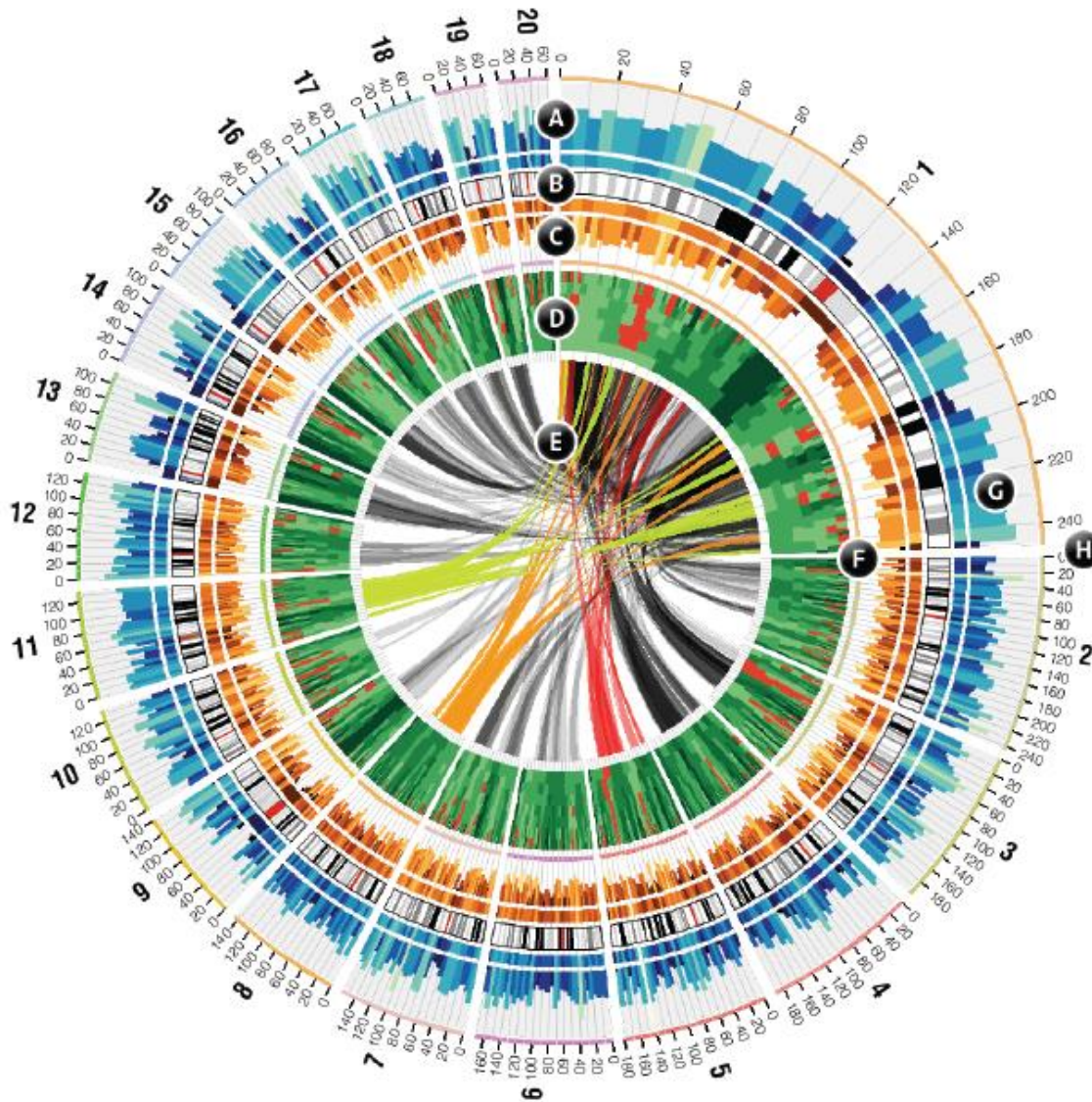# Visualizing Genomes: Creating Circos Plots

Overview

- Visualization Challenges

- Circos Plots – What is it?

- Circos Plots Applications

- Creating Circos Plots
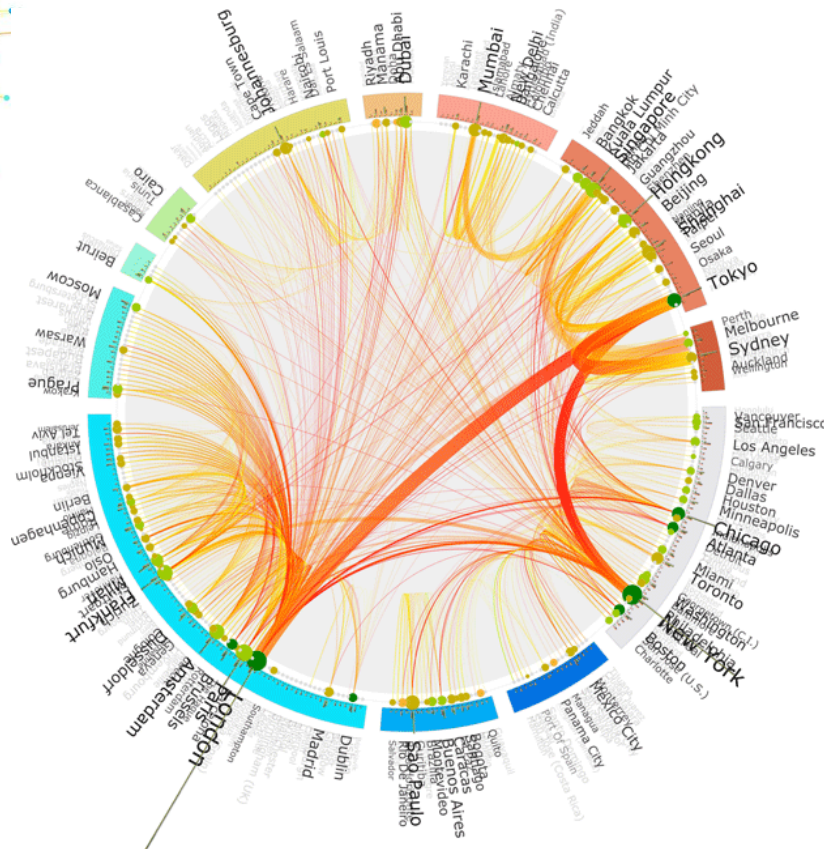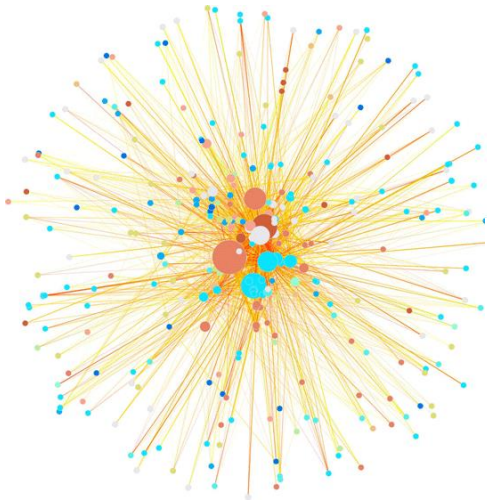 - ➢ Hands-on: commands and scripting

