



## SOFTWARE TOOL ARTICLE

# REVISED Contextual Hub Analysis Tool (CHAT): A Cytoscape app for identifying contextually relevant hubs in biological networks [version 2; referees: 2 approved]

Tanja Muetze<sup>1\*</sup>, Ivan H. Goenawan<sup>1\*</sup>, Heather L. Wiencko<sup>2</sup>, Manuel Bernal-Llinares<sup>1</sup>, Kenneth Bryan<sup>1</sup>, David J. Lynn<sup>1,3</sup>

<sup>1</sup>EMBL Australia Biomedical Informatics Group, Infection & Immunity Theme, South Australian Medical and Health Research Institute, Adelaide, Australia

<sup>2</sup>Animal and Bioscience Research Department, Animal and Grassland Research and Innovation Centre, Teagasc, Meath, Ireland

<sup>3</sup>School of Medicine, Flinders University, Bedford Park, Australia

\* Equal contributors

**v2** First published: 19 Jul 2016, 5:1745 (doi: [10.12688/f1000research.9118.1](https://doi.org/10.12688/f1000research.9118.1))  
Latest published: 30 Aug 2016, 5:1745 (doi: [10.12688/f1000research.9118.2](https://doi.org/10.12688/f1000research.9118.2))

## Abstract

Highly connected nodes (hubs) in biological networks are topologically important to the structure of the network and have also been shown to be preferentially associated with a range of phenotypes of interest. The relative importance of a hub node, however, can change depending on the biological context. Here, we report a Cytoscape app, the Contextual Hub Analysis Tool (CHAT), which enables users to easily construct and visualize a network of interactions from a gene or protein list of interest, integrate contextual information, such as gene expression or mass spectrometry data, and identify hub nodes that are more highly connected to contextual nodes (e.g. genes or proteins that are differentially expressed) than expected by chance. In a case study, we use CHAT to construct a network of genes that are differentially expressed in Dengue fever, a viral infection. CHAT was used to identify and compare contextual and degree-based hubs in this network. The top 20 degree-based hubs were enriched in pathways related to the cell cycle and cancer, which is likely due to the fact that proteins involved in these processes tend to be highly connected in general. In comparison, the top 20 contextual hubs were enriched in pathways commonly observed in a viral infection including pathways related to the immune response to viral infection. This analysis shows that such *contextual hubs* are considerably more biologically relevant than degree-based hubs and that analyses which rely on the identification of hubs solely based on their connectivity may be biased towards nodes that are highly connected in general rather than in the specific context of interest.

Availability: CHAT is available for Cytoscape 3.0+ and can be installed via the Cytoscape App Store (<http://apps.cytoscape.org/apps/chat>).

## Open Peer Review

Referee Status:

	Invited Referees	
	1	2
<b>REVISED</b> version 2 published 30 Aug 2016		 report
version 1 published 19 Jul 2016	 report	

1 **Sandra Orchard**, European Molecular Biology Laboratory-European Bioinformatics Institute UK, **Pablo Porras Millán**, European Molecular Biology Laboratory-European Bioinformatics Institute UK

2 **Christopher K. Tuggle**, Iowa State University USA, **Haibo Liu**, Iowa State University USA

## Discuss this article

Comments (0)



This article is included in the [Cytoscape apps](#) channel.

**Corresponding author:** David J. Lynn ([David.Lynn@sahmri.com](mailto:David.Lynn@sahmri.com))

**How to cite this article:** Muetze T, Goenawan IH, Wiencko HL *et al.* **Contextual Hub Analysis Tool (CHAT): A Cytoscape app for identifying contextually relevant hubs in biological networks [version 2; referees: 2 approved]** *F1000Research* 2016, 5:1745 (doi: [10.12688/f1000research.9118.2](https://doi.org/10.12688/f1000research.9118.2))

**Copyright:** © 2016 Muetze T *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

**Grant information:** The research leading to these results received funding from the European Union Seventh Framework Programme (FP7/2007-2013) PRIMES project under grant agreement number FP7-HEALTH-2011-278568. The Lynn Group is also supported by EMBL Australia.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

**Competing interests:** No competing interests were disclosed.

**First published:** 19 Jul 2016, 5:1745 (doi: [10.12688/f1000research.9118.1](https://doi.org/10.12688/f1000research.9118.1))

**REVISED** Amendments from Version 1

Incorporated Sandra Orchard's and Pablo Porras's suggestions to add a reference, to clarify that interactions between input list genes or proteins are considered in the network creation and calculations and to elucidate that CHAT can not only be applied to gene lists but also works for protein lists.

