CoExp Web, a web tool for the exploitation of co-expression networks

Sonia García[1,2,3], Ana L. Gil[1], Alejandro Cisterna[4], Federico Jurado[4], Regina Reynolds[1,2,3], NABEC (North America Brain Expression Consortium), Mark R. Cookson[5], John Hardy[1,2,3], Mina Ryten[1,2,3] and Juan A. Botía[3,4]

Affiliations:

1 Institute of Neurology, University College London (UCL), London, UK.

2 NIHR Great Ormond Street Hospital Biomedical Research Centre, University College London, London, UK.

3 Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London WC1E 6BT, UK.

4 Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, Spain

5 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA.

Summary

Gene co-expression networks [1,2] are a powerful type of analysis to construct genome wide gene groupings based on expression profiling. Co-expression networks allow discovery of groups of genes whose mRNA levels are highly correlated. Subsequent annotation of those modules often reveals biological functions and specificity to the cell types implicated in the tissue being studied. There are multiple approaches to perform such analyses with WGCNA amongst one of the most widely used R packages. While investigating a few network models can be done manually, it is often more advantageous to study a wider set of network models derived from the ever increasing quantities of transcriptomic data being generated (e.g. multiple networks built from many transcriptomic sources). However, there is no software tool available to make such a task easy. Furthermore, the visual nature of co-expression networks in combination with the coding skills that exploring networks requires, makes the construction of a web-based platform for their management highly desirable. Here, we present CoExp Web page, a user-friendly online tool that allows the exploitation of the full collection of 109 co-expression networks provided by the CoExpNets R package.

Availability and Implementation:

The back and front ends code of CoExp Web application is fully available for download on GitHub at https://github.com/SoniaRuiz/CoExp_Web.

The docker images of CoExp Webpage are available for download on Docker Hub at https://hub.docker.com/r/soniaruiz/coexp.

Contact: *[full E-mail address to be supplied, preferably an institution address.]*

Supplementary information: *[in case we have any supplementary data, here we will include links to additional figures/data available on a web site, or reference to online-only Supplementary data available at the journal's web site]*

Introduction

Gene co-expression network analysis has been widely used to identify biologically important patterns in gene expression in a hypothesis free and genome-wide manner[3–6]. The driving principle behind co-expression network analysis is that genes whose expression levels are highly correlated, are also likely to share functional and biological relationships[7,8]. Thus, co-expression networks are models of how genes cluster together into modules of highly co-expressed genes, generally by using graph-based approaches to reveal their similarities[2].

Currently, WGCNA is amongst the most widely used R package to build co-expression networks and therefore CoExpNets has been based on WGCNA with a refinement of the gene modules using a k-means based approach[2]. These packages generate networks in the form of text files. As a result, to use, share and exploit network models with proficiency is a manual task which is prone to error, and requires coding skills to be performed effectively. Furthermore, R´s command-line environment reduces its usability in a world where the web-page format has become the most well-known and accepted way of browsing information. However, most importantly the underlying graph-based model in which most co-expression networks are based makes it most natural to explore this type of data in a visual and interactive manner.

In order to address these issues, we propose CoExp Webpage: a web-page application to increase the usability and accessibility of co-expression network data. We illustrate the use of CoExp Webpage through the release of 109 different co-expression networks offered by the CoExpNets R package. We extend the functionality provided by the CoExpNets R package, through CoExp Webpage's 'Plot Network' option, which generates a directed graph to visualise the most important genes from a preferred module. Therefore, CoExp is a convenient way to deploy, share and use any co-expression models suite.

Description

CoExp Webpage consists of three separate tabs, corresponding to the three different ways of exploiting networks models: network catalogue browsing, network-based annotation of gene sets, and network module visualization through active graphs.

