $\psi(x)$ was estimated numerically via Eq. (2) using aggregated data from the final time point. Color online.

It has previously been argued, based upon analysis of changing proportions of cells in the Sca1 high and low states, that the observed dynamics are characterized by slow 'sigmoidal' relaxation towards the stationary state [6]. Since a constant probability flux across a barrier naturally leads to exponential relaxation, it was suggested that these dynamics indicate deviation from expected first-order kinetics, possibly due to regulation of Sca1 fluctuations by cell-cell communication or autocrine sig-

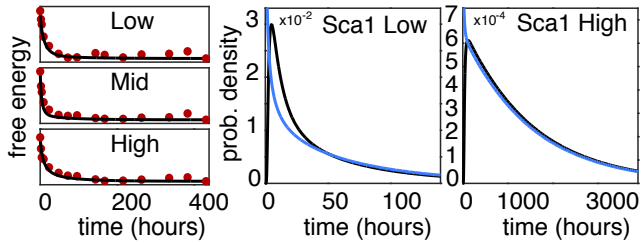

Figure 2. (*Left*) Convergence to equilibrium with respect to the free energy. Exponential convergence was observed from all three initial conditions for large time, in accordance with Eq. (1). First passage time (FPT) distributions in the Sca1 low (*Middle*) and high (*Right*) states. The FPT distributions $F_X(x_X, t)$ starting at the local minima of the potential $\psi(x)$ are shown in black; the expected FPT distributions $\langle F_X \rangle(t)$ averaging over all initial conditions in $X \in \{L, H\}$ are shown in blue. Color online.

naling. However, it is apparent that such recourse is not needed since in all 3 cases the experimental system is initially far from equilibrium, and therefore far from the regime in which first-order kinetics apply. Rather, in accordance with standard reaction-rate theory, the dynamics are characterized by an initial transient period during which local equilibrium is first established within each potential well, before transitions between wells occur [8]. Examination of the free energy (which is a natural way to assess convergence to equilibrium [9]) shows that this separation of timescales naturally generates the observed convergence dynamics without the need to include additional regulatory mechanisms in the model (see Fig. 2, left). These results indicate that the observed Sca1 expression dynamics are well described by a simple ergodic process in which individual cells behave independently with respect to Sca1 fluctuations.

This ergodic property is useful since it allows inference of the behavior of individual cells from the population dynamics. While stochastic excursions into the Sca1 high and low states have previously been seen to transiently confer different lineage biases to individual progenitor cells in culture, the timescales upon which these excursions occur at the single cell level are not known. Thus, the distribution of first passage times (FPTs) out of the Sca1 low and high states are of particular interest. Defining the ranges of Sca1 low and high expression as $L = (-\infty, x_0)$ and $H = (x_0, \infty)$ respectively, where x_0 is the intermediate maxima in $\psi(x)$, the FPT $T(x)$ out of X for a cell initially at $x \in X$ (where $X \in \{L, H\}$) may be obtained from the backward Fokker-Planck equation associated with Eq. (1). Denoting $G(x,t) = P(T(x) \geq t)$ we solve:

$$\frac{\partial G}{\partial t} = -\frac{\mathrm{d}\psi}{\mathrm{d}x}\frac{\partial G}{\partial x} + \sigma \frac{\partial^2 G}{\partial x^2},$$

with initial conditions $G(x, 0) = 1$ for $x \in X$ and boundary conditions $G(x_0, t) = \partial G/\partial x(\pm\infty, t) = 0$, from which the FPT distributions $F_X(x,t) = -\partial G/\partial t$ for $X \in \{L, H\}$ may be obtained. Conventionally, the FPT distribution $F_X(x,t)$ is evaluated from the local minima x_X of $\psi(x)$ in X, since this is the state of highest probability. Alternatively, we can weight each initial position within X according to the probability that the cell is at this position at equilibrium. We thus define the expected FPT distribution with respect to the Gibbs measure,

$$\langle F_X \rangle(t) = \int_{x \in X} \frac{\rho_\infty(x)}{w_X} F_X(x,t)\,\mathrm{d}x,$$

