Methods 85 (2015) 1–2

Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

Guest Editor's Introduction

Inferring gene expression regulatory networks from high-throughput measurements

While molecular biology has meticulously and successfully built the catalog of components for a large number of cell types, recent technological developments have broadened the spectrum and resolution of measurement techniques. These have led to a flourishing of a number of subfields, including mathematical biology, computational biology, systems biology, synthetic biology, etc. Although the precise definitions and boundaries of these partially overlapping subfields can be debated, it is clear that the general availability of high-throughput approaches of increasing quantitative accuracy has shifted the focus away from single components toward quantitative modeling of whole-cell behaviors. The vision behind this volume was to illustrate some of these approaches and the insights that they have brought to the field. We focused on gene expression, which in eukaryotic cells is a very complex process of many steps, all of which are subject to regulation. We hope that readers find this perspective motivating. I am grateful to the contributing authors that participated in this endeavor, to Dr. Adolf for the invitation to edit such an issue, and to Tiffany Hicks and Liz Weishaar for their great help in seeing the project to completion.

Gene expression starts with transcription, the synthesis of the pre-mRNA by RNA polymerase II, which typically occurs in bursts [1]. Analysis of time-lapse microscopy data with stochastic models of gene expression allows the inference of transcriptional kinetic parameters for individual genes at the single-cell level [2]. To enable such studies, Blanchoud and colleagues have developed CAST (Cell Automated Segmentation and Tracking), a tool that performs automated detection and tracking of cell nuclei—including through cell division—as well as quantification of gene expression [3]. Another approach to the inference of transcriptional kinetics makes use of single-molecule RNA fluorescence in situ hybridization (smRNA-FISH) [4] to obtain the distribution of RNA molecules across cells. An approximation of the solution to the chemical master equation, called finite state projection, can then fit parameters of specific models of transcription dynamics. Illustrations of this approach are shown in the paper by Munsky, Fox, and Neuert [5]. Bronstein, Zechner, and Koeppel discuss more general approaches to inferring parameters of biochemical reaction networks. Their work introduces a particular class of algorithms, which employ marginalization of extrinsic factors, to infer parameters of reaction networks based on data from heterogeneous cell populations [6].

Eukaryotic genes generally have a multi-exon structure, and removal of introns from pre-mRNAs by the spliceosome is necessary for the production of mature mRNAs. Studying this process, which takes place largely co-transcriptionally [7], requires appropriate methods for profiling and quantifying nascent mRNAs. Herzel and Neugebauer [8] discuss these methods.

Global quantification of transcript levels in single cells is an area of very active methodology development. The two main classes of approaches that have been proposed so far are based on either sequencing or imaging of individual transcripts with fluorescent oligomers. The paper by Stoeger and colleagues [9] describes the experimental setup for imaging-based quantification of transcript abundance in single cells, with specific regard to cellular subcompartments. Although these methods are still in flux, many groups have started to investigate the heterogeneity in gene expression, both between cells of the same type and across cell types [10]. It has thus become apparent that a substantial degree of heterogeneity can be attributed to differences in the physiological state of the cells, which is due, for example, to their being in different phases of the cell cycle. The work of Scialdone and colleagues addresses this, investigating methods for annotating the cell-cycle phase based on the gene-expression pattern of individual cells [11].

The transcriptional activity of individual genes and the resulting mRNA abundance is modulated by transcription factors, and these factors have a cell-type-specific pattern of expression and activity. As genome sequences became available for a multitude of species, comparative-genomics-based methods have become increasingly accurate in predicting transcription factor binding sites genome-wide. Combining these predictions with mRNA expression data, the so-called motif activity response analysis [12] aims to identify the key drivers of gene expression in specific samples or cell types. The paper by Pemberton-Ross, Pachkov, and van Nimwegen now extends this method to allow inference of causal regulatory interactions from time series expression data [13].

