

REVIEW

Predictive power of cell-to-cell variability

Bochong Li¹ and Lingchong You^{1,2,3,*}

¹ Department of Biomedical Engineering, Duke University, Durham, NC 27708, USA

² Center for Systems Biology, Duke University, Durham, NC 27708, USA

³ Institute for Genome Sciences and Policy, Duke University, Durham, NC 27708, USA

* Correspondence: you@duke.edu

Received January 26, 2013; Revised March 20, 2013; Accepted March 20, 2013

Much of our current knowledge of biology has been constructed based on population-average measurements. However, advances in single-cell analysis have demonstrated the omnipresent nature of cell-to-cell variability in any population. On one hand, tremendous efforts have been made to examine how such variability arises, how it is regulated by cellular networks, and how it can affect cell-fate decisions by single cells. On the other hand, recent studies suggest that the variability may carry valuable information that can facilitate the elucidation of underlying regulatory networks or the classification of cell states. To this end, a major challenge is determining what aspects of variability bear significant biological meaning. Addressing this challenge requires the development of new computational tools, in conjunction with appropriately chosen experimental platforms, to more effectively describe and interpret data on cell-cell variability. Here, we discuss examples of when population heterogeneity plays critical roles in determining biologically and clinically significant phenotypes, how it serves as a rich information source of regulatory mechanisms, and how we can extract such information to gain a deeper understanding of biological systems.

INTRODUCTION

Phenotypic heterogeneity in cell populations is a widespread phenomenon observed across physiologic processes and pathological conditions [1–9]. Such heterogeneity stems from fluctuations in cellular components (e.g., RNA and protein abundance) and states (e.g., cell-cycle phase and stress), compounded by differences in the microenvironments experienced by the cell (e.g., local cell density and cell-cell contact) [8–18]. The fluctuation in internal cellular state fundamentally arises from the discrete nature of molecules and the randomness in their interactions, also known as “noise” [19–21]. Phenotypic heterogeneity often manifests as differential fate determination, including cell growth, senescence, and death in response to both natural (e.g., growth factors) and artificial (e.g., therapeutics) stimulus. Functional implications of cell-to-cell variability under various biological contexts have been firmly established in the past decade [3,7,22–30].

Historically, biological models (e.g., cellular networks and signal processing) have been conceived based on

experimental measurements of population-average behavior (e.g., western blot). However, given the large degree of phenotypic cell-to-cell variability even between genetically homogeneous individuals and in response to identical stimulus, conclusions reached as such may be misleading [9,31,32]. For example, p53 dynamics in response to DNA damage was previously thought of as damped oscillations, until later single-cell experiments showed that individual cells give rise to varying numbers of p53 pulses of fixed amplitude and duration [33]. In another case, single-cell analysis of E2F responses to varying levels of Myc expression, which was enabled through viral gene delivery with drastically heterogeneous efficiency, revealed a biphasic dependence of the former on the latter, shedding new light on cell-cycle regulation [34].

Most studies to date have been focusing on the explanation of cell-to-cell variability (the forward problem), such as the delineation of its source, mechanisms of generation, propagation along cellular networks, as well as the development of computational methods to predict variability given a particular mechanistic structure

[21,35–46]. The reverse problem, that is, if and how critical information can be gained from cell-to-cell variability, is only starting to be appreciated and will be the subject under review here. We focus on two specific aspects of the information capacity in heterogeneity observation: First, in constraining mathematical models of the biological system (e.g., signaling pathways); second, in associating with and predicting functional population-level phenotypes (e.g., drug efficacy).

USING CELL-TO-CELL VARIABILITY TO INFER REGULATORY MECHANISMS

Efforts in addressing the aforementioned forward problem have offered evidence of regulatory mechanisms underlying cellular pathways dictating the characteristic cell-to-cell variability pattern in single-cell attributes, such as gene or protein expression levels (Figure 1A)[35,39,47–50]. Such regulatory mechanisms are often described in the form of mathematical models that quantitatively capture the production, degradation, and interaction between cellular components. Different mechanisms correspond to different model structures and parameterization. Given such a model, the variability in single-cell behaviors can be attributed to either intrinsic or extrinsic noise sources. Intrinsic noise can be explicitly accounted for by the mechanistic model itself and the assumed probability distribution dictating the random events involved in the model. Extrinsic noise, on the other hand, refers to the fluctuation in the cellular state or the microenvironment not included in the model, yet affecting system dynamics in a “predetermined manner,” usually by influencing model parameters. Great efforts have been dedicated to identifying such extrinsic factors, the success in which would eliminate the uncertainty in such sources and result in what’s called “regulated cell-to-cell variability.” We refer the readers elsewhere for extensive review on this topic [18].

The inverse problem of inferring the regulatory mechanism is much more challenging than the forward problem [51], largely due to the daunting complexity incurred by the inclusion of stochastic elements in the modeling framework and the requirement of more sophisticated statistical toolbox to handle distributional data.

Several studies have made a case for a general framework whereby *a priori* construction of a mapping between the regulatory mechanism (e.g., network structure or parameter) and the distribution of some output measure (e.g., protein abundance) allows subsequent inference of the mechanism based on variability observation. In one case, analytically derived relationships between the moments of distribution from an expanded

chemical master equation and the structure of the cis-regulatory input function are established to infer the cooperativity as well as the combinatorial logic (‘AND’ vs. ‘OR’) between transcription factors in promoter regulation [52,53]. Different sources of intrinsic noise, namely promoter fluctuation and mRNA birth/death fluctuation in a transcriptional system, are also distinguishable based on variability measures [54]. It was shown in this case that the coefficient of variance in the protein product as a function of time after transcription blockade changes more slowly if promoter fluctuation dominates. It is interesting to note that the steady-state distribution of the protein alone is not sufficient in discriminating between these two noise sources. This is partly because the variability in protein expression mainly scales with the natural protein abundance, which can be treated as a constant extrinsic noise regardless of how the intrinsic noise is generated [55].

