Bioinformatics, 37(2), 2021, 276–278
doi: 10.1093/bioinformatics/btaa1107
Advance Access Publication Date: 8 January 2021
Applications Note

Gene expression

Interactive gene networks with KNIT

D. S. Magruder [1,2,*], A. M. Liebhoff[1], J. Bethune[1] and S. Bonn [1,*]

[1]Institute for Medical Systems Biology, bAIome – Center for Biomedical AI, Center for Molecular Neurobiology, University Medical Center Hamburg-Eppendorf, 20246 Hamburg, Germany and [2]Genevention GmbH, 37079 Goettingen, Germany

*To whom correspondence should be addressed.

Associate Editor: Pier Luigi Martelli

Received on May 26, 2020; revised on December 28, 2020; editorial decision on December 29, 2020; accepted on January 3, 2021

Abstract

Summary: KNIT is a web application that provides a hierarchical, directed graph on how a set of genes is connected to a particular gene of interest. Its primary aim is to aid researchers in discerning direct from indirect effects that a gene might have on the expression of other genes and molecular pathways, a very common problem in omics analysis. As such, KNIT provides deep contextual information for experiments where gene or protein expression might be changed, such as gene knock-out and overexpression experiments.
Availability and implementation: KNIT is publicly available at http://knit.ims.bio. It is implemented with Django and Nuxtjs, with all major browsers supported.
Contact: sumner.magruder@zmnh.uni-hamburg.de or sbonn@uke.de
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

A common method to understand the functional role of a gene is by altering its expression status and measuring the subsequent molecular changes of the cell. Alterations in gene expression have been functionally grouped into gene deletion, expression attenuation or gene over-expression using various Molecular Biology methods. Subsequent measurements of molecular changes are usually obtained using omics technologies, such as next generation sequencing in the case of gene expression.

While these studies present molecular changes with unprecedented depth they tend to lack information on which functional changes are directly caused by the gene of interest, and which ones are compensatory or indirect changes to sustain cell homeostasis. Therefore, differentially expressed genes (DEGs) in a gene deletion study are not necessarily a direct consequence of initial perturbation of the system. Identifying which of these downstream effects are relevant to the primary gene of interest (e.g. the gene which was deleted, over-expressed, etc.) can be cumbersome.

To contextualize a set of genes in relation to the primary gene of interest, gene-gene network tools like GeneMANIA, Pathway Commons, STRING and STITCH may be leveraged (Cerami *et al.*, 2006, 2011; Kuhn *et al.*, 2008; Mostafavi *et al.*, 2008; Szklarczyk et al., 2015). However, these tools do not provide a query matching the paradigm, as they search for connections between the set of all requested genes rather than querying for pathways to or from the primary gene from or to the rest of the genes in the set. In conjunction with the use of force-based graph layouts, visualizing the directional relationship from a primary gene of interest to a set of genes becomes convoluted.

Here, we present KNIT, a web application that provides visual and query-able information on how a set of genes is connected to a particular gene of interest. KNIT uses hierarchical, directed layouts to provide visual cues of potentially direct effects, aiding researchers in defining the true molecular function of a gene of interest. In addition, KNIT supports enrichment analysis of a given graph, guiding the formulation of hypotheses on the underlying biology. While KNIT was designed with the gene knock-in (KI) and knock-out (KO) paradigm in mind, KNIT clearly generalizes to any exploratory question between a gene of interest and set of genes.

2 Materials and methods

KNIT facilitates the exploration of the directional relationship between a gene of interest (e.g. KI or KO) and the entries in a gene list (e.g. DEGs) by constructing a composite graph from the human data collected via cPath, which is made accessible by Pathway Commons and through metadata from NCBI (Cerami *et al.*, 2006, 2011). As the goal of KNIT is the identification of targeted relationships between a gene of interest and another set of genes, KNIT constructs a composite graph by utilizing the *k*-shortest paths to and from the gene of interest and each entry in the gene list, where *k* is set by the user. As querying for the pathways may be computationally expensive, each pathway request (source, target, *k*) is conducted asynchronously on the backend, allowing users to see their composite graph develop in real time. In case the user supplies gene expression fold change information or *P*-values for the gene set genes, KNIT will incorporate this information in the resultant graph (Fig. 1). While individual connections are valid, as they are retrieved from an established database (Pathway Commons), not every sequence of connections in the graph is a pathway, as KNIT displays aggregated

