MODELING CELL HETEROGENEITY: FROM SINGLE-CELL VARIATIONS TO MIXED CELLS

ERIC BATCHELOR
Center for Cancer Research, National Cancer Institute, National Institutes of Health
Bethesda, MD 20894, USA
Email: eric.batchelor@nih.gov

MARICEL G. KANN
Department of Biological Sciences, University of Maryland, Baltimore County
Baltimore, MD, 21250, USA
Email: mkann@umbc.edu

TERESA M. PRZYTYCKA
National Center for Biotechnology Information, National Institutes of Health
Bethesda, MD 20894, USA
Email: przytyck@ncbi.nlm.nih.gov

BENJAMIN J. RAPHAEL
Department of Computer Science, Brown University
Providence, RI 02912, USA
Email: braphael@cs.brown.edu

DAMIAN WOJTOWICZ
National Center for Biotechnology Information, National Institutes of Health
Bethesda, MD 20894, USA
Email: wojtowda@ncbi.nlm.nih.gov

Emerging technologies such as single cell gene expression analysis and single cell genome sequencing provide an unprecedented opportunity to quantitatively probe biological interactions at the single cell level. This new level of insight has begun to reveal a more accurate picture of cellular behavior, and to highlight the importance of understanding cellular variation in a wide range of biological contexts. The aim of this workshop is to bring together researchers working on identifying and modeling cell heterogeneity that arises by a variety of mechanisms, including but not limited to cell-to-cell noise, cell-state switches and cell differentiation, heterogeneity in immune responses, cancer evolution, and heterogeneity in disease progression.

1. Background

Quantifying the molecular mechanisms underlying cell behaviors and functions is one of the ultimate goals of biology and medicine. Until recently, most current measures to classify and characterize cellular behavior have been performed on the average of all cells in a sample instead of a single cell. However, measurements derived from pooled populations of cells lack the specificity to capture outlier cell behavior that might explain cell differentiation and transitions

from normal to disease cellular states. The noise, or variance, between the genomic state of different cells -- even among cells assumed to be homogenous -- has been shown to be strongly correlated with protein expression and function [1]. Furthermore, it has been argued that neglecting cell heterogeneity is one of the major causes of error in disease classification [2]. Emergence of cell heterogeneity might be sporadic (e.g., cell-to-cell variation in an isogenic cell population [3]), programmed (e.g., cell differentiation [4]), or, to some extent, driven by selection pressure (e.g., in cancer [5]).

Emerging technologies like single cell gene expression and single cell sequencing provide unprecedented opportunities to quantify single cell level differences. These technologies will be able to provide a wealth of various forms of new information including protein abundance, methylation patterns, promoter structure, gene expression, copy number variations, gene function and essentiality, DNA structure, evolutionary plasticity, and selective advantage. These data can all be leveraged in the quest to understand the emergence and consequences of heterogeneity. However, simple accrual of data from these various single cell experimental techniques is not enough to obtain a clear understanding of the diverse range of biological processes affected by cellular heterogeneity. Synthesis and interpretation of the wealth of single cell-level data depends on novel computational approaches to uncover and model the biological principles that underlie the emergence of cell heterogeneity. Most importantly, computational methods are needed to provide a system-level view of the interplay of diverse, fluctuating biological components.

The increasing interest in the mechanisms underlying diseases is complemented by the emergence of single cell technologies. Due to these factors, we expect increasing efforts to develop computational models and tools for the analysis of cell heterogeneity. We believe that this workshop gives the community of computational biologists the chance to discuss this theme, which may soon become dominant in biomedical science.

2. Main directions and challenges

The focus of this workshop is on uncovering and modeling cell heterogeneity that arises by any of the above-mentioned mechanisms – sporadically, programmed, and through evolution. The main topics covered by this workshop are questions related to cell-to-cell noise, cell-state switches and cell differentiation, heterogeneity in immune responses, cancer evolution, and disease progression and heterogeneity.

2.1. *Cell-to-cell noise*

The stochastic nature of gene expression leads to cell-to-cell differences in protein level, commonly referred to as noise. Expression noise can be disadvantageous, by affecting the precision of performing biological functions, but it may also be advantageous by enabling heterogeneous stress-response programs to environmental changes [6]. Therefore, various genes and gene groups might display various levels of expression noise. Importantly, gene expression is a multi-step process and the stochasticity of its individual steps, including transcription and translation, contributes to the resulting variability. Untangling different components of expression

noise is highly nontrivial and requires a concerted effort of experiment and computational modeling.

