

SOFTWARE

Open Access

RCircos: an R package for Circos 2D track plots

Hongen Zhang, Paul Meltzer and Sean Davis*

Abstract

Background: Circos is a Perl language based software package for visualizing similarities and differences of genome structure and positional relationships between genomic intervals. Running Circos requires extra data processing procedures to prepare plot data files and configure files from datasets, which limits its capability of integrating directly with other software tools such as R. Recently published R Bioconductor package *ggbio* provides a function to display genomic data in circular layout based on multiple other packages, which increases its complexity of usage and decreased the flexibility in integrating with other R pipelines.

Results: We implemented an R package, *RCircos*, using only R packages that come with R base installation. The package supports Circos 2D data track plots such as scatter, line, histogram, heatmap, tile, connectors, links, and text labels. Each plot is implemented with a specific function and input data for all functions are data frames which can be objects read from text files or generated with other R pipelines.

Conclusion: *RCircos* package provides a simple and flexible way to make Circos 2D track plots with R and could be easily integrated into other R data processing and graphic manipulation pipelines for presenting large-scale multi-sample genomic research data. It can also serve as a base tool to generate complex Circos images.

Keywords: Software, *RCircos*, R package, Circos, Genomic data visualization

Background

Circos is a Perl language based software package for visualizing similarities and differences of genome structure and positional relationships between genomic intervals [1]. Although many tools for genomic data visualization have been developed [2-5], Circos is commonly used by the genome research community to present large-scale multi-sample genomic research data (http://circos.ca/in_literature/scientific/). While Circos is powerful and flexible in displaying genomic data it requires extra data procedures to prepare plot data files and configuration files from datasets, which limits its capability of integrating directly with other software tools such as R, one of the most commonly used toolsets in processing and statistical analysis of genomic data.

Recently, Yin et al. [6] published a Bioconductor package, *ggbio*, that includes a function to display genomic data in a circular layout and covers many of the basic Circos-like plots. The *ggbio* package relies on multiple other packages and offers some integration with other

Bioconductor packages. However, *ggbio* is somewhat complex (but powerful) and relies on high-level plotting packages. *RCircos* was developed as a simple and flexible approach to Circos-like plots that uses base R graphics.

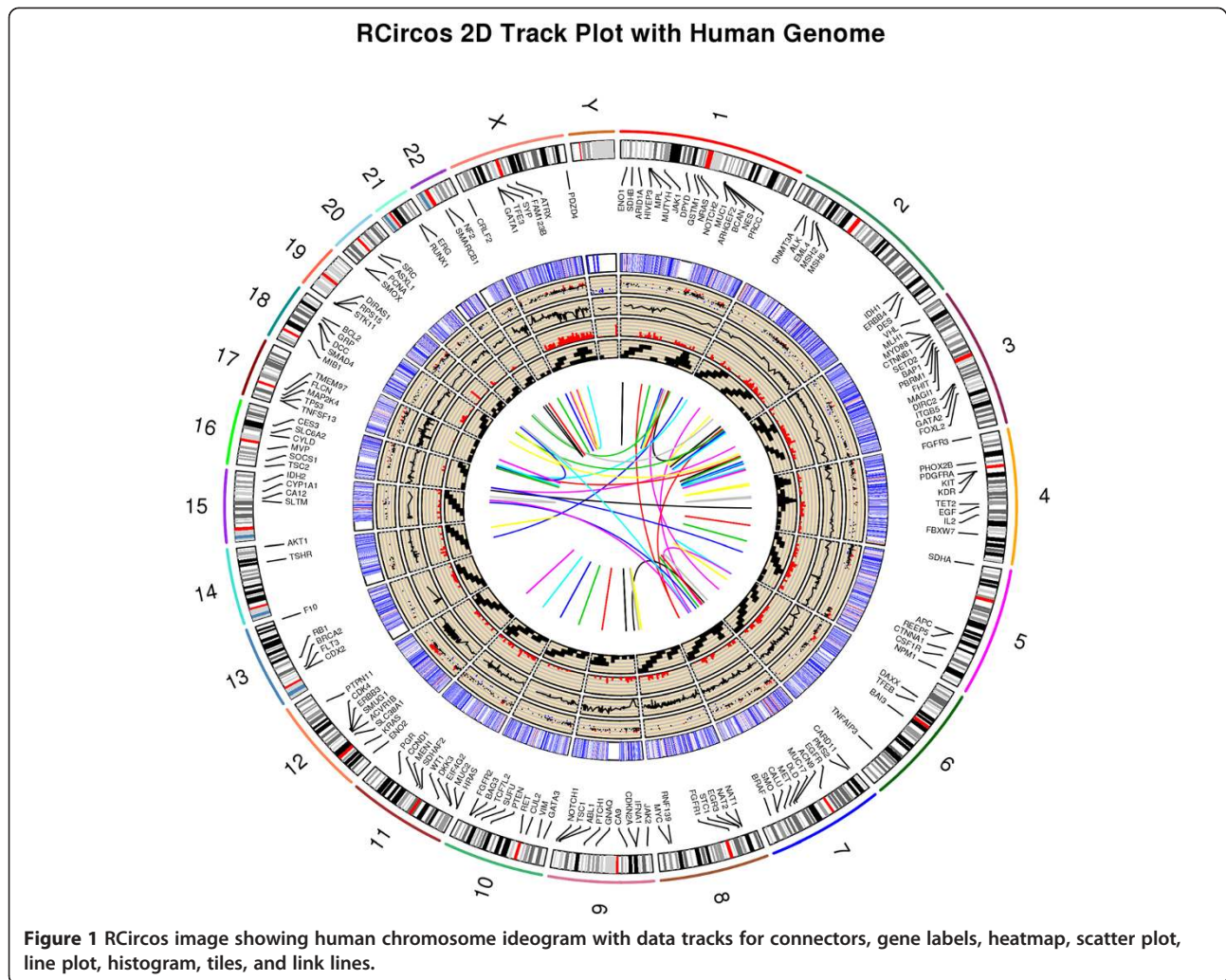
To make Circos 2D track plots simple and flexible, we implemented an R package, *RCircos*, that relies on base graphics and R data structures. With *RCircos*, Circos 2D track plots could be easily generated and the procedures can be effectively integrated with other R pipelines including graphics output manipulation.

