# J-Circos Manual

J-Circos: A Java Graphic User Interface for Circos Plot

Jiyuan An[1], John Lai[1], Atul Sajjanhar[2], Jyotsna Batra[1],Chenwei Wang[1] and Colleen C Nelson[1]

[1]Australian Prostate Cancer Research Centre-Queensland, Institute of Health and Biomedical Innovation, Queensland University of Technology and Translational Research Institute, Princess Alexandra Hospital, Level 1, Building 1, Ipswich Road, Brisbane, QLD 4102, Australia. [2]School of Information Technology, Deakin University, 221 Burwood Highway Burwood, VIC 3125, Australia,

j.an@qut.edu.au

Web page: http://www.australianprostatecentre.org/research/software/jcircos

# Table of Contents

# List of Figures

## 1.1 Java Requirement

J-Circos requires Java 7 or later. Instructions to download and install Java Platform (Java SE Development Kit) are available from

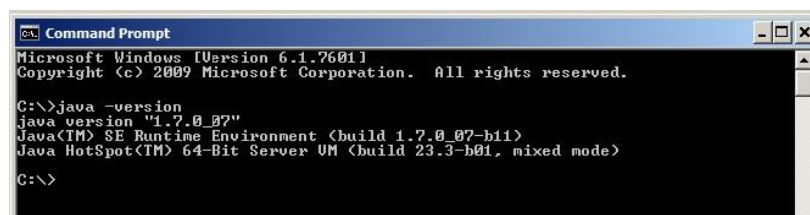http://www.oracle.com/technetwork/java/javase/downloads/index.html

## 1.2 Preparation

1.  Dowload J-Circos source files as a zip archive from:

    http://www.australianprostatecentre.org/research/software/jcircos

2.  Save the zip file into a folder – e.g. C:\J-Circos (Windows), or /home/xxx/J-Circos (Linux, MacOS).

3.  Once the files are extracted from the zip archive, a new folder "J-Circos" will appear. The folder contains four files:

    *   **J-Circos.bat**, is a batch file, which runs in the Windows environment. A shortcut can be created on the desktop to allow J-Circos to run by simply clicking the shortcut.
    *   **J-Circos.sh**, is a bash file for linux and MacOS.
    *   **J-Circos.jar**, stores Java classes of J-Circos. This is the main file to run this tool.

    And two subdirectories:

    *   **Lib folder**, contains all Java libraries, used by J-Circos.
    *   **Data.** Under the data folder you will find:
        *   hg18.chrom.sizes and hg19.chrom.sizes list the chromosome sizes for initial cycle band. It will be indicated in the command line.
        *   cytogenetic_hg18 and cytogenetic_hg19 are folders where two cytogentic files and value.txt are located. Corresponding "cytogentic" file needs to be indicated when plotting the cytogentic.
        *   Demo data files: demo_Circos_bridge.txt, demo_Circos_wiggle.txt and etc.

## 1.3 Start J-Circos

1.  Ensure that Java is executable using the approach shown in Figure 1. If Java fails to execute, edit the PATH system environment variable.



**Figure 1 Test whether Java is executable**

2.  In Windows, double click on **_J-Circos.bat_** to start J-Circos. You can also start J-Circos manually from the command line.

3.  In Linux or MacOS, type

- chmod +x J-Circos.sh
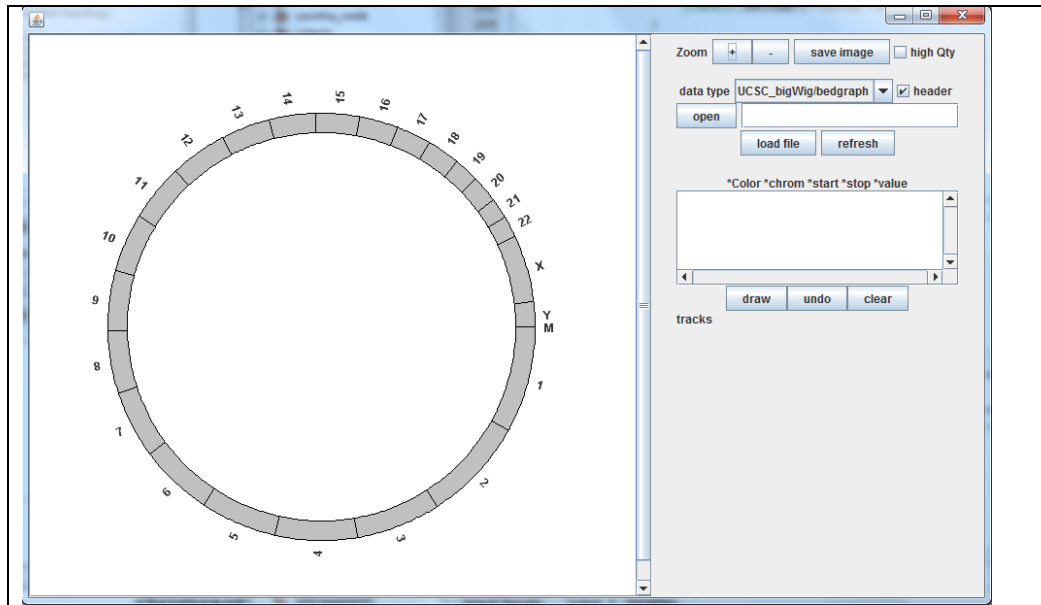
- *J-Circos.sh*.