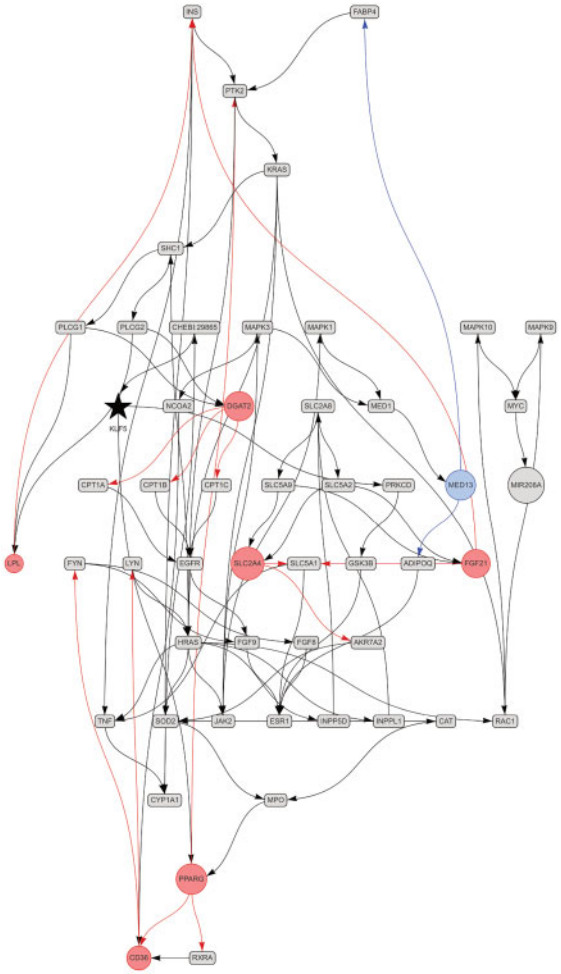

Fig. 1. The composite graph returned by KNIT. The primary gene of interest (KLF5) is highlighted as a black star, while a set of DEGs (circles) are colored and scaled by their metadata (expression change and significance, where red and blue symbols indicate up- and down-regulation, respectively). The graph suggests that the up-regulation of Pparg may be a direct consequence of KLF knockout. Together with the interactively accessible metadata, the graph allows users to formulate hypotheses of how their primary gene interacts with a given set of genes

information from various sources. Pathway information can, however, be retrieved from provided metadata. Therefore, KNIT should be understood as a simple-to-use tool for hypothesis formulation and defining the focus in follow up research.

2.1 Path-finding

Path-finding is non-trivial and depends on various preferences such as the length of the path or the cost to travel the path, as highlighted in Madkour *et al.* (2017) for shortest-path algorithms and Pascoal et al. (2006) for *k*-shortest-path algorithms. In addition, viewing the singular 'best' path between two entities in an interaction network provides an incomplete image of how these two entities interact. Therefore KNIT computes the *k*-best paths utilizing Yen's algorithm (Yen, 1971), which utilizes Dijstkra's algorithm for shortest-path finding. Currently, KNIT weights arcs in the graph proportionally to the number of publications supporting the arc.

2.2 Web application

KNIT's architecture consists of a single page Nuxt.js application and a singular module Django application. A Node.js server provides the frontend, while Nginx serves the backend. Non-blocking asynchronous requests to build the composite graph for each source-target pair and for the composite graph's metadata are sent from the frontend via the axios.js library to the backend, allowing for scalability (Supplementary Fig. S1). The *k*-shortest paths for each source—target pair are calculated by the backend and as this data is returned to the frontend, the graph is rendered utilizing vis.js. An overview of this architecture can be seen in the Supplementary Materials. Once all queried paths are found, the meta information for the resultant sub-graph is requested. Four main types of meta information are provided: (i) data sources: origin of evidence for the composite graph together with a brief overview of the sources, (ii) interaction types: summary of the interaction information between entities of the sub-graph, as well as the relative percentage of publications that support that interaction type, (iii) pathways: known pathways of edges of the returned sub-graph are a part of and (iv) publications: list of the supporting publications. In addition, KNIT provides a feature for interactive computation of enrichment. The user can select meta-information which will update the graph. As an edge may have multiple sources of evidence which support it, the edge will only be removed from the graph if every supporting evidence is deselected.

3 Usage and case study

KNIT has a rich online documentation, explaining its basic functionality and how to interpret analysis results. In addition, KNIT supports batch upload of data, which makes it easy to query a list of e.g. 50 differentially expressed genes with *P*-value and fold change information. To exemplify KNIT's salient features we used cardiomyocyte data to compare analysis results for KNIT, STRING, STITCH, Genemania and Pathway Commons using default settings (Cerami *et al.*, 2006, 2011; Kuhn *et al.*, 2008; Mostafavi *et al.*, 2008; Szklarczyk *et al.*, 2015) (Fig. 1, Supplementary Fig. S2). More specifically, the data by Pol *et al.* (2019) highlights the effect of cardiomyocyte KLF5 signaling (black star in Fig. 1) on white adipose tissue using a murine *Klf5* knocked-out model. The *Klf5* knock-out resulted in increased weight of the mice and increased mRNA levels of genes involved in the adipocyte lipid metabolism: Pparg1, Pparg2, Lpl, Cd36 and Dgat2 (Fig. 1, red and blue circles). As can be seen in Figure 1 and Supplementary Figures S2 and S3, KNIT shows clearly which genes might be directly affected by the *Klf5* KO and which are not, while visualizing positive and negative interactions, the interaction type and all relevant meta-information interactively. Additional validation is provided in Supplementary Materials and in Supplementary Table S1.

4 Conclusion

KNIT provides users an intuitive GUI to readily find interactions to their primary gene of interest along with associated meta-data. Further KNIT interactive exploration aids researchers in framing the relationship between the conditions of their experiment and the results. KNIT is the first web application that allows to query a target gene and a gene set of interest for potential directed signaling, supporting researchers in differentiating direct from indirect interactions.

Financial Support: none declared.

Conflict of Interest: none declared.

References

Cerami,E.G. *et al.* (2006) cPath: open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics*, **7**, 497.

Cerami,E.G. *et al.* (2011) Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.*, **39**, D685–D690.

Kuhn,M. *et al.* (2008) STITCH: interaction networks of chemicals and proteins. *Nucleic Acids Res.*, **36**, D684–8.

Madkour,A. *et al.* (2017) A Survey of Shortest-Path Algorithms. Technical Report. arXiv.

Mostafavi,S. *et al.* (2008) GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol.*, **9**, S4.

Pol,C.J. *et al.* (2019) Cardiac myocyte KLF5 regulates body weight via alteration of cardiac FGF21. *Biochim. Biophys. Acta Mol. Basis Dis.*, **1865**, 2125–2137.

Pascoal,M.B. *et al.* (2006) A comprehensive survey on the quickest path problem. *Ann. Oper. Res.* **147**, 5–21.

Szklarczyk,D. *et al.* (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.*, **43**, D447–52.

Yen,J.Y. (1971) Finding the K shortest loopless paths in a network. *Manag. Sci.*, **17**, 712–716.