Week 5, Lecture 9

István Albert
Bioinformatics Consulting Center
Penn State
Genomic Data Visualization

• Online websites data are also repositories → 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

• Visualization is surprisingly complex matter – users’ needs diverge and can be mutually exclusive
Tools developed in a lab tend to suit the tasks frequent in that environment. Many include some tools as well:

- Genome variation $\rightarrow$ IGV
- ChipSeq $\rightarrow$ MochiView
- DNA Methylation $\rightarrow$ ChipMonk and SeqMonk
- Structural variation detection: inGAP (developed at PSU based, Schuster Lab, Illumina IDEA award 2011)
More project ideas

• Compare one or more short read aligners:
  – Bowtie, BWA, Novoalign, SOAP
  – generate data then compare performance, accuracy, sensitivity

• Compare one or more genome visualizers:
  – usability, features, performance
Files to start with

- annot.gff, yeast.fasta downloaded for lecture 6

- hw8.bam is the paired end alignment file that you created for homework 8
Isolating the lines that match genes

common error: filtering too much, it is harder to catch than filtering too little → be defensive
Find patterns that are unique

Sometimes you have to work a bit more on finding what matches uniquely.

Always double check!

There are better ways of doing this (see later lectures) nevertheless quick matching is very convenient.
IGV: Integrative Genomics Viewer

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

Integrated Genome Browser
Visualization for genome-scale data

What is IGB?

The Integrated Genome Browser (IGB, pronounced Ig-Bee) is an interactive, zoomable, scrollable software program you can use to visualize and explore genome-scale data sets, such as tiling array data, next-generation sequencing results, genome annotations, microarray designs, and the sequence itself. IGB is implemented using the Java programming language and should run on any computer.

More generic but also a bit quirkyer than IGV. It has a detailed user guide.
Create a new genome (IGV)

We are using a custom genome that was not available by default

We need to create a custom genome

Import Genome

ID: Yeast
Name: Yeast Genome
Sequence as FASTA: ygwin/home/albert/docs/web/bioinfo-courses/source/597D-2011/down/yeast.fasta
Gene Annotations: cygwin/home/albert/docs/web/bioinfo-courses/source/597D-2011/down/genes.gff

Save or Cancel
common “gotcha”: you see nothing → 1. zoom in → 2. check chromosome ids
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/hw8.bam

(this is a way to share data with other people)
Homework 9

Visualize the results of Homework 8 in a genome viewer: IGV IGB or another tool of your choice.

1. Load the genome
2. A separate track for the genes
3. A separate track for LTR retrotransposons
4. Load the paired end alignment for homework 8 into the browser.
5. Provide a screenshot of a region where aligned reads overlap with the LTR retrotransposon.

What is a retrotransposon? (as defined by the Sequence Ontology)

What can you see about reads that overlap with retrotransposons?