Circos Plots: Example

Available Tracks/Display:
- A: Histogram
- B: Ideogram
- C: Histogram (inverted)
- D: Heatmap
- E: Links
- F: Highlights
- G: Grid
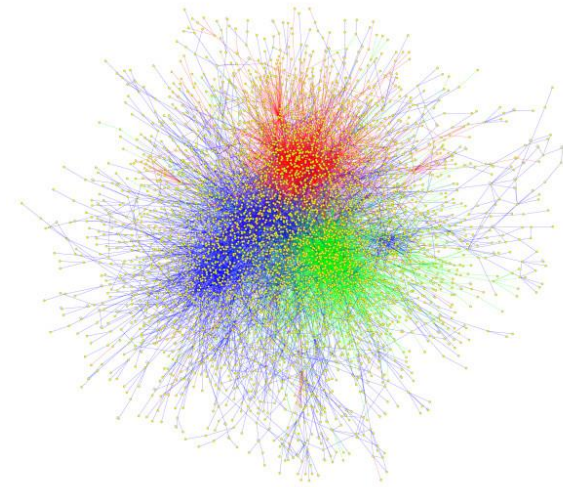- H: Ticks

3

Circos Plots: Example

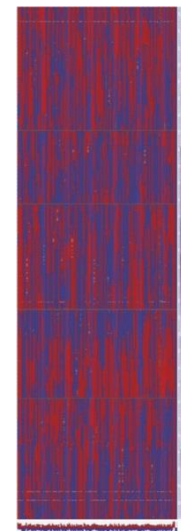- Visualization of information-rich geographical network

Visualization Challenges

- Rate limiting step is not data generation but the analysis (including visualization)

- Understanding and interpreting complex data

- Information dense figures can be overwhelming
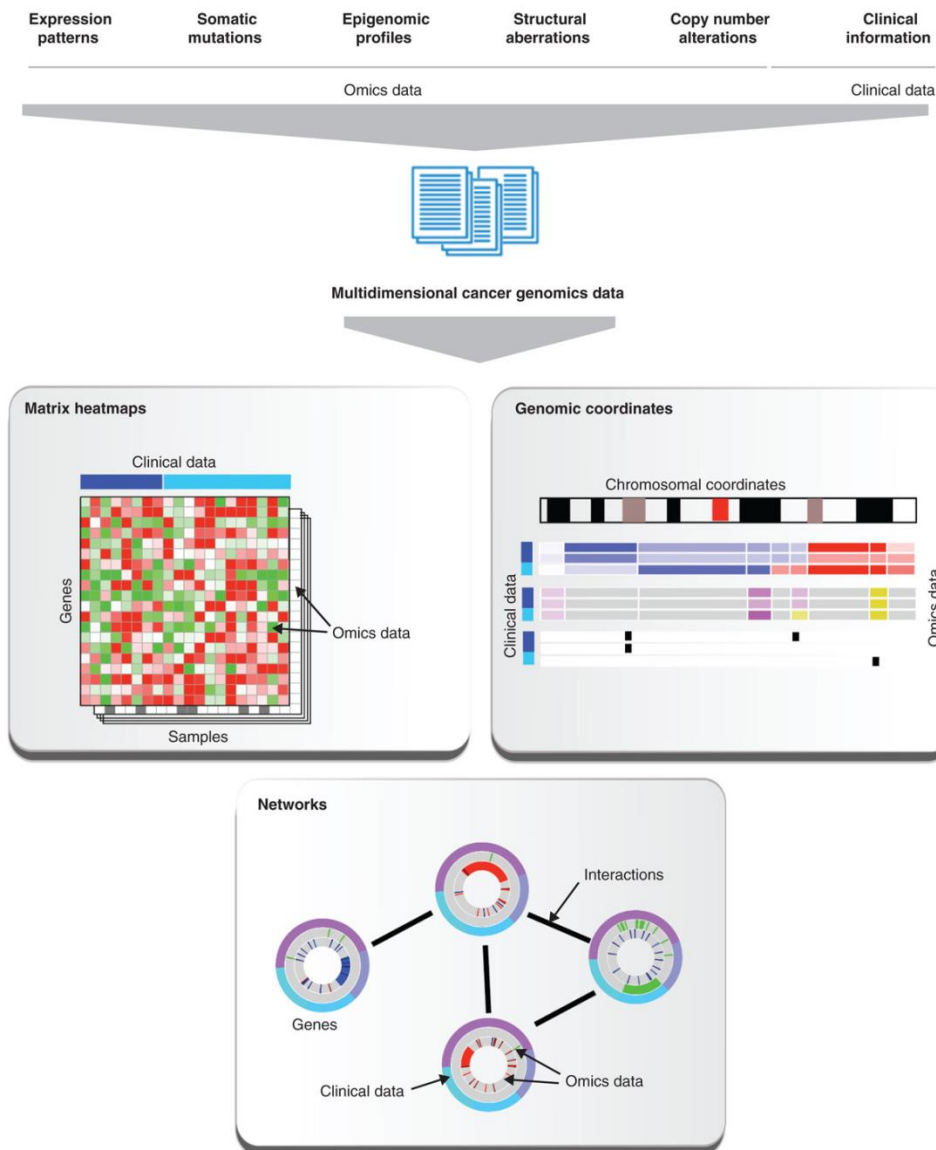
BMC Biology 2010, 8:40

