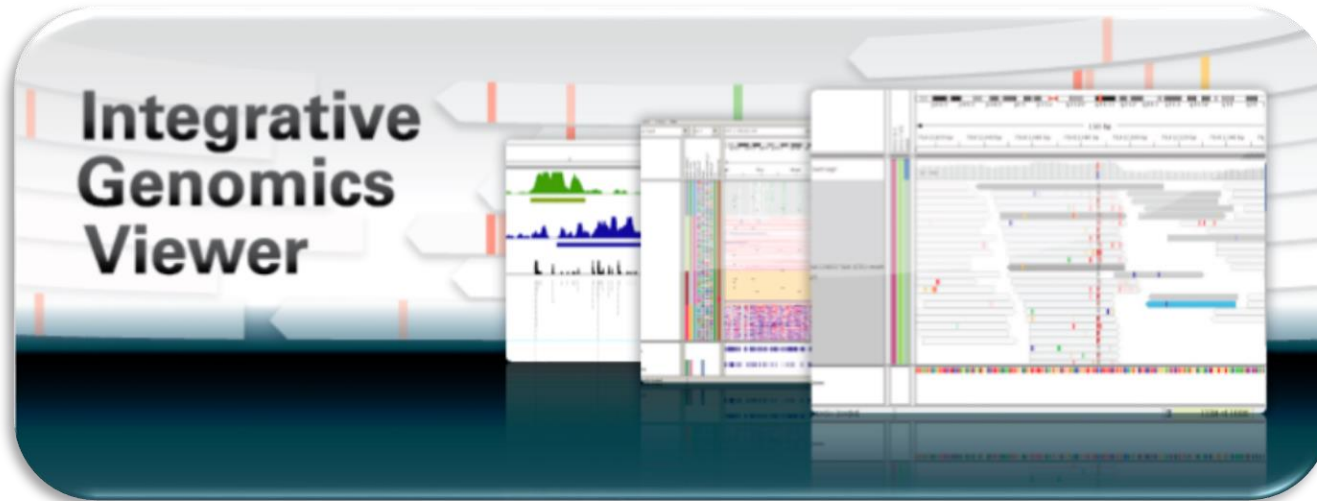# URPP tutorial

Genomic data visualization

Dr. Heidi E.L. Tschanz-Lischer
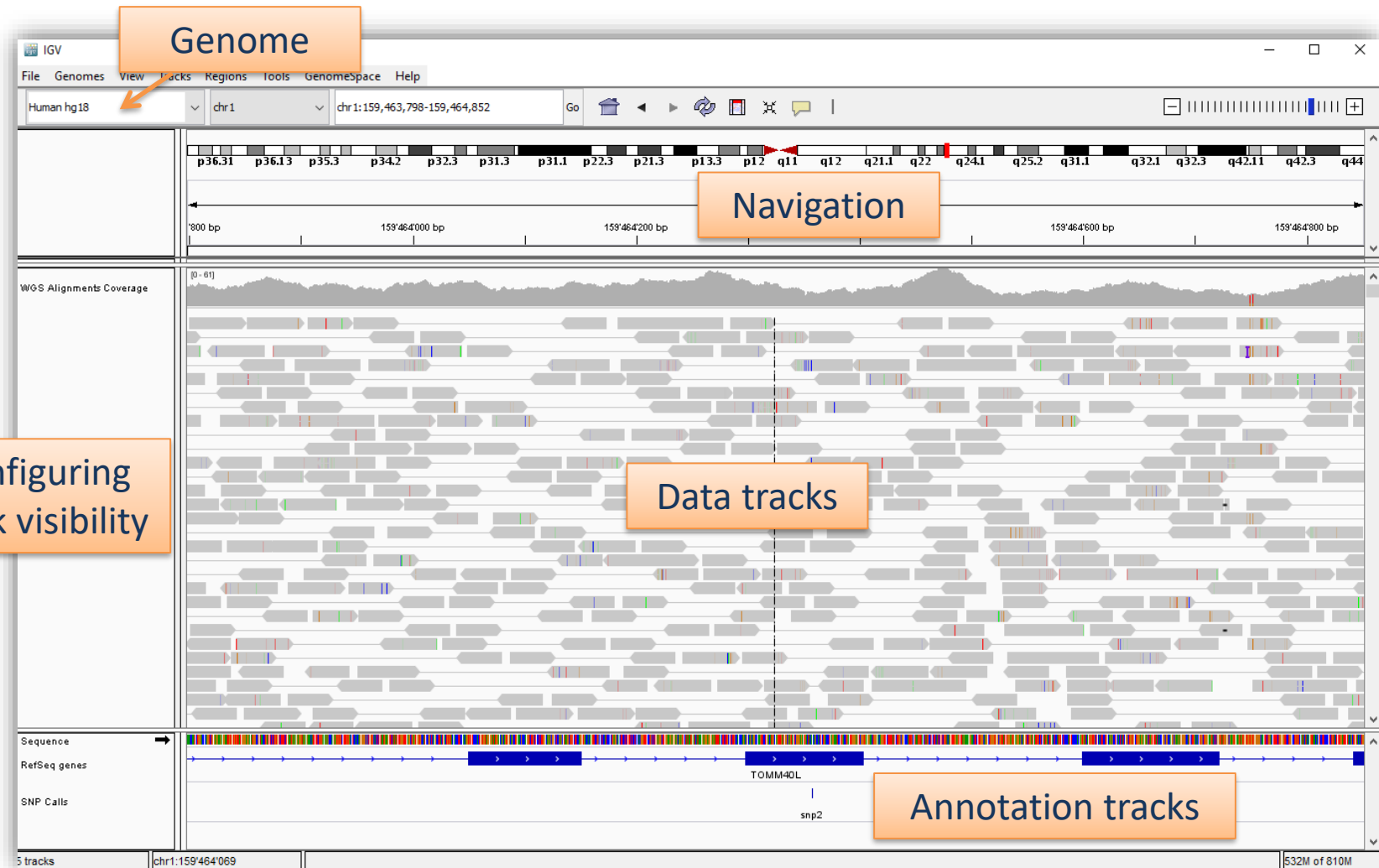**University of Zurich**
Switzerland

10 January, 2018

IGV

- high-performance visualization tool
- interactive exploration of large, integrated genomic datasets
- Runs "locally" (on your computer)

Robinson *et al*. 2011, Thorvaldsdóttir *et al*. 2013
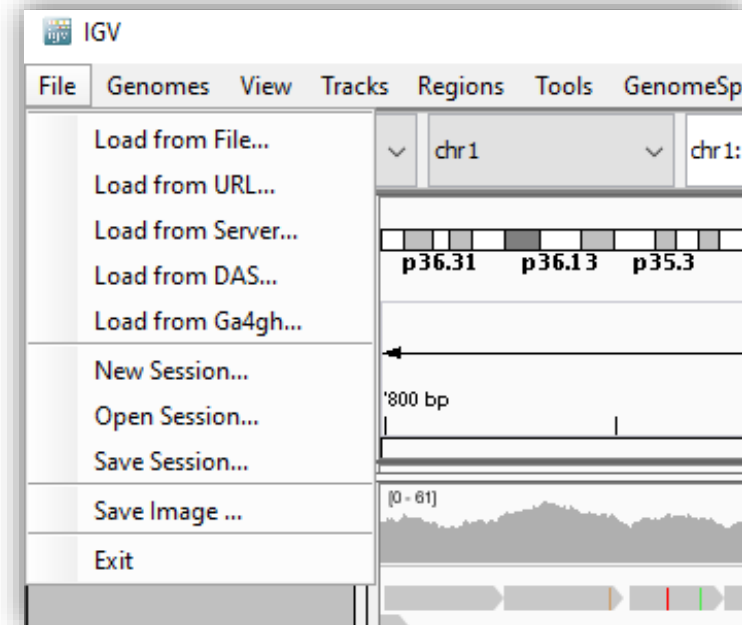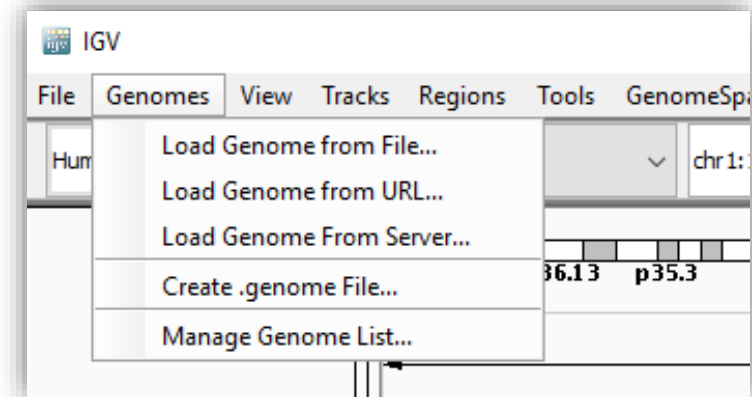
Basic view
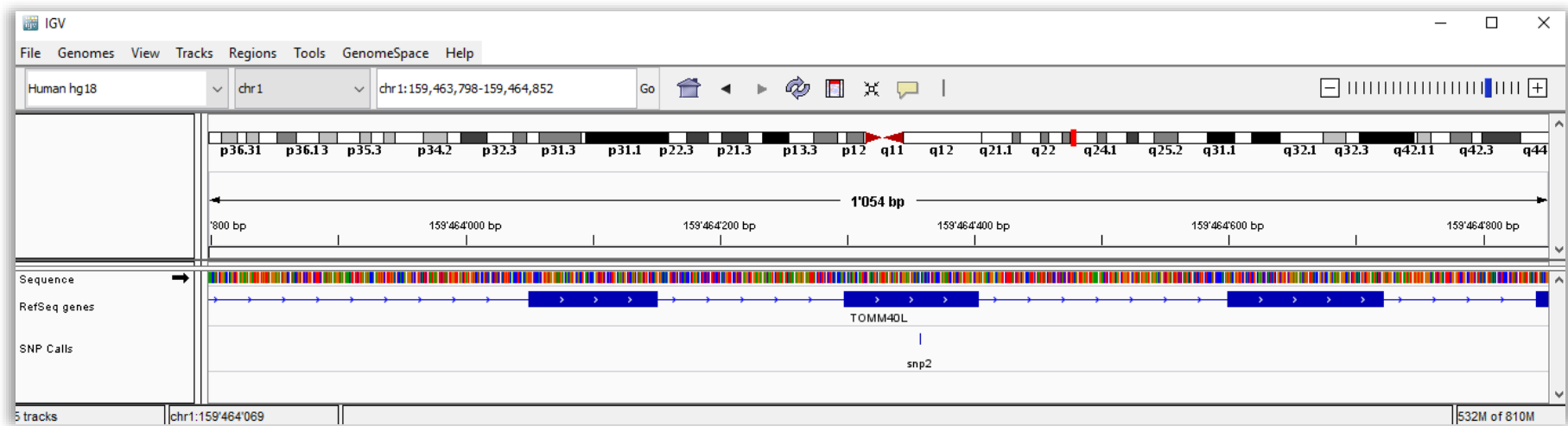
Load data

Genome

- Load from server
  → IGV provides a number of genomes
- Load from file:
  – FASTA file (not zipped)
  – index file (.fai) → samtools

Data

- Load from file:
  – BAM, SAM, BED, FASTA, GFF, GTF, GWAS, VCF, WIG and many more
- Load from server:
  – IGV data data (1000 genomes project)
- Load from URL, …

Track types

- Feature track:
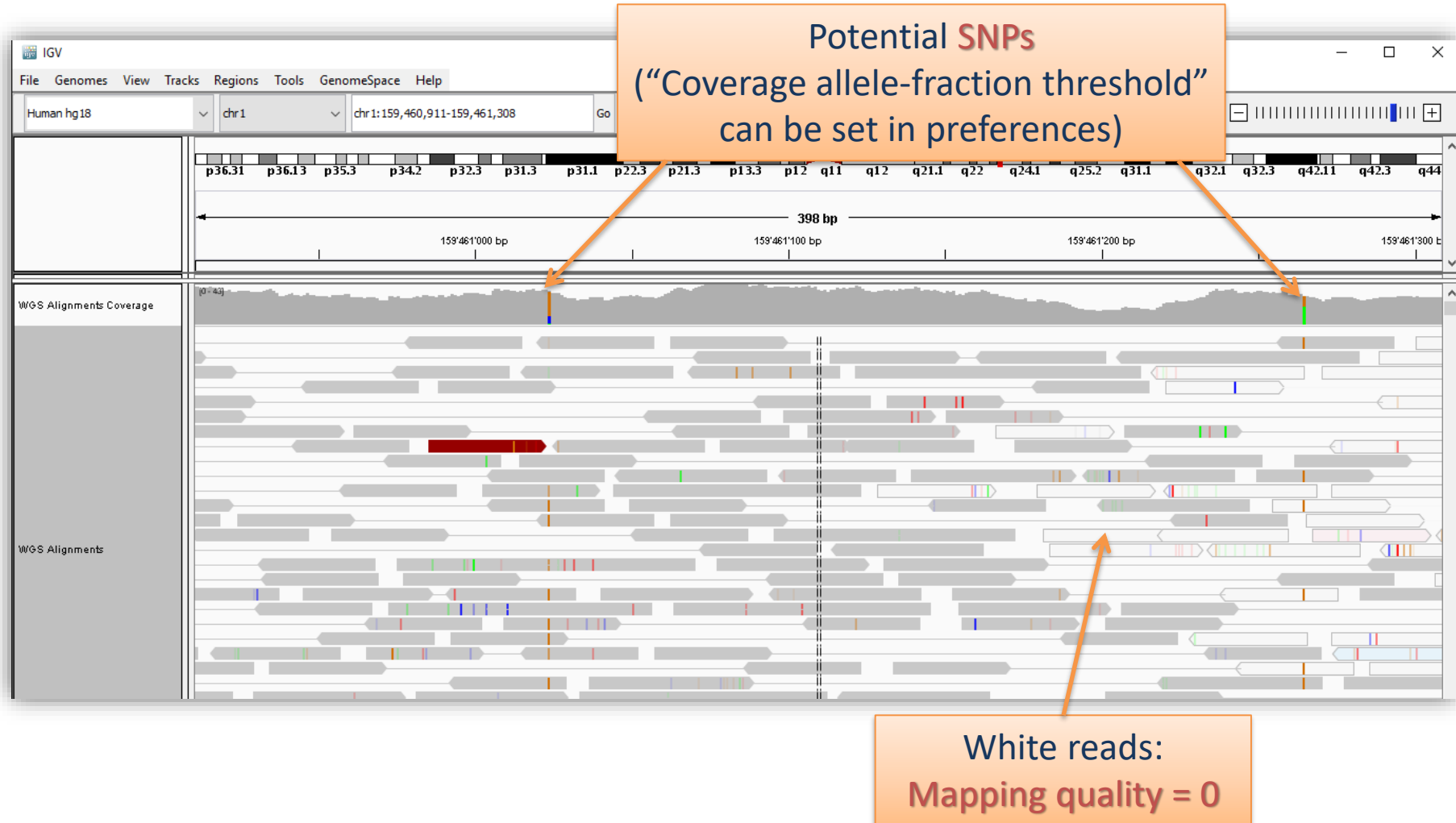 - Genomic features, e.g. gene annotations

Track types

- Alignment track:
  - Display alignments
    (with a lot of display options → right click on track)

Alignment track
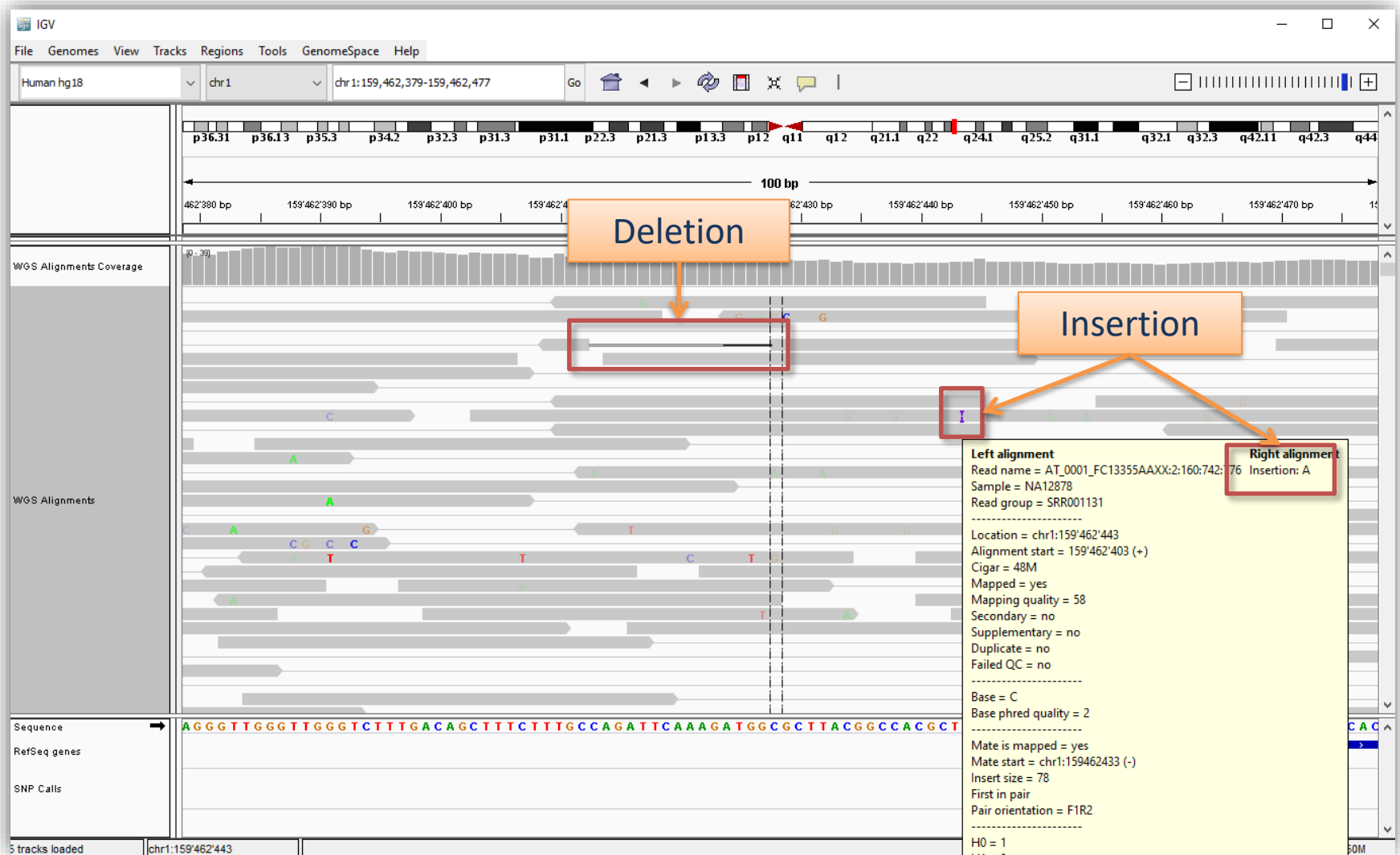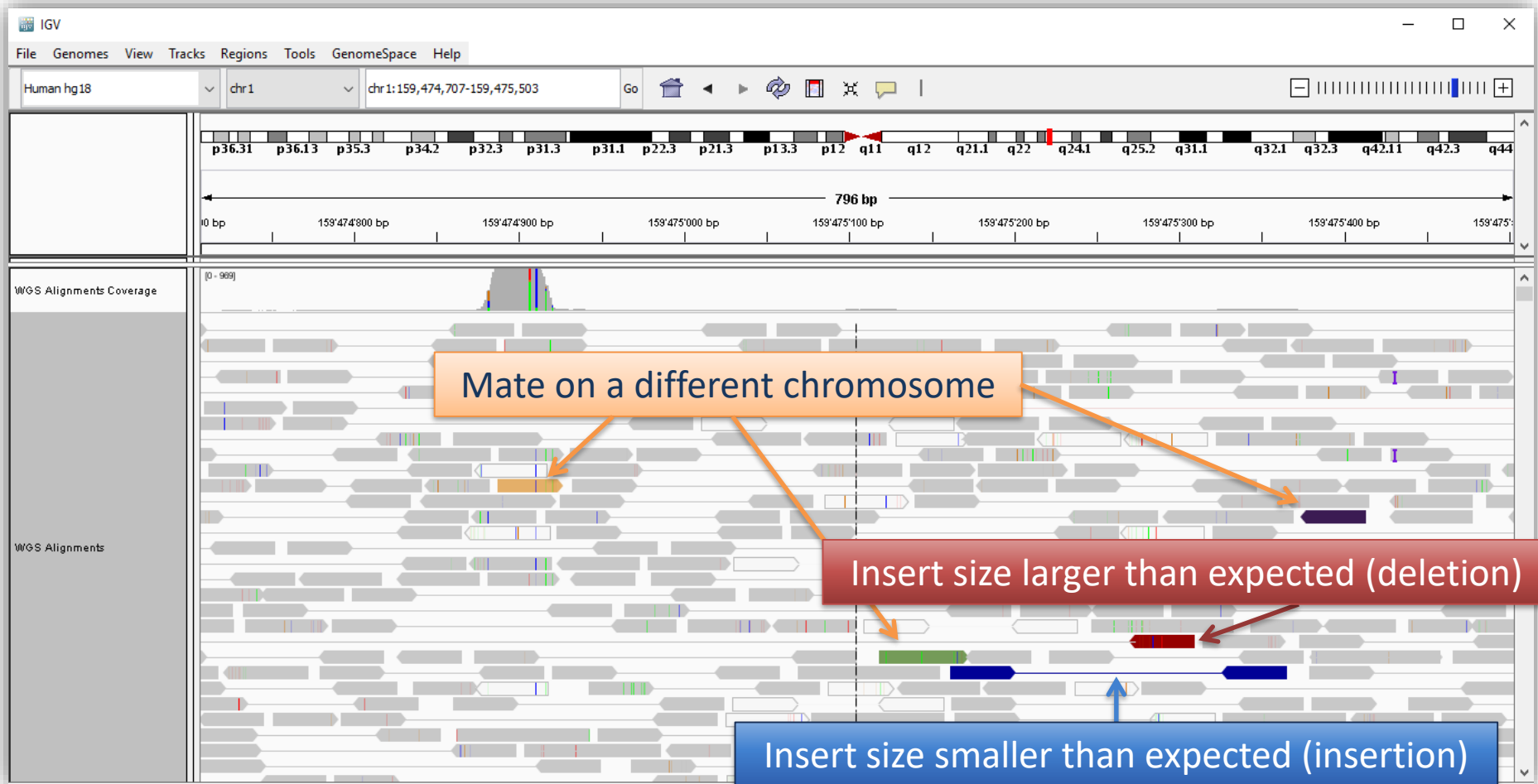## - SNPs

Potential SNPs
("Coverage allele-fraction threshold" can be set in preferences)

White reads:
Mapping quality = 0

Alignment track
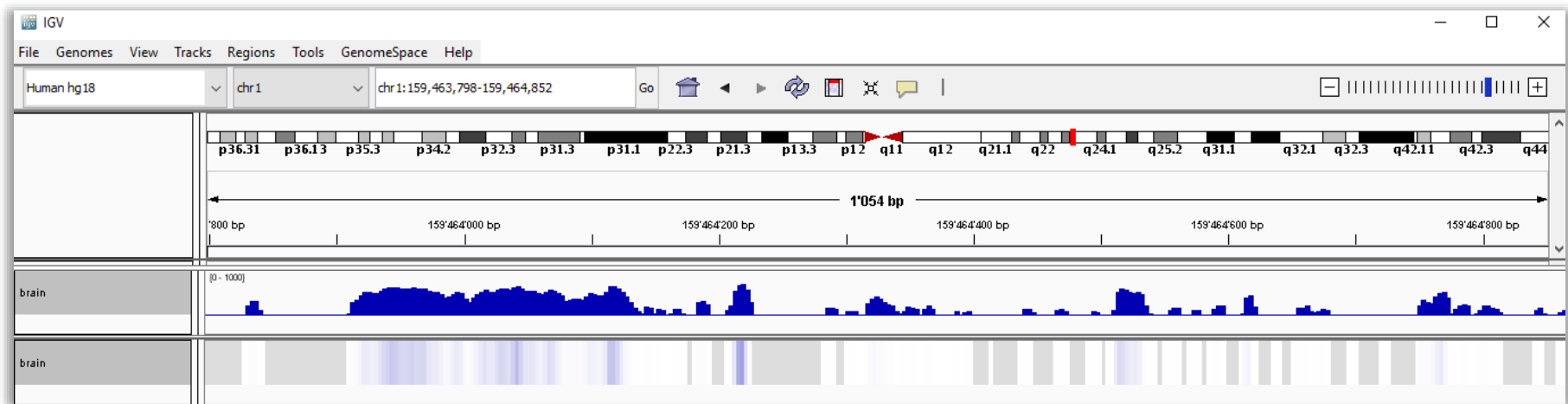## - insertions/deletions

Alignment track
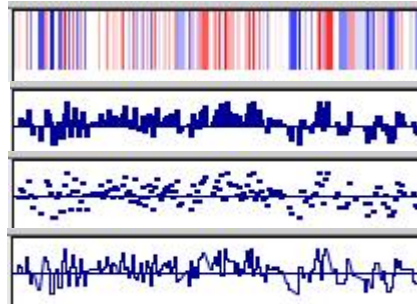## - structural variants

Chromosome color code:

Track types

- **Data track:**
  - Display numeric data, e.g. expression counts
  - Displayed as (choose type → right click on track)
    - Heatmap
    - Bar chart
    - Points (scatter plot)
    - or line plot

View settings

A lot of display settings:

- Menu View → **Preferences**
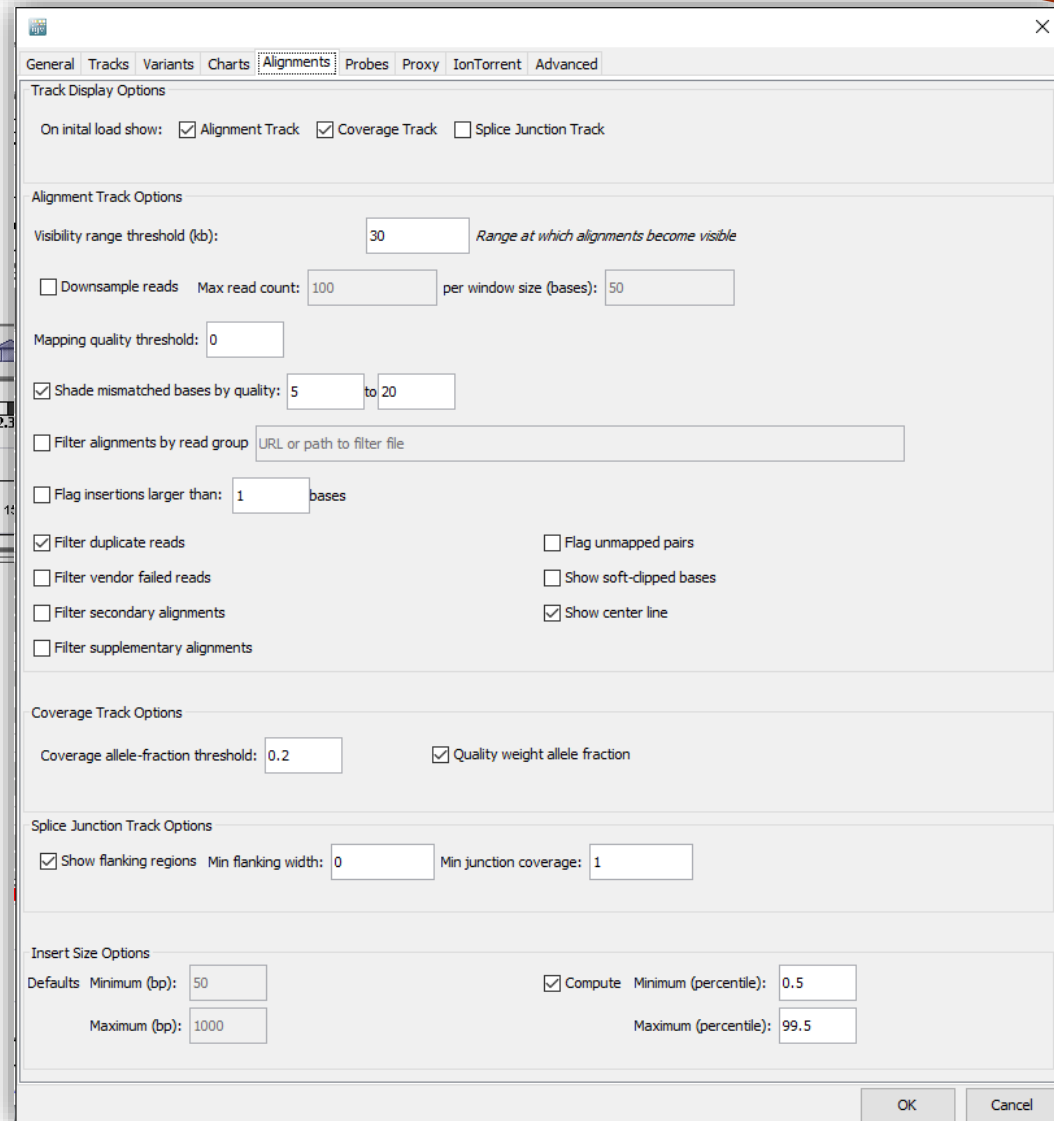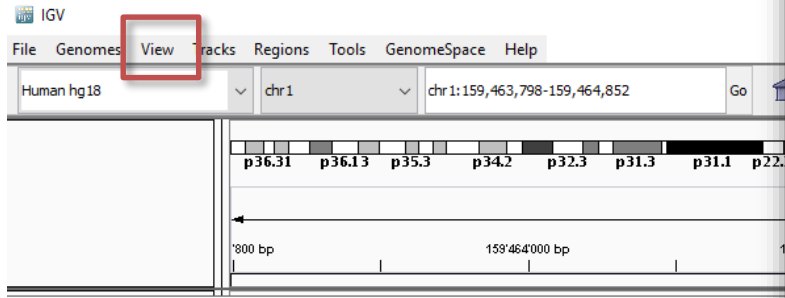
Splice junctions

- Visualize splice junctions: splice junctions track
  Menu View → Preferences → Alignment → tick "Splice Junction Track" box

- Right click on an alignment track → Sashimi Plot

View multiple regions at once

Type a few genes or locations

Batch file

- Users can load a text file to execute series of sequential tasks
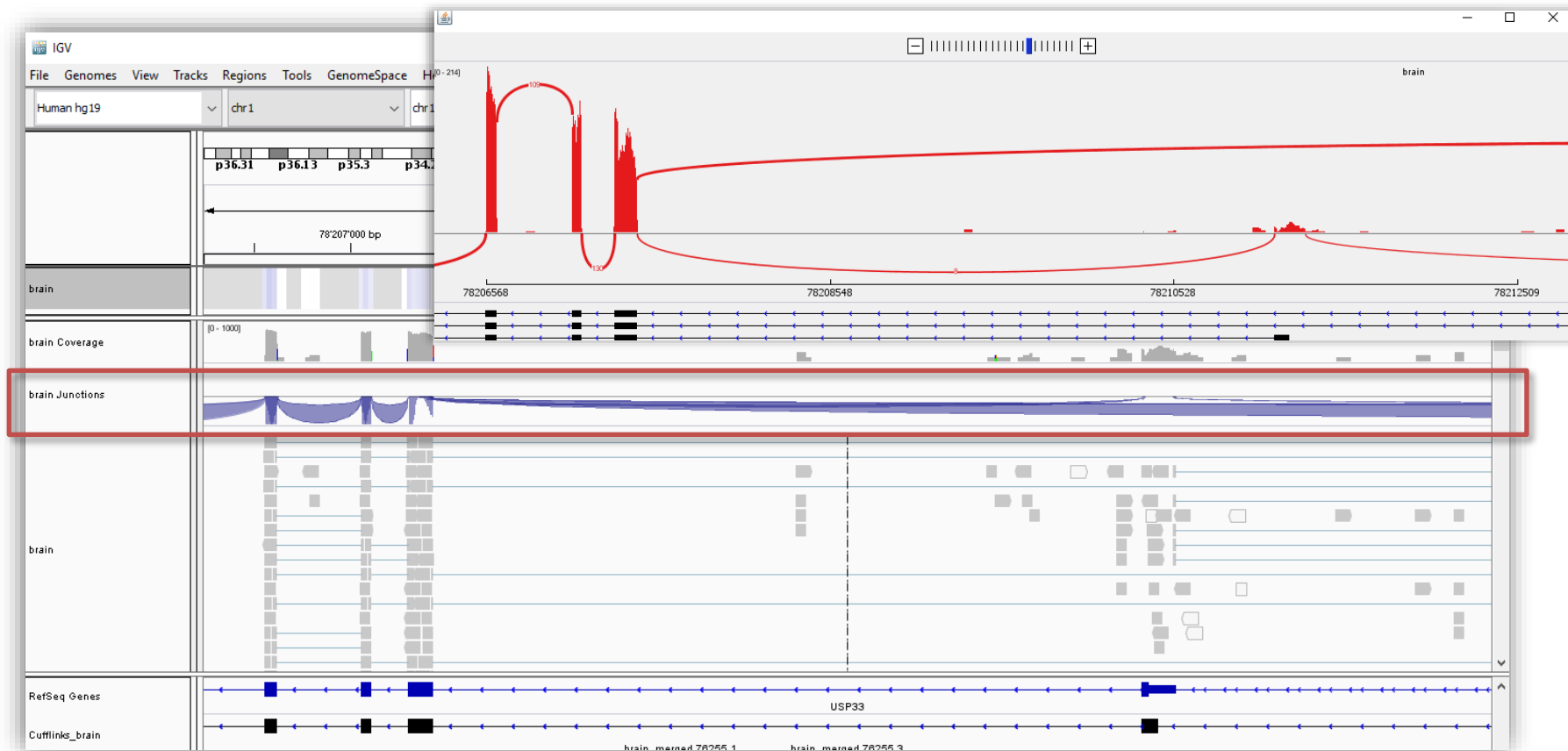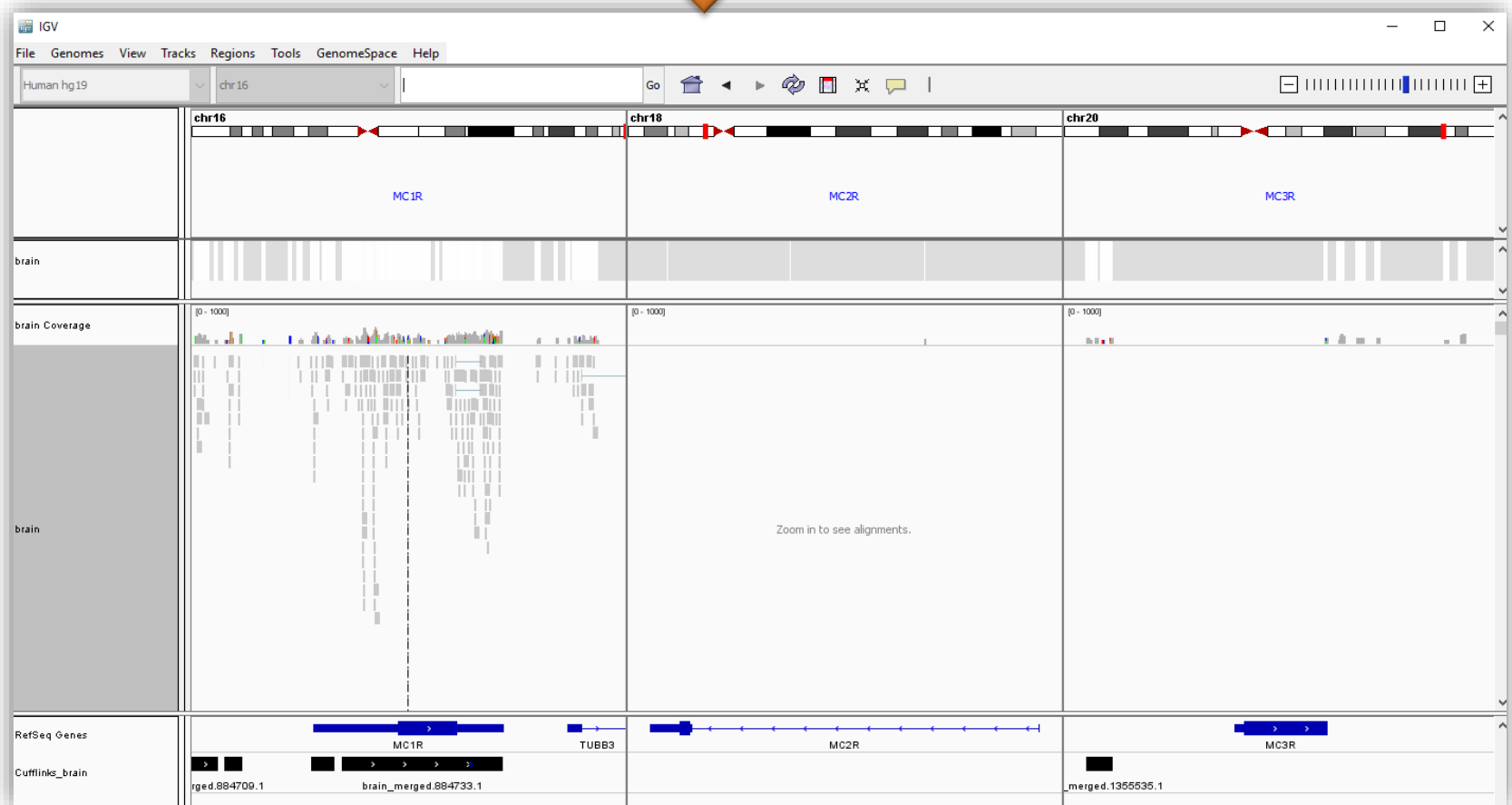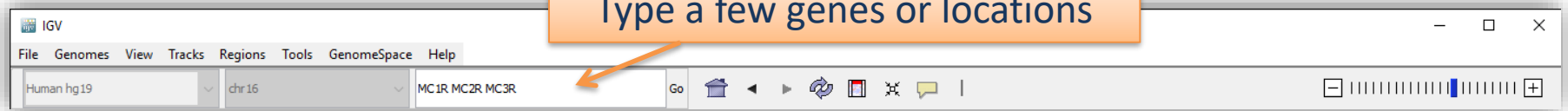 → Menu Tools → Run Batch Script

- Own simple scripting language (18 commands)
  → see http://www.broadinstitute.org/software/igv/PortCommands

```
# Example script                          → Comment (ignored)
genome hg18                               → Select genome
load  myfile.bam                          → Load alignment file
snapshotDirectory myDirectory             → Set snapshot directory
goto chr1:65,289,335-65,309,335           → Jump to specific locus
sort position
collapse
snapshot
goto chr1:113,144,120-113,164,120
sort base
collapse
snapshot
```

Sort by position, collapse all tracks and take a snapshot of the screen

Go to next locus, sort by base, collapse all tracks and take again a snapshot of the screen

Circos

- A software package for visualizing data in circular layout
  - → Ideal for imaging relationship between positional data
  - → Combines traditional 2D plot types with links to positions
- Plain text configuration
- GFF data input
- Very flexible
- Can be automated
- Requires Perl
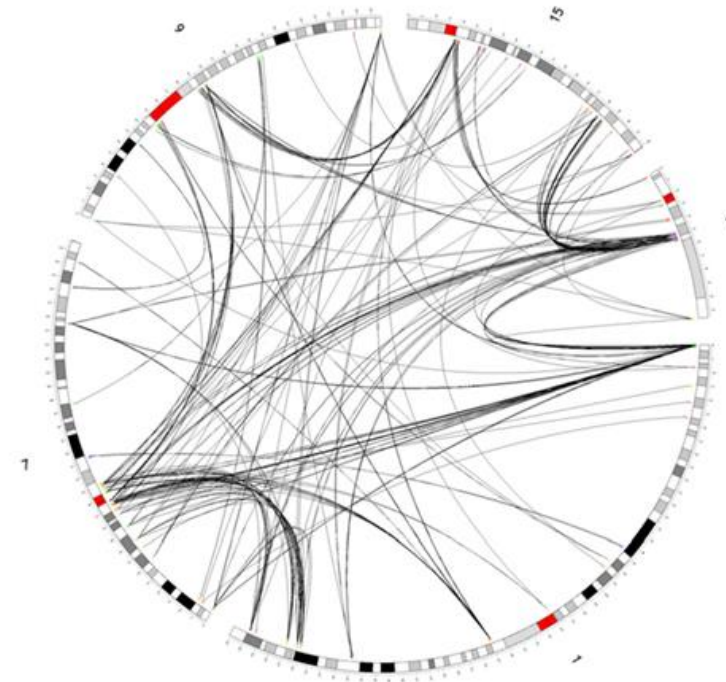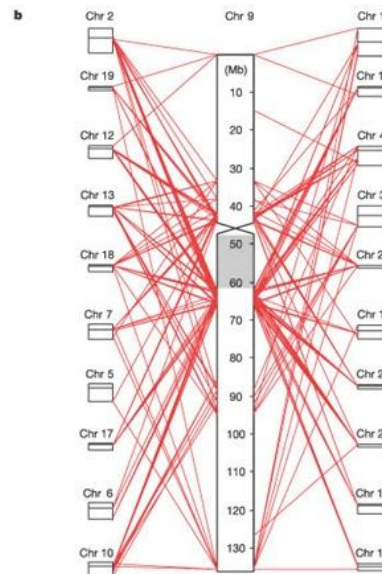
Humphray, S. J., K. Oliver, et al. (2004).
"DNA sequence and analysis of human chromosome 9."
Nature 429(6990): 369-74.

Circos image
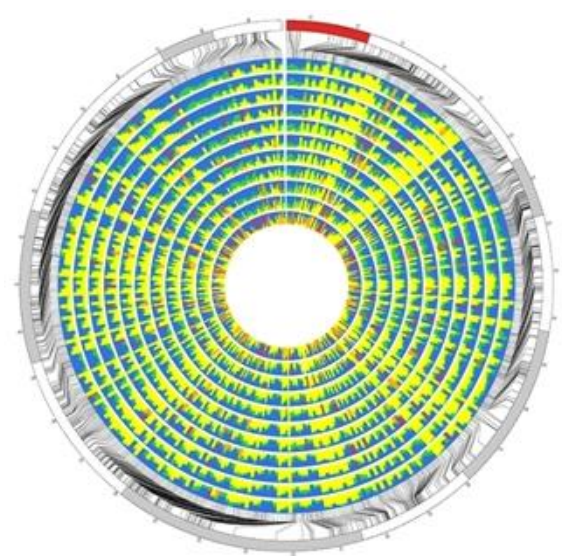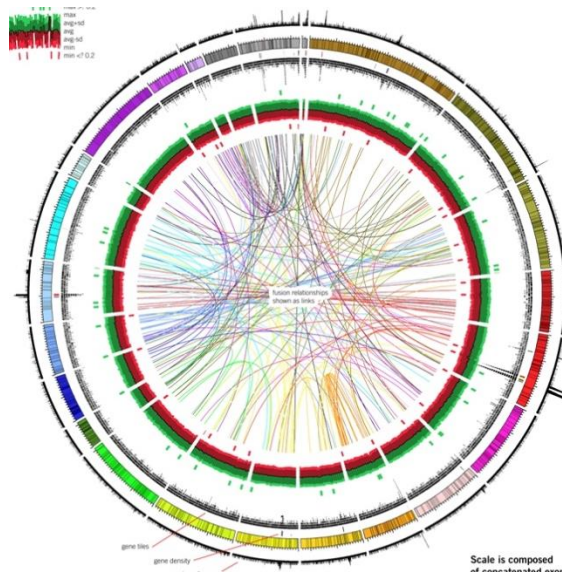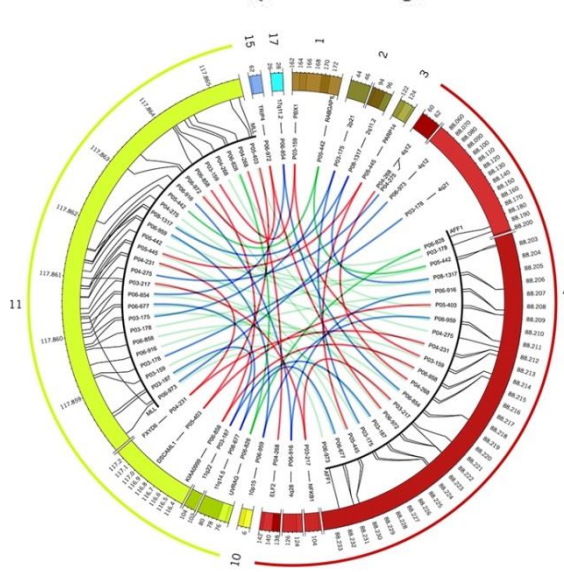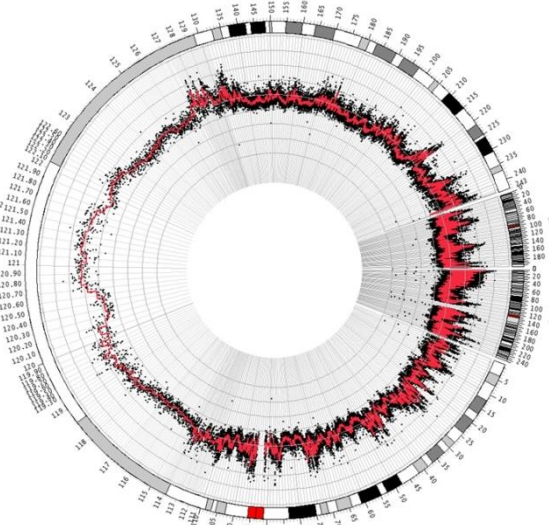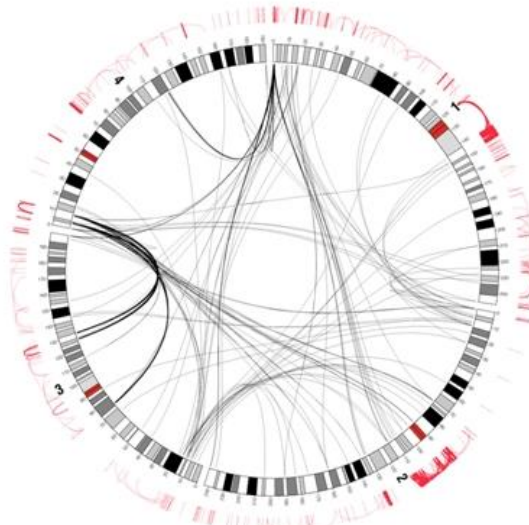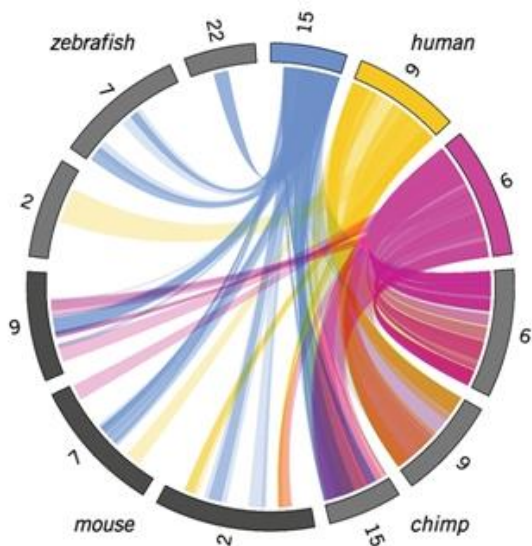
Circos

- Run circos:

```
./circos -conf circos.conf -outputdir . -outputfile image.png
```

Configuration file
with graphic
information

Output folder

Graphic file name

Data track types:
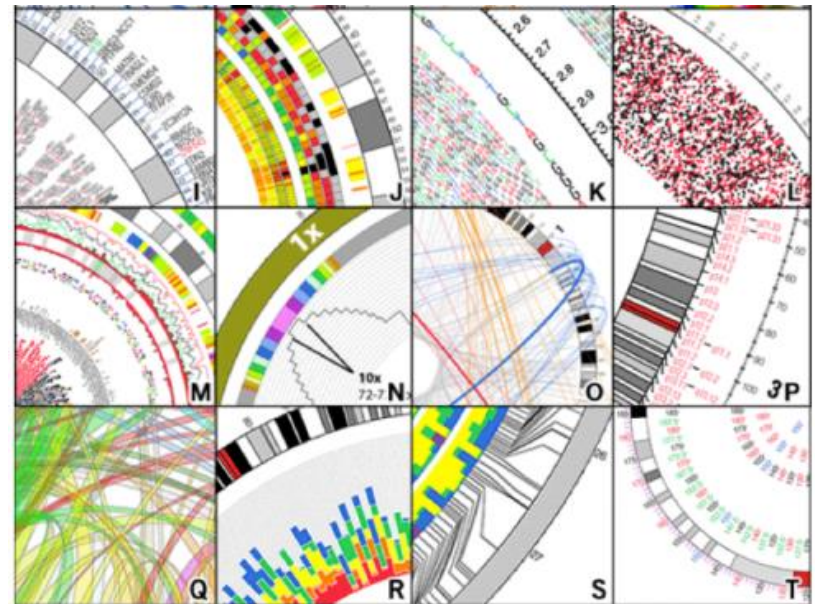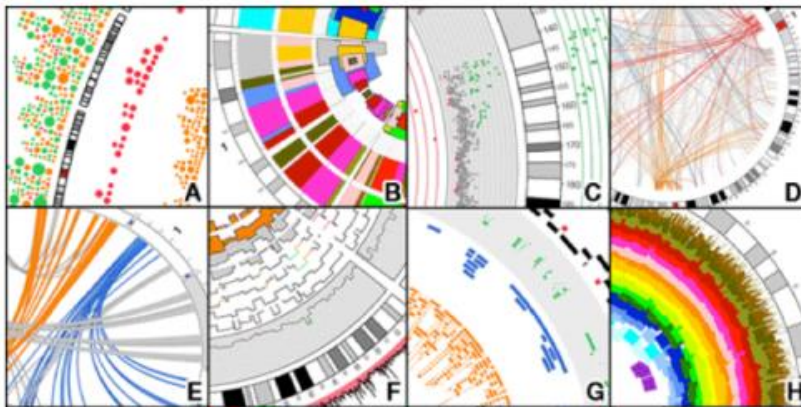
Configuration file
- minimum configuration file

```
karyotype = data/karyotype/karyotype.human.txt
<ideogram>
  <spacing>
    default = 0.005r
  </spacing>
  radius     = 0.9r
  thickness = 20p
  fill       = yes
  stroke_thickness=0p
</ideogram>


########################           #
# The remaining content is standard and required.
# Included from Circos distribution.
<image>
  <<include etc/image.conf>>
</image>
# RGB/HSV color definitions, color lists, fonts, fill patterns
<<include etc/colors_fonts_patterns.conf>>
# Debugging, I/O and other system parameters
<<include etc/housekeeping.conf>>
```

Chromosome name, size and color definition

Spacing between ideograms (r=relative; 0.5% of circumference)

Radial position of ideogram (r=relative)

Thickness of ideogram (p=pixels)

Draw ideogram filled

Ideogram without outline
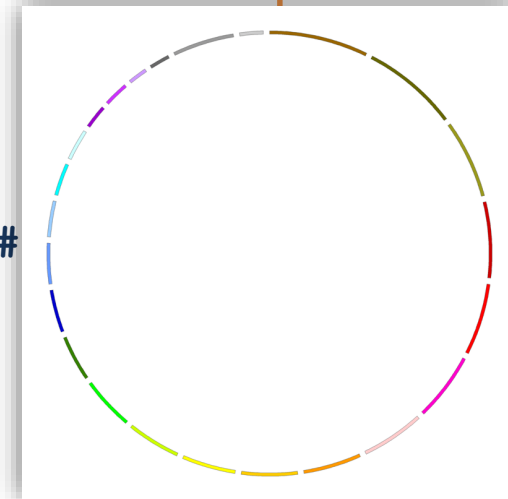
circos.conf

Configuration file
- labels

```
...
<ideogram>
  <spacing>
    default = 0.005r
  </spacing>
  radius          = 0.90r
  thickness       = 20p
  fill            = yes
  stroke_color    = dgrey
  stroke_thickness = 2p

  show_label      = yes
  label_font      = default
  # using image dimensions
  label_radius    = dims(image,radius)-60p
  label_size      = 30
  label_parallel  = yes
</ideogram>
...
```

Labels for ideograms can be placed at any radial position and formatted flexibly
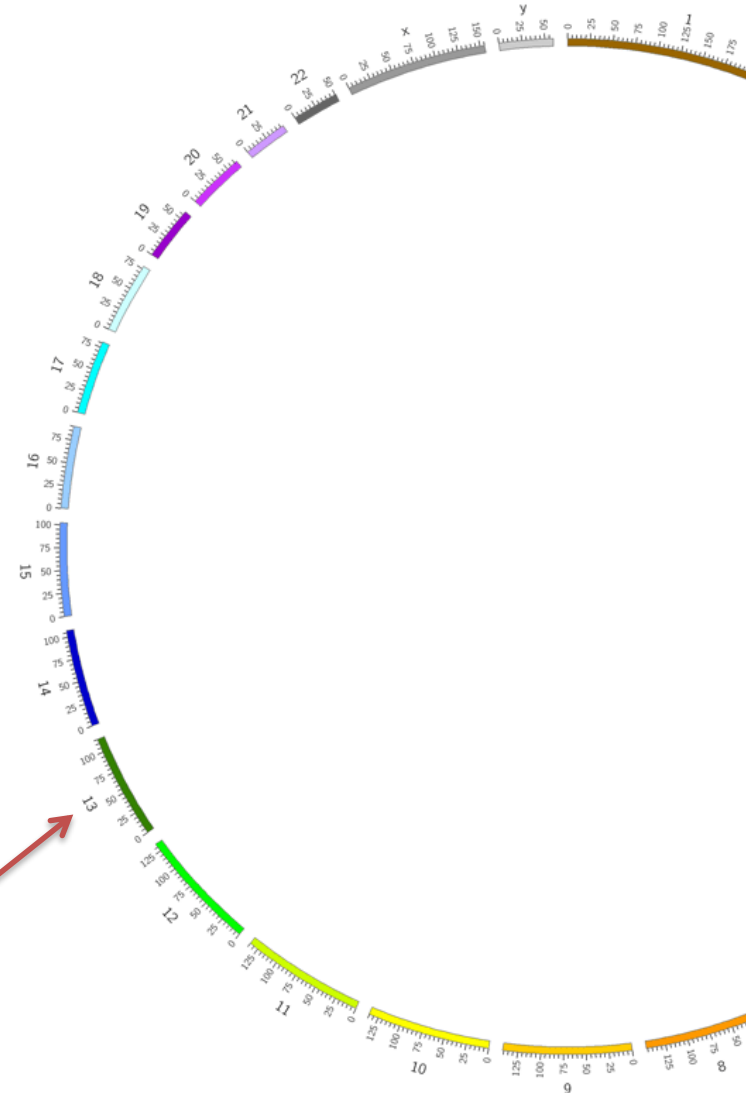
Configuration file
- ticks

```
...
chromosomes_units = 1000000          ← Define units
show_ticks        = yes
show_tick_labels  = yes
<ticks>
  radius          = 1r
  color           = black
  thickness       = 2p
  multiplier      = 1e-6             ← Tick label = tick position * multiplier
  format          = %d              ← Tick format (%d = integer)
  <tick>
    spacing       = 5u
    size          = 10p             ← "Small" ticks
  </tick>
  <tick>
    spacing       = 25u
    size          = 15p
    show_label    = yes
    label_size    = 20p             ← "Large" ticks
    label_offset  = 10p
    format        = %d
  </tick>
</ticks>
...
```
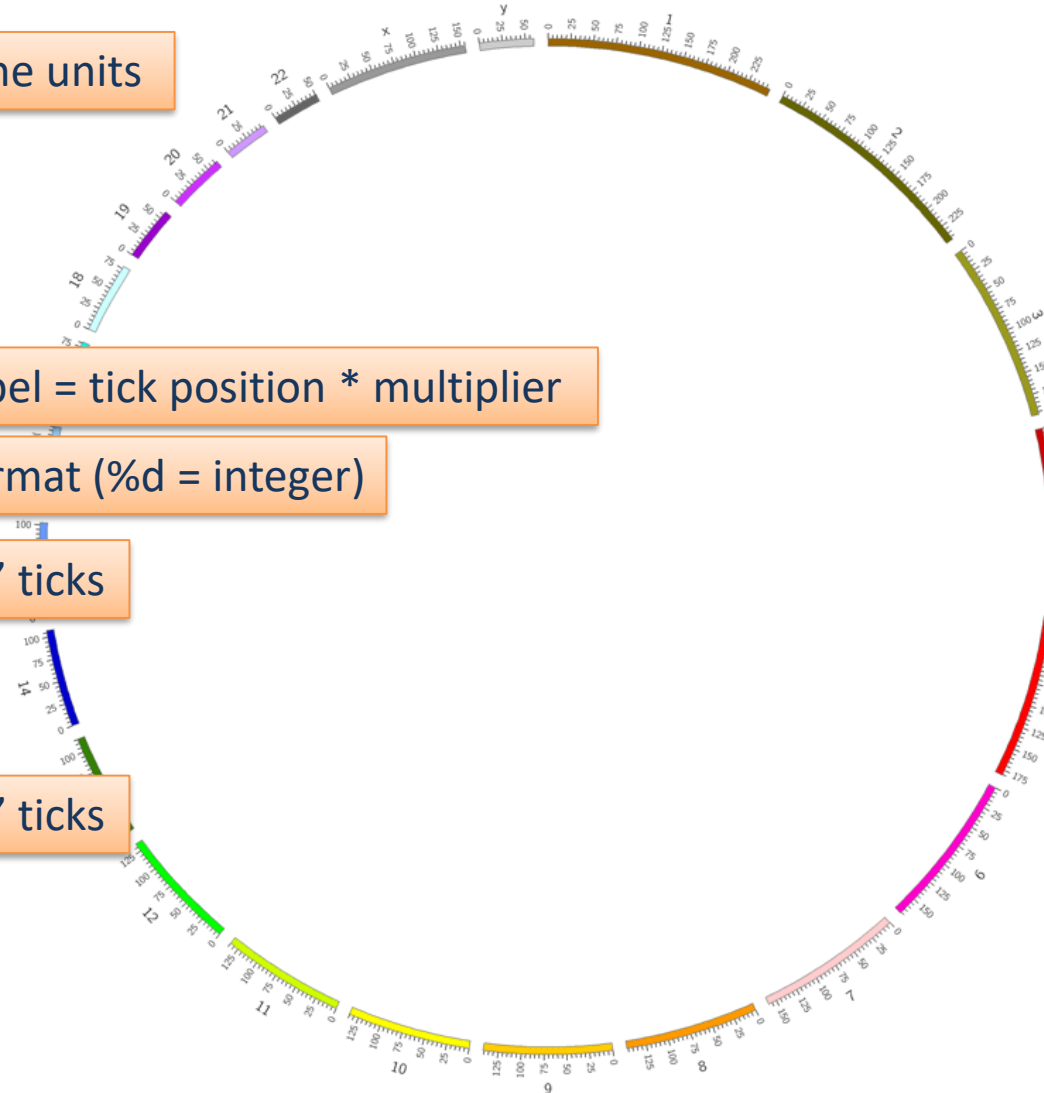
Configuration file
- ideogram selection, scale, color & orientation

```
karyotype = data/karyotype/karyotype.human.txt
chromosomes_units = 1000000
chromosomes_display_default = no
chromosomes        = hs1;hs2;hs3;h4

chromosomes_scale = hs1=0.5r,/hs[234]/=0.5rn

chromosomes_reverse = /hs[234]/

# change chromosome color
chromosomes_color = hs1=red,hs2=orange,hs3=green,hs4=blue

# change radial position of one chromosome
chromosomes_radius  = hs4:0.9r
...
```

Subset of chromosomes

hs1: 50% of figure
hs2-4: 50% of figure
(n=each evenly distributed)

Reverse chr hs2-4

Configuration file
- links and rules

```
<links>
  <link>
    file          = data/5/segdup.txt
    radius        = 0.8r
    bezier_radius = 0r
    color         = black_a4
    thickness     = 2
# Rule blocks define how data points are formatted
    <rules>
      <rule>
        condition     = var(intrachr)
        show          = no
      </rule>
      <rule>
        condition     = 1
        color         = eval(var(chr2))
        flow          = continue
      </rule>
      <rule>
        condition     = from(hs1)
        radius1       = 0.99r
      </rule>
      <rule>
        condition     = to(hs1)
        radius2       = 0.99r
      </rule>
    </rules>
  </link>
</links>
```
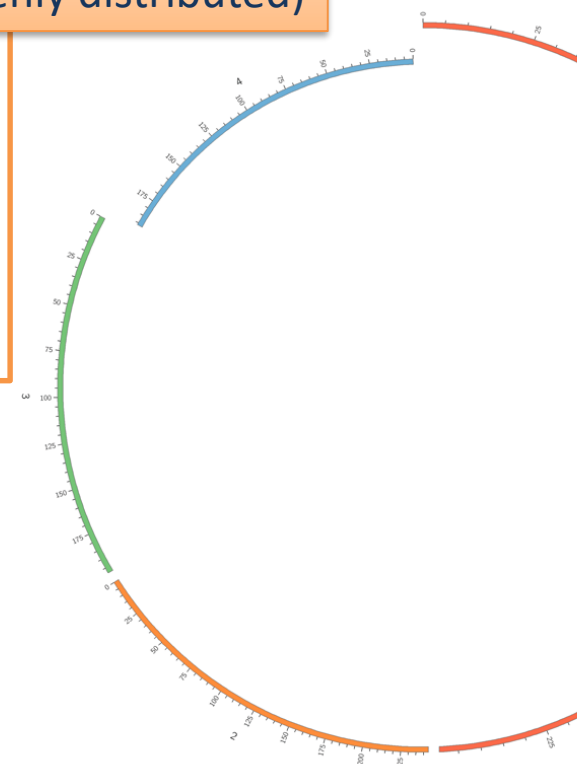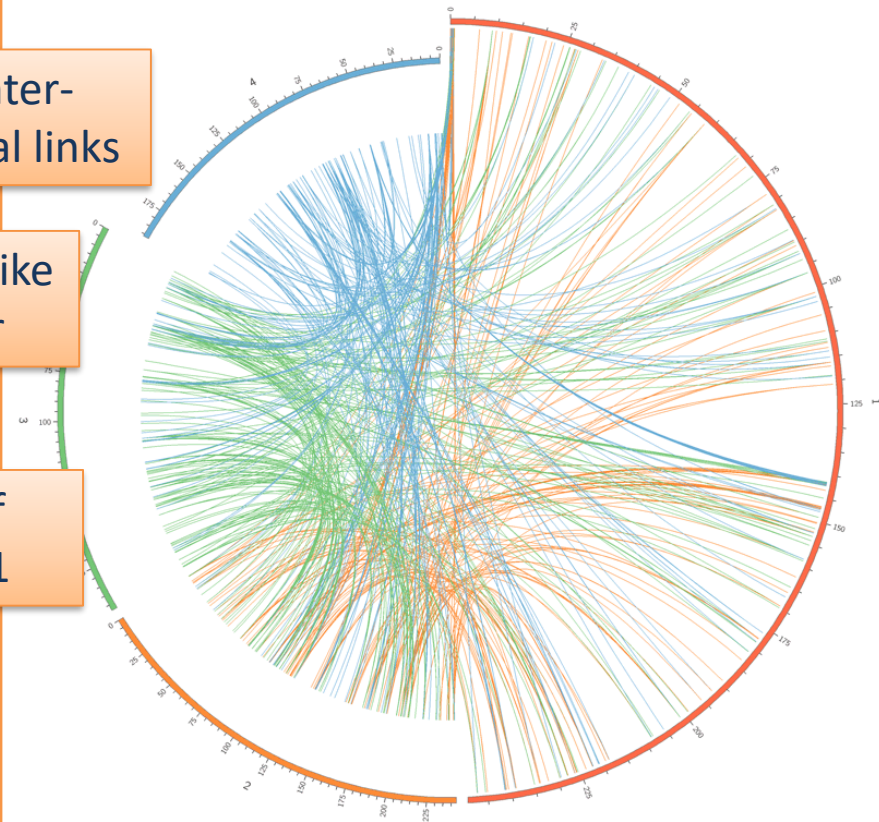
File with links

Radial position of start

curvature of the line

Show only inter-chromosomal links

Color link like second chr

Change radius of links from/to hs1

Configuration file
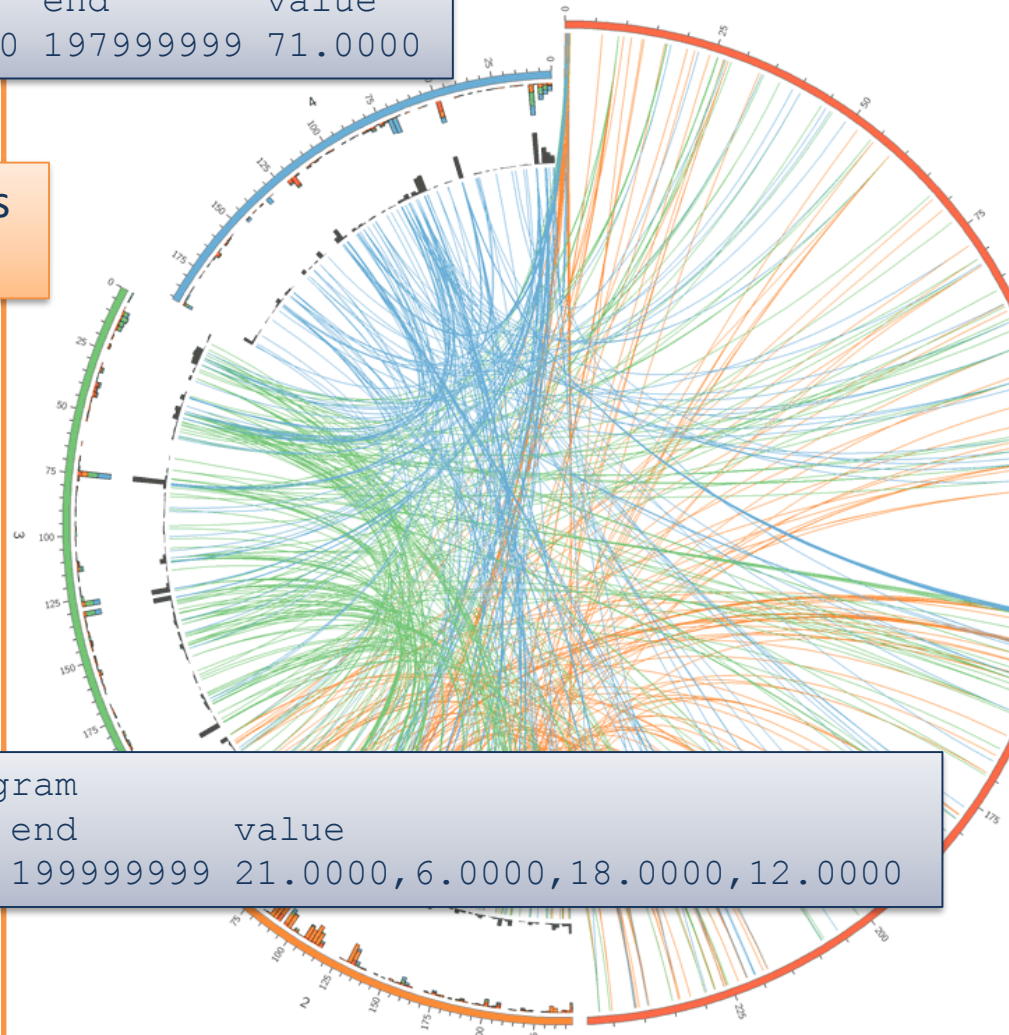## - Histograms

```
<plots>
  <plot>
    type = histogram
    file = data/5/segdup.hs1234.hist.txt
    r1    = 0.88r
    r0    = 0.81r
    fill_color = vdgrey
    extend_bin = no
    <rule>
      condition = on(hs1)
      show      = no
    </rule>
  </plot>
  <plot>
    type = histogram
    file = data/5/segdup.hs1234.stacked.txt
    r1    = 0.99r
    r0    = 0.92r
    fill_color  = hs1,hs2,hs3,hs4
    orientation = in
    extend_bin  = no
    <rules>
      condition = on(hs1)
      show      = no
    </rule>
  </plot>
</plots>
```

```
#chr start       end        value
hs3  196000000 197999999 71.0000
```

inner/outer radius of track

```
#stacked histogram
#chr start       end        value
hs3  198000000 199999999 21.0000,6.0000,18.0000,12.0000
```

Configuration file
- heat maps

```
<plots>
  <plot>
    type  = heatmap
    file  = data/5/segdup.hs1234.heatmap.txt
    r1    = 0.89r
    r0    = 0.88r

    color = hs1_a5,hs1_a4,hs1_a3,hs1_a2,hs1_a1,hs1
    scale_log_base = 0.25
    <rule>
      condition = on(hs1)
      show      = no
    </rule>
  </plot>
  <plot>
...
</plots>
```
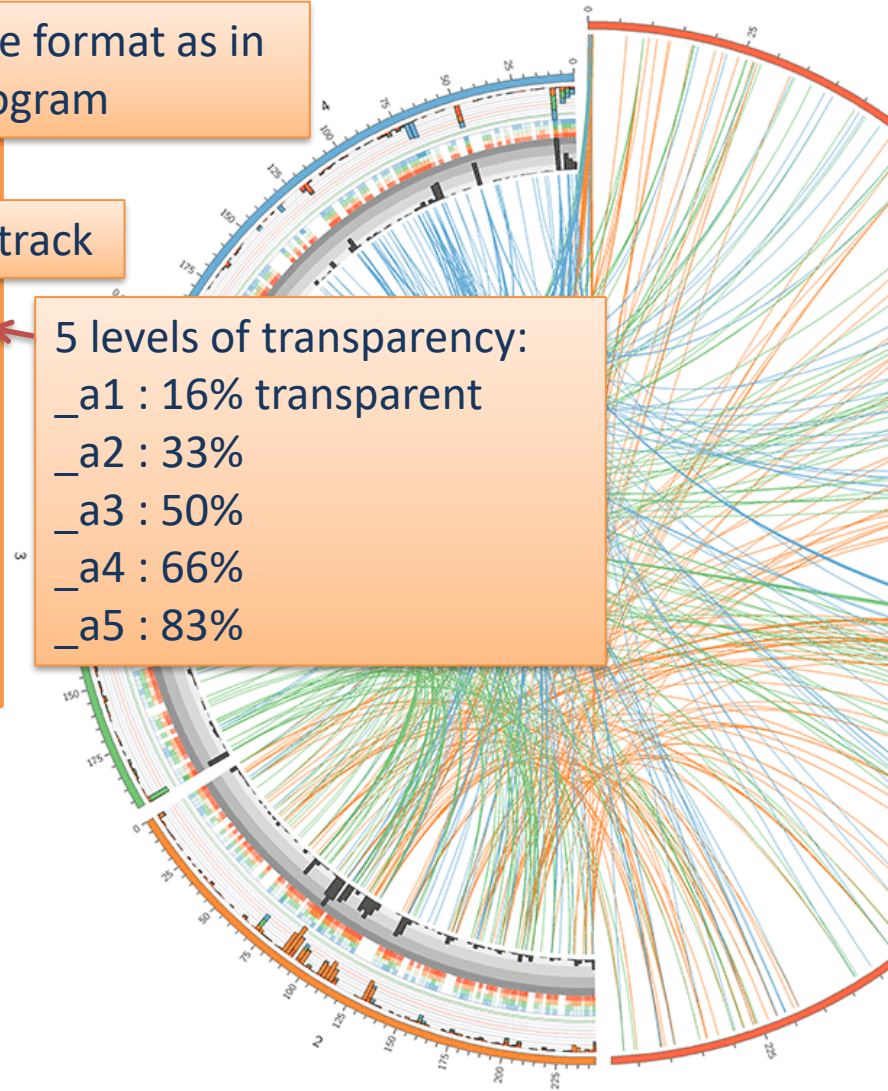
Same format as in histogram

inner/outer radius of track

Non-linear scaling

5 levels of transparency:
_a1 : 16% transparent
_a2 : 33%
_a3 : 50%
_a4 : 66%
_a5 : 83%

Configuration file
- text
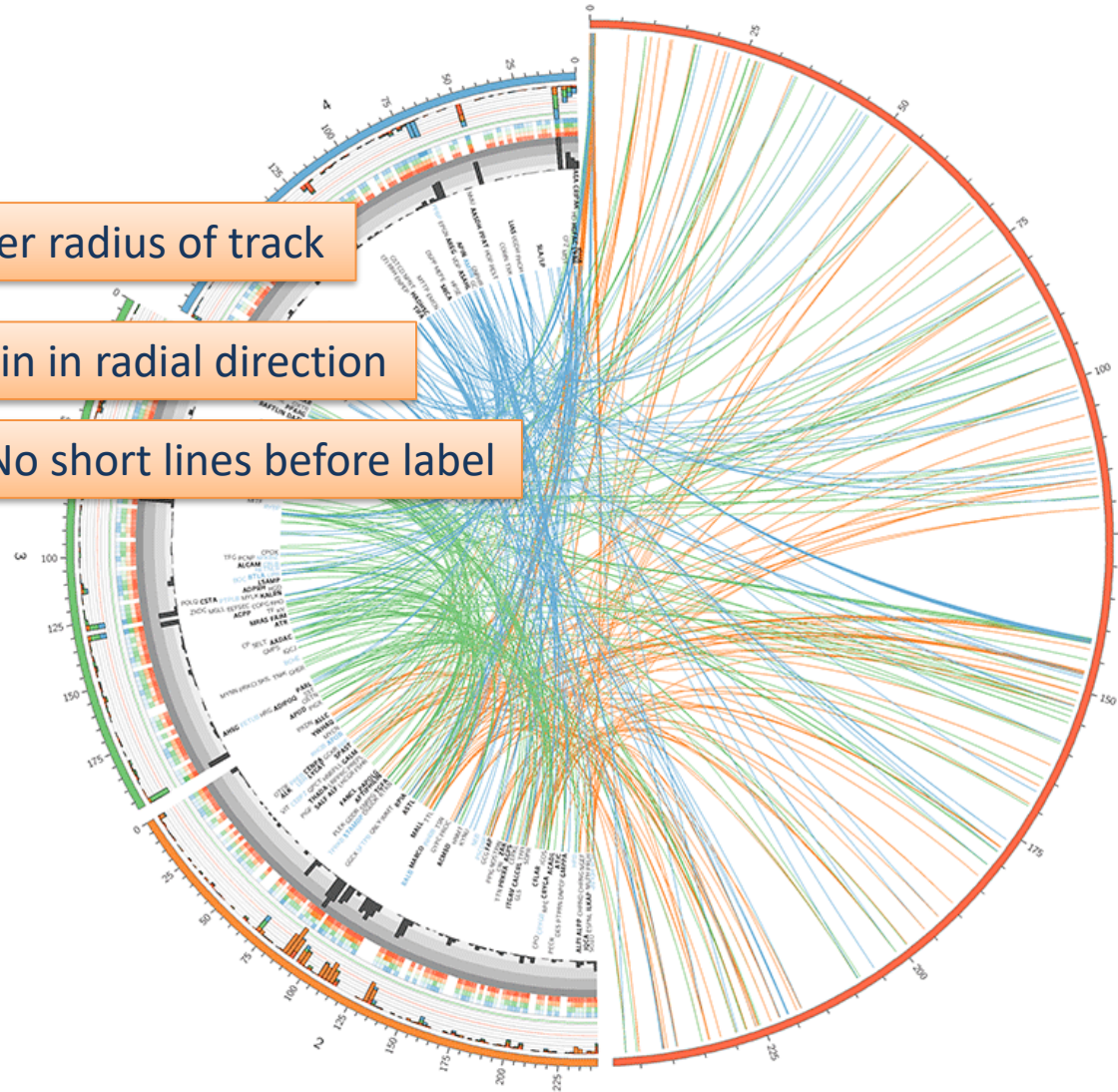
```
<plots>
  <plot>
    type  = text
    file  = data/6/genes.labels.txt
    r1    = 0.8r
    r0    = 0.6r
    label_font = light
    label_size = 12p
    rpadding   = 5p

    show_links      = no
    link_dims       = 0p,2p,5p,2p,2p
    link_thickness = 2p
    link_color      = black
    <rules>
      <rule>
        condition = on(hs1)
        show      = no
      </rule>
      ...
  </plot>
...
</plots>
```
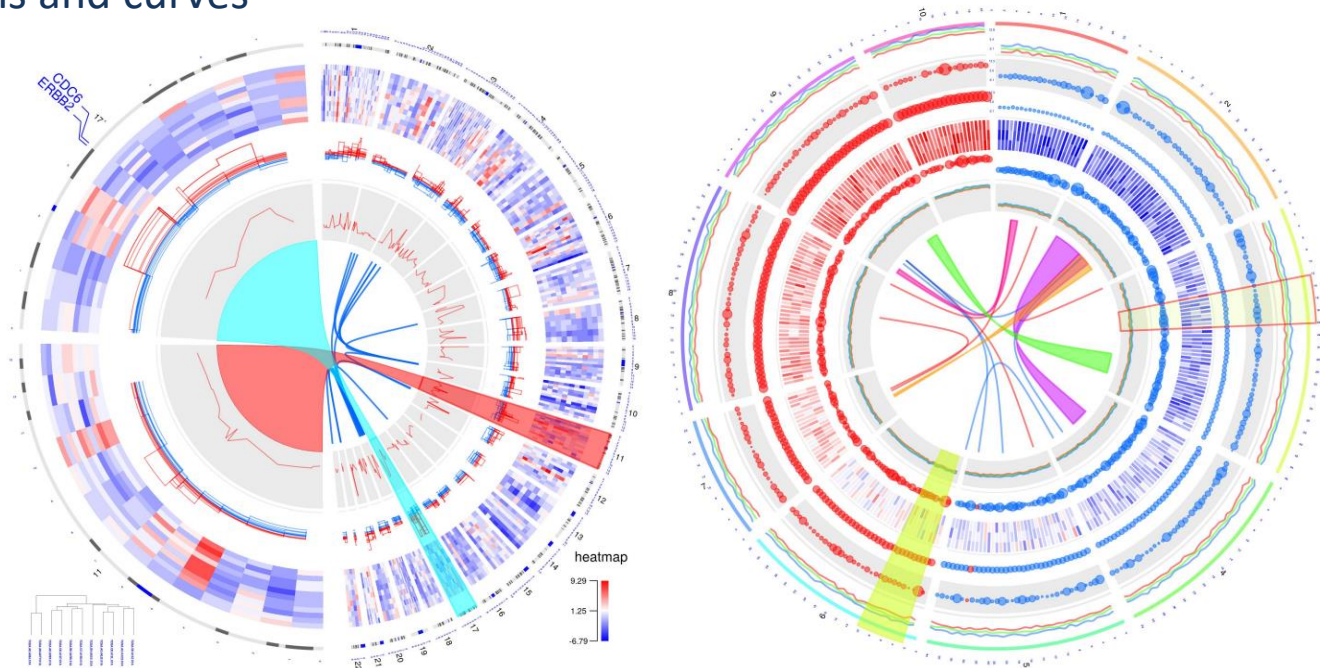
inner/outer radius of track

Text margin in radial direction

No short lines before label

Circos

- Many more possibilities: http://circos.ca/documentation/tutorials/

- Good R package: OmicCircos
  - Easy to use → Each track is drawn independently
  - Gene or chromosome position based display
  - Links as polygons and curves
  - Scatterplots
  - Lines
  - Text labels
  - Boxplots
  - Histograms
  - Heatmaps

Acknowledgment

- IGV: https://www.broadinstitute.org/software/igv/home

- Circos:
  - http://circos.ca/
  - http://circos.ca/documentation/tutorials/
  - http://circos.ca/presentations/talks/circos_intro/

- OmicCircos:
  - https://www.bioconductor.org/packages/release/bioc/html/OmicCircos.html