**Figure 2 Initial interface of J-Circos.**

4. Parameters

   The parameters can be changed by editing the parameters in .bat or .sh files:

   - −r: Radius of Chromosome circos. The default is 200, which means the chromosome circos has Radius 200 (pixels).

   - −c: the origin of the circos. The default values are 100 (pixels),100(pixels).

   - The chromosome size file. The default file is "hg19.chrom.sizes". User can download "fetchChromSizes" from http://hgdownload.cse.ucsc.edu/admin/exe/.

## 1.4  Interface of J-Circos (**Figure 2**)

1. **Zoom function**: "+" and "-" are two buttons to zoom in and zoom out.

2. **Save image**: the image in the left canvas is saved into eps format. For high quality images, users can tick the "high Qty" box which is beside the "save image" box.

3. **Data type**: Before loading data files, users need to select data type from the drop-down menu "data type".

4. **Header**: Needs to be checked if the first line of the data file has a header row.

5. **File**: select data file to load. The selected filename will appear in the textbox beside the *File* button.

6. **Load file**: the file in the textbox above is loaded and the result will appear in the canvas.

## 1.5  Steps to load data

```
┌────────────────────────────────────────────────────────┐
│ Select data type                                       │
└────────────────────────────────────────────────────────┘
                          │
                          ▼
┌────────────────────────────────────────────────────────┐
│ Tick header if the first line of the data file is header│
└────────────────────────────────────────────────────────┘
                          │
                          ▼
┌────────────────────────────────────────────────────────┐
│ Select data file name                                  │
└────────────────────────────────────────────────────────┘
                          │
                          ▼
┌────────────────────────────────────────────────────────┐
│ Load file                                              │
└────────────────────────────────────────────────────────┘
```
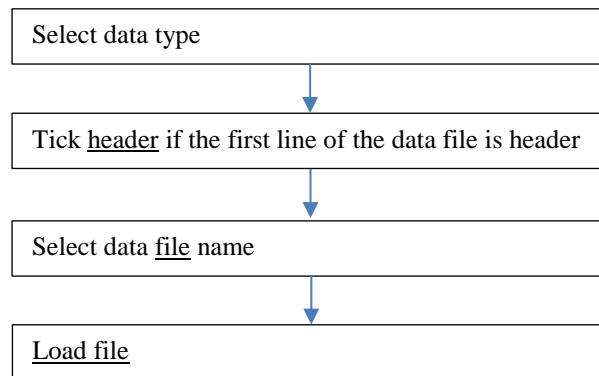
## 1.6  J-Circos tracks

### 1.6.1   Plot UCSC_wiggle/bedgraph

J-Circos supports two wiggle formats from the UCSC genome browser: bedgraph
(http://genome.ucsc.edu/goldenpath/help/bedgraph.html); and bigwig
(http://genome.ucsc.edu/goldenpath/help/bigWig.html).

For the bedgraph file, the lines starting with "#" are filtered out. When "load file" is clicked, a popup window queries the user to input the Radius and the wiggle color. Figure 3 shows the LNCaP RNA-seq data for plus (red) and minus (blue) DNA strands. Since RNA-seq data is usually big, this may take several seconds to appear. The mouse over display displays the genomic loci followed by the wiggle value, e.g. "10000-20000 val:-100" which means that the genomic loci of the wiggle box are from 10,000 to 20,000 and the wiggle value is -100.  The chromosome name is not included because it is apparent from the circos plot.
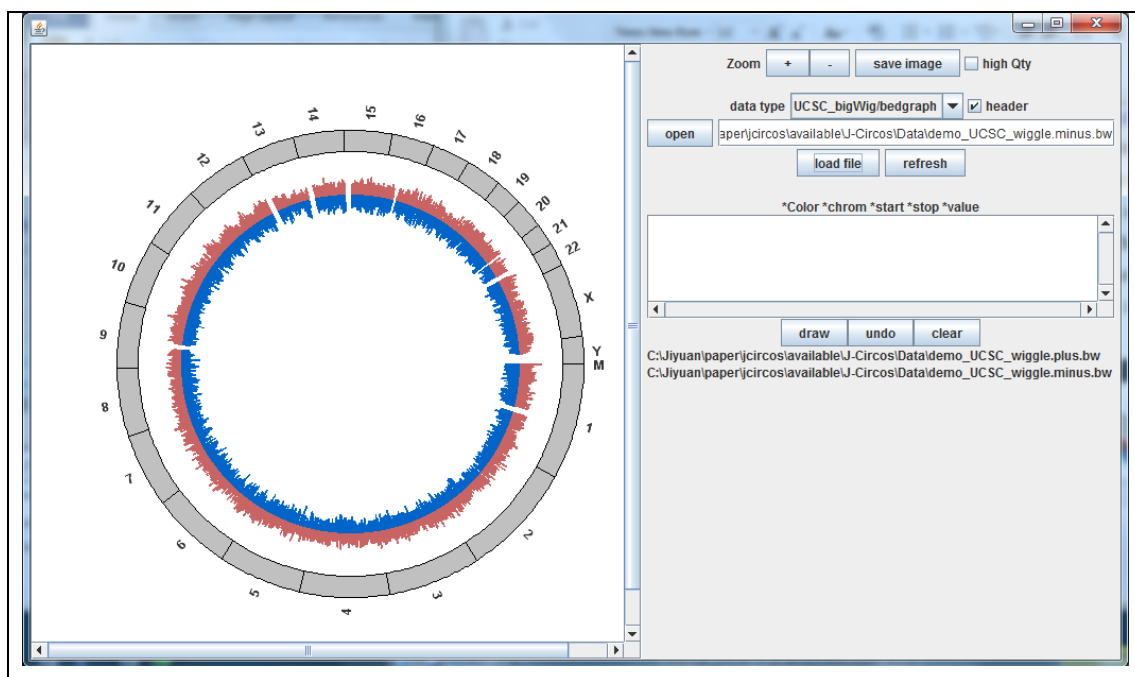


