Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Systems Biology

Review

Computational methods to dissect gene regulatory networks in cancer

Archana S. Iyer, Hatice U. Osmanbeyoglu and Christina S. Leslie

Abstract

Cancer is a disease of gene dysregulation, where cells acquire genetic alterations that drive aberrant signaling. These alterations adversely impact transcriptional programs and cause profound changes in gene expression. Large international consortia have generated massive tumor profiling data sets across many cancer types, collecting mutation and copy number variation, mRNA expression, and in some cases epigenomic and proteomic profiles. An overarching goal of these tumor-profiling efforts is to identify genes that are essential drivers of cellular processes in cancer. Here we review diverse computational methodologies that have sought to interpret somatic alterations and gene expression data through models of gene regulatory networks. Early work in the field used expression data alone to infer regulatory networks, and expression-only network inference continues to be an active area of research. Once catalogs of somatic mutations and copy number variations became available, another class of methods tried to interpret these alterations either in terms of a prior interaction network or an inferred regulatory network. More recently, the tools of regulatory genomics have been applied to cancer data sets, integrating regulatory sequence information and epigenomic data with gene expression, often through supervised methods. Finally, newer cross-cutting algorithms link upstream signaling changes – through a prior network or proteomics data – to downstream transcriptional programs and interpret somatic alterations in terms an integrative model. Currently, clinical trials use specific somatic alterations alone to direct patients towards pathway-targeted therapies. Ultimately, methods that interpret patient mutations profiles through the lens of gene regulatory networks may be better able to identify patient populations who will benefit from therapy.

Addresses

Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

Corresponding author: Leslie, Christina S (cleslie@cbio.mskcc.org)

Current Opinion in Systems Biology 2017, **2**:114–121

This review comes from a themed issue on **Regulatory and metabolic networks (2017)**

Edited by **Sang Yup Lee** and **Christina Leslie**

For a complete overview see the Issue and the Editorial

Available online 21 April 2017

http://dx.doi.org/10.1016/j.coisb.2017.04.004

2452-3100/© 2017 Published by Elsevier Ltd.

Keywords

Cancer genomics, Transcriptional regulatory networks, Machine learning.

Interpreting tumor profiling data sets through the lens of gene regulation

Cancer cells acquire the capacity for uncontrolled growth and proliferation through a multistep process that results in the deregulation of homeostatic signals [1,2]. The genetic alterations that cause deregulation eventually causes aberrant gene expression and genomic instability. In the last decade, a monumental effort to molecularly profile tumors was undertaken by consortia such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC). These groups and others have generated large gene expression compendia of different tumor types and catalogued their genetic alterations (recurrent mutations and copy number variations (CNVs)) [3–5]. Early computational analyses focused on subtype identification within tumor cohorts, usually through clustering patient gene expression profiles. Meanwhile, somatic alterations were mapped to known cellular processes and also associated to specific cancer subtypes. Together, these efforts have led to the molecular characterization of different types of tumors and support the idea that cancer is not a single disease, but a group of diseases [5].

Profiling tumors for genetic alterations has the potential to improve treatment options available to patients [6]. While considerable effort is being spent on patient stratification to achieve optimal therapeutic outcome, a mechanistic understanding of the drivers of cancer is also crucial. This is particularly important in the era of pathway-targeted therapies, where not all patients with a "targeted mutation" respond, and the tumor eventually develops resistance mechanisms to evade therapy. More recently, additional multimodal sets including profiling for chromatin and phosphoprotein expression have enabled the integration of transcriptional states with upstream signaling pathways. Ultimately, interpreting patient genetic alterations within context-specific regulatory networks will facilitate therapeutic decisions for each individual.

Numerous and diverse computational algorithms have been developed to mine and interpret massive tumor profiling data sets. In this review, we focus on methods that interpret cancer data sets through the lens of gene regulation, linking genetic alterations to changes in the transcriptional program. Namely, we highlight methods that model transcriptional regulatory networks in tumors — where the altered activity of transcription factors changes the expression of target genes and drives aberrant cellular phenotypes — and thus identify drivers of gene regulation. We describe the development of these approaches in four categories, roughly based on the chronology of when each approach was first applied in cancer: (1) network inference based on gene expression data; (2) interpretation of genetic alterations with respect to a prior network; (3) network integration of 'cis' regulatory information with tumor gene expression; and (4) cross-cutting algorithms that link changes in upstream signaling, through a prior network or proteomics data, to downstream transcriptional programs and thus decipher the regulatory impact of somatic alterations (Figure 1). While several of these categories are areas of active computational research outside of cancer, we highlight recent methods in each class that have advanced our understanding of cancer regulatory networks. We further discuss the challenges associated with transforming data from heterogeneous tumor profiles into meaningful biological insights.

