Article

Stochastic Sensitivity Analysis and Kernel Inference via Distributional Data

Bochong Li[1] and Lingchong You[1,2,*]
[1]Department of Biomedical Engineering and [2]Center for Systems Biology, Duke University, Durham, North Carolina

ABSTRACT Cellular processes are noisy due to the stochastic nature of biochemical reactions. As such, it is impossible to predict the exact quantity of a molecule or other attributes at the single-cell level. However, the distribution of a molecule over a population is often deterministic and is governed by the underlying regulatory networks relevant to the cellular functionality of interest. Recent studies have started to exploit this property to infer network states. To facilitate the analysis of distributional data in a general experimental setting, we introduce a computational framework to efficiently characterize the sensitivity of distributional output to changes in external stimuli. Further, we establish a probability-divergence-based kernel regression model to accurately infer signal level based on distribution measurements. Our methodology is applicable to any biological system subject to stochastic dynamics and can be used to elucidate how population-based information processing may contribute to organism-level functionality. It also lays the foundation for engineering synthetic biological systems that exploit population decoding to more robustly perform various biocomputation tasks, such as disease diagnostics and environmental-pollutant sensing.

INTRODUCTION

As information-processing units, cellular networks transform diverse stimuli, such as DNA damage and pathogenic infection, to appropriate responses (2). In each cell, this process can be highly noisy due to the discrete nature of biochemical reactions (3–10). As a result, single-cell responses are often highly heterogeneous even in an isogenic population (11–14). Phenotypic heterogeneity often manifests as differential fate determination, including cell growth, senescence, and death, which can be triggered by both natural (e.g., growth factors) and artificial (e.g., therapeutics) stimuli (1,15). Functional implications of cell-to-cell variability have been established in various biological contexts (4,5,7,9,16–22). Historically, biological models (e.g., network structure) have been conceived based on cell-population average measurements (e.g., Western blot) (23). Given the large degree of phenotypic heterogeneity, however, biologically relevant information may be lost in the process of averaging (11,12,24). For example, it was thought formerly that p53 underwent damped oscillation in response to DNA damage. However, single-cell experiments showed that individual cells give rise to varying numbers of p53 pulses of fixed amplitude and duration (25). In another case, single-cell analysis revealed a biphasic dependence of E2F on Myc expression, shedding new light on cell-cycle regulation (13).

The presence of noise may fundamentally limit the information-processing capacity at the single-cell level (26). However, some tissue- and organism-level responses rely on the constructive use of noise (27–31). One example is the generation of robust acute and recall immunity in response to infection via the diversification of individual naïve antigen-specific T cells (30). Another example is the decoding of motion components via the firing-rate distribution over the population of noisy neurons in the middle temporal visual area (30,33). The key rationale behind noise-aided information processing is statistical regularity. The shape of the distribution is governed by network properties, including structure and parameterization, and can be captured by observing a sufficiently large sample population (28,34–37). This population-level determinism justifies the use of distribution data as a quantitative phenotype of a signaling network and its corresponding cellular output (e.g., cell proliferation). The implications of this perspective are twofold. First, distribution data can be used to gain insight into and constrain the regulatory property of the underlying network model. Second, it provides the basis for deciphering information-processing mechanisms in naturally occurring cell populations and for engineering synthetic gene circuits that exploit stochastic dynamics and population codes. These applications will advance our understanding in basic biological principles and clinical practices such as disease diagnostics (36,38,39).

Here, we present a streamlined computational framework to quantify stochastic network sensitivity to parameter perturbations, using distribution as the readout. We further implement a kernel regression model with a probability-distance measure that enables accurate inference of external stimuli. Although our framework is established using a mechanistic model of a well-defined biological network, it is entirely data-driven and does not require a priori mechanistic knowledge. In addition, we use perturbations on

Submitted April 4, 2014, and accepted for publication July 15, 2014.

*Correspondence: you@duke.edu

Editor: Rong Li.

0006-3495/14/09/1247/9 $2.00

http://dx.doi.org/10.1016/j.bpj.2014.07.025

network parameters as an emulation of realistic external stimuli, such as growth factors and environmental pollutants. This mechanism-free nature renders our method particularly suitable for complex biological phenomena like cancer, where a simple mechanistic model would not do the whole system justice and population heterogeneity is physiologically or pathologically relevant.

Various methods have been proposed to exploit stochastic dynamics in estimating model parameters and selecting optimal model structures (40–56). Our method expands the tool kit. It allows intuitive exploration of biological systems via characterization of their distributional responses to environmental stimuli. Of more importance, our method exploits the totality of the distribution, whereas past methods have primarily relied on low-order statistics and are limited by the analytical tractability of the mechanistic model.

METHODS

Stochastic simulation of the MYC/Rb/E2F model

We adopt a previously developed stochastic model for this network (57,58). It consists of a set of stochastic differential equations, which has the general form

$$\frac{dX_i(t)}{dt} = \sum_{j=1}^{M} v_{ji} a_j[X(t)] + \sum_{j=1}^{M} v_{ji} a_j^{\frac{1}{2}}[X(t)]\Gamma_j(t) + \omega_i(t), \quad (1)$$

where $X_i(t)$ represents the number of molecules of a molecular species i ($i = 1, \ldots, N$) at time t, and $X(t) = (X_1(t), \ldots, X_N(t))$ is the state of the entire system at time t. $X(t)$ evolves over time at the rate of $a_j[X(t)]$ ($j = 1, \ldots, M$), and the corresponding changes in the number of individual molecules are described in v_{ji}. $\Gamma_j(t)$ and $\omega_i(t)$ are temporally uncorrelated, statistically independent Gaussian noises. $\Gamma_j(t)$ is the standard normal distribution with mean 0 and variance 1. $\omega_i(t)$ tunes the level of empirical additive extrinsic noise (1). The stochastic differential equations are simulated using Matlab. The distributional output corresponding to a network state (parameterization) is generated by performing temporal simulations of the network 5000 times and taking the level of the nodal activity of interest (E2F) at a fixed time point (24 h). Therefore, the immediate raw distributional output is a histogram with 5000 samples.