Efforts have also been made in developing parameter estimation schemes based on variability measures to better constrain the mathematical model of a biologic system, as the majority of parameters are difficult to measure directly through experimentation. Although stochastic time-series data in single cells may be highly informative [56], we focus on the analysis of snapshot distributional data over the population at fixed time points. Such distributional data can be easily made available using mature platforms such as flow cytometry and microscopy with low cost and high efficiency. Note that when time series data of high temporal resolution are of interest, especially when correlation in the temporal dynamics of single cell attributes within individual cells is under investigation, microscopy or other continuous-time measurement platforms are better suited. Various experimental methodologies for studying single cell dynamics and population heterogeneity are reviewed in great detail in the recent literature [57].

In one of the pioneering studies, Minsky et al. developed a scheme that takes single-cell readout distributions at multiple time points before reaching steady-state and applies discrete state space probability matching [58]. Importantly, they showed that stationary moments of any arbitrary order are insufficient in uniquely identifying model parameters. For simple systems, analytical expressions of parameters as functions of distribution statistics can be derived. For complex systems, a parameter search process, which identifies the parameters generating distributions most consistent with the experimental observation, proved to be equally effective (Figure 1B). Using IPTG-induced gene expression as a test model, the authors showed that the fitted model captures gene expression distribution at unlearned induction levels despite drastically different overall shape

of the distributions. The efficiency and accuracy of the scheme, however, do depend on the choice of the molecular species and the order of the statistics.

Similar to the earlier examples of regulatory mechanism inference, extrinsic noise, which has been shown to be dominating in most biological systems, is ignored

