Bioinformatics, 31(23), 2015, 3865–3867
doi: 10.1093/bioinformatics/btv450
Advance Access Publication Date: 30 July 2015
Applications Note

OXFORD

Systems biology

stringgaussnet: from differentially expressed genes to semantic and Gaussian networks generation

Emmanuel Chaplais[1] and Henri-Jean Garchon[1,2,*]

[1]Inserm U1173 and University of Versailles Saint-Quentin, 78180 Montigny-le-Bretonneux, France and [2]Division of Genetics, Ambroise Paré Hospital, 92100 Boulogne-Billancourt, France

[*]To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on June 25, 2015; revised on July 23, 2015; accepted on July 24, 2015

Abstract

Motivation: Knowledge-based and co-expression networks are two kinds of gene networks that can be currently implemented by sophisticated but distinct tools. We developed stringgaussnet, an R package that integrates both approaches, starting from a list of differentially expressed genes.
Contact: henri-jean.garchon@inserm.fr
Availability and implementation: Freely available on the web at http://cran.r-project.org/web/packages/stringgaussnet.

1 Introduction

Analysis of genes differentially expressed (DE) depending on a condition has become a standard procedure in current biology. However, identification of biologically relevant DE genes is far from being trivial. Yet efficient prioritization of DE genes is an essential step before undertaking rate-limiting wet lab experiments (Smyth, 2004). In this regard, the network theory appears as a powerful framework. The aim is to connect genes (the nodes) by means of their interactions, the edges (Dong and Horvath, 2007). These interactions may be based on prior knowledge, curated and stored in databases, or extracted from the experimental dataset, e.g. using coexpression information (Cotney *et al.*, 2015; Lin *et al.*, 2015; Verfaillie *et al.*, 2015; Xue *et al.*, 2014). Sophisticated but distinct tools are available to implement either one of these approaches separately. We introduce stringgaussnet, an R package that allows inferring gene networks starting from a list of DE genes by integrating both of these approaches with ease and flexibility.

2 Initial objects and example data

Stringgaussnet requires two data frames that store, one the expression measurements, the other DE gene statistics. They are combined to create an object of class DEGeneExpr. Expression data are usually normalized for efficient correlation computation. In the example data, we provided transcriptomic profiles of monocyte-derived dendritic cells from nine patients affected with ankylosing spondylitis and 10 healthy controls. Transcriptomic data were obtained using microarrays. Gene expression levels in patients and in controls were then compared with LIMMA (Talpin *et al.*, 2014). We limited the number of DE genes to 75.

3 String network construction

Stringgaussnet allows constructing a protein–protein interaction network using the DE gene names (Ensembl IDs or HGNC symbols) with the use of the STRING API with specific URIs (Franceschini *et al.*, 2013; Szklarczyk *et al.*, 2015). The number of additional nodes can be set by the user, and these are useful to detect indirect relationships between input genes. The default value is twice the number of initial DE genes. Different species can be curated, the default being *Homo sapiens*. This process generates an object of class STRINGNet, a network with multiple edges depending on sources and combined scores. One can select specific sources of interactions and filter on the scores given by STRING. Stringgaussnet can calculate a new combined score, based on the algorithm provided for STRING version 8 (http://string-stitch.blogspot.fr/2010/03/combining-scores-right-way.html).

4 Short paths from string networks

The generated network can be large and dense. As a STRINGNet object, it can be reduced by computing shortest paths between genes

of a user's list. To this aim, combined scores S are converted to distances D for each node pair i with $D_i = \max(S_i) + 1 - S_i$, where $\max(S_i)$ is the maximum of S over all interactions. Shortest paths between each pair of nodes are computed with the Dijkstra's algorithm. This method creates an object of class ShortPathSTRINGNet, with unique edges giving distances and intermediates as attributes. It is also possible to filter edges on D.

5 SIMoNe network inference

Alternatively, stringgaussnet can help infer Gaussian networks from expression data, using the R package SIMoNe (Chiquet *et al.*, 2009). Default parameters are set for easy use. The number of edges that is selected is the mean of the number of edges corresponding to maximal AIC and BIC scores. By default, the algorithm computes a network under two models, with or without clustering constraints, and picks edges common to both models. Stringgaussnet performs a Spearman's test for each inferred edge. This generates an object of class SIMoNeNet, a network of unique edges including theta, Spearman's rho and P-value as attributes. One can filter on edge attributes, notably on Spearman's rho.

