

## Gene expression

# GENIUS: web server to predict local gene networks and key genes for biological functions

Tomas Puelma<sup>1,2</sup>, Viviana Araus<sup>1</sup>, Javier Canales<sup>1,†</sup>, Elena A. Vidal<sup>1</sup>, Juan M. Cabello<sup>1</sup>, Alvaro Soto<sup>2</sup> and Rodrigo A. Gutiérrez<sup>1,\*</sup>

<sup>1</sup>Millennium Nucleus Center for Plant Systems and Synthetic Biology, FONDAP Center for Genome Regulation, Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Santiago 8331150, Chile and <sup>2</sup>Millennium Nucleus Center for Plant Systems and Synthetic Biology, Departamento de Ciencias de la Computación, Facultad de Ingeniería, Pontificia Universidad Católica de Chile, Santiago 7820436, Chile

\*To whom correspondence should be addressed.

<sup>†</sup>Present address: Instituto de Bioquímica y Microbiología, Facultad de Ciencias, Universidad Austral de Chile, Campus Isla Teja s/n, Valdivia 5090000, Chile

Associate Editor: Janet Kelso

Received on June 3, 2016; revised on October 14, 2016; editorial decision on November 3, 2016; accepted on November 4, 2016

## Abstract

**Summary:** GENIUS is a user-friendly web server that uses a novel machine learning algorithm to infer functional gene networks focused on specific genes and experimental conditions that are relevant to biological functions of interest. These functions may have different levels of complexity, from specific biological processes to complex traits that involve several interacting processes. GENIUS also enriches the network with new genes related to the biological function of interest, with accuracies comparable to highly discriminative Support Vector Machine methods.

**Availability and Implementation:** GENIUS currently supports eight model organisms and is freely available for public use at <http://networks.bio.puc.cl/genius>.

**Contact:** [genius.psbl@gmail.com](mailto:genius.psbl@gmail.com)

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

## 1 Introduction

Biologists often pursue the difficult task of finding key genes to modulate biological functions or complex traits of interest (e.g. disease, growth, yield, plant nitrogen-use efficiency, water-use efficiency). These functions and traits are governed by networks of functionally interacting genes and gene products. Unfortunately, the large number of genes, and the many and intricate interactions among them, make it difficult to discern which genes are relevant. As a consequence, biologists use a mix of prior knowledge and intuition to choose which genes to focus and which to ignore for detailed experimental work (Moreau and Tranchevent, 2012). This has motivated the development of many reverse engineering computational methods and tools that use experimental data to infer and analyze gene networks (Aoki *et al.*, 2007; Franceschini *et al.*, 2013; Hruz *et al.*, 2008; Jupiter *et al.*, 2009; Obayashi and Kinoshita, 2010;

Obayashi *et al.*, 2013; Stuart *et al.*, 2003; Usadel *et al.*, 2009; Vandepoele *et al.*, 2009; Warde-Farley *et al.*, 2010). An inconvenience of current methods is that they use a wide and often fixed set of experimental data and genes, which limits their predictive capabilities.

To tackle these problems, in previous work we presented DLS, a machine learning method to infer local and discriminative gene networks that can pinpoint new and key genes related to specific biological functions or even complex traits of interest (Puelma *et al.*, 2012). DLS accomplishes this by using supervised machine learning to find expression signatures: expression patterns present in particular genes and experimental conditions that are distinctive and relevant for the biological function defined by the user. These signatures are used by DLS to find coexpression associations and construct a gene network focused in specific biological contexts that are relevant

for the biological function of interest. This ability is the key feature that distinguishes DLS from other network inference algorithms and popular tools like GeneMANIA (Warde-Farley *et al.*, 2010), allowing it to find relevant genes and relationships that these tools may miss.

Here we present GENIUS (GEnE Networks Inference Using Signatures), a web server with a user-friendly interface to DLS that allows the scientific community to fully exploit its capabilities. GENIUS incorporates Gene Ontology annotations and thousands of microarrays experiments from Gene Expression Omnibus (GEO) for eight model organisms: *A. thaliana*, *C. elegans*, *D. rerio*, *D. melanogaster*, *E. coli*, *H. sapiens*, *M. musculus* and *S. cerevisiae*. Depending on existing annotations, the prediction performance ranges from 54 to 94% (Supplementary Fig. S5). In addition, GENIUS adds tools to visualize and analyze gene networks, including integration with Cytoscape (Smoot *et al.*, 2011), an advanced network analysis software platform.

## 2 Results and discussion

Starting a new prediction in GENIUS is simple. The first step is to define a query list of genes related to the biological function or trait of interest. This can be done by directly adding a list of gene identifiers and/or by selecting a list of GO biological processes. Users can select which GO evidence codes to use when importing GO annotations. They can also select the stringency of predictions, allowing them to obtain a more exploratory or reliable network prediction (Supplementary Fig. S1).