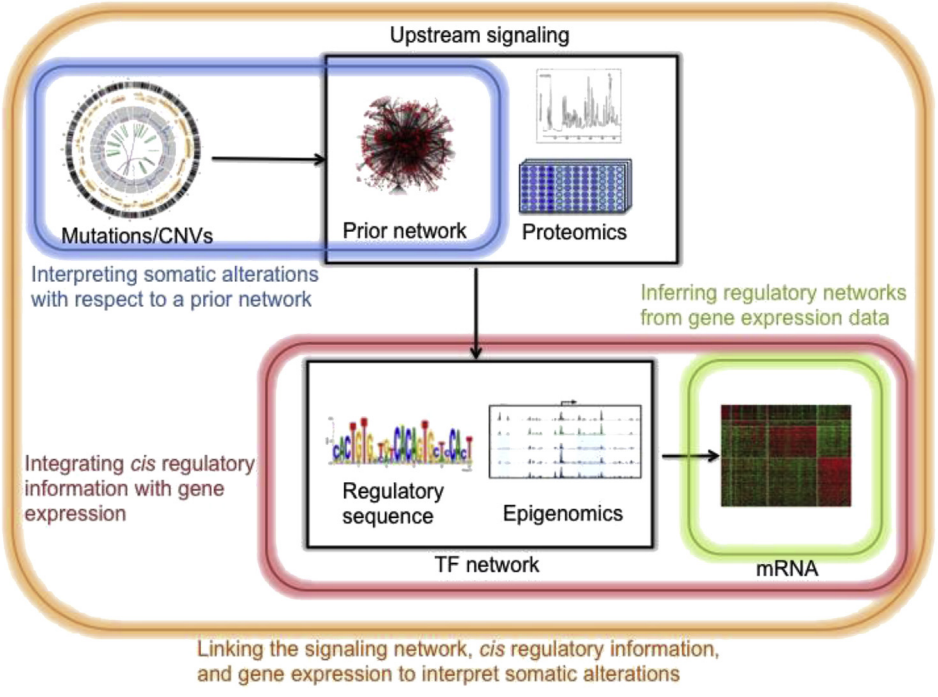
Network inference based on gene expression data

Since the completion of the Human Genome Project in 2001, gene expression profiling has advanced through various microarray platforms to standard RNA sequencing technology today. Historically, tumor profiles were among the first large-scale gene expression data sets to become available, and early algorithms inferred regulatory networks in cancer solely based on this data. We describe three groups of network inference methods based on tumor gene expression data: unsupervised network inference, regression-based network inference, and probabilistic module learning.

Unsupervised network inference

One of the first human transcriptional networks was built for B cell malignancies using ARACNe, an information theoretic network inference approach [7,8]. The algorithm computes pairwise mutual information between transcription factors (TFs) and potential target genes and tests if these interactions are significant. The rationale is that high mutual information between a given TF and other genes may suggest a transcriptional regulatory relationship. Since the information content between direct interactions will be higher than indirect ones, the network is further pruned to enrich for direct interactions through the data processing inequality. To infer upstream signaling modulators, an approach based

Figure 1



Methods that integrate data from tumor profiles to develop a molecular framework of cellular interactions and infer drivers of gene regulation: (a) Network inference based on gene expression data (b) Interpretation of genetic alterations with respect to a prior network (c) Network integration of 'cis' regulatory information with tumor gene expression, and (d) Cross-cutting algorithms that link upstream signaling changes to downstream alterations and transcriptional programs, through a prior network or proteomics data.

Download English Version:

https://daneshyari.com/en/article/8918196

Download Persian Version:

https://daneshyari.com/article/8918196

Daneshyari.com