

Genome analysis

J-Circos: an interactive Circos plotter

Jiyuan An^{1,*}, John Lai¹, Atul Sajjanhar², Jyotsna Batra¹,
Chenwei Wang¹ and Colleen C Nelson¹

¹Australian Prostate Cancer Research Centre-Queensland, Institute of Health and Biomedical Innovation, Queensland University of Technology, Princess Alexandra Hospital and Translational Research Institute, Brisbane, QLD 4102, Australia and ²School of Information Technology, Deakin University, Burwood, VIC 3125, Australia

*To whom correspondence should be addressed.

Associate Editor: John Hancock

Received on September 23, 2014; revised on December 12, 2014; accepted on December 16, 2014

Abstract

Summary: Circos plots are graphical outputs that display three dimensional chromosomal interactions and fusion transcripts. However, the Circos plot tool is not an interactive visualization tool, but rather a figure generator. For example, it does not enable data to be added dynamically nor does it provide information for specific data points interactively. Recently, an R-based Circos tool (RCircos) has been developed to integrate Circos to R, but similarly, Rcircos can only be used to generate plots. Thus, we have developed a Circos plot tool (J-Circos) that is an interactive visualization tool that can plot Circos figures, as well as being able to dynamically add data to the figure, and providing information for specific data points using mouse hover display and zoom in/out functions. J-Circos uses the Java computer language to enable, it to be used on most operating systems (Windows, MacOS, Linux). Users can input data into J-Circos using flat data formats, as well as from the Graphical user interface (GUI). J-Circos will enable biologists to better study more complex chromosomal interactions and fusion transcripts that are otherwise difficult to visualize from next-generation sequencing data.

Availability and implementation: J-circos and its manual are freely available at <http://www.australianprostatecentre.org/research/software/jcircos>

Contact: j.an@qut.edu.au

Supplementary information: [Supplementary data](#) are available at *Bioinformatics* online.

1 Introduction

Circos plot (Krzywinski *et al.*, 2009) has been widely used in genomics research, especially in conjunction with next-generation sequencing to detect genomic rearrangements and gene fusions (Hagege *et al.*, 2007; Maher *et al.*, 2009). Circos is a plotting tool that was developed in the Perl language. However, there are limitations to Circos, some of which are inherent in Perl computer language, such as the complexity of installation, i.e. many Perl packages have to be installed. Furthermore, some packages are version sensitive, making it difficult even for bioinformaticians to install. Because Circos plot has the difficult task of representing the whole genome length as a circular diagram (e.g. ~3 billion

nucleotides in the human genome), there is no way to accurately determine the genes and/or loci that each line represents in the Circos plot. Thus, it is advantageous to have a Circos plot with mouse hover function that provides information for specific data points. The recently developed Circlize (Gu *et al.*, 2014), OmicCircos (Hu *et al.*, 2014), ggbio (Yin *et al.*, 2012) and RCircos (Zhang *et al.*, 2013) are R language versions of Circos. They are much easier to import R-generated data into Circos, but it has the same limitations, such as lack of interactive functionality.

In this article, we provide an interactive visualization Circos plot tool that only requires Java installation, which is typically installed on most computers. The input data are in tab-delimited format.



Fig. 1. (a) User interface of J-Circos. A circle chromosome band concatenated by hg19 chromosome sizes is the default plot. Users can change genome, circle size and its origin in the command line. There are also other data types (Circos-wiggle, bridge, label, text, cytogenetic and UCSC wiggle). Users can input data from files (click 'file') or copy/paste into the textbox. Click 'load file' or 'draw' to plot the inputted data. (b) Types of Circos-wiggle pictures. Four pictures in the left hand side are generated by setting the data column 'filled' to 'false'. The circle picture was generated setting the wiggle range from 0 to whole size. Arcs are generated by setting identical 'height1' and 'height2' values. (c) Circos bridge is drawn as a parabola. c is the constant in the quadratic equation. w is the distance between two sides of the wiggle curve.

Users can either import data through text files or copy/paste in the interface from MS Excel sheets.

2 Features

J-Circos is mainly used for genome visualization. Multiple concentric circles can be used to show different types of genomic information, such as expression level (wiggle plots), and gene fusions and chromosome conformation capture data (bridge plots) (Fig. 1a). J-Circos uses widely accepted UCSC genome browser data formats, such as bigWig, bigBed, bedgraph and Bed format. J-Circos also uses a new data type called Circos wiggle that is an extended wiggle data type that can construct more complex visual objects. J-Circos also uses another type called Circos bridge for visualizing genomic translocation and gene fusions. J-Circos also has interactive functions such as zoom in/out and mouse hover display, and circos tracks can be loaded dynamically.