**Figure 3 RNA-seq expression shown in UCSC wiggle data format.**

## 1.6.2 Plot UCSC_BigBed/Bed

The UCSC-bed format was originally used to illustrate transcripts. All exons are represented by rectangles, and the rectangles for the 5' and 3' UTRs are thinner than coding exons. All exons are connected by a line. Recently, the bed format is frequently used to show differentially expressed genes. J-Circos accepts both formats:

Bed (http://genome.ucsc.edu/FAQ/FAQformat.html#format1);

and Bigbed (http://genome.ucsc.edu/goldenpath/help/bigBed.html).

Following the definition of Bed as per the UCSC genome browser, the first three columns are necessary and the others are optional. The color of the Bed rectangle is determined by the column "ItemRgb". The value of "ItemRgb" consists of three integers delimited by comma. If column "ItemRgb" is 0 or empty, the Bed rectangle is colored black. The column "name" in the Bed/bigBed file will be shown in mouse-over display. Since circos plot displays massive genomic scales, all genes in the circos plot usually becomes a line. In Figure 4, we use Galaxy to convert miRNA data into Bed files and show these 1870 miRNAs as a bed format track.
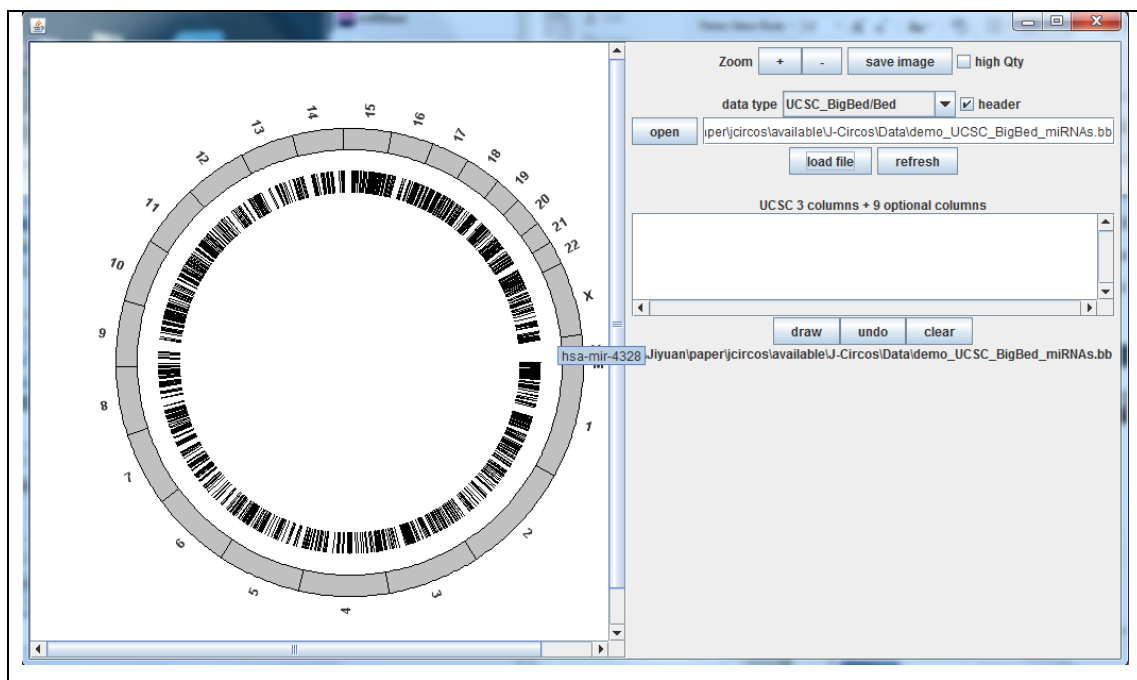