Genome Res. 2006 Jun;16(6):787-95

5

Nielsen, C.B., et al. *Visualizing genomes: techniques and challenges* Nature Methods 7: S5-S15 (2010)
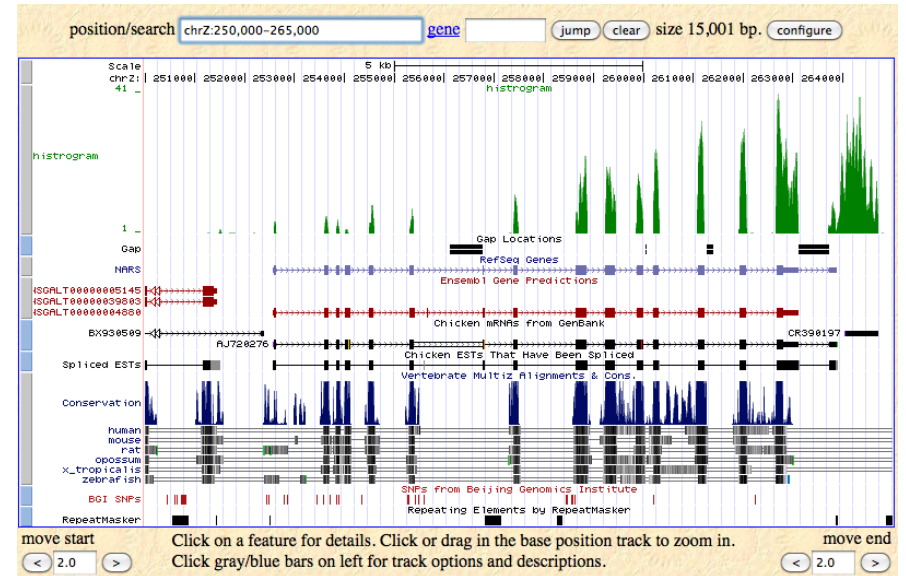
Visualization Challenges

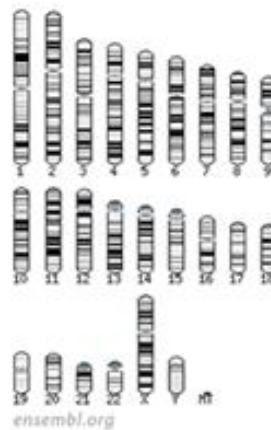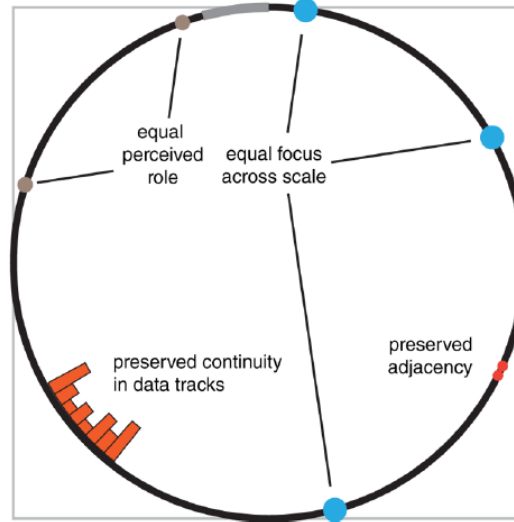- Viewing multidimensional data

6

Visualization Challenges

- Traditional browsers are linear: good for visualizing specific regions but difficult to get a *global* view

- Viewing genomic regions that are not adjacent is not easy on a regular browser

- *Stacked tracks* may require scrolling up and down

Visualization Challenges: Linear vs Circular

- Continuity and focus

Circos Plots: Overview

- No relationship to circular DNA, however, that too can be displayed

- Over ~350 citations (May 2013)

- Not limited to biological or genomic data, almost any kind of relationship data can be visualized in Circos

- Too many tracks on a Circos plot can be difficult to understand
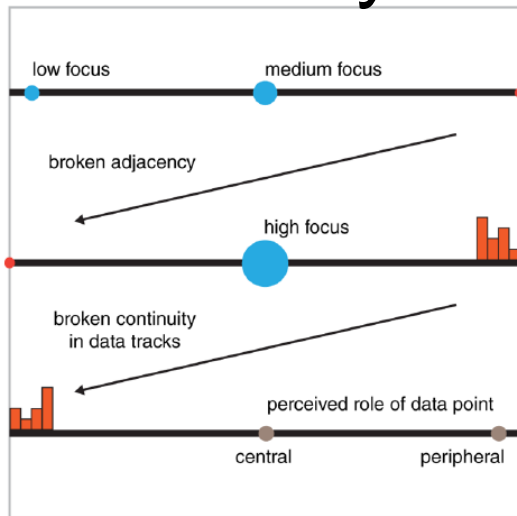
Circos: Software