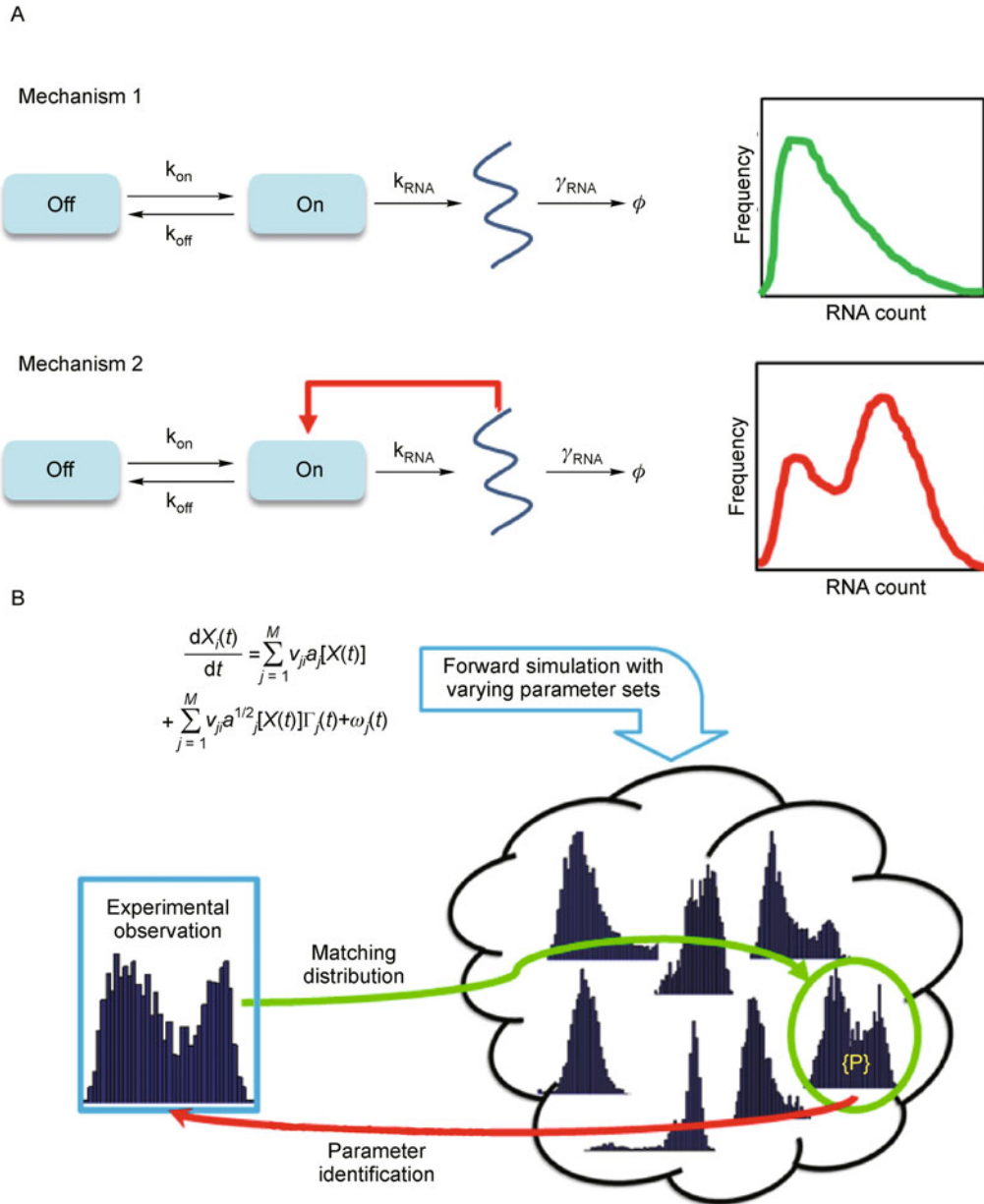


Figure 1. Inference of regulatory mechanisms from population variability measurements. (A) Two slightly different transcription regulatory mechanisms generate distinctive distribution of mRNA molecules under stochastic reaction dynamics over a population. Above, the promoter of the gene randomly switches between on and off states. When the promoter is on, mRNA molecules are produced, which in turn undergoes decay. Below, the mRNA molecule (or its protein product) and feedback to the promoter and enhance the transcription rate when it's on, resulting in a bimodal distribution of the mRNA molecule. By counting RNA molecules in single cells and obtain its variability measure over the population, the regulatory mechanism of transcription may be distinguished between alternative hypotheses. (B) Given a regulatory mechanism, a mathematical model capable of simulating stochastic chemical dynamics (such as stochastic differential equations), whose structure best represents the current understanding of the biological system, can be constructed. The parameters (or parameter distributions) of the model and hence the biological system can be more efficiently inferred by comparing the experimentally observed distribution of one or more system components with that generated by the model with varying parameter values and selecting one generating the best matching distribution.

[6,42,55]. Subsequent studies took on the task of incorporating extrinsic noise into the inference scheme [6,42,55]. This can be accomplished by imposing a distribution over the model parameters and initial conditions [59]. Instead of finding the best point estimation of each parameter, a distribution of parameter values is obtained [60]. Despite the constraints on the extrinsic noise structure and the necessity to parameterize its density, results from these studies would allow better statistical modeling of experimentally observed variability.

One concern in performing such inference is the scalability of the scheme: Whether the algorithm can handle systems of much greater size than what is demonstrated as a proof of concept with reasonably high efficiency and low computational cost. To address this issue, moment-based approaches and convolution techniques have been proposed to combine intrinsic and extrinsic noise [45,61]. Conventional parameter search algorithms, such as the Metropolis-Hastings MCMC sampler, are used to maximize the likelihood of the parameter posterior. The power of such a framework has been demonstrated by fitting a stochastic model of osmotic stress induced transcriptional activation in budding yeast using only the mean and the variance of the measured distribution. The fitted model successfully predicted the transient bimodality in gene expression, and provided evidence for the role of chromatin remodeling in regulating population variability. Such moment-based frameworks have the potential of alleviating the curse of dimensionality and the prohibitive computational cost of the integer-valued state-space models [34,62]. However, their success may depend on the choice of moment closure method, the accuracy of which depends on both the network structure and the unknown parameters. Aside from such heuristic aspects, existing work seems to warrant certain level of confidence in moment-based parameter estimation, which has been argued to be inadequate to handle complicated distributions [51].

Others have proposed various moment or density based frameworks to perform parameter estimation, as well as model selection, using cell-to-cell heterogeneity distributional data as input [63–65]. Alternative approaches include noise autocorrelation function [66] and spectral analysis [67]. Studies examining relationships between noise properties and network topology have been examined as well [68]. These alternative approaches simply take advantage of different statistical features of the distribution of cellular readouts. Regardless of the specific implementation, inference based on variability measures has been proven advantageous over population mean measures, and even necessary under certain circumstances.

For the mechanistic model selection task, a ‘library’ of

models has to be constructed *a priori*, which is limited to our current biological knowledge base and the scope of our conception. It would be interesting to further investigate more efficient, exhaustive, and less biased ways of constructing such model candidate library.

The explicit modeling of extrinsic noise statistics enables the computational model to shed light on the roles certain modulators play in the regulation of cell-cell variability, whose mechanism of noise generation is not directly captured by the mechanistic model. We note that regulated cell-to-cell variability falls exactly under the category of extrinsic noise. The fundamental idea in regulated cell-to-cell heterogeneity is that there exists high degree of correlation between molecular or dynamical readouts, which may be hidden from the observer, and single cell behaviors, as well as between molecular readouts (e.g., different signaling molecules). As almost all studies in identifying hidden determinants of cell-to-cell variability take a single-cell readout to single-cell activity mapping approach, adequate and flexible modeling of extrinsic noise may be exactly what we need to bridge the gap between population level manifestation of regulated variability and underlying regulatory mechanisms. The regularity and determinism at the single-cell level may be reflected in the exact pattern of the extrinsic noise parameter distribution [69].

FUNCTIONAL CHARACTERIZATION OF POPULATION THROUGH HETEROGENEITY PROFILING

So far, we have discussed how cell-to-cell variability can provide useful information on the underlying regulatory mechanisms. However, the complexity of cellular heterogeneity often stretches beyond our knowledge base and cognitive boundaries. Nonetheless, as we will show in this section, such complexity does not keep us from exploiting cell-to-cell variability as an effective tool to capture population-level phenotypes (Figure 2A) that may be difficult to characterize or even ill-posed at the single-cell level.

The awareness of phenotypic heterogeneity at the population or tissue level being a critical issue in disease (e.g., cancer) diagnostics and treatment has been raised substantially [17,45,57,70–75]. Subpopulations of cancer cells are believed to underlie the emergence of drug resistance [57,70,72,75–77]. It has been suggested that the dynamic maintenance of a specific fractional composition of different subpopulation, each with different cellular characteristics, distinguishes different cancer cell populations [78]. Most studies on cancer heterogeneity so far aim at associating subpopulations with differentially activated pathways, and consequently targeting them with

