

WebGestalt 2017 Manual

April 08, 2017

The WEB-based GENE SeT AnaLysis Toolkit (WebGestalt) is a suite of tools for functional enrichment analysis in various biological contexts. The original version of WebGestalt was described in the paper “WebGestalt: an integrated system for exploring gene sets in various biological contexts” (Nucleic Acids Res. 2005 Jul 1;33(Web Server issue): W741-8). The updated version, WebGestalt 2013, was described in the paper “WEB-based GENE SeT AnaLysis Toolkit (WebGestalt): update 2013” (Nucleic Acids Res. 2013 Jul 1;41(Web Server issue): W77-83). Since 2013, WebGestalt has been visited 217,382 times by 87,380 unique users from 143 countries and territories according to Google Analytics. The two papers have also been cited in 1229 scientific papers since 2013, according to Google Scholar.

For WebGestalt 2017, we performed the following major changes compared to the WebGestalt 2013 version:

- The new version places all parameters in the same page to reduce the operation time and complexity. This design also facilitates easy submission of multiple analyses.
- In addition to Over-Representation Analysis (ORA), the new version now includes Gene Set Enrichment Analysis (GSEA) for analyzing the ranked gene list and Network Topology-based Analysis (NTA) for analyzing the system-level dependencies and interactions among genes.
- The new version increases the number of organisms from 8 to 12. It now includes a plant species, *Arabidopsis thaliana*, which extends impact of WebGestalt to the plant research community.
- The new version increases the number of gene identifiers from 201 to 324 (+61.2%), with at least 15% increase for each organism. Specially, the new version significantly increased the number of supported Affymetrix platforms from 68 to 101 (+48.5%), Agilent platforms from 18 to 25 (+38.9%) and Illumina platforms from 10 to 18 (+80.0%). These expansions fill the gap between technology platform developments and user's ability to interpret data generated from the new platforms.
- When the organisms or gene identifiers are not covered by the WebGestalt database, WebGestalt 2017 allows users to upload their own functional categories and experimental data with the same gene identifiers, which provides the maximum flexibility for functional enrichment analysis. In fact, this new development makes it even possible to perform enrichment analysis for non-biological data sets in WebGestalt.
- The new version significantly increases the number of functional categories from 78,612 to 150,937 (+92.0%), which includes the new pathway databases Reactome and PANTER, new disease database DisGeNET, new drug database DrugBank, new hierarchical TCGA RNASeq co-expression modules for 33

cancer types and hierarchical protein interaction modules from the protein-protein interaction networks of 11 organisms.

- For the existing organism, users can also upload the new functional database not included in the WebGestalt for the enrichment analysis.
- The new version includes a user-friendly HTML output report containing analysis summary, ID mapping information, GO Slim summary and enrichment results.
- The R code was developed to plot the high-resolution GO Slim summary bar chart for biological process, cellular component and molecular function ontologies, which can directly be used for publication or presentation.
- Based on cutting-edge javascript technology, the new version develops an interactive DAG (directed acyclic graph) visualization system that can visualize the enrichment results under the DAG, zoom the DAG in and out, search and highlight the interesting nodes and related paths in the DAG and download the high-resolution DAG image.
- Because the input to the GSEA analysis contains values for individual genes, the new version can set node colors according to the values of corresponding genes in network visualization

At the top of the WebGestalt interface (blue box in Figure 1), the user can find the link about the sample run, external examples, the manual, GOView link, citation and user forum. Clicking the “ORA Sample Run”, “GSEA Sample Run” or “NTA Sample run” can automatically fill out all parameters in the interface related to ORA, GSEA or NTA method, respectively. “External Examples” link provides some examples that can guide users prepare their own data for the analysis. Clicking the GOView link can jump to the GOView web tool for GO list comparison. The right part of the page also includes the introduction of the WebGestalt 2017, Data source introduction and News (red box in Figure 1). The following sections will introduce the detailed information about the input parameters and output report of WebGestalt.

WEB-based Gene SeT Analysis Toolkit
WebGestalt *Translating gene lists into biological insights...*

[ORA Sample Run](#) | [GSEA Sample Run](#) | [NTA Sample Run](#) | [External Examples](#) | [Manual](#) | [GOView](#) | [Citation](#) | [User Forum](#) | [WebGestalt 2013](#)

Basic Parameters

Select Organism of Interest ⓘ

Select Method of Interest ⓘ

Select Functional Database ⓘ

Gene List

Select Gene ID Type ⓘ

Choose File | No file chosen | Reset

Upload Gene List (max size: 5 MB) ⓘ **OR**
Please enter gene ids... Clear

Reference Gene List

Select Reference Set for Enrichment Analysis ⓘ Reset

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type ⓘ
Choose File | No file chosen | Reset

Advanced parameters

Submit

Introduction

WebGestalt (WEB-based Gene SeT Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

Data Source

News

Figure 1. General Information

1. Select an interesting organism

A user needs to select an organism of interest from the drop-down menu that includes 12 organisms plus an “others” option. We will introduce “others” in the “Select a functional database” section.



ORA Sample Run | GSEA Sample Run | NTA Sample Run | External Examples | Manual | GOView | Citation | User Forum | WebGestalt 2013

Basic Parameters

Select Organism of Interest ⓘ

Select Method of Interest ⓘ

Select Functional Database ⓘ

Gene List

Select Gene ID Type ⓘ

Upload Gene List (max size: 5 MB) ⓘ

Reference Gene List

Select Reference Set for Enrichment Analysis ⓘ

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type ⓘ

Advanced parameters

Submit

WebGestalt (WEB-based Gene Set Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

Introduction

Data Source

News

Figure 2. Selection of an interesting organism.

2. Select an interesting method

A user needs to select a method of interest from the drop-down menu (see Figure 3).

WebGestalt *Translating gene lists into biological insights...*

ORA Sample Run | GSEA Sample Run | NTA Sample Run | External Examples | Manual | GOView | Citation | User Forum | WebGestalt 2013

» Basic Parameters

Select Organism of Interest hsapiens

Select Method of Interest **Overrepresentation Enrichment Analysis (ORA)**
Gene Set Enrichment Analysis (GSEA)
Network Topology-based Analysis (NTA)

Select Functional Database -- Functional Database Name --

Gene List

Select Gene ID Type -- Gene ID Type --

Choose File No file chosen

Upload Gene List (max size: 5 MB) **OR**
Please enter gene ids...

