
Gene expression

iDINGO—integrative differential network analysis in genomics with *Shiny* application

Caleb A. Class, Min Jin Ha*, Veerabhadran Baladandayuthapani and Kim-Anh Do

Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

*To whom correspondence should be addressed.

Associate Editor: Bonnie Berger

Received on August 31, 2017; revised on October 30, 2017; editorial decision on November 14, 2017; accepted on November 28, 2017

Abstract

Motivation: Differential network analysis is an important way to understand network rewiring involved in disease progression and development. Building differential networks from multiple ‘omics data provides insight into the holistic differences of the interactive system under different patient-specific groups. DINGO was developed to infer group-specific dependencies and build differential networks. However, DINGO and other existing tools are limited to analyze data arising from a single platform, and modeling each of the multiple ‘omics data independently does not account for the hierarchical structure of the data.

Results: We developed the iDINGO R package to estimate group-specific dependencies and make inferences on the integrative differential networks, considering the biological hierarchy among the platforms. A *Shiny* application has also been developed to facilitate easier analysis and visualization of results, including integrative differential networks and hub gene identification across platforms.

Availability and implementation: R package is available on CRAN (<https://cran.r-project.org/web/packages/iDINGO>) and *Shiny* application at <https://github.com/MinJinHa/iDINGO>.

Contact: mjha@mdanderson.org

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

The analysis of differential networks has led to a deeper understanding of network rewiring, which explains changing molecular relationships associated with a characteristic of interest, such as disease states or progression, clinical treatments or environmental stress (Bandyopadhyay *et al.*, 2010; Califano, 2011). Most of the previous approaches for differential network analysis have relied on different correlation-based metrics to measure the dependencies between pairs of nodes in a network (Hudson *et al.*, 2009; Reverter *et al.*, 2006). However, these methods are limited to marginal correlation networks (i.e. two nodes at a time) that are estimated separately using observations within each group, and generally do not consider relationships that are conserved across multiple groups. This has been refined in the DINGO framework (Ha *et al.*, 2015) that separates group-specific conditional dependencies into global and group-specific components,

and this method has been shown to improve performance over other existing methods in simulation studies and with real data.

At the same time, integrative analyses (including genomics, epigenomics, proteomics and others) have provided biological and clinical insight into a variety of diseases (Gerstung *et al.*, 2015; Qin, 2008). Differential network analysis of integrative datasets introduces new opportunities, as an understanding of the relationships of elements across platforms can provide a more complete biological understanding of the characteristic of interest. Our integrative approach identifies a set of edges between nodes that are differentially connected between patient groups, including directed edges between platforms and undirected edges within platforms. Using data from additional platforms allows us to adjust for the upstream data, providing a more refined network than the original DINGO method. The resulting network allows us to identify ‘hub nodes’ (nodes with the greatest number of

outgoing or undirected edges) across platforms, which may have the greatest effect on the clinical/grouping variable (Flintoft, 2004).

In this paper, we present the R package ‘iDINGO’ (with accompanying *Shiny* application) as an expansion of the ‘DINGO’ package. This package integrates relationships between different ‘omics levels in the analysis using a chain graph model. Parallelization is implemented to improve computation time, and a multiple-testing correction is also included to improve inference on differential edges. Finally, we introduce a *Shiny* application to facilitate easier analysis and visualization.

2 The iDINGO R package

We can integrate ordered data platforms using the chain graph model. For example, we can integrate microRNA, mRNA and protein data using the assumed ordering,

$$\text{microRNA} < \text{mRNA} < \text{Protein},$$

which means that microRNA can affect mRNA and protein, and mRNA can affect protein, but not vice versa. In this case, we have a set of nodes $V = V_M \cup V_R \cup V_P$, where $V_M = \{M_1, M_2, \dots\}$ is a set of microRNA nodes, $V_R = \{R_1, R_2, \dots\}$ is a set of mRNA nodes and $V_P = \{P_1, P_2, \dots\}$ is a set of protein nodes, and a set of edges E that may contain both directed (\rightarrow) and undirected ($-$) edges between and within V_M , V_R and V_P , respectively. Following the Markov property for chain graphs (Frydenberg, 1990; Lauritzen and Wermuth, 1989), the within- and between-platform conditional independence is defined as follows:

$$(D1) M_i \perp\!\!\!\perp M_j \mid V_M \setminus \{M_i, M_j\}$$

$$(D2) M_i \perp\!\!\!\perp R_j \mid V_M \cup V_R \setminus \{M_i, R_j\}$$

$$(D3) R_i \perp\!\!\!\perp R_j \mid V_M \cup V_R \setminus \{R_i, R_j\}$$

$$(D4) M_i \perp\!\!\!\perp P_j \mid V \setminus \{M_i, P_j\}$$

$$(D5) R_i \perp\!\!\!\perp P_j \mid V \setminus \{R_i, P_j\}$$

$$(D6) P_i \perp\!\!\!\perp P_j \mid V \setminus \{P_i, P_j\},$$

where the conditional dependencies in (D1), (D3) and (D6) encode undirected edges within microRNA, mRNA and protein, respectively, and those in (D2), (D4) and (D5) represent directed edges microRNA \rightarrow mRNA, microRNA \rightarrow protein and mRNA \rightarrow protein. In our iDINGO framework, we investigate the differential network between those integrative dependencies for random variables from multiple platforms, which follows a multivariate normal distribution. This approach allows us to consider more refined biological relationships between platforms than the original DINGO package, which did not consider dependencies between different data platforms (the implementation of this chain graph model is described in Supplementary Section S1, along with a notation table).