6 WGCNA approach to compare results

SIMoNe is a powerful tool to infer unsupervised Gaussian networks. For comparison, we also propose the use of the popular WGCNA package (Langfelder and Horvath, 2008). Stringgaussnet performs a Spearman's test between all pairs of genes. The respective rho coefficients are converted to similarity scores $\sigma = (1 + \rho)/2$ that are then converted to adjacency scores $A = 1/(1 + e - \alpha^*(\sigma - 0.5))$ where α is the soft power threshold; its default is set to 8. A method helps adjust this parameter by plotting relationships between A and rho. Then a filtering step is done using a threshold t, A being superior to t or inferior to $1 - t$. Dissimilarity and module computation are not implemented, because the main purpose is to compare with SIMoNe results. The network is saved in a WGCNANet class object. A function is provided to compare networks inferred by SIMoNe and WGCNA, with a Venn diagram displaying the numbers of shared and specific edges, and a series of plots showing correlation coefficients of selected interactions.

7 Adding annotations to genes

While networks are being generated, it is possible to add gene annotations as node attributes. They are of two kinds, including genomic localization and brief gene description, using the biomaRt R package, and cellular component terms, using the GO.db package. In addition, genes are ranked depending on the localization of their protein product, from nuclear, the most relevant, to extracellular, to plasma membrane and last to cytoplasm.

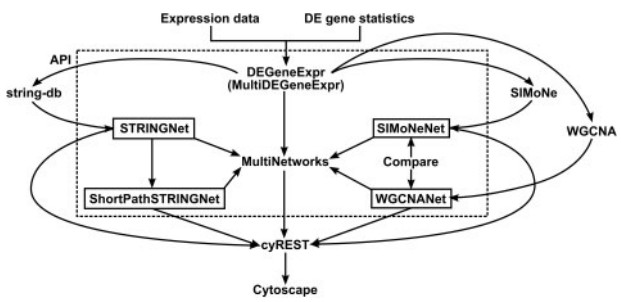

Fig. 1. Stringgaussnet operating principle. Starting from expression data and DE gene statistics, both semantic and Gaussian networks can be inferred and then exported into Cytoscape. The package environment is circumscribed by the dashed rectangle

8 Automatic network creations in one step

An overlay of functions allows the user to create multiple networks in only one step, with all options configurable in the same method. One can create multiple Gaussian networks from the same DEGeneExpr object, depending on a grouping factor and for a given list of genes. The package then allows comparing networks inferred for multiple levels of the factor, and for the same DE genes list. One can create an object of class MultiDEGeneExpr, a list of DEGeneExpr objects. Then, both kinds of networks, semantic or Gaussian, can be generated for each data set and stored in a MultiNetworks object.

9 Automatic export to Cytoscape

In addition to be exported in standard file formats, generated networks can be imported automatically from R objects into Cytoscape, without requiring an intermediate language (Cline *et al.*, 2007). This is performed in an operating-system independent manner through the plugin cyREST (http://apps.cytoscape.org/apps/cyrest). Stringgaussnet proposes predefined styles for displaying the exported networks; they can be easily customized in Cytoscape or directly in the package.

Acknowledgement

We thank Maxime Breban and Gilles Chiocchia for discussion and support and Christophe Ambroise for advice.

Funding

E.C. was funded by a PhD scholarship from the doctoral school 'Du génome aux organismes'. This work was supported by an ANR grant (2010 GEMISA).

Conflict of Interest: none declared.

References

Chiquet,J. *et al.* (2009). SIMoNe: Statistical Inference for MOdular NEtworks. *Bioinforma. Oxf. Engl.*, **25**, 417–418.

Cline,M.S. *et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, **2**, 2366–2382.

Cotney,J. *et al.* (2015). The autism-associated chromatin modifier CHD8 regulates other autism risk genes during human neurodevelopment. *Nat. Commun.*, **6**, 6404.

Dong,J., and Horvath,S. (2007). Understanding network concepts in modules. *BMC Syst. Biol.*, **1**, 24.

Franceschini,A. *et al.* (2013). STRING v9.1: protein–protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.

Langfelder,P., and Horvath,S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559.

Lin,Y. *et al.* (2015). MiRNA and TF co-regulatory network analysis for the pathology and recurrence of myocardial infarction. *Sci. Rep.*, **5**.

Smyth,G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.*, **3**, Article3.

Szklarczyk,D. *et al.* (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–D452.

Talpin,A. *et al.* (2014). Monocyte-derived dendritic cells from HLA-B27 + axial spondyloarthritis (SpA) patients display altered functional capacity and deregulated gene expression. *Arthritis Res. Ther.*, **16**, 417.

Verfaillie,A. *et al.* (2015). Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat. Commun.*, **6**, 6683.

Xue,J. *et al.* (2014). Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity*, **40**, 274–288.