Reference Gene List

Select Reference Set for Enrichment Analysis -- Reference Gene Set --

OR

Upload User Reference Set File (max size: 5 MB) -- Reference Gene ID Type --

Choose File No file chosen

» Advanced parameters

» Introduction

WebGestalt (WEB-based Gene Set Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,004 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool *GOView* that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

» Data Source

» News

Figure 3. Selection of an interesting method.

The different methods will have the different parameter input.

3. Select a functional database

3.1 For “others” organism

If a user selects “others” from the “Select Organism of Interest” menu, the user can analyze any kind of data not currently provided by WebGestalt. The user needs to upload the functional database, genes of interest, and reference genes (for the ORA method). Because WebGestalt will not perform any ID mapping for the uploaded data (all ID type related options are disabled (blue box in Figure 4)), the user needs to make sure that the ID type in all uploaded data is the same. The uploaded functional database file (red box in the Figure 4) should have an extension of “**gmt**”. As shown in Table 1, the first column of the file is the gene set ID, the second column is the external link to the gene set and other columns are the gene IDs annotated to this gene set. The file should be tab-delimited. If each gene set ID also has a description (e.g. the name of the gene set ID), the user can also upload a “**des**” file (this is **optional**, green box in the Figure 4). As shown in Table 2, the first column is the gene set ID which should be identical to the ID in the gmt file, and the second column is the description for each gene set. All columns should be separated by a **tab**.



ORA Sample Run | GSEA Sample Run | NTA Sample Run | External Examples | Manual | GOView | Citation | User Forum | WebGestalt 2013

Introduction

WebGestalt (WEB-based Gene SeT Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

Data Source

News

Basic Parameters

Select Organism of Interest: others

Select Method of Interest: Overrepresentation Enrichment Analysis (ORA)

Upload Functional Database (max size: 5 MB): Choose File No file chosen

Upload Database Description File (optional, max size: 5 MB): Choose File No file chosen

Gene List

Select Gene ID Type: Gene ID Type

Upload Gene List (max size: 5 MB): OR Please enter gene ids... Clear

Reference Gene List

Select Reference Set for Enrichment Analysis: Reference Gene Set

Upload User Reference Set File (max size: 5 MB) and Select ID Type: Reference Gene ID Type

Advanced parameters

Submit

Figure 4. The “others” option for “Select Organism of Interest”.

Table1. An example of the “gmt” file

Geneset1	http://www.webgestalt.org/Geneset1	Gene11	Gene12	Gene13
Geneset2	http://www.webgestalt.org/Geneset2	Gene21	Gene22	
Geneset3	http://www.webgestalt.org/Geneset3	Gene31	Gene32	

Table 2. An example of the “des” file

Geneset1	The description of Geneset1
Geneset2	The description of Geneset2
Geneset3	The description of Geneset3

3.2 For one of the 12 organisms

If a user selects one of the 12 organisms, there is a drop-down menu to show eight categories: geneontology, pathway, network, phenotype, disease, drug, chromosomalLocation and “others” option (red box in Figure 5). **Except “others”**, after selecting one of the other seven classes, the detailed database name in the class will be shown in another drop-down menu (see Figure 6). For gene ontology categories, we used an algorithm to remove the redundant terms and create three more databases only containing non-redundant terms. Considering the running time, the user can only use the non-redundant version for the **GSEA** GO enrichment analysis.



ORA Sample Run | GSEA Sample Run | NTA Sample Run | External Examples | Manual | GOView | Citation | User Forum | WebGestalt 2013

Basic Parameters

Select Organism of Interest:

Select Method of Interest:

Select Functional Database:

Gene List

Select Gene ID Type:

Upload Gene List (max size: 5 MB):

Reference Gene List

Select Reference Set for Enrichment Analysis:

Upload User Reference Set File (max size: 5 MB) and Select ID Type:

Advanced parameters

Submit

Introduction

WebGestalt (WEB-based Gene SeT Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.


WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.


Data Source


News

Figure 5. Selection of the functional database class.


Basic Parameters

Select Organism of Interest :

Select Method of Interest :

Select Functional Database :

Gene List

Select Gene ID Type :

-- Functional Database Name --

Biological_Process

Cellular_Component

Molecular_Function


Biological_Process_noRedundant

Cellular_Component_noRedundant

Molecular_Function_noRedundant

Choose file No file chosen

Reset


Upload Gene List (max size: 5 MB) :

OR

Please enter gene ids...

Clear


Reference Gene List

Select Reference Set for Enrichment Analysis :

-- Reference Gene Set --

Reset

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type :

-- Reference Gene ID Type --

Choose File No file chosen

Reset

Advanced parameters

Submit

Introduction

WebGestalt (WEB-based Gene SeT Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.


Data Source


News


Figure 6. Selection of the functional database.


If the user selects the “others” category, the user can upload a new functional database not included in WebGestalt (see red box in Figure 7). Because the user has selected one of 12 organisms instead of the “others” organism, WebGestalt will perform the ID mapping for all uploaded files (all ID type related options are still active (blue box in Figure 7)), which means it is not necessary to upload the files with the same ID types. Thus, the user also needs to select the ID type of the uploaded functional database file (see Figure 9).

Basic Parameters

Select Organism of Interest :

Select Method of Interest :


Select Functional Database :

Select ID Type of the uploaded database :


-- Gene ID Type --

Choose File No file chosen

Reset


Upload Functional Database (max size: 5 MB) :

Choose File No file chosen

Upload Database Description File (optional, max size: 5 MB) :

Choose File No file chosen


Gene List

Select Gene ID Type :

-- Gene ID Type --

Choose File No file chosen

Reset


Upload Gene List (max size: 5 MB) :

OR

Please enter gene ids...

Clear


Reference Gene List

Select Reference Set for Enrichment Analysis :

-- Reference Gene Set --

Reset

OR

Upload User Reference Set File (max size: 5 MB) and Select ID Type :

-- Reference Gene ID Type --

Choose File No file chosen

Reset

Advanced parameters

Submit

Introduction

WebGestalt (WEB-based Gene SeT Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.


WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.


Data Source


News


Figure 7. The “others” class selection


Basic Parameters


Select Organism of Interest 

Select Method of Interest 


Select Functional Database 


Select ID Type of the uploaded database 

Upload Functional Database (max size: 5 MB) 


Upload Database Description File (optional, max size: 5 MB) 


Gene List

Select Gene ID Type 

Upload Gene List (max size: 5 MB) 

Reference Gene List

Select Reference Set for Enrichment Analysis 

Upload User Reference Set File (max size: 5 MB) and Select ID Type 

Advanced parameters

WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.


[Data Source](#)


[News](#)


Figure 8. The ID type selection for the uploaded functional database file.

