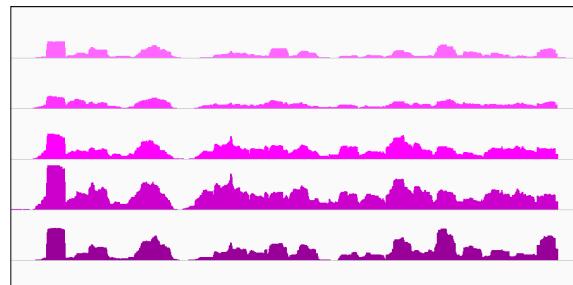


Getting Started with IGV

Programming for Biology 2015



Madelaine Gogol

Programmer Analyst

Computational Biology Core

Stowers Institute

Kansas City, Missouri

Outline

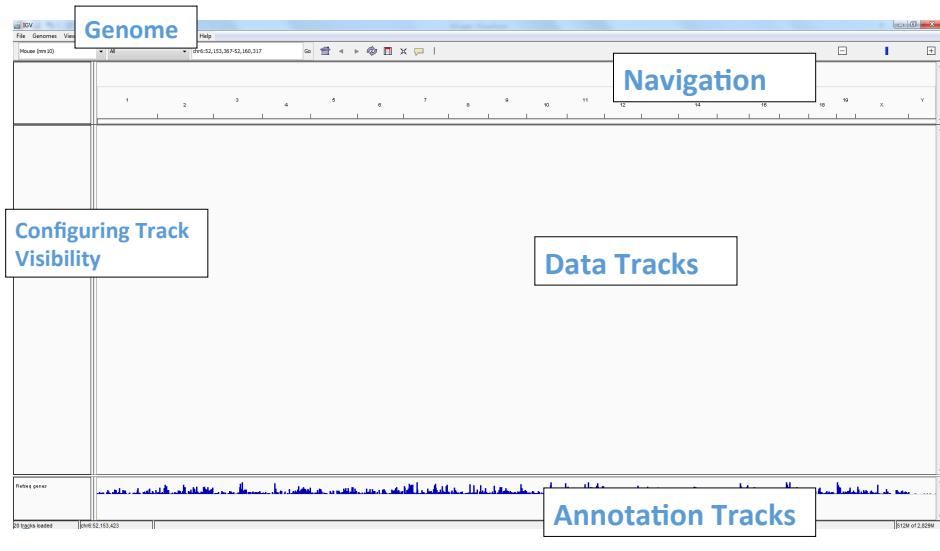
- First, a short presentation
 - Introduction to IGV
 - Files and Formats
 - Installation and Execution tips
- Mostly, a hands-on workshop
 - Navigation
 - Loading Data
 - Visualization Options
 - Saving Sessions
 - Scripting

What is IGV?

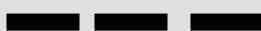


- Integrative Genomics Viewer
- Desktop genome browser – to view genomic data in context
- Runs “locally” (on your computer or a server)
- Developed by James Robinson, et. al, Broad Institute

Basic Orientation



Common track file types (all work in IGV)

	Text format	Binary format
Rectangular 	bed, gff, gtf	bigBed, BAM
Wiggle 	bedGraph, wig	bigWig

Also accepts: Birdsuite, broadPeak/narrowPeak (macs), CBS, CN, Cufflinks, CytoBand, FASTA, GCT, genePred, GFF, GISTIC, Goby, GWAS, IGV, LOH, MAF, MUT, PSL, RES, SAM, SEG, SNP, TAB, TDF, VCF

Installation and Starting IGV

- Mac – Download, unzip Mac App Archive
 - if you need more memory just get binary distribution
- Windows / Linux – Download, unzip binary distribution
 - Then start it with igv.bat (win) or igv.sh (linux/mac)
- Ipad version also available (apple app store)
- “Java Web Start” – cutting edge, could be unstable

Memory requirements

- Often a good idea to up the memory IGV can use
- First see how much memory you have available
 - Windows – right click “Computer”
 - Mac – Apple, About this mac
 - Linux – cat /proc/meminfo
- If your computer is 64-bit, make sure you have 64-bit Java installed so you can use more memory
- Edit the igv.bat or igv.sh file with a text editor

Before editing

```
java -Xmx1200m -Dproduction=true -Djava.net.preferIPv4Stack=true -  
Dsun.java2d.noddraw=true -jar %BatchPath%\igv.jar %*
```

After editing

```
java -Xmx6g -Dproduction=true -Djava.net.preferIPv4Stack=true -  
Dsun.java2d.noddraw=true -jar %BatchPath%\igv.jar %*
```

What is IGV better at (compared to other genome browsers)

My opinion!

- SNPs / structural event examination
- Viewing or troubleshooting the details of “weird” alignments
- Non-model organisms or other “odd” situations
- Local, so doesn’t require hosting data files or passing things around the web if that’s a concern

What is IGV worse at (compared to other genome browsers)

My opinion!

- Loading LOTS of data at once (maybe okay if you have LOTS of memory)
- Not as “pretty” as UCSC?
- Configuring the visualization can be a bit fiddly – can’t always change all tracks at once

Alternatives to Consider:

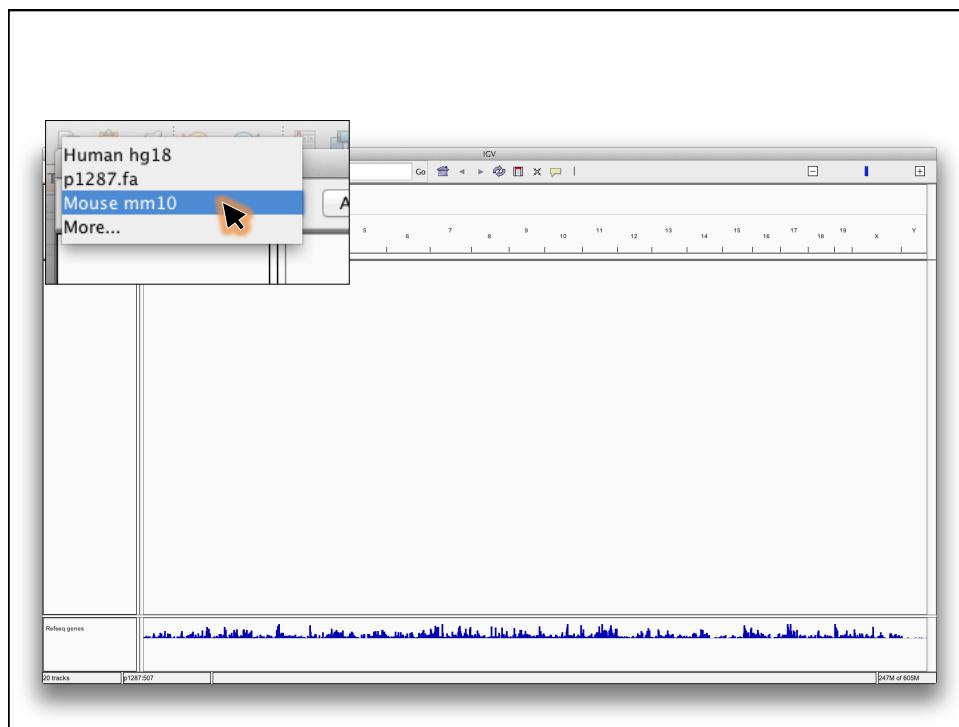
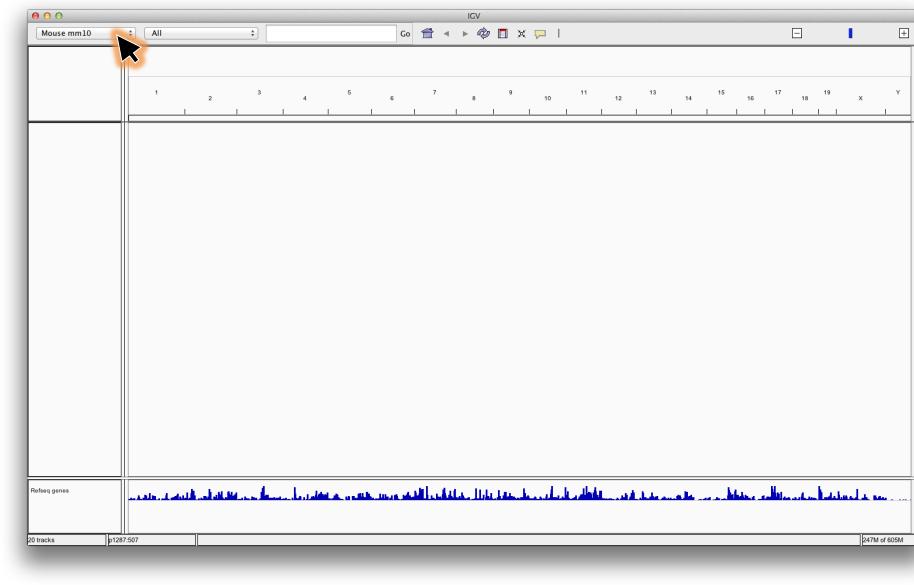
UCSC genome browser, Gbrowse/Jbrowse, IGB, Circos, R-based genome plotting packages (ggbio, GenomeGraphs)

Workshop Time!



<https://www.flickr.com/photos/pennua/5363515039/>

Select the genome - Mouse (mm10)



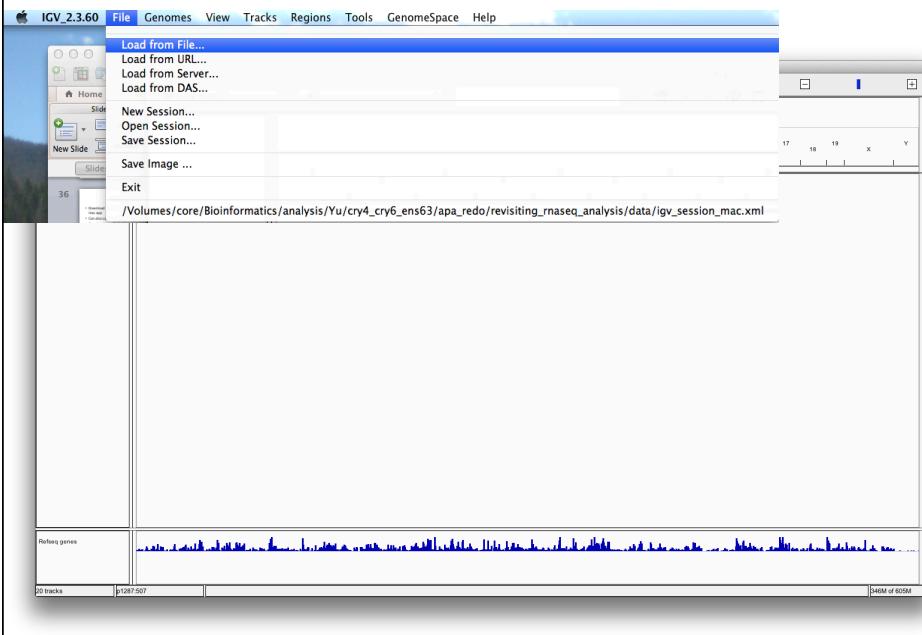
About the data we'll use...

