



2013 - BMMB 597D: Analyzing Next Generation Sequencing Data

# Week 9, Lecture 18

István Albert

Biochemistry and Molecular Biology  
and Bioinformatics Consulting Center

Penn State

# Sequence duplication

Type: **Natural**

**Artificial**

Source: **Sample**

**Contaminant**

Detection: **Sequence identity**    **Alignment identity**

Some tools/methods on the course webpage and course repository

# 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 performed in that environment:

- Genome variation → **IGV, IGB, Tablet**
- ChipSeq → **MochiView** (really nice tool)
- DNA Methylation → **ChipMonk** and **SeqMonk**

# A review paper

## Visualizing genomes: techniques and challenges

Cydney B Nielsen<sup>1</sup>, Michael Cantor<sup>2</sup>, Inna Dubchak<sup>2,3</sup>, David Gordon<sup>4</sup> & Ting Wang<sup>5</sup>

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.

Nature Methods 7, S5 - S15 (2010) doi:10.1038/nmeth.1422

# IGV: Integrative Genomics Viewer

The screenshot shows the homepage of the Integrative Genomics Viewer (IGV) website. The header features the IGV logo and the text "Integrative Genomics Viewer". The main content area has a dark blue header with the word "Home". Below it is a large banner with the text "Integrative Genomics Viewer" and a screenshot of the software interface. The interface displays multiple tracks of genomic data, including tracks for genes, RNA expression levels, and other biological features. To the left of the banner is a sidebar with a navigation menu and a search bar. The navigation menu includes links for Home, Downloads, Documents, FAQ, Hosted Genomes, IGV Quick Start, IGV User Guide, File Formats, Release Notes, Credits, Contact, and a search bar. The search bar has a "search" button. Below the search bar are links to "Broad Home" and "Cancer Program". At the bottom of the sidebar is the "BROAD INSTITUTE" logo and the text "© 2011 Broad Institute".

Developed by the Broad Institute – focused on genetic variation studies

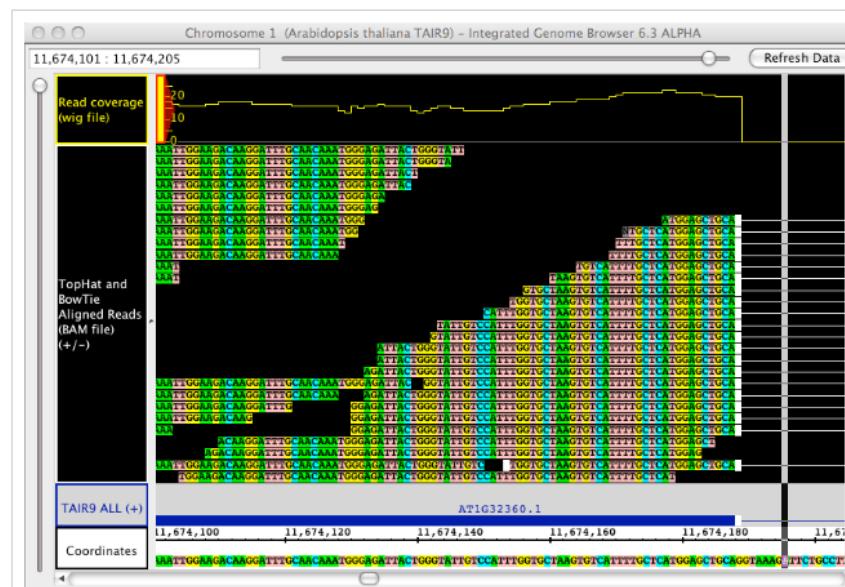
# IGB (Ig-Bee) Integrated Genome Browser

Integrated Genome Browser  
Visualization for genome-scale data

[Home](#)  
[Overview](#)  
[Download](#)  
[Documentation](#)  
[FAQ](#)  
[Contact](#)  
[Credits](#)  
[News](#)  
[Data Sets](#)  
[Other Tools](#)  
[Cite IGB](#)

**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.



Chromosome 1 (Arabidopsis thaliana TAIR9) – Integrated Genome Browser 6.3 ALPHA

11,674,101 : 11,674,205

Read coverage (wig file)

TopHat and BowTie Aligned Reads (BAM file) (+/-)

TAIR9 ALL (+)

