Methods 124 (2017) 36–45

Contents lists available at ScienceDirect

Methods

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

Reconstructing context-specific gene regulatory network and identifying modules and network rewiring through data integration

Tianle Ma, Aidong Zhang *

Department of Computer Science and Engineering, University at Buffalo (SUNY), Buffalo, NY 14260-2500, United States

ARTICLE INFO

Article history:
Received 2 April 2017
Accepted 5 May 2017
Available online 18 May 2017

Keywords:
Gene regulatory network
Context-specific network
Data integration
Transcriptional profiling
Network rewiring
Regulatory module
Collaborative clustering

ABSTRACT

Reconstructing context-specific transcriptional regulatory network is crucial for deciphering principles of regulatory mechanisms underlying various conditions. Recently studies that reconstructed transcriptional networks have focused on individual organisms or cell types and relied on data repositories of context-free regulatory relationships. Here we present a comprehensive framework to systematically derive putative regulator-target pairs in any given context by integrating context-specific transcriptional profiling and public data repositories of gene regulatory networks. Moreover, our framework can identify core regulatory modules and signature genes underlying global regulatory circuitry, and detect network rewiring and core rewired modules in different contexts by considering gene modules and edge (gene interaction) modules collaboratively. We applied our methods to analyzing Autism RNA-seq experiment data and produced biologically meaningful results. In particular, all 11 hub genes in a predicted rewired autistic regulatory subnetwork have been linked to autism based on literature review. The predicted rewired autistic regulatory network may shed some new insight into disease mechanism.

Published by Elsevier Inc.

1. Introduction

Gene regulatory network is crucial to understand human biology and disease mechanism, yet fine-grained, context-specific regulatory circuitry remain largely unknown. Large-scale projectes such as ENCODE [1] and Roadmap Epigenomics [2] have used genomic and epigenomic profiling to pinpoint genome-wide regulatory elements. However, they are still largely constrained by a limited number of cell types and epigenomic markers studied, and the results cannot be directly translated into a context-specific gene regulatory network, which can be viewed as a graphical representation of causal relationships between regulator genes and their targets.

Computational methods have been developed to construct gene regulatory networks by majorly using transcriptional profiling [3,4] or integrating with multi-omic data [5–7]. However, for most small-scale projects studying genetic diseases, comprehensive multi-omic profiling is not often performed due to budget constraint, while only transcriptional profiling is available in many cases.

The general framework of analyzing transcriptional data in genetic disease studies is first identifying differentially expressed genes/transcripts, and then performing network and pathway analysis, including gene set enrichment analysis and co-expression network analysis to prioritize disease genes and identify disease gene modules, and finally detecting disrupted pathways that contribute to etiology. If data is available, transcriptomic profiling can be integrated with mutation data to infer driver mutations and pathway expression levels. For example, CONEXIC [8] used Bayesian network to integrate gene expression and copy number variation data to uncover drivers of cancer. PARADIGM [9] integrates gene expression, copy number variation, and pathway databases to infer patient-specific pathway expression levels in cancer patients. WGCNA [10] is a widely used R package for co-expression analysis. Co-expression network is undirected, and thus does not represent causal relationships.

Cytoscape [11] is a powerful platform for network visualization and analysis. There are many useful Cytoscape plugins available for regulatory network analysis (http://apps.cytoscape.org/apps/with_tag/regulatorynetworks). For example, CyTargetLinker [12] can automatically add known regulatory interactions from public databases to user-defined biologically networks. iRegulon [13] majorly relies on motif information from public data repositories for constructing gene regulatory networks. CyTRANSFINDER [14] can extract signaling transduction pathways involving user-defined regulators and targets from public databases.

* Corresponding author.
E-mail addresses: tianlema@buffalo.edu (T. Ma), azhang@buffalo.edu (A. Zhang).

http://dx.doi.org/10.1016/j.ymeth.2017.05.002
1046-2023/Published by Elsevier Inc.

However, these methods still have limitations: some of them only use transcriptional profiling for inference [3,4], which is too restricted. Some methods rely on multi-omic data that may not be available in small-scale genetics studies [5], or they use complementary omic data from public repositories that are context-free [6]. Many Cytoscape plugins for constructing regulatory networks are for exploratory analysis and can not infer weighed context-specific networks [12–14].

Currently most publicly available gene regulatory networks are not context-specific: that is, the regulatory relationships are derived from various context but the contextual information are largely lost. However, constructing context-specific gene regulatory network is crucial to reveal regulatory circuit rewiring under various conditions.

Here we present a comprehensive framework (Fig. 1) to reconstruct context-specific regulatory network (RCRN) and identify core gene regulatory modules and network rewiring by combining transcriptomic data collected in a specific study and publicly available gene regulatory networks. Specifically, we use a multi-step process (Fig. 2) to construct and refine context-specific regulatory network from a global regulatory network assembled from multiple data repositories. We also present a new perspective for identifying core regulatory modules and network rewiring by integrating node clustering and edge clustering on predicted weighted regulatory networks (Fig. 3).