**Figure 4 Plot bed format with human miNRA data**

However, if the genome is very small, detailed transcript structures can be shown (Figure 5).
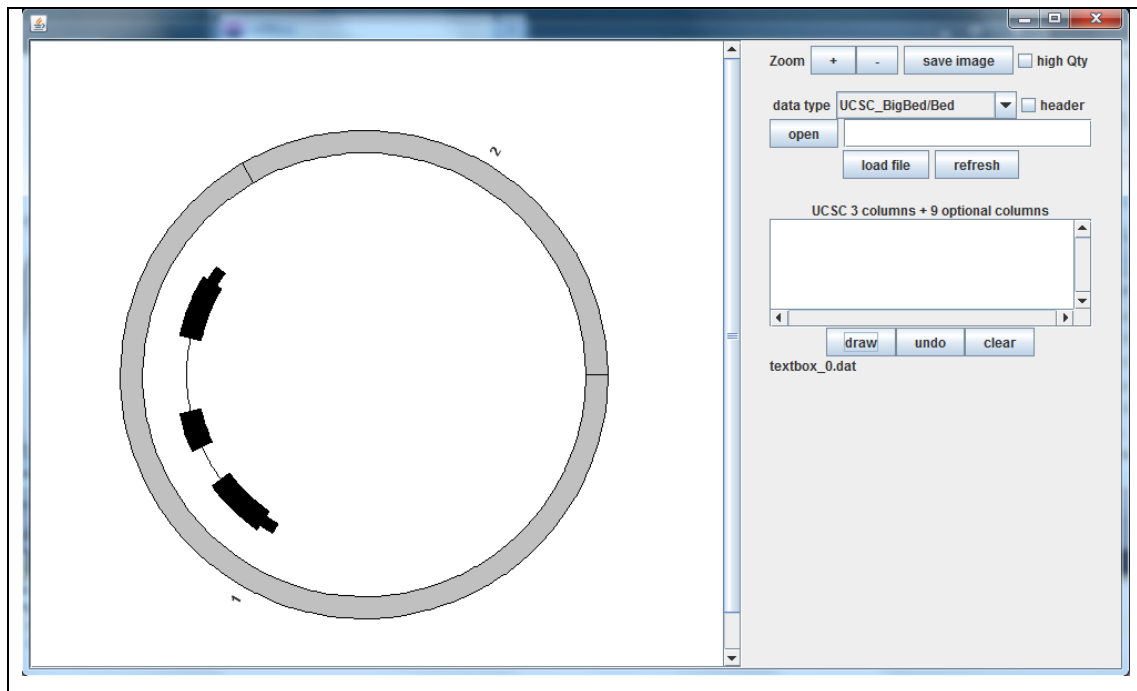
**Figure 5 Display of transcript in circos style**

### 1.6.3 Plot Circos_wiggle

Circos wiggle is the most functional data format. Unlike the UCSC wiggle format, where the wiggle boxes are always fixed to the base line, Circos wiggle can elevated boxes above or below the base line. Moreover, each box can have different colors.

The first line "Radius 150" is the Radius of the base line. This is optional. A popup window will appear for users to input the baseline radius if the line is missing.

1. Color: RGB format, i.e. three integers delimited by comma
2. Chrom: chromosome name
3. Start: the wiggle box start locus
4. Length: the width of the wiggle box
5. Height1: the distance between the baseline and the wiggle box. If the value is minus, the bottom of the wiggle box will be plotted below the baseline.
   **The following columns are optional.**
6. Height2: the distance between the baseline and the top of wiggle box. The top of the wiggle box will be plotted below the baseline if the value is minus. If this column is omitted, the height1 value becomes height2 and height1 becomes 0.
7. Description: the contents will be appear when the mouse hovers over the wiggle box. If this column is omitted, the value becomes empty "".
8. Filled: the wiggle box is filled when the "TRUE" value is set, otherwise the wiggle box is empty. If the column is omitted, The value is assumed to be "TRUE".

Figure 6 and Figure 7 show examples to plot wiggle in different styles.

| Radius | 150 | | | | | | |
|---|---|---|---|---|---|---|---|
| Color | chrom | start | length | height1 | height2 | description | filled |
| 120,0,0 | chr5 | 20000000 | 10000000 | 0 | 20 | I_am_in_chr5 | TRUE |
| 0,0,249 | chr6 | 30000000 | 10000000 | 0 | 30 | I_am_in_chr6 | TRUE |
| 0,230,0 | chr8 | 30000000 | 20000000 | 30 | 50 | I_am_in_chr8 | TRUE |
| 0,230,0 | chr10 | 50000000 | 30000000 | -50 | -30 | I_am_in_chr10 | TRUE |
| 255,0,0 | chr9 | 1 | 100000000 | 10 | 40 | I_am_in_chr9 | FALSE |

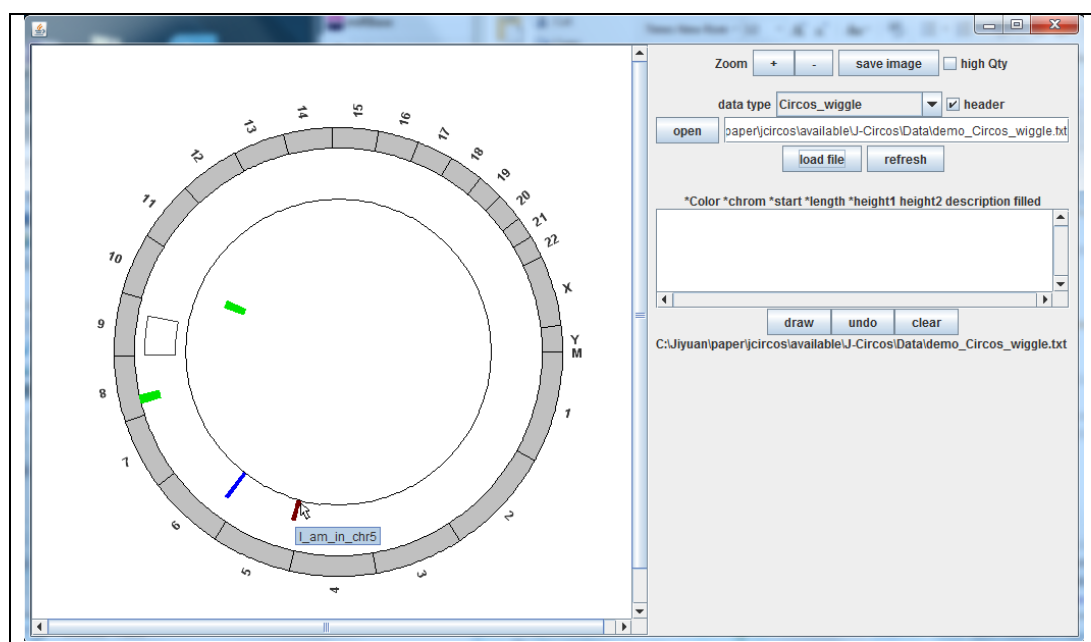**Figure 6 Data of demo_Circos_wiggle.txt.**



**Figure 7 Plot wiggle from demo data. Mouse hover displays the information of description**

**column.**

## 1.6.4  Plot Circos_bridge

The Circos_bridge is mainly used for displaying genomic translocations and gene fusions.