See referee reports

## Introduction

Network analysis has emerged as a powerful approach to elucidate biological and disease processes<sup>1</sup>. Biological networks (and many other types of networks) have been shown to have a power law distribution of node connectivity, with most nodes having few connections and a few nodes being highly connected<sup>2</sup>. The identification of such highly connected nodes, termed hubs, is often of interest as hubs have been shown to be topologically and functionally important. The deletion of genes encoding hub proteins, for example, has been shown to correlate with lethality in yeast (the centrality-lethality rule)<sup>3</sup>. Hubs have also been found to be preferentially targeted by both bacterial and viral pathogens<sup>4</sup> and may be master regulators of biological processes<sup>5</sup>. Biological networks, such as the human interactome, however, are not static entities<sup>6</sup>, and the extent to which a node acts as a hub can change depending on the biological context e.g. the network present in a specific cell type at a particular point in time<sup>7,8</sup>. Integrating contextual information, such as gene or protein expression data, with standard network analysis can provide insight into what are the most relevant network features in a particular study or context<sup>9-11</sup>.

Cytoscape has a number of applications to identify hubs in networks including cytoHubba<sup>12</sup>, APID2Net<sup>13</sup>, PinnacleZ<sup>14</sup>, NetworkAnalyzer<sup>15,16</sup> and CentiScaPe<sup>17</sup>, however, only the latter two are compatible with Cytoscape 3+. All of the applications available to date identify hubs based on node connectivity (degree) in a network of interest. To construct a network, users frequently query interaction databases to identify the interactors of a list of genes of interest, e.g. differentially expressed genes, and then identify the high degree nodes in this network. This approach to constructing a network is useful because it identifies a more fully connected network for analysis than would be the case if one restricted interactions to only those that occur between nodes in the gene list. Analysis of these networks can, for example, identify subnetworks that are enriched in (but do not exclusively consist of) differentially expressed genes, or identify non-differentially expressed nodes that are topologically important in the network, both of which would otherwise not be identified. Identifying hubs in these networks, however, is biased towards identifying nodes that are highly connected in general such as promiscuous, ubiquitous or well-studied nodes, because nodes with many interactions in the query database have a higher probability of being included in the network by chance alone. Analysis of these degree-based hubs, for example identifying what biological processes or pathways these

nodes are enriched in, tells us little about the experimental context of interest and more about the properties of highly connected nodes in general. A more appropriate analysis is to determine which nodes interact with relevant nodes in the network (which we term contextual nodes) more than is statistically expected.

Here, we introduce the Contextual Hub Analysis Tool (CHAT), a Cytoscape App that identifies hub nodes that interact with more “contextual” nodes (e.g. differentially expressed genes or proteins) than statistically expected in networks integrated with user-supplied contextual data (e.g. gene expression data). We term these nodes *contextual hubs*. We show that such *contextual hubs* are considerably more relevant than degree-based hubs to the specific experimental context under investigation. As such, these nodes are promising candidates for further functional validation studies and potentially represent important points in the network for drug targeting.

## Methods

### Implementation

CHAT was written in Java 8 as an Open Services Gateway Initiative (OSGi) bundle for Cytoscape 3.0+<sup>18</sup>. It adds a “CHAT” option in the “Apps” menu that launches a popup window, which allows users to adjust different network initialization parameters. CHAT prompts users to input a list of gene identifiers (the supported ID types are dependent on the database selected by the user) and any associated contextual data, e.g. gene expression data associated with the genes. While the focus of this paper is on genes, CHAT can equally be applied to proteins. The OK button triggers Cytoscape's TaskManager to run a task that initiates the network construction and adds a tab to the results panel that provides functionality to further modify and analyze the network. To create the network, CHAT finds all the first neighbor interactors of the user-provided genes (or their encoded products). Interaction data is retrieved from one of the databases included in the PSICQUIC registry<sup>19</sup>, which the user can select. Note that interactions between the first neighbors are considered by CHAT but these are not included in the network visualization for clarity reasons. Once the network has been constructed, CHAT performs a hypergeometric test on each node in the network to identify nodes that interact with contextual nodes more than expected by chance. The probability that a given hub has  $k$  or more contextual interactors among its  $n$  interactors is given by the hypergeometric distribution:

$$p(X \geq k) = \sum_{x=k}^n \frac{\binom{K}{x} \binom{N-K}{n-x}}{\binom{N}{n}}$$