Coordinates

AT1G32360.1

11,674,100 11,674,120 11,674,140 11,674,160 11,674,180

175.9 MB / 1,018.8 MB

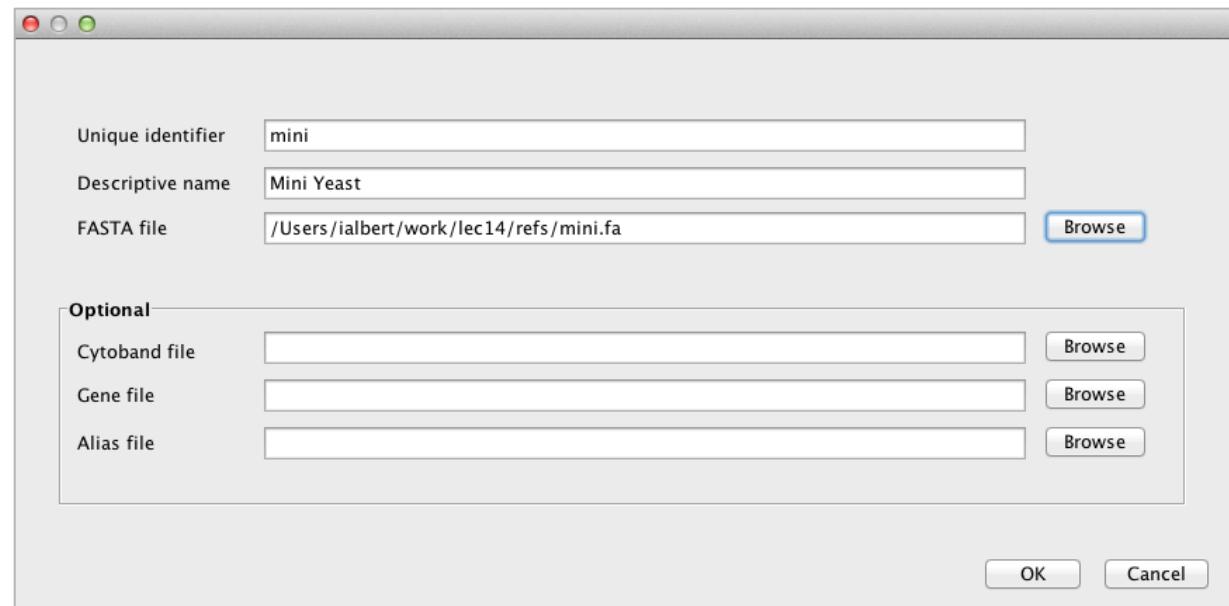
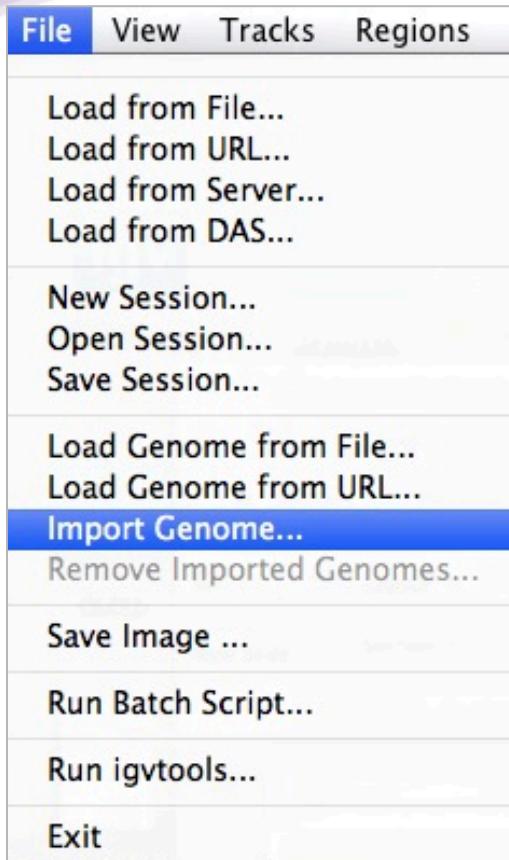
[http://bioviz.org/quickload/A\\_thaliana\\_Jun\\_2009/nextgen\\_samples/wigsample.wig.gz](http://bioviz.org/quickload/A_thaliana_Jun_2009/nextgen_samples/wigsample.wig.gz)

Seems to offers more options than IGV  
It is a great tool though with a detailed user guide