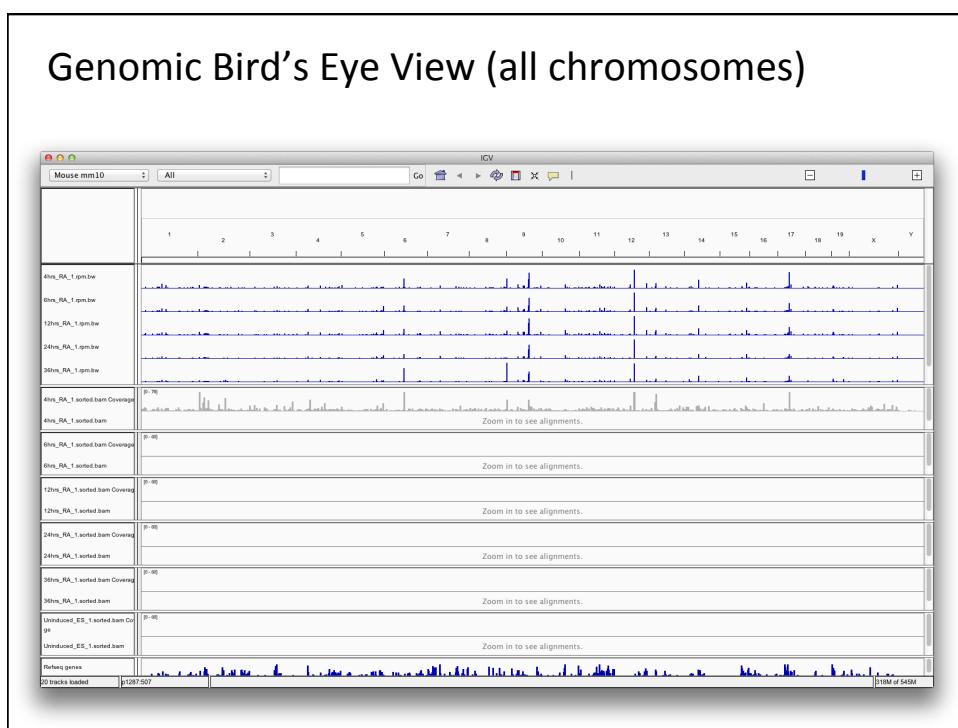
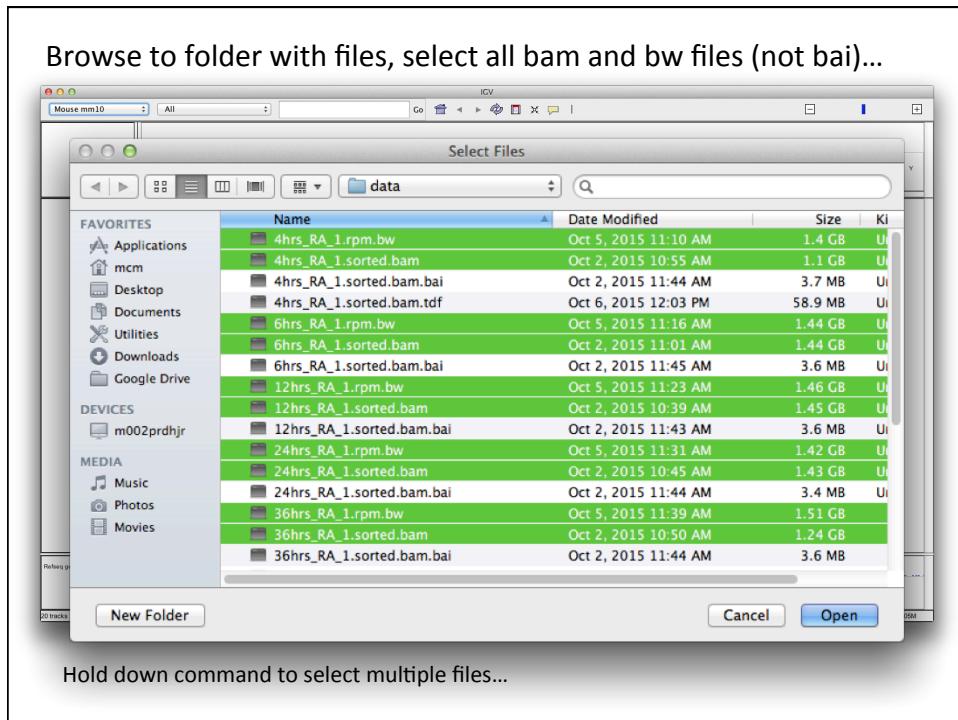
Analysis of dynamic changes in retinoid induced transcription and epigenetic profiles of murine *Hox* clusters in ES cells

Bony De Kumar¹, Mark E. Parrish¹, Brian D Slaughter¹, Jay R Unruh¹, Madelaine Gogoi¹, Christopher Seidel¹, Ariel Paulson¹, Hua Li¹, Karin Gaudenz¹, Allison Peak¹, William McDowell¹, Brian Fleharty¹, Youngwook Ahn¹, Chengqi Lin^{1,4}, Edwin Smith^{1,5}, Ali Shilatifard^{1,5} and Robb Krumlauf^{1,2,3}

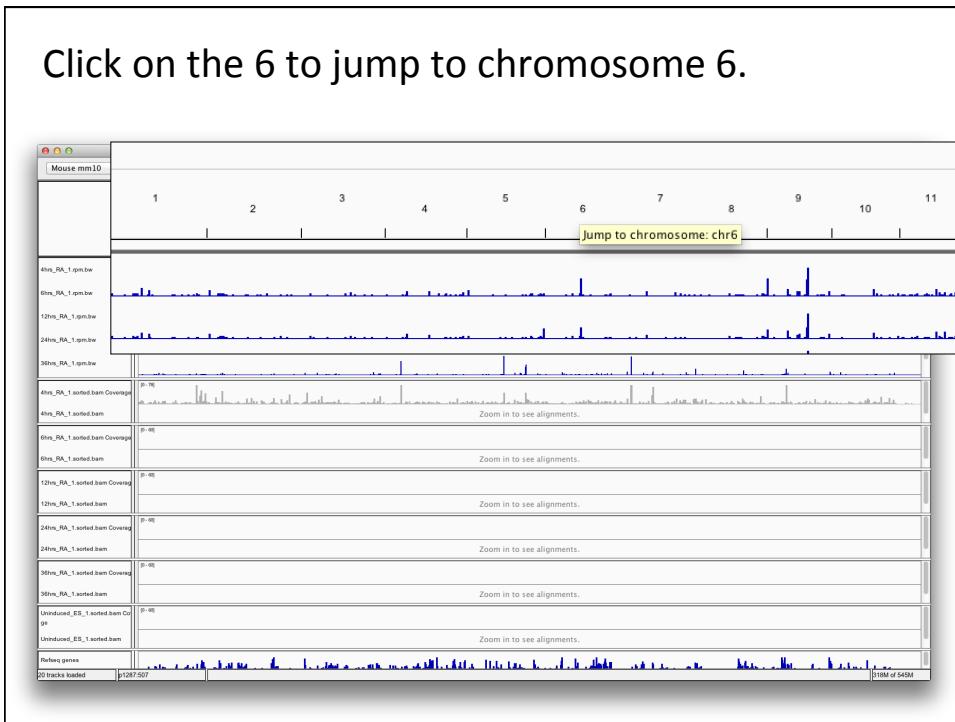
GSE67610

Let's load some data!

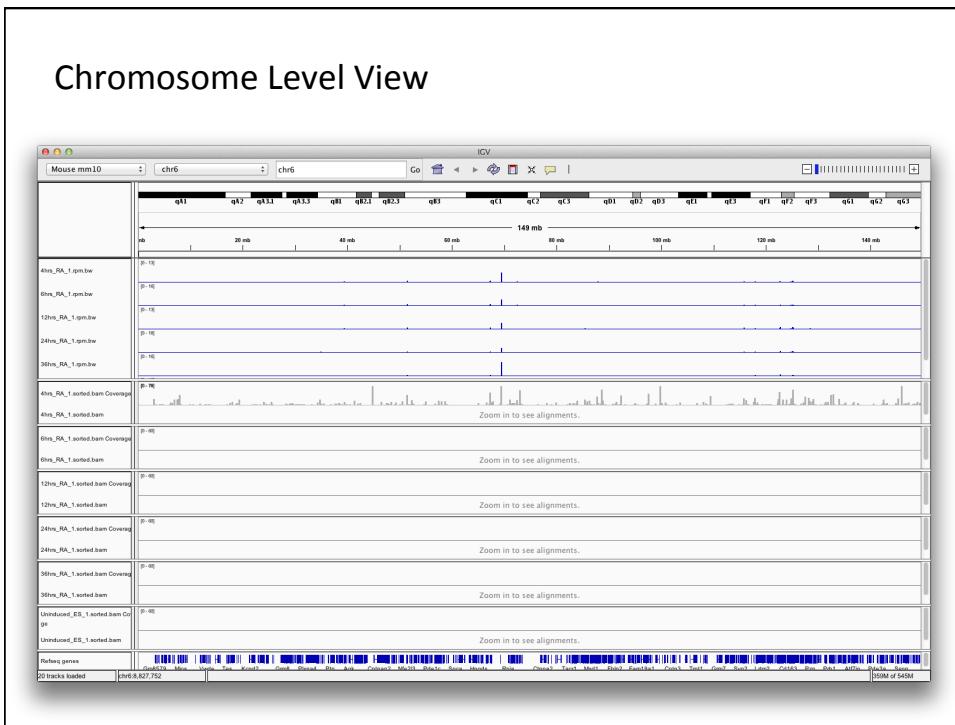




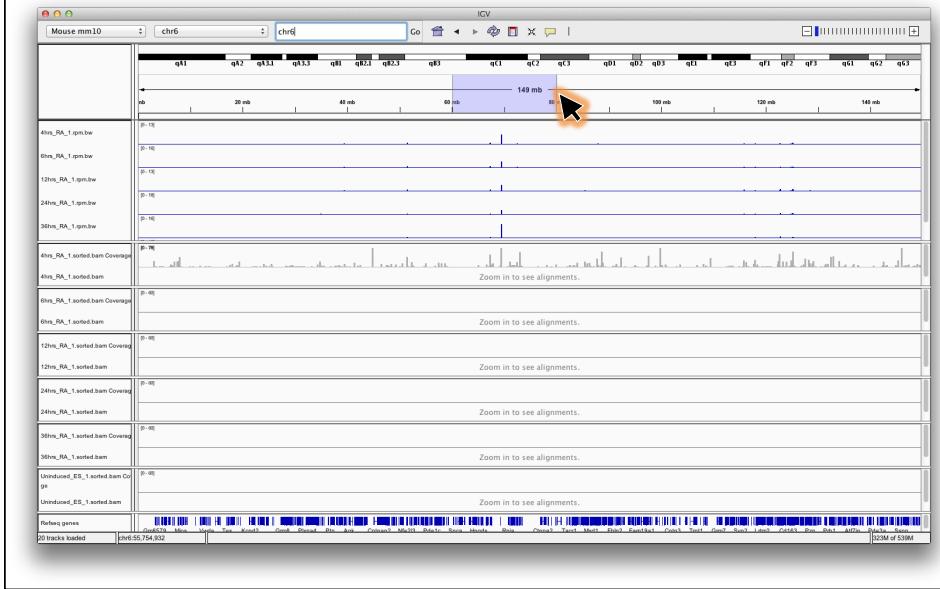
Click on the 6 to jump to chromosome 6.



