Genomic Data Visualization

• Online websites data are also repositories → these run in a web browser: UCSC, 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”

• There are probably hundreds of applications with various features/applications

• Genomic data visualization is a surprisingly complex matter – users’ needs diverge and can be mutually exclusive
Many are domain specific

Tools developed in a lab tend to suit the tasks frequent in that environment. Many include some tools as well:

• Genome variation → IGV

• ChipSeq → MochiView

• DNA Methylation → ChipMonk and SeqMonk
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.

IGV: Integrative Genomics Viewer

Developed by the Broad Institute – focused on genetic variation studies
IGB (Ig-Bee) Integrated Genome Browser

Seems to offer more options than IGV (but perhaps slightly quirkier).

It is a great tool though with a detailed user guide.
Choosing a genome browser

• Model organisms have more options of data being “pre-filled”

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

• Import your genome if you are NOT using a standardized genome build!
Homework

• Create a custom (mini.fa) genome that only contains the first two chromosomes of the yeast genome

• Generate high coverage sequencing data from it

• Align the generated data to the mini.fa genome

• Visualize the results in IGV
Homework checklist

Think or Squint!
My directory now looks like this

```
ialbert@porthos ~/work/lec14
$ ls refs/
mini.fa mini.fa.ann mini.fa.fai mini.fa.sa sc.fasta
mini.fa.amb mini.fa.bwt mini.fa.pac mini.genome select.sh

ialbert@porthos ~/work/lec14
$ ls
aln.sh r2.fq results.bam.bai
mutations.txt refs select.sh
r1.fq results.bam

ialbert@porthos ~/work/lec14
$ 
```
Import your genome into IGV

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.
Visualizing the BAM file

You will need to zoom in to see the data.

This is paired view.
Right click on the view to set options
Create your annotations

Extract the genes from the annotation file of lecture 2

```bash
# a GFF file needs to start with the header
# it is just ##gff-version 3
# extract it from the sc.gff file
#
head -1 sc.gff > genes.gff
#
# put the genes in the annotation file
# this matches every line that has the word GENE in it
cat sc.gff | grep gene | head -5
#
# a better solution is match on column 3
# and extract lines that have the type gene
#
cat sc.gff | awk ' $3 =="gene" { print $0 }' >> gene.gff
```
Visualize it as a new track
You should understand the capabilities of at least one genome browser!

This guide describes the Integrative Genomics Viewer (IGV).

- To start IGV, go to the IGV downloads page: [http://www.broadinstitute.org/igv/download](http://www.broadinstitute.org/igv/download).
- For a 10-minute hands-on introduction, see the Quick Start.

Look at a printer-friendly HTML version of the whole User Guide.

- To generate a PDF of the UG, look at the HTML of the whole UG, then Print it. The Print dialog should offer you the ability to print to PDF.

- User Interface
- Starting IGV
- Navigating
- Loading a Genome
- Viewing the Reference Genome
- Loading Data and Attributes
- Viewing Data
- Viewing Alignments
- Viewing Variants
- Gene List View
- Regions of Interest
- Sample Attributes
- Sorting, Grouping, and Filtering
- Saving and Restoring Sessions
- Server Configuration
- External Control of IGV
- Creating HTML Links to IGV
- igvtools
- IGV 2.0 Feature Guide

[User Interface]
Nice features

• BAM paired end data support

• Supports opening remote data (data on a webserver)
  – place some data on a web location
  – see  http://bcc.bx.psu.edu/tmp/
  – IGV → open URL:  http://bcc.bx.psu.edu/tmp/results.bam

  (this is a way to share data with other people)
Homework 14

- Generate the data presented in the previous slides and display the alignment in IGV. Show a screenshot of your data.

- What should the average base coverage be?

- What is the highest and lowest observed coverage (hint: `samtools depth`)