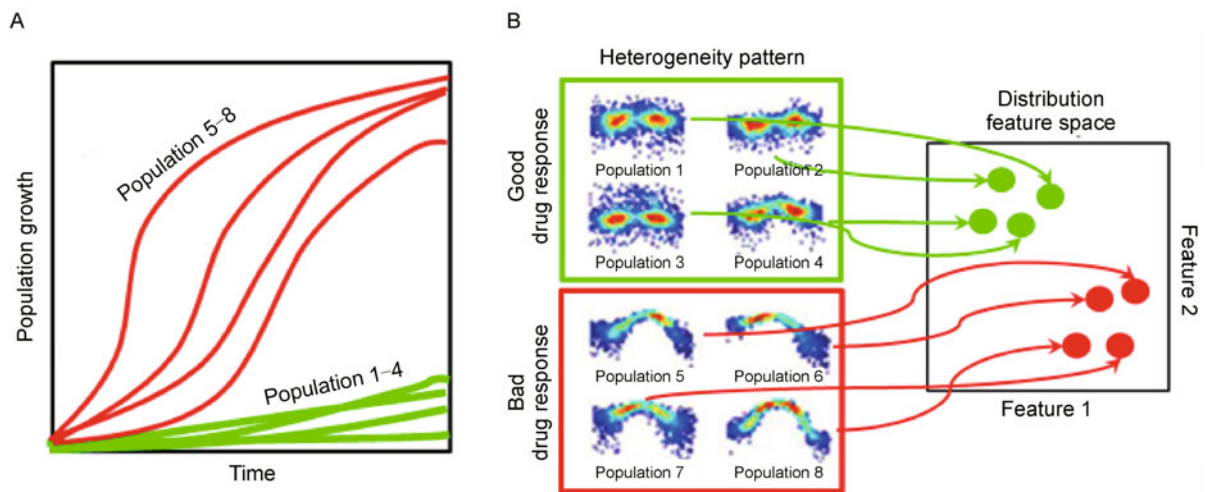


Figure 2. Characterizing population phenotypes with heterogeneity measures. (A) Cell populations manifest various population level emergent phenotypes, for example, the overall growth rate of various cancer cell populations (of different origins and malignancy) in response to anti-cancer therapeutics. Such phenotypes can be classified into qualitatively different categories, such as good or poor responses to therapeutics shown here. (B) Population phenotypes may correlate with population heterogeneity in certain single-cell attributes, including gene express, surface marker, and signaling molecule co-localization mentioned in the text. That is, similarity in the population heterogeneity measure (shape of the distribution clustering in the distribution feature space) coincides with the classification in population phenotype. Here the heterogeneity measure is represented by 2D joint distributions (red, higher probability density; blue, lower probability density). Consequently, after curating a large database of phenotype-heterogeneity correlations, one can predict population phenotype by measuring population heterogeneity and determining which distribution pattern category the observation belongs to.

rationally designed drugs [57,73,74,79–83]. Though this strategy may seem promising in many cases, the limitation is still the confinement to the existing biological knowledge base. As demonstrated in a recent study [84], functional subpopulations may be masked by ‘blind’ adoption of previously established gating strategy and marker set. This is an example of how hidden subpopulations, and thus information critical to the functional output of the entire population, can be lost in naïve handling of distributional heterogeneity profiles. As discussed in the previous section, cell-to-cell variability patterns contain an enormous amount of information regarding the regulatory network underlying cellular behavior; heterogeneity profiling may therefore serve as a robust predictor of population phenotype, which must emerge from the aggregation of single cell behaviors. This is a largely under-appreciated field. However, we point to a couple of pioneering studies from the Altschuler and Wu group along this line of inquiry.

In one study, Slack et al. developed an imaging-assay based platform to extract measurements on the co-localization patterns of fluorescently labeled cellular components from a large number of cells under diverse conditions. For different cells in the same population under identical environmental condition, such co-localization patterns differ significantly and represent phenotypic variability in the signaling state of individual cells.

Here, each cell is represented by a point in the pattern space and the phenotypically heterogeneous population hence corresponds to a distribution in such a space (Figure 2B). By applying an unsupervised learning algorithm to phenotypic distributions of a large number of cell populations under diverse conditions, unique phenotypic “stereotypes” are assigned to cells residing in different regions of the pattern space. Each population is subsequently represented by the fractional composition, or a heterogeneity profile, of each stereotype. They found that drugs of similar mechanism usually induce similar patterns in the population heterogeneity profile (Figure 2). Although the markers used to establish the subpopulation profile may not be directly involved in the pathways relevant to the perturbation of interest (such as drug action), the proximity in the heterogeneity profile does suggest a high likelihood of resemblance in the underlying biological process between different perturbations. This correspondence can facilitate the generation of novel biological insights (such as modes of drug action) by referencing to established biological knowledge base (compared to the heterogeneity profile that is induced by drugs of known action mode). The authors also explored the potential benefit of increasing the number of phenotypic stereotypes (the number of subpopulations), and found that the performance of classification plateaus at a relative small number of subpopulations, around 5.

This diminished marginal return may point to the drawback of introducing artificial non-biological separation in phenotypes. That is, not all dimensions in the heterogeneity profile are informative. One can potentially incur unnecessary computational cost by non-selectively keeping all facets of such distributions. Note that the lack of clear mechanistic understanding of each subpopulation and their association with environmental perturbation does not jeopardize the predictive power of the framework. The operation of the computational platform does not depend on how the heterogeneity profile is generated, as long as the single-cell level measurements can be converted into distributional data.

