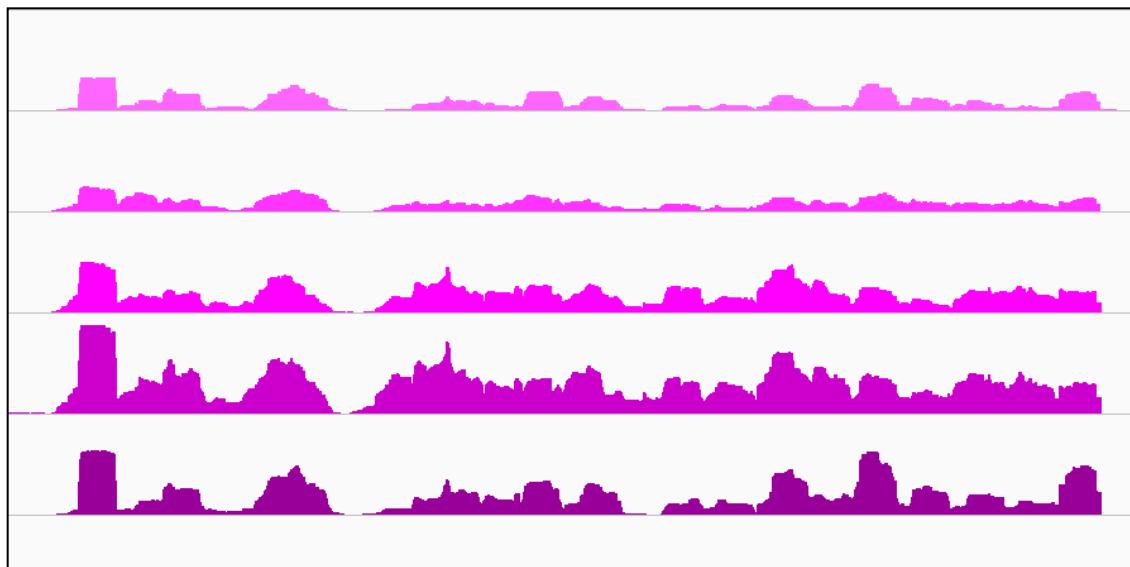


Getting Started with IGV

Programming for Biology 2015



Madelaine Gogol

Programmer Analyst

Computational Biology Core

Stowers Institute

Kansas City, Missouri

Outline

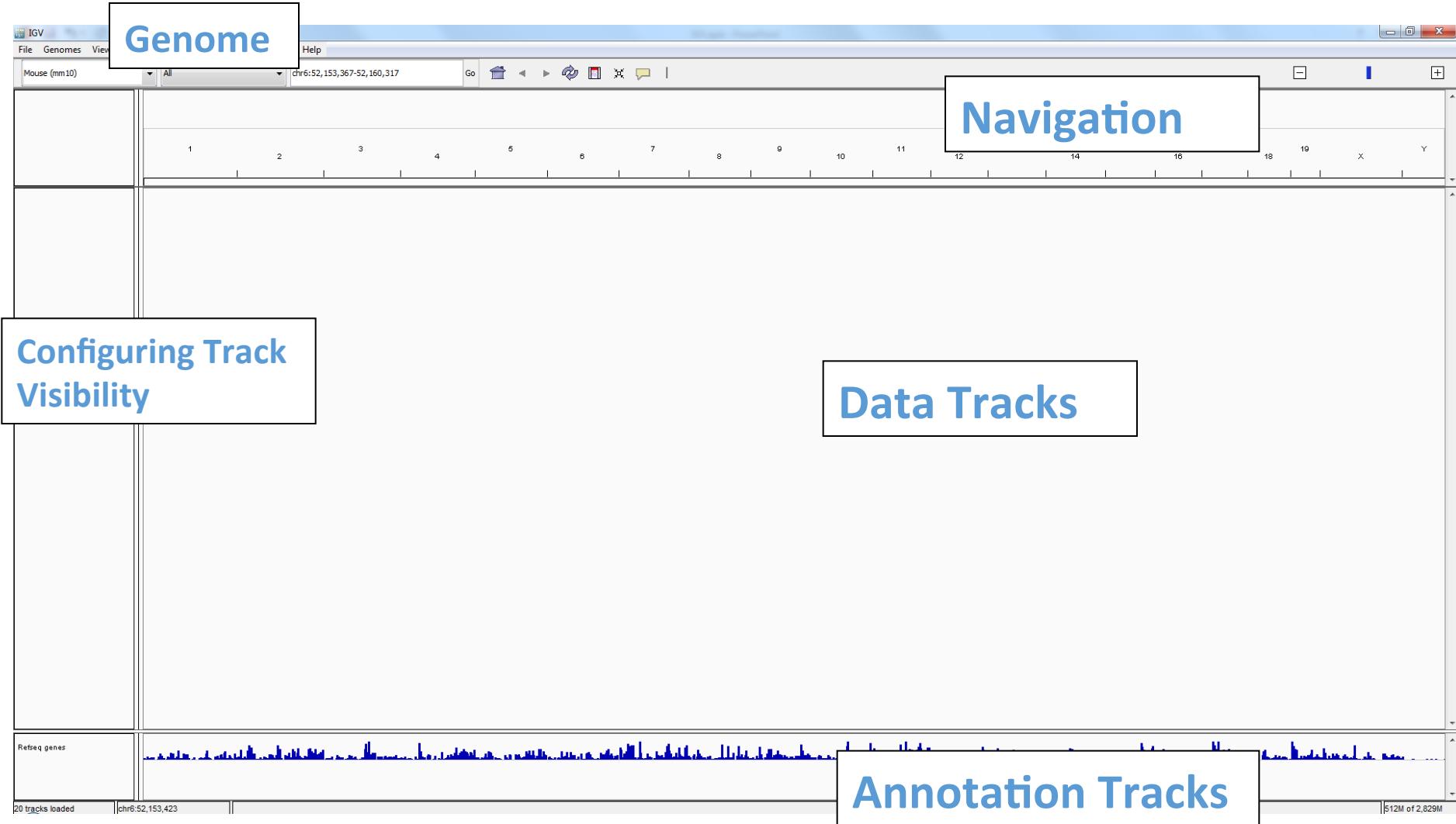
- First, a short presentation
 - Introduction to IGV
 - Files and Formats
 - Installation and Execution tips
- Mostly, a hands-on workshop
 - Navigation
 - Loading Data
 - Visualization Options
 - Saving Sessions
 - Scripting

What is IGV?

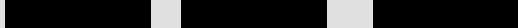
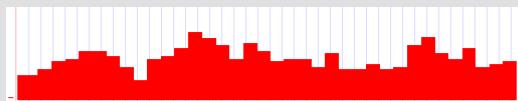


- Integrative Genomics Viewer
- Desktop genome browser – to view genomic data in context
- Runs “locally” (on your computer or a server)
