



ELSEVIER

Contents lists available at ScienceDirect

Data in Brief

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



Data Article

Data and programs in support of network analysis of genes and their association with diseases

Panagiota I. Kontou^{a,1}, Athanasia Pavlopoulou^{a,1},
Niki L. Dimou^a, Georgios A. Pavlopoulos^b, Pantelis G. Bagos^{a,*}

^a Department of Computer Science and Biomedical Informatics, University of Thessaly, Papasiopoulou 2-4, Lamia 35100, Greece

^b Lawrence Berkeley Lab, Joint Genome Institute, United States Department of Energy, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

ARTICLE INFO

Article history:

Received 2 June 2016

Received in revised form

6 July 2016

Accepted 13 July 2016

Available online 19 July 2016

Keywords:

Gene-disease associations

Gene-gene networks

Disease-disease networks

ABSTRACT

The network-based approaches that were employed in order to depict the relationships between human genetic diseases and their associated genes are described. Towards this direction, mono-partite disease-disease and gene-gene networks were constructed from bipartite gene-disease association networks. The latter were created by collecting and integrating data from three diverse resources, each one with different content, covering from rare monogenic disorders to common complex diseases. Moreover, topological and clustering graph analyses were performed. The methodology and the programs presented in this article are related to the research article entitled "Network analysis of genes and their association with diseases" [1].

© 2016 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

DOI of original article: <http://dx.doi.org/10.1016/j.gene.2016.05.044>

* Coresponding author. Fax: +30 223 106 6915.

E-mail address: pbagos@compgen.org (P.G. Bagos).

¹ These authors contributed equally to this work.

<http://dx.doi.org/10.1016/j.dib.2016.07.022>

2352-3409/© 2016 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Specifications Table

Subject area	Systems biology
More specific subject area	Gene-disease networks
Type of data	Figure, text files, Cytoscape Network file
How data were acquired	Data were acquired from the publicly available databases: OMIM, GAD, GWAS, UniProtKB, ICD, HGNC
Data format	Processed, analyzed
Experimental factors	Gene-disease association data were analyzed using Perl and R scripts and Cytoscape.
Experimental features	Gene-gene and disease-disease networks were constructed.
Data source location	Department of Computer Science and Biomedical Informatics, University of Thessaly, Lamia, Greece
Data accessibility	Data are provided with this article.

Value of the data

- The need for integrating complementary data from different sources to biological networks is further highlighted in this study.
- Important, previously unknown, associations between genes and diseases were revealed.
- Based on the constructed disease-disease networks, diseases with apparently distinct phenotypic manifestations were found to share a common genetic background. This finding could be utilized in network pharmacology.

1. Data

The overall procedure of the data analysis is shown illustratively in Fig. 1. The Perl ([Supplementary Files 1-5](#)) and R ([Supplementary File 6](#)) programs used for data analysis are indicated. A complete description of the data and methodology is presented in [1].

2. Experimental design, materials and methods

2.1. Data collection

Disease-gene association data were collected and integrated from three diverse publicly available, comprehensive resources (NCBI's OMIM [2], NIH's GAD [3] and NHRI GWAS Catalog [4]). As a given disease can be associated with more than one gene, a script was written in Perl to separate the multiple entries ([Supplementary File 1](#); separate.pl).

2.2. Disease and gene nomenclature

In order to maintain a consistent nomenclature and classification for diseases in our analysis, the naming conventions described in the International Classification of Diseases (ICD) were used. The disease terms from the three databases were converted to ICD terms with the use of a Perl script ([Supplementary File 2](#); ICD.pl). Moreover, in order to maintain a uniform nomenclature across all datasets, all genes from our three databases along with the ones from UniProtKB [5] were converted to the official HGNC (HUGO Gene Nomenclature Committee) [6] gene symbols using a Perl script ([Supplementary File 3](#); Hugo.pl).

Download English Version:

<https://daneshyari.com/en/article/174731>

Download Persian Version:

<https://daneshyari.com/article/174731>

[Daneshyari.com](https://daneshyari.com)