An online tool for measuring and visualizing phenotype similarities using HPO

Jiajie Peng1†, Hansheng Xue2†, Weiwei Hui1, Junya Lu1, Bolin Chen1, Qinghua Jiang3, Xuequn Shang1* and Yadong Wang2,4*

From 13th International Symposium on Bioinformatics Research and Applications (ISBRA 2017)
Honolulu, Hawaii, USA. 30 May - 2 June 2017

Abstract

Background: The Human Phenotype Ontology (HPO) is one of the most popular bioinformatics resources. Recently, HPO-based phenotype semantic similarity has been effectively applied to model patient phenotype data. However, the existing tools are revised based on the Gene Ontology (GO)-based term similarity. The design of the models are not optimized for the unique features of HPO. In addition, existing tools only allow HPO terms as input and only provide pure text-based outputs.

Results: We present PhenoSimWeb, a web application that allows researchers to measure HPO-based phenotype semantic similarities using four approaches borrowed from GO-based similarity measurements. Besides, we provide a approach considering the unique properties of HPO. And, PhenoSimWeb allows text that describes phenotypes as input, since clinical phenotype data is always in text. PhenoSimWeb also provides a graphic visualization interface to visualize the resulting phenotype network.

Conclusions: PhenoSimWeb is an easy-to-use and functional online application. Researchers can use it to calculate phenotype similarity conveniently, predict phenotype associated genes or diseases, and visualize the network of phenotype interactions. PhenoSimWeb is available at http://120.77.47.2:8080.

Keywords: Human phenotype ontology, Web tool, Phenotype similarity

Background

Since the successful completion of the Human Genome Project, significant improvement has been made in genome sequencing technologies, which benefit the Mendelian disease and cancer diagnosis [1–9]. Even so, it remains challenging to make correct diagnosis only based on sequencing technologies for many diseases. Because the relationships between genetic variants and clinical phenotypes are difficult to understand for diseases with high genetic heterogeneity and complex phenotypes [10, 11].

Patient phenotypes are the observable features of a patient, such as anatomy and biomedical properties [12]. Phenotypes are usually determined by both genetic and environmental factor. To improve the efficiency of disease diagnosis, several methods have been developed to analyse the relationships between patient phenotypes and known phenotypes related with a gene based on Human Phenotype Ontology (HPO) recently [13–15]. The Human Phenotype Ontology (HPO) is one of the most popular bioinformatics resources, which was constructed by Robinson et al. in 2008 [12]. HPO provides the unique and structured vocabulary to represent the phenotypic characteristics and their relationships with a directed acyclic graph (DAG). In recent study, quantifying the phenotypic similarity based on HPO is usually integrated with sequencing technologies to aid disease diagnosis [16–20].
As a kind of widely used resource, HPO contains abundant information and researchers could study phenotype semantic similarity conveniently. In recent years, various methods have been proposed to compute HPO-based phenotype similarities by comparing HPO terms with their annotations and topological information, such as Phenomizer [21], OWLSim [22] and HPOSim [23]. However, Most of these methods are modified based on GO-based similarity measurements that have been widely utilized and studied by many researchers [24–31]. Phenomizer applied information content to compute the phenotype semantic similarity based on HPO. Based on the IC-based method, PhenomeNet [32] and OWLSim [22] exploit simGIC [33] to measure the semantic similarity of two phenotype sets. HPOSim [23] implements seven commonly used ontology-based semantic similarity measurements to compute the phenotype similarities, such as Jiang measure [34], Schlicker measure [35] and Wang measure [31].

Although the aforesaid methods have been widely used to measure the phenotype similarity, none of these measurements take into account the unique features of HPO. To fill this gap, we have recently presented a measurement named PhenoSim to compute the phenotype similarities [36, 37] considering the unique properties of HPO. Our method can simulate the noise in the patient phenotype dataset and compute the similarities using a novel path-constrained Information Content based measurement in three steps. Firstly, construct a phenotype network. Secondly, reduce noise in the patient’s phenotype set using PageRank [38] algorithm. Thirdly, compute phenotype set similarities using a novel path-constrained Information Content based measurement. And the experiment result shows that PhenoSim performs better than existing methods.

In addition, existing tools have two main drawbacks: firstly, none of existing tools allow text that describes phenotype features as input, neglecting that symptoms of patients are always described as text not HPO terms; secondly, most existing tools ignore the effect of visualization, which is necessary for result interpretation, and simply list the experimental results as the final output. Thereby, it is very urgent and essential to research an easy-to-use and functional web application.

