

ClueGO - CluePedia

Frequently asked questions

Gabriela Bindea, Bernhard Mlecnik
Laboratory of Integrative Cancer Immunology
INSERM U872 Cordeliers Research Center
Paris, France

Contents

License	2
Documentation	2
How to cite ClueGO and CluePedia	3
Installation	3
Organisms and ontologies	4
Network generation	5
Stats	8
Results: network, tables, figures	9
Others	10

License

Why is the license key needed?

ClueGO and CluePedia are freely available for non-commercial users. The license key is just used to estimate the number of the ClueGO/CluePedia users.

I would like a ClueGO commercial license key.

For a commercial license key please contact jerome.galon@crc.jussieu.fr.

I would like to request a license key for the ClueGO plugin; however, I am unable to access the license key request site.

This could be due to a blocked outgoing port (29080) on your side.

Contact us directly (gabriela.bindea@crc.jussieu.fr). For the license we need: the name, institution, department, address (street, city, post code, country), email. Optional: the title, telephone, fax, url of the institution.

I did not get the license key within one week time.

Please recontact us (gabriela.bindea@crc.jussieu.fr). Sometimes the email submitted in the form is not correct, and we cannot write you back.

I just requested the license key for the ClueGO plugin but after pasting it it tells me that my license expired.

Verify if there were not added some empty spaces to the key you are pasting in.

If this is not the case, rename the `.clugoplugin` folder (e.g. `.cluegoplugin.old`). Paste the latest ClueGO plugin version into the `plugins` folder of Cytoscape and restart Cytoscape.

Documentation

Where can be found ClueGO and CluePedia documentation?

Documentation and step-by-step example analysis are available:

- in our site: <http://www.ici.upmc.fr/>

- in Cytoscape app store: <http://apps.cytoscape.org/apps/>

<http://www.ici.upmc.fr/cluego/cluegoDocumentation.shtml>

<http://www.ici.upmc.fr/cluepedia/cluepediaDocumentation.shtml>

How to cite ClueGO and CluePedia

If you find our tools useful, please cite us:

Bindea G*, Mlecnik B*, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman WH, Pages F, Trajanoski Z and Galon J. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology annotation networks. *Bioinformatics* 2009.25(8):1091-1093

CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data. Bindea G, Galon J and Mlecnik B. *Bioinformatics*. 2013 Jan 16. [Epub ahead of print]

Installation

How I start ClueGO/CluePedia?

Download the latest versions of the plugins and save them in the Plugins folder of Cytoscape. Restart Cytoscape. The plugins will be visible under the Cytoscape Plugins menu. The plugins can be downloaded via the Cytoscape manager as well.

Where are saved the ClueGO files in my system?

The ClueGO files are saved in the .cluegoplugin folder, a hidden folder placed in the user home directory.

Organisms and ontologies

How can I create ClueGO files for an organism not supported by ClueGO?

A new organism can be created by the user. A new .properties file containing the tax id, organism identifier in GO, KEGG, Reactome as well as links for the source of ontologies and annotation files should be placed in a dedicated folder in the ClueGO Organism File Source. The naming of the new folders and files should be similar to the ones created for other organisms. Upon request, we provide files for new organisms. The files are based on information from KEGG, GO, UniProt-GOA annotations and NCBI (for gene info). The download of the new organism is automatically done.

There are new ontologies/interaction files available?

Starting with ClueGO v1.5 new ontologies (CORUM,) can be automatically downloaded.

How can I update the ontologies?

The user can automatically extract informations from GO, KEGG and Reactome and create new ClueGO source files. The update of the gene annotations (NCBI based) should be done before the update.

All the species will be updated in the same time?

The update is made separately for each specie. Only the selected organism will be updated.

I'm trying to update the GO annotations, but a message appears saying time out in connection.

Verify that you have the latest version of the plugins. Some time, the structure of the annotation files used as source of information changes. ClueGO changes accordingly. Please let us know if you notice this problem.

How can I select genes annotated with any of the experimental evidence codes: EXP, IDA, IPI, IMP, IGI and IEP? Would it be possible to add an option in the GO evidence menu comprising "All experimental evidence codes"?

This option was included in ClueGO 1.5. All experimental evidence code can be directly selected. In a ClueGO network showing all types of codes, for a better visibility the experimental evidence code are draw thicker compared to IEA codes, for example. The type of the code can be visualized as edge labels in the network.

Network generation

How the ClueGO network is created?

The ClueGO input is a list of genes. Those genes can be uploaded from a .txt or from a network. Genes are mapped on selected ontologies. From all possible terms, the ones meeting the set criteria (GO level, # genes, % genes) are selected, and for those a pval is calculated using the fisher exact test (an example of pval calculation is included in the documentation). Terms are connected based on shared genes (kappa score). Kappa score is used to define groups.

I try to do a ClueGO analysis and I get the error message "There were no GO Terms found for this selection!". What can I do?

The error occurs if the ID type selected in ClueGO is different from the IDs from user's list. In case the id type is not supported by ClueGO, a file containing entrezIDs and the corresponding new id type can be included in ClueGO as additional resource. The file format is described here: (<http://www.ici.upmc.fr/cluego/cluegoDocumentation.shtml>). This error message occurs also e.g. when a small list of genes is analyzed with too restrictive selection criteria. In this case, the number of genes/term and the percentage that those represent in the term should be reduced, or the GO interval used enlarged.

