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
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

Visualization Challenges

• Viewing multidimensional data
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

![Linear vs Circular Continuity and Focus Diagram]

Classical ideogram layout

Circos ideogram layout

circos.ca
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

[Image of .png and .svg icons]

http://commons.wikimedia.org/wiki
Creating Circos Plots: Pipeline

**Usage:**
circos -conf <configFile>
  eg. circos -conf circos.conf

**ADDITIONAL CONFIGURATION FILES**
ticks.conf  ideogram.conf

**CONFIGURATION FILE**
circos.conf

**TRACK DATA FILES**
histogram.txt scatter.txt heatmap.txt links1.txt links2.txt

**POST PROCESS**
- Illustrator
- Inkscape
- Photoshop
- Gimp
- ImageMagick
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

```xml
<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
```

Global to all plots

Specific to data.1.txt plot
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
Circos Plot Applications: Tiles

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

Tiles are placed in layer with smallest index that can accommodate 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).

<table>
<thead>
<tr>
<th>layer index</th>
<th>orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>in 6 out 5</td>
</tr>
<tr>
<td>2</td>
<td>5 3</td>
</tr>
<tr>
<td>3</td>
<td>4 1</td>
</tr>
<tr>
<td>4</td>
<td>3 2</td>
</tr>
<tr>
<td>5</td>
<td>2 4</td>
</tr>
<tr>
<td>6</td>
<td>1 6</td>
</tr>
</tbody>
</table>
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 (e.g., read count) connected by a single line

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

<table>
<thead>
<tr>
<th>Track/Data</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideogram</td>
<td>chr – id label start end color</td>
</tr>
<tr>
<td>Line/Heatmap</td>
<td>chr(id) start end value</td>
</tr>
<tr>
<td>Tile</td>
<td>chr(id) start end</td>
</tr>
<tr>
<td>Text (eg. label)</td>
<td>chr(id) start end label</td>
</tr>
</tbody>
</table>
More Information

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