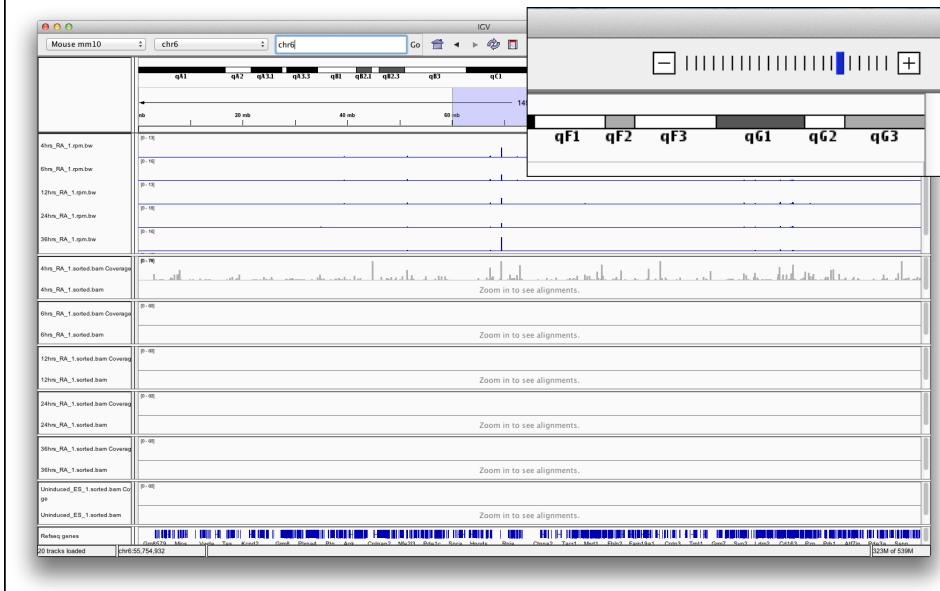
Chromosome Level View



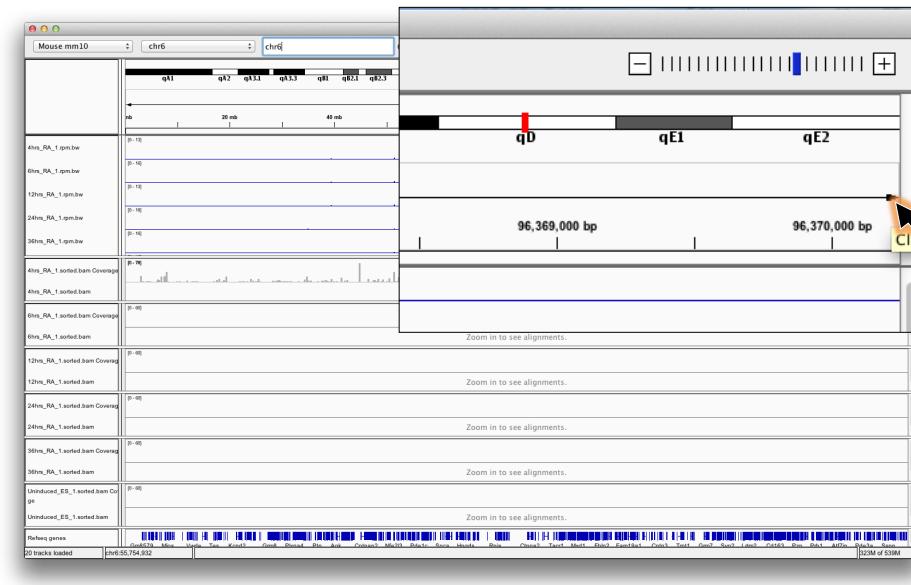
Click and drag to zoom in on a region of chromosome 6.



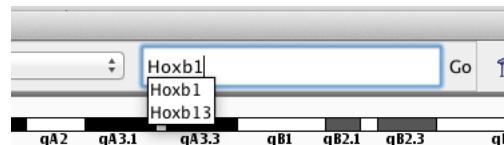
Keep click and dragging or use railroad to zoom until you can see some alignments.



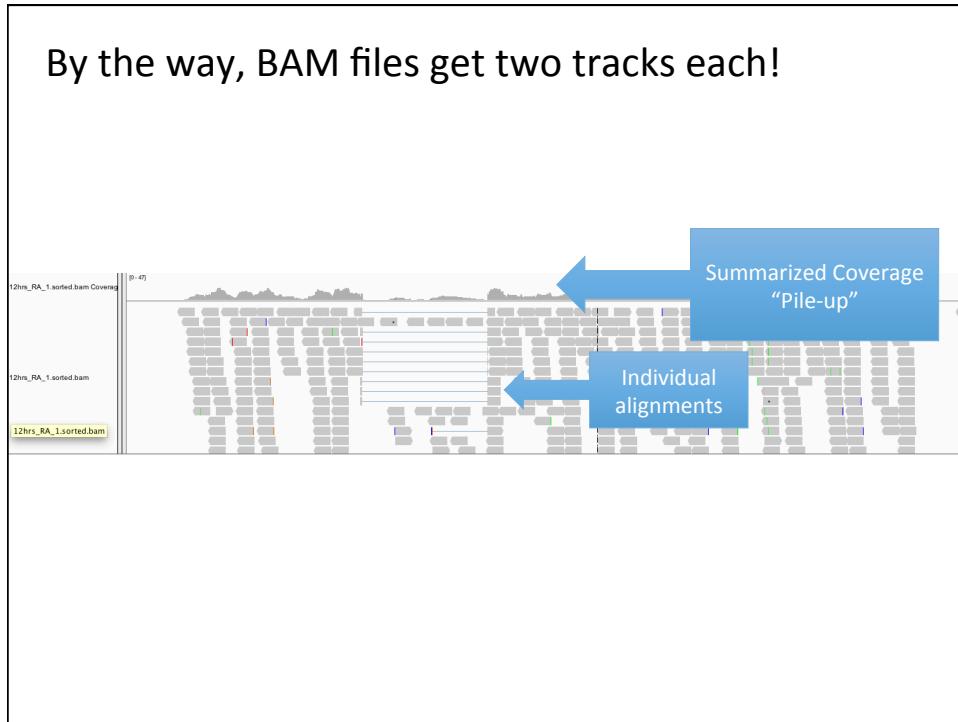
Move left and right along the chromosome by left-click and drag on tracks or click the tiny arrows.



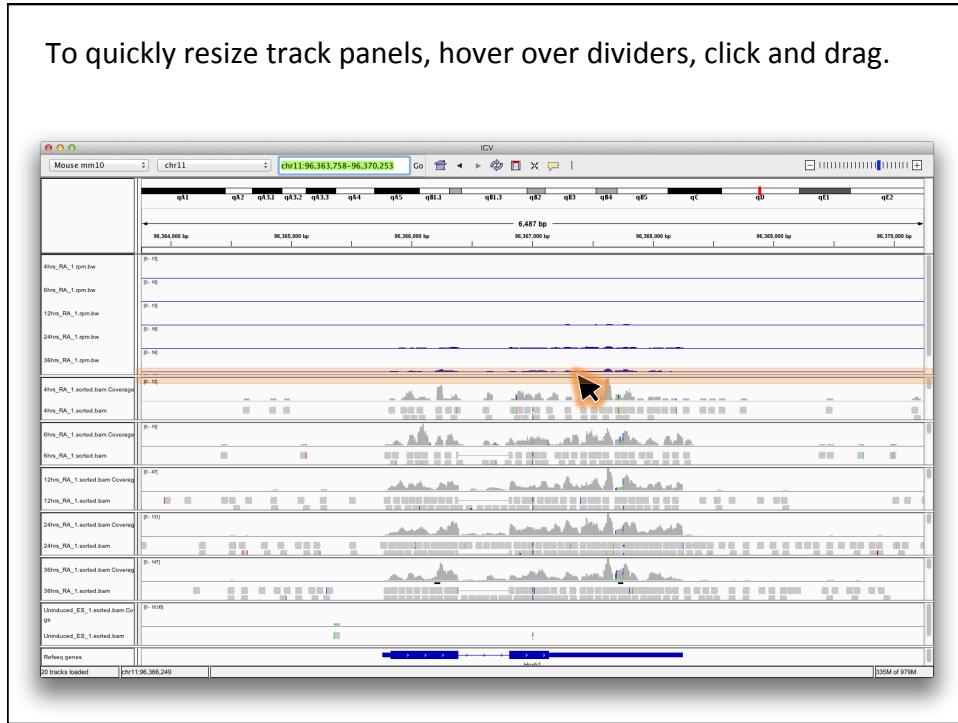
Try typing a mouse gene name in the search box and navigate to it. Are there any reads aligning there?



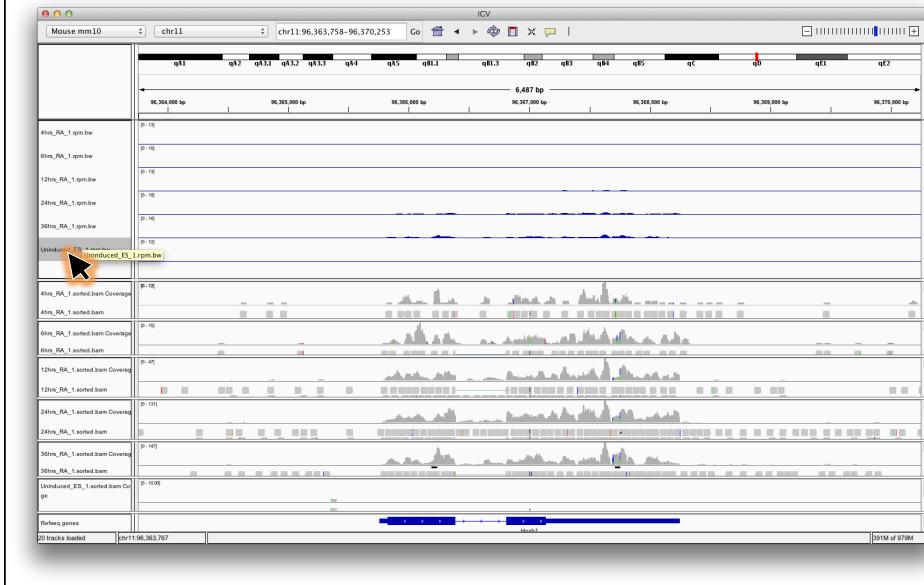
By the way, BAM files get two tracks each!



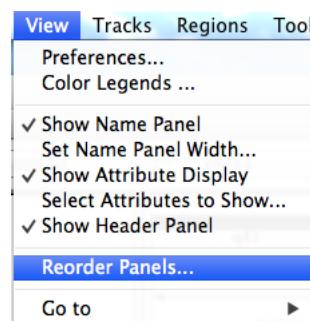
To quickly resize track panels, hover over dividers, click and drag.