For the NTA method, users only need to select the networks from the drop down menu (see Figure 9).


Basic Parameters


Select Organism of Interest 

Select Method of Interest 


Select Functional Database 


Gene List

Select Gene ID Type 

Upload Gene List (max size: 5 MB) 

Reference Gene List

Select Reference Set for Enrichment Analysis 

Upload User Reference Set File (max size: 5 MB) and Select ID Type 

Advanced parameters

WebGestalt (WEB-based Gene SeT AnaLysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

[Data Source](#)

[News](#)

Figure 9. Network selection for the NTA method.

4. Upload gene list

After selecting or uploading the functional database, the user needs to upload or paste a gene list. As shown in Figure 10, the user should first select an ID type related to the gene list from the drop-down menu (red box, the same with the ID type of functional data in Figure 8). Then, if the user selects the “ORA” method, the user can upload a “txt” file with only one column or paste a gene list to the text box (blue box in Figure 10). If the user selects the “GSEA” method, the user should upload a “rnk” file with two columns: gene IDs and the scores separated by tab (see Table 3) or paste gene IDs and the scores separated by tab in the text box.

WEB-based GENE SeT Analysis Toolkit
WebGestalt Translating gene lists into biological insights...

ORA Sample Run | GSEA Sample Run | NTA Sample Run | External Examples | Manual | GOView | Citation | User Forum | WebGestalt 2013

Basic Parameters

Select Organism of Interest hsapiens

Select Method of Interest Overrepresentation Enrichment Analysis (ORA)

Select Functional Database geneontology

Gene List

Select Gene ID Type -- Gene ID Type --

Upload Gene List (max size: 5 MB) Choose File No file chosen Reset

OR

Please enter gene ids... Clear

Reference Gene List

Select Reference Set for Enrichment Analysis -- Reference Gene Set -- Reset

OR

Upload User Reference Set File (max size: 5 MB) -- Reference Gene ID Type -- Choose File No file chosen Reset

Advanced parameters

Submit

Introduction

WebGestalt (WEB-based Gene SeT Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

Data Source

News

Figure 10. Gene list upload.

Table 3. An example of “rnk” file

Gene1	1.2
Gene2	-1.1
Gene3	-3.3
Gene4	5.5

5. Upload the reference gene list

For the “ORA” method, the user needs to select a reference gene list from a drop-down menu (see Figure 11) that contains most existing high-throughput platforms or select ID type and upload a “txt” file.

The “GSEA” method does not need a reference gene list.



ORA Sample Run | GSEA Sample Run | NTA Sample Run | External Examples | Manual | GOView | Citation | User Forum | WebGestalt 2013

» Basic Parameters

Select Organism of Interest [?]

Select Method of Interest [?]

Select Functional Database [?]

Gene List

Select Gene ID Type [?]

Upload Gene List (max size: 5 MB) [?] **OR**

Reference Gene List

Select Reference Set for Enrichment Analysis [?]

Upload User Reference Set File (max size: 5 MB) and Select ID Type [?]

» Advanced parameters

genome_protein-coding
genome
afly_Axiom_BioBank1.na35_probe
afly_Axiom_BioBank1.na35_rsid
afly_Axiom_PMA.na35_probe
afly_Axiom_PMA.na35_rsid
afly_Axiom_tx.v1.na35_probe
afly_Axiom_tx.v1.na35_rsid
afly_GenomeWideSNP_5.na35_SNP_probe
afly_GenomeWideSNP_5.na35_SNP_rsid
afly_GenomeWideSNP_6.na35_SNP_probe
afly_GenomeWideSNP_6.na35_SNP_rsid
afly_hc_g110
afly_hg_focus
afly_hg_v133_plus_2

» Introduction

WebGestalt (WEB-based Gene Set Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

» Data Source

» News

Figure 11. Selection of the reference gene list.


6. Advanced parameters


The user can also set some advanced parameters for the different methods.


Figure 12 shows the advanced parameters for the ORA and GSEA methods.

- Setting the “Minimum Number of Genes for a Category” will remove the categories with a size less than this number. The category size is calculated based on the number of overlapping genes between the annotated genes in the category and the reference gene list for the “ORA” method (or ranked gene list for the “GSEA” method).
- Setting the “Maximum Number of Genes for a Category” will remove the categories with a size greater than this number.
- “Multiple Test Adjustment” will set the FDR method from “BH”, “BY”, “bonferroni”, “Holm” and “Hommel”. This option is only for the “ORA” method.
- The “Significance Level” parameter has two options. “FDR” means the enriched categories will be identified based on FDR threshold and “TOP” means the categories will be first ranked based on the FDR and then the top most significant categories will be selected. **For the GSEA method, the “TOP” method will select the top most significant categories from positive and negative related categories separately.**
- “Number of Permutations”, “Number of Categories with Leading-Edge Genes” and “Collapse Method” parameters are just for the “GSEA” method. These indicate how many permutations to perform, how many categories will have leading edge genes outputted from GSEA and which method will be used for collapsing any duplicate IDs.
- “Number of categories visualized in the report” represents how many significant categories will be shown in the report. **For the GSEA method, this number represents how many positive related categories or how many negative related categories will be visualized. If this number is 40, the report can at most contain 80 significant categories.**
- If “Color in DAG” is continuous, the categories of the DAG structure in the output report will be colored based on the FDR. Otherwise, WebGestalt will use red color for ORA method or red/blue color for GSEA method to color the significant categories.


Basic Parameters


Select Organism of Interest 

Select Method of Interest 


Select Functional Database 

Gene List


Select Gene ID Type 

Upload Gene List (max size: 5 MB)  **OR**


Reference Gene List


Select Reference Set for Enrichment Analysis 


OR


Upload User Reference Set File (max size: 5 MB) 


Advanced parameters


Minimum Number of Genes for a Category 


Maximum Number of Genes for a Category 


Multiple Test Adjustment 


Significance Level  ☐ FDR ☐ TOP

Number of Permutations 

Number of Categories with Leading-Edge Genes 

Collapse Method 

Number of categories visualized in the report 

Color in DAG  ☐ Continuous ☐ Binary

Introduction

WebGestalt (WEB-based Gene Set Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool [GOView](#) that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

Data Source

News

Figure 12. Advanced parameters for the ORA and GSEA methods.

Figure 13 and 14 shows the advanced parameters for the two NTA methods.