# Choosing a genome browser

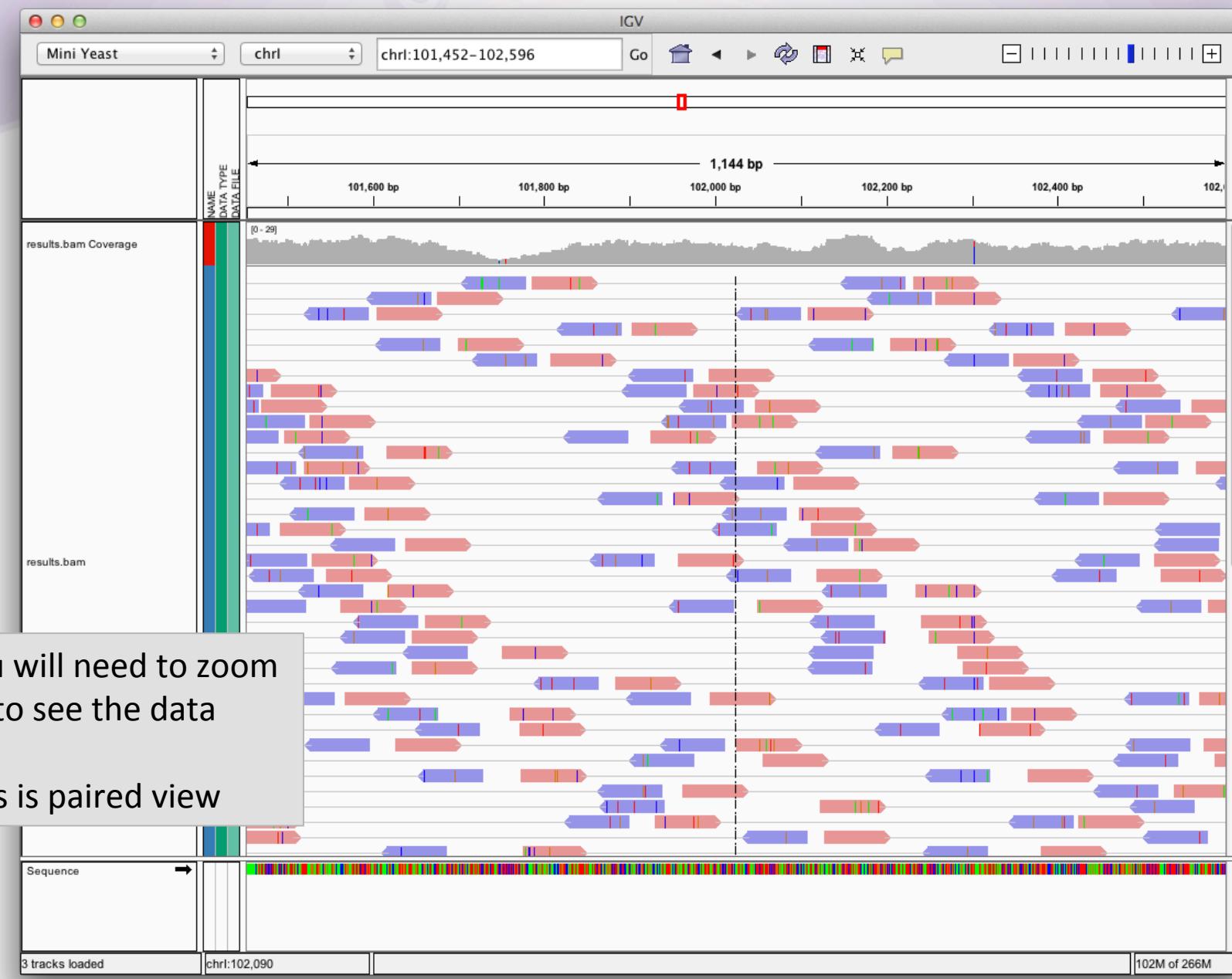
- Data for model organisms may be “pre-filled”
- Custom or less common type of data will need to be loaded manually (we will do this)
- Import your own genome if you are not using a standardized genome build