Implementation

Packages used to build *RCircos* are all included in the R base installation (<http://www.r-project.org/>). Graphics functionality is accomplished using base R graphics. No other package is required unless input data is associated with special data structure such as *GenomicRanges* objects and need to be processed separately.

To reduce the complexity of the usage, all functions in *RCircos* use a simple data frame as input. The first three columns of the data frame are genomic position data in the order of chromosome name, start position, and end position followed by one or more data columns except of link data which requires paired chromosome positions

* Correspondence: sdavis2@mail.nih.gov
Genetics Branch, Center for Cancer Research, National Cancer Institute,
National Institutes of Health, Building 37, Room 6138, 37 Convent Drive,
Bethesda, MD 20892-4265, USA



for each row. Data set in data frame is directly passed to the plot function without need of further processing. Sample data are included in the package to show the input data formats and can be easily explored with data (package = "RCircos") function.

We follow the layout paradigm set forth by Circos and arrange data plots by tracks. The core track is the chromosome ideogram track with highlighting and labels. Data plot tracks can be placed inside or outside of chromosome ideogram track. A set of parameters is used to control the plot pattern such as chromosome width, number of base pairs per chromosome unit, track height, and point type. These parameters are initialized prior to plotting but can be customized to meet the requirements of different plot types.

RCircos is designed such that each type of Circos 2D track plot is drawn with a separate and dedicated function call. To make RCircos more flexible in integrating with other R pipelines, we chose low level plot functions of R including points(), lines(), polygon(), and text() to implement graphic plot functions of RCircos. All RCircos

plots work on an existing plot facilitating plot customization using standard R plot functionality.

Result and discussion

RCircos implements most of Circos 2D track plots including scatter, line, histogram, heatmaps, tiles, connectors, and text labels. We use the chromosome ideogram tables from UCSC genome browser to generate chromosome ideogram images and currently human, mouse, and rat are available in RCircos, but other species can be supported if relevant ideogram table is provided in a same format as cytoBandIdeo table in the UCSC genome browser [7].