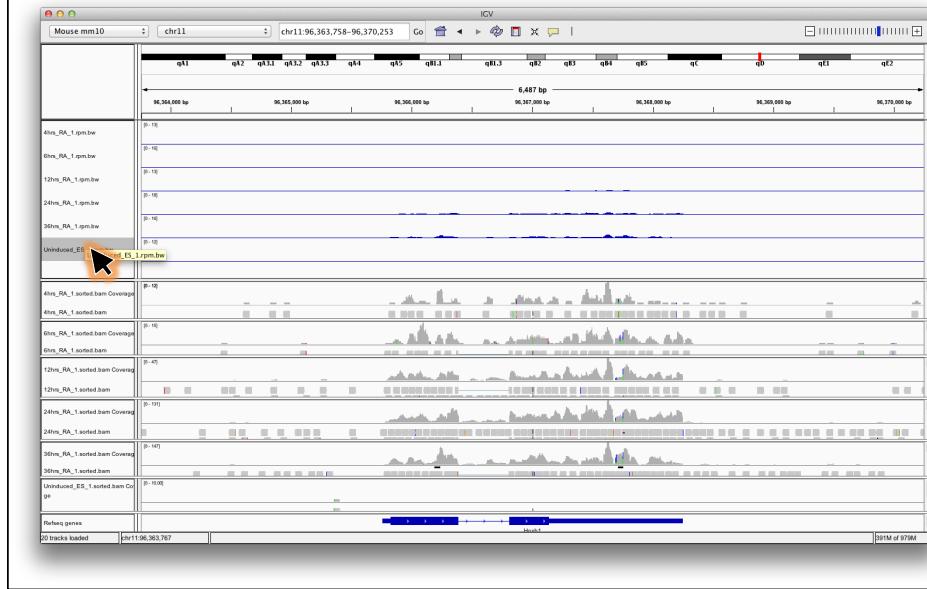
To reorder tracks, click and drag the track name on the left side.



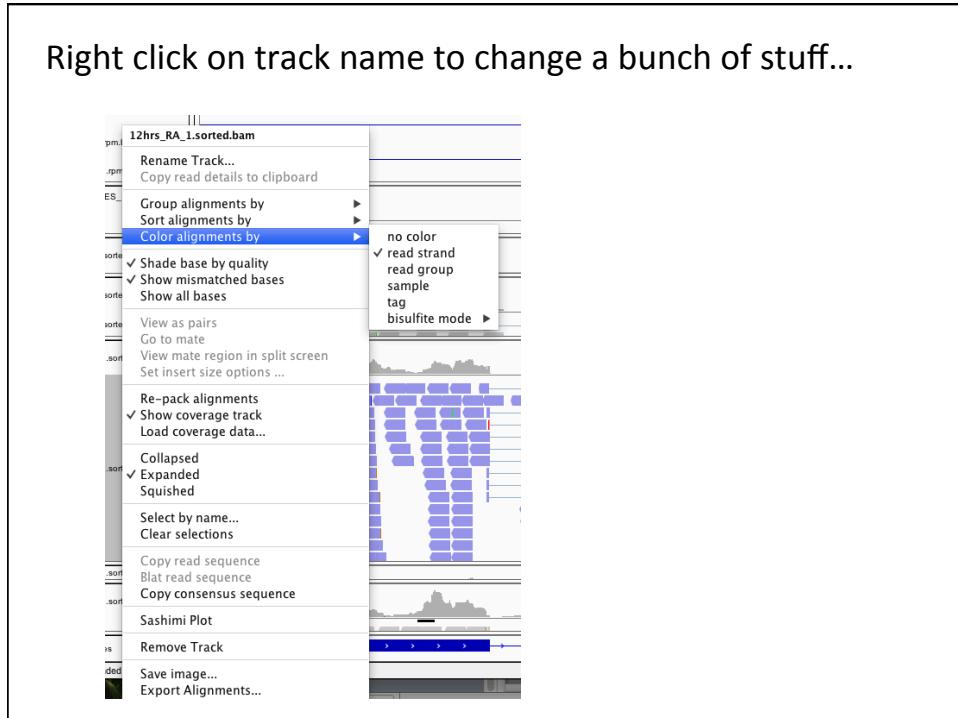
To reorder “panels” easier – View menu – Reorder Panels...



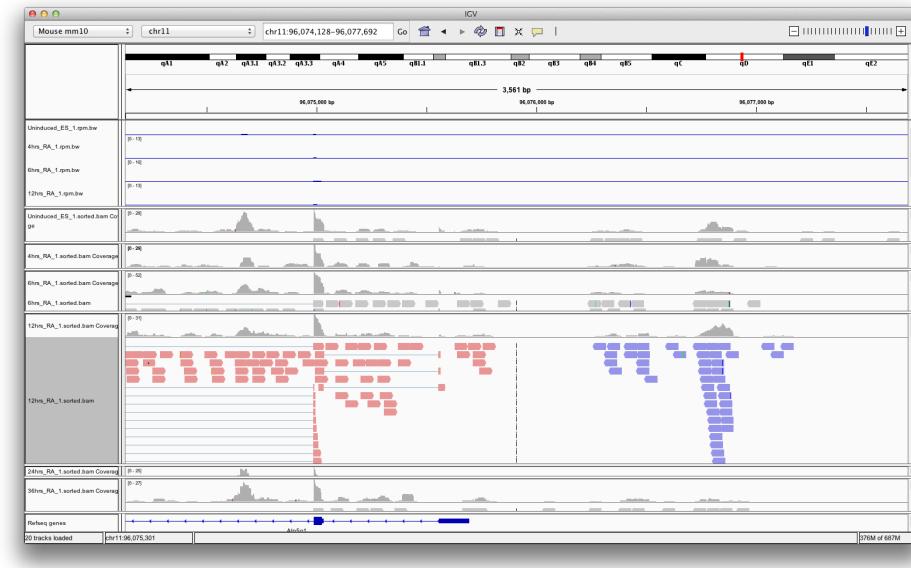
Hold command and click to select multiple tracks at once.



Right click on track name to change a bunch of stuff...



Color alignments by strand

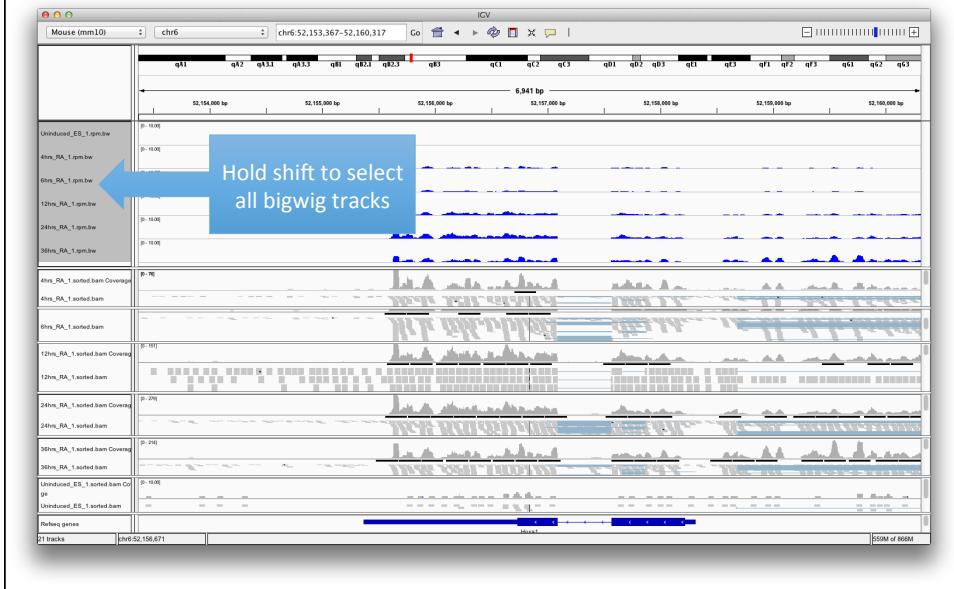


Exploring Hoxa1

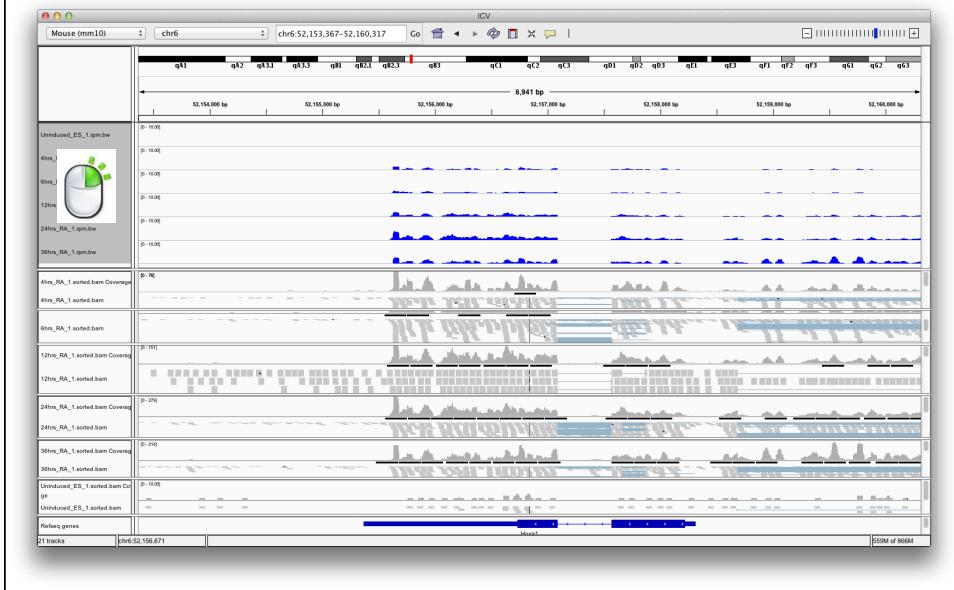
You try it!

