



# WGCNA: an R package for weighted correlation network analysis

Peter Langfelder and Steve Horvath  
with help of many other contributors

Semel Institute for Neuroscience and Human Behavior, UC Los Angeles (PL),  
Dept. of Human Genetics and Dept. of Biostatistics, UC Los Angeles (SH)

Peter (dot) Langfelder (at) gmail (dot) com, SHorvath (at) mednet (dot) ucla (dot) edu

[BMC Bioinformatics, 2008 9:559](#)  (link opens in a new tab/window)

## Quick navigation

- [Article abstract](#)
- [Introduction to WGCNA](#)
- [Tutorials](#)
- [Further reading](#)
- [Automatic installation from CRAN](#)
- [Manual download and installation](#)
- [FAQ: Problems installing or using the package](#)
- [Old versions of the R package](#)
- [Citing the WGCNA package](#)
- [Acknowledgments](#)

## Abstract

Correlation networks are increasingly being used in bioinformatics applications. For example, weighted gene co-expression network analysis is a systems biology method for describing the correlation patterns among genes across microarray samples. Weighted correlation network analysis (WGCNA) can be used for finding clusters (modules) of highly correlated genes, for summarizing such clusters using the module eigengene or an intramodular hub gene, for relating modules to one another and to external sample traits (using eigengene network methodology), and for calculating module membership measures. Correlation networks facilitate network based gene screening methods that can be used to identify candidate biomarkers or therapeutic targets. These methods have been successfully applied in various biological contexts, e.g. cancer, mouse genetics, yeast genetics, and analysis of brain imaging data. While parts of the correlation network methodology have been described in separate publications, there is a need to provide a user-friendly, comprehensive, and consistent software implementation and an accompanying tutorial.

The WGCNA R software package is a comprehensive collection of R functions for performing various aspects of weighted correlation network analysis. The package includes functions for network construction, module detection, gene selection, calculations of topological properties, data simulation, visualization, and interfacing with external software. Along with the R package we also present R software tutorials. While the methods development was motivated by gene expression data, the underlying data mining approach can be applied to a variety of different settings.

## Getting started with R and Weighted Gene Co-expression Network Analysis

The package described here is an add-on for the [statistical language and environment R](#) (free software). Our [tutorials](#) contain step by step instructions such that even complete novice users should be able to get started in R immediately.

Readers wishing to learn about the theory and published applications of WGCNA are invited to visit the [WGCNA main page](#).

## R Tutorials

A [comprehensive set of tutorials](#) that illustrate various aspects of WGCNA is available. We offer not only introductory tutorials that introduce basic functionality of the package, but also more advanced analyses in which we used the WGCNA package in our own research.

[Click here](#) to access the tutorial page.

## Further reading

Peter Langfelder occasionally writes about [WGCNA features and other topics](#) relating to data analysis. The articles are written for a general audience and try to avoid deep technical details. We also have a few [technical reports](#) that discuss a selected deeply technical aspects of the WGCNA methodology - these are more mathematical and targeted primarily to die-hard statistician geeks.

## Automatic installation from CRAN

The WGCNA package is now available from the Comprehensive R Archive Network (CRAN), the standard repository for R add-on packages. Currently, some of the required packages is only available from [Bioconductor](#) and need to be installed using Bioconductor's installation tools. The easiest way to do this is

```
install.packages("BiocManager")
BiocManager::install("WGCNA")
```

The first command (`install.packages("BiocManager")`) can be skipped if the package BiocManager is already installed.

This will install the WGCNA package and all necessary dependencies. The catch is that this only installs the newest version of WGCNA if your R version is also the newest (minor) version. Users using older versions of R will need to follow the manual download and installation instructions below.

**Note for Mac users:** CRAN occasionally fails to compile the WGCNA package for Mac OS X. This leads to the error message "Package WGCNA is not available..." when calling `BiocManager::install()`. If this occurs, please download the binary version from here and follow the installation instructions (or, if you are able to compile packages locally, download the source and install that).

**Note of caution:** The newest versions of WGCNA is available from CRAN only for the current R version and (usually) one older version. For example, if your R version is 3.2.1 and the current R version on CRAN is 3.5.0, the automatic installation and update will not use the newest version of WGCNA. Please update your R to the newest version or use the manual download below.

**Problems installing or using the package?** Please see our [list of frequently asked questions](#). Your problem and the solution may already be posted there.

## Manual download and installation

Please follow these steps only if the automatic package installation above does not work.

### Prerequisites:

The current version of the WGCNA package **will only work with R version 3.0.0 and higher**. If you have an older version of R, please upgrade your R.