A set of demos and a complete vignette are included in the package to show the RCircos plot procedures for each Circos 2D track plot type. Figure 1 was generated using the code below with build-in datasets and default parameters showing the human chromosome ideogram track along with data tracks for connectors, gene labels, heatmap, scatter plot, line plot, histogram, tiles, and link lines.

```
# Initialize core components used for calculation of positions of all plot
items. All chromosomes
# are included for plot and total of 10 tracks inside of chromosome ideogram
will be held for
# data plots (except of link lines). There is no data track outside of
chromosome ideogram.
data(UCSC.HG19.Human.CytoBandIdeogram)
cyto.info <- UCSC.HG19.Human.CytoBandIdeogram
RCircos.Set.Core.Components(cyto.info, chr.exclude=NULL, tracks.inside=10,
tracks.outside=0)

# Initialize graphic device (here a pdf file instead of R GUI)
pdf(file="RCircos.Demo.Human.pdf", height=8, width=8)

# Setup size of plot area with default values
RCircos.Set.Plot.Area()

# Plot chromosome ideogram
RCircos.Chromosome.Ideogram.Plot()

# Plot connectors and gene names on the first and second track inside of
chromosome ideogram.
# Gene names are held in the fourth column of input data.
data(RCircos.Gene.Label.Data)
RCircos.Gene.Connector.Plot(RCircos.Gene.Label.Data, track.num=1, side="in")
RCircos.Gene.Name.Plot(RCircos.Gene.Label.Data, name.col=4, track.num=2,
side="in")

# Heatmap plot on the fifth track inside of chromosome ideogram with data
values in the sixth
# column of input data.
data(RCircos.Heatmap.Data)
RCircos.Heatmap.Plot(RCircos.Heatmap.Data, data.col=6, track.num=5,
side="in")

# Scatterplot on the sixth track inside of chromosome ideogram with data
values in the fifth
# column of input data. Point colors will be red for any data values which
are >= 1 and blue for
# values <=-1.
data(RCircos.Scatter.Data)
RCircos.Scatter.Plot(RCircos.Scatter.Data, data.col=5, track.num=6, "in",
fold=1)

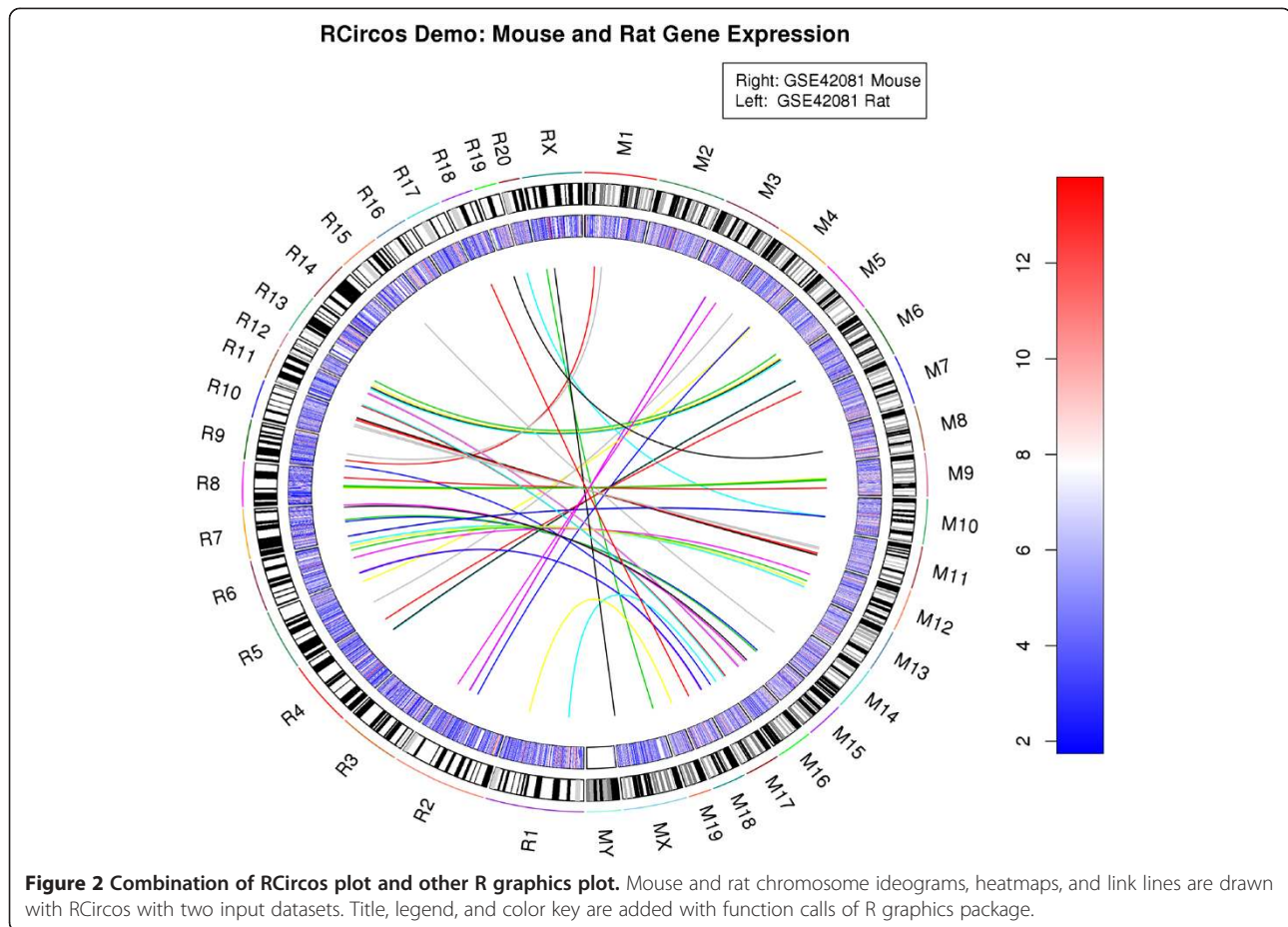
# Line plot on the seventh track inside of chromosome ideogram with data
values in the fifth
# column of input data.
data(RCircos.Line.Data)
RCircos.Line.Plot(RCircos.Line.Data, data.col=5, track.num=7, side="in")

# Histogram plot on the eight track inside of chromosome ideogram with data
values in the
# fourth column of input data.
data(RCircos.Histogram.Data)
RCircos.Histogram.Plot(RCircos.Histogram.Data, data.col=4, track.num=8,
side="in")

# Tile plot on the ninth track inside of chromosome ideogram.
data(RCircos.Tile.Data)
RCircos.Tile.Plot(RCircos.Tile.Data, track.num=9, side="in")

# Link line plot. Line colors are randomly assigned.
data(RCircos.Link.Data)
RCircos.Link.Plot(RCircos.Link.Data, track.num=11, by.chromosome=FALSE)

# close the graphic device
dev.off()
```