where $w_X = \int_{x \in X} \rho_\infty(x)\mathrm{d}x \in [0,1]$ is the weight of the population in X. Numerical approximations to $F_X(x_X, t)$ and $\langle F_X \rangle(t)$ are shown in Fig. 2. These distributions yield mean FPTs of 60/56 hours for the low state and 1573/1487 hours for the high state using $F_X(x_X, t)$ and $\langle F_X \rangle(t)$ respectively. These timescales are substantially longer than the EML cell cycle time (approx. $10-14$ hours [10]), and therefore suggest that Sca1 fluctuations are not simply a consequence of the cell-cycle. Rather, by setting the expected length of time that a pair of cells initially at the same position (e.g. daughter cells from the same cell division) will forget their common origin – and therefore the expected length of time that their identities will be coupled – Sca1 switching appears to encode an elementary form of epigenetic memory that endows individual cells with a transient functional identity. Since the rate of switching is slower than the rate of cell division this allows the formation of communities of cells that maintain the same characteristics though divisions, and are therefore able to adopt a temporarily stable functional phenotype. Yet, by allowing mixing between the communities on a feasible time-scale, Sca1 fluctuations also safeguard long-term cell-cell variability and ensure that a robustly heterogeneous population, able to rapidly respond to a range of environmental challenges and resilient to removal of cellular sub-types, is maintained.

These results indicate that regulated fluctuations in Sca1 levels may be an intrinsic feature of EML cells in culture since they provide a mechanism by which the population hedges against unforeseen future environmental challenges and thereby retains the capacity to differentiate along both erythroid and/or myeloid lineages as required. If this is the case, then it is natural to ask if the experimentally observed stationary Sca1 distribution is optimal for this purpose; that is, if it is *maximally* variable in some appropriately defined way. To investigate this, it is convenient to introduce a parameterization of the potential $\psi(x)$, in order to compare distributions. A parsimonious model, which allows for observed bimodality without introducing large numbers of parameters, is:

$$\frac{\mathrm{d}\psi}{\mathrm{d}x} = \beta x - \alpha_0 - \frac{\alpha_1 x^n}{K^n + x^n},$$

where n is a positive even integer [11]. Intuitively, this is a simple model of a positive-feedback based bistable

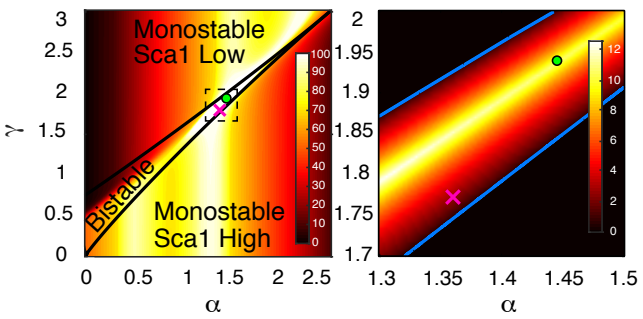

Figure 3. (*Left*) Entropy of the stationary distribution relative to the maximum entropy distribution over the $\alpha\gamma$-plane. The empirical distribution is marked with a magenta cross and the maximum entropy distribution $\rho_\infty^{\max}(x)$ is marked with a green circle. Color shows percentiles. (*Right*) Minimum MFPT τ in the vicinity of the maximum entropy distribution (close-up over the dashed box in the left panel). The critical lines separating the bistable and monostable regimes are shown in blue. The empirical distribution lies in the small region of the $\alpha\gamma$-plane that is both close to critical and of high entropy. Color shows dimensionless time. Color online.

switch of the kind that commonly regulate cell fate changes [12]. The stationary distribution $p_\infty(x)$ is then characterized by four nonnegative dimensionless parameters: $\boldsymbol{\theta} = [n,\, \alpha = \alpha_0/\alpha_1,\, \gamma = \beta K/\alpha_1,\, \sigma_{\mathrm{d}} = \sigma\beta/\alpha_1^2]$.

