

Visualization II Circular plot

R package **circlize**

CIRCOS – an information aesthetic for comparative genomics

M Krzywinski, J Schein, I Birol, S Jones, M Marra

Canada's Michael Smith Genome Sciences Centre // British Columbia Cancer Research Centre // www.bcgsc.ca



circos - round is the new square

We have created a visualization tool, called CIRCOS, to facilitate exploring relationships between genomes and in general any kind of position in maps that relate genomic intervals. Structural variation data such as these, produced by sequence alignment and hybridization arrays, underpin comparative studies but are opaque to conventional visualization methods designed for 2D data sets. Compared to other tools [1,2], CIRCOS is unique in its combination of **circular data domain layout**, support for a large number of **diverse data tracks**, **global and local length scale control**, **viewer customization** and **automation**, and maintaining a **high data-to-ink ratio** [3] without sacrificing clarity of presentation. CIRCOS has been used within the genomics community [4-6] and its flexibility and aesthetic has garnered interest from mainstream periodicals and newspapers [7-9] and, recently, illustrate the dynamics of a US presidential debate [10].

DOWNLOAD CIRCOS AT <http://mkweb.bcgsc.ca/circos>

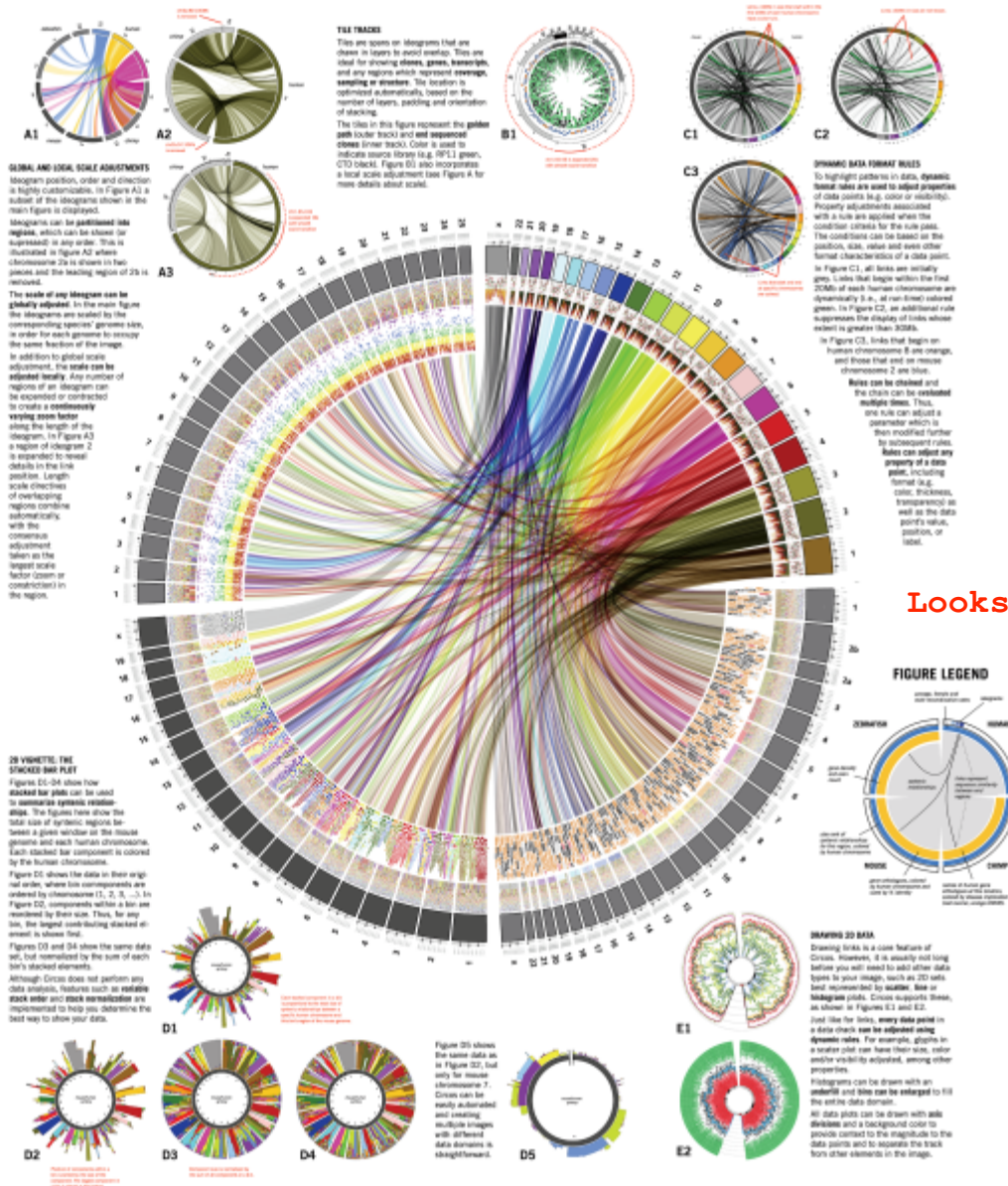
At present, laboratories are hard-pressed not only to store and analyze, but to visualize the reams of data produced by ultra-high-throughput technologies, such as massively parallel sequencing. Because analytically extracting informative patterns from these large data sets is very difficult, automated visualization tools that generate informative vignettes of the data are valuable in data mining and formulating hypotheses.