Not all the genes are annotated in GO. Using CluePedia, the role of those genes without GO associations could be estimated/predicted by using in silico scores, or calculating experimentally derived interrelations with genes known to be involved in pathways/terms.

What means "# genes" and "% genes" per term? Do the two parameters act independently of each other?

The # and % of genes, together with the GO level, are used to select terms from ontologies. Those features can be used independently or together, depending on the analysis you want to perform and the number of genes to analyze. In case of large lists of genes, we usually start with general settings, to see how the genes associate in functions, and once interesting directions are defined, we use more precise settings. Usually we use all those parameters in analysis.

If you have to analyze a list of 200 genes there will be many GO terms found. Between those, terms with many genes associated (e.g. 500 genes), but with only 1 gene from your list. Those general terms are maybe not the most relevant for your list. By setting a min nr of genes, those terms can be discarded.

The % shows how much represent the initial genes mapped to a term compared to all the other genes associated to this term. With the % you can find more "specific" terms. For example, 1 gene found can represent different %, depending on how many genes are associated to the terms: a term T1 with 300 genes ("general" term) and 1 gene (0.3%) found. Many other genes

are involved in this function. A term T2 with 10 genes (more "specific" term) and 1 gene (10%) found.

To restrict the network you can use in addition the fusion feature (that removes the redundant parent-child terms) or you can visualize only the significant terms (feature available starting with ClueGO v1.5).

In ClueGO v1.4 there was a button "Show Associated Genes" that is missing in v1.5. Is there another way to add the associated genes to the graph?

This function was included in CluePedia. Beside showing genes associated to the terms, gene-gene known interaction as well as methods to calculate interactions from custom data are now available. Gene/miRNA enrichments of networks are possible, as well as in detail investigation of ClueGO pathways/terms (in nested networks). The Save button from ClueGO v1.4 is missing in ClueGO v1.5. This was fixed. Please download the latest plugin (www.ici.upmc.fr).

I created a network with GO term connection restriction of 0.3 and I see on my network that some edges are thicker than others. What does this mean?

The edges show the kappa score: how many genes in common have two terms. If a line is thicker means that the connected terms have more genes in common than others.

Why some terms are included more than once in the overview pie chart? Is there a way out to avoid the repeats?

The overview pie chart shows the proportion that represents each group from the total (number of the terms in each group reported to the rest). The name of the group is by default the most significant term of the group. A term can be part of several groups (the groups are overlapping) and it can be the most significant in each of those groups. This is the reason for which the overview chart can have several time the term (Term, Term_1, Term_2). To avoid this, one could

increase the kappa score and/or the % for group merge. In this way the groups will not overlap. On the other hand, those restrictions will result in many not grouped terms.

The selection of the leader term was improved to avoid duplicates in group names and in the overview chart starting with ClueGO v2.0.0 working with Cytoscape 3.0.0.

It would be possible to combine experimental (e.g. expression) data with ClueGO functional analysis? How visualize all proteins belonging to a certain group?

To visualize and analyze experimental data we developed CluePedia Cytoscape plugin. Expression data can be extracted, normalized and visualized as node label on the network. Built-in statistical methods can be used to derive gene-gene interrelations specific for the experimental data used. CluePedia extends ClueGO functional analysis. Expression data of genes associated to a term/pathway can be visualized. More, the pathway can be enriched with related genes reflecting experimental settings. Validated or predicted miRNAs that could influence the expression levels can be visualized on the network as well.

How to predict functions of genes not annotated in GO?

Not all genes are annotated in GO. The role of genes without GO associations could be estimated/predicted using CluePedia. Within CluePedia, in silico scores and/or scores calculated from experimental data can be used to derive interrelations between genes without functional annotations and genes known to be involved in pathways/terms. Based on this, the role of the not annotated genes can be estimated, role that should be afterwards experimentally validated.

Stats

Can I restrict the P values or the adjusted P values within ClueGO?

Starting with ClueGO 1.5 it is possible to set a pvalue threshold for the terms to be visualized in the network.

What the red stars (single and double) in the histogram mean?

The stars represent the level of significance of the term. You can find how is this calculated in the Statistics section of the ClueGO documentation.

How to use a different reference than the selected ontologies?

Starting with ClueGO v1.5, several options for selecting the reference were made available.

Results: network, tables, figures

For big networks, the output file resolution is low. How can I improve the quality of the figures?

The quality of the network figure can be improved using Cytoscape features (Cytoscape menu bar). Select File/Export/Current Network View as Graphics, set the location of the image and the format e.g.: .png. A window with image resolution settings will appear. Set e.g. Zoom: 300% and Resolution: 300 DPI.

Starting with ClueGO 1.6 the charts are saved in .svg as well.

The genes associated to the pathway are not linked in the nested network.

How can I get their interactions?

Select first the interaction file you are interested in, and then create the nested network. The interactions will be automatically visualized.

The update command applies to both, the main network and the nested network. How to apply the update on the nested network only?

The update is now applying separately on the main and nested networks. Please download the latest plugin (www.ici.upmc.fr).

How to zoom in the Cerebral view?

Use the middle mouse button.

In the "node attribute" table, only some terms are "true", the rest are false.

Why is that?

The group term leaders (by default the most significant terms from each group) are marked in the node attribute table as "true". Those terms will be shown in the overview chart. The histogram instead shows all the terms from each group.

Others

Is there a web forum where users can exchange tips and tricks to use ClueGO and CluePedia?

We recommend the use the excellent frame of Cytoscape App Store: <http://apps.cytoscape.org/apps/>