- All input files are text
- Output are image files (png and svg format)
- Requires configuration file(s) to specify Circos layout and data tracks
- Comment lines begin with hash tag, #
- Circos does *not* do any analysis, it's only for visualization
- Created images are static, image details must be specified in the configuration files
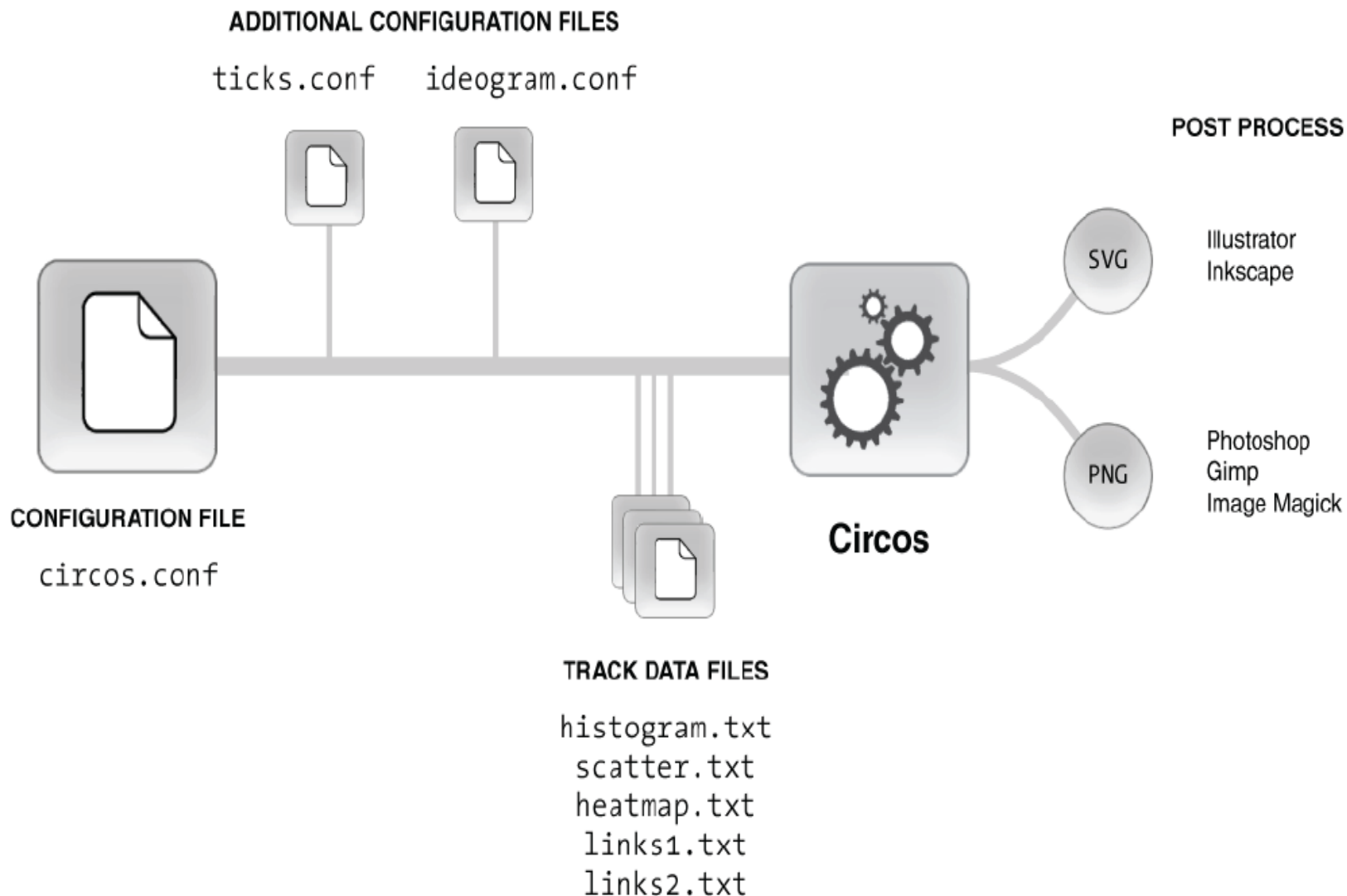- Run on command-line

http://commons.wikimedia.org/wiki

Creating Circos Plots: Pipeline

Usage: `circos –conf <configFile>`
     eg. `circos –conf circos.conf`

Creating Circos Plots:
Circos Distribution Contents

- bin/   Circos executable
- etc/   Configuration files
- fonts/ Fonts used by Circos
- lib/   Libraries
- tiles/   Tiles for pattern fills
- tools/ Helper tools for Circos

➤ **On tak,** `/usr/local/share/circos`

Creating Circos Plots: conf files

- Configuration files specifies the image rendering (eg. color, font, etc.)