Calculating modified Kullback-Leibler divergence and stochastic sensitivity

We fit each histogram, as depicted in the paragraph above, by a Gaussian mixture model (GMM) with 20 components to capture its totality and ease downstream computational analysis. A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. Suppose there are two such GMM density functions, f and g. We define a modified Kullback-Leibler (KL) divergence (D), a statistical distance, to quantify the difference between two distributions (i.e., f and g) (59). By its original definition, KL divergence is asymmetric: the KL divergence of g from f, $D_{KL}(f||g)$, is in general different from that of f from g, $D_{KL}(g||f)$. Our metric symmetrizes the divergence by taking the average of the two (Fig. S1). There is no analytical equation to calculate the KL divergence between two GMM density functions. To approximate the KL divergence, we draw n samples $\{x_i\}_{i=1}^{n}$ from f, from which we calculate $D_{KL}(f||g) = (1/n)\sum_{i=1}^{n}\log(f(x_i)/g(x_i))$. In a similar way, we can calculate $D_{KL}(g||f)$. Note that the samples here are from the fitted GMM density, not from the original stochastic simulations. For large n, which we set to be 10,000, this Monte Carlo sampling approximation converges to the true divergence between f and g (59) We define the symmetrical divergence measure, $D(f||g) = D_{KL}(f||g) + D_{KL}(g||f)/2$. This is also known as the Jensen-Shannon divergence and is routinely used in bioinformatics analysis (60,61).

We bootstrap the raw distribution to emulate the variability in the distributional response over multiple observations. The original distribution is sampled with replacement N times, with N equal to the number of samples in the original distribution. Fig. S1, *A* and *B*, shows the KL divergence calculated in reverse directions (*green* and *red lines*), with the black line being the average of the two, or the symmetric divergence, D. The statistics on intracondition variability generated by bootstrapped distributions are consistent with those generated by independent SDE simulations. We also reason that the sufficient sampling of the distribution generated by a single run of the simulation makes bootstrapping an appropriate procedure to represent the variability of the distribution under the same condition. Exemplary codes can be found at http://www.genome.duke.edu/labs/YouLab/software/index.php.

KL divergence can be calculated directly from the observed sample distribution using the empirical cumulative distribution function (cdf) without an intermediate density estimation step (62). However, the exploitation of an intermediate parametric GMM does convey convenience and computational efficiency in evaluating the probability of samples, which will be performed repeatedly during prediction.

RESULTS

The basic computational framework

To generate simulated data, we employ a well-established stochastic model of the Myc/Rb/E2f network (Fig. 1; see also Methods), which plays critical roles in regulating cell-cycle progression and cell-fate decisions (22). The model consists of stochastic differential equations accounting for both intrinsic and extrinsic variability associated with the network dynamics (63). Conventional computational tools for analyzing distributions are often developed using data generated by relatively simple stochastic models, which are more amenable to analytically tractable approaches (46,52,64). Natural biological systems, however, are more complex and may generate less regular distributions, which demand a different computational treatment.

We carry out 5000 rounds of independent simulations to generate a distribution of network output (E2F protein) at a fixed time point. Experimentally, such a distribution can be obtained via flow cytometry. We perform bootstrapping to characterize intrastate variability due to finite sample size and experimental noise. Parameter perturbation is carried out over three orders of magnitude on a log scale centered around the base value (63), with the other parameters held constant. This perturbation range is divided into 1000 equal units on the log scale with unit width Δs, the smallest perturbation step in our study.

Sensitivity analysis using distributional data

Fig. 2 *A* shows representative distribution responses to parameter perturbations. The stochastic sensitivity at a

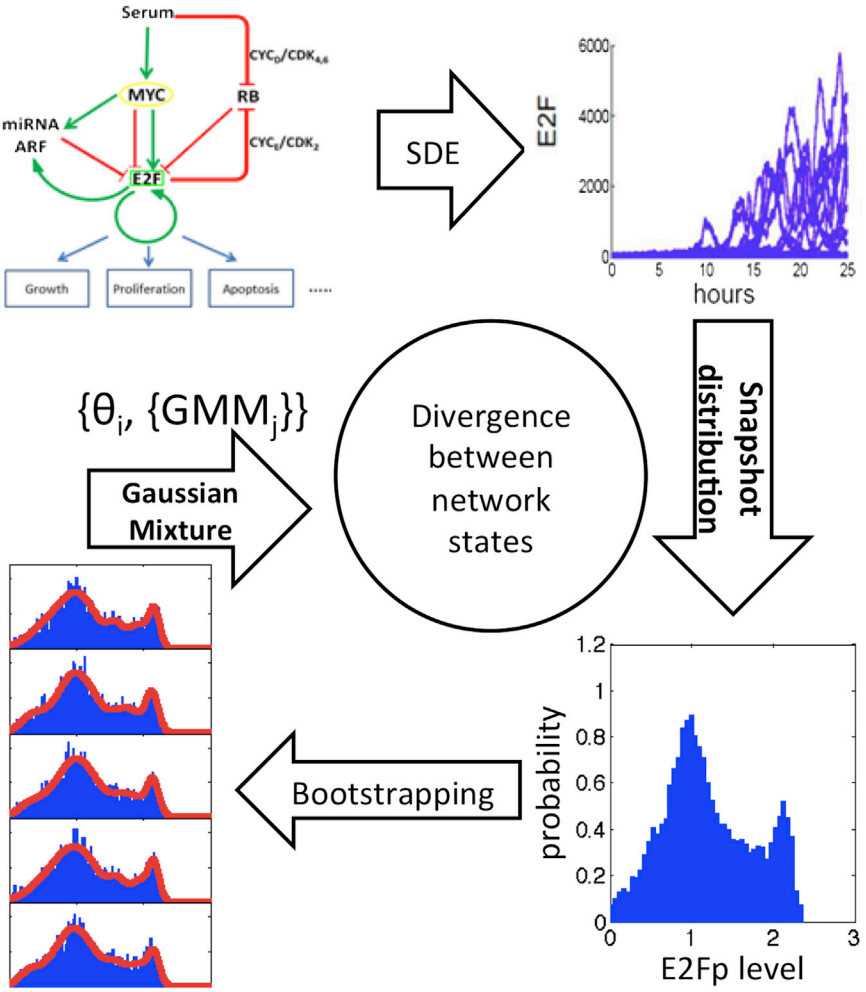

FIGURE 1 Summary of our computational framework. We use a stochastic model of the Myc/Rb/E2F network dynamics to generate distributional data. For each parameter set, we generate a distribution of E2F protein levels at 24 h from 5000 simulations. The distribution is bootstrapped with replacements to generate 30 replicates to represent potential variability in the distributional output for the same condition. Each distribution is fitted with a GMM with 20 components, from which pairwise KL divergence between distributions within the same condition and between different conditions is calculated. To see this figure in color, go online.

parameter value is defined as the divergence between distributions corresponding to that parameter value and one perturbation step apart (in either direction defined by the user). The goal is to quantify the change in distribution shape with respect to a unit perturbation. Using sets of distributional data similar to those shown in Fig. 2 *A*, but with a much smaller perturbation step (Δs), we obtain stochastic sensitivity curves over the entire perturbation range for various parameters (Fig. 2, *B* and *C*). It is critical that the interstate divergence is greater than intrastate variability. The sensitivity level (*black curves*) varies with parameter value.

