

Rcircle: an R package for Integrating and Visualizing multiple “–omics” data for Knowledge Discovery

**Xing Li^{1*}, Almudena Martinez-Fernandez², Terry Therneau¹, Jean-Pierre Kocher¹,
Timothy J. Nelson²**

1. Division of Biomedical Statistics and Informatics, Department of Health Sciences Research, 200 First Street, S.W., Rochester, MN 55906, USA.
2. Department of General Medicine, Center of Regenerative Medicine, Transplant Center, 200 First Street, S.W., Rochester, MN 55906, USA.

*Contact author: Li.Xing@mayo.edu

Keywords: R package, Rcircle, visualization, genome, transcriptome

Biomedical science has entered the big data era and biologists have access to an overwhelming abundance of data due to the rapid advance of high-throughput technology in sequencing and microarray[1]. The tremendous volume and high dimensions pose an unprecedented challenge on data visualization and integration for efficient data exploration and effective scientific communication. Herein, we developed an R package, **Rcircle**, based on **grid** package to integrate and visualize interactome, time-course transcriptome, disease information, disease-affected pathways or networks to facilitate knowledge discovery[2]. Starting with a curated list of congenital heart disease (CHD) genes, we identified their top 10 partners for each CHD gene and built a network for both disease genes and their partners. Pathway analysis is performed on the entire gene list. The R package visualized the gene network in the inner circle with line width representing the interaction confidence. Hub genes in the network were represented by the size of bubbles circling the genes. Transcription profile, disease information, and pathways are shown in the outer layers linked directly to the genes in the network. By integrating different types of information together, we discovered disease hub genes and established critical pathways turned on at different stages of cardiogenesis. Furthermore, the **Rcircle** package is able to reveal vital genes in charge of the crosstalk among those disease associated pathways. The case application of **Rcircle** package in CHD analysis indicates that our strategy goes beyond visualization of information to highlight the pattern, prioritize vital candidates, and facilitate scientific discovery.

References

- [1]. Wong B. Points of Views: Visualizing Biological Data. Nat Methods 9, 1131 (2102)
- [2]. R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.