- Configuration syntax (html-like format)

  ➢ variable assignment

```
variable = value
```

  ➢ Blocks

```
<ideogram>
 thickness = 30p
 fill       = yes
 ...
</ideogram>
```

  ➢ Nested Blocks

```
<plots>
  <plot>
   file       = data/set1.txt
   color      = black
  </plot>

  <plot>
   file       = data/set2.txt
   color      = red
   ...
  </plot>
</plots>
```

Creating Circos Plots: conf files

- ## Global vs Local

```
<plots>            #start of plots block


type  = heatmap
min   = 0
max   = 1                  Global to all plots


<plot>             #start of inner plot block
file = data.1.txt
r1    = 0.6r
r0    = 0.5r               Specific to data.1.txt plot
...
</plot>            #end of inner plot block


<plot>
file = data.2.txt
r1    = 0.7r
r0    = 0.6r

...
</plot>
</plots>            #end of plots block
```

Creating Circos Plots: conf files

- ## Units

  - ➢ b (bases) - used to indicate distance along the ideogram
  - ➢ p (pixels) - used for quantities defined in absolute pixel size, such as track radius, label size, glyph size, and others.
  - ➢ r (relative) - quantifies a parameter relative to another value, which is sometimes more intuitive than using absolute pixel values.
  - ➢ u (chromosome units) - special relative unit which expresses distance long ideogram in terms of the chromosomes_unit value
  - ➢ Examples:

```
# 1 pixel padding
padding = 1p
# relative padding (e.g. relative to label width)
padding = -0.25r
# radius of track (relative to inner ideogram radius)
r0 = 0.5r
# combination of relative and pixel values
r1 = 0.5r+200p
```

Creating Circos Plots: conf files

- Imports

  ➢ Should always be imported

  ```
  # colors, fonts and fill patterns
  <<include etc/colors_fonts_patterns.conf>>
  # system and debug parameters
  <<include etc/housekeeping.conf>>
  ```

  ➢ Others as needed

  ```
  <<include ideogram.conf>>
  <<include ticks.conf>>
  ```

Creating Circos Plots: Hands-on

- ## Ideograms

  - ### Chromosome

  `chr - ID LABEL START END COLOR`

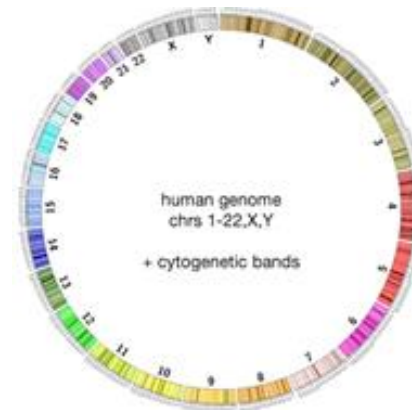 ## Example:
  ```
  chr - hs1 1 0 247249719 brown
  chr - hs2 2 0 242951149 green
  ...
  ```
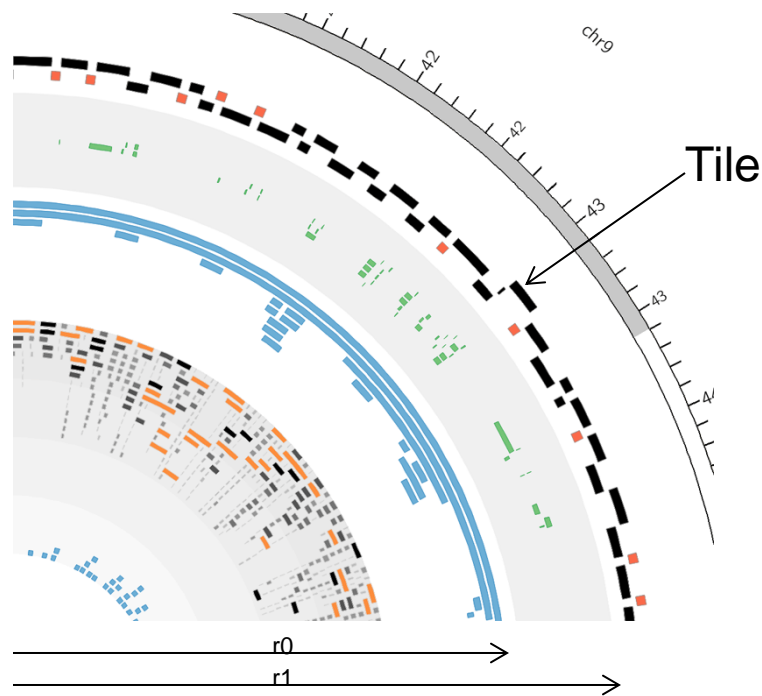
  - ### Cytogenetic Bands
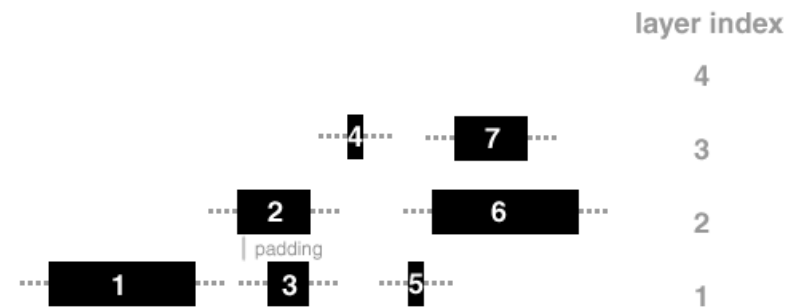
  `band ID parentChr parentChr START END COLOR`

human genome
chrs 1-22,X,Y

human genome
chrs 1-22,X,Y
+ cytogenetic bands

Circos Plot Applications: Tiles

- Tracks used to show spans or genomic regions (eg. genes, reads, etc.)

Tile

r0
r1

r0 and r1 used to specify the position of the track on the Circos plot.

Tiles are placed in layer with smallest index that can accomodate tile's extent without overlap with other tiles in the layer. Tile's extent is defined as the region **[start-margin,end+margin]**. Spacing between layers is defined by **padding**. Relationship between layer index and layer distance from center of circle is defined by tile plot **orientation** (*in*, *out*, or *center*).