Additionally, all three tabs support the exploitation of the same collection of networks. This collection consists of 109 different co-expression networks that belong to four different network groups: (1) the Religious Orders Study and Memory and Aging Project **(**ROSMAP) [9–11] composed of four co-expression networks derived from post-mortem human frontal cortex originating from control individuals, as well as those with cognitive impairment and Alzheimer's disease; (2) The Genotype-Tissue Expression project (GTEx) V6 and V7 [12] composed of two suites of co-expression networks on 47 post-mortem control human tissue samples; (3) UK Brain Expression Consortium (UKBEC) [4,13] composed of 10 microarray-based gene expression profiling networks derived from post-mortem control human brain tissue; (4) North America Brain Expression Consortium (NABEC), composed of one gene co-expression network derived from post-mortem control human frontal cortex.

Data preparation

Only the 'Gene Set Annotation' tab requires data preparation. The user is required to provide a list of genes of interest in a comma-separated or blank-space-separated fashion. That gene list should

be pasted within the 'Genes' text area placed in the left-hand side of the page. When the formatting of the gene list is incorrect, a user-friendly prompt message will inform the user.

Network Catalogue Browser

The user can become familiar with the co-expression models available upon by navigating through the catalogue. With this in mind, the first tab available on the upper menu corresponds to the 'Network Catalogue' tab through which the user can inspect and download the whole network catalogue to obtain information about any network or any module within a network. To browse the catalogue, the first step consists of selecting a network category in the menu placed in the left-hand side of the webpage. The second step is the selection of a co-expression network of interest within that category. Finally, the user can select one of two different views: the 'Ontology Classification' or the 'Cell Type Classification'. The 'Ontology Classification' view returns a data table in which each module from the selected network occupies one row. The columns provide summarized information about annotation terms enriched for the genes in the modules. The p-value column shows the enrichment obtained from gProfileR [14], which incorporates data from well recognised ontologies, including Gene Ontology, REACTOME and KEGG. The 'Cell Type Classification' view, returns a data table in which the rows correspond now to the sets of gene markers of brain cell types tested for enrichment (Fisher´s Exact test) and each module occupies a column. Each cell within the table contains the Bonferroni corrected p-value for the enrichment of a set of cell type markers within a module. In all cases, the raw data that supports the generation of the data table can be downloaded under the three buttons placed on the upper-left-side corner of the table.

Gene Set Annotation

Fo-expression models are often used to annotate a gene set of interest, in the context of a specific condition (e.g. a tissue of interest). 'Gene Set Annotation' is the second tab within the main menu. Using this function the user can investigate whether his/her own gene set of interest is enriched within a single or multiple co-expression modules across all the co-expression networks from amongst those available in the catalogue. In this way, those genes can be annotated based on how they spread across the network modules and their biological context explored within just a couple of clicks. If a gene of interest has not been found in any module a pop-up view will inform the user. The user is supplied a results table in which each row relates to a gene of interest which has been successfully found in any of the modules belonging to the network or networks selected. The columns provide information on the module in which the gene has been found, including the statistical significance of the overlap between the input genes and the genes int the module and a brief description of the module's function based on the top 5 GO terms. All the outputs associated with this type of analysis is available for download using the three buttons placed on the upper-left-corner of the table.

Plot Network

Once the user has decided which network is of interest, and which module within the network requires of a detailed visualization, the genes can be plotted. The third tab, 'Plot Network', enables the graph-based visualisation of the genes within a network module of interest, whether identified by browsing through the catalogue or because the user's gene set of interest significantly clusters within a module. The 'Plot Network' tab generates an interactive directed graph formed by the hub genes within a module. The user can select how many of the most relevant genes will appear at the plot. The resulting plot is interactive in the sense that it can be zoomed, rotated and the direct

neighbours of any gene highlighted by just clicking on the gene of interest. Both the raw data and a high quality PNG image of the graph are available for download.