The first line of "Radius 150" in Figure 8 means that the base line has 150 Radius. The first line is optional. A popup window will appear for users to input the baseline radius if the line is missing.

1. Color: RGB format, i.e. three integers delimited by commas.

2. Chrom1: chromosome name of one side of the Circos bridge.

3. Start1: start locus of one side of the circos bridge.

4. Length1: genomic length of one side of the circos bridge.

5. Description1: for mouse-over display.

6. Chrom2: chromosome name of other side of the Circos bridge.

7. Start2: start locus of other side of the circos bridge.

8. Length2: genomic length of other side of the circos bridge.

9. Description2: description1 and description2 are concatenated to show up when the mouse hovers on either root sides of the circos bridge.

| Radius | 150 | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Color | chrom1 | start1 | length1 | description1 | chrom2 | start2 | length2 | description2 |
| 255,0,0 | chr2 | 30000000 | 10000000 | from_chr2 | chr3 | 40000000 | 50000000 | to_chr3 |
| 0,0,225 | chr5 | 20 | 20 | from_chr5 | chrX | 10 | 10 | to_chrX |
| 0,255,0 | chr6 | 10000000 | 1 | from_chr6 | chr9 | 999999 | 1 | to_chr9 |
| 0,200,0 | chr1 | 10000000 | 1 | from_chr1 | chrY | 999999 | 1 | to_chrY |
| 120,30,100 | chr5 | 100000000 | 1 | from_chr5 | chr13 | 99999999 | 1 | to_chr13 |
| 20,230,200 | chr8 | 10000000 | 1 | from_chr8 | chr17 | 99999999 | 1 | to_chr17 |
| 100,0,120 | chr4 | 1000000 | 1 | from_chr4 | chr10 | 9999999 | 1 | to_chr10 |

**Figure 8 Bridge demo data**
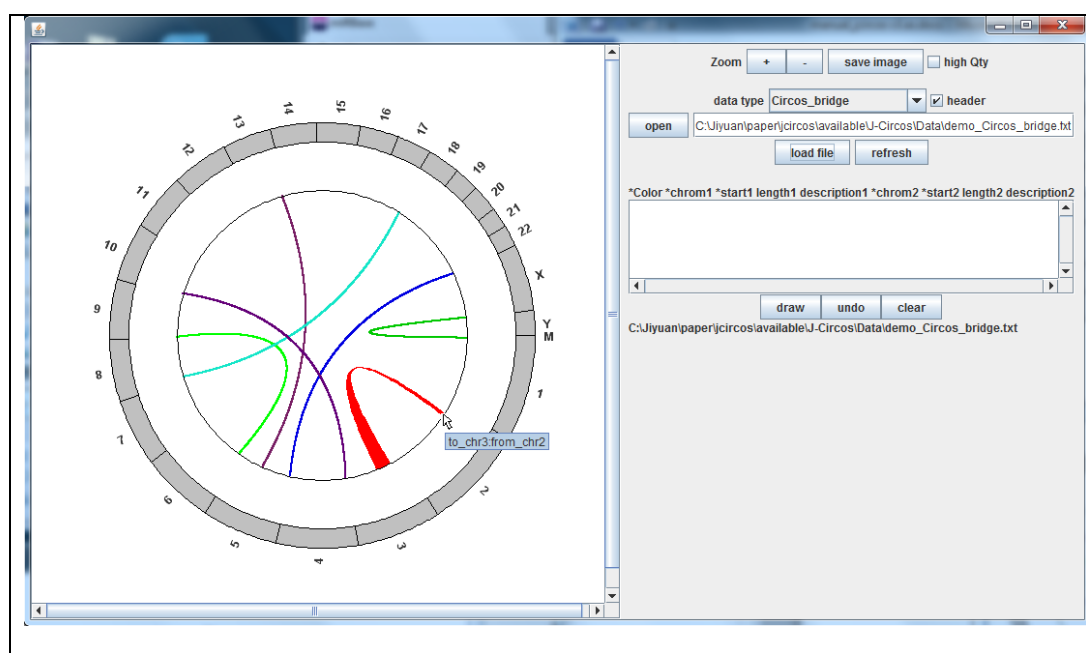


**Figure 9 Bridge plot output.**

### 1.6.5 Plot Label

A specific locus is indicated to enable detailed information of circos plots/points.

The first line of "Radius 200" in Figure 10 means the base line has 200 Radius. The first line is optional. A popup window will appear for users to input the baseline radius if the line is missing.

1. color: RGB format i.e. three integers delimited by commas.

2. chrom: chromosome name of the label to point.

3. start: the genomic locus for the label to point.