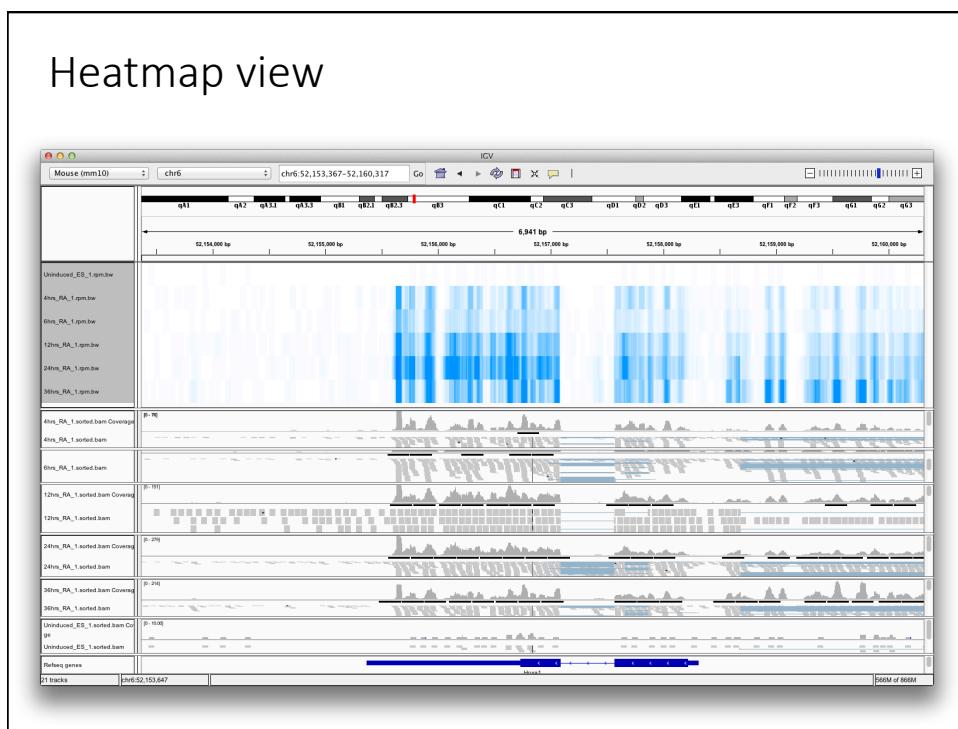
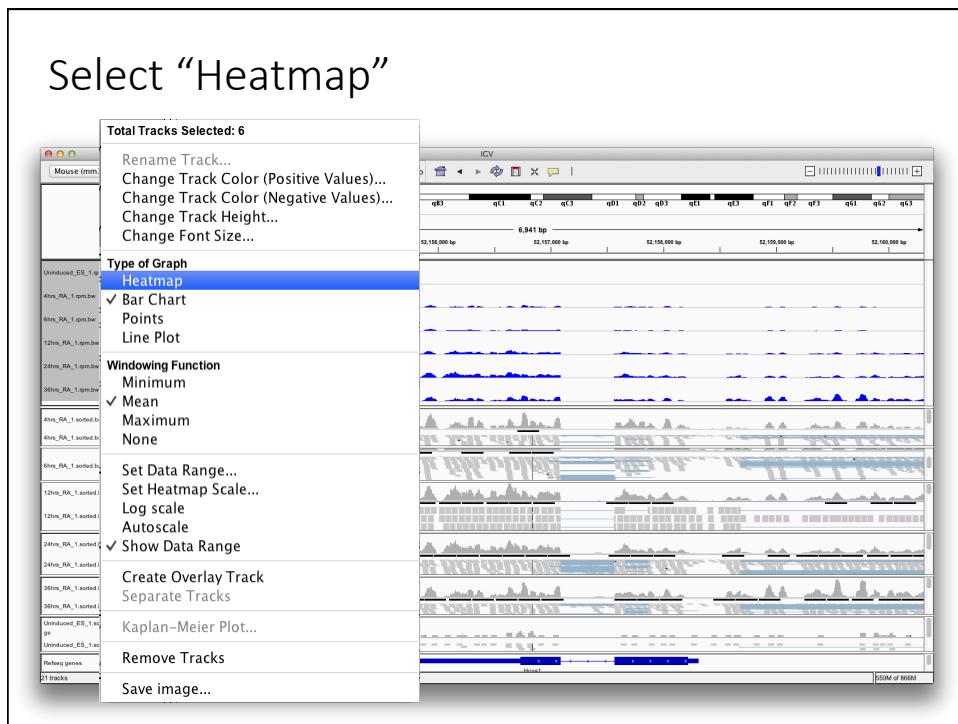
- Go to gene Hoxa1 using the search bar
- Is the expression level of Hoxa1 generally increasing or decreasing from uninduced to 36hrs?
- Color one of the bam file read tracks by strand. Which strand are the reads aligning to? Is this the “expected” strand?

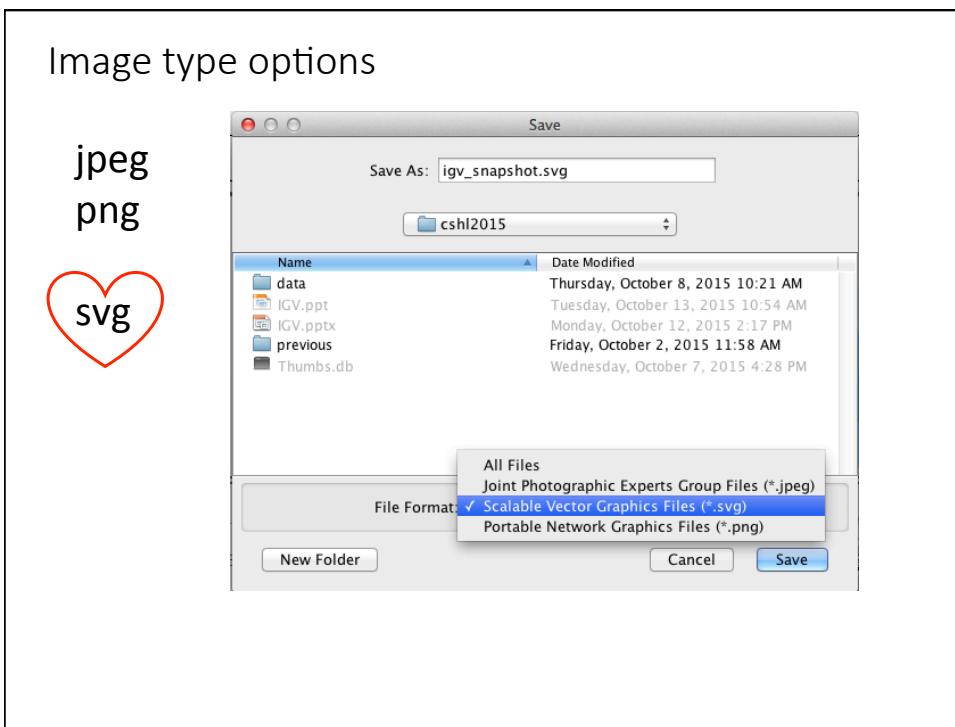
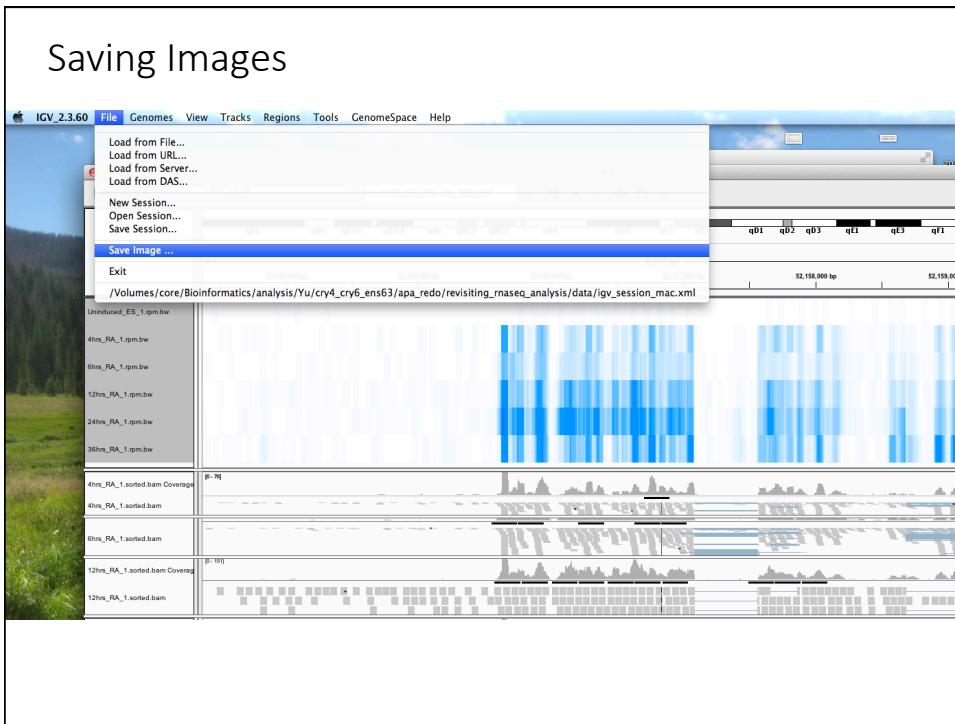
From barplot to heatmap