2.2. *Cell state switches and cell differentiation*

Heterogeneity has profound implications for cellular differentiation, in which cells must commit to one of a finite number of possible cell states. It has long been appreciated that cells must have molecular mechanisms for counteracting fluctuations in both environmental conditions and cellular components to reliably affect developmental programs [7]. However, recent work suggests that utilization of heterogeneity in the activation of cell differentiation programs in a population of cells can be evolutionarily advantageous. Such advantages are being found in a broad range of biological systems. For example, combined experimental and computational work in *B. subtilis* has begun to characterize the benefits of cell-to-cell variability for the survival of prokaryote populations [4,8]. The importance of heterogeneity in differentiation programs in higher organisms is also being highlighted by recent advances in single cell technologies. New data from analysis of both embryonic stem cells (e.g., [9]) and adult stem cells (e.g., [10]) highlight the necessity for novel computational approaches to provide a deeper understanding of the role of heterogeneity in these important areas of biology and medicine.

2.3. *Heterogeneity in immune responses*

Innate and adaptive immune responses depend on the proper utilization and regulation of cellular heterogeneity. Recognition of a wide variety of antigens necessitates a heterogeneous population of immune cells. However, if the heterogeneity is too great, discrimination between "self" and "other" antigens may be compromised leading to autoimmune diseases. Advances in flow cytometry and single cell proteomics are beginning to elucidate the molecular mechanisms governing the proper regulation on this heterogeneity [11]. Complementary approaches in computational methods to analyze this important aspect of cellular immunity will be key to further our understanding of immune responses in healthy and disease states [12].

2.4. *Cancer evolution*

A tumor is formed from a heterogeneous mass of cells with different complements of somatic mutations and possibly different differentiated states. The implications of this heterogeneity for cancer progression and treatment are not well understood. While mutational heterogeneity could merely be a consequence of the somatic mutation process that drives cancer development, heterogeneity may itself be an important, or even essential, contributor to tumor evolution [5,13,14]. Single-cell analyses are necessary to understand the role heterogeneity plays in cancer evolution. For instance, single-cell sequencing is likely to profoundly impact cancer diagnostics and prognosis through the detection of rare tumor cells or through the monitoring of circulating tumor cells [15]. Single-cell next generation sequencing can also be used to investigate tumor subpopulations and to delineate the differences between primary and metastasic tumors [5,14].

3. Workshop Contributions

The workshop includes two invited speakers and five accepted submissions.

Dana Pe'er (invited speaker) is an Associate Professor in the Department of Computer Science and the Department of Biological Sciences at Columbia University. Dr. Pe'er received her doctoral degree in computational biology from the Hebrew University. Her research focuses on understanding the organization, function and evolution of molecular networks. Dr. Pe'er and her team develop computational methods to integrate diverse high-throughput genomic data and to discover the general principles governing cellular signal processing, propagation of small changes in regulatory networks and how they alter cellular functioning that can lead to diseases such as autoimmune disease and cancer.

Sylvia Plevritis (invited speaker) is an Associate Professor in the Department of Radiology in the Stanford School of Medicine. Dr. Plevritis received her doctoral degree in Electrical Engineering and master's degree in health services from Stanford University. Her research focuses on computational and mathematical modeling of cancer biology and cancer outcomes. Using diverse sources of information from genomic and proteomic data to clinical data, her laboratory was able to infer natural histories of cancer, to estimate cell subpopulations after cancer treatment, and to identify perturbations of molecular networks during different stages of cancer. Dr. Plevritis is also the Program Director of the Stanford Center for Cancer Systems Biology (CCSB), and the co-Section Chief of Information Sciences in Imaging at Stanford (ISIS).

Julian Candia, Jayanth Banavar and Wolfgang Losert. "From molecules to cells to organisms: understanding health and disease with multidimensional single-cell methods." This works describes a newly developed framework to investigate multicolor data from fluorescence-activated cell sorting (FACS). The method integrates several approaches to gain different perspectives on the data: singular value decomposition to reduce data representation, machine learning to separate patients into classes and improve diagnosis, and network analysis to infer cell subpopulations.