Where  $N$  is the number of genes with at least one interaction in the database queried and  $K$  is the number of contextually relevant nodes provided by the user (with at least one interaction in the database queried). Overrepresentation analysis heavily depends on the choice of background dataset for the determination of  $N$ . To

estimate the background frequency  $K/N$ , CHAT provides access to interaction data from databases available in the PSICQUIC registry. Databases with less than 10,000 interactions are excluded. The number of genes in the user-selected database that have at least one interaction (of the specified type) in which both interactors match the user-selected criteria for constructing the network (species, interaction type and ID type) determine the node population size  $N$ . Self-interactions are disregarded. Interactions between input genes and between their first neighbors are considered in the CHAT analysis. P-values calculated by CHAT are automatically corrected for multiple testing using the Benjamini-Hochberg procedure<sup>20</sup>, a method widely used in bioinformatics to avoid high false discovery rates. The Bonferroni approach is widely considered to be too strict<sup>21</sup>.

A right click on a node brings up an option to activate the “Node Analyzer” mode, which allows the user to analyze the connectivity pattern of individual hubs of interest. Using this function will display the node analyzer table on the results panel and all nodes except the selected node and its interactors will be hidden in the network visualization. The execution time of CHAT varies between a few seconds and a few minutes based on the number of user-supplied (contextual) genes, the size of the chosen database and its connection speed as well as the user-selected network layout. These factors also influence memory consumption.

### Operation

The identification of the top contextual hubs consists of three primary steps: 1) input of a user-supplied gene list and contextual data, 2) network construction and statistical analysis to identify nodes that preferentially interact with contextual nodes and 3) visualization of the top contextual hubs and their interactions and comparison to the top degree-based hubs. To construct a network using CHAT, the user must provide a list of gene identifiers and associated numerical or categorical attributes in the text box in tab-delimited format, or upload the data as a csv or tab-delimited file via the upload button (Figure 1) (.csv or .txt file types). The user can then specify which genes in the uploaded list are contextually important based on the user-provided contextual data (e.g. genes with > 2 fold-change in expression). The user then selects one of the databases in the PSICQUIC registry to query, and specifies the relevant species, ID type and interaction type for the query. The user can then choose to visualize the network using any of the layout algorithms available in Cytoscape. Clicking the OK button creates the network and a new tab in the results panel, which allows the user to visualize the network and to analyze the results further (Figure 2). The results panel is split into several parts. In the first part, the parameters used to generate the network (database, species, id type and interaction type(s)) are displayed. The second panel allows the user to compare the top contextual hubs and the top degree-based hubs at the click of a button. By default, node size and node color are proportional to the node’s corrected p-value calculated by CHAT, such that the smaller the p-value (i.e. more statistically significant), the larger the node size and the darker the red coloring of the node. The user can customize the color scheme, however. In contrast, if the users selects “Show degree hubs”, the

visualization changes and the node size and coloring will now be proportional to each node’s degree in the selected database. By default, CHAT displays the top 20 contextual hubs but the user can adjust this by using the slider provided. To investigate a single node in detail the user can employ CHAT’s “Node Analyzer” by right clicking on a node. This will limit the network view to show only the selected node and its interactors and will display a table at the bottom of the results panel tab with information on the node’s name, p-value and its interactors.

### Use case

**Dataset 1. Use case data: 462 genes that have been reported to be up-regulated during Dengue fever infection**

<http://dx.doi.org/10.5256/f1000research.9118.d128126>

As a demonstration of its potential utility and as validation, CHAT was used to construct a network using a dataset of 462 genes that have been reported to be up-regulated during Dengue fever, a mosquito-borne viral infection<sup>22</sup> (Ensembl gene IDs for these 462 genes are provided in Dataset 1). These 462 genes represent the contextual data for this case study. CHAT was used to construct a network of these genes and their first neighbor interactors using interaction data that was sourced from InnateDB<sup>23,24</sup> via the PSICQUIC web service (InnateDB-All). A network of 4,910 nodes was generated. CHAT was then used to identify the top 20 conventional hub nodes (based solely on degree) and the top 20 contextual hub nodes in the network (Figure 3). No nodes were in common in the two top 20 lists. InnateDB pathway analysis<sup>23,24</sup> revealed that the top 20 degree-based hubs were enriched in pathways related to the cell cycle and cancer (Supplementary Table 1), which is likely due to the fact that proteins involved in these processes tend to be highly connected in general. In comparison to degree-based hubs, the top 20 contextual hubs were statistically enriched in pathways related to the immune response to viral infection, such as the interferon signaling pathway; the Retinoic acid inducible gene-I (RIG-I) pathway; the Toll-like receptor (TLR) pathway; and the Janus kinase (JAK) - Signal Transducer and Activator of Transcription (STAT) pathway (Supplementary Table 2). All of these pathways have been shown to play key roles in the host response to Dengue infection<sup>25,26</sup>. Indeed, many of the top 20 contextual hubs (but not degree-based hubs) were well-known transcription factors involved in the host interferon response including STAT1, STAT2 and the interferon regulatory factors (IRFs); IRF1, 3, 8 and 9, which is a key cellular response to viral infection including Dengue<sup>27,28</sup>. Another gene identified in the contextual hub analysis but not the degree-based analysis was interferon-stimulated gene 15 (ISG15). Cells in which ISG15 has been silenced have been shown to have significantly higher Dengue viral loads<sup>29</sup>. The results of the pathway analysis were reinforced by a Gene Ontology analysis using innatedb.com<sup>23,24</sup>, which identified terms including cytokine-mediated signaling pathway, type I interferon signaling pathway, and innate immune response among the top 10 enriched terms (FDR < 0.05) for the contextual hubs but not the degree-based hubs (Supplementary Table 3 and Supplementary Table 4).