4. radius_display: the Radius of the location to plot the label.

5. chrom_display: the chromosome of the location to plot the label.

6. start_display: the location to plot the label.

7. str: what to display.

| Radius | 200 | | | | | |
|---|---|---|---|---|---|---|
| Color | chrom | start | Radius_display | Chrom_display | start_display | str |
| 155,0,0 | chr11 | 20000000 | 180 | chr11 | 69000000 | AAA |
| 152,55,122 | chr11 | 27000000 | 180 | chr11 | 100000000 | BBB |

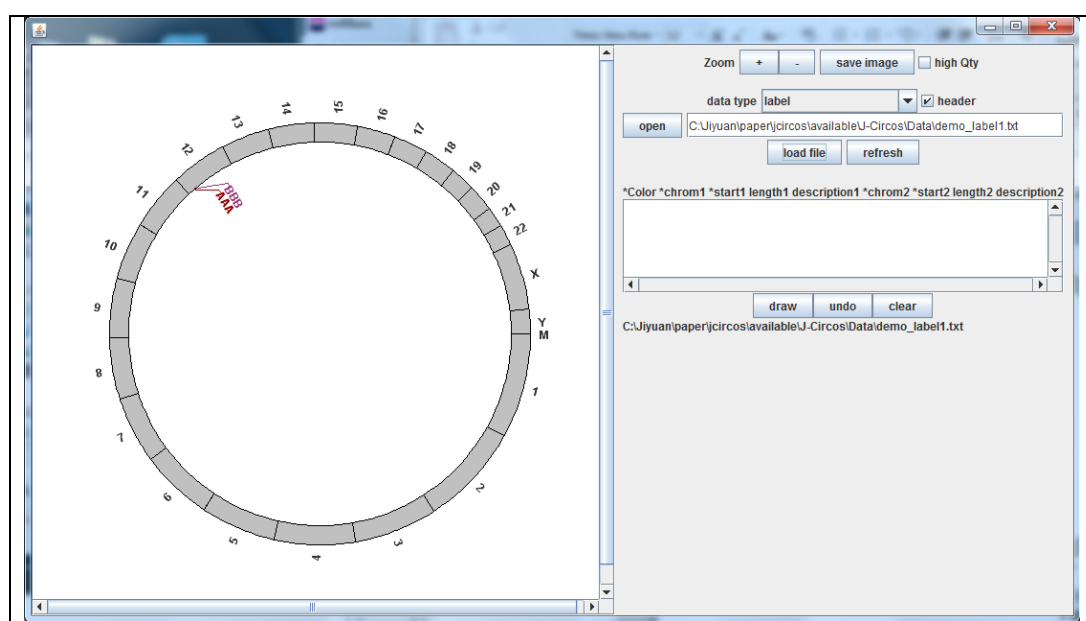**Figure 10 Demo data for plot the label inside the cycle**



**Figure 11 Plot inside lable**

Figure 13 shows the label can be put outside of the chrmosmsom circos using data from Figure 12.

| Radius | 220 | | | | | |
|---|---|---|---|---|---|---|
| Color | chrom | start | Radius_display | chrom_display | start_display | str |
| 155,0,0 | chrX | 20001000 | 270 | chr22 | 1 | CCC |
| 15,5,255 | chrX | 25000000 | 270 | chr22 | 20000000 | DDD |

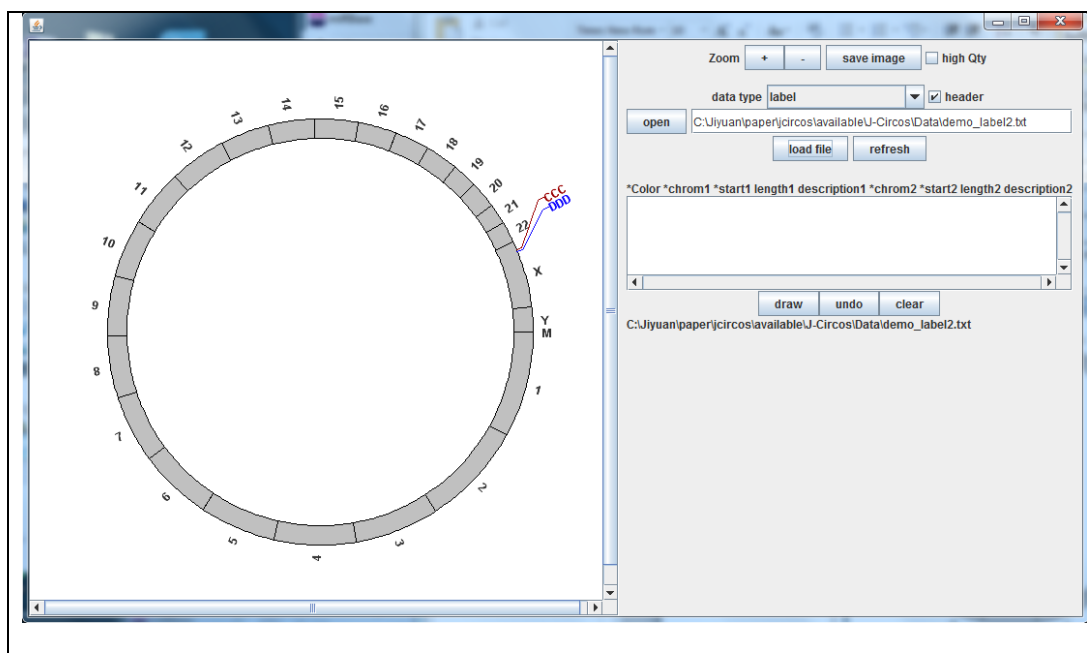**Figure 12 Demo data for plot the label inside the cycle**

**Figure 13 Plot label**

## 1.6.6 Plot text

The Circos text function is for printing text in the circos plot.