- NTA includes two network construction method: “Network Retrieval & Prioritization” and “Network Expansion”. “Network Retrieval & Prioritization” first uses random walk analysis to calculate random walk probability for the input seeds, then identified the relationships among the seeds in the selected network and returns a retrieval sub-network. The seeds with the top random walk probability are highlighted in the sub-network. “Network Expansion” first uses random walk analysis to rank all genes in the selected network based on their network proximity to the input seeds and then return an expanded sub-network in which nodes are the input seeds and their top ranking neighbors and edges represent their relationships.
- For “Network Retrieval & Prioritization”, users need to set how many seeds will be highlighted in the retrieved sub-network. For “Network Expansion” method, users need to set how many top ranking neighbors will be included in the expanded sub-network and also to set which type of genes will be highlighted in the sub-network (seeds or neighbors).
- These two methods finally will perform the enrichment analysis for the identified sub-networks based on the GO BP ontology. Thus, users need to select the criteria for identifying the significant categories.

[Basic Parameters](#)

Select Organism of Interest

hsapiens

Select Method of Interest

Network Topology-based Analysis (NTA)

Select Functional Database

network

-- Functional Database Name --

Gene List

Select Gene ID Type

-- Gene ID Type --

Choose File No file chosen Reset

OR

Please enter gene ids...

Clear

Reference Gene List

Select Reference Set for Enrichment Analysis

-- Reference Gene Set --

Reset

OR

-- Reference Gene ID Type --

Choose File No file chosen Reset

Upload User Reference Set File (max size: 5 MB) and Select ID Type

Advanced parameters

Select Network Construction Method

Network Retrieval & Prioritization

Set Number of Highlighted Seed Genes

10

Select Significant Level Method for Enrichment Analysis

☒ FDR ☐ Top

Set FDR Threshold

0.05

Submit

[Introduction](#)

WebGestalt (WEB-based Gene Set Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

[Data Source](#)

[News](#)

Figure 13. Advanced parameters for the Network Retrieval & Prioritization method.

[Basic Parameters](#)

Select Organism of Interest

hsapiens

Select Method of Interest

Network Topology-based Analysis (NTA)

Select Functional Database

network

-- Functional Database Name --

Gene List

Select Gene ID Type

-- Gene ID Type --

Choose File No file chosen Reset

OR

Please enter gene ids...

Clear

Reference Gene List

Select Reference Set for Enrichment Analysis

-- Reference Gene Set --

Reset

OR

-- Reference Gene ID Type --

Choose File No file chosen Reset

Upload User Reference Set File (max size: 5 MB) and Select ID Type

Advanced parameters

Select Network Construction Method

Network Expansion

Set Number of Top Ranking Neighbors

10

Select Significant Level Method for Enrichment Analysis

☒ FDR ☐ Top

Set FDR Threshold

0.05

Highlight

☒ Seeds ☐ Neighbors

Submit

[Introduction](#)

WebGestalt (WEB-based Gene Set Analysis Toolkit) is a functional enrichment analysis web tool, which has been visited 209,028 times by 84,024 unique users from 144 countries and territories since 2013 according to Google Analytics. The WebGestalt 2005 and WebGestalt 2013 papers have been cited in 1179 scientific papers since 2013 according to Google Scholar.

WebGestalt 2017 significantly increased the number of supported organisms, gene identifiers, and functional categories in WebGestalt. Notably, experimental data from organisms or with gene identifiers not covered by the WebGestalt database can also be analyzed in WebGestalt. WebGestalt also supports three well-established and complementary methods for enrichment analysis, including Over-Representation Analysis (ORA), Gene Set Enrichment Analysis (GSEA), and Network Topology-based Analysis (NTA). To facilitate easy exploration and better understanding of the enrichment results, we have revamped the output interface with a user-friendly, tab-based, and interactive report. We have also developed a companion tool GOView that can help visualize and compare multiple Gene Ontology (GO) enrichment results under the GO Directed Acyclic Graph (DAG) structure.

[Data Source](#)

[News](#)

Figure 14. Advanced parameters for the Network Expansion method.

7. Output report for 12 organisms based on ORA or GSEA method

As shown in Figure 15, if the ID type of the uploaded data is from one of 12 organisms, the output report will contain four tabs: “Summary”, “User ID Mapping Table”, “GOSlim Summary” and “Enrichment Results”. the Summary tab includes the upload information, parameters used in the analyses and the result summary. Clicking the “Result Download” link will download the mapping tables, the GO Slim summary figure, enrichment results and the summary report.

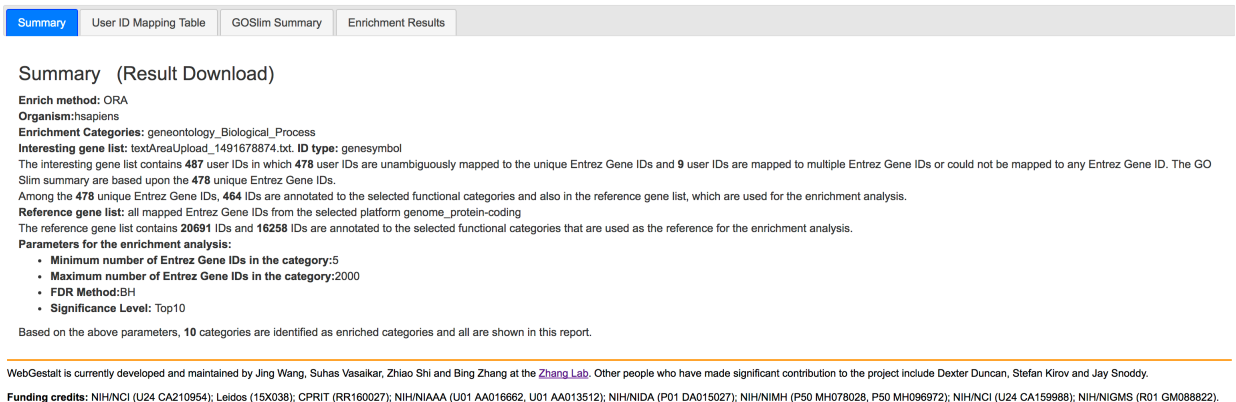


Figure 15. The Summary tab of the output HTML report.

The “User ID Mapping Table” tab includes two tables (see Figure 16). The “Mapped User IDs” table contains the uploaded gene list that maps to unique Entrez gene IDs. Any IDs that map to multiple Entrez gene IDs or that do not map to any Entrez gene IDs are included in the second table.