Shift to select all, right click





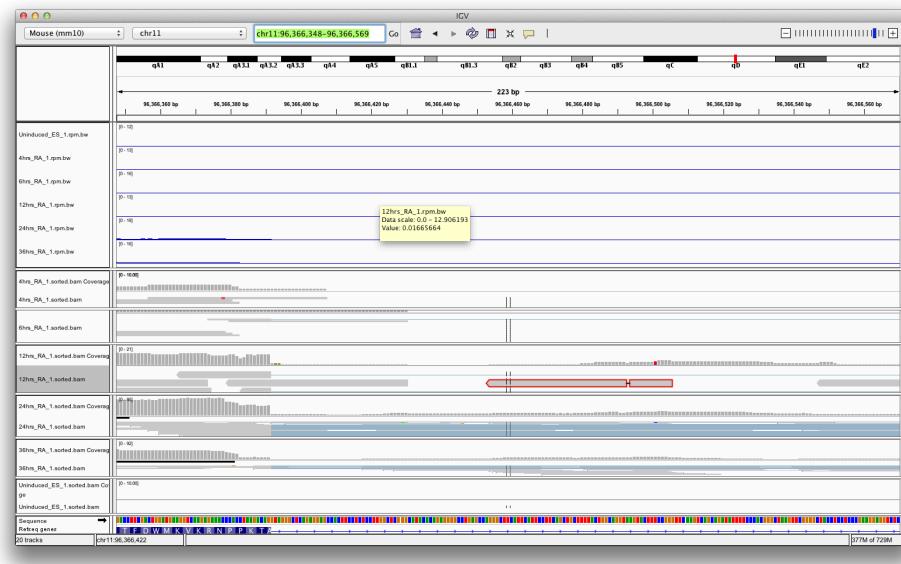


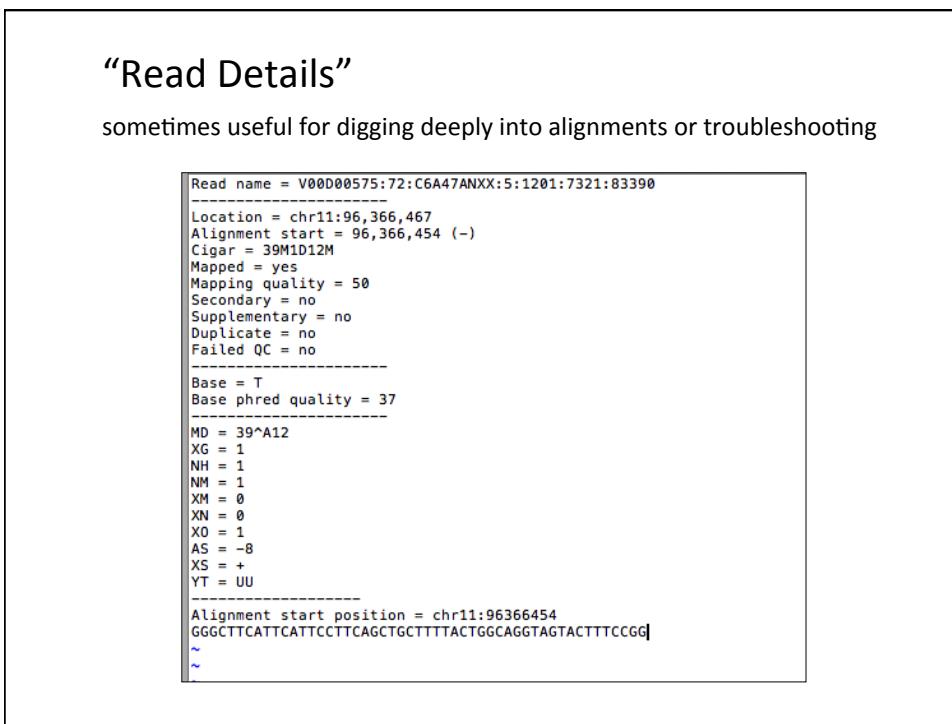
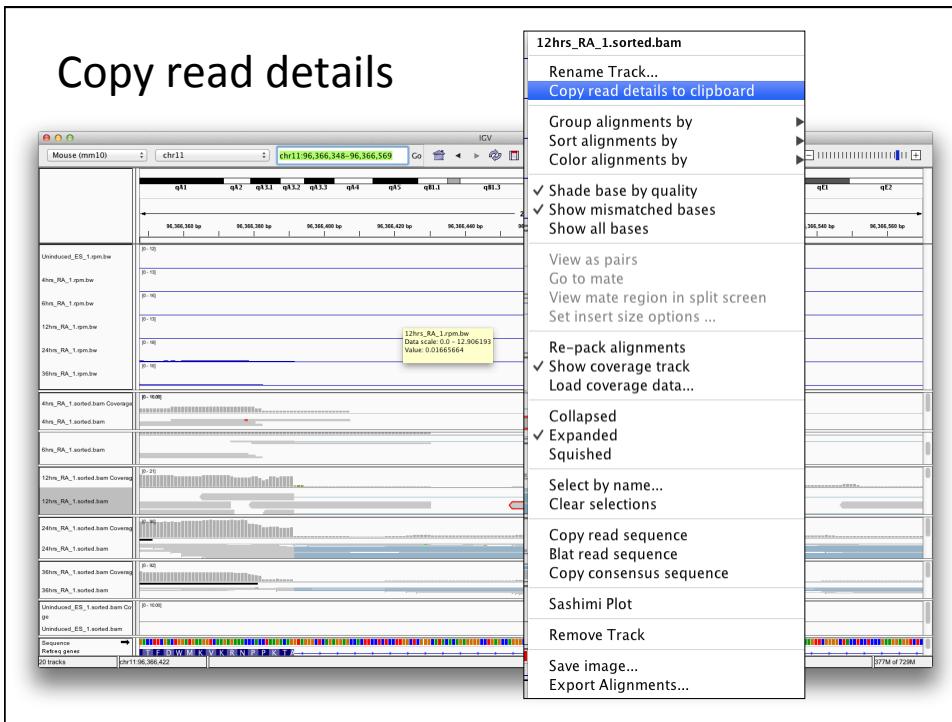
Editing SVG

- Open in illustrator or inkscape (free, open source)
- Ungroup, edit individual elements

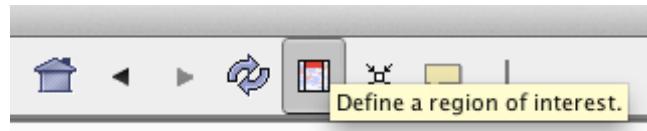


Zoom in until you can see reads and right-click on a read

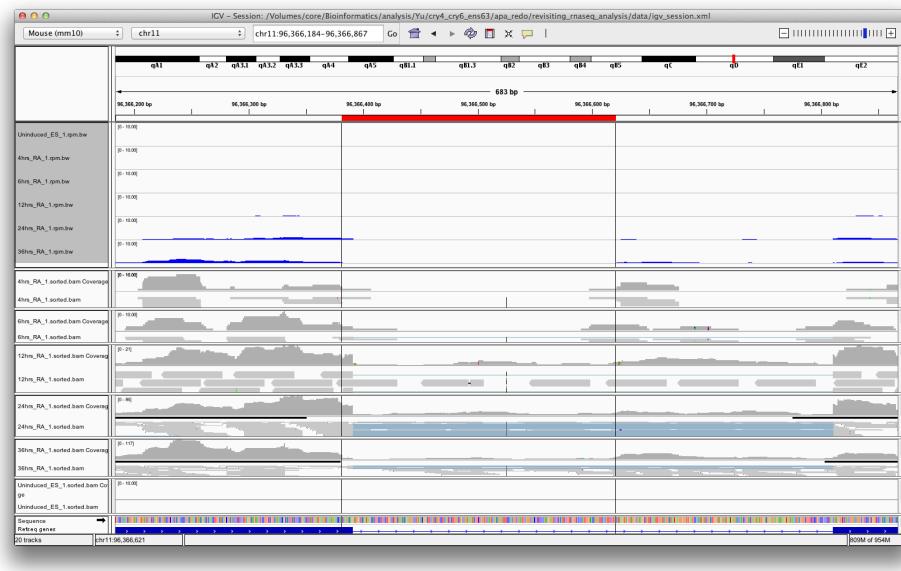




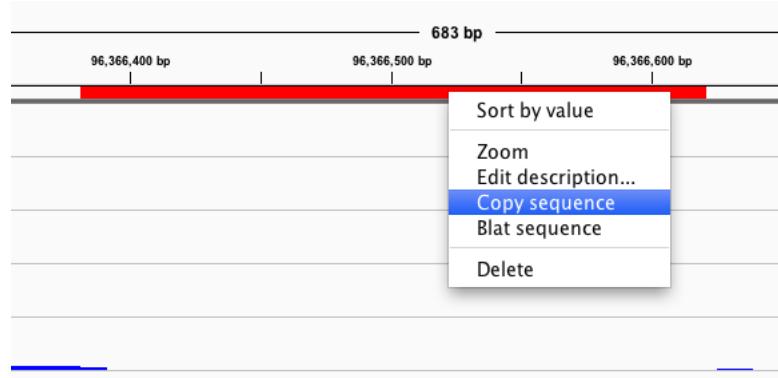
Exporting sequence for a region



Define the left and right ends of the region by clicking on the tracks

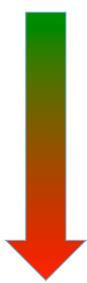


Right click on the red rectangle



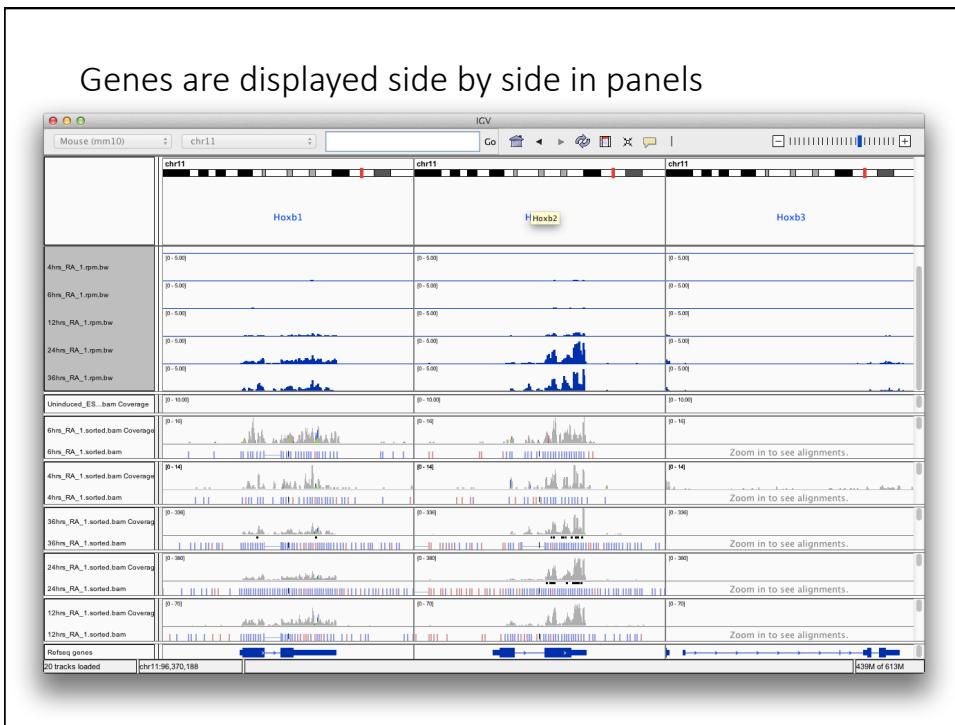
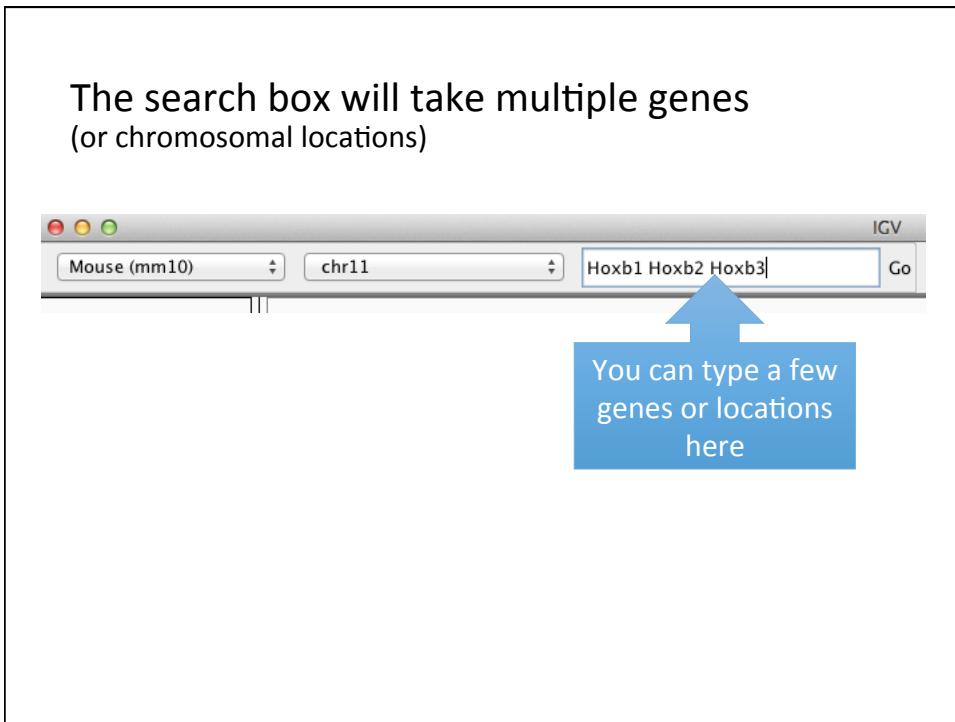
Viewing multiple regions at once

Quick & Easy, Less Powerful



Search box
Constructing links
Batch scripting

Harder, More Powerful



Looking at Hoxa genes...

Search for Hoxa1, Hoxa2, Hoxa3...

You try it!

Constructing links to regions

You can create a link that will open IGV at a specific location.