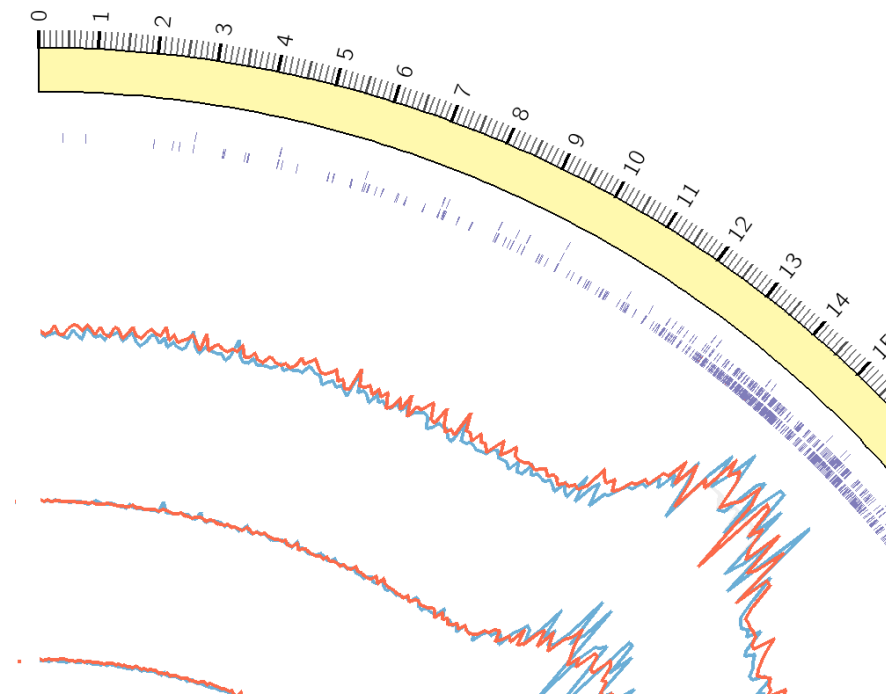# Circos Plot Applications Hands-on: Studying Variants
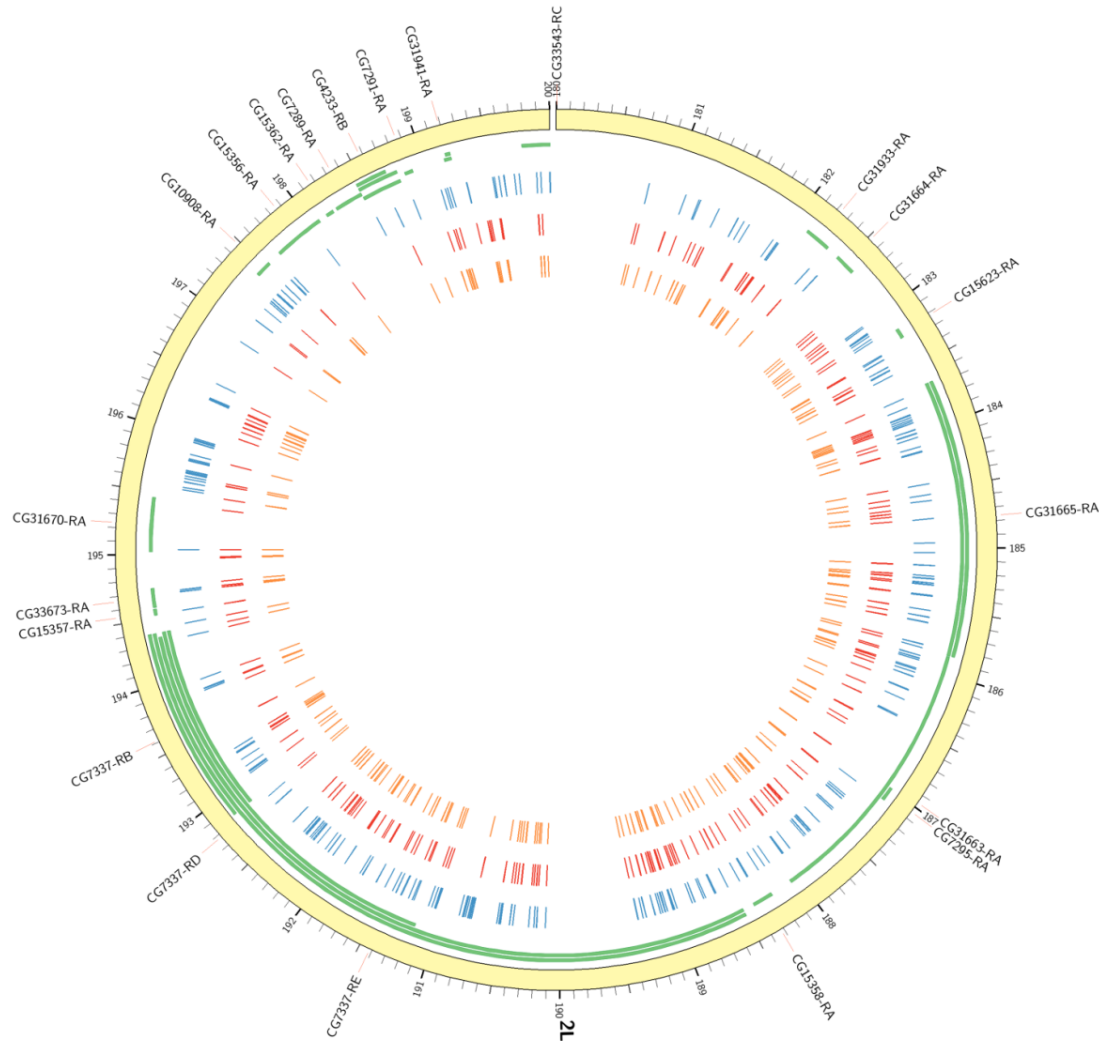
Display of 200kb region in fly chr2L showing variants in 3 strains (orange, red, blue) along with genes (green) in the region (*Orr-Weaver Lab*).
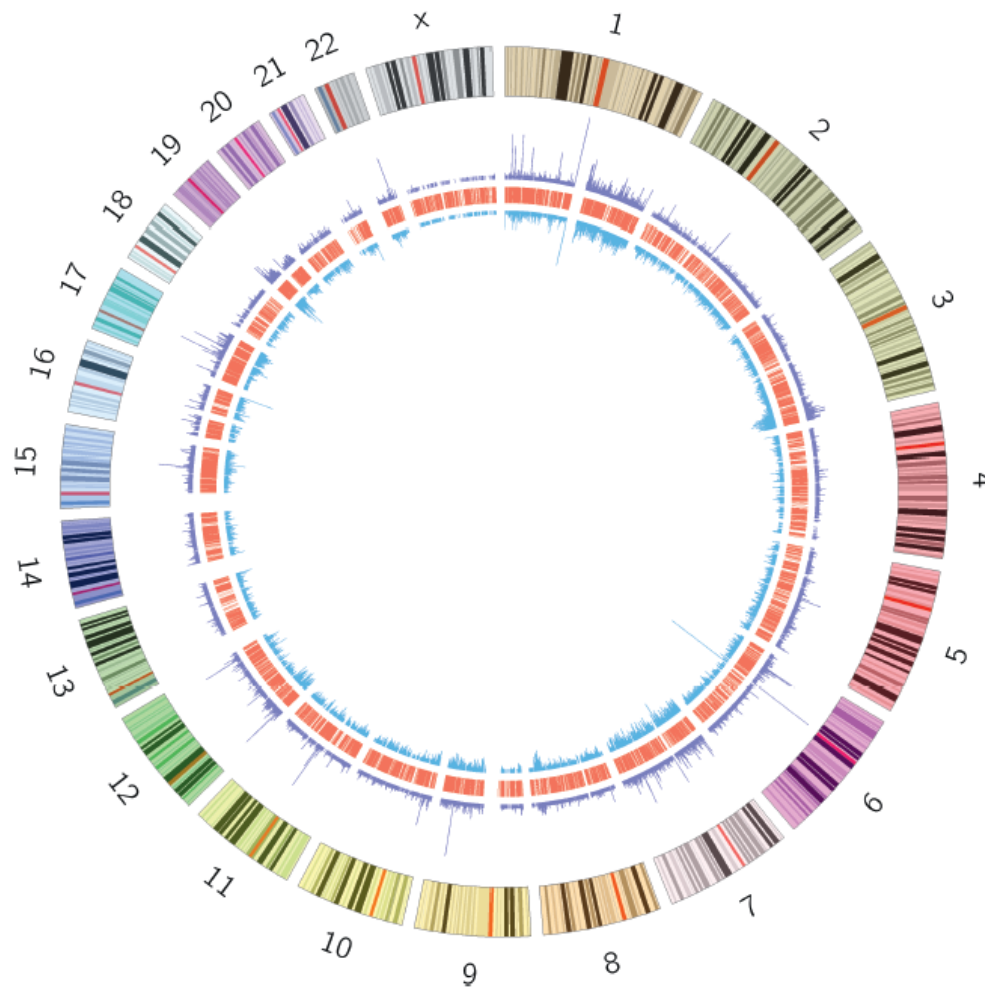
Circos Plot Applications: Line Plots