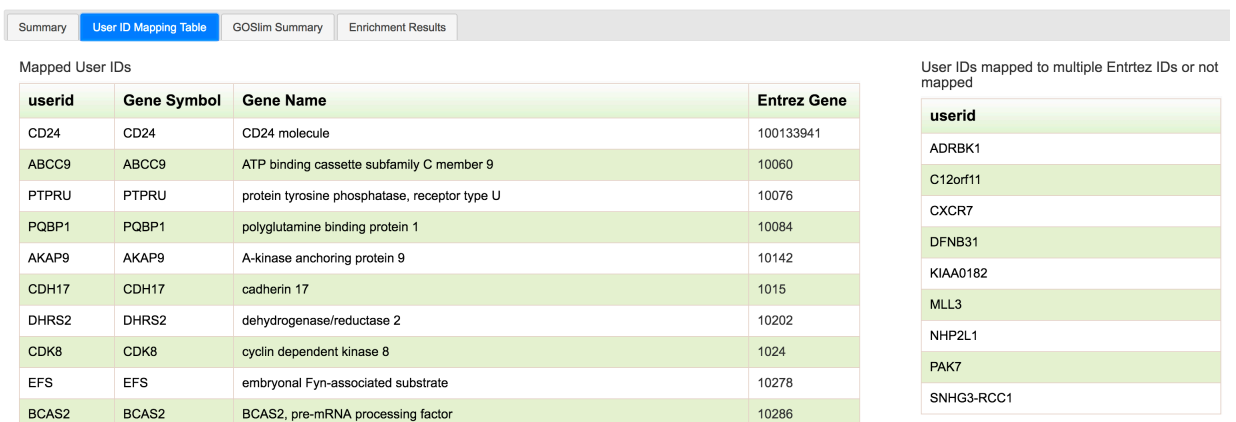


Figure 16. The User ID Mapping Table tab of the output HTML report.

The GOSlim summary tab includes three bar plots that represent the number of genes in the uploaded gene list that overlap with the annotated genes in the GO Slim terms from biological process (red bar plot), cellular component (blue bar plot) and molecular function (green bar plot) ontologies, respectively (see Figure 17).



WEB-based GENE SeT Analysis Toolkit

WebGestalt Translating gene lists into biological insights...

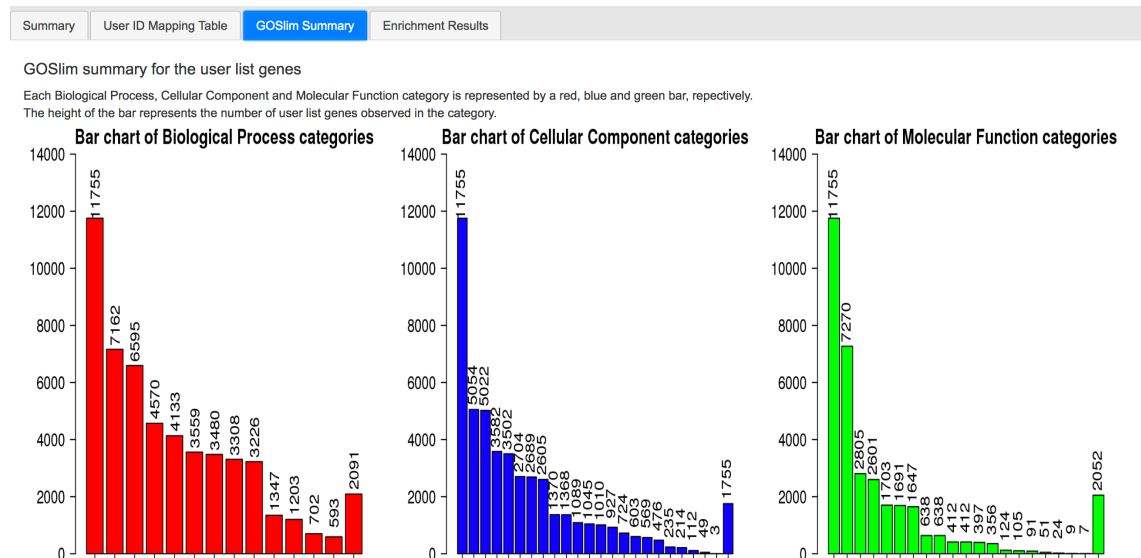


Figure 17. GOSlim Summary tab of the output HTML report.

The “Enrichment Results” tab contains the detailed information of the enrichment results (see Figure 18). If the functional database contains the DAG or tree structure, the structure will be visualized in this tab. If the “Color in DAG” is continuous, the significant categories will be colored by red gradient based on the FDR. Otherwise, the categories will be colored by red. For the GSEA method, the positive related categories will be colored by red gradient while the negative related categories will be colored by blue gradient if the “Color in DAG” is continuous. Otherwise, the positive and negative related categories will be colored by red and blue, respectively. The user can scroll the mouse to zoom-in and zoom-out and can view different portions of the DAG by clicking and dragging the small box in the right-bottom corner (red box). The user can also right-click one node to select the neighbors or all related paths (orange box). Clicking the save button in the left-bottom corner (blue box) can save the DAG in the PDF, SVG or PNG format. The other two buttons allow searching for a node in the DAG or viewing the entire DAG. Clicking the colored nodes (significant categories) can make the right part of the tab jump to the corresponding table (brown box). The table contains the enrichment statistic and the annotated genes uploaded by the user. Clicking the link in the table can access the external link of the category. The “Download Table” link can let the user download only the table

content of this category. The user can also search the category by typing a key word in the search box (black box).