CHAT (Contextual Hub Analysis Tool) Setup

CHAT enables users to construct and visualize a network of interactions, integrate contextual information such as gene expression data, and identify nodes that are more highly connected to these contextual nodes than expected by chance.

Identifier and attributes

Select an ID type: **ensembl** (checked)  create network from only contextual genes

uniprotkb	
ensembl	
hgnc	
refseq	
innatedb	
mgc	
ENSG00000093009	2.00999999
ENSG00000112343	2.00999999
ENSG00000072121	2.00999999
ENSG00000164713	2
ENSG00000103966	2

Upload (two columns, no header) Demo (N=462)

Contextually important if:

Database settings

Active databases:  Interaction type (none=select all):

BAR	acetylation reaction
BioGrid	adp ribosylation reaction
DIP	association
EBI-GOA-nonIntAct	cleavage reaction
I2D	colocalization
InnateDB	covalent binding
<b>InnateDB-All</b>	deacetylation reaction
IntAct	deneddylation reaction

NCBI Taxonomy ID:

Network characteristics

Network name:

Layout:

If your network has more than 1500 nodes or 9000 edges, a grid layout will be applied to reduce computational time. You can always change it later.

Cancel OK

**Figure 1. CHAT network analysis.** To construct a network using CHAT, the user provides a list of gene identifiers and associated numerical or categorical attributes relevant in the context of interest.

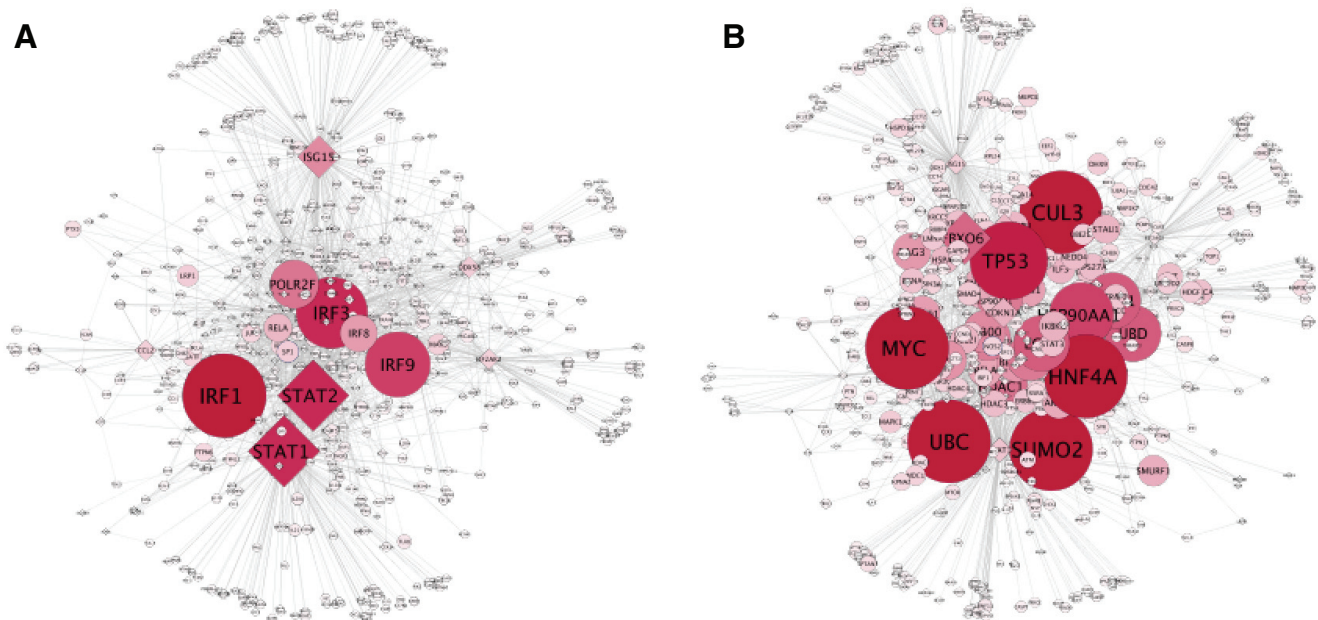
The screenshot shows the 'Results Panel' for 'CHAT Network'. It contains several sections for user configuration:

- User-supplied settings:** Database: InnateDB-All, Species: 9606, ID type: ensembl. Interaction type(s): acetylation reaction, adp ribosylation reaction.
- Compare contextual with degree-based hubs:** Buttons for 'Show contextual hubs' (active) and 'Show degree hubs'. A checkbox for 'Show contextually important nodes' is unchecked.
- Visualize the top n hubs:** A slider set to 20, with a text input field containing '20'. Below it, a dropdown menu for 'Re-layout network using:' is set to 'Edge-weighted Spring Embedded Layout'.
- Change the colour schema:** A table of color swatches:
 

Hub colour gradient - max	Red
Hub colour gradient - min	White
Contextual node colour	Green
Non-contextual node col...	Grey
- Node Analyzer:** A section for analyzing a specific node. It shows:
  - Node: IRF1
  - P-value: 1.1863153881799472E-11
  - First neighbors:
 Below this is a table of neighbors:
 

Name	Context Attribute	Contextual
DDX58		3.74

**Figure 2. Network visualization.** CHAT provides a number of options to customize the network visualization.



**Figure 3. Visualization of a Dengue gene expression dataset.** A CHAT network visualization comparing contextual hubs (A) to degree-based hubs (B) in a network constructed using InnateDB<sup>23,24</sup>.

## Conclusion

Through the integration of contextual information, such as gene or protein expression, contextual hub analysis as implemented in CHAT can identify context-specific hubs more relevant to the biological context under study, such as disease, treatment or cellular state. As shown in the above case study, these hubs are of more functional relevance than genes found through analysis based on degree only. Given the current emphasis on the importance of considering the network model of biological pathways and the ever-increasing abundance of high-throughput data, CHAT provides a valuable addition to the biologists' computational toolkit in using a network-based approach to help prioritize genes of interest for further investigation or drug discovery. In the future, CHAT can be extended to include the contextual analysis of other network features such as network bottlenecks.

## Data availability

*F1000Research*: Dataset 1. Use case data: 462 genes that have been reported to be up-regulated during Dengue fever infection, [10.5256/f1000research.9118.d128126](https://doi.org/10.5256/f1000research.9118.d128126)<sup>30</sup>

## Software availability

Software available from: <http://apps.cytoscape.org/apps/chat>

Latest source code: <https://bitbucket.org/dynetteam/chat>

Archived source code at time of publication: <http://www.doi.org/10.5281/zenodo.56496><sup>31</sup>

Manual/Tutorial: <https://bitbucket.org/dynetteam/chat/downloads>

License: [Lesser GNU Public License 3.0](https://www.gnu.org/licenses/lesser.html)

## Author contributions

TM and IHG jointly developed the App under the supervision of KB and DJL. TM and DJL wrote the paper with contributions from the other authors. HLW developed an earlier unpublished Python version of CHAT that inspired the development of this App. MBL provided computational and systems support for the project. DJL conceived of the idea, supervised the App's development and co-wrote the paper.

## Competing interests

No competing interests were disclosed.

## Grant information

The research leading to these results received funding from the European Union Seventh Framework Programme (FP7/2007-2013) PRIMES project under grant agreement number FP7-HEALTH-2011-278568. The Lynn Group is also supported by EMBL Australia.

*The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

## Supplementary material

Supplementary Table 1. Pathway analysis of the top 20 degree-based hubs.

Supplementary Table 2. Pathway analysis of the top 20 contextual hubs.

Supplementary Table 3. Gene ontology terms overrepresented among the top 20 contextual hubs in the Dengue fever network.

Supplementary Table 4. Gene ontology terms overrepresented among the top 20 degree-based hubs in the Dengue fever network.