The same group later used the same platform to demonstrate that basal signaling heterogeneity contains information predictive of population response to perturbation, specifically drug sensitivity of both cancer and noncancerous clonal populations. For multiple clones derived from the same population of parental clone of H460 cancer cells, the resemblance in drug sensitivity highly correlate with the similarity in the pattern of basal signaling heterogeneity (Figure 2). Moreover, the predictive information resides in multiple marker sets examined, implying the potential of using patterns of general nonspecific markers in reflecting “deeper” similarities of underlying regulatory networks. This may not be too surprising given the densely connected nature of cellular networks [85]. Interestingly, a collection of noncancerous clone populations (human bronchial epithelial cell, HBEC) showed reduced ranges of overall heterogeneity and drug sensitivities in comparison with the cancer cell clones. This observation supports the notion that cancer cells possess much greater and clinically relevant phenotypic heterogeneity compared to normal cells. They also demonstrated the power of heterogeneity profiling in differentiating different cancer types. It was noted that the stereotypes of subpopulations are common between different cell populations, and the difference lies in the exact composition of different subpopulation. Critically, increase or decrease in the fraction of a certain subpopulation may be sufficient to account for functional differences between populations in some cases, resembling conventional gating practices in subpopulation identification [74]; however, higher order characterization of the entire ensemble of subpopulations is generally required.

It is important to note that molecular states of each specific subpopulation and their relationship to population phenotype (drug response) is far from clear, highlighting the potential of unbiased and systematic heterogeneity profiling of general biomarkers in serving as a robust predictor of complex population phenotype as well as a tool to aid the generation of novel biological insights. Such mapping may not be captured by characterization of

single-cell behavior, as it is by definition a statistical property of the population. Several subsequent studies have applied similar imaging based heterogeneity profiling platforms in different biological context to classify population phenotypes [86,87]. We note that imaging only represents one of many mature and emerging techniques of acquiring single-cell attributes. Others include flow cytometry and single-cell sequencing. In fact, distributional data collected through flow cytometry has been strongly suggested to provide diagnostically useful information in myelodysplastic syndrome and related marrow diseases, where differential mixtures of various cell types and maturation stages influence disease progression and treatment outcomes [88–91]. However, in such cases, the analysis of the distribution still focus on the extraction of fractional composition of various predefined cell types rather than through an unbiased and unsupervised learning algorithm. On the other hand, several studies have laid some computational ground work for carrying out such systematic unbiased analysis on flow cytometry data [92–95]. We speculate that by applying such computational platforms to various biological and clinical contexts, where analysis of distributional data has been of paramount importance, such as immunology, stem cell biology and cancer, major advances will be on the horizon.

Combining the herein reviewed two aspects of heterogeneity-based inquiries, we think that the major gap in making sense of cell-to-cell variability and population-level phenotype remains in relating single-cell activity output to population-level phenotypes where single-cell phenotypic heterogeneity is functionally relevant. For example, in the aforementioned signal colocalization heterogeneity measure, answering questions like how different signaling pathways are differentially activated within individual cells when such cells reside in different regions of the colocalization feature distribution or whether the colocalization feature can be quantitatively related to the probability of cell death or proliferation is key to transforming the statistical correlation between heterogeneity measures and population phenotypes into a more mechanistically insightful relationship amenable to quantitative modeling at the single-cell level using first principles, as illustrated in the first section of the review.

If such a mapping can be defined, one can potentially predict population-level phenotype more accurately and efficiently by first predicting single-cell activity output using either deterministic or stochastic model simulations, and then applying the mapping function. This union may be confounded by spatial factors in cell-to-cell communication and cooperation in defining population context and phenotype, which dictate and depend on single-cell level activities. We envision that the integration of

stochastic modeling, parameter inference, and functional mapping between heterogeneity profile and population phenotype will provide novel insights into how variable single-cell activities are determined by underlying molecular regulatory mechanisms and how they cooperate to generate a complex population.

ACKNOWLEDGEMENTS

The work described was partially supported by a National Science Foundation CAREER Award (BES-0625213), a DuPont Young Professorship, a David and Lucile Packard Fellowship, North Carolina Biotechnology Center (2012-MRG-1102), and Office of Naval Research (N00014-12-1-0631).