Finally, three papers describe approaches to analyzing post-transcriptional steps of gene expression. The work by Aeschimann and colleagues illustrates an approach to estimating ribosome occupancy, and thereby protein synthesis rates, transcriptome-wide [14]. The paper by Breda and team uses various types of experimental data to infer the strength of interaction between miRNAs (small RNAs that act within ribonucleoprotein complexes to repress gene expression) and their mRNA targets [15]. Finally, Ahrné and colleagues describe a mass spectrometry method that uses isobaric tandem mass tags to accurately quantify protein abundance while achieving high coverage of the proteome [16].

http://dx.doi.org/10.1016/j.ymeth.2015.07.006

We hope that this collection of papers, covering the entire spectrum of analyses, from high-throughput experiments to algorithms for the analysis of the generated data and mathematical models that provide mechanistic insights, will serve as a good reference, and that it will inspire the development of novel methodologies for the analysis of cellular systems, particularly at the single-cell level.

References

[1] J.R. Chubb, T.B. Liverpool, Bursts and pulses: insights from single cell studies into transcriptional mechanisms, Curr. Opin. Genet. Dev. 20 (2010) 478–484.

[2] D.M. Suter, N. Molina, D. Gatfield, K. Schneider, U. Schibler, F. Naef, Mammalian genes are transcribed with widely different bursting kinetics, Science 332 (2011) 472–474.

[3] S. Blanchoud, D. Nicolas, B. Zoller, O. Tidin, F. Naef, CAST: an automated segmentation and tracking tool for the analysis of transcriptional kinetics from single-cell time-lapse recordings, Methods 85 (2015) 3–11.

[4] S. Itzkovitz, A. van Oudenaarden, Validating transcripts with probes and imaging technology, Nat. Methods 8 (2011) S12–9.

[5] B. Munsky, Z. Fox, G. Neuert, Integrating single-molecule experiments and discrete stochastic models to understand heterogeneous gene transcription dynamics, Methods 85 (2015) 12–21.

[6] L. Bronstein, C. Zechner, H. Koeppl, Bayesian influence of reaction kinetics from single-cell recordings across a heterogeneous cell population, Methods 85 (2015) 22–35.

[7] K.-M. Lee, W.-Y. Tarn, Coupling pre-mRNA processing to transcription on the RNA factory assembly line, RNA Biol. 10 (2013) 380–390.

[8] L. Herzel, K.M. Neugebauer, Quantification of co-transcriptional splicing from RNA-Seq data, Methods 85 (2015) 36–43.

[9] T. Stoeger, N. Battich, M.D. Herrmann, Y. Yakimovich, L. Pelkmans, Computer vision for image-based transcriptomics, Methods 85 (2015) 44–53.

[10] D.A. Jaitin, E. Kenigsberg, H. Keren-Shaul, N. Elefant, F. Paul, I. Zaretsky, et al, Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types, Science 343 (2014) 776–779.

[11] A. Scialdone, K.N. Natarajan, L.R. Saraiva, V. Proserpio, S.A. Teichmann, O. Stegle, J.C. Marioni, F. Buettner, Computational assignment of cell-cycle stage from single-cell transcriptome data, Methods 85 (2015) 54–61.

[12] P.J. Balwierz, M. Pachkov, P. Arnold, A.J. Gruber, M. Zavolan, E. van Nimwegen, ISMARA: automated modeling of genomic signals as a democracy of regulatory motifs, Genome Res. 24 (2014) 869–884.

[13] P.J. Pemberton-Ross, M. Pachkov, E. van Minwegen, ARMADA: Using motif activity dynamics to infer gene regulatory networks from gene expression data, Methods 85 (2015) 62–74.

[14] F. Aeschimann, J. Xiong, A. Arnold, C. Dieterich, Helge Großhans, Transcriptome-wide measurement of ribosomal occupancy by ribosome profiling, Methods 85 (2015) 75–89.

[15] J. Breda, A.J. Rzepiela, R. Gumienny, E. van Nimwegen, M. Zavolan, Quantifying the strength of miRNA-target interactions, Methods 85 (2015) 90–99.

[16] E. Ahrné, A. Martinez-Segura, A.P. Sayed, A. Vina-Vilaseca, A.J. Gruber, S. Marguerat, Al Schmidt, Exploiting the multiplexing capabilities of tandem mass tags for high-throughput estimation of cellular protein abundances by mass spectrometry, Methods 85 (2015) 100–107.

Mihaela Zavolan
Biozentrum, University of Basel, Switzerland
E-mail address: mihaela.zavolan@unibas.ch

Available online 15 July 2015