## References

- Barabasi AL, Gulbahce N, Loscalzo J: **Network medicine: a network-based approach to human disease.** *Nat Rev Genet.* 2011; **12**(1): 56–68.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Barabasi AL, Oltvai ZN: **Network biology: understanding the cell's functional organization.** *Nat Rev Genet.* 2004; **5**(2): 101–113.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Jeong H, Mason SP, Barabási AL, et al.: **Lethality and centrality in protein networks.** *Nature.* 2001; **411**(6833): 41–42.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Dyer MD, Murali TM, Sobral BW: **The landscape of human proteins interacting with viruses and other pathogens.** *PLoS Pathog.* 2008; **4**(2): e32.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Borneman AR, Leigh-Bell JA, Yu H, et al.: **Target hub proteins serve as master regulators of development in yeast.** *Genes Dev.* 2006; **20**(4): 435–448.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Przytycka TM, Singh M, Slonim DK: **Toward the dynamic interactome: It's about time.** *Brief Bioinform.* 2010; **11**(1): 15–29.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Rachlin J, Cohen DD, Cantor C, et al.: **Biological context networks: a mosaic view of the interactome.** *Mol Syst Biol.* 2006; **2**: 66.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Agarwal S, Deane CM, Porter MA, et al.: **Revisiting date and party hubs: Novel approaches to role assignment in protein interaction networks.** *PLoS Comput Biol.* 2010; **6**(6): e1000817.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Gao S, Wang X: **Identification of highly synchronized subnetworks from gene expression data.** *BMC Bioinformatics.* 2013; **14**(Suppl 9): S5.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Zinman GE, Naiman S, O'Dee DM, et al.: **ModuleBlast: identifying activated sub-networks within and across species.** *Nucleic Acids Res.* 2015; **43**(3): e20.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Soul J, Hardingham TE, Boot-Handford RP, et al.: **PhenomeExpress: a refined network analysis of expression datasets by inclusion of known disease phenotypes.** *Sci Rep.* 2015; **5**: 8117.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Chin CH, Chen SH, Wu HH, et al.: **cytoHubba: identifying hub objects and sub-networks from complex interactome.** *BMC Syst Biol.* 2014; **8**(Suppl 4): S11.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hernandez-Toro J, Prieto C, De Las Rivas J: **APID2NET: Unified interactome graphic analyzer.** *Bioinformatics.* 2007; **23**(18): 2495–2497.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Chuang HY, Lee E, Liu YT, et al.: **Network-based classification of breast cancer metastasis.** *Mol Syst Biol.* 2007; **3**: 140.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Assenov Y, Ramirez F, Schelhorn SE, et al.: **Computing topological parameters of biological networks.** *Bioinformatics.* 2008; **24**(2): 282–284.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Doncheva NT, Assenov Y, Domingues FS, et al.: **Topological analysis and interactive visualization of biological networks and protein structures.** *Nat Protoc.* 2012; **7**(4): 670–85.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Scardoni G, Tosadori G, Faizan M, et al.: **Biological network analysis with CentiScaPe: centralities and experimental dataset integration [version 2; referees: 2 approved].** *F1000Research.* 2014; **3**: 139.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Shannon P, Markiel A, Ozier O, et al.: **Cytoscape: A software Environment for integrated models of biomolecular interaction networks.** *Genome Res.* 2003; **13**(11): 2498–2504.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Aranda B, Blankenburg H, Kerrien S, et al.: **PSICQUIC and PSIScore: accessing and scoring molecular interactions.** *Nat Methods.* 2011; **8**(7): 528–529.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical and powerful approach to multiple testing.** *J R Stat Soc B.* 1995; **57**(1): 289–300.  
[Reference Source](#)
- Noble WS: **How does multiple testing correction work?** *Nat Biotechnol.* 2009; **27**(12): 1135–1137.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hoang LT, Lynn DJ, Henn M, et al.: **The early whole-blood transcriptional signature of dengue virus and features associated with progression to dengue shock syndrome in Vietnamese children and young adults.** *J Virol.* 2010; **84**(24): 12982–94.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Breuer K, Foroushani AK, Laird MR, et al.: **InnoDB: systems biology of innate immunity and beyond—recent updates and continuing curation.** *Nucleic Acids Res.* 2013; **41**(Database issue): D1228–1233.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- del-Toro N, Dumousseau M, Orchard S, et al.: **A new reference implementation of the PSICQUIC web service.** *Nucleic Acids Res.* 2013; **41**(Web Server issue): W601–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nasirudeen AM, Wong HH, Thien P, et al.: **RIG-I, MDA5 and TLR3 synergistically play an important role in restriction of dengue virus infection.** *PLoS Negl Trop Dis.* 2011; **5**(1): e926.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Souza-Neto JA, Sim S, Dimopoulos G: **An evolutionary conserved function of the JAK-STAT pathway in anti-dengue defense.** *Proc Natl Acad Sci U S A.* 2009; **106**(42): 17841–6.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- De La Cruz Hernández SI, Puerta-Guardo H, Flores-Aguilar H, et al.: **A strong interferon response correlates with a milder dengue clinical condition.** *J Clin Virol.* 2014; **60**(3): 196–199.  
[PubMed Abstract](#) | [Publisher Full Text](#)
- Morrison J, García-Sastre A: **STAT2 signaling and dengue virus infection.** *JAKSTAT.* 2014; **3**(1): e27715.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Dai J, Pan W, Wang P: **ISG15 facilitates cellular antiviral response to dengue and west nile virus infection in vitro.** *Virology.* 2011; **418**: 468.  
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Muetze T, Goenawan IH, Wiencko HL, et al.: **Dataset 1 in: Contextual Hub Analysis Tool (CHAT): A Cytoscape app for identifying contextually relevant hubs in biological networks.** *F1000Research.* 2016.  
[Data Source](#)
- Muetze T, Goenawan IH, Wiencko HL, et al.: **Contextual Hub Analysis Tool (CHAT): A Cytoscape app for identifying contextually relevant hubs in biological networks.** *Zenodo.* 2016.  
[Data Source](#)