To fairly compare the stochastic sensitivity in perturbing different parameters, we choose two ends of the perturbation spectrum for both k_{Rb} and k_{E2Fm} independently, such that the distributional responses (i.e., distribution shapes) of the networks at the endpoints match up as closely as possible (Fig. S2). One endpoint for both perturbation spectra is by default the basal state of the network. The other endpoint of the two spectra is determined by finding the minimum pairwise divergence between all network distributional outputs resulting from the perturbation of the two parameters. The perturbation range of k_{Rb} required for the network distribution to morph from P_{low} to P_{high} is much smaller than that of k_{E2Fm} (Fig. S2 *C*). For Rb synthesis rate, the sensitive region is restricted to a narrower domain compared to the E2F mRNA synthesis rate. There are also differences in the gross sensitivity of the network distributional output to the perturbation of different parameters (Fig. S3). For example, it is lower for k_{CD} perturbation (Fig. S3 *K*) than for k_{EFm} or k_{EFp} perturbation (Fig. S3, *C* and *O*).

The sensitivity of either the mean or the variance to parameter perturbation is lower than that of the whole distribution (Fig. 2, *D* and *E*). This suggests that relying solely on lower-order statistics fails to capture all the changes in distribution, incurring a loss of information. This is expected given the large deviation of the distributions from simple statistical models such as Gaussian or Poisson (Fig. 2 *A*). Note that the degree of discrepancy varies between parameters, being larger for k_{E2Fm} than for k_{Rb}. This suggests that for certain stimuli, it is critical to examine the totality of distributions.

Although our computational framework is defined using simulated data, it is applicable to any distributional data.

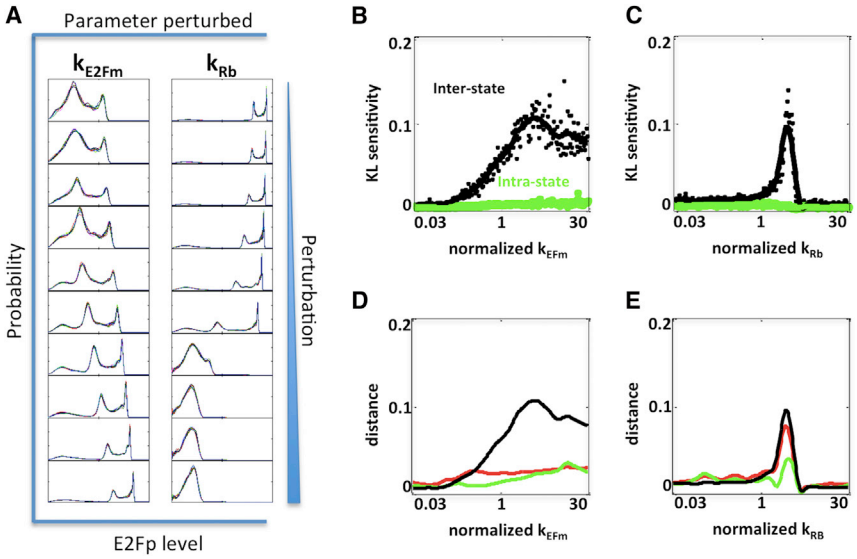

FIGURE 2 E2F distribution changes with differential sensitivity for perturbations to different parameters. (*A*) E2F distributions corresponding to 10 different perturbation levels for either E2F mRNA synthesis rate (k_{E2Fm}) or Rb synthesis rate (k_{Rb}). At each perturbation level, GMMs are fitted over bootstrapped distributions and 10 are plotted. The variation in the distributions within each condition (same network state) exists. However, the variation between perturbation levels is much larger for certain perturbation increments. The higher the divergence, the more drastic the network distributional output changes in response to an incremental change in the perturbation level. (*B*) The black dots are average divergence between all pairs of bootstrapped distributions generated by adjacent parameter values (interstate divergence); the green line represents the average divergence between all pairs of distributions generated by the same network state (i.e., intrastate variability). The error bar shows the standard error. The solid lines are smoothing curves fitted over the corresponding dots. (*C*) The inter- and intrastate divergence for E2F distributions resulting from different k_{Rb} values. (*D*) Comparison of divergence-based stochastic sensitivity and lower-order statistic sensitivities (*red*, mean; *green*, variance) to k_{E2Fm} perturbation. There are significant discrepancies in the functional form of the sensitivity curves. (*E*) Comparison of divergence-based stochastic sensitivity and lower-order statistic sensitivities (*red*, mean; *green*, variance) to k_{Rb} perturbation. To see this figure in color, go online.

To illustrate the versatility of our method, we analyze a previously collected data set on the distribution of E2F activation by serum stimulation (65). At different serum concentrations, the evolution of the E2F activity distribution as a function of time (here, time is the stimulus) is qualitatively different. Specifically, bimodal distribution emerges at low serum concentration (0.3%), whereas monomodal distribution persists at high serum concentration (5%), with the mean increasing in a graded manner (Fig. 3 *A*). Applying our sensitivity analysis allows the identification of sensitive time intervals during which more drastic distribution changes occur (Fig. 3, *B* and *C*). This information will facilitate the design of further experiments that would reveal more information regarding the functionality of the network or constrain mathematical models by zooming into the more sensitive regions. The increased sensitivity of KL divergence to parametric perturbations (compared to the mean and variance) allows more quantitative constraints when comparing results from modeling to single-cell experiments.