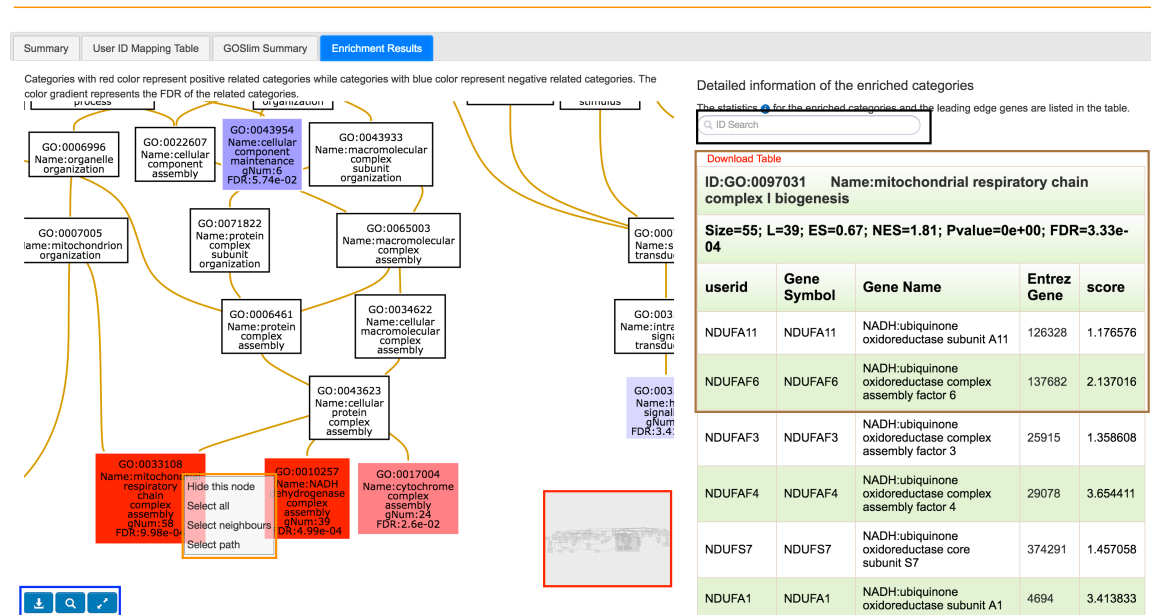


Figure 18. Enrichment result tab for the categories with DAG structure.

If the user performs the enrichment analysis for TCGA co-expression modules or protein interaction modules, clicking the “Network View” link (red box in Figure 19) can visualize the sub-network related to the enriched module. If the enriched method is “ORA”, the overlap genes between the uploaded gene list and the annotated genes in the module will be colored as green in the network. If the user selects the “GSEA” method, the leading-edge genes will be colored by a color gradient based on the corresponding scores in the upload file. The genes with the positive values will be colored from white to red while genes with the negative values will be colored from blue to white.

Note: For the Safari browser user, please allow the flash first before visualizing the network. Detailed information about how to allow the flash can be found in <http://osxdaily.com/2013/12/18/enable-flash-plugin-specific-websites-safari-mac/>.

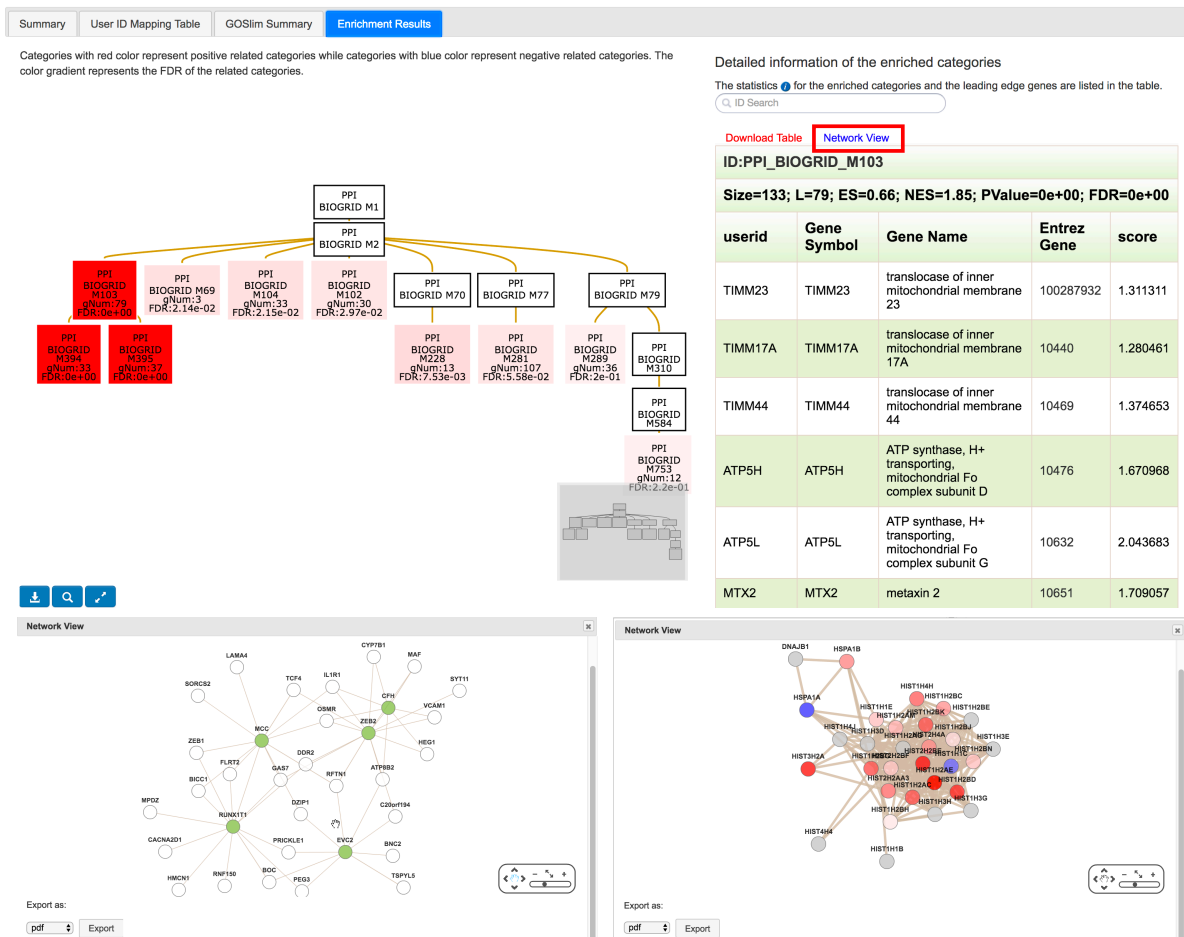


Figure 19. Enrichment result tab for the protein interaction modules or TCGA co-expression modules.

If the categories do not contain the DAG structure, the new version will use a table to summarize the significant categories (see Figure 20). For the GSEA method, the positive related categories in the table will be colored as red while the negative related categories will be colored as blue. Clicking the ID link (blue box) can make the right part of the tab jump to the corresponding table. If the functional database is KEGG, clicking the link in the table can highlight the annotated genes uploaded by the user in the pathway by changing the label to red (see left-bottom figure of Figure 20). If the functional database is Wikipathway, clicking the link can show not only the pathway but also highlight the genes by changing the color of the gene's border to red (see right-bottom figure of Figure 20).

If the user selects the functional database category as "others", the output report will be the same as Figure 20.

Summary of the enriched categories

This table lists the enriched categories, number of entrez genes in the user gene list and also in the categories and FDR. Categories with red color represent the positive related categories while categories with blue color represent the positive related categories.