In this article, we present a novel and easy-to-use online application, termed as PhenoSimWeb, to compute phenotype similarities based on HPO and to visualize the similarity using an intuitive graphical interface. Comparing with the existing online tools, the main contributions of our work can be summarized as:

- PhenoSimWeb supplies researchers with a measurement based on the design optimized for unique features of HPO.
- PhenoSimWeb allows text that describes phenotype features as input.
- PhenoSimWeb contains an intuitive and functional visualization interface to visualize phenotype association network.

**Methods**

PhenoSimWeb is a Browser/Server architecture-based online application which can be used to calculate the phenotype similarities based on HPO, visualize the association between phenotypes, and predict the associated gene/diseases given a set of phenotypes. The back-end of PhenoSimWeb is implemented using Java SDK 7, Python 2.7 and web framework web.py. And PhenoSimWeb uses MySQL to manage dataset. In part of data transmission between the browser and server, the web application applies JavaScript Object Notation (JSON) and Asynchronous JavaScript and XML (AJAX) and so on. Besides, PhenoSimWeb uses cytoscape.js and HTML5 canvas as the graphics engine for the association network visualization. The Human Phenotype Ontology (HPO) dataset was downloaded from the HPO official website (http://humanphenotype-ontology.github.io/) on January, 2016. PhenoSimWeb was tested on Chrome, Firefox and Internet Explorer.

**Results and discussion**

PhenoSimWeb mainly contains two operations to execute: 1) to type in a set of phenotypes and specify the corresponding parameters, 2) to visualize and download the phenotype similarities. Besides, users can submit a list of phenotypes to predict the genes or diseases associated with the given phenotype set.

**User inputs**

The user interface of PhenoSimWeb can be divided into three parts: phenotypes input (Fig. 1a), similarity measurement selection (Fig. 1b), and user information input (Fig. 1c).

PhenoSimWeb mainly contains three functional modules: (1) given a list of phenotypes, calculate the pairwise similarities among the input phenotypes; (2) given a list of genes or diseases, calculate the pairwise similarities by aggregating the similarities of phenotypes associated to given genes or diseases; (3) given a list of phenotypes, identify the most associated genes or diseases with the given phenotypes based on their HPO-based similarity. The input interface for each functional module is introduced as follows.

**Input interface for phenotype similarity calculation**

PhenoSimWeb provides three methods for user to input a phenotype list. User can input text that describes phenotypes, select phenotypes from existing databases, and
Fig. 1 The main input webpage of PhenoSimWeb. The whole process can be divided into three parts, including: a) inputting phenotype, gene, or disease dataset, b) choosing phenotype similarity measurement, c) typing in experimental user information optionally.

input a set of phenotypes directly (see Fig. 1a). Allowing text input is important, since patients’ phenotypes are always described in text, such as clinical records. PhenoSimWeb uses annotation tool Annotator [39] of the National Center for Biomedical Ontology (NCBO) to convert input text to corresponding HPO terms. For the other two input methods, Only HPO ID and Name are allowed in current version.

Input interface for gene (or disease) similarity calculation
PhenoSimWeb provides two methods for user to input a gene (or disease) list. User can select genes or diseases from existing databases, and input a set of genes or diseases directly (see Fig. 2). Currently, PhenoSimWeb can only calculate similarities for genes or diseases that annotated by HPO terms, since their similarities are based on the HPO-based phenotype similarities.

Input interface for phenotype associated gene or disease prediction
In this part (Fig. 3), users can input phenotype set in the left text box and select the type of target to be predicted, such as gene or disease. Users can also provide a list of target genes or diseases in the right text box to check whether these genes or diseases are associated with the input phenotype set. If the user do not provide a specific gene or disease set, PhenoSimWeb would compare the phenotype set with all the genes or diseases involved in HPO.

After the data input step, users can select a semantic similarity measurement for phenotype similarity calculation. A new proposed measurement named PhenoSim and other four widely-used similarity measurements are available to choose. The detailed descriptions of these measurements are in the following subsection.

In the last step, users can input email address and the experimental user name optionally. And if users do it, the application will send a notification to the specified mailbox when all the job has been done. And the application will validate it for error checking if all the input information is submitted. The validation process mainly checks the format of input phenotypes, phenotype lists, phenotype texts, genes, genes lists, diseases, diseases lists and all the user specified parameters. And if the input exists any errors, users would be notified immediately. After the validation process, the application will calculate the similarity using specified measurement, which users chose in step two, among phenotypes, genes or diseases, and visualize the phenotype associated network.