REFERENCES

- Balázsi, G., van Oudenaarden, A. and Collins, J. J. (2011) Cellular decision making and biological noise: From microbes to mammals. *Cell*, 144, 910–925.
- Nachman, I., Regev, A. and Ramanathan, S. (2007) Dissecting timing variability in yeast meiosis. *Cell*, 131, 544–556.
- Raj, A. and van Oudenaarden, A. (2008) Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell*, 135, 216–226.
- Spencer, S. L. and Sorger, P. K. (2011) Measuring and modeling apoptosis in single cells. *Cell*, 144, 926–939.
- Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. and Huang, S. (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature*, 453, 544–547.
- Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C. G. and Brent, R. (2005) Regulated cell-to-cell variation in a cell-fate decision system. *Nature*, 437, 699–706.
- Raj, A., Rifkin, S. A., Andersen, E. and van Oudenaarden, A. (2010) Variability in gene expression underlies incomplete penetrance. *Nature*, 463, 913–918.
- Spencer, S. L., Gaudet, S., Albeck, J. G., Burke, J. M. and Sorger, P. K. (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*, 459, 428–432.
- Tay, S., Hughey, J. J., Lee, T. K., Lipniacki, T., Quake, S. R. and Covert, M. W. (2010) Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature*, 466, 267–271.
- Khan, M., Vaes, E. and Mombaerts, P. (2011) Regulation of the probability of mouse odorant receptor gene choice. *Cell*, 147, 907–921.
- Zeng, L., Skinner, S. O., Zong, C., Sippy, J., Feiss, M. and Golding, I. (2010) Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell*, 141, 682–691.
- Vlamakis, H., Aguilar, C., Losick, R. and Kolter, R. (2008) Control of cell fate by the formation of an architecturally complex bacterial community. *Genes. Dev.*, 22, 945–953.
- St-Pierre, F. and Endy, D. (2008) Determination of cell fate selection during phage lambda infection. *Proc. Natl. Acad. Sci. USA*, 105, 20705–20710.
- Zong, C., So, L. H., Sepúlveda, L. A., Skinner, S. O. and Golding, I. (2010) Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Mol. Syst. Biol.*, 6, 440.
- Snijder, B., Sacher, R., Rämö, P., Damm, E. M., Liberali, P. and Pelkmans, L. (2009) Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*, 461, 520–523.
- Feinerman, O., Jentsch, G., Tkach, K. E., Coward, J. W., Hathorn, M. M., Sneddon, M. W., Emonet, T., Smith, K. A. and Altan-Bonnet, G. (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.
- Cohen, A. A., Geva-Zatorsky, N., Eden, E., Frenkel-Morgenstern, M., Issaeva, I., Sigal, A., Milo, R., Cohen-Saidon, C., Liron, Y., Kam, Z., et al. (2008) Dynamic proteomics of individual cancer cells in response to a drug. *Science*, 322, 1511–1516.
- Snijder, B. and Pelkmans, L. (2011) Origins of regulated cell-to-cell variability. *Nat. Rev. Mol. Cell Biol.*, 12, 119–125.
- Mettetal, J. T., Muzzey, D., Pedraza, J. M., Ozbudak, E. M. and van Oudenaarden, A. (2006) Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. USA*, 103, 7304–7309.
- Swain, P. S., Elowitz, M. B. and Siggia, E. D. (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA*, 99, 12795–12800.
- Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science*, 297, 1183–1186.
- Losick, R. and Desplan, C. (2008) Stochasticity and cell fate. *Science*, 320, 65–68.
- Acar, M., Mettetal, J. T. and van Oudenaarden, A. (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.*, 40, 471–475.
- Wernet, M. F., Mazzoni, E. O., Celik, A., Duncan, D. M., Duncan, I. and Desplan, C. (2006) Stochastic spineless expression creates the retinal mosaic for colour vision. *Nature*, 440, 174–180.
- Johnston, R. J. Jr and Desplan, C. (2010) Stochastic mechanisms of cell fate specification that yield random or robust outcomes. *Annu. Rev. Cell Dev. Biol.*, 26, 689–719.
- Chabot, J. R., Pedraza, J. M., Luitel, P. and van Oudenaarden, A. (2007) Stochastic gene expression out-of-steady-state in the cyanobacterial circadian clock. *Nature*, 450, 1249–1252.
- Weinberger, L. S., Burnett, J. C., Toettcher, J. E., Arkin, A. P. and Schaffer, D. V. (2005) Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell*, 122, 169–182.
- Maamar, H., Raj, A. and Dubnau, D. (2007) Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science*, 317, 526–529.
- Eldar, A. and Elowitz, M. B. (2010) Functional roles for noise in genetic circuits. *Nature*, 467, 167–173.
- Yao, G., Lee, T. J., Mori, S., Nevins, J. R. and You, L. C. (2008) A bistable Rb-E2F switch underlies the restriction point. *Nat. Cell Biol.* 10, 476–482.
- Batchelor, E., Loewer, A., Mock, C. and Lahav, G. (2011) Stimulus-dependent dynamics of p53 in single cells. *Mol. Syst. Biol.*, 7, 488.
- Pelkmans, L. (2012) Cell Biology. Using cell-to-cell variability—a new era in molecular biology. *Science*, 336, 425–426.
- Wang, C., Tian, Y. W., Wu, X. W. and Zhao, X. Z. (1990) Genetic polymorphisms of HLA class III and GLO1 in Chinese Yao nationality. *Gene Geography: A Computerized Bulletin on Human Gene Frequencies* 4, 29–34.
- Wong, J. V., Yao, G. A., Nevins, J. R. and You, L. C. (2011) Viral-mediated noisy gene expression reveals biphasic E2f1 response to MYC. *Mol. Cell*, 41, 275–285.
- Austin, D. W., Allen, M. S., McCollum, J. M., Dar, R. D., Wilgus, J. R., Sayler, G. S., Samatova, N. F., Cox, C. D. and Simpson, M. L. (2006) Gene network shaping of inherent noise spectra. *Nature*, 439, 608–611.