The first line of "Radius 150" in Figure 14 means the base line has 150 Radius. The first line is optional. A popup window will appear for users to input the baseline radius if the line is missing.

1. Color: RGB format i.e. three integers delimited by commas.
2. Chrom: chromosome name where the text is to be placed.
3. Start: the genomic locus for the text.
4. Text: what is to display.
5. Type: there are two types: tangent and perpendicular, which show the direction of the text in the circos plot (Figure 15).

| Radius | 150 | | | |
|--------|-------|---------|-------------------------|---------------|
| Color | chrom | start | text | type |
| 255,0,0 | chr3 | 1000000 | text_radio_demo | radio |
| 255,0,0 | Chr13 | 1 | text_circumference_demo | circumference |

**Figure 14 Demo text data**

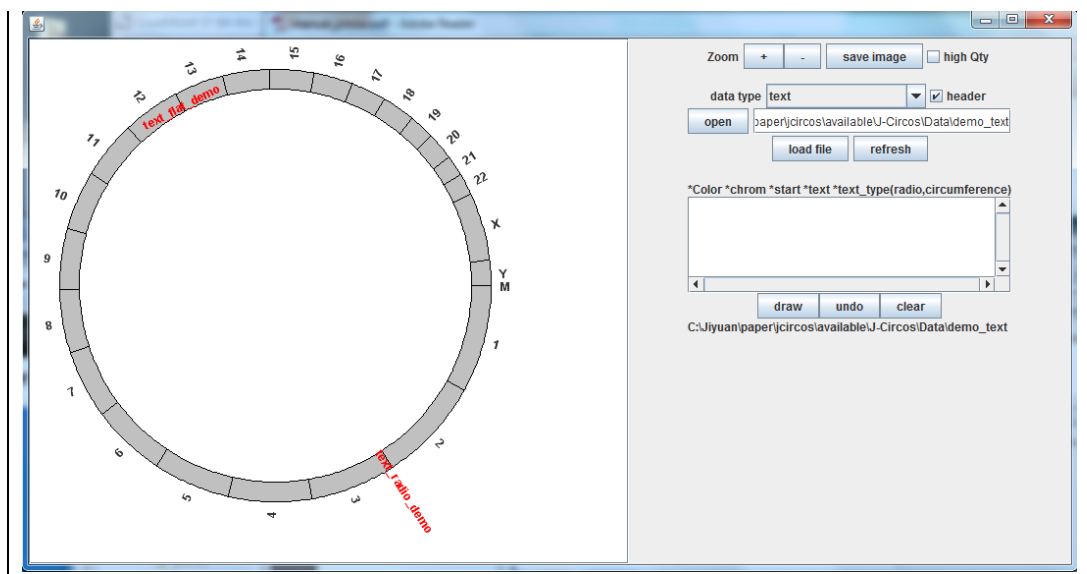**Figure 15 Plot output of demo text data**

## 1.6.7 Plot cytogenetic

Cytogenetic files can be downloaded from the UCSC genome browser (http://genome.ucsc.edu) or galaxy (http://galaxyproject.org/).

Figure 16 shows the cytogenetic map for hg19. The cytogentic data has the data format shown in Figure 17. The Circos cytogentic for hg19 is shown in Figure 18.
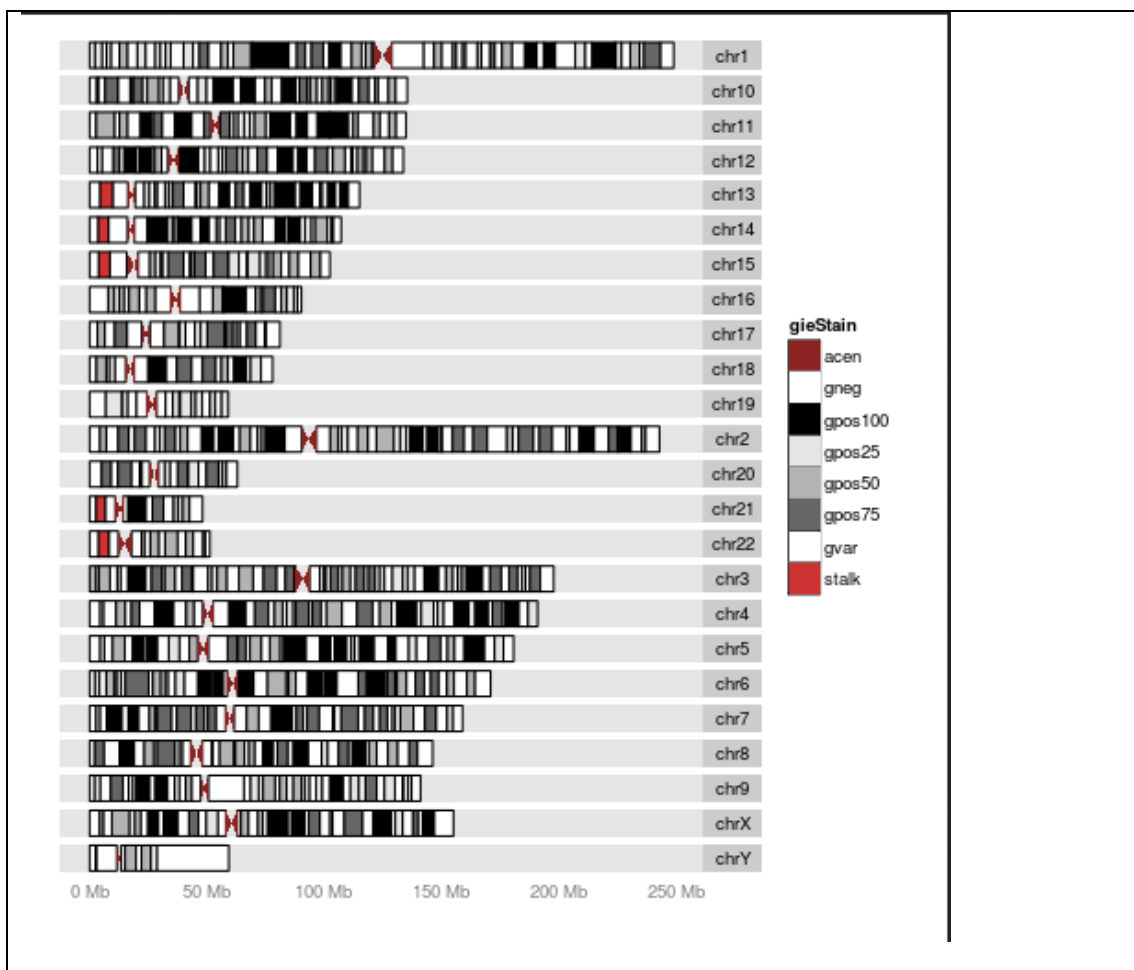
**Figure 16 Cytogenetic for hg19**