All the submitted jobs are executed by a job scheduler on the back-end server of PhenoSimWeb. Once all the jobs
are finished, a notification email will be sent to the specified mailbox, if users typed in email address in step three. Also, the web will jump to the experimental result’s webpage, if the user unclose the submission webpage and keep it on.

The experimental result webpage displays the detailed similarity calculation results and corresponding p-values (Fig. 4). The other detailed information in the calculation process, such as the calculation method, is also displayed on the result webpage. Besides, users can download the experimental result and corresponding information by the links on the webpage.

**Visualization interface**

*PhenoSimWeb* supplies an intuitional and functional visual webpage to display the similarity results. The visualization interface of *PhenoSimWeb* (see Fig. 5) displays the resulting phenotype association network, and gene or disease association network based on corresponding phenotype similarities in the visualization webpage, in which a node represents a term, such as phenotype or gene or disease, and an edge between any two interconnected terms indicates that the edge similarity score is greater than the edge similarity threshold, which users input in Fig. 5a. Users can implement interactive browsing of the visual interface using the mouse conveniently (Fig. 5c). Besides, users can also activate the node operation panel by long-right clicking a node (see panel in Fig. 5e). Using the node operation panel, users can execute multiple node operations, such as: insert current term into selected list on panel A, display term info in top-right panel D, insert term into locked list, delete current term from locked list and set current node’s background color into green.

Users can drag the threshold bar or type in a specific value directly to adjust the edge similarity threshold, and the network will change simultaneously (see Fig. 5a). *PhenoSimWeb* also provides several different graph layouts for graph visualization (see Table 1). Figure 5b shows the overall distribution of similarity scores for all the input term pairs, users can regulate the edge similarity threshold in panel A by this distribution intuitively. The resulting
An illustrative example of using PhenoSimWeb
In this section, we take the sample list of phenotypes in the website as input to demonstrate how to use PhenoSimWeb this web application to calculate the pairwise similarity for a set of phenotypes. We select the “PhenoSim” as the HPO similarity measurement in Fig. 1b. And the parameters in Fig. 1c are optional, user can type in an email address and leave the corresponding user name or

![Figure 4](image1.png) The calculation results of the phenotype list. And PhenoSim calculated the corresponding P-Value in addition to the semantic similarity

term association network is browsed in the network displaying panel (see Fig. 5c). Besides, users can specify a term group in panel A or node operation panel to select subnetworks (see Fig. 5f). And the term information panel (Fig. 5d) displays the neighbors of current selected term. By clicking a term ID or name on the information panel, users can get more comprehensive information about this term from website (http://compbio.charite.de).

![Figure 5](image2.png) The visualization interface of PhenoSimWeb to explore phenotype, gene or disease functional similarities based on HPO. Panel c shows the term association network, and among the association network a node indicates a term(phenotype, gene or disease) and an edge between any two interconnected terms indicates that the edge similarity score is greater than the edge similarity threshold, which users input in panel a. Panel b shows the overall distribution of similarity scores for all the input term pairs, users can regulate the edge similarity threshold in panel a by this distribution intuitively. The neighbors of the recently chosen terms are shown in panel d. Users can add, info, lock, unlock and flag a term in the node operation panel e. Besides, panel f show user’s selected subnetwork
implemented similarity measurements
PhenoSimWeb provides five HPO-based semantic similarity measures for all the users. We will briefly introduce these five measurements in the following part.

1) PhenoSim
In briefly, PhenoSim is a path-constrained Information Content-based method for phenotype semantic similarity measurement and includes a noise reduction component to model the noisy patient phenotype data [36]. The whole process of PhenoSim contains three steps: First, it constructs a phenotype network $N$ using phenotype ontologies and gene-phenotype associations. Second, given a set of clinical phenotypes of a patient, it filters noises based on $N$ using PageRANK. Finally, it computes the phenotype similarities with a novel path-constrained Information Content-based method.

Compared with other existing approaches, PhenoSim effectively improves the performance of the phenotype similarity measurement, and enhances the accuracy of phenotype-based causative gene and disease prediction.