ID	Name	#Gene	FDR
hsa00190	Oxidative phosphorylation - Homo sapiens (human)	106	0e+00
hsa04260	Cardiac muscle contraction - Homo sapiens (human)	38	0e+00
hsa04932	Non-alcoholic fatty liver disease (NAFLD) - Homo sapiens (human)	122	0e+00
hsa05012	Parkinson's disease - Homo sapiens (human)	109	0e+00
hsa05010	Alzheimer's disease - Homo sapiens (human)	134	1.67e-04
hsa05016	Huntington's disease - Homo sapiens (human)	151	2e-04
hsa03010	Ribosome - Homo sapiens (human)	131	8.56e-04
hsa00640	Propanoate metabolism - Homo sapiens (human)	27	1.51e-02
hsa00020	Citrate cycle (TCA cycle) - Homo sapiens (human)	29	1.7e-02
hsa04978	Mineral absorption - Homo sapiens (human)	36	2.79e-02
hsa04810	Complement and coagulation cascades - Homo sapiens (human)	24	1.3e-01

Detailed information of the enriched categories

The statistics for the enriched categories and the leading edge genes are listed in the table.

ID Search

Download Table

ID:hsa00190 Name:Oxidative phosphorylation - Homo sapiens (human)

Size=106; L=78; ES=0.7; NES=1.96; Pvalue=0e+00; FDR=0e+00

userid	genesymbol	genename	entrezgene	score
NDUFC2-KCTD14	NDUFC2-KCTD14	NDUFC2-KCTD14 readthrough	100532726	1.119929
ATP5H	ATP5H	ATP synthase, H+ transporting, mitochondrial Fo complex subunit D	10476	1.670968
ATP5L	ATP5L	ATP synthase, H+ transporting, mitochondrial Fo complex subunit G	10632	2.043683
UQCRC1	UQCRC1	ubiquinol-cytochrome c reductase, complex III subunit XI	10975	1.747905

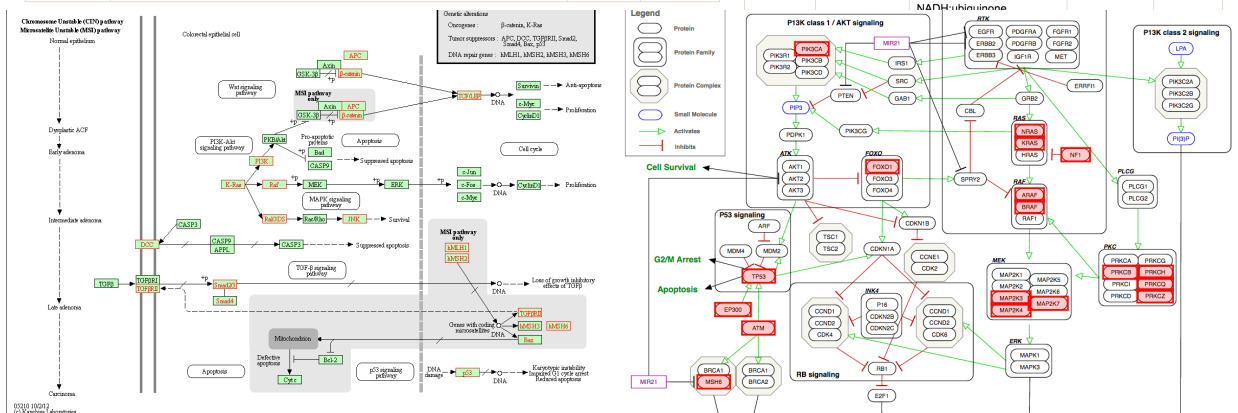


Figure 20. Enrichment result tab for the categories without DAG structure.

NOTE: output report for each analysis is associated with a unique Uniform Resource Locator (URL) and thus can be easily shared.

8. Output report for “others” organisms

If the user selects the organism as “others”, there will only have the summary section and the enriched result table that contains the statistic and the annotated genes in the uploaded list (see Figure 21). For the GSEA method, the positive related categories in the table will be colored as red while the negative related categories will be colored as blue.



Summary [\(Result Download\)](#)

Enrich method: GSEA
Organism: Others
Enrichment Categories: uploads/hsapiens_pathway_Wikipathway_1487804492.gmt Description File: uploads/hsapiens_pathway_Wikipathway_1487804492.des
Interesting List: GncfKunL1st_gencid_1487804492.rnk
The file contains 11755 user IDs (no ID mapping). All these IDs were used to perform the enrichment analysis.
Parameters for the enrichment analysis:
• Minimum number of Entrez Gene IDs in the category:5
• Maximum number of Entrez Gene IDs in the category:2000
• Significance Level: 10p10
• Number of permutation:1000
• Number of categories with the outputted leading edge genes:20

Based on the above parameters, 10 positive related categories were identified as enriched categories and all were shown in this report.
Based on the above parameters, 10 negative related categories were identified as enriched categories and all were shown in this report.

Detailed information of the enriched categories

The statistics [●](#) for the enriched categories and the leading edge genes are listed in the table.
Categories with red color represent positive related categories while categories with blue color represent negative related categories.