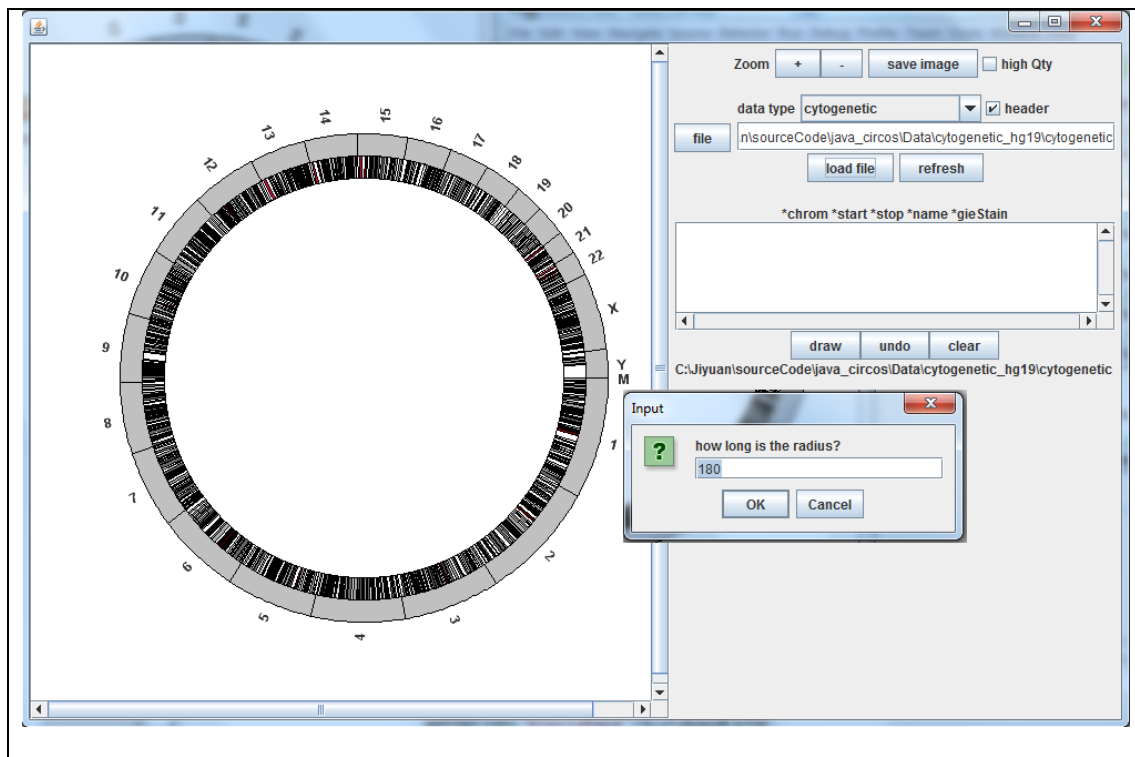| #chrom | chromStart | chromEnd | name | gieStain |
|--------|-----------|----------|------|----------|
| chr1 | 0 | 2300000 | p36.33 | gneg |
| chr1 | 2300000 | 5400000 | p36.32 | gpos25 |
| chr1 | 5400000 | 7200000 | p36.31 | gneg |
| chr1 | 7200000 | 9200000 | p36.23 | gpos25 |
| chr1 | 9200000 | 12700000 | p36.22 | gneg |
| chr1 | 12700000 | 16200000 | p36.21 | gpos50 |

**Figure 17 G-band data for**



**Figure 18 Plot of cytogenetic. A pop up window for inputting the radius of cycle of cytogenetic.**