36. Lestas, I., Vinnicombe, G. and Paulsson, J. (2010) Fundamental limits on the suppression of molecular fluctuations. *Nature*, 467, 174–178.
37. Volfson, D., Marciniak, J., Blake, W. J., Ostroff, N., Tsimring, L. S. and Hasty, J. (2006) Origins of extrinsic variability in eukaryotic gene expression. *Nature*, 439, 861–864.
38. Bialek, W. and Setayeshgar, S. (2008) Cooperativity, sensitivity, and noise in biochemical signaling. *Phys. Rev. Lett.*, 100, 258101.
39. Friedman, N., Cai, L. and Xie, X. S. (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.*, 97, 168302.
40. Elf, J., Li, G. W. and Xie, X. S. (2007) Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 316, 1191–1194.
41. Pedraza, J. M. and Paulsson, J. (2008) Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319, 339–343.
42. Pedraza, J. M. and van Oudenaarden, A. (2005) Noise propagation in gene networks. *Science*, 307, 1965–1969.
43. Suter, D. M., Molina, N., Gatfield, D., Schneider, K., Schibler, U. and Naef, F. (2011) Mammalian genes are transcribed with widely different bursting kinetics. *Science*, 332, 472–474.
44. To, T.L. and Maheshri, N. (2010) Noise can induce bimodality in positive transcriptional feedback loops without bistability. *Science*, 327, 1142–1145.
45. Hallen, M., Li, B. C., Tanouchi, Y., Tan, C. E. M., West, M. and You, L. C. (2011) Computation of steady-state probability distributions in stochastic models of cellular networks. *PLoS Comput. Biol.*, 7, e1002209.
46. Ma, R., Wang, J. C., Hou, Z. H. and Liu, H. Y. (2012) Small-number effects: a third stable state in a genetic bistable toggle switch. *Phys. Rev. Lett.*, 109, 248107.
47. Cağatay, T., Turcotte, M., Elowitz, M. B., Garcia-Ojalvo, J. and Süel, G. M. (2009) Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139, 512–522.
48. Johnston, R. J. Jr, Otake, Y., Sood, P., Vogt, N., Behnia, R., Vasilias, D., McDonald, E., Xie, B., Koenig, S., Wolf, R., et al. (2011) Interlocked feedforward loops control cell-type-specific Rhodopsin expression in the *Drosophila* eye. *Cell*, 145, 956–968.
49. Dobrzynski, M. and Bruggeman, F. J. (2009) Elongation dynamics shape bursty transcription and translation. *Proc. Natl. Acad. Sci. USA*, 106, 2583–2588.
50. Tan, C., Marguet, P. and You, L. C. (2009) Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.*, 5, 842–848.
51. Munsky, B. and Khammash, M. (2010) Identification from stochastic cell-to-cell variation: a genetic switch case study. *IET Syst. Biol.*, 4, 356–366.
52. Warmflash, A. and Dinner, A. R. (2008) Signatures of combinatorial regulation in intrinsic biological noise. *Proc. Natl. Acad. Sci. USA*, 105, 17262–17267.
53. Maienschein-Cline, M., Warmflash, A. and Dinner, A. R. (2010) Defining cooperativity in gene regulation locally through intrinsic noise. *IET Syst. Biol.*, 4, 379–392.
54. Singh, A., Razoooky, B. S., Dar, R. D. and Weinberger, L. S. (2012) Dynamics of protein noise can distinguish between alternate sources of gene-expression variability. *Mol. Syst. Biol.*, 8, 607.
55. Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O’Shea, E., Pilpel, Y. and Barkai, N. (2006) Noise in protein expression scales with natural protein abundance. *Nat. Genet.*, 38, 636–643.
56. Lopes, F. M., de Oliveira, E. A. and Cesar, R. M. Jr. (2011) Inference of gene regulatory networks from time series by Tsallis entropy. *BMC Syst. Biol.*, 5, 61.
57. Bendall, S. C. and Nolan, G. P. (2012) From single cells to deep phenotypes in cancer. *Nat. Biotechnol.*, 30, 639–647.
58. Munsky, B., Trinh, B. and Khammash, M. (2009) Listening to the noise: random fluctuations reveal gene network parameters. *Mol. Syst. Biol.*, 5, 318.
59. Hasenauer, J., Waldherr, S., Doszczak, M., Radde, N., Scheurich, P. and Allgöwer, F. (2011) Identification of models of heterogeneous cell populations from population snapshot data. *BMC Bioinformatics*, 12, 125.
60. Bonassi, F. V., You, L. C. and West, M. (2011) Bayesian learning from marginal data in bionetwork models. *Stat. Appl. Genet. Mol.*, 10.
61. Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J. and Koepl, H. (2012) Moment-based inference predicts bimodality in transient gene expression. *Proc. Natl. Acad. Sci. USA*, 109, 8340–8345.
62. Lim, C. A., Yao, F., Wong, J. J. Y., George, J., Xu, H., Chiu, K. P., Sung, W. K., Lipovich, L., Vega, V. B., Chen, J., et al. (2007) Genome-wide mapping of RELA(p65) binding identifies E2F1 as a transcriptional activator recruited by NF-kappaB upon TLR4 activation. *Mol. Cell*, 27, 622–635.
63. Lillacci, G. and Khammash, M. (2012) A distribution-matching method for parameter estimation and model selection in computational biology. *International Journal of Robust and Nonlinear Control*, 22, 1065–1081.
64. Kügler, P. (2012) Moment fitting for parameter inference in repeatedly and partially observed stochastic biological models. *PLoS One*, 7, e43001.
65. August, E. (2012) Using noise for model-testing. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology* 19, 968–977.
66. Cox, C. D., McCollum, J. M., Allen, M. S., Dar, R. D. and Simpson, M. L. (2008) Using noise to probe and characterize gene circuits. *Proc. Natl. Acad. Sci. USA*, 105, 10809–10814.
67. Kim, D., Debusschere, B. J. and Najm, H. N. (2007) Spectral methods for parametric sensitivity in stochastic dynamical systems. *Biophys. J.*, 92, 379–393.
68. Ren, J., Wang, W. X., Li, B. and Lai, Y. C. (2010) Noise bridges dynamical correlation and topology in coupled oscillator networks. *Phys. Rev. Lett.*, 104, 058701.
69. Feinerman, O., Veiga, J., Dorfman, J. R., Germain, R. N. and Altan-Bonnet, G. (2008) Variability and robustness in T cell activation from regulated heterogeneity in protein levels. *Science*, 321, 1081–1084.
70. Brock, A., Chang, H. and Huang, S. (2009) Non-genetic heterogeneity—a mutation-independent driving force for the somatic evolution of tumours. *Nat. Rev. Genet.*, 10, 336–342.
71. Creixell, P., Schoof, E. M., Erler, J. T. and Linding, R. (2012) Navigating cancer network attractors for tumor-specific therapy. *Nat. Biotechnol.*, 30, 842–848.
72. Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.*, 366, 883–892.
73. Irish, J. M., Hovland, R., Krutzik, P. O., Perez, O. D., Bruserud, O., Gjertsen, B. T. and Nolan, G. P. (2004) Single cell profiling of