Inference of network state using distributional data

Here, we demonstrate how distributions can be used as landmarks to fingerprint populations with unique network states or those subjected to different external stimuli. Similar to the development of stochastic sensitivity analysis framework, we use parameter perturbation to represent both scenarios.

The foundation of our distributional fingerprinting method is an empirically established knowledge base in which each distribution, denoted as a fingerprint, corresponds to a known parameter value (Fig. 4 *A*). We ask whether the parameter corresponding to a query distribution can be inferred based on this knowledge base and using probability divergence as a distance measure. To this end, we combine divergence with a kernel method in a statistical framework and apply kernel regression. The kernel method allows interpolation between learnt network states (66). This becomes useful in real biological settings, given the large network-state space and the often limited ability to acquire a large enough data set for other algorithms, such as nearest-neighbor matching.

We define the kernel function to be $k(f,g) = e^{-\alpha D(f,g)}$. This kernel function is symmetric with respect to the two input probability functions. It converts the similarity (measured by D) between two distributions to a real number between zero and one. Specifically, it is equal to 1 if $f = g$ and diminishes to 0 as the dissimilarity between the two probability densities increases. Furthermore, the conversion is nonlinear so as to better represent the correlation between similarity in distributional output and proximity in network state.

With this kernel function, we apply weighted kernel regression to predict the parameter s corresponding to a query distribution using the equation $s = (\sum_n s_n k(q,f_n)/\sum_n k(q,f_n)) + \beta$, where $n \in \Omega$. Each reference fingerprint (f_n) corresponds to a known parameter (s_n). β is an offset parameter. To demonstrate the ability of this kernel to interpolate between reference fingerprints, we endow the kernel with 125 reference fingerprints that are $8\Delta s$ apart (Δs as defined previously); the kernel is then trained and used to

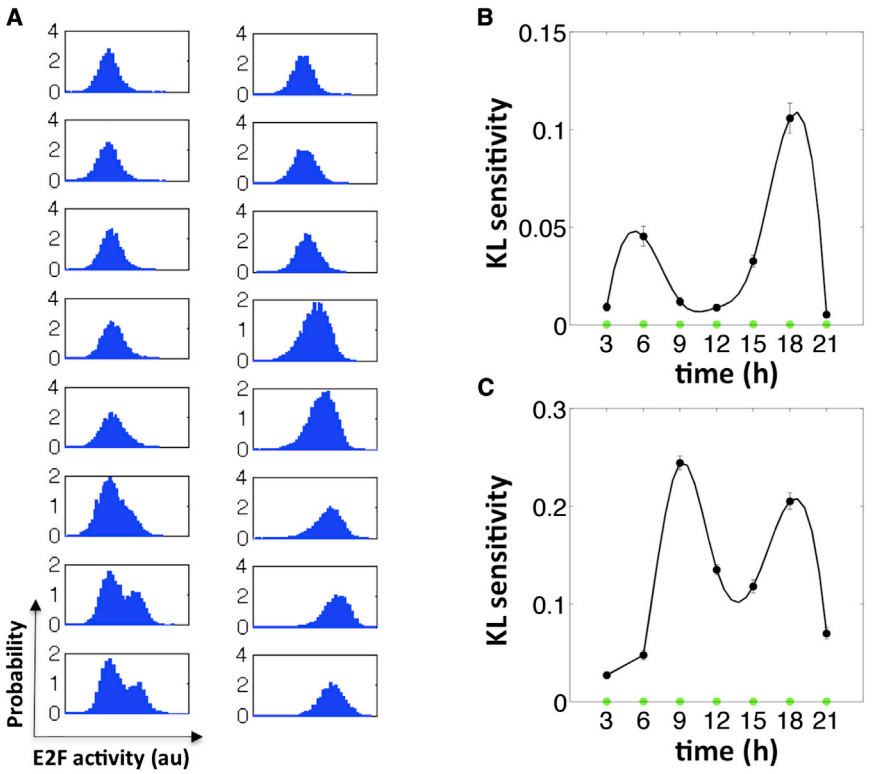

FIGURE 3 Stochastic sensitivity analysis of experimentally measured E2F activity distribution over time. (*A*) E2F activity distributions after serum stimulation at 0.2% (*left*) and 5% (*right*) were measured at consecutive time points (1). (*B*) The black dots represent the average divergence between all pairs of bootstrapped distributions measured at adjacent time points (interstate divergence) for low serum concentration; the green line represents the average divergence between all pairs of distributions measured at the same time point (i.e., intrastate variability). The error bar shows the standard error. The solid lines are smoothing curves fitted over the corresponding dots. (*C*) Same as in *B* for high serum concentration. To see this figure in color, go online.

infer all 1000 network states (Δs apart). There are seven network states that are not explicitly included in the kernel between two reference fingerprints. Using the training set, we optimize the values for α, β, and Ω associated with each reference fingerprint. These parameters determine the weighing of the fingerprints in the neighborhood of the signal. Specifically, for α, β, and Ω associated with the ith distributional fingerprint, we minimize an error function: $L_i = \sum_j \left| s_j - s_j^{ker} \right|$, $\forall j{>}0$, $j = (\Delta s'/\Delta s)\ (i + (m'/2))$, $m' \in \{0, \pm 1, \ldots, \pm m\}$, where $\Delta s' = 8\Delta s$ is the perturbation step size, and i is the index of the ordered perturbation input. We choose m to be 4. In other words, this algorithm optimizes the kernel parameters for the ith distributional fingerprint by minimizing the prediction error of the signal in the vicinity of the ith fingerprint, specifically within $2\Delta s'$ on either side of s_i. The optimized number of fingerprints ranges from ~5–15. There is an inverse correlation between the stochastic sensitivity and the optimal number of kernel fingerprints for a given network state.