The input to iDINGO includes up to three matrices containing expression data on the same samples from different platforms, as well as a vector denoting the group membership of each sample. The final iDINGO object contains all of the possible edges among and between platforms, along with their respective partial correlations for each group, differential scores and P -values. The iDINGO methodology is general to be applied with just one, two or more than three platforms with known ordering information between them (in the case of one platform, the regular DINGO algorithm will be used).

Examples of the iDINGO implementation have been provided in Supplementary Section S5. In addition, we discuss additional features added to iDINGO, including parallelization, false discovery

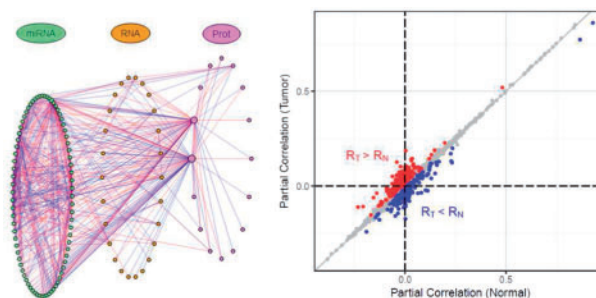


Fig. 1. Left: iDINGO differential network plot (with microRNA nodes on the left, mRNA nodes in the center and protein nodes on the right), containing partial correlations increased in breast cancer (red edges) or normal tissue (blue edges), with $FDR < 0.1$. Right: A scatterplot comparing the partial correlations of the edges in the two groups, colored as in the network

rate (FDR) corrected P -values and differential network plotting (Supplementary Section S1).

3 Shiny web application

iDINGO has been implemented in a web application using the *Shiny* R package (Chang et al., 2017), to provide a user-friendly integrative analysis method. A description of its graphical user interface and a usage example are provided in Supplementary Section S4.

One, two or three omics datasets (matched samples) are to be provided as text files. The sample group classifiers are input as another text file, containing the binary group information for the samples. We recommend generating pathway-based iDINGO networks with no more than a few hundred elements, due to the computational resources and time required. More information about pathway-level analysis is provided in Supplementary Section S2, as well as other considerations in platform integration in Supplementary Section S3.

After running iDINGO, the resulting network is presented in the main panel along with a table of the top hub elements (those with the greatest number of differential edges). In the network plot (Fig. 1), nodes are colored by platform level, and the differential network can be depicted in multiple layouts. The differential score threshold can be set to further filter which edges are considered ‘differential’ and included in the plotted network. A scatterplot of group-specific partial correlations (Fig. 1) is provided to compare the magnitude of the group-specific dependencies.

Funding

This work was supported by the National Institutes of Health [R01-CA194391 and P50-CA070907-18 to M.J.H. and V.B., R01160736 to V.B. and P30-CA016672 to V.B. and K.A.D.], the National Science Foundation [1463233 to V.B.] and the Moon Shot Grant Program at MD Anderson Cancer Center [to C.A.C. and K.A.D.].

Conflict of Interest: none declared.

References

- Bandyopadhyay, S. et al. (2010) Rewiring of genetic networks in response to DNA damage. *Science*, **330**, 1385–1389.
- Califano, A. (2011) Rewiring makes the difference. *Mol. Syst. Biol.*, **7**, 463.
- Chang, W. et al. (2017) Shiny: web application framework for R. *R package version 1.0.5*.
- Flintoft, L. (2004) Rewiring the network. *Nat. Rev. Genet.*, **5**, 808–808.
- Frydenberg, M. (1990) The chain graph Markov property. *Scand. J. Stat.*, **17**, 333–353.

- Gerstung, M. *et al.* (2015) Combining gene mutation with gene expression data improves outcome prediction in myelodysplastic syndromes. *Nat. Commun.*, **6**, 5901.
- Ha, M.J. *et al.* (2015) DINGO: differential network analysis in genomics. *Bioinformatics*, **31**, 3413–3420.
- Hudson, N.J. *et al.* (2009) A differential wiring analysis of expression data correctly identifies the gene containing the causal mutation. *PLoS Comput. Biol.*, **5**, e1000382.
- Lauritzen, S.L. and Wermuth, N. (1989) Graphical models for associations between variables, some of which are qualitative and some quantitative. *Ann. Stat.*, **17**, 31–57.
- Qin, L.-X. (2008) An integrative analysis of microRNA and mRNA expression—a case study. *Cancer Inform.*, **6**, 369–379.
- Reverter, A. *et al.* (2006) Simultaneous identification of differential gene expression and connectivity in inflammation, adipogenesis and cancer. *Bioinformatics*, **22**, 2396–2404.