- potentiated phospho-protein networks in cancer cells. *Cell*, 118, 217–228.
74. Kotecha, N., Flores, N. J., Irish, J. M., Simonds, E. F., Sakai, D. S., Archambeault, S., Diaz-Flores, E., Coram, M., Shannon, K. M., Nolan, G. P., et al. (2008) Single-cell profiling identifies aberrant STAT5 activation in myeloid malignancies with specific clinical and biologic correlates. *Cancer Cell*, 14, 335–343.
 75. Hayashi, M., Okabe-Kado, J. and Hozumi, M. (1994) Flow-cytometric analysis of *in vivo* induction of differentiation of WEHI-3B myelomonocytic leukemia cells by recombinant granulocyte colony-stimulating factor. *Exp. Hematol.*, 22, 393–398.
 76. Marusyk, A., Almendro, V. and Polyak, K. (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, 12, 323–334.
 77. Li, W., Cui, L. B. and Ng, M. K. (2012) On computation of the steady-state probability distribution of probabilistic Boolean networks with gene perturbation. *J. Comput. Appl. Math.*, 236, 4067–4081.
 78. Gupta, P. B., Fillmore, C. M., Jiang, G., Shapira, S. D., Tao, K., Kuperwasser, C. and Lander, E. S. (2011) Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146, 633–644.
 79. Krutzik, P. O., Crane, J. M., Clutter, M. R. and Nolan, G. P. (2008) High-content single-cell drug screening with phosphospecific flow cytometry. *Nat. Chem. Biol.*, 4, 132–142.
 80. Ghimire, P., Wu, G. Y. and Zhu, L. (2010) Primary esophageal lymphoma in immunocompetent patients: two case reports and literature review. *World Journal of Radiology*, 2, 334–338.
 81. Song, Y., Yang, Z., Ke, Z., Yao, Y., Hu, X., Sun, Y., Li, H., Yin, J. and Zeng, C. (2012) Expression of 14–3-3gamma in patients with breast cancer: correlation with clinicopathological features and prognosis. *Cancer Epidemiology*, 36, 533–536.
 82. Bosch, J., Gerstein, H. C., Dagenais, G. R., Díaz, R., Dyal, L., Jung, H., Maggioni, A. P., Probstfield, J., Ramachandran, A., Riddle, M. C., et al. and the ORIGIN Trial Investigators. (2012) n-3 fatty acids and cardiovascular outcomes in patients with dysglycemia. *N. Engl. J. Med.*, 367, 309–318.
 83. Gerstein, H. C., Bosch, J., Dagenais, G. R., Díaz, R., Jung, H., Maggioni, A. P., Pogue, J., Probstfield, J., Ramachandran, A., Riddle, M. C., et al. and the ORIGIN Trial Investigators. (2012) Basal insulin and cardiovascular and other outcomes in dysglycemia. *N. Engl. J. Med.*, 367, 319–328.
 84. Canham, M. A., Sharov, A. A., Ko, M. S. and Brickman, J. M. (2010) Functional heterogeneity of embryonic stem cells revealed through translational amplification of an early endodermal transcript. *PLoS Biol.*, 8, e1000379.
 85. Novershtern, N., Subramanian, A., Lawton, L. N., Mak, R. H., Haining, W. N., McConkey, M. E., Habib, N., Yosef, N., Chang, C. Y., Shay, T., et al. (2011) Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*, 144, 296–309.
 86. Caie, P. D., Walls, R. E., Ingleston-Orme, A., Daya, S., Houslay, T., Eagle, R., Roberts, M. E., and Carragher, N. O. (2010) High-content phenotypic profiling of drug response signatures across distinct cancer cells. *Mol. Cancer Ther.*, 9, 1913–1926.
 87. Sutherland, J. J., Low, J., Blosser, W., Dowless, M., Engler, T. A. and Stancato, L. F. (2011) A robust high-content imaging approach for probing the mechanism of action and phenotypic outcomes of cell-cycle modulators. *Mol. Cancer Ther.*, 10, 242–254.
 88. Elghetany, M. T. (2002) Diagnostic utility of flow cytometric immunophenotyping in myelodysplastic syndrome. *Blood*, 99, 391–392.
 89. Stetler-Stevenson, M., Arthur, D. C., Jabbour, N., Xie, X. Y., Molldrem, J., Barrett, A. J., Venzon, D. and Rick, M. E. (2001) Diagnostic utility of flow cytometric immunophenotyping in myelodysplastic syndrome. *Blood*, 98, 979–987.
 90. Finn, W. G., Harrington, A. M., Carter, K. M., Raich, R., Kroft, S. H. and Hero, A. O. 3rd. (2011) Immunophenotypic signatures of benign and dysplastic granulopoiesis by cytomic profiling. *Cytometry B Clin. Cytom.*, 80, 282–290.
 91. Stachurski, D., Smith, B. R., Pozdnyakova, O., Andersen, M., Xiao, Z., Raza, A., Woda, B. A., and Wang, S. A. (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.
 92. Pyne, S., Hu, X., Wang, K., Rossin, E., Lin, T. I., Maier, L. M., Baecher-Allan, C., McLachlan, G. J., Tamayo, P., Hafler, D. A., et al. (2009) Automated high-dimensional flow cytometric data analysis. *Proc. Natl. Acad. Sci. USA*, 106, 8519–8524.
 93. Finn, W. G., Carter, K. M., Raich, R., Stoolman, L. M. and Hero, A. O. (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.
 94. Rogers, W. T. and Holyst, H. A. (2009) FlowFP: A bioconductor package for fingerprinting flow cytometric data. *Adv. Bioinformatics*, 193947.
 95. Rogers, W. T., Moser, A. R., Holyst, H. A., Bantly, A., Mohler, E. R., 3rd, Scangas, G. and Moore, J. S. (2008) Cytometric fingerprinting: quantitative characterization of multivariate distributions. *Cytometry. Part A: the Journal of the International Society for Analytical Cytology*, 73, 430–441.