Since we implemented the RCircos plots with base R graphics, combining RCircos with other R plot functions is straightforward. Figure 2 show a heatmap generated with demo (“R.Circos.Demo.Mouse.And.Rat”) with blue and red colors for comparison of gene expression between mouse and rat (GEO data accession number: GSE42081) and link lines between top 50 highly expressed genes in mouse and the same genes in rat. Legend and color key for the heatmap were added with the legend() and image() functions.

Conclusions

The RCircos package provides simple and flexible functionality to generate Circos 2D track plots with R and can be easily integrated into other R data processing and graphic manipulation pipelines to present large-scale multi-sample genomic research data.

Availability and requirements

The package and source code of RCircos are available for download and install from CRAN website (<http://www.r-project.org>) with the license of GPL (> = 2).

Competing interests

The authors declare that they have no competing interests.

Author's contributions

HZ designed and implemented the software package, and wrote manuscript. SD participated in the software design and drafted the manuscript. PM revised the manuscript critically. All authors read and approved the final manuscript.

Acknowledgements

This study was supported by NCI intramural research Funding in Center for Cancer Research, National Cancer Institute, National Institutes of Health, USA. The package was developed and tested with the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (<http://biowulf.nih.gov>).

Received: 18 March 2013 Accepted: 7 August 2013

Published: 10 August 2013

References

1. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.** *Genome Res* 2009, **19**:1639–1645.
2. Nielsen CB, Cantor M, Dubchak I, Gordon D, Wang T: **Visualizing genomes: techniques and challenges.** *Nat Methods* 2010, **7**(3 Suppl):S5–S15.
3. Schroeder MP, Gonzalez-Perez A, Lopez-Bigas N: **Visualizing multidimensional cancer genomics data.** *Genome Med* 2013, **5**(1):9.

4. Miller CA, Anthony J, Meyer MM, Marth G: **Scribl: An HTML5 Canvas-based graphics library for visualizing genomic data over the web.** *Bioinformatics* 2013, **29**:381–383.
5. Zhu Z, Niu B, Chen J, Wu S, Sun S, Li W: **MGViewer: A desktop visualization tool for analysis of metagenomics alignment data.** *Bioinformatics* 2013, **29**:122–123.
6. Yin T, Cook D, Lawrence M: **ggbio: an R package for extending the grammar of graphics for genomic data.** *Genome Biol* 2012, **13**:R77.
7. Kuhn RM, Haussler D, Kent WJ: **The UCSC genome browser and associated tools.** *Brief Bioinform* 2012. doi:10.1093/bib/bbs038.

doi:10.1186/1471-2105-14-244

Cite this article as: Zhang et al.: RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics* 2013 **14**:244.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