To carry out the inference, an anchor reference fingerprint is first determined by finding among all (125) reference fingerprints the one with the smallest divergence from the query. The optimal parameters associated with this anchor fingerprint are then applied to the kernel function. The intuition behind distributional fingerprinting is that the more similar the query distribution is to a reference fingerprint, the closer the query network state is to that underlying the reference fingerprint (Fig. 4 *A*). Subsequently, including multiple reference fingerprints in the vicinity of the query distribution increases the resolution, analogous to triangulation of the physical location of a signal. On the other hand, including fingerprints corresponding to network states too disparate from the query offers little information yet incurs computational cost. As shown in Fig. 4, *B* and *D*, and Fig. S3, *A* and *E*, the predicted parameter values (*black dots*) lie closely on top of the red line, which indicates the true network states. The average deviation from true parameter values is 1.3 Δs for k_{E2Fm} and 12.8 Δs for k_{Rb}. Moreover, the prediction accuracy in the midrange of the perturbation spectrum (Fig. 4 *D*, *inset*) is much higher than at the two ends (Fig. S3 *E*) for k_{Rb}. This poorer performance is due to low stochastic sensitivity and a lack of balanced reference fingerprints on both sides of the query at either end of the spectrum.

To examine whether using distributions is superior to using lower-order moments in predicting network state, we apply the same weighted kernel regression using the Euclidean distance between the mean of distributions in constructing the kernel function. The kernel regression parameters are similarly optimized. The accuracy of mean fingerprinting is significantly lower than that of distributional fingerprinting, with an average deviation from true parameter values of 4.3 Δs for k_{E2Fm} and 18.1 Δs for k_{Rb} (Fig. 4, *C* and *E*, and Fig. S3, *B* and *F*).

A key requirement for accurate fingerprinting of the network state is for the interstate divergence (or Euclidean distance) to be significantly higher than intrastate divergence. This relative sensitivity is represented by the

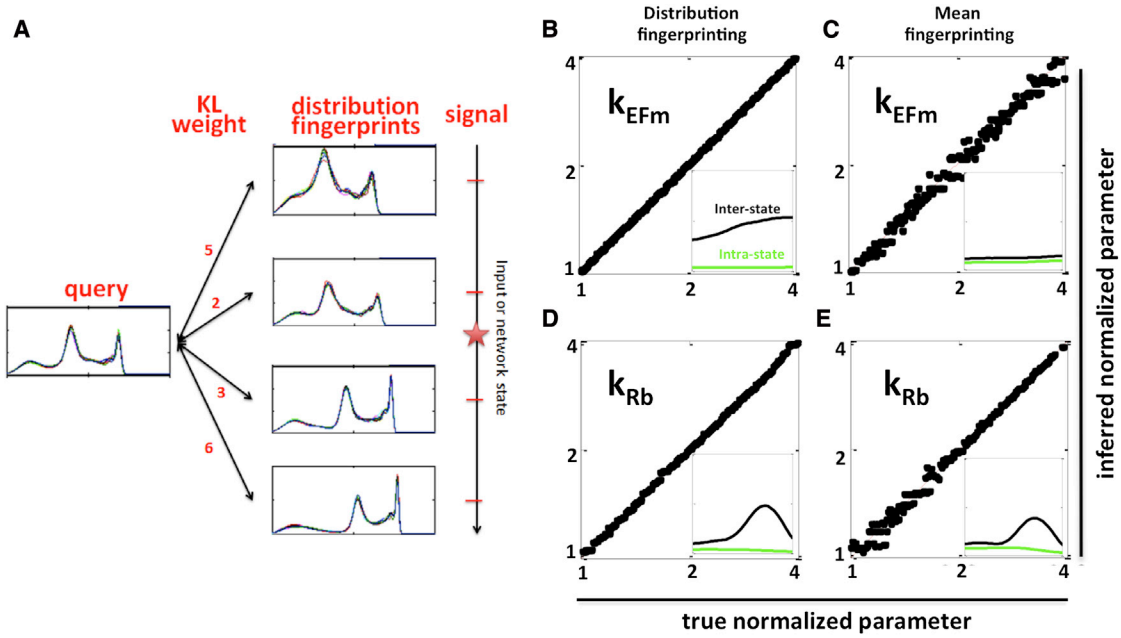

FIGURE 4 Distributional fingerprinting via divergence-based kernel regression is advantageous over mean fingerprinting via lower-order statistic-based regression using the same kernel. (*A*) Illustration of distributional fingerprinting. The more the query distribution resembles the fingerprint distribution, the closer the network state corresponding to the query lies to the state of that particular fingerprint. The inclusion of multiple fingerprints increases the accuracy of the kernel inference via training, which essentially instructs the optimal weighing of fingerprints. (*B*) Prediction of network state (k_{E2Fm}) using distributional fingerprints and divergence-based kernel regression. The predicted network states (*black dots*) lie closely on top of the true network state (*red line*). The distribution-based stochastic sensitivity (interstate divergence (*black*)) to k_{E2Fm} perturbation is much higher than the intrastate variability (*green*). The axes represent state indices, with corresponding adjacent states Δs apart. Only a fraction of all states are plotted here, to better show the difference between distributional fingerprinting and mean fingerprinting. The inferences on all states are shown in the Supporting Material. (*C*) Inference of k_{E2Fm} using only the mean and Euclidean distance-based kernel regression. The predicted network states (*black dots*) are scattered further away from the true network state (*diagonal line*), corresponding to lower inference accuracy. The mean sensitivity (interstate difference (*black*)) is much closer to the intrastate variability (*green*). (*D*) Inference of k_{Rb} using distributional fingerprints and divergence-based kernel regression. The predicted network states (*black dots*) lie closely on top of the true network state (*diagonal line*). (*E*) Inference of k_{Rb} using only the mean and Euclidean distance-based kernel regression. Larger separation between the mean sensitivity and the intrastate variability results in more accurate inference compared to the k_{E2Fm} perturbation in *C*. To see this figure in color, go online.

ratio between the sensitivity curve and the intrastate variability curve. As shown in Fig. 4, *B–E*, *inset*, and Fig. S3, *C*, *D*, *G*, and *H*, the relative sensitivity is much higher for the distribution than for the mean, explaining the better performance of the kernel via distributional fingerprinting. Even within each mode of fingerprinting, the accuracy correlates with the relative sensitivity, as in the case of k_{Rb}. This pattern also manifests on a more global level between different parameters. For example, the relative sensitivity for k_{CD} perturbation is low throughout the range of perturbation; consequently, the accuracy of inference is significantly lower than that of other parameters, yet still outperforms mean fingerprinting (Fig. S3, *I–L*).