# Open Peer Review

Current Referee Status:  

Version 2

Referee Report 01 November 2016

doi:10.5256/f1000research.10240.r16856

 **Christopher K. Tuggle, Haibo Liu**

Bioinformatics and Computational Biology Program, Iowa State University, Ames, IA, USA

This Cytoscape app “CHAT” is valuable given its improvement over conventional biological network analysis methods by considering the context of network analysis. It is a good addition to the Cytoscape toolkits. However, we suggest some modifications that might make the paper more readable and the method more applicable.

Suggested Modifications to text

1. In the “Operation” section, “The identification of the top contextual hubs consists of three primary steps: 1) input of a user-supplied gene list and contextual data, 2) network construction and statistical analysis to identify nodes that preferentially interact with contextual nodes and 3) visualization of the top contextual hubs and their interactions and comparison to the top degree-based hubs.” should include four steps to be consistent with the following statements: 1) input a gene list, **2) database selection**, 3) network construction and statistical analysis, 4) visualization of the top conceptual hubs and their interactions and comparison to the top degree-based hubs.
2. The “implementation” and “Operation” sections in the Methods are somewhat redundant. For example, “The OK button triggers Cytoscape’s TaskManager to run a task that initiates the network construction and adds a tab to the results panel that provides functionality to further modify and analyze the network” and “A right click on a node brings up an option to activate the “Node Analyzer” mode”. It will be more appropriate if both these sentences are moved to the “Operation” section. The “Implementation” section should focus on the description of functionalities provide by CHAT and the application interfaces (APIs) are implemented, while how to conduct network analysis should be in the “Operation” part.

Suggested Modifications to future versions of CHAT

1. One question is the background selection criteria. As the authors mentioned, this step is very important and directly affects the later statistical test results. An alternative/better background might be the set of genes/proteins expressed under the condition used to create the original dataset, such as genes expressed in given the cell type, tissue, or treatment. This might eliminate the irrelevant nodes and reduce the number of tests needed.

2. In the current CHAT version, only the first-order neighbor-interactors are allowed to be considered which is generally most important and might be enough in most situations. But when the resulting network is small, the user might not be able to perform further analysis. So if the option of higher-order interactors is provided, the tool will be more versatile.
3. The authors mentioned that only databases in the PSICQUIC registry with at least 10,000 interactions are included. This is kind of arbitrary. While it is appreciated that other DB available in Cytoscape can be used with this tool, the usefulness of the databases should be determined case by case. So we suggest the author provide the user all the available choices of interaction databases.
4. We suggest a different method might be more appropriate for multiple testing correction. The authors state that “a method widely used in bioinformatics to avoid high false discovery rates. The Bonferroni approach is widely considered to be too strict.” This could be reworded as “a method widely used in bioinformatics to avoid high false discovery rates, instead of the Bonferroni approach which is widely considered to be too strict.”  
However, the author assumed the genes in the network are independent by using the “BH” method. To be more realistic, we suggest the “BY” method for correction of the multiple testing. See [Benjamini and Yekutieli](#): The control of the false discovery rate in multiple testing under dependency *Ann. Statist.* 2001; **29** (1165-1188)