# Import your genome into IGV

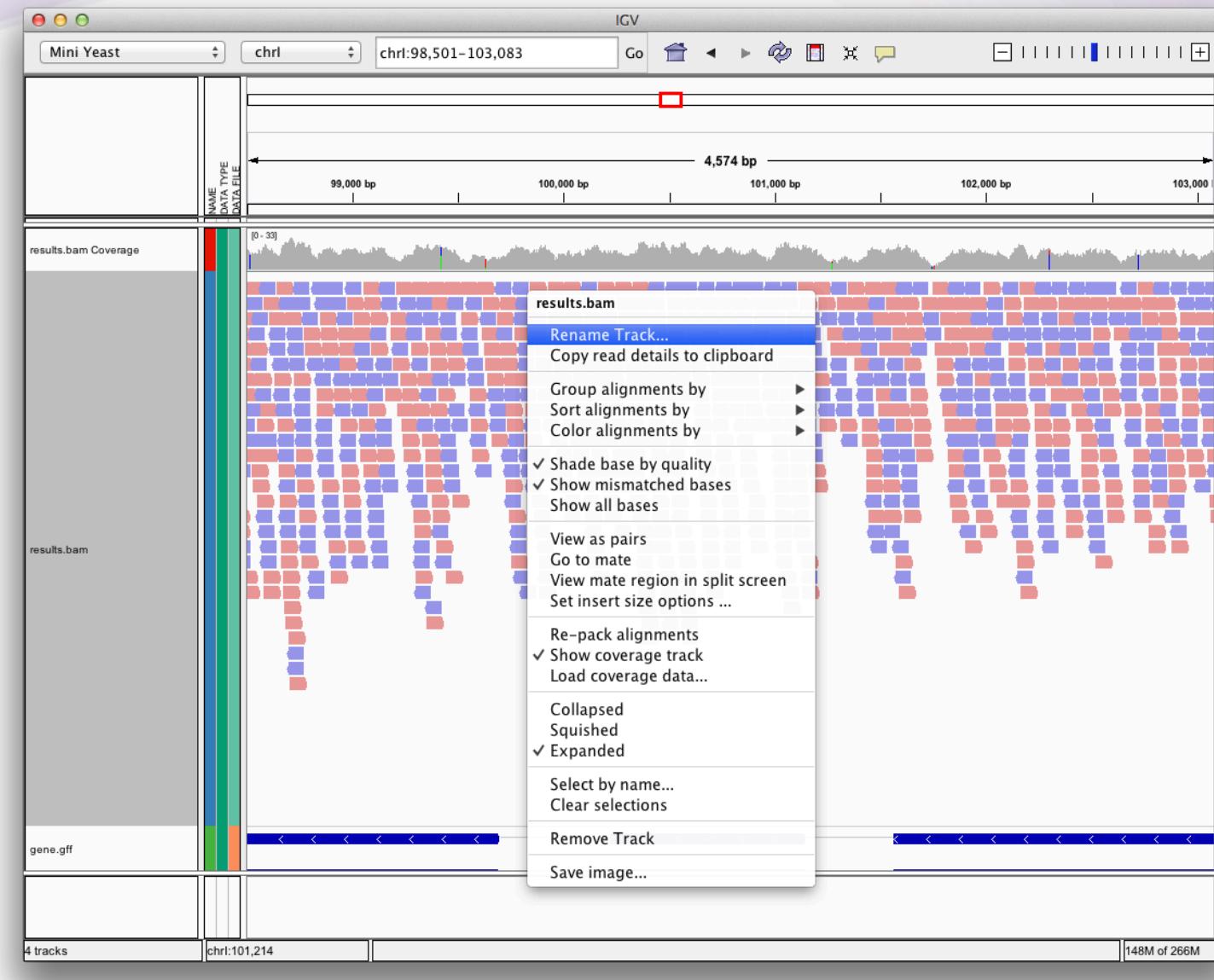


If you use a model organism with a well defined genomic build that IGV already knows about then you don't need to these steps