Michael Januszyk, Jason P. Glotzbach, Michael Sorkin, Atul J. Butte and Geoffrey C. Gurtner. "Automated Functional Profiling of Progenitor Cell Populations using High-Resolution Single Cell Gene Expression Data." The authors show how information from thousands of publicly available microarray datasets of gene expression can be used to increase the power of single cell gene expression data analysis, which has high-resolution but is limited to only several dozen target genes that can be measured at the same time. The method is based on higher-order covariance of gene expression retrieved from the Gene Expression Omnibus (GEO) database, and functional classification based on Gene Ontology. Applied to murine bone marrow-derived mesenchymal stem cells, the method finds significant associations between cell subpopulations and known functional categories.

Layla Oesper, Ahmad Mahmoody and Ben Raphael. "Estimating Tumor Clonal Populations from Copy Number Data." The authors introduce an algorithm to infer tumor subpopulations directly from high-throughput DNA sequencing data. Their method decomposes a mixture of normal cells, clonal cells, and sub-clonal populations to maximize the probability of observed data using techniques from convex optimization. The algorithm was applied to nine

breast cancer samples and successfully recovered the heterogeneity of these tumor cell populations.

Kyoungmin Roh and Stephen Proulx. "The role of positive and negative feedback loops of p53 pathway." This work describes simulations of different scenarios built from deterministic p53 feedback loops, both negative and positive, based on Puszynski's model. Using simulated annealing to find the optimal response of p53 to DNA damage, the authors demonstrate the ability of p53 feedback loops to reduce the chance of cell apoptosis, making the cell less sensitive to DNA damage. The analysis provides new insight on p53 feedback loops and DNA damage pathways.

Damian Wojtowicz, Daniela Ganelin, Raheleh Salari, Jie Zheng, David Lavens, Yitzhak Pilpel and Teresa Przytycka. "Teasing apart sources of stochastic variations in eukaryotic gene expression." In this project, the authors develop a novel computational approach to delineate the relative impact of transcription and translation processes on cell-to-cell variations, noise, in protein abundance, and apply it to large-scale gene expression data from yeast (Newman et al., 2006). Interestingly, they show that translation-related genomic features, such as codon usage and 5'UTR secondary structure, have higher impact on noise than previously appreciated.

4. Acknowledgments

We would like to thank Dana Pe'er and Sylvia Plevritis for kindly accepting our invitation and generously supporting our workshop. M.G.K was supported by the National Institutes of Health (NIH), 1K22CA143148. E.B. was supported in part by Intramural Research Program of the NIH, NCI, Center for Cancer Research. T.M.P. and D.W. research was supported in part by the Intramural Research Program of the NIH, National Library of Medicine. B.J.R. is supported by the National Institutes of Health grant R01HG5690, a CAREER Award from the National Science Foundation (CCF-1053753), and Burroughs Wellcome Career Award at the Scientific Interface.

References

1. Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, et al. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441: 840-846.
2. Marko NF, Quackenbush J, Weil RJ (2011) Why is there a lack of consensus on molecular subgroups of glioblastoma? Understanding the nature of biological and statistical variability in glioblastoma expression data. PLoS One 6: e20826.
3. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. Nature 459: 428-432.
4. Suel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. Science 315: 1716-1719.
5. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, et al. (2011) Tumour evolution inferred by single-cell sequencing. Nature 472: 90-94.
6. Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467: 167-173.
7. Waddington CH (1942) Canalization of development and the inheritance of acquired characters. Nature 150: 563-565.

8. Kuchina A, Espinar L, Garcia-Ojalvo J, Suel GM (2011) Reversible and noisy progression towards a commitment point enables adaptable and reliable cellular decision-making. PLoS Comput Biol 7: e1002273.
9. Phanstiel DH, Brumbaugh J, Wenger CD, Tian S, Probasco MD, et al. (2011) Proteomic and phosphoproteomic comparison of human ES and iPS cells. Nat Methods 8: 821-827.
10. Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, et al. (2011) Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell 144: 296-309.
11. Feinerman O, Jentsch G, Tkach KE, Coward JW, Hathorn MM, et al. (2010) Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. Molecular systems biology 6: 437.
12. Coward J, Germain RN, Altan-Bonnet G (2010) Perspectives for computer modeling in the study of T cell activation. Cold Spring Harb Perspect Biol 2: a005538.
13. Michor F, Polyak K (2010) The origins and implications of intratumor heterogeneity. Cancer Prev Res (Phila) 3: 1361-1364.
14. Navin N, Krasnitz A, Rodgers L, Cook K, Meth J, et al. (2010) Inferring tumor progression from genomic heterogeneity. Genome Res 20: 68-80.
15. Navin N, Hicks J (2011) Future medical applications of single-cell sequencing in cancer. Genome Med 3: 31.