**We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

---

### Version 1

Referee Report 09 August 2016

doi:[10.5256/f1000research.9812.r15065](https://doi.org/10.5256/f1000research.9812.r15065)



**Sandra Orchard, Pablo Porras Millán**

Wellcome Trust Genome Campus, European Molecular Biology Laboratory-European Bioinformatics Institute, Hinxton, UK

This is a well written technical paper, clearly outlining a new Cytoscape App in terms that would make it easy for a new user, with some familiarity with Cytoscape, to download, install and use. The ability to generate contextual hubs is currently not possible with existing Cytoscape Apps, so this is a valuable addition to the collection. A couple of queries and some minor points for correction:

1. The application searches for first-neighbour interactions of molecules in the list presented to it. It did not appear to search for interaction between members of the list, which should not affect the contextual nodes selection much, but will alter the degree-based hubs. This should be commented on, or the documentation made clearer if we are incorrect with this observation. To bypass this problem and make the user more aware of this limitation, the tool should be able to provide more control over how the network is constructed, for example providing the option to exclude first neighbours.

2. Can this application be made to work with an existing network?
3. The text is entirely gene-centric and may leave an inexperienced user under the impression it is only usable for gene-expression data whereas it is equally useful for the analysis of proteomic data and works with UniProtKB identifiers. Whilst I realise this is apparent to anyone who downloads the app, it may well be worth adding a sentence to both the Summary or Introduction of this paper, and also the description in the App store just to make this very clear to naive users.
4. It may also be worth adding the reference to the 2013 PSICQUIC paper as well as the original as I personally find it more informative and again, may be helpful to the inexperienced user.

## References

1. del-Toro N, Dumousseau M, Orchard S, Jimenez RC, Galeota E, Launay G, Goll J, Breuer K, Ono K, Salwinski L, Hermjakob H: A new reference implementation of the PSICQUIC web service. *Nucleic Acids Res.* 2013; **41** (Web Server issue): W601-6 [PubMed Abstract](#) | [Publisher Full Text](#)

**We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.**

**Competing Interests:** No competing interests were disclosed.

Author Response 22 Aug 2016

**Tanja Muetze**, SAHMRI, Australia

Thank you very much for your thoughtful review. Below we have addressed each of the points raised.

***The application searches for first-neighbour interactions of molecules in the list presented to it. It did not appear to search for interaction between members of the list, which should not affect the contextual nodes selection much, but will alter the degree-based hubs. This should be commented on, or the documentation made clearer if we are incorrect with this observation. To bypass this problem and make the user more aware of this limitation, the tool should be able to provide more control over how the network is constructed, for example providing the option to exclude first neighbours.***

This observation is incorrect. CHAT does include interactions between members of the uploaded list. Perhaps what you mean is that interactions between the first neighbors of the uploaded list of genes/proteins are not considered? These are actually also considered in CHAT's calculations – they are just excluded from the visualization to avoid unhelpful "hairball" visualizations. We have now clarified this point in the paper. We are not sure why a user would want to exclude first neighbors – this information is needed in CHAT to identify the contextual hubs.

***Can this application be made to work with an existing network?***

We agree that this would be a very nice feature, unfortunately it is actually very difficult to do with the current design of CHAT. CHAT identifies hub nodes that interact with more "contextual" nodes than statistically expected using a hypergeometric test. This test is reliant on calculating,  $N$ , the number of genes with at least one interaction in the database queried to estimate the background

expectation. This parameter would not be known or easily estimated in a user-supplied network as CHAT wouldn't know what database or the data in the database when the network was constructed by the user. One of the nice features of CHAT is that the network is constructed using the latest data available via a PSICQUIC query.

***The text is entirely gene-centric and may leave an inexperienced user under the impression it is only usable for gene-expression data whereas it is equally useful for the analysis of proteomic data and works with UniProtKB identifiers. Whilst I realise this is apparent to anyone who downloads the app, it may well be worth adding a sentence to both the Summary or Introduction of this paper, and also the description in the App store just to make this very clear to naive users.***

Good point. We have now edited the text to clarify that protein as well as gene ids can be used to construct the network in CHAT.

***It may also be worth adding the reference to the 2013 PSICQUIC paper as well as the original as I personally find it more informative and again, may be helpful to the inexperienced user.***

We have now added the suggested reference to the paper.

Again, we thank you for comments and suggestions.

***Competing Interests:*** No competing interests were disclosed.

---