The predicted network contains genes from the query list, as well as newly predicted genes functionally related to them, which can be ranked according to various properties (Supplementary Fig. S2). Section 1.1 of Supplementary Information presents a case study comparing the predictions of GENIUS and GeneMANIA for nitrate response in *A. thaliana*. Both tools correctly predict a total of 25 genes that have been validated by published evidence. Of these, 16 genes are common for both tools, while 9 are exclusive to GENIUS and 9 to GeneMANIA, showing that GENIUS can provide complementary and valuable predictions.

### 2.1 GENIUS automatically selects relevant expression signatures

A key aspect of GENIUS is its ability to automatically select genes and experimental conditions containing discriminative expression patterns or ‘expression signatures’. Examining the discovered signatures can be useful to identify key genes for the biological function of interest, as well as experimental conditions under which these genes may be experimentally verified. Section 1.1.3 of Supplementary Information shows that GENIUS selects relevant and biologically sound expression signatures for the nitrate response case study.

### 2.2 GENIUS can uncover key regulators for complex traits

A key aspect of GENIUS is that it can infer local functional networks that expose central genes acting at the intersection of several biological processes. GENIUS allows users to easily identify these central genes by ranking them according to centrality indicators used in network theory, like degree centrality (DC) and betweenness centrality (BC) (Section 1.2.1 of Supplementary Information) (Yu *et al.*, 2007).

As an example, we experimentally validated a prediction performed with GENIUS to find genes modulating nitrogen-use efficiency in *A. thaliana* (Araus *et al.*, 2016). Additionally, Section 1.2 of Supplementary Information illustrates this use of GENIUS in detail, applied to the drought tolerance trait in *A. thaliana*. In contrast to GeneMANIA, GENIUS predicts a network with a scale-free topology, allowing it to take advantage of network theory centrality indicators to pinpoint relevant genes.

Thus, we believe the use of GENIUS web server can facilitate biologists to generate sophisticated hypothesis, make novel discoveries, and ultimately, improve our molecular understanding of biological systems.

## Funding

This work was funded by grants from the Howard Hughes Medical Institute, Fondo de Desarrollo de Areas Prioritarias (FONDAP) Center for Genome Regulation (15090007), Millennium Nucleus Center for Plant Systems and Synthetic Biology (NC130030) and Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT) 1141097 to RAG.

*Conflict of Interest:* none declared.

## References

- Aoki, K. *et al.* (2007) Approaches for extracting practical information from gene co-expression networks in plant biology. *Plant Cell Physiol.*, **48**, 381–390.
- Araus, V. *et al.* (2016) Members of BTB gene family regulate negatively nitrate uptake and nitrogen use efficiency in *Arabidopsis thaliana* and *Oryza sativa*. *Plant Physiol.*, **171**, 1523–1532.
- Franceschini, A. *et al.* (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.*, **41**, D808–D815.
- Hruz, T. *et al.* (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. *Adv. Bioinform.*, **2008**, 420747.
- Jupiter, D. *et al.* (2009) STARNET 2: a web-based tool for accelerating discovery of gene regulatory networks using microarray co-expression data. *BMC Bioinform.*, **10**, 332.
- Moreau, Y. and Tranchevent, L.C. (2012) Computational tools for prioritizing candidate genes: boosting disease gene discovery. *Nat. Rev. Genet.*, **13**, 523–536.
- Obayashi, T. *et al.* (2013) COXPRESdb: a database of comparative gene coexpression networks of eleven species for mammals. *Nucleic Acids Res.*, **41**, D1014–D1020.
- Obayashi, T. and Kinoshita, K. (2010) Coexpression landscape in ATTED-II: usage of gene list and gene network for various types of pathways. *J. Plant Res.*, **123**, 311–319.
- Puelma, T. *et al.* (2012) Discriminative local subspaces in gene expression data for effective gene function prediction. *Bioinformatics*, **28**, 2256–2264.
- Smoot, M.E. *et al.* (2011) Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, **27**, 431–432.
- Stuart, J.M. *et al.* (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science*, **302**, 249–255.
- Usadel, B. *et al.* (2009) Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. *Plant. Cell Environ.*, **32**, 1633–1651.
- Vandepoele, K. *et al.* (2009) Unraveling transcriptional control in *Arabidopsis* Using cis-regulatory elements and coexpression networks. *Plant Physiol.*, **150**, 535–546.
- Warde-Farley, D. *et al.* (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.*, **38**, W214–W220.
- Yu, H. *et al.* (2007) The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.*, **3**, e59.