The design of CIRCOS is based on the fact that a circularly composited ideogram layout can encode relationships between genomic regions more informatively than a linear layout. These relationships are visually encoded by links which can be either straight lines or Bézier curves whose control point location can be highly customizable. Other data types that are supported are **scatter plots**, **line plots**, **histograms**, **file plots**, **heatmaps**, and **glyph labels**, **highlights**, **ideograms**, and **labeled ticks**. The radial position of data tracks is controlled by the user and their angular extent is a function of the extent of the data domain. Data tracks such as the file and tick label have their individual elements automatically positioned to avoid overlap.

CIRCOS uses **plain text files** for both **input data** and **configuration**. The latter controls the placement and format of each data track. The ability to generate both data and configuration files automatically makes CIRCOS highly amenable to incorporation in web-based database mining and visualization.

A feature unique to CIRCOS is not only the ability to adjust the length scale for each ideogram (e.g. display chr 17 at 3x normal size), but to **smoothly vary the length scale** locally, effectively zooming (or contracting) regions of interest while still displaying the entire data domain (Figures A1-A3, B1). This global and local scale adjustment is useful when illustrating genomic regions in which data density is highly variable. Furthermore, to help draw attention to important data, the ideograms can be divided into any number of disjoint regions, which in turn can be drawn in any color. The resulting **aria breaks** can be marked up in various styles on the final image to clearly mark the disruption.

Every aspect of the final image is customizable and output can be generated in either **bitmap** or **SVG** format. For example, the thickness, outline and color of the ideogram track is customizable, as are the corresponding features of the cytogenetic bands. The radial position for each ideogram can be independently set. Each data track, and individual primitives within a track, has an associated z-depth value, which controls how elements stack. Finally, every data type format characteristic, such as color, thickness, data value, label and visibility, can be adjusted by **dynamic formatting rules** based on data position and other format values at run-time (Figures C1-C3). These rule sets are stored in the configuration files and separate the definition and storage of formatting rules from the raw data.



<http://circos.ca>

R package **circlize**

Looks like Poincare Metrics

Circos plot in R: **circlize**

- *Circlize* implements circos plot in R, which is an ideal environment providing seamless connection between analysis and visualization
- *Circlize* is not a front-end wrapper for circos
- *Circlize* keeps the flexibility and configurability of circos, but also makes it more straightforward to use and enhance its compatibility with other types of graphical designs



Design Principle

- *Circlize* implements low-level graphic functions for adding graphics in circular layout

<code>circos.points</code>	add points in a cell, similar as points.
<code>circos.lines</code>	add lines in a cell, similar as lines.
<code>circos.rect</code>	add rectangle in a cell, similar as rect.
<code>circos.polygon</code>	add polygon in a cell, similar as polygon.
<code>circos.text</code>	add text in a cell, similar as text.
<code>circos.axis</code>	add axis in a cell, functionally similar as axis but with more features.
<code>circos.link</code>	this maybe the unique feature for circular layout to represent relationships between elements.

More handy functions

circos.trackPoints	can be replaced by <code>circos.points</code> through a for loop.
circos.trackLines	can be replaced by <code>circos.lines</code> through a for loop.
circos.trackText	can be replaced by <code>circos.text</code> through a for loop.

Functions for layout:

circos.trackPlotRegion	create plotting regions for cells in a track.
circos.updatePlotRegion	update an existed cell
circos.par	graphic parameters
circos.info	print general parameters of current circos plot
circos.clear	reset graphic parameters and internal variables



Some Examples

1. Close-Ups of the Genome, Species by Species by Species (
http://www.nytimes.com/imagepages/2007/01/22/science/20070123_SCI_ILLO.html)
2. The Global Flow of People (
<http://www.global-migration.info/>)
3. [Not R] Circos Tools to get started (
<http://mkweb.bcgsc.ca/tableviewer/>)