For fixed $\boldsymbol{\theta}$, the conditional probability $\rho_\infty(x\,|\,\boldsymbol{\theta})$ is the minimizer of the free energy $F(\rho)$, and may therefore be viewed as the most non-committal way to assign probabilities subject to the particular constrains imposed upon the dynamics by $\psi(x;\boldsymbol{\theta})$ (i.e. an instance of maximum entropy statistical inference) [13]. As each set of model parameters defines a different potential, which places different constraints upon the dynamics, we may therefore determine the extent to which Sca1 fluctuations optimize population diversity by assessing the proximity of the empirical stationary Sca1 distribution to the maximum entropy distribution $\rho_\infty^{\max}(x) = \rho_\infty(x\,|\,\boldsymbol{\theta}^*)$, where $S(\rho_\infty(x\,|\,\boldsymbol{\theta}^*)) = \max_{\boldsymbol{\theta}} S(\rho_\infty(x\,|\,\boldsymbol{\theta}))$. The relative entropy,

$$D(\rho_\infty \,||\, \rho_\infty^{\max}) = \int \rho_\infty \log\left(\frac{\rho_\infty}{\rho_\infty^{\max}}\right) \mathrm{d}x,$$

is a natural way to measure this proximity. Since the Hill coefficient n is, informally, a measure of the sensitivity of the underlying switch to input stimulus, it primarily affects the curvature of the potential around the local minima x_0 (where present) and does not have a strong effect on the entropy. However, by governing a cusp bifurcation that determines whether the underlying switch is in a monostable or bistable state, α and γ can affect the entropy of the stationary distribution considerably. Fig. 3 shows how the relative entropy of $p_\infty(x)$ varies over the biologically relevant bistable region of the $\alpha\gamma$-plane [14]. It can be seen that the point estimate for the experimentally observed Sca1 distribution is remarkably close to the maximum entropy distribution $\rho_\infty^{\max}(x)$. However, while the maximum entropy distribution is in the center of the bistable regime, the empirical distribution is close to one of the critical lines that separate the bistable and monostable regimes (shown in blue in Fig. 3, right). It has long been suggested that such criticality may emerge naturally in biological systems via self-organizing evolutionary processes without the need for fine-tuning (i.e. as an attractor of the evolutionary dynamics) since critical states provide the dual benefits of stability and adaptability [15]. Here, proximity to criticality specifically regulates the rate of mixing between the Sca1 high and low subpopulations, and therefore the response time of the population to environmental changes. To illustrate this, Fig. 3 also shows how $\tau = \min[\tau_-, \tau_+]$, where τ_- and τ_+ are the mean first passage times (MFPTs) in the low and high states respectively, varies in the vicinity of the maximum entropy state in the $\alpha\gamma$-plane. It can be seen that the minimum MFPT in the maximum entropy state is approximately an order of magnitude greater than that of the empirical distribution. Thus, while a population distributed according to the maximum entropy distribution would ultimately able to adapt better to environmental changes than the empirical population, it could not do so as rapidly. In this regard, close proximity to criticality is vital since it ensures that a diverse population is produced, yet mixing between subpopulations occurs on a physically relevant time-scale. These results suggest that Sca1 levels are regulated by fitness constraints that involve a trade-off between maximizing cell-cell variability and maintaining the ability to respond rapidly to environmental changes.

In summary, we have proposed an information-theoretic interpretation of stem cell dynamics that views cellular multipotency as an instance of maximum entropy statistical inference. Although we have focused on Sca1 dynamics, comperable expression fluctuations are known to generate functional diversity in other mammalian stem cell systems [2, 16], and similar ergodic dynamics have been observed to give rise to universal protein expression distributions in microorganisms [17]. Thus, the generation of ergodic expression fluctuations may be a generic way in which cell populations maintain robust multilineage differentiation potential under environmental uncertainty. If so, then molecular noise processing could be particularly important in regulating stem cell function in a range of contexts. A better understanding of the relationship between molecular noise and stem cell identity should help to distinguish variability due to interchangeable subpopulations of cells from that due to the presence of distinct, non-interconvertible, cell types (i.e. to determine which underlying stochastic processes are ergodic) [18]. We anticipate that advances in single cell profiling techniques will help to address these issues in the near future.