Connection to Bayesian methodology

Given the popularity of conventional Bayesian methodologies, and to better facilitate the understanding of our proposed methodology, we briefly establish a connection between the two. To facilitate the argument, we'll use discrete distributions. Let $\boldsymbol{Q}$ be the type of histogram representing each observed distributional response. $\boldsymbol{Q}$ is expressed in terms of histograms $\{h_1, h_2, \ldots, h_k\}$, where h_k is the number of observations in the kth bin. Note that the binning would be the same between all distributional fingerprints. Let f_n be the distribution at each stimulus level, s_n, where n is the index of the stimuli. Then the probability of observing $\boldsymbol{Q}$ conditional on the stimulus level s_n can be expressed by a multinomial distribution:

$$p(\boldsymbol{Q}|s_n) = \prod_{k=1}^{L} f_n(k)^{h_k},$$

where f_n is the true distributional fingerprint corresponding to the nth stimulus. Let α be the total count of the observed histogram, $\boldsymbol{Q}$, and $q(k) = h_k/\alpha$ the normalized observed distribution. By definition of KL divergence and Shannon entropy, we have

$$-\frac{1}{\alpha}\log p(\boldsymbol{Q}|s_n) - H(q) = KL(q||f_n),$$

where $H(q) = -\sum_k q(k) \log q(k)$.

Then we can apply the Bayes rule, assuming uniform prior distribution on $p(s)$, to obtain the probability of stimulus s_n conditional on the observation $\boldsymbol{Q}$:

$$p(s_n|\boldsymbol{Q}) = \frac{p(\boldsymbol{Q}|s_n)p(s_n)}{p(\boldsymbol{Q})} \propto p(\boldsymbol{Q}|s_n).$$

With normalization by a partition function, and using the equations above, we can derive the expected stimulus,

$$\mathrm{E}[s_n] = \frac{\sum_n s_n e^{-\alpha[KL(q||f_n)+H(q)]}}{\sum_n e^{-\alpha[KL(q||f_n)+H(q)]}}.$$

Compared to our proposed kernel model, $s_n = (\sum_n s_n k(q,f_n)/\sum_n k(q,f_n)) + \beta$, the term derived using the Bayesian methodology is asymmetric. Such asymmetry incurs difficulty in applying various kernel methods such as the support vector machine. However, we also want to note the similarity between the two models and emphasize the similarity in functional form between the two. Our proposed kernel model does provide the flexibility to accommodate a wide range of machine learning algorithms, as well as efficiency in computation.

DISCUSSION

We present a streamlined framework to characterize distributional responses of a cell population. The distributional sensitivity analysis examines how the entire distribution changes in response to perturbation. The divergence-based kernel inference allows one to use a finite set of distributional fingerprints to infer the system state or external stimulus level, even when they have not been explicitly learnt during the training phase. Here, the distribution can be that of any quantifiable single-cell readout, such as surface markers and intracellular concentrations of signaling proteins.

Our method does not impose any simple statistical model structure (such as Gaussian or Poisson) onto the observed distribution. The rationale is that a significant amount of information is actually embedded in the higher-order moments of the distribution (i.e., the fine features of its shape), which cannot be adequately represented by conventional distribution models. Thus, we chose the GMM to capture as many features of the distribution as possible. However, such statistical rendering and overfitting makes the application of conventional Bayesian methods difficult and computationally expensive. For example, it is difficult to extract a distribution of statistical model parameters (e.g., mean and variance) to characterize the intracondition variability of total distribution (i.e., a distribution of distributions), as can be easily done for simple statistical models such as Gaussian and Poisson. Our kernel method offers one practical solution that simultaneously preserves the totality of the distribution and analyzes the data with a systematic and coherent machine-learning framework. The various practices used in establishing and characterizing our method, such as bootstrapping and measuring prediction error, are standard in the field of machine learning. To our knowledge, kernel regression using distribution divergence has not been demonstrated in studying nonneuronal biological systems.

For some biological systems, the behavior of an entire cell population is the sum of subpopulations with clearly defined functions, such as the immune system consisting of multiple cell types (e.g., different T cells and B cells), each with its own responsibilities. A common practice in analyzing population heterogeneity data, such as those obtained via flow cytometry, is to decompose the entire population into subpopulations. However, such decomposition may not always be possible, as in the case of motion perception. Neither is it always desirable, as doing so may hinder the discovery of novel subpopulations or incur a loss of information embedded in higher-order structures of the distribution (67). Under such circumstances, one should consider the possibility that the totality of the distribution is of informational value and biological significance.

In fact, population distribution has been implicated in multiple biological systems. One example is the distribution of cardiac clonal populations in the development of the heart and its correlation with the overall size and the shape of the heart (68). Large variation in the size of each clone has been observed, yet quantification of the exact distribution of the sizes or how such distribution would change in response to genetic or environmental perturbation has not been carried out. Another example is the development of immune repertoire. It is well appreciated that naïve T cells can give rise to diverse effector cell types with different phenotypes and functionalities, implicating the functional role of regulated population heterogeneity. Yet the focus so far has been placed on assigning cells to one of several defined categories based on a set of surface markers. Recent studies have demonstrated that the constituent cells within such a heterogeneous population differ from each other on a rather continuous spectrum (27,47). These observations blur the exact boundaries between discrete cell types. In the context of cancer, it has been shown that one differentiating feature between tumor samples is the distribution of cellular subtypes. Furthermore, this distribution is robustly associated with each sample (36,69). Our methodology equips experimentalists with an intuitive platform to address questions on how populations of somatic cells can potentially act as a unit to process environmental cues and how the whole organism can benefit from population-level information processing.

We demonstrated the inference of a continually varying signal (a single network parameter) using divergence-based kernel regression. However, the same framework can be applied in categorical learning and inference. Different categories may represent experimental conditions that cannot

be connected in any network-state parameter space. Yet they can be ordered based on their pairwise divergence and indexed. The index will be treated as the parameter to be inferred, and the corresponding category can then be retrieved.