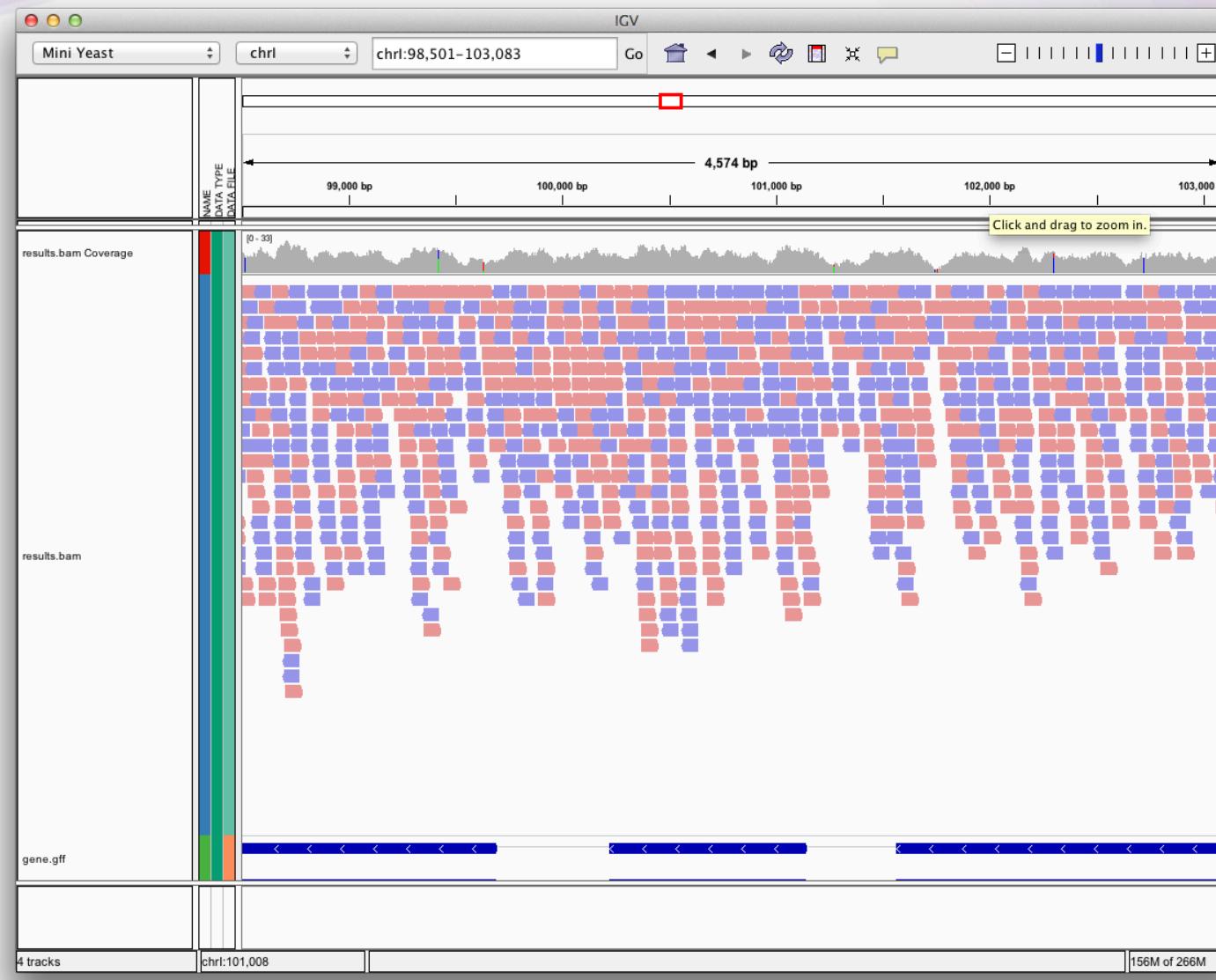
# Visualizing the BAM file



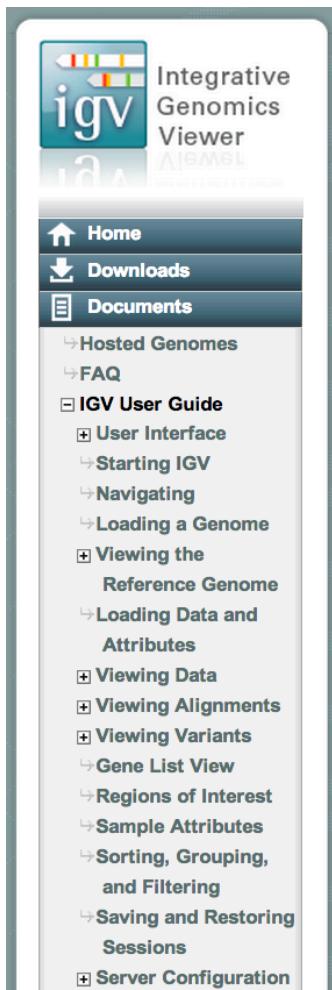
# Right click on the view to set options



# Visualize annotations as a new track



# One needs to become familiar with at least one genome browser!



The screenshot shows the IGV website's navigation menu. The menu items include Home, Downloads, Documents, Hosted Genomes, FAQ, IGV User Guide (which is expanded to show sub-sections like User Interface, Starting IGV, Navigating, etc.), Viewing Data, Viewing Alignments, Viewing Variants, Gene List View, Regions of Interest, Sample Attributes, Sorting, Grouping, and Filtering, Saving and Restoring Sessions, and Server Configuration.

Home > IGV User Guide

## IGV User Guide

This guide describes the Integrative Genomics Viewer (IGV).

- To start IGV, go to the IGV downloads page: <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](#)

[User Interface >](#)

# Useful features

- BAM paired end data support
- Supporting 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 18

- Create a custom genome in IGV using the yeast genome
- Visualize the alignments that you produced for homework 17 (1 million reads from the yeast genome). Show a screenshot of your data that covers a genomic region.
- What is the theoretical base coverage (C)? What is the actually observed coverage?
- Find the highest and lowest observed (nonzero) coverage (hint: **samtools depth**), show a screenshot of these locations