2) Information content based (Resnik)
Resnik et al. [40] proposed a method to calculate Ontology-based semantic similarity between any two phenotype ontologies, by integrating Information Content (IC) with the Ontology structure. The information content of any term represents the specificity of the term. The terms at a lower level of Ontology tend to have higher IC, and vice verse. In addition, the IC of two phenotype terms is the lowest common ancestor of these two terms in the ontology structure. Given ontology term $t$, and the corresponding information content of $t$ could be defined as $IC(t) = -\log(|D_t|/|D|)$, where $D_t$ and $D$ are sets of diseases annotated to $t$ and the root term. Mathematically, given any two ontology terms $t_a$ and $t_b$, let $t_{MICA}$ represents their Most Informative Common Ancestor (MICA), the semantic similarity of $t_a$ and $t_b$ is calculated as follows:

$$Sim_{Resnik}(t_a, t_b) = IC(t_{MICA}) = -\log\frac{|D_{SMICA}|}{|D|}$$

where $D_{SMICA}$ and $D$ represent the set of annotations of $t_{MICA}$ and the set of all the annotations involved in the Ontology, respectively.

3) Enhanced information content based (Lin)
Lin et al. [41] considered the Information Content (IC) of two terms $t_a$ and $t_b$ besides the IC of their most informative common ancestor, comparing with the Resnik

not. In the end, we click the “submit” button to submit the job.

Once all the back-end programs are finished, the calculation results will be displayed on the website (Fig. 4). Users can also download the calculation results by clicking the “Click here to download result file of this run”. Besides, users can click “Display” button to view the graphical visualization of corresponding experimental results (see Fig. 5). By adjusting the phenotype-to-phenotype similarity threshold in panel A, we could obtain two contrasting phenotype association networks (see Fig. 6). In addition, we can also display the association network with different layouts, which are interpreted in Table 1, i.e., cola and grid, by selecting graph layouts in panel A (see Fig. 7).

In addition to the above functions, we can also choose several phenotypes (i.e., HP:0000080, HP:0000069, HP:0030037 and HP:0000025) as the interested phenotypes and append them into the blank box in panel A using node operation panel. Then the corresponding sub-network, which contains the selected phenotypes, are highlighted (the right figure in Fig. 8). Besides, users can also add all the neighbors of interested phenotypes into the highlighted network by clicking “Toggle Neighbor Display” in panel A (the left figure in Fig. 8). Furthermore, users could see the detail of each phenotype by clicking nodes among the network in panel C, and the detailed information of the chosen phenotype will be shown in Fig. 5d.

### Table 1

| Name        | Description                                                                 |
|-------------|-----------------------------------------------------------------------------|
| Concentric  | This layout displays the association network in a concentric circle.          |
| Breadthfirst| This layout displays the association network in a hierarchical structure.     |
| Circle      | This layout displays the association network in a circle.                     |
| Cose        | Cose is an abbreviation for Compound Spring Embedder. This layout lays out compound graphs using a force-directed simulation. |
| Cola        | This layout displays the association network using a force-directed physics simulation. |
| Grid        | This layout displays the association network in a well-spaced grid.          |
measure. And the equation of calculating the Ontology term similarity is defined as:

$$\text{Sim}_{\text{Lin}}(t_a, t_b) = \frac{2 \times IC(t_{MICA})}{IC(t_a) + IC(t_b)}$$  \hspace{1cm} (2)

4) Normalized information content based (Schlicker)
Schlicker et al. [35] normalized the Information Content based measure (Resnik) and utilized a weighting function to regulate the overall score:

$$\text{Sim}_{\text{Schlicker}}(t_a, t_b) = \frac{2 \times IC(t_{MICA})}{IC(t_a) + IC(t_b)} \times \left(1 - \frac{|D_{MICA}|}{|D|}\right)$$  \hspace{1cm} (3)

5) Jiang-Conrath Measure (JC)
Comparing the Resnik measure, Jiang-Conrath [34] considered the information content of term $t_a$ and $t_b$ and the distance between the most public common ancestor besides the information content of $t_a$ and $t_b$. And Jiang-Conrath calculates semantic similarity as:

$$\text{Sim}_{\text{JC}}(t_a, t_b) = \frac{1}{\text{dist}(t_a, t_b) + 1}$$  \hspace{1cm} (4)

$$\text{dist}(t_a, t_b) = IC(t_a) + IC(t_b) - 2 \times IC(t_{MICA})$$  \hspace{1cm} (5)