The WGCNA package requires the following packages to be installed: stats, grDevices, utils, matrixStats (0.8.1 or higher), Hmisc, splines, foreach, doParallel, fastcluster, dynamicTreeCut, survival, parallel, preprocessCore, GO.db, impute, and AnnotationDbi. If your system does not have them installed, the easiest way to install them is to issue the following command at the R prompt:

```
install.packages(c("matrixStats", "Hmisc", "splines", "foreach", "doParallel", "fastcluster",
"dynamicTreeCut", "survival"))
source("http://bioconductor.org/biocLite.R")
biocLite(c("GO.db", "preprocessCore", "impute"))
```

Please note that GO enrichment calculations in WGCNA are deprecated; we recommend using the R package [anRichment](#) which provides replacement for WGCNA functions `GOenrichmentAnalysis()` and `userListEnrichment()`.

If you run an older version of R, the above may not install the newest version of the `dynamicTreeCut` package. Should you encounter this problem, please manually download and install `dynamicTreeCut` from [this web page](#).

**R package download and installation:** Package `WGCNA_1.64-1` (last updated 2018/02/27) is available here as source code and several pre-compiled versions for various platforms. In general it is preferable to download the source and compile the package locally; however, if this is not practical, please select an appropriate compiled version.

- Source for **Linux** and all users able to compile the package locally: [WGCNA\\_1.64-1.tar.gz](#)
- Compiled binary Mac OS X: [WGCNA\\_1.64-1.tgz](#)
- Compiled binary for **Windows** running **R-3.4.0** or higher: [WGCNA\\_1.64-1.zip](#)
- [Reference manual](#) in pdf format
- Quick reference: [overview table of most important functions](#)
- A terse [changelog](#)

If you require a compiled version, please make sure you select the correct version. We are unable to provide compiled binaries for other versions of R and/or operating systems; please upgrade your R if you are running an old version not listed here.

The package version numbers follow the format `packageName_major.minor-revision`. Minor versions typically add or change some functionality; revisions typically contain bugfixes or minor enhancements.

Should you discover bugs (of which there are most likely plenty), please report them to Peter Langfelder.

## Problems installing or using the package

Please see our [list of Frequently Asked Questions](#) (and frequently given answers); the solution to your problem may lie there. In particular, you can find answers about spurious Mac errors, compatibility problems when upgrading WGCNA, and others. If you still cannot solve the problem, email Peter Langfelder.

## Old versions of R package WGCNA

Older version of the packages presented on this page are available [here](#).


## Citing the WGCNA package

If you use WGCNA in published work, please cite it to properly credit people who have created it.

The WGCNA as an analysis method is described in

- Zhang B and Horvath S (2005) A General Framework for Weighted Gene Co-Expression Network Analysis, *Statistical Applications in Genetics and Molecular Biology*: Vol. 4: No. 1, Article 17 PMID: 16646834

The package implementation is described in the article

- Langfelder P, Horvath S (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* 2008, 9:559 ([link to paper](#)) 

If you use any q-value (FDR) calculations, please also cite at least one of the following articles:

- Storey JD. (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society, Series B*, 64: 479-498.
- Storey JD and Tibshirani R. (2003) Statistical significance for genome-wide experiments. *Proceedings of the National Academy of Sciences*, 100: 9440-9445.
- Storey JD, Taylor JE, and Siegmund D. (2004) Strong control, conservative point estimation, and simultaneous conservative consistency of false discovery rates: A unified approach. *Journal of the Royal Statistical Society, Series B*, 66: 187-205.

If you use the `collapseRows` function to summarize/convert probe-level data to gene-level data, please cite

- Miller JA, Cai C, Langfelder P, Geschwind DH, Kurian SM, Salomon DR, Horvath S (2011) Strategies for aggregating gene expression data: The `collapseRows` R function. *BMC Bioinformatics* 12:322.

If you use module preservation calculations, please cite

- Langfelder P, Luo R, Oldham MC, Horvath S (2011) Is my network module preserved and reproducible? PloS Comp Biol. 7(1): e1001057

If you use functions `rgcolors.func`, `plotCor`, `plotMat`, `stat.bwss`, or `stat.diag.da`, please also cite the article

- Sandrine Dudoit, Yee Hwa Yang, Matthew J. Callow, and Terence P. Speed, Statistical methods for identifying differentially expressed genes in replicated cDNA microarray experiments, STATISTICA SINICA 2002, 12:111

## Acknowledgments

The core of the functions and other code was written by Peter Langfelder and Steve Horvath, partly based on older code written by Steve Horvath and Bin Zhang. Multiple people contributed additional code, most prominently Jeremy Miller, Chaochao (Ricky) Cai, Lin Song, Jun Dong, and Andy Yip. The package also contains code adapted from external packages that were either orphaned (such as package `sma`) or their development has made the code difficult to use in WGCNA (such as package `qvalue`). A big thanks goes out to people who continue report the many bugs in the package.

The package is currently maintained by Peter Langfelder.