- Tracks used to show adjacent discrete data points (eg. read count) connected by a single line

Methylation profiles (red and blue) on a chromosome segment (*Gehring Lab*)

20

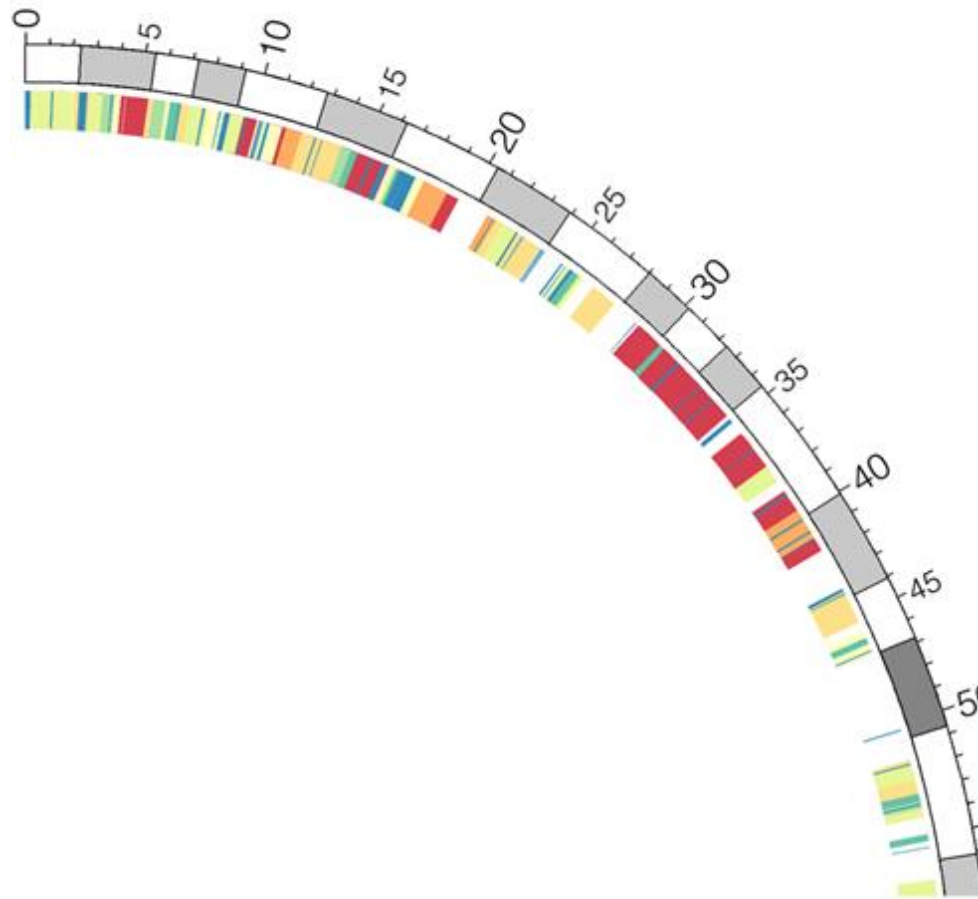# Circos Plot Applications Hands-on: Profiling

Visualization of co-bound regions profile from 2 ChIP-Seq experiments (purple and blue) along with genes (red). (*Sabatini Lab*)
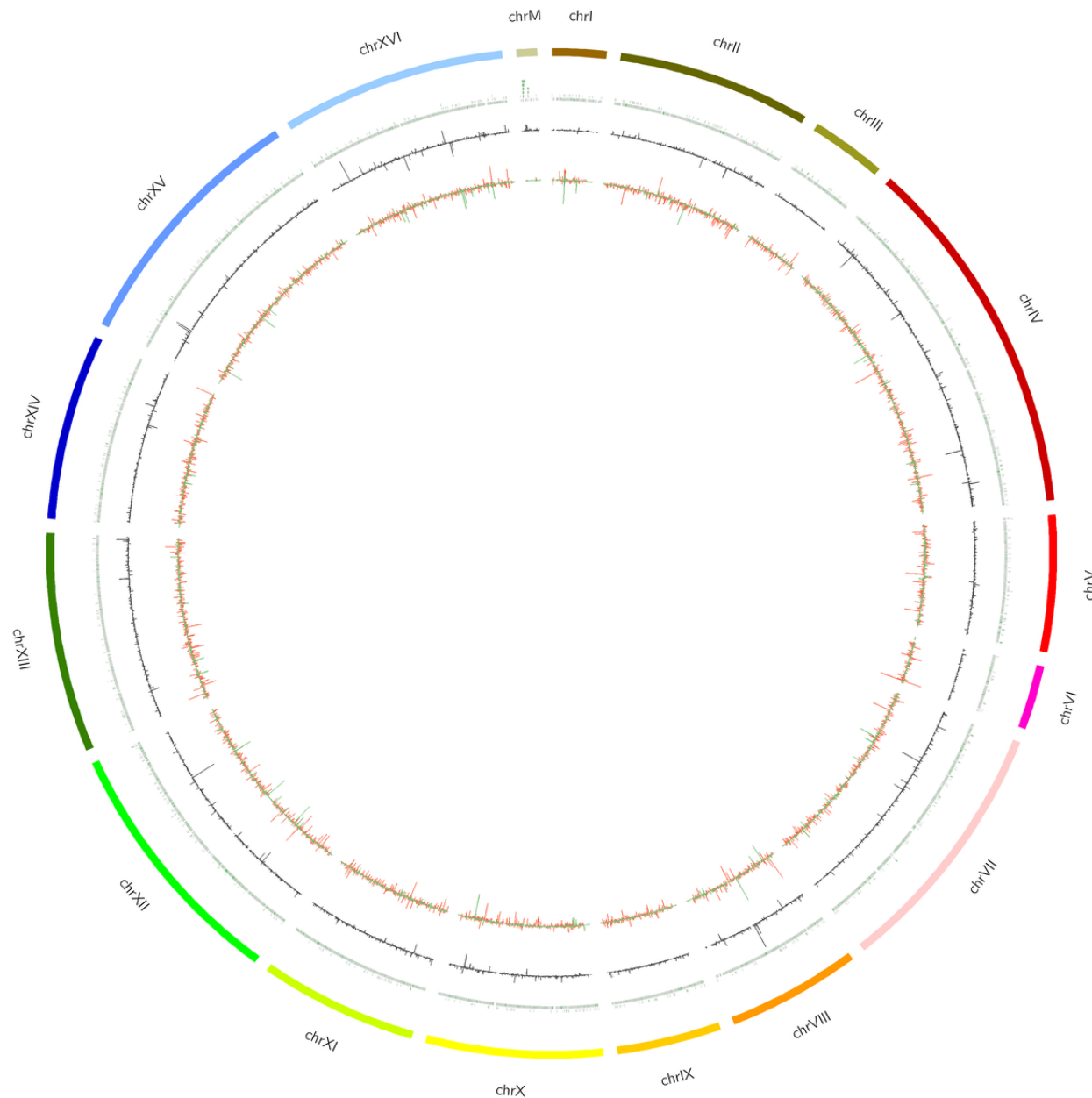
Circos Plot Applications Hands-on: Heatmap

- Tracks used to highlight genomic regions whose color is function of the value

Circos Plot Applications

Translational Efficiency (*Lindquist Lab*)

Circos Plots Summary

- .conf file(s) contains all the parameters needed for the display

- karyotype data required to draw the ideogram

- other data tracks (eg. genes, SNPs) must be specified in the conf file

- File formats:

Track/Data	Format
Ideogram	chr – id label start end color
Line/Heatmap	chr(id) start end value
Tile	chr(id) start end
Text (eg. label)	chr(id) start end label

More Information

- CIRCOS. http://circos.ca
  - ➢ Includes extended documentation and in-depth tutorials

- Krzywinski, M., et al. *Circos: An information aesthetic for comparative genomics* Genome Research 19:1639-1645 (2009)