Conclusions
The Human Phenotype Ontology (HPO) is a kind of widely used bioinformatics resources. Recently, various approaches and online or offline tools have been developed to calculate phenotype semantic similarities based on HPO. In this paper, we developed and presented a novel and functional web application, named PhenoSimWeb, which allows researchers to compute phenotype similarity with five different measurements conveniently and visualize the resulting phenotype association networks with an easy-to-use and powerful web visualization interface. PhenoSimWeb allows text that describes phenotype features as input. PhenoSimWeb includes three main functional modules: calculate the
pairwise similarities for the input phenotypes; calculate the gene or disease similarities by aggregating the similarities of phenotypes corresponding to the given genes or diseases; identify the most associated genes or diseases with the given phenotype set. In summary, PhenoSimWeb is a novel and convenient web application for users to calculate and visualize HPO-based phenotype similarities.

Acknowledgements
We thank all anonymous reviewers.

Funding
The publication costs for this article were funded by the corresponding author’s institution. This work was supported by National Natural Science Foundation of China (No. 61702421), Natural Science Basic Research Plan in Shaanxi Province of China (No. 2017QJ6047), China Postdoctoral Science Foundation (No. 2017M610651), Fundamental Research Funds for the Central Universities (3102018T033), National Natural Science Foundation of China (Grant No. 61602386 and 61332014).

Availability of data and materials
All data sets are available at http://120.77.47.2:8080/.

About this supplement
This article has been published as part of BMC Genomics Volume 19 Supplement 6, 2018: Selected articles from the 13th International Symposium on Bioinformatics Research and Applications (ISBRA 2017): genomics. The full contents of the supplement are available online at https://bmcgenomics.biomedcentral.com/articles/supplements/volume-19-supplement-6.

Declaration
The abridged abstract of this work was previously published in the Proceedings of the 13th International Symposium on Bioinformatics Research and Applications (ISBRA 2017): genomics. The full contents of the supplement are available online at https://bmcgenomics.biomedcentral.com/articles/supplements/volume-19-supplement-6.

Authors’ contributions
YW and XS designed the web tool framework; JP and HX implemented the web tool; JP and HX wrote this manuscript; BC helped design the visualization interface; QJ helped design the input interface. All authors read and approved the final manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher’s Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
1 School of Computer Science, Northwestern Polytechnical University, Xi’an 710072, China. 2 Department of Computer Science and Technology, Harbin Institute of Technology, Shenzhen 518055, China. 3 School of Life Science and Technology, Harbin Institute of Technology, Harbin 150001, China. 4 School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China.

Published: 13 August 2018

References
1. Jiang Q, Jin S, Jiang Y, Liao M, Feng R, Zhang L, Liu G, Hao J. Alzheimer’s disease variants with the genome-wide significance are significantly enriched in immune pathways and active in immune cells. Mol Neurobiol. 2017;54(1):.
2. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17(6):333–51.
3. Liu G, Jiang Q. Alzheimer’s disease CD33 rs3865444 variant does not contribute to cognitive performance. Proc Natl Acad Sci. 2016;113(12):1589–90.
4. Deciphering Developmental Disorders Study. Large-scale discovery of novel genetic causes of developmental disorders. Nature. 2015;519(7542):223–8.
5. Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, et al. Molecular findings among patients referred for clinical whole-exome sequencing. Jama. 2014;312(18):1870–9.
6. Peng J, Lu J, Shang X, Chen J. Identifying consistent disease subnetworks using dnet. Methods. 2017;131:104–10.
7. Hu Y, Zhou M, Shi H, Ju H, Jiang Q, Cheng L. Measuring disease similarity and predicting disease-related ncRNAs by a novel method. BMC Med Genomics. 2017;10(5):71. https://doi.org/10.1186/s12920-017-0315-9.
8. Hu J, Shang X. Detection of network motif based on a novel graph canonization algorithm from transcriptional regulation networks. Molecules. 2017;22(12):2194.
9. Hu L, Gao Y, Zheng Y, Shang X. Kf-finder: identification of key factors from host-microbial networks in cervical cancer. BMC Syst Biol. 2018;12(4):54.
10. Liu G, Zhang F, Hu Y, Jiang Y, Gong Z, Liu S, Chen X, Jiang Q, Hao J. Genetic variants and multiple sclerosis risk gene SCLC9A9 expression in distinct human brain regions. Mol Neurobiol. 2017;54(9):6820–6.
diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. Sci Transl Med. 2014;6(252):252–123252123.

12. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. Am J Hum Genet. 2008;83(3):610–5.