Conclusions

CoExp Webpage is a web platform that enables the exploitation of co-expression networks. CoExp currently offers 109 co-expression models focused on brain transcriptomics with plans to expand its scope. It is a powerful, easy to use and innovative tool for gene set annotation across a variety of brain specific transcriptomicdata sets. CoExp makes co-expression models visually manageable, accessible and easily exploitable to the scientific community. Everything is shareable in CoExp, from the network models themselves to the software used to build front and back-ends of CoExp so any research laboratory can construct their own CoExp web site. This makes it a powerful tool for the wider research community interested in producing, using or sharing co-expression models to support their research.

Funding

This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging.

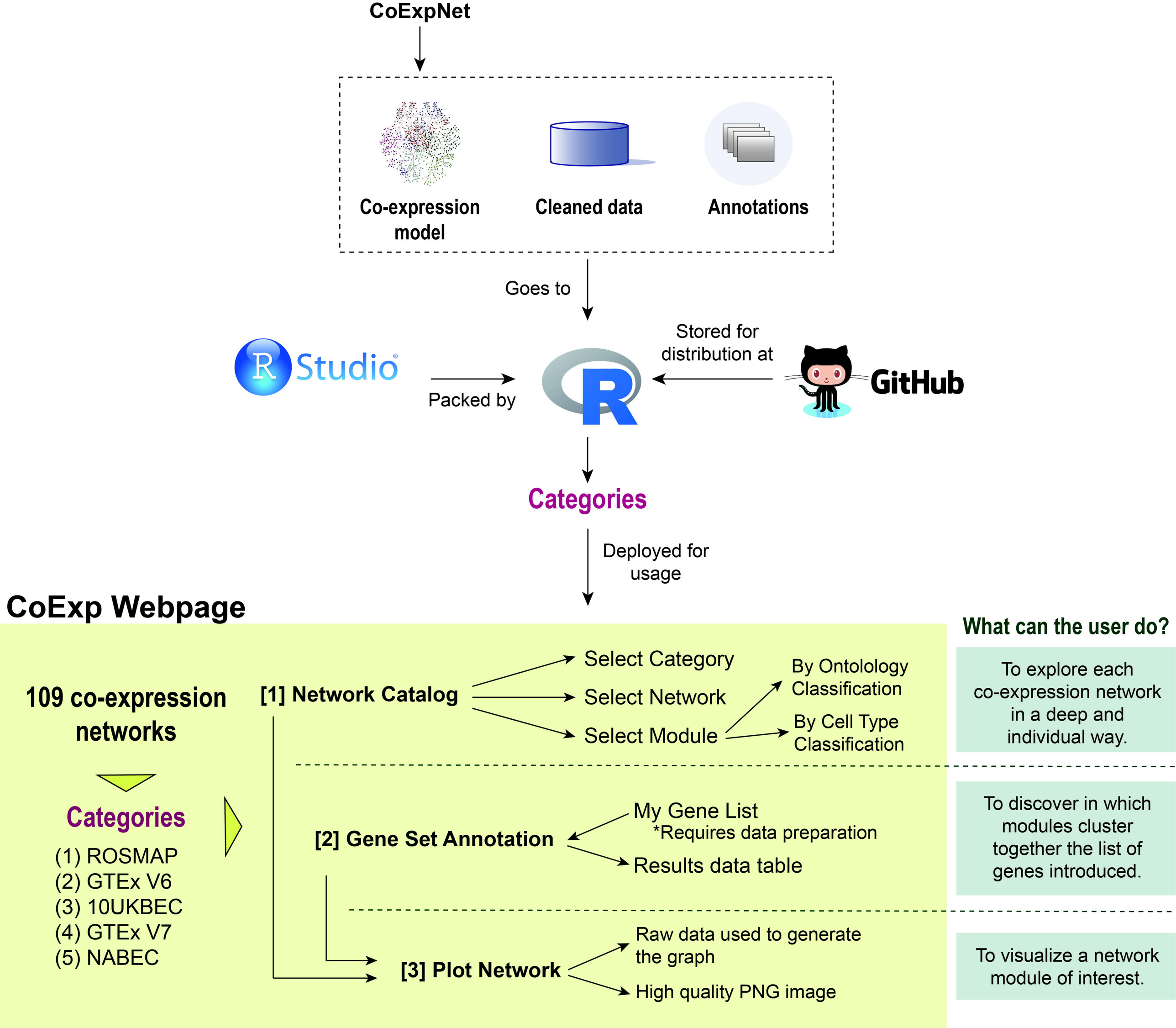
R.H.R. was supported through the award of a Leonard Wolfson Doctoral Training Fellowship in Neurodegeneration. J.H. and M.R. were supported by the UK Medical Research Council (MRC), with J.H. supported by a grant (MR/N026004/) and M.R. through the award of a Tenure Track Clinician Scientist Fellowship (MR/N008324/1). J.H. was also supported by the UK Dementia Research Institute, The Wellcome Trust (202903/Z/16/Z), the Dolby Family Fund, and the NIHR.