3 Data structure

The outermost circle in J-Circos represents the chromosomes (Fig. 1a), which are concatenated according to the chromosomal sizes listed in the chromosome size file. All chromosome loci in the data file are converted into an angle degree. Data are represented in J-Circos as outlined in Figure 1a. The data format for J-Circos are arranged in columns for (i) color, which is entered as three integer numbers separated by commas, (RGB), (ii) chromosome, which should be consistent with the chromosome size file, (iii) start

nucleotide position, (iv) length, which corresponds to the thickness of UCSC-bigWig/bedgraph/Circos-wiggle/Circos-bridge lines and (v) description, which is the information for mouse hover display. Further details on the data format can be found in Supplementary file S1 and file S2.

The seven major features of J-Circos plot are:

- (1) Circos-wiggle plots: This is typically used to show expression levels at genomic loci. In J-Circos, this is shown as a box located on a base line that is defined by height1 and height2 as shown in Figure 1a. Of note, the height value of wiggle plots can be either positive or negative values, and these are displayed above or below the base line, respectively. Furthermore, if the wiggle height1 value is zero, then the wiggle plot is drawn on the base line. Another important column is 'filled', which provides an option to draw an empty wiggle. Many different types of shapes for wiggle tracks as shown in Figure 1b.
- (2) Circos-bridge plots: As mentioned, bridge tracks are used to display connections of two chromosomal loci (e.g. from fusion genes or chromosome conformation capture data) in J-Circos. J-Circos uses the following parabolic curve formula to draw these bridges:

$$y = ax^2 + c,$$
 c is the distance from the top of the parabolic curve to the line that connects the two roots of the inner or outer curve. We set c to two-thirds of the distance from the origin to the line that connects the two roots of the innermost or outermost curve, so that we can induce $a = -c/w^2$, where w is half of the width of the two roots of the innermost or outermost curve. This is illustrated in Figure 1c.
- (3) Label: Labels are used to annotate specific data points in J-Circos. Labels are displayed perpendicular to a baseline as shown in Figure 1a.
- (4) Text: Text is used to annotate J-Circos plots. The text feature is distinct from the 'label' feature as the 'text' feature is intended for purposes such as titles for circos-wiggle/UCSC-wiggle or Circos-bridge tracks. The two types of orientation for text plotting are (a) tangent or (b) perpendicular (Fig. 1a), and it is determined by column 'type' (Supplementary file S1).
- (5) Cytogenetic: This feature shows G-band gene density. J-Circos uses the data format from the UCSC genome browser (Meyer et al., 2013). J-Circos requires a value.txt and cytogenetic file to display cytogenetic bands. The value.txt file contains the g-band value information, and the cytogenetic file contains the chromosome ranges corresponding to giesma staining.
- (6) USCS bigWig/bedgraph: The data format is simple: chromosome, nucleotide start position, nucleotide end position and value integer corresponding to expression level. This expression value can be positive or negative, corresponding to either the plus or minus strand, respectively. Of note, bigWig is uploaded more efficiently.
- (7) UCSC bigBed/Bed: bigBed/Bed files have three columns (chromosome name, start, end) and nine optional columns. The exons will be displayed along the circle. Because Circos plots typically represent a massive scale, the whole transcript is consequently displayed as a line.

4 Summary

In this article, we introduce J-Circos, which not only covers most functions of the original Circos tool, but also provides an interactive function which provides further detail for specific data points using

the zoom function and mouse hovering over these data points. However, unlike the original Circos that needs installation of other packages, J-Circos does not require installation of any third party package. Furthermore, J-Circos has a very user-friendly interface and can produce complex figures with a few clicks. Thus, we propose that J-Circos is an easy-to-use tool that is also readily accessible for biologists with limited bioinformatics skills. All data are kept in flat text file and users can also copy/paste data from MS excel files directly into J-Circos.

Acknowledgements

This work is supported by the Australian Government Department of Health and a Queensland Government Smart Futures Premier's Fellowship.

Conflict of Interest: none declared.

References

- Gu,Z. *et al.* (2014) Circlize implements and enhances circular visualization in R. *Bioinformatics*, **30**, 2811-2812.
- Hagege,H. *et al.* (2007) Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc.*, **2**, 1722-1733.
- Hu,Y. *et al.* (2014) OmicCircos: a simple-to-use R package for the circular visualization of multidimensional omics data. *Cancer Inform.*, **13**, 13-20.
- Krzywinski,M. *et al.* (2009) Circos: an information aesthetic for comparative genomics. *Genome Res.*, **19**, 1639-1645.
- Maher,C.A. *et al.* (2009) Transcriptome sequencing to detect gene fusions in cancer. *Nature*, **458**, 97-101.
- Meyer,L.R. *et al.* (2013) The UCSC genome browser database: extensions and updates 2013. *Nucleic Acids Res.*, **41**, D64-69.
- Yin,T. *et al.* (2012) ggbio: an R package for extending the grammar of graphics for genomic data. *Genome Biol.*, **13**, R77.
- Zhang,H. *et al.* (2013) RCircos: an R package for Circos 2D track plots. *BMC Bioinformatics*, **14**, 244.