13. Groza T, Köhler S, Moldenhauer D, Vasilevsky N, Bayram G, Zemotet J, Schriml LM, Kibbe WA, Schofield PN, Beck T, et al. The human phenotype ontology: semantic unification of common and rare disease. Am J Hum Genet. 2015;97(1):111–24.

14. Köhler S, Doelken SC, Mungall CJ, Bauer S, Firth HV, Baillée-Forestier I, Black GC, Brown DL, Brudno M, Campbell J, et al. The human phenotype ontology project: linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014;42:966–74.

15. Petrovski S, Goldstein DB. Phenornets: the implementation and the interpretation of personal genomics. Sci Transl Med. 2014;6(254):254–3525435.

16. Peng J, Hui W, Shang X. Measuring phenotype-phenotype similarity through the interactome (J). BMC Bioinformatics. 2018;19(5):114.

17. Peng J, Wang T, Wang J, Wang Y, Chen J. Extending gene ontology with gene association networks. Bioinformatics. 2015;31(20):3274–81.

18. Smidley D, Jacobsen JO, Järvelin MR, Köhler S, Holmgren M, Schubach M, Siragusa E, Zemotet J, Buske OJ, Washington NL, et al. Next-generation diagnostics and disease- gene discovery. Nat Rev Genet. 2017;18(16):573.

19. Vissers LE, Veltman JA. Standardized phenotyping enhances mendelian disease gene identification. Nat Genet. 2015;47(1):122–4.

20. Köhler S, Schulz MH, Krawitz P, Bauer S, Dölken S, Ott CE, Mundlos C, Horn D, Mundlos S, Robinson PN. Clinical diagnostics in human genetics with semantic similarity searches in ontologies. Am J Hum Genet. 2009;85(4):457–64.

21. Washington NL, Haendel MA, Mungall CJ, Ashburner M, Westerfield M, Lewis SE. Linking human diseases to animal models using ontology-based phenotype annotation. PLoS Biol. 2009;7(11):1000247.

22. Deng Y, Gao L, Wang B, Guo X. Hpsim: an r package for phenotypic similarity measure and enrichment analysis based on the human phenotype ontology. PLoS ONE. 2015;10(2):0115692.

23. Peng J, Zhang X, Hui W, Lu J, Li Q, Liu S, Shang X. Improving the measurement of semantic similarity by combining gene ontology and co-functional networks: a random walk based approach. BMC Syst Biol. 2018;12(2):18.

24. Peng J, Li H, Liu Y, Juan L, Jiang Q, Wang Y, Chen J. IntegO2: a web tool for measuring and visualizing gene semantic similarities using gene ontology. BMC Genomics. 2016;17(5):530.

25. Cheng L, Jiang Y, Wang Z, Shi H, Sun J, Yang H, Zhang S, Hu Y, Zhou M. Dissim: an online system for exploring significant similar diseases and exhibiting potential therapeutic drugs. Sci Rep. 2016;6:30024.

26. Peng J, Uygun S, Kim T, Wang Y, Rhee SY, Chen J. Measuring semantic similarities by combining gene ontology annotations and gene co-function networks. BMC Bioinformatics. 2015;16(1):1.

27. Peng J, Wang H, Lu J, Hui W, Wang Y, Shang X. Identifying term relations across different gene ontology categories. BMC Bioinformatics. 2017;18(16):573.

28. Teng Z, Guo M, Liu X, Dai Q, Wang C, Xuan P. Measuring gene functional similarity based on group-wise comparison of go terms. Bioinformatics. 2013;29(11):1424–1432.

29. Caniza H, Romero AE, Heron S, Yang H, Devoto A, Frasca M, Mestiri M, Valentini G, Paccanaro A. Gossto: a stand-alone application and a web tool for calculating semantic similarities on the gene ontology. Bioinformatics. 2014;30(15):2235–6.

30. Wang JZ, Du Z, Payattakoos R, Philip SY, Chen C-F. A new method to measure the semantic similarity of go terms. Bioinformatics. 2007;23(10):1274–81.

31. Hoehndorf R, Schofield PN, Gkoutos GV. PhenoMent: a whole-phenome approach to disease gene discovery. Nucleic Acids Res. 2011;39(18):119.

32. Pesquita C, Faria D, Bastos H, Falcao A, Couto F. Evaluating go-based semantic similarity measures. In: Proc. 10th Annual Bio-Ontologies Meeting, vol. 37; 2007. p. 38.