Examples:

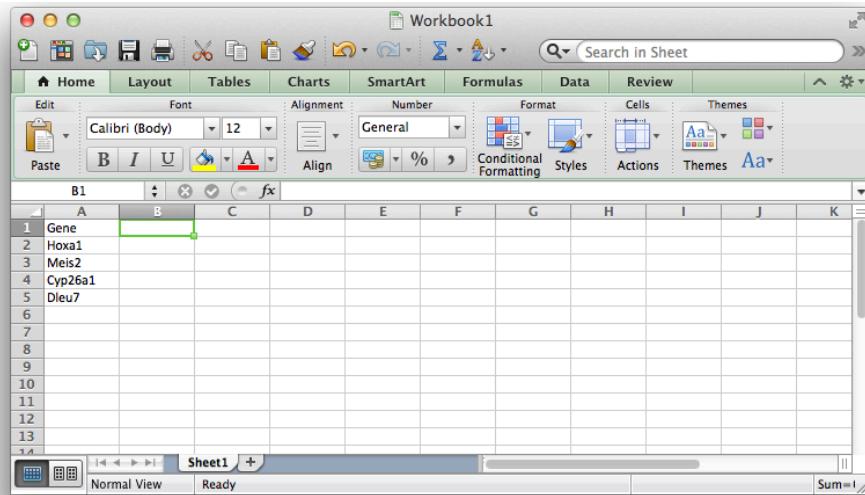
<http://localhost:60151/goto?locus=Hoxa1>

<http://localhost:60151/goto?locus=chr1:1-500>

You can do other things with this, like load data. For more information:

<https://www.broadinstitute.org/igv/ControlIGV>

Constructing links to genes (in Excel)



The screenshot shows a Microsoft Excel spreadsheet titled "Workbook1". The "Home" tab is selected. The data is organized into two columns: "Gene" and "Link". The "Gene" column contains the following entries:

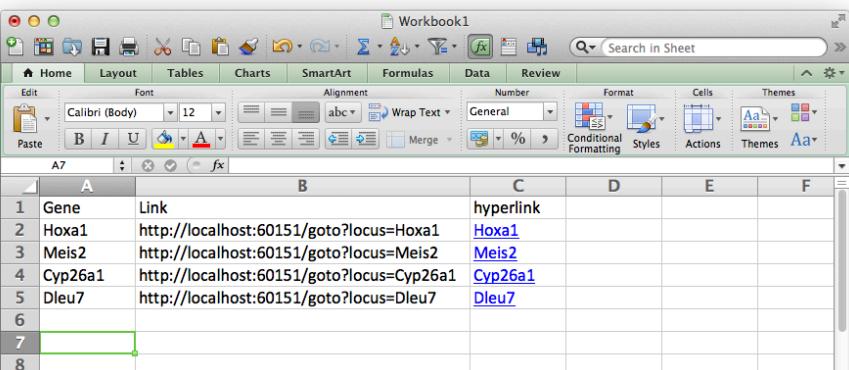
	Gene
1	Gene
2	Hoxa1
3	Meis2
4	Cyp26a1
5	Dleu7
6	
7	
8	
9	
10	
11	
12	
13	
14	

The "Link" column contains the following URLs:

	Link
1	
2	http://localhost:60151/goto?locus=Hoxa1
3	http://localhost:60151/goto?locus=Meis2
4	http://localhost:60151/goto?locus=Cyp26a1
5	http://localhost:60151/goto?locus=Dleu7
6	
7	
8	

Excel demo

Use concatenate and hyperlink functions to construct links to genes by name...



The screenshot shows a Microsoft Excel spreadsheet titled "Workbook1". The "Home" tab is selected. The data is organized into three columns: "Gene", "Link", and "hyperlink". The "Gene" column contains the following entries:

	Gene
1	Gene
2	Hoxa1
3	Meis2
4	Cyp26a1
5	Dleu7
6	
7	
8	

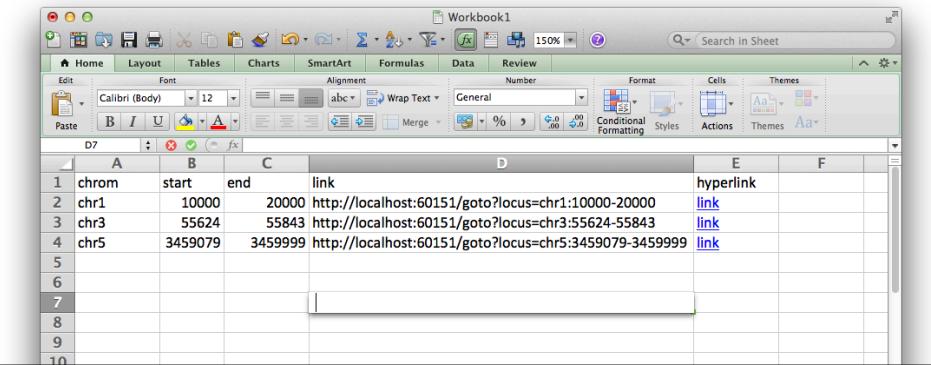
The "Link" column contains the following URLs:

	Link
1	
2	http://localhost:60151/goto?locus=Hoxa1
3	http://localhost:60151/goto?locus=Meis2
4	http://localhost:60151/goto?locus=Cyp26a1
5	http://localhost:60151/goto?locus=Dleu7
6	
7	
8	

The "hyperlink" column contains the gene names from the "Link" column, which are formatted as hyperlinks.

Excel demo

Or link to chromosomal locations (peak calls, SNPs, etc)...



A screenshot of Microsoft Excel showing a table with genomic data. The table has columns labeled 'chrom', 'start', 'end', and 'link'. The 'link' column contains URLs that are formatted as hyperlinks. Row 1 is a header. Rows 2, 3, and 4 are data rows. Row 5 is blank. Row 6 is blank. Row 7 is selected.

	A	B	C	D	E	F
1	chrom	start	end	link	hyperlink	
2	chr1	10000	20000	http://localhost:60151/goto?locus=chr1:10000-20000	link	
3	chr3	55624	55843	http://localhost:60151/goto?locus=chr3:55624-55843	link	
4	chr5	3459079	3459999	http://localhost:60151/goto?locus=chr5:3459079-3459999	link	
5						
6						
7						
8						
9						
10						

Batch Scripting IGV Demo

IGV has its own simple scripting language! (18 commands)

```
new
genome hg18
Load myfile.bam
snapshotDirectory mySnapshotDirectory
goto chr1:65,289,335-65,309,335
sort position
collapse
snapshot
goto chr1:113,144,120-113,164,120
sort base
collapse
snapshot
```

<https://www.broadinstitute.org/software/igv/batch>

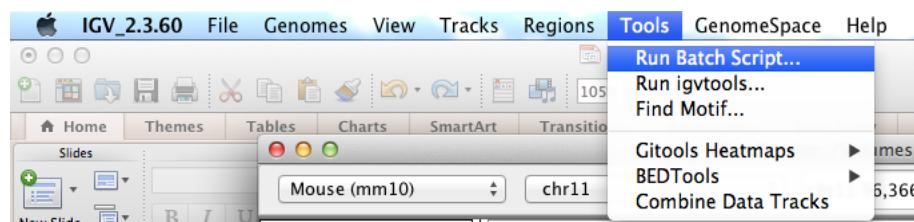
Let's drop to the terminal...

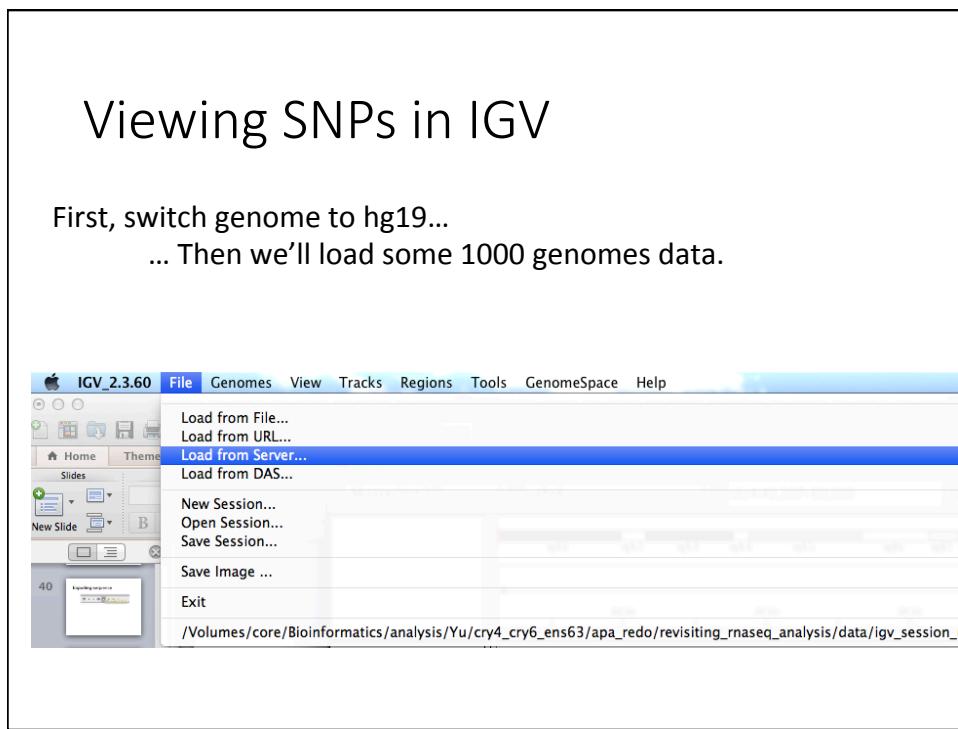
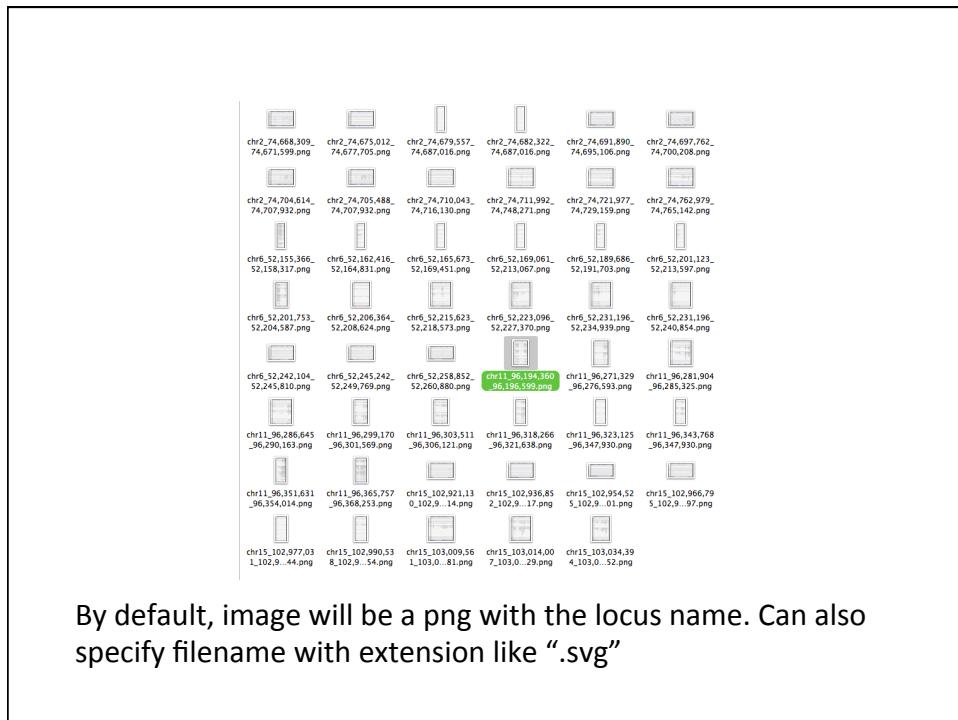
`hox.bed` – a bed file listing the location of all mouse hox genes