ID	Name	Statistic	Genes
WP111	Electron Transport Chain	Size=88; L=71; ES=0.73; NES=2.02; PValue=0.00e+00; FDR=0.00e+00	4694 4725 6390 9551 1345 4714 522 521 518 7184 1347 1329 515 539 4710 1340 4731 10632 4702 9377 1349 7386 7381 4701 4704 4708 6389 27089 7388 10975 513 4713 1337 10476 506 4726 4729 1327 4696 516 55967 514 7385 6392 29796 4697 4709 374291 4718 9481 4716 4698 4695 509 1351 4706 4722 4707 1355 293 4720 4723 4700 517 9167 93974 4711 4717 4719 1350 6391
WP623	Oxidative phosphorylation	Size=51; L=41; ES=0.73; NES=1.98; PValue=0.00e+00; FDR=0.00e+00	4725 9551 4714 522 521 518 515 539 4710 4731 10632 4702 4701 4704 4708 513 4713 10476 506 4726 4729 4696 516 514 4697 374291 4718 4716 4698 4695 4706 4722 4707 4720 126328 4723 4700 517 4711 4717 4719
WP477	Cytoplasmic Ribosomal Proteins	Size=88; L=71; ES=0.63; NES=1.75; PValue=0.00e+00; FDR=1.92e-02	6206 6231 25873 6161 6235 6229 6230 6205 7311 6218 6167 6169 6168 6173 6227 6158 6171 9045 6181 6197 6136 6234 6201 6157 9801 6135 6128 6176 6146 6134 6155 6159 6232 6191 11224 6207 6195 2197 6196 6165 6152 6170 6204 6223 6208 6233 6175 6142 6193 6224 6125 6187 6188 6209 6194 6222 3921 6210 4736 6203 6228 6156 6160 6138 6217 6129 6134 6143 6130 6139 6147
WP78	TCA Cycle	Size=17; L=13; ES=0.67; NES=1.6; PValue=5.15e-03; FDR=1.87e-01	6390 4191 8801 3419 8802 3418 6389 1431 6392 1743 50 6391 4967
WP299	Nuclear Receptors in Lipid Metabolism and Toxicity	Size=16; L=5; ES=0.62; NES=1.46; PValue=1.43e-02; FDR=5.54e-01	7421 8856 5468 1555 5243
WP1602	Nicotine Activity on Dopaminergic Neurons	Size=10; L=7; ES=0.69; NES=1.5; PValue=2.04e-02; FDR=5.66e-01	1813 6571 84152 1644 5566 5499 1138
WP2895	Differentiation of white and brown adipocyte	Size=11; L=7; ES=0.66; NES=1.47; PValue=2.13e-02; FDR=5.97e-01	5468 655 1050 650 652 10205 1054
WP2453	TCA Cycle and Deficiency of Pyruvate Dehydrogenase complex (PDHc)	Size=16; L=11; ES=0.62; NES=1.47; PValue=2.97e-02; FDR=6.52e-01	5160 8801 3419 5105 6389 1431 1737 1743 4967 47 2271
WP311	Synthesis and Degradation of Ketone Bodies	Size=5; L=1; ES=0.75; NES=1.42; PValue=5.57e-02; FDR=6.63e-01	3158
WP704	Methylation Pathways	Size=6; L=2; ES=0.72; NES=1.42; PValue=5.40e-02; FDR=6.97e-01	3176 7172
WP1423	Ganglio Sphingolipid Metabolism	Size=6; L=4; ES=0.68; NES=1.7; PValue=8.06e-03; FDR=2.22e-01	2523 8869 30815 6482
WP561	Heme Biosynthesis	Size=8; L=2; ES=0.54; NES=1.58; PValue=1.96e-02; FDR=2.38e-01	2235 7389
WP338	Complement and Coagulation Cascades	Size=20; L=8; ES=0.35; NES=1.48; PValue=0.00e+00; FDR=2.75e-01	1213 3998 2149 5054 5627 5327 5329 7035

Figure 21. The output report for the “others” organism.

9. Output for the NTA method

Figure 22 shows the output for the NTA method. The detailed information about this output can be found in the www.gene2net.org.

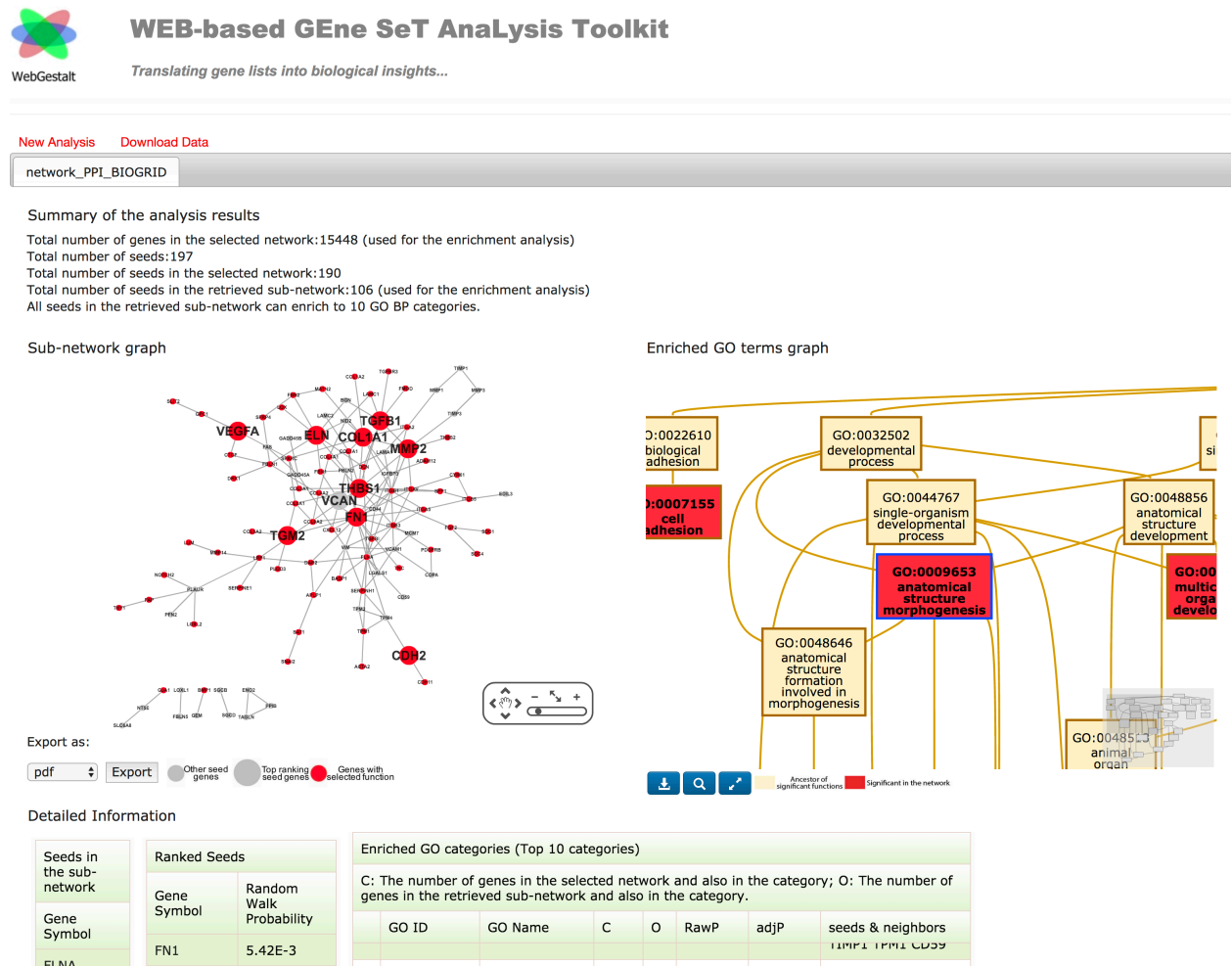


Figure 22. Output for the NTA method.