We applied our method to an autism RNA-seq experiment dataset. Our predicted autistic regulatory network rewiring has revealed many known risk disease genes based on literature review. Moreover, the predicted rewired network provide rich information about the regulatory circuitry, which may shed new insight on disease mechanism. The detailed predictions of autistic regulatory network as well as source code and data are available at https://www.github.com/beautyofweb/RCRN for reproducible research.

2. Methods

2.1. Overview

With abundant epigenomic data sources from publicly funded projects, researchers have assembled a few data repositories of gene regulatory relationships. This makes it possible to leverage a global regulatory network to infer a fine-grained, context-specific network with transcriptomic data collected in individual studies (many small-scale genetics studies leverage RNA-seq for transcriptomic profiling without comprehensive epigenomic profiling). Our framework (Fig. 1) consists of two major components: context-specific gene regulatory subnetwork construction (Fig. 2),

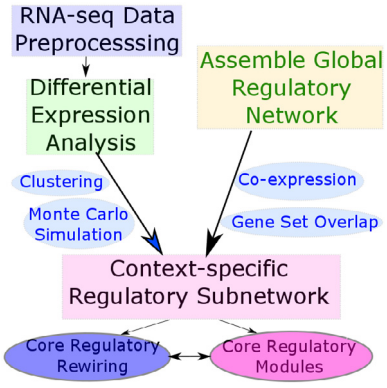

Fig. 1. Overview of our framework.

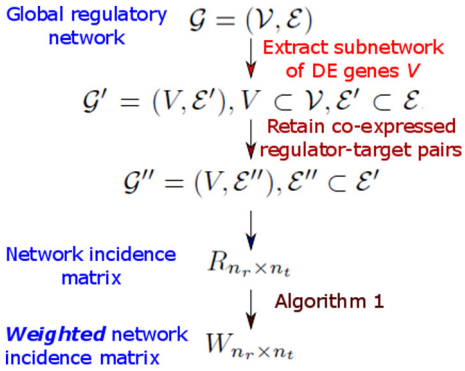

Fig. 2. Flowchart of constructing weighted context-specific regulatory network.

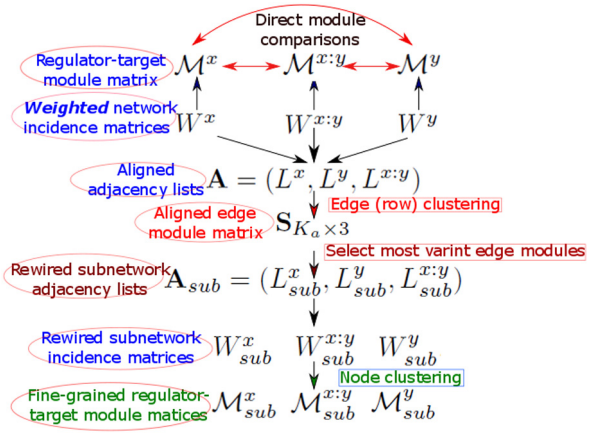

Fig. 3. Flowchart of identifying core rewired regulatory network modules.

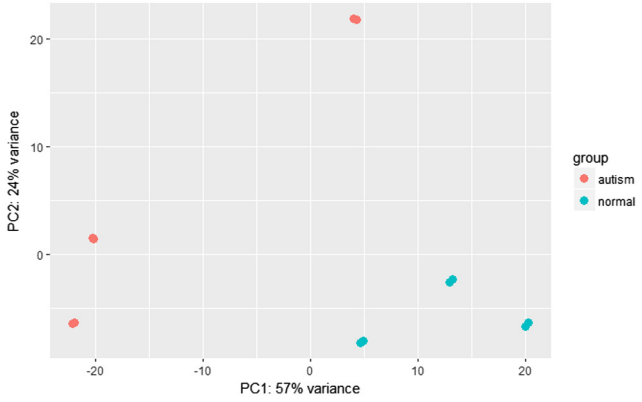

Fig. 4. Sample PCA plot using 7101 DE genes.

and rewired regulatory subnetwork and network modules identification (Fig. 3).

The overall framework is shown in Fig. 1. We first use bioinformatic pipelines to process RNA-seq data to quantify gene expression. We specifically point out some aspects in preprocessing sometimes overlooked by researchers.

In order to use transcriptomic profiling to construct context-specific regulatory networks, we need to identify regulator-target edges. Perturbation techniques are mostly often used to identify causal relationships between regulators and targets, but the experimental cost is high. Here we identify a set of differentially expressed genes to "simulate perturbation". For instance, an