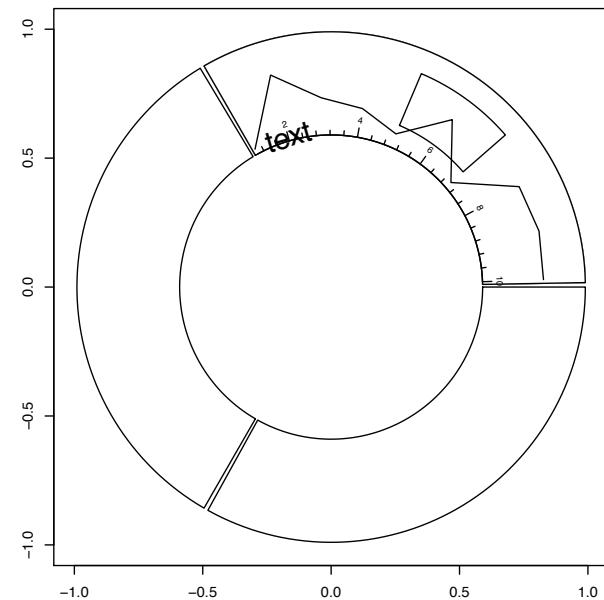
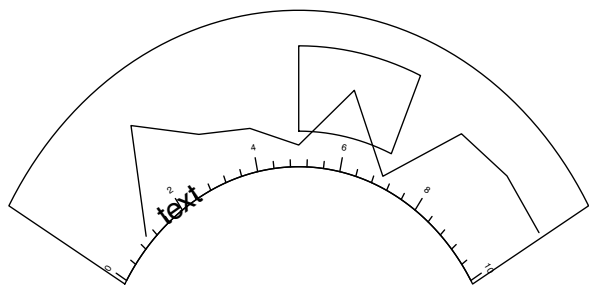
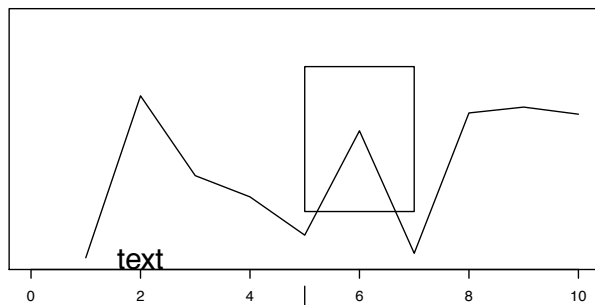


A step-by-step *Circlize* Example in R

Practical session

Coordinate Transformation

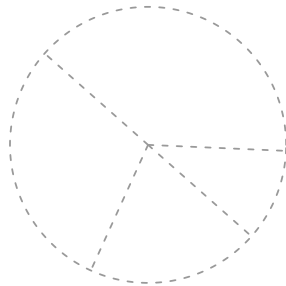
- Data coordinate \rightarrow polar coordinate \rightarrow canvas coordinate



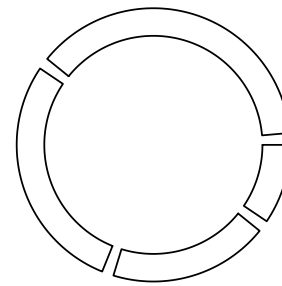
Steps for making circular layout

Initialize → create track → add graphics → create track →
add more graphics → ... → **clear**

`circos.initialize`

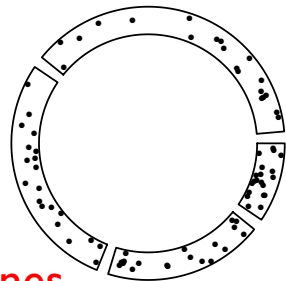


`circos.trackPlotRegion`

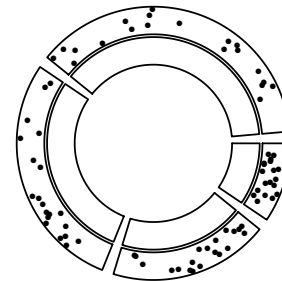


**I: use low-level graphics
(basic)**

`circos.points`
`circos.lines`
`circos.text`
...

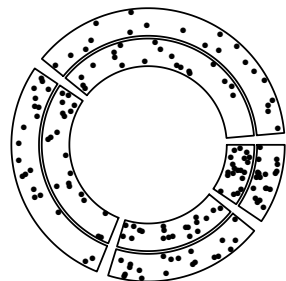


`circos.trackPlotRegion`



**II: use `circos.trackPoints/Lines`
(not recommended)**

`circos.points`
`circos.lines`
`circos.text`
...



**III: `panel.fun` in `circos.trackPlotRegion`
(recommended)**

...
`circos.clear`

▶ Use `circos.updatePlotRegion` to overwrite

Poincare Metrics? The length of line has something to do with rank
Not the Poincare Measure.
Overlapping lines, orders.
Pull one piece up to observe the link, like a pizza

More *Circlize* Examples in R

Circlize.R