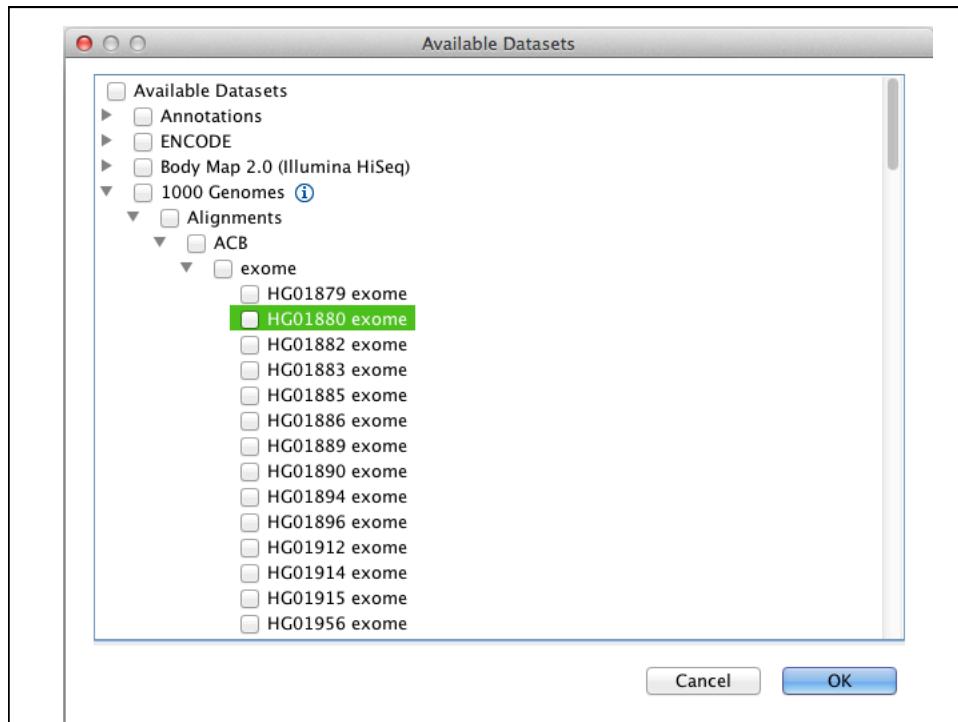
`batch_igv.pl` – a perl script to turn that bed file into an igv batch file

`igv.batch` – the resulting igv batch file

Loading the batch file into igv

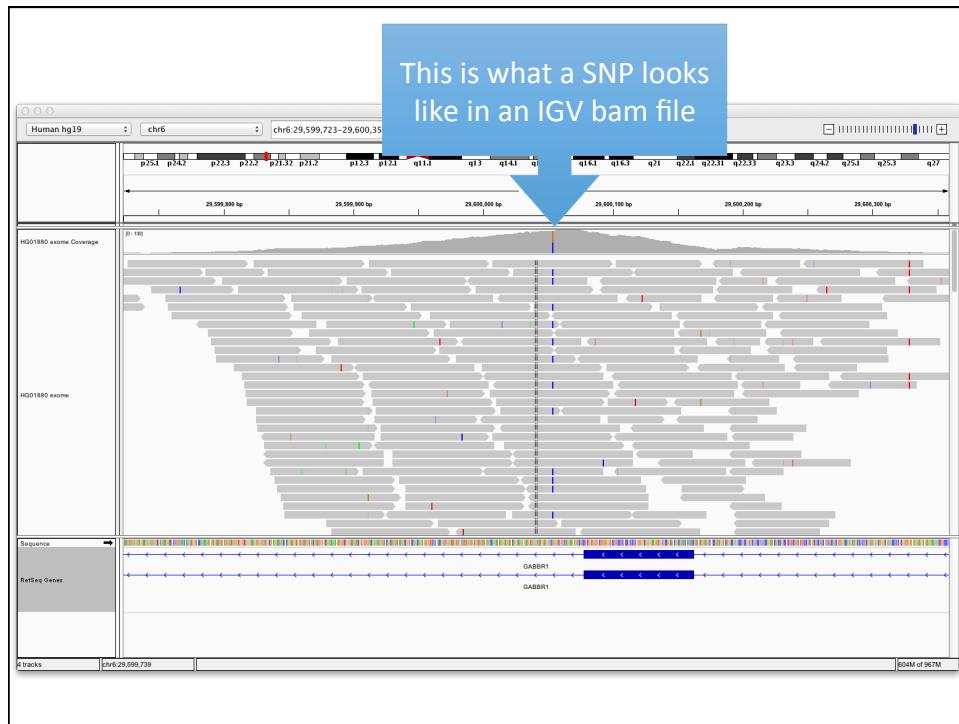




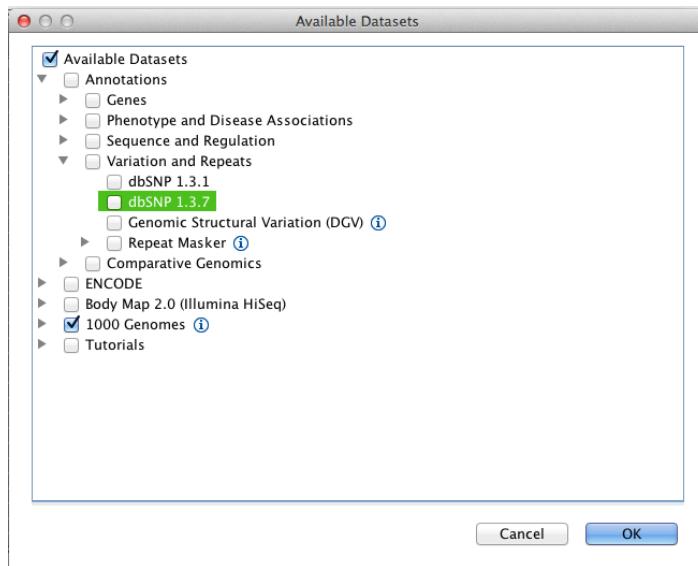


The slide has a white background with a blue diagonal banner in the upper right corner containing the text "You try it!". The main title "Looking at a SNP" is centered at the top. Below the title, there is a bulleted list:

- Go to GABBR1 gene, and zoom in on the last few exons...

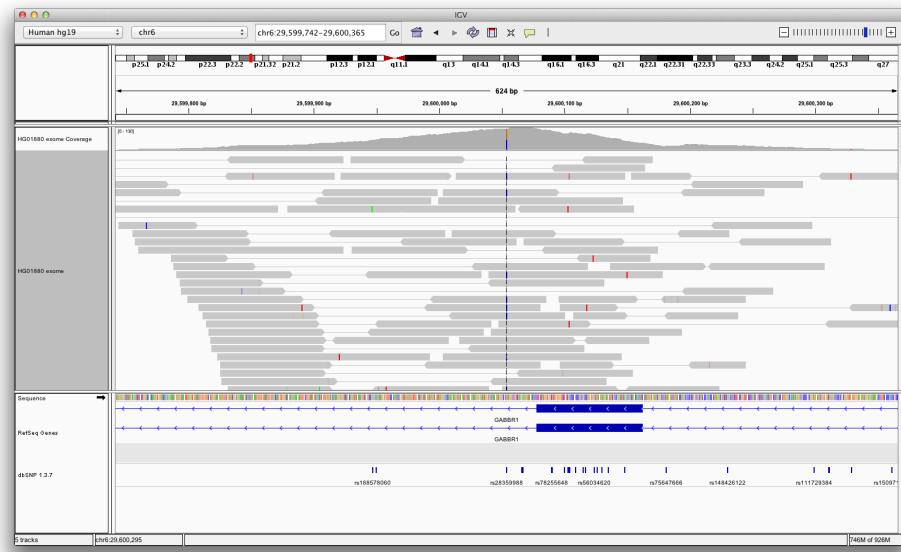


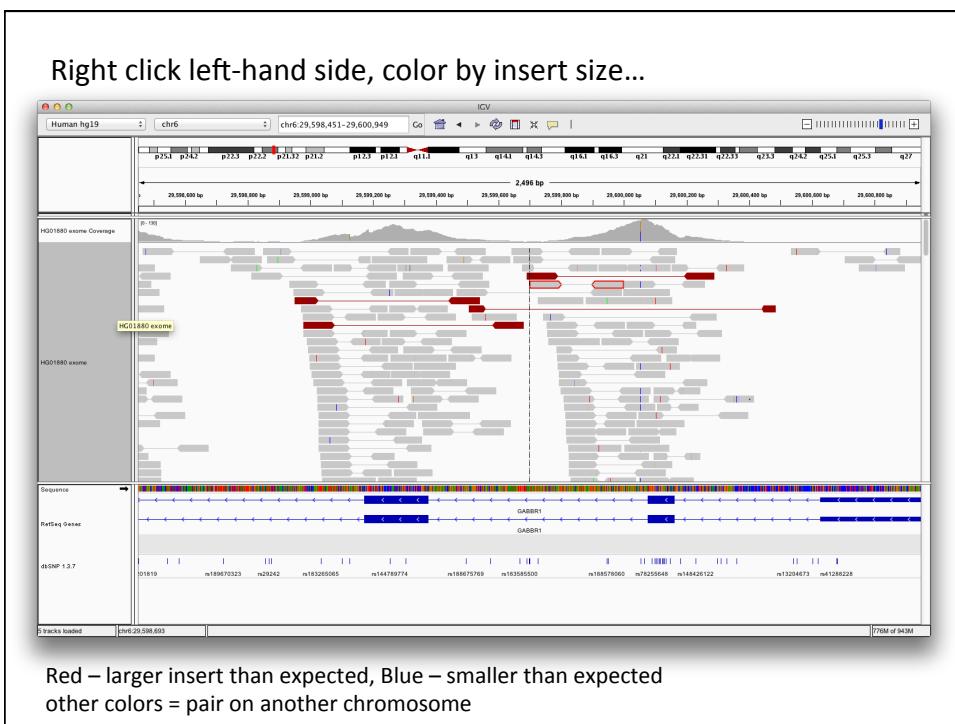
Load dbSNP annotation



While we have some paired-end data to look at...

Right-click, View as Pairs





Summary

- IGV is a “Desktop” or “local” genome browser
- You may need to up the default memory
- Good for SNPs / Structural anomalies / Non-model genomes
- Visualizations are flexible
- When in doubt, right click or hover
- Comprehensive documentation available at
<http://www.broadinstitute.org/software/igv>
 - Also a google group mailing list

Get out there and view some genomes!



Any Questions?

Thanks to Bony & the Krumlauf Lab for the data, Sofia & Simon for inviting me to teach, and Jim Robinson and the IGV Team for making and supporting IGV!

James T. Robinson, Helga Thorvaldsdottir, Wendy Winckler, Mitchell Guttman, Eric S. Lander, Gad Getz, Jill P. Mesirov.

[Integrative Genomics Viewer. Nature Biotechnology 29](#), 24-26 (2011)