Acknowledgements. The authors would like

to thank Sui Huang for providing the experimental data. This work is partly funded by BBSRC grant BB/L000512/1.

* Correspondence to bdm@soton.ac.uk

[1] J. W. Veening, E. J. Stewart, T. W. Berngruber, F. Taddei, O. P. Kuipers, and L. W. Hamoen, Proc. Natl. Acad. Sci. USA **105**, 4393 (2008); J. W. Veening, W. K. Smits, and O. P. Kuipers, Annu. Rev. Microbiol. **62**, 193 (2008).

[2] T. Graf and M. Stadtfeld, Cell Stem Cell **3**, 480 (2008); P. Cahan and G. Q. Daley, Nat. Rev. Mol. Cell. Biol. **14**, 357 (2013); T. Kalmar, C. Lim, P. Hayward, S. Muñoz-Descalzo, J. Nichols, J. Garcia-Ojalvo, and A. M. Arias, PLoS Biol. **7**, e1000149 (2009).

[3] J. Lei, S. Levin, and Q. Nie, Proc. Natl. Acad. Sci. USA **111**, E880 (2014); B. MacArthur and I. R. Lemischka, Cell **154**, 484 (2013); J. Garcia-Ojalvo and A. M. Arias, Curr. Opin. Genet. Dev. **22**, 619 (2012).

[4] M. Osawa, K. I. Hanada, H. Hamada, and H. Nakauchi, Science **273**, 242 (1996); Z. Wang and R. Jaenisch, Dev. Biol. **275**, 192 (2004).

[5] S. Huang, G. Eichler, Y. Bar-Yam, and D. Ingber, Phys. Rev. Lett. **94**, 128701 (2005).

[6] H. Chang, M. Hemberg, M. Barahona, D. Ingber, and S. Huang, Nature **453**, 544 (2008).

[1] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, O. E., Y. Pilpel, and N. Barkai, Nat. Genet. **38**, 636 (2006); D. R. Sisan, M. Halter, J. B. Hubbard, and A. L. Plant, Proc. Natl. Acad. Sci. USA **109**, 19262 (2012).

[8] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).

[9] P. A. Markowich and C. Villani, Mat. Contemp. **19**, 1 (2000); R. Jordan, D. Kinderlehrer, and F. Otto, SIAM J. Math. Anal. **29**, 1 (1998).

[10] M. B. Bowie, K. D. McKnight, D. G. Kent, L. McCaffrey, P. A. Hoodless, C. J. Eaves, *et al.*, J. Clin. Invest. **116**, 2808 (2006); J. M. Nygren, D. Bryder, and S. E. W. Jacobsen, J. Immunol. **177**, 201 (2006).

[11] This restriction ensures that $\psi(x)$ is continuous and real for all $x \in \mathbb{R}$. Although in principle x may be negative, Sca1 levels are sufficiently high that we did not observe negative values in practice.

[12] A. Becskei, B. Séraphin, and L. Serrano, EMBO J. **20**, 2528 (2001); W. Xiong and J. E. Ferrell, Nature **426**, 460 (2003); J. E. Ferrell, Curr. Opin. Cell. Biol. **14**, 140 (2002).

[13] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).

[14] Note that $p_\infty(x)$ also depends upon σ_d, the relative strength of stochastic fluctuations. However, since this parameter includes the effects of unregulated extrinsic noise, we assume that it is not within the cells capacity to regulate and fix it at the experimentally determined value.

[15] M. Pal, S. Ghosh, and I. Bose, Phys. Biol. **12**, 016001 (2015); M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, Proc. Natl. Acad. Sci. USA **105**, 1897 (2008); S. A. Kauffman, *The origins of order: self-organization and selection in evolution* (Oxford University Press, 1993).

[16] I. Chambers, J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold, and A. Smith, Nature **450**, 1230 (2007); K. Hayashi, S. M. C. de Sousa Lopes, F. Tang, and M. A. Surani, Cell Stem Cell **3**, 391 (2008); Y. Toyooka, D. Shimosato, K. Murakami, K. Takahashi, and H. Niwa, Development **135**, 909 (2008); T. Kobayashi, H. Mizuno, I. Imayoshi, C. Furusawa, K. Shirahige, and R. Kageyama, Genes Dev. **23**, 1870 (2009); M. A. Canham, A. A. Sharov, M. Ko, and J. M. Brickman, PLoS Biol. **8**, e1000379 (2010); T. S. Macfarlan, W. D. Gifford, S. Driscoll, K. Lettieri, H. M. Rowe, D. Bonanomi, A. Firth, O. Singer, D. Trono, and S. L. Pfaff, Nature **487**, 57 (2012); J. Trott, K. Hayashi, A. Surani, M. M. Babu, and A. Martinez-Arias, Mol. Biosyst. **8**, 744 (2012).

[17] H. Salman, N. Brenner, C.-k. Tung, N. Elyahu, E. Stolovicki, L. Moore, A. Libchaber, and E. Braun, Phys. Rev. Lett. **108**, 238105 (2012); N. Brenner, E. Braun, J. S. Rotella, and H. Salman, arXiv:1503.01046 (2015).

[18] C. Pina, C. Fugazza, A. J. Tipping, J. Brown, S. Soneji, J. Teles, C. Peterson, and T. Enver, Nat. Cell. Biol. **14**, 287 (2012); P. Rué and A. Martinez Arias, Mol. Syst. Biol. **11** (2015).

Supplementary Material

We assume that the intracellular dynamics of Sca1 expression level $z(t)$ are given by the following generic stochastic differential equation:

$$\frac{\mathrm{d}z}{\mathrm{d}t} = a(z) + \sqrt{2d(z)}\xi(t),$$

where $\xi(t)$ is a standard one-dimensional white noise process $[\langle\xi(t)\rangle = 0$ and $\langle\xi(t)\xi(s)\rangle = \delta(t-s)]$ and $d(z)$ accounts for fluctuations in Sca1 levels due to both intrinsic and extrinsic sources. This equation can be written in a more convenient form by introducing an appropriate reaction coordinate $x(z)$ such that the dynamics are mapped to those of a Brownian particle in a one-dimensional potential field. Such a transformation may be achieved by application of Itō's lemma, which reads:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = a(z)\frac{\mathrm{d}x}{\mathrm{d}z} + d(z)\frac{\mathrm{d}^2x}{\mathrm{d}z^2} + \sqrt{2d(z)}\frac{\mathrm{d}x}{\mathrm{d}z}\xi(t).$$

Firstly, the reaction coordinate $x(z)$ can be chosen such that the noise term in this equation is constant, say $\sqrt{2\sigma}$, which gives the transformation

$$x = \int \sqrt{\frac{\sigma}{d(z)}}\mathrm{d}z. \tag{S3}$$

Since the dynamics are one-dimensional we may also introduce a potential $\psi(x)$ such that

$$-\frac{\mathrm{d}\psi}{\mathrm{d}x} = a(z)\frac{\mathrm{d}x}{\mathrm{d}z} + d(z)\frac{\mathrm{d}^2x}{\mathrm{d}z^2},$$

to obtain

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\mathrm{d}\psi}{\mathrm{d}x} + \sqrt{2\sigma}\xi(t),$$

which is the stochastic differential equation corresponding to the Fokker-Planck equation given in the main text. Experimental data suggests that protein expression fluctuations often scale linearly with expression level [1]. Thus, a natural choice for the noise term is $d(z) = \sigma z^2$. Substituting this into Eq. (S3) gives $x = \log(z)$. This approach is similar to that taken in Ref. [1].

* Correspondence to bdm@soton.ac.uk

[1] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, O. E., Y. Pilpel, and N. Barkai, Nat. Genet. **38**, 636 (2006); D. R. Sisan, M. Halter, J. B. Hubbard, and A. L. Plant, Proc. Natl. Acad. Sci. USA **109**, 19262 (2012).