- Developed by James Robinson, et. al, Broad Institute

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 %*
```

My opinion!

What is IGV better at

(compared to other genome browsers)

- 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

My opinion!

What is IGV worse at

(compared to other genome browsers)

- 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

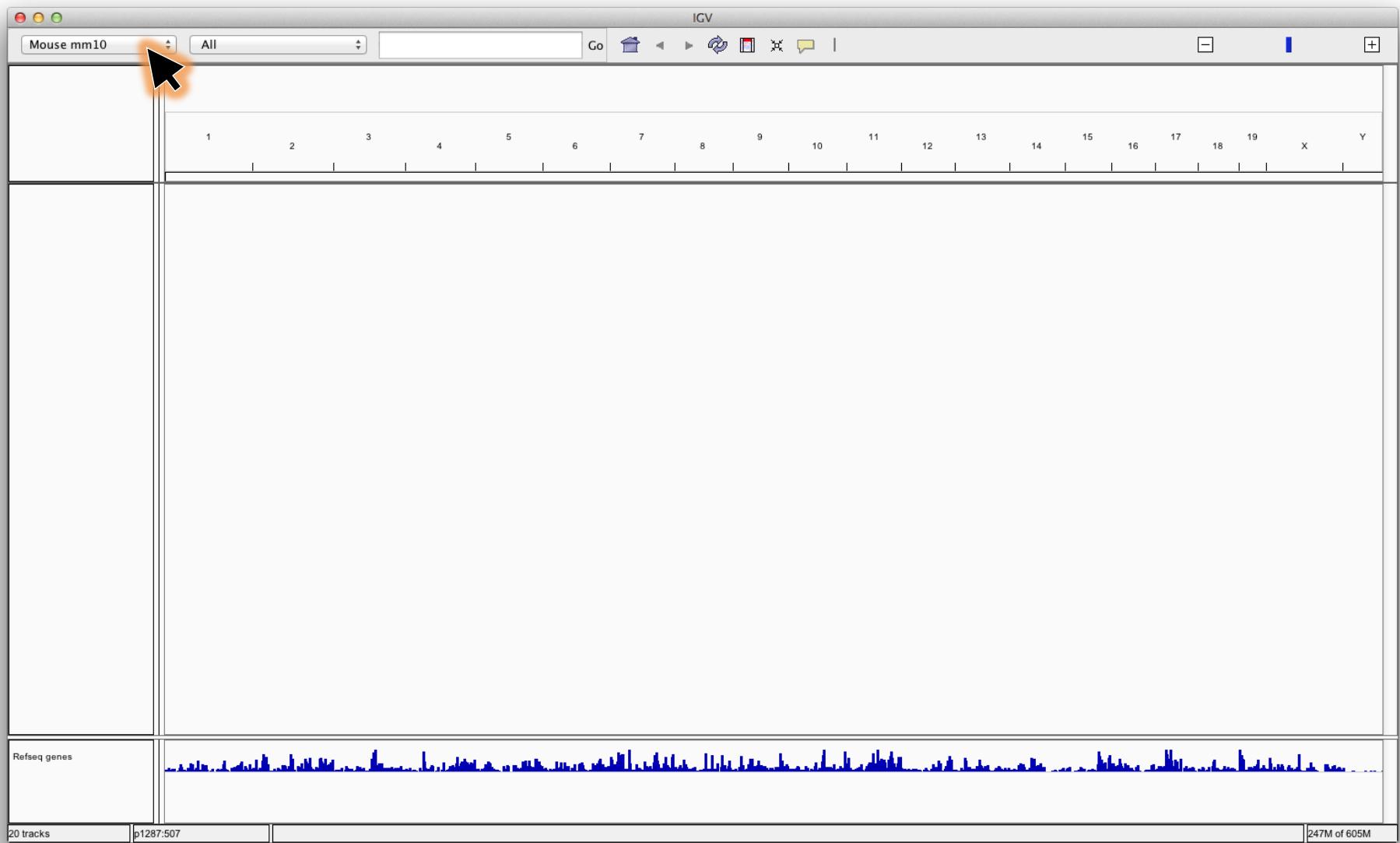
Alternatives to Consider:

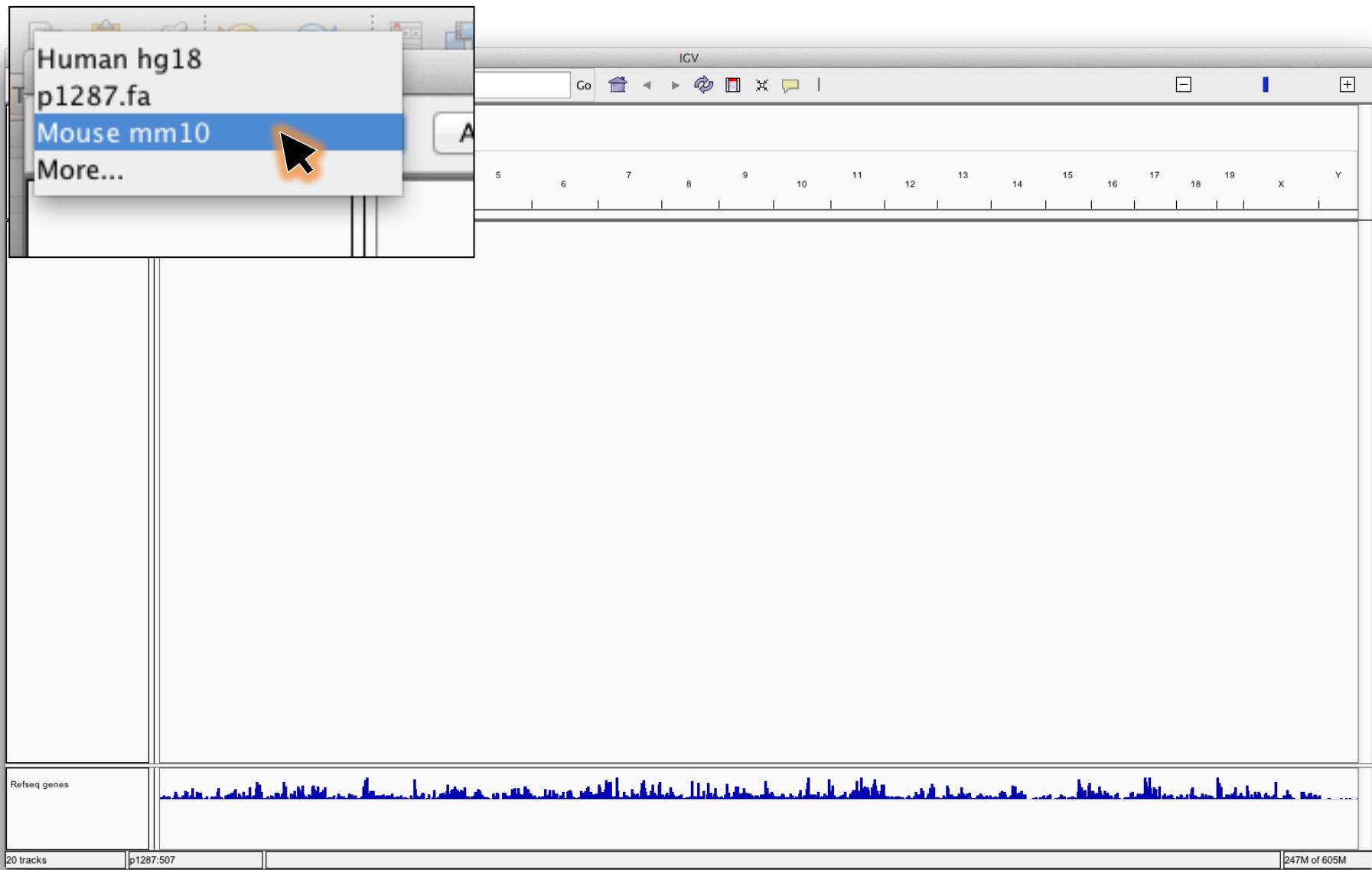
UCSC genome browser, Gbrowse/Jbrowse, IGB, Circos, R-based genome plotting packages (ggbio, GenomeGraphs)

Workshop Time!



Select the genome - Mouse (mm10)





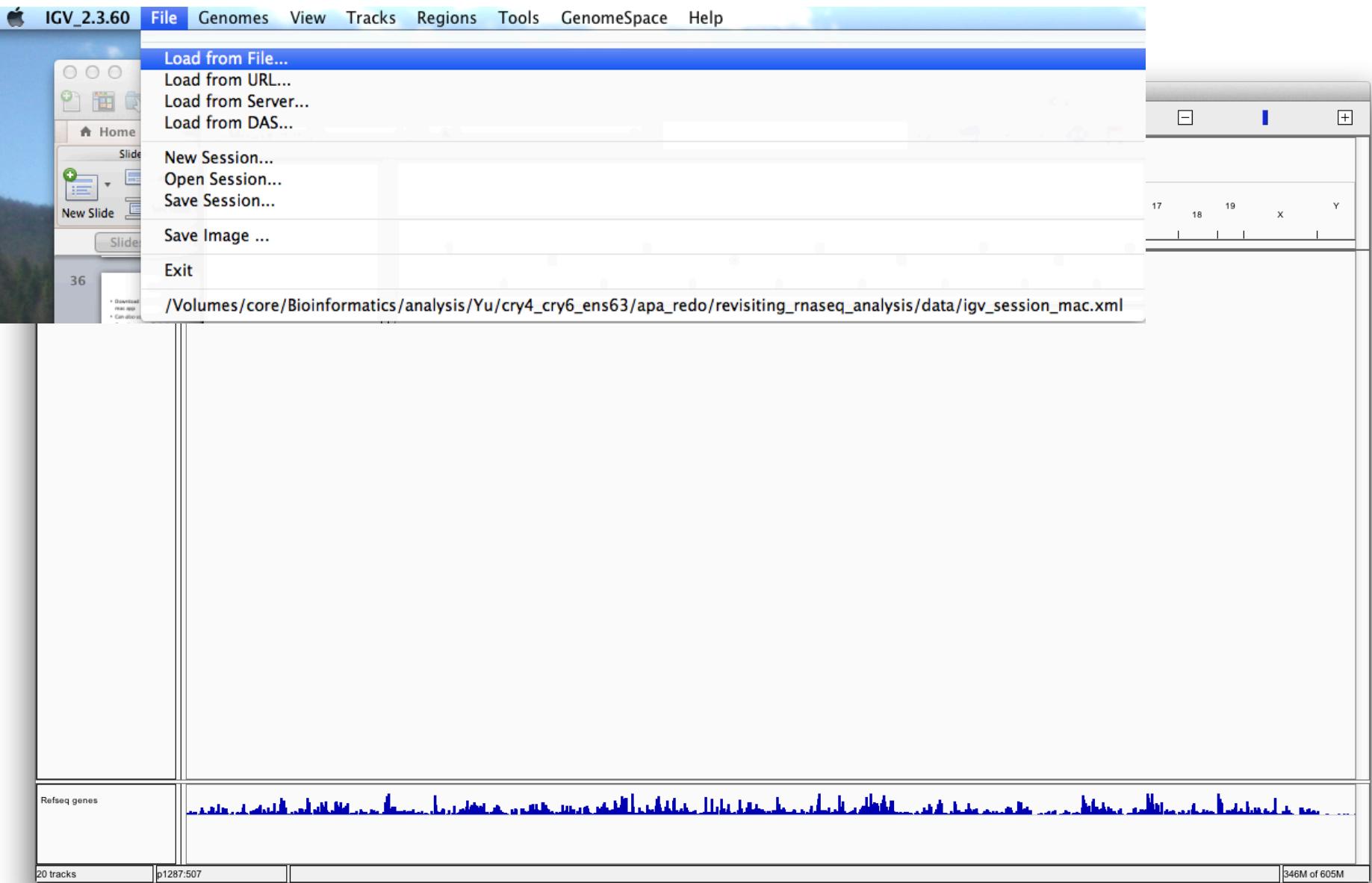
About the data we'll use...

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

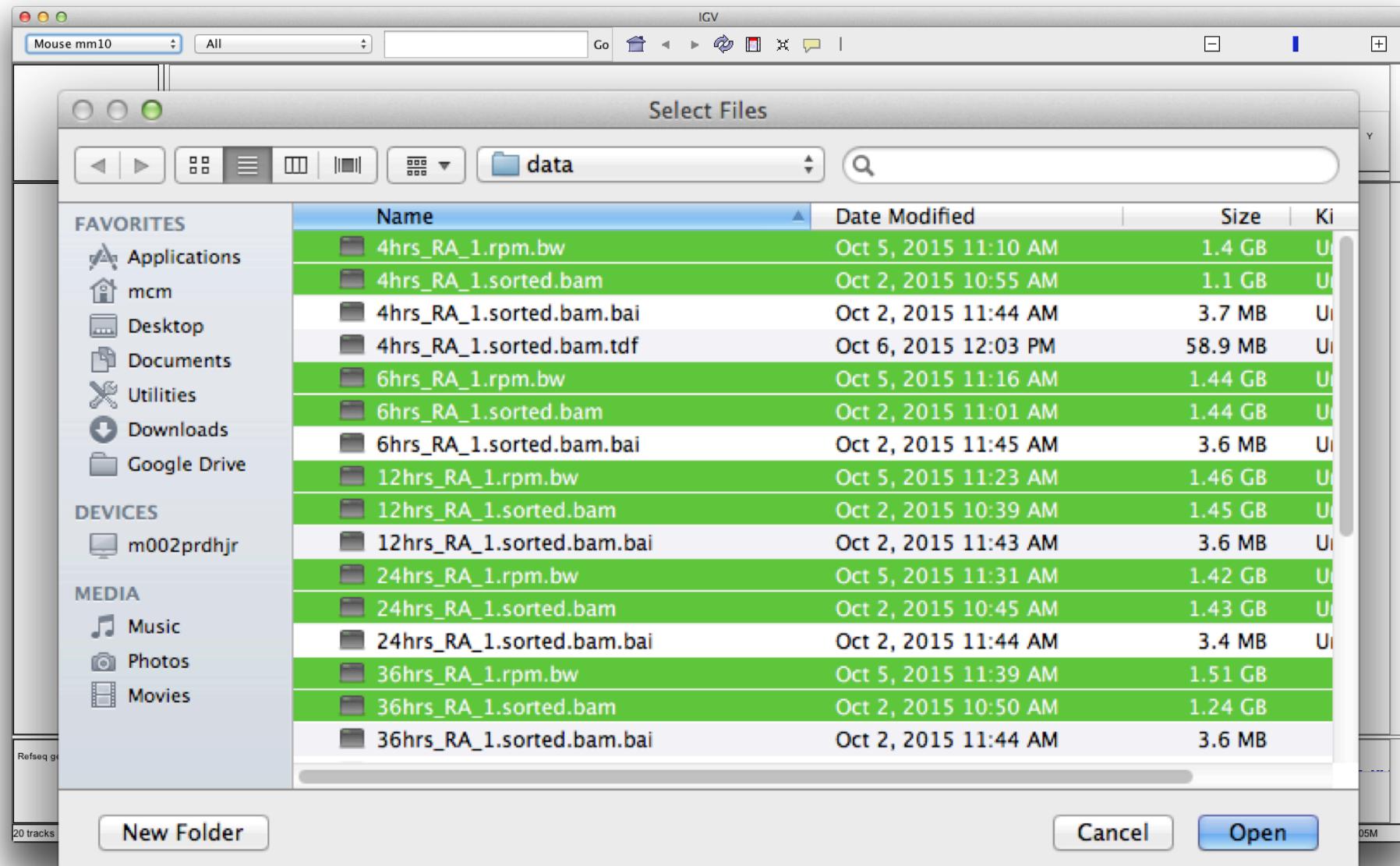
Bony De Kumar¹, Mark E. Parrish¹, Brian D Slaughter¹, Jay R Unruh¹, Madelaine Gogol¹, Christopher Seidel¹, Ariel Paulson¹, Hua Li¹, Karin Gaudenz¹, Allison Peak¹, William McDowell¹, Brian Fleharty¹, Youngwook Ahn¹, Chengqi Lin^{1,4}, Edwin Smith^{1,5}, Ali Shilatifard^{1,5} and Robb Krumlauf^{1,2,3}

GSE67610

Let's load some data!

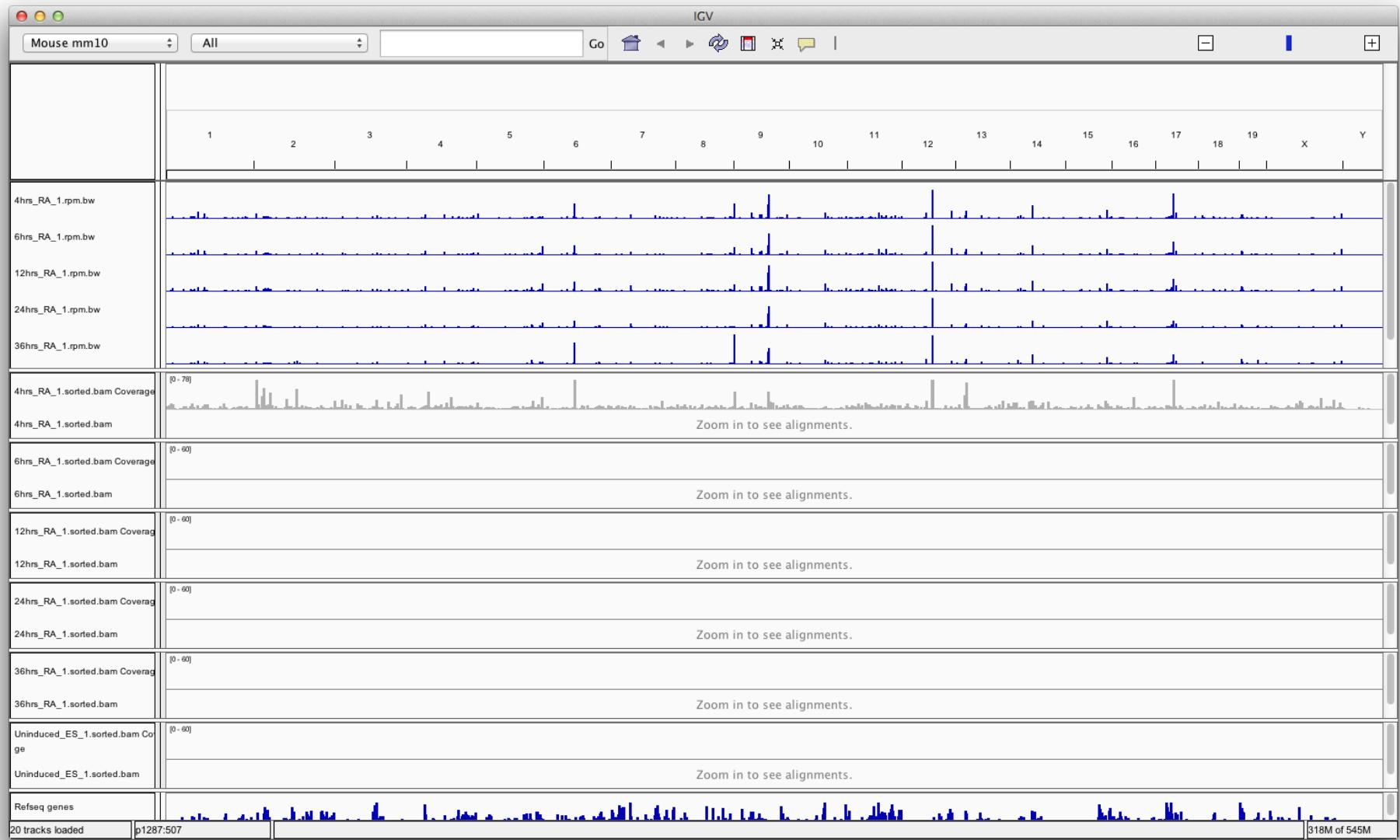


Browse to folder with files, select all bam and bw files (not bai)...

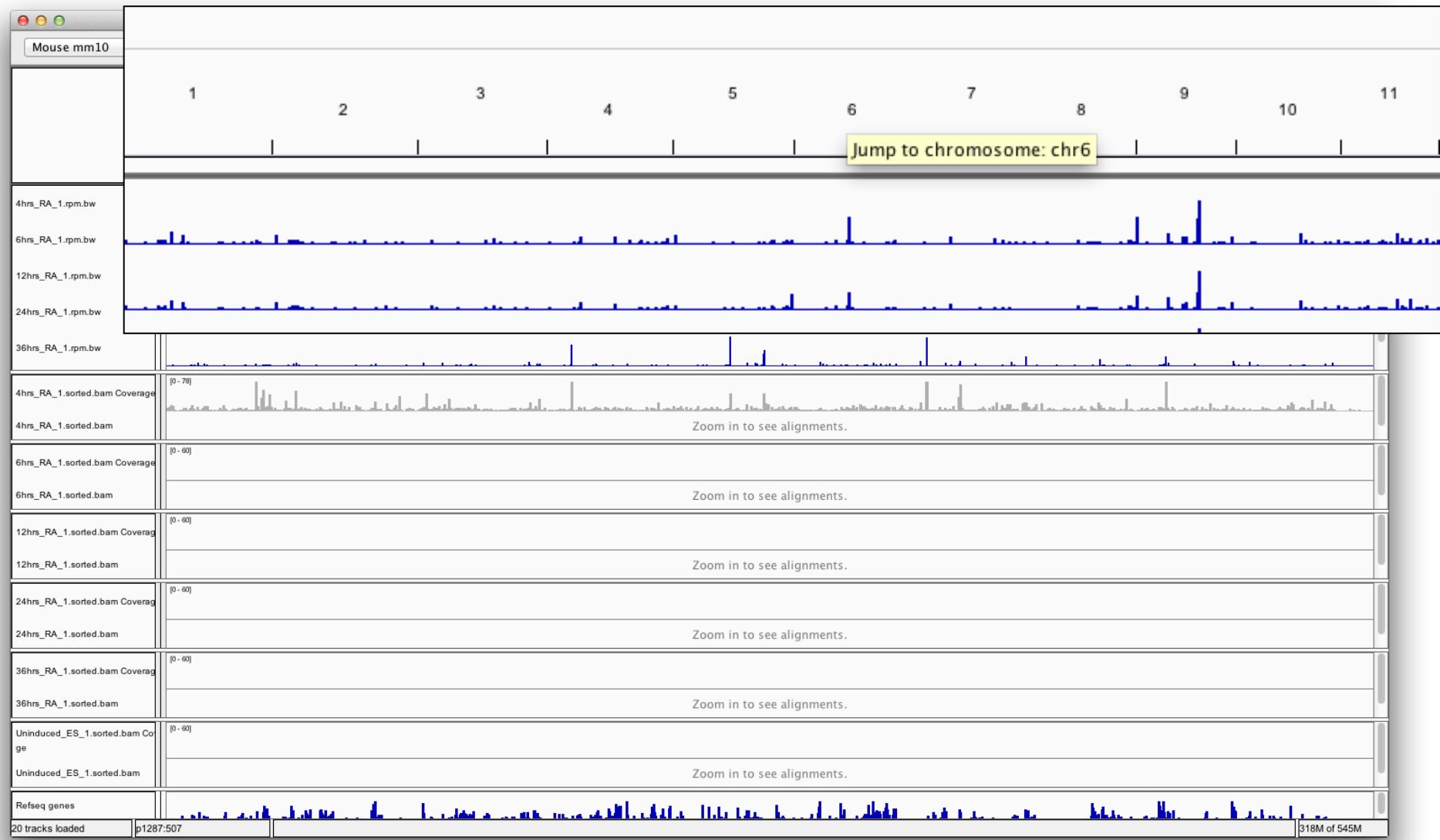


Hold down command to select multiple files...

Genomic Bird's Eye View (all chromosomes)



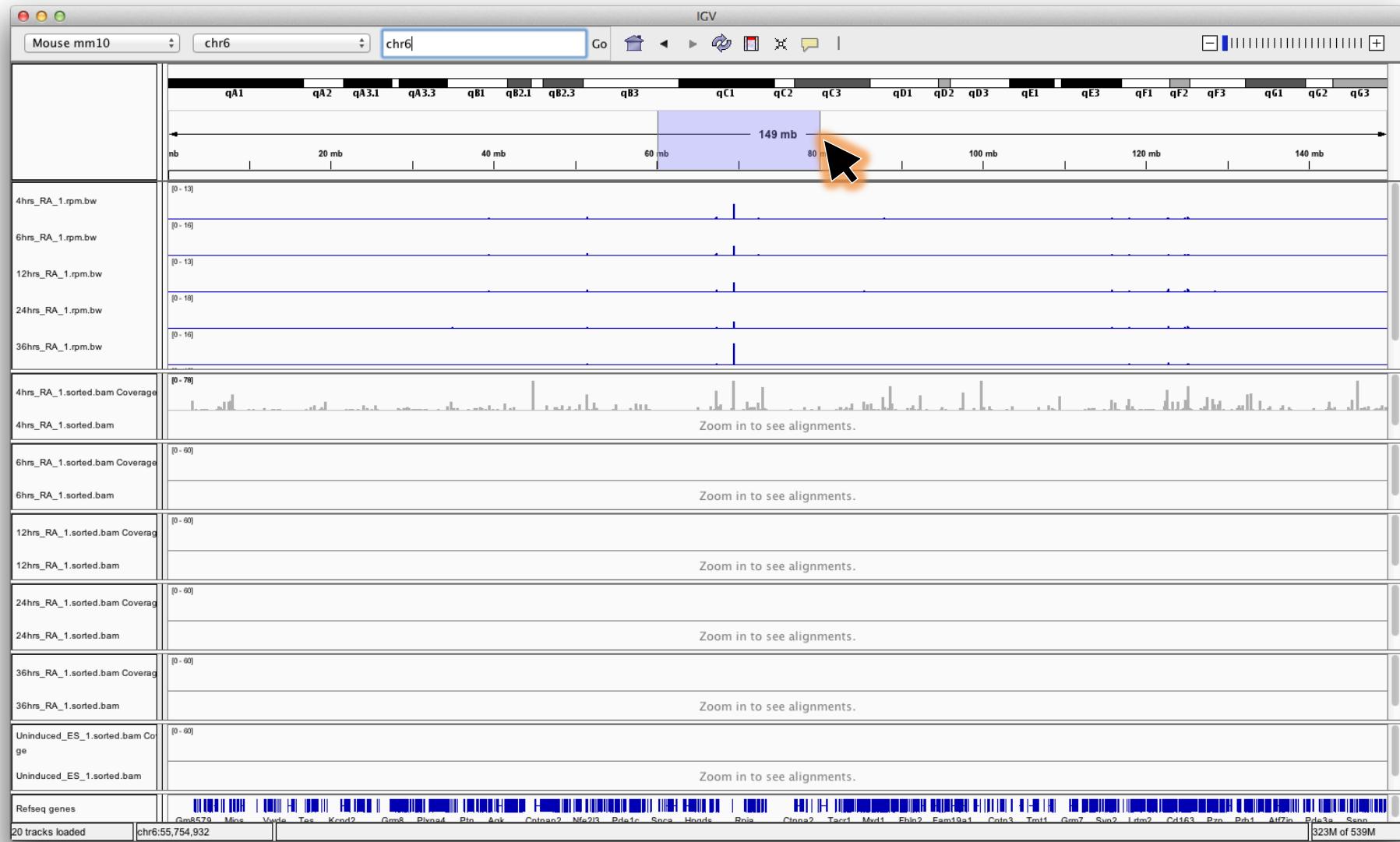
Click on the 6 to jump to chromosome 6.



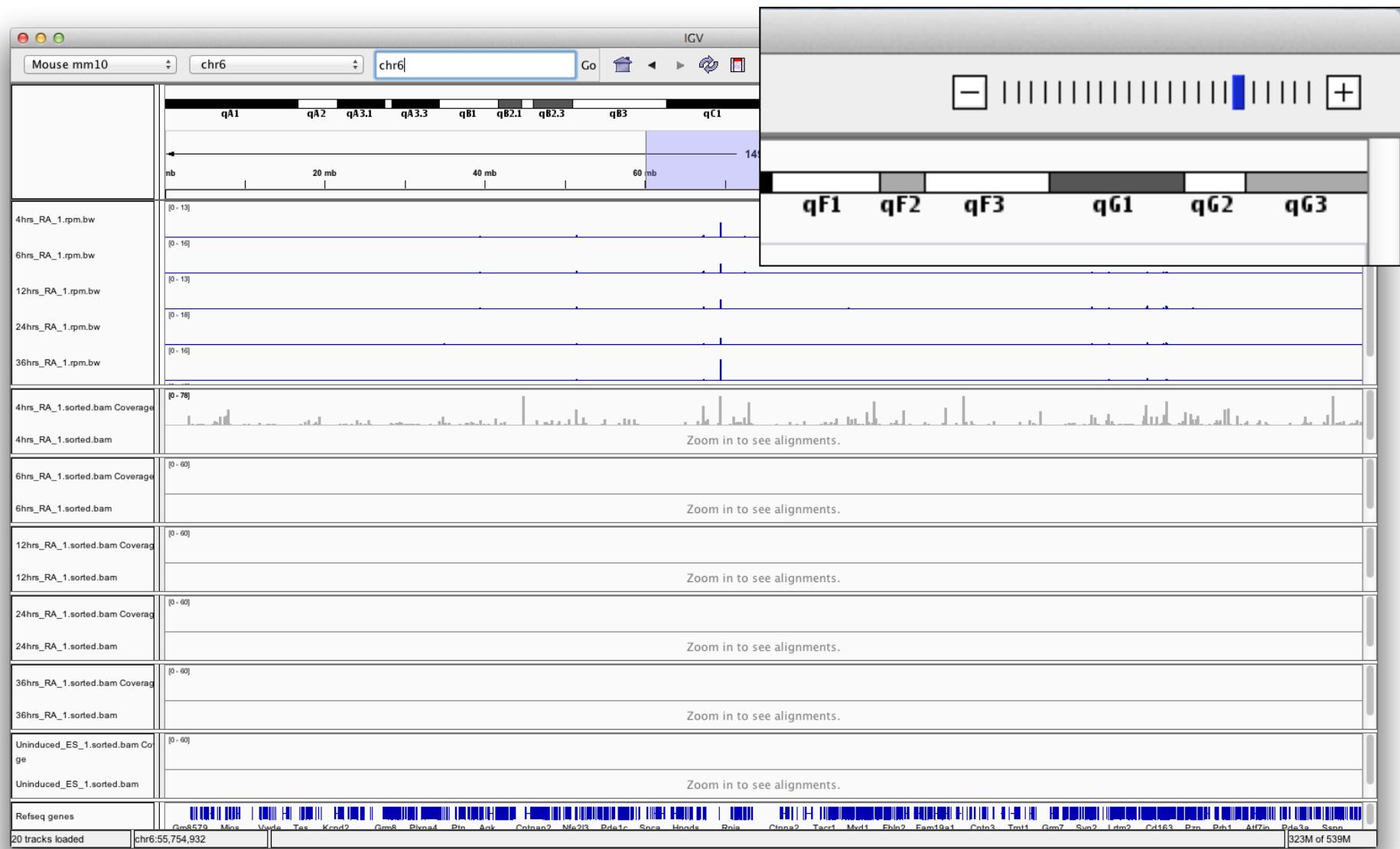
Chromosome Level View



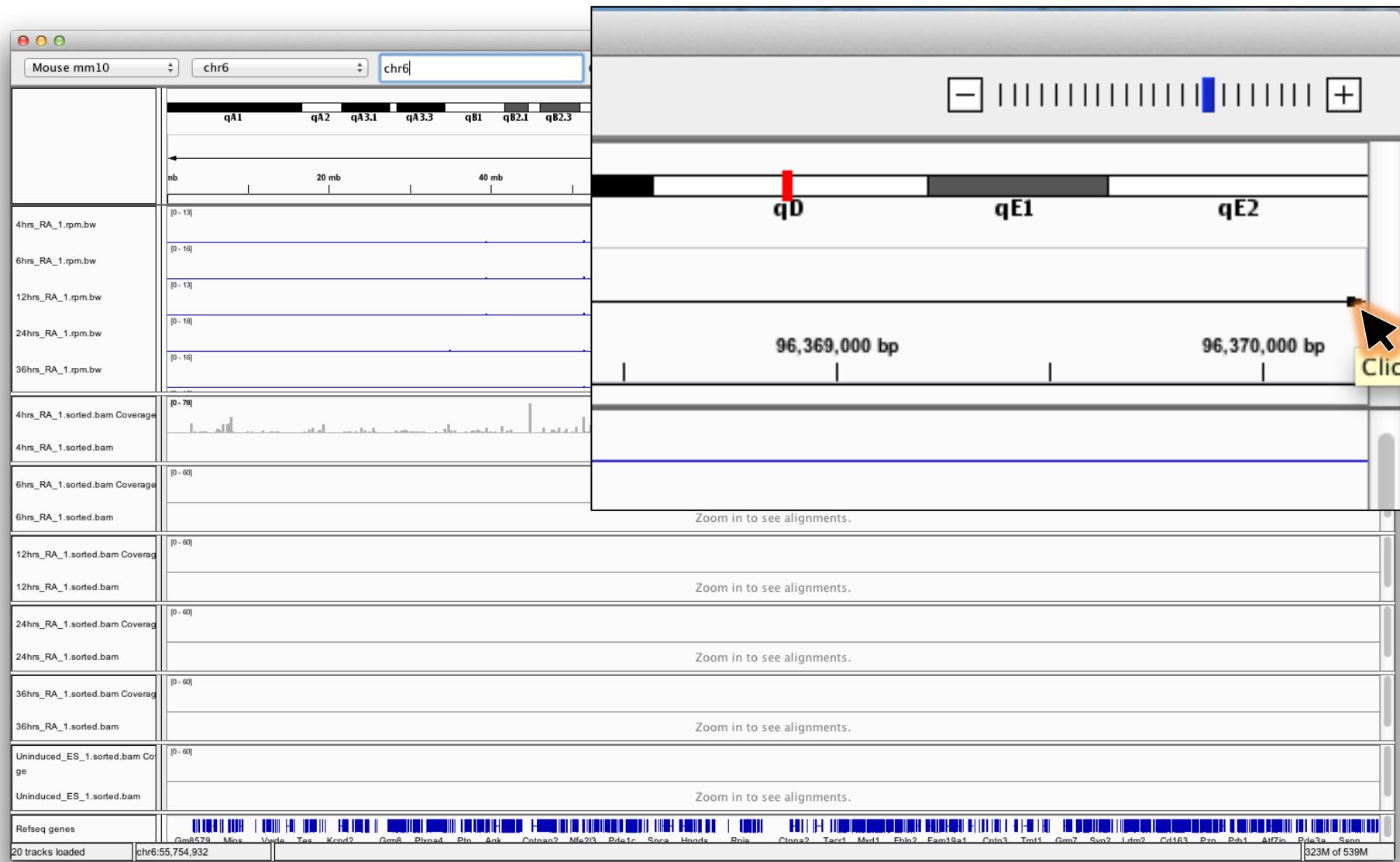
Click and drag to zoom in on a region of chromosome 6.



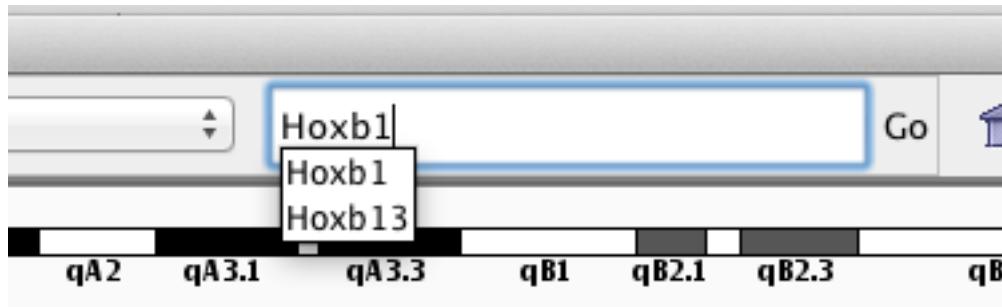
Keep click and dragging or use railroad to zoom until you can see some alignments.



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



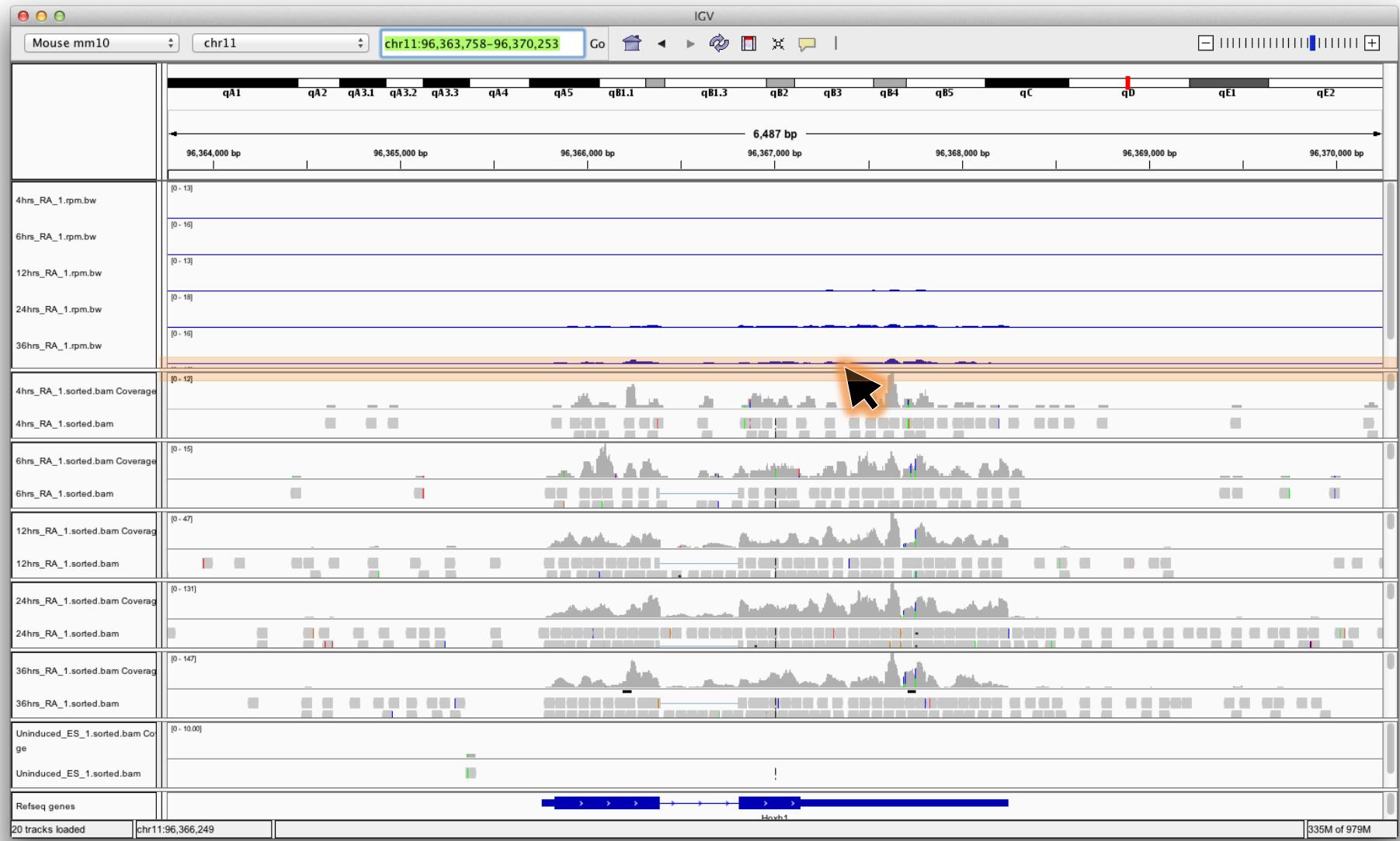
Try typing a mouse gene name in the search box and navigate to it. Are there any reads aligning there?



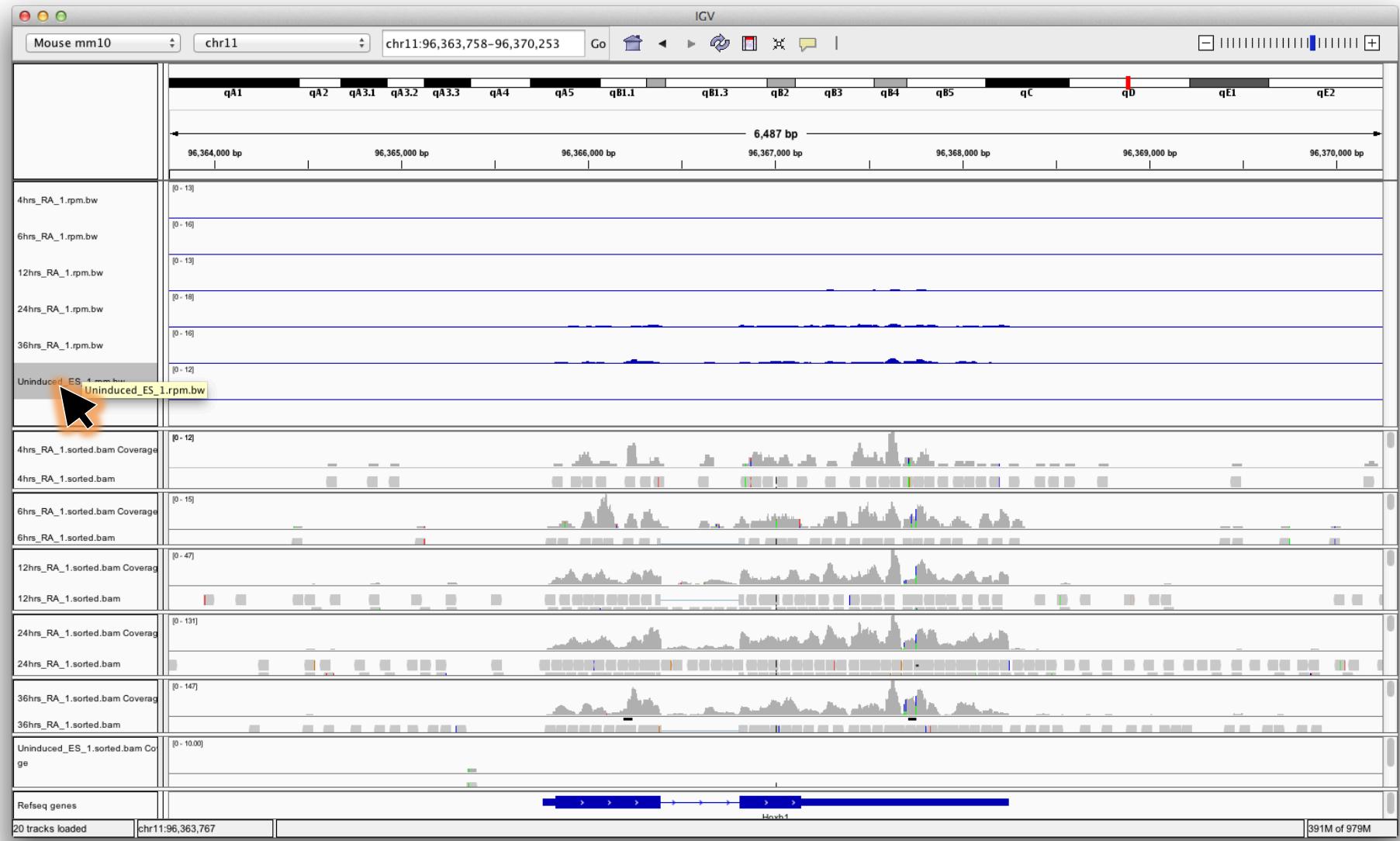
By the way, BAM files get two tracks each!



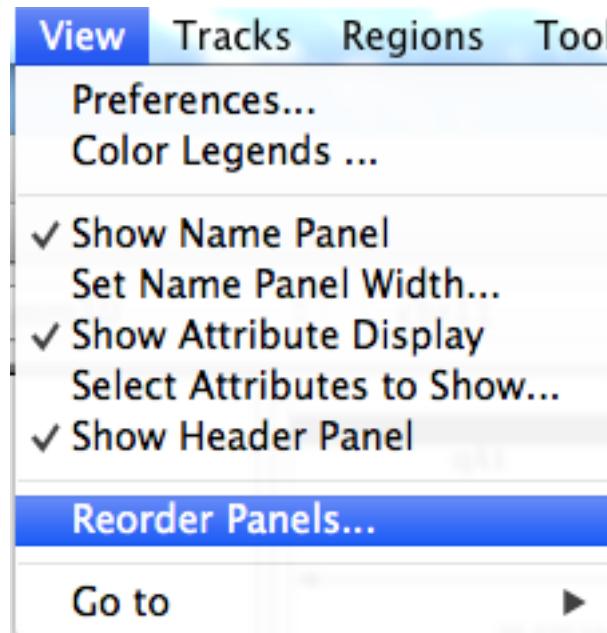
To quickly resize track panels, hover over dividers, click and drag.



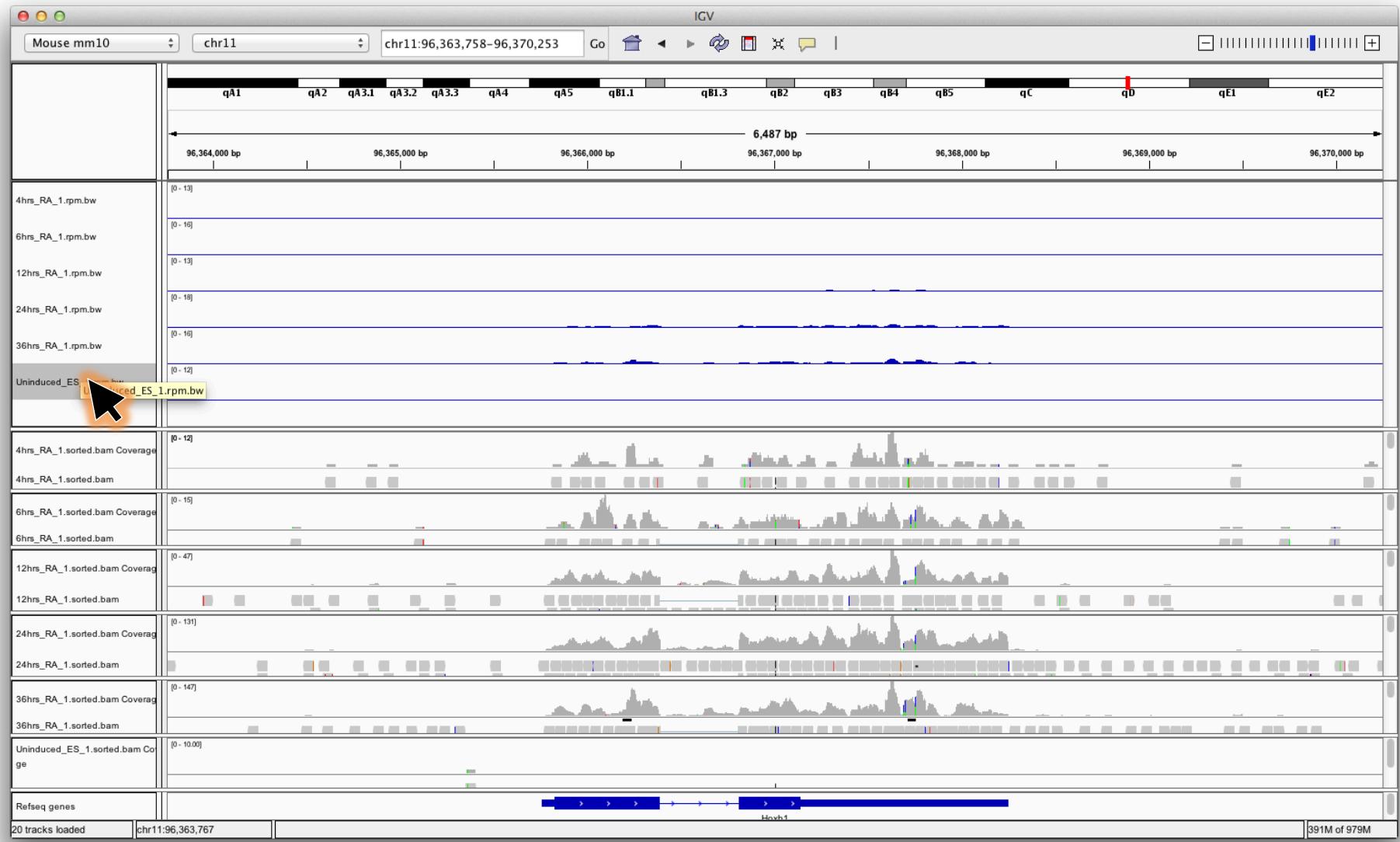
To reorder tracks, click and drag the track name on the left side.



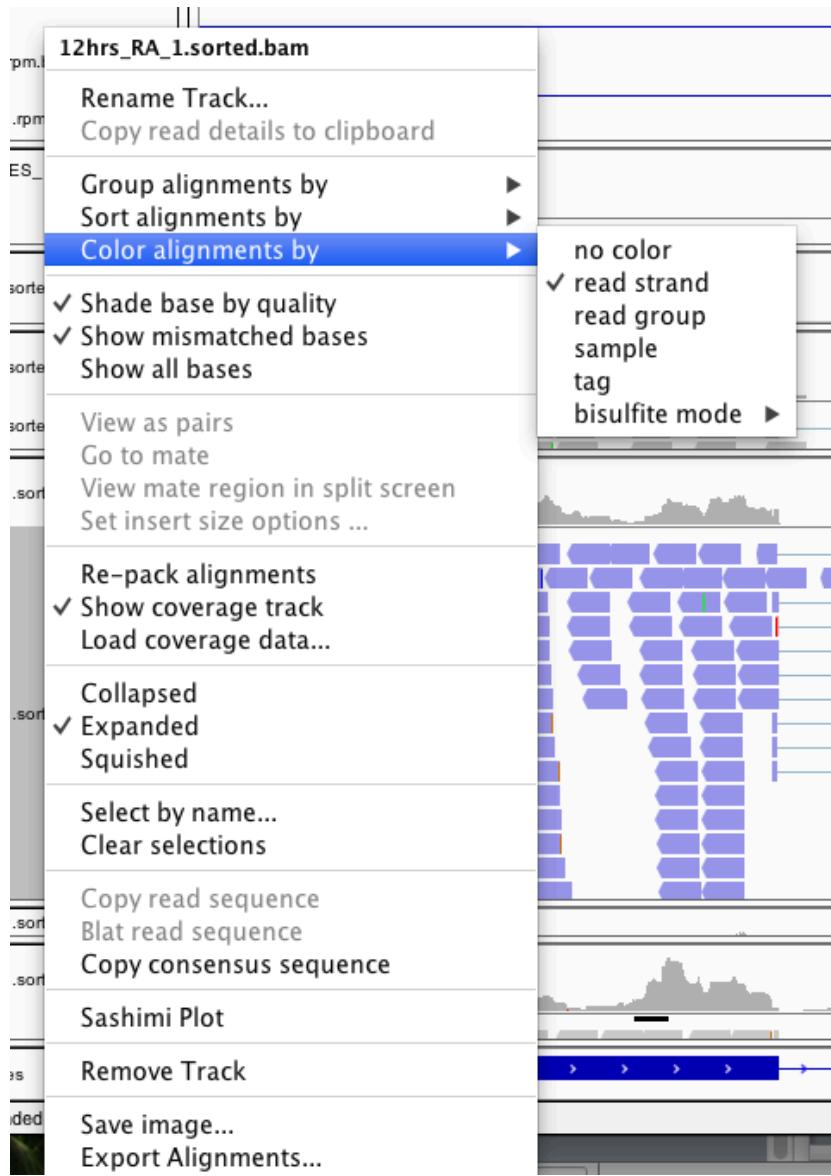
To reorder “panels” easier – View menu – Reorder Panels...



Hold command and click to select multiple tracks at once.



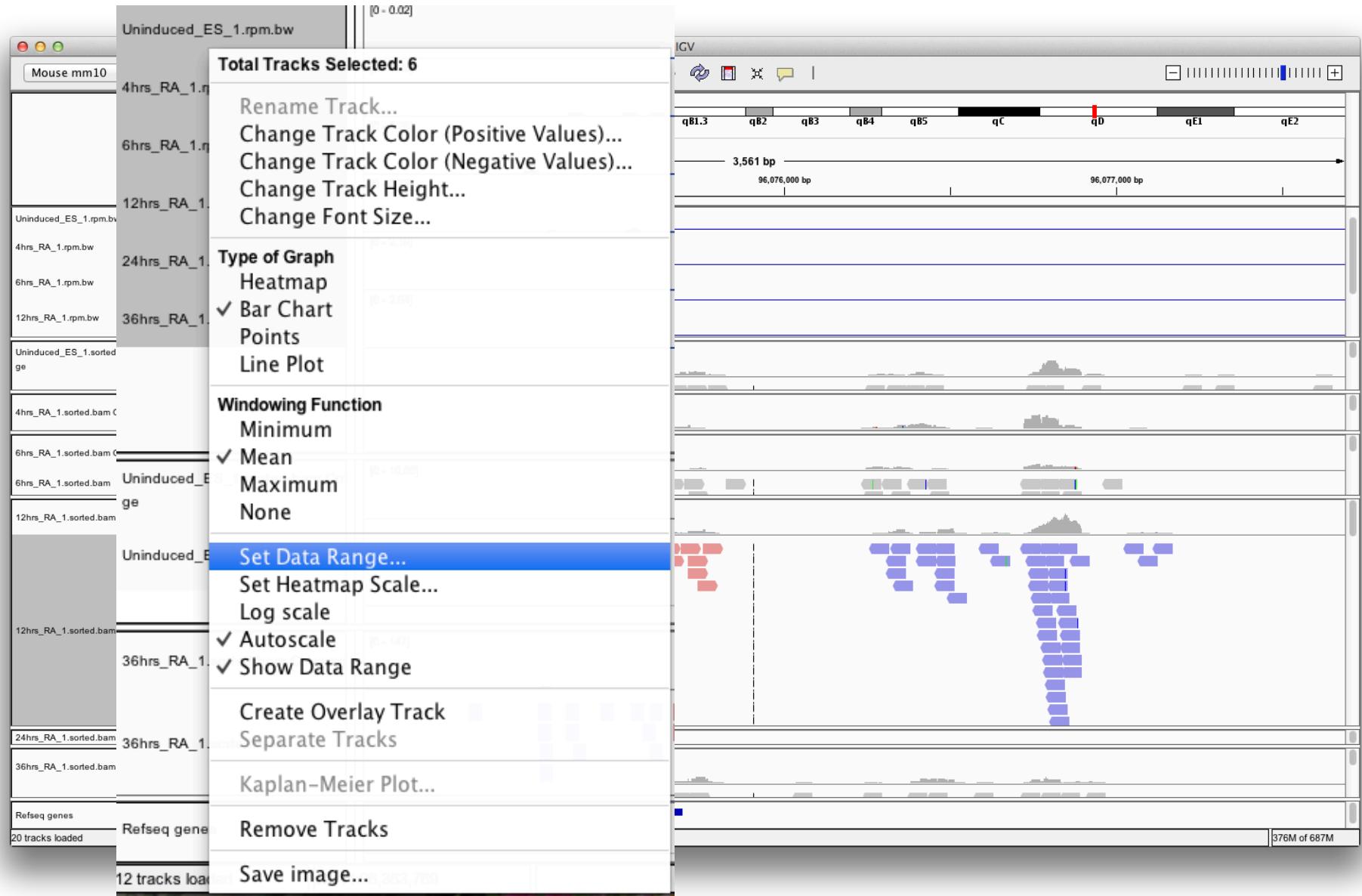
Right click on track name to change a bunch of stuff...



Color alignments by strand

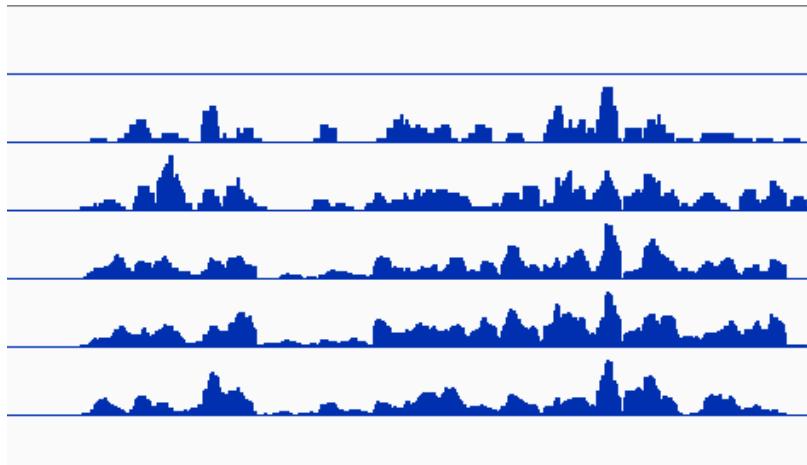


Scale

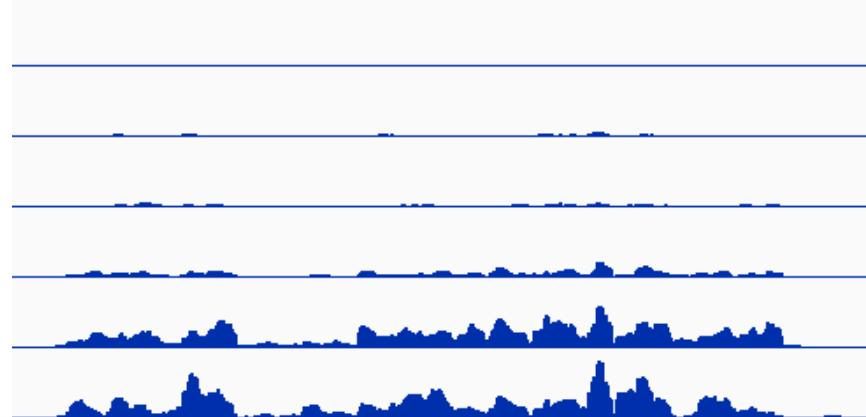


Know your scale...

Autoscale



Set Data Range



You try it!

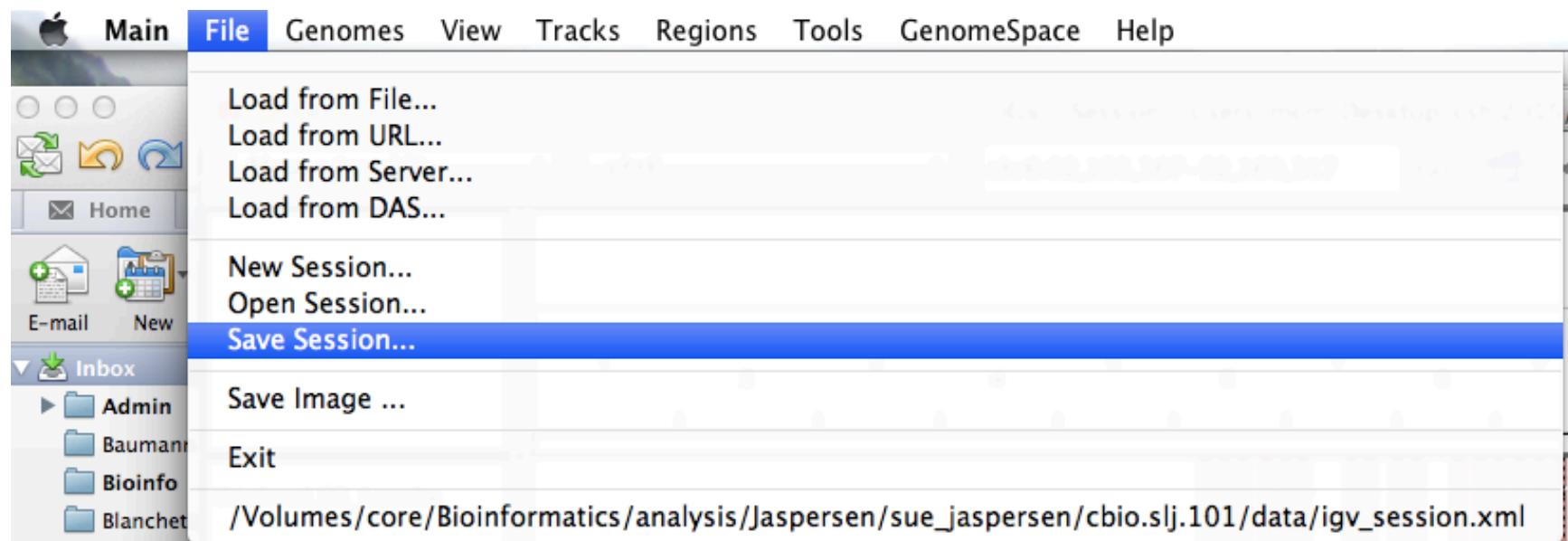
Exploring Hoxa1

- 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?

Saving Sessions

- Once you have all the data loaded and looking the way you like, you can save a session
- Loading the session when you (or someone else) starts IGV will load your data and settings.

File, Save Session...

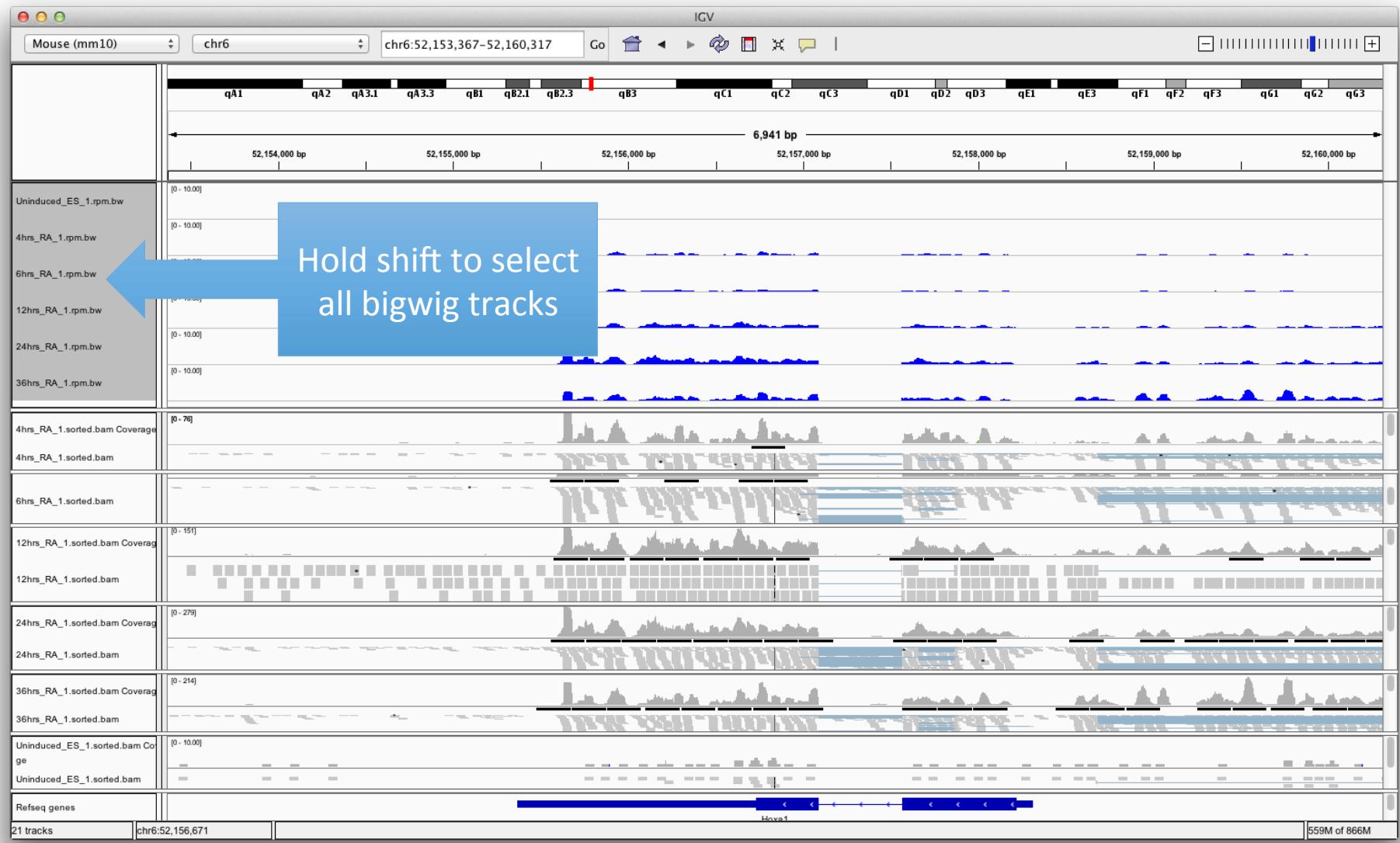


You try it!

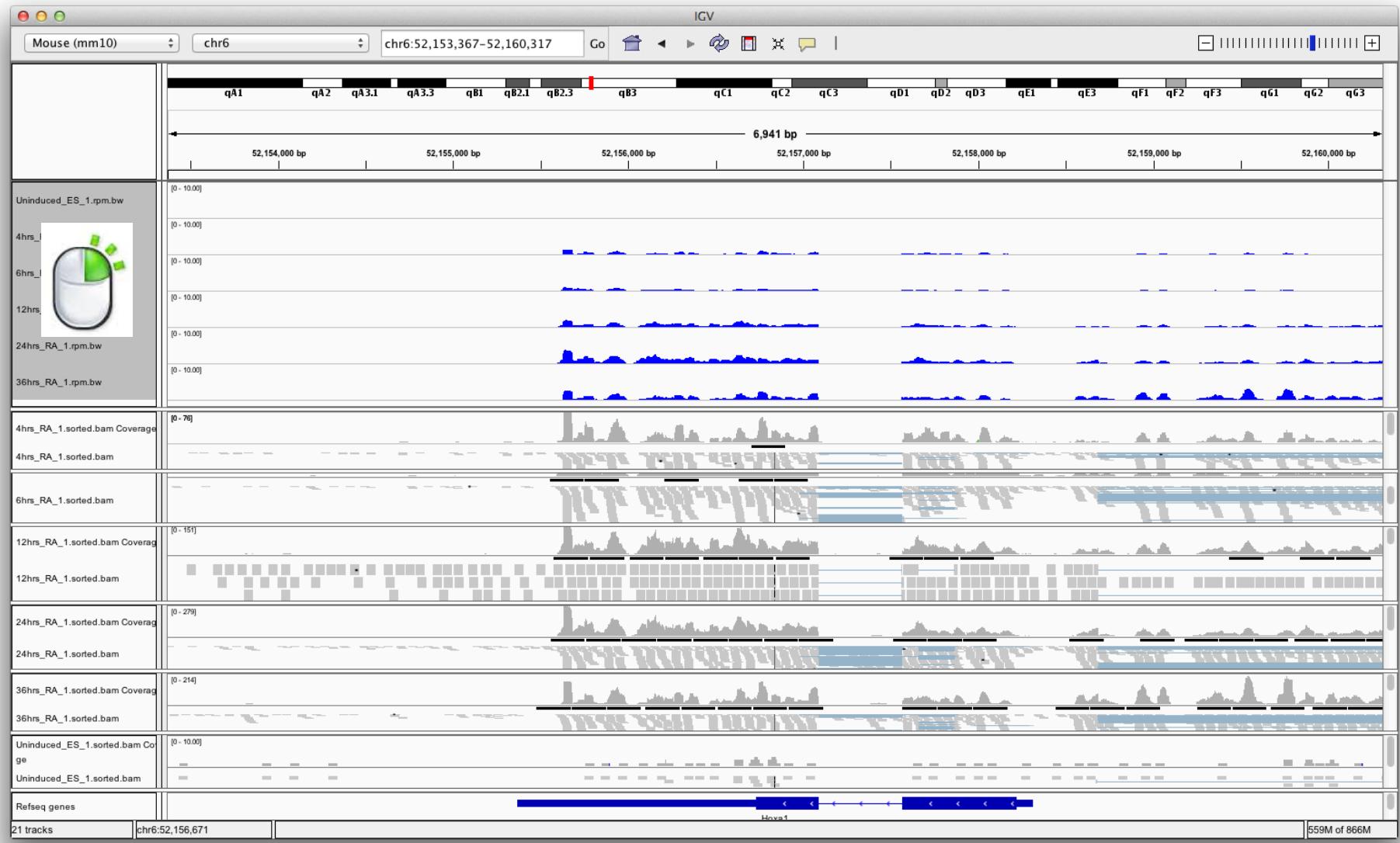
Now, change some stuff...

- Change location, visualization settings, colors...
- Then save a new session under a different name
- Then Open your first session again!

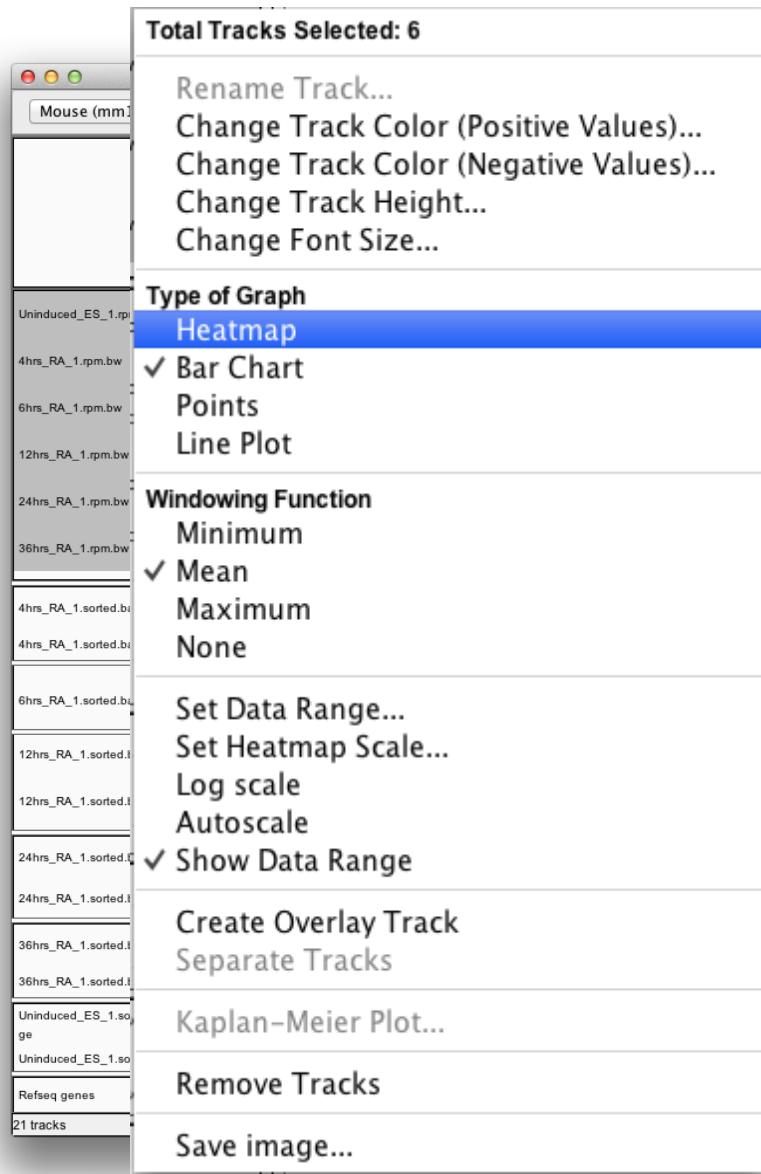
From barplot to heatmap