Conflict of Interest: none declared.

References

1. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
2. An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Systems Biology* (2017) doi:https://doi.org/10.1186/s12918-017-0420-6.
3. de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289-304.e18 (2018).
4. Forabosco, P. *et al.* Insights into TREM2 biology by network analysis of human brain gene expression data. *Neurobiol. Aging* **34**, 2699–2714 (2013).

5. Bettencourt, C. *et al.* White matter DNA methylation profiling reveals deregulation of HIP1, LMAN2, MOBP, and other loci in multiple system atrophy. *Acta Neuropathol. (Berl.)* (2019) doi:10.1007/s00401-019-02074-0.
6. UK Brain Expression Consortium (UKBEC) *et al.* Frontotemporal dementia: insights into the biological underpinnings of disease through gene co-expression network analysis. *Mol. Neurodegener.* **11**, (2016).
7. Xuelian Ma. Co-expression Gene Network Analysis and Functional Module Identification in Bamboo Growth and Development. *frontiers in Genetics* (2018) doi:https://doi.org/10.3389/fgene.2018.00574.
8. Mohammad Reza Bakhtiarizadeh. Weighted Gene Co-expression Network Analysis of Endometriosis and Identification of Functional Modules Associated With Its Main Hallmarks. *frontiers in Genetics* (2018) doi:https://doi.org/10.3389/fgene.2018.00453.
9. A. Bennett, D., A. Schneider, J., Arvanitakis, Z. & S. Wilson, R. Overview and Findings from the Religious Orders Study. *Curr. Alzheimer Res.* **9**, 628–645 (2012).
10. A. Bennett, D. *et al.* Overview and Findings from the Rush Memory and Aging Project. *Curr. Alzheimer Res.* **9**, 646–663 (2012).
11. De Jager, P. L. *et al.* A multi-omic atlas of the human frontal cortex for aging and Alzheimer's disease research. *Sci. Data* **5**, 180142 (2018).
12. The GTEx Consortium *et al.* The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
13. UK Brain Expression Consortium *et al.* Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat. Neurosci.* **17**, 1418–1428 (2014).
14. Reimand, J., Kull, M., Peterson, H., Hansen, J. & Vilo, J. g:Profiler—a web-based toolset for functional profiling of gene lists from large-scale experiments. *Nucleic Acids Res.* **35**, W193–W200 (2007).


CoExpNet
Co-expression model
Cleaned data
Annotations
Goes to
RStudio
Packed by
Stored for distribution at
GitHub
Categories
Deployed for usage
CoExp Webpage
109 co-expression networks
Categories
(1) ROSMAP
(2) GTEx V6
(3) 10UKBEC
(4) GTEx V7
(5) NABEC
[1] Network Catalog
Select Category
Select Network
Select Module
By Ontolology Classification
By Cell Type Classification
[2] Gene Set Annotation
My Gene List
*Requires data preparation
Results data table
[3] Plot Network
Raw data used to generate the graph
High quality PNG image
What can the user do?
To explore each co-expression network in a deep and individual way.
To discover in which modules cluster together the list of genes introduced.
To visualize a network module of interest.