Finally, our sensitivity analysis method can help better integrate modeling and experimentation. The stochastic sensitivity analysis can be applied as an initial constraint on parameters of the mechanistic model to check its overall validity. The rationale here is that a sound mechanistic model should adequately capture how the distribution of a network output responds to experimental perturbation. In this context, it is possible to restrict the perturbation to a single parameter and examine the distributional response, both computationally and experimentally. The model prediction on stochastic sensitivity can also direct experimental perturbation to a targeted parameter domain to achieve either improved resolution in parameter estimation or enhanced system sensitivity to subsequent perturbations. Examining the stochastic sensitivity curve corresponding to different cellular states will also enable optimal sampling of fingerprints and enhance the accuracy and efficiency of the kernel predictor.

SUPPORTING MATERIAL

Three figures are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(14)00743-7.

We thank Tae Lee for providing experimental data and Cheemeng Tan for providing helpful comments. B.L. and L.Y. conceived and designed the experiments, B.L. performed the experiments, and B.L. and L.Y. analyzed the data, contributed materials/analysis tools, and wrote the article.

This work was partially supported by the National Science Foundation (CBET-0953202), the National Institutes of Health (1R01GM098642), a DuPont Young Professorship (L.Y.), and a David and Lucile Packard Fellowship (L.Y.).

REFERENCES

1. Lee, T. J., G. Yao, …, L. You. 2010. Stochastic E2F activation and reconciliation of phenomenological cell-cycle models. *PLoS Biol.* 8:e1000488.
2. Cheong, R., A. Rhee, …, A. Levchenko. 2011. Information transduction capacity of noisy biochemical signaling networks. *Science.* 334: 354–358.
3. Pedraza, J. M., and A. van Oudenaarden. 2005. Noise propagation in gene networks. *Science.* 307:1965–1969.
4. Maamar, H., A. Raj, and D. Dubnau. 2007. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science.* 317:526–529.
5. Losick, R., and C. Desplan. 2008. Stochasticity and cell fate. *Science.* 320:65–68.
6. Elowitz, M. B., A. J. Levine, …, P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science.* 297:1183–1186.
7. Wernet, M. F., E. O. Mazzoni, …, C. Desplan. 2006. Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature.* 440:174–180.
8. Volfson, D., J. Marciniak, …, J. Hasty. 2006. Origins of extrinsic variability in eukaryotic gene expression. *Nature.* 439:861–864.
9. Raj, A., S. A. Rifkin, …, A. van Oudenaarden. 2010. Variability in gene expression underlies incomplete penetrance. *Nature.* 463:913–918.
10. Lu, T., M. Ferry, …, J. Hasty. 2008. A molecular noise generator. *Phys. Biol.* 5:036006.
11. Batchelor, E., A. Loewer, …, G. Lahav. 2011. Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.* 7:488.
12. Tay, S., J. J. Hughey, …, M. W. Covert. 2010. Single-cell NF-κB dynamics reveal digital activation and analogue information processing. *Nature.* 466:267–271.
13. Wong, J. V., G. Yao, …, L. You. 2011. Viral-mediated noisy gene expression reveals biphasic E2f1 response to MYC. *Mol. Cell.* 41: 275–285.
14. Lu, T., T. Shen, …, J. Hasty. 2007. Phenotypic variability of growing cellular populations. *Proc. Natl. Acad. Sci. USA.* 104:18982–18987.
15. Spencer, S. L., and P. K. Sorger. 2011. Measuring and modeling apoptosis in single cells. *Cell.* 144:926–939.
16. Acar, M., J. T. Mettetal, and A. van Oudenaarden. 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* 40:471–475.
17. Johnston, Jr., R. J., and C. Desplan. 2010. Stochastic mechanisms of cell fate specification that yield random or robust outcomes. *Annu. Rev. Cell Dev. Biol.* 26:689–719.
18. Chabot, J. R., J. M. Pedraza, …, A. van Oudenaarden. 2007. Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature.* 450:1249–1252.
19. Weinberger, L. S., J. C. Burnett, …, D. V. Schaffer. 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell.* 122:169–182.
20. Raj, A., and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell.* 135:216–226.
21. Eldar, A., and M. B. Elowitz. 2010. Functional roles for noise in genetic circuits. *Nature.* 467:167–173.
22. Yao, G., T. J. Lee, …, L. You. 2008. A bistable Rb-E2F switch underlies the restriction point. *Nat. Cell Biol.* 10:476–482.
23. Altschuler, S. J., and L. F. Wu. 2010. Cellular heterogeneity: do differences make a difference? *Cell.* 141:559–563.
24. Pelkmans, L. 2012. Cell Biology. Using cell-to-cell variability—a new era in molecular biology. *Science.* 336:425–426.
25. Wang, C., Y. W. Tian, X. W. Wu, …, X. Z. Zhao. 1990. Genetic polymorphisms of HLA class III and GLO1 in Chinese Yao nationality. *Gene Geogr.* 4:29–34.
26. Lestas, I., G. Vinnicombe, and J. Paulsson. 2010. Fundamental limits on the suppression of molecular fluctuations. *Nature.* 467:174–178.
27. Feinerman, O., J. Veiga, …, G. Altan-Bonnet. 2008. Variability and robustness in T cell activation from regulated heterogeneity in protein levels. *Science.* 321:1081–1084.
28. Regan, E. R., and W. C. Aird. 2012. Dynamical systems approach to endothelial heterogeneity. *Circ. Res.* 111:110–130.
29. Buchholz, V. R., M. Flossdorf, …, D. H. Busch. 2013. Disparate individual fates compose robust CD8+ T cell immunity. *Science.* 340: 630–635.
30. Tubo, N. J., A. J. Pagán, …, M. K. Jenkins. 2013. Single naive CD4+ T cells from a diverse repertoire produce different effector cell types during infection. *Cell.* 153:785–796.
31. Gupta, P. B., C. M. Fillmore, …, E. S. Lander. 2011. Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell.* 146:633–644.
32. Reference deleted in proof.
33. Treue, S., K. Hol, and H. J. Rauber. 2000. Seeing multiple directions of motion-physiology and psychophysics. *Nat. Neurosci.* 3:270–276.
34. MacArthur, B. D., and I. R. Lemischka. 2013. Statistical mechanics of pluripotency. *Cell.* 154:484–489.

35. Wang, C. L., D. C. Yang, and M. Wabl. 2006. Slow, stochastic transgene repression with properties of a timer. *Genome Biol.* 7:R47.

36. Singh, D. K., C. J. Ku, …, S. J. Altschuler. 2010. Patterns of basal signaling heterogeneity can distinguish cellular populations with different drug sensitivities. *Mol. Syst. Biol.* 6:369.

37. Perlman, Z. E., M. D. Slack, …, S. J. Altschuler. 2004. Multidimensional drug profiling by automated microscopy. *Science.* 306:1194–1198.

38. Loo, L. H., H. J. Lin, …, L. F. Wu. 2009. Heterogeneity in the physiological states and pharmacological responses of differentiating 3T3-L1 preadipocytes. *J. Cell Biol.* 187:375–384.

39. Stachurski, D., B. R. Smith, …, S. A. Wang. 2008. Flow cytometric analysis of myelomonocytic cells by a pattern recognition approach is sensitive and specific in diagnosing myelodysplastic syndrome and related marrow diseases: emphasis on a global evaluation and recognition of diagnostic pitfalls. *Leuk. Res.* 32:215–224.

40. Bonassi, F. V., L. C. You, and M. West. 2011. Bayesian learning from marginal data in bionetwork models. *Stat. Appl. Genet. Mol. Biol.* 10:49.

41. Zechner, C., J. Ruess, …, H. Koeppl. 2012. Moment-based inference predicts bimodality in transient gene expression. *Proc. Natl. Acad. Sci. USA.* 109:8340–8345.

42. Cox, C. D., J. M. McCollum, …, M. L. Simpson. 2008. Using noise to probe and characterize gene circuits. *Proc. Natl. Acad. Sci. USA.* 105:10809–10814.

43. Kügler, P. 2012. Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. *PLoS ONE.* 7:e43001.

44. Ren, J., W.-X. Wang, …, Y.-C. Lai. 2010. Noise bridges dynamical correlation and topology in coupled oscillator networks. *Phys. Rev. Lett.* 104:058701.

45. Singh, A., B. S. Razooky, …, L. S. Weinberger. 2012. Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. *Mol. Syst. Biol.* 8:607.

46. Munsky, B., B. Trinh, and M. Khammash. 2009. Listening to the noise: random fluctuations reveal gene network parameters. *Mol. Syst. Biol.* 5:318.

47. Feinerman, O., G. Jentsch, …, G. Altan-Bonnet. 2010. Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Mol. Syst. Biol.* 6:437.

48. August, E. 2012. Using noise for model-testing. *J. Comput. Biol.* 19:968–977.

49. Ge, H., S. Pressé, …, K. A. Dill. 2012. Markov processes follow from the principle of maximum caliber. *J. Chem. Phys.* 136:064108.

50. Lillacci, G., and M. Khammash. 2012. A distribution-matching method for parameter estimation and model selection in computational biology. *Int. J. Robust Nonlin.* 22:1065–1081.

51. Munsky, B., and M. Khammash. 2010. Identification from stochastic cell-to-cell variation: a genetic switch case study. *IET Syst. Biol.* 4:356–366.

52. Maienschein-Cline, M., A. Warmflash, and A. R. Dinner. 2010. Defining cooperativity in gene regulation locally through intrinsic noise. *IET Syst. Biol.* 4:379–392.

53. Finn, W. G., A. M. Harrington, …, A. O. Hero, 3rd. 2011. Immunophenotypic signatures of benign and dysplastic granulopoiesis by cytomic profiling. *Cytometry B Clin. Cytom.* 80:282–290.

54. Finn, W. G., K. M. Carter, …, A. O. Hero. 2009. Analysis of clinical flow cytometric immunophenotyping data by clustering on statistical manifolds: treating flow cytometry data as high-dimensional objects. *Cytometry B Clin. Cytom.* 76:1–7.

55. Rogers, W. T., A. R. Moser, …, J. S. Moore. 2008. Cytometric fingerprinting: quantitative characterization of multivariate distributions. *Cytometry A.* 73:430–441.

56. Cağatay, T., M. Turcotte, …, G. M. Süel. 2009. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell.* 139:512–522.

57. Gillespie, D. T. 2000. The chemical Langevin equation. *J. Chem. Phys.* 113:297–306.

58. Lee, T. J., C. M. Tan, …, L. C. You. 2007. Modeling cellular networks. *In* Bioinformatics: An Engineering Case-Based Approach. G. Alterovitz and M. F. Ramoni, editors. Artech House, Boston, pp. 151–172.

59. Hershey, J. R., and P. A. Olsen. 2007. Approximating the Kullback Leibler divergence between Gaussian mixture models. *IEEE Int. Conf. Acoustics, Speech, and Signal Proc., Honolulu.* 4:317–320.

60. Itzkovitz, S., E. Hodis, and E. Segal. 2010. Overlapping codes within protein-coding sequences. *Genome Res.* 20:1582–1589.

61. Sims, G. E., S. R. Jun, …, S. H. Kim. 2009. Alignment-free genome comparison with feature frequency profiles (FFP) and optimal resolutions. *Proc. Natl. Acad. Sci. USA.* 106:2677–2682.

62. Perez-Cruz, F. 2008. Kullback-Leibler divergence estimation of continuous distributions. *Proc. IEEE Int. Symp. Inform. Theory, Toronto.* 2008:1666–1670.

63. Hallen, M., B. Li, …, L. You. 2011. Computation of steady-state probability distributions in stochastic models of cellular networks. *PLOS Comput. Biol.* 7:e1002209.

64. Warmflash, A., and A. R. Dinner. 2008. Signatures of combinatorial regulation in intrinsic biological noise. *Proc. Natl. Acad. Sci. USA.* 105:17262–17267.

65. Balázsi, G., A. van Oudenaarden, and J. J. Collins. 2011. Cellular decision making and biological noise: from microbes to mammals. *Cell.* 144:910–925.

66. Shawe-Taylor, J., and N. Cristianini. 2004. Kernel methods for pattern analysis. Cambridge University Press, Cambridge, United Kingdom.

67. Canham, M. A., A. A. Sharov, …, J. M. Brickman. 2010. Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.* 8: e1000379.

68. Gupta, V., and K. D. Poss. 2012. Clonally dominant cardiomyocytes direct heart morphogenesis. *Nature.* 484:479–484.

69. Slack, M. D., E. D. Martinez, …, S. J. Altschuler. 2008. Characterizing heterogeneous cellular responses to perturbations. *Proc. Natl. Acad. Sci. USA.* 105:19306–19311.