Genomic Data Visualization

• Online in web based data repositories → UCSC, NCBI Ensembl, GBrowse ...

• Downloadable applications with graphical user interface: IGV, IGB, BamView, Savant, Tablet, GenoViewer, MochiView, SeqMonk, inGAP ...

• Installable web applications: Anno-J, JBrowse
Towards a “better” genome browser

• Writing a better genome browser used to be a “rite of passage” for many bio-informaticians.

• There are probably hundreds of applications with various features/applications – each one is the best as long as you define best in a narrow way!

• Genomic data visualization is a surprisingly complex matter – Users’ needs diverge and can be mutually exclusive
Epitomes of information overload
Many are domain specific

Tools developed in a lab tend to suit the tasks performed in that environment:

- Genome variation → IGV, IGB, Tablet
- Generic visualizer → Artemis
- Targeted use cases
  - ChipSEQ → MochiView
  - DNA Methylation → ChipMonk and SeqMonk
  - and many others
A review paper

Visualizing genomes: techniques and challenges

Cydney B Nielsen¹, Michael Cantor², Inna Dubchak²,³, David Gordon⁴ & Ting Wang⁵

As our ability to generate sequencing data continues to increase, data analysis is replacing data generation as the rate-limiting step in genomics studies. Here we provide a guide to genomic data visualization tools that facilitate analysis tasks by enabling researchers to explore, interpret and manipulate their data, and in some cases perform on-the-fly computations. We will discuss graphical methods designed for the analysis of de novo sequencing assemblies and read alignments, genome browsing, and comparative genomics, highlighting the strengths and limitations of these approaches and the challenges ahead.

What do we visualize?

• Horizontal span (intervals) → horizontal
  – gene location, alignments ...

• Values over intervals → vertical
  – coverages, probabilities ...

• Attributes at certain locations → colors
  – mutation, junctions ...
Developed by the Broad Institute – focused on genetic variation studies
IGB (Ig-Bee) Integrated Genome Browser

It has more options than IGV – it is more of an analysis platform
Choosing a genome browser

• Data for model organisms may be “pre-filled”

• Custom or less common type of data will need to be loaded manually (we will do this)

• Import your own genome if you are not using a standardized genome build
Import your genome into IGV

Genomes → Create .genome file

If you use a model organism with a well defined genomic build that IGV already knows about then you don’t need to these steps
Investigate the feature files

Create feature files with ReadSeq.

Investigate subtle details – understand the type of features that you see

This can be more difficult than you anticipate!
Visualizing alignment BAM file

You may need to zoom in to see the data

This is paired view
Right click on the view to set options
One needs to become familiar with at least one genome browser!
Homework 16

• Create a custom genome in IGV using the ebola genome

• Visualize the alignments that you produced for homework 15. Show a screenshot of your data that covers a genomic region.

• Find what seems to be the highest and lowest covered regions by visual inspection.

• Find the highest and lowest observed coverage using the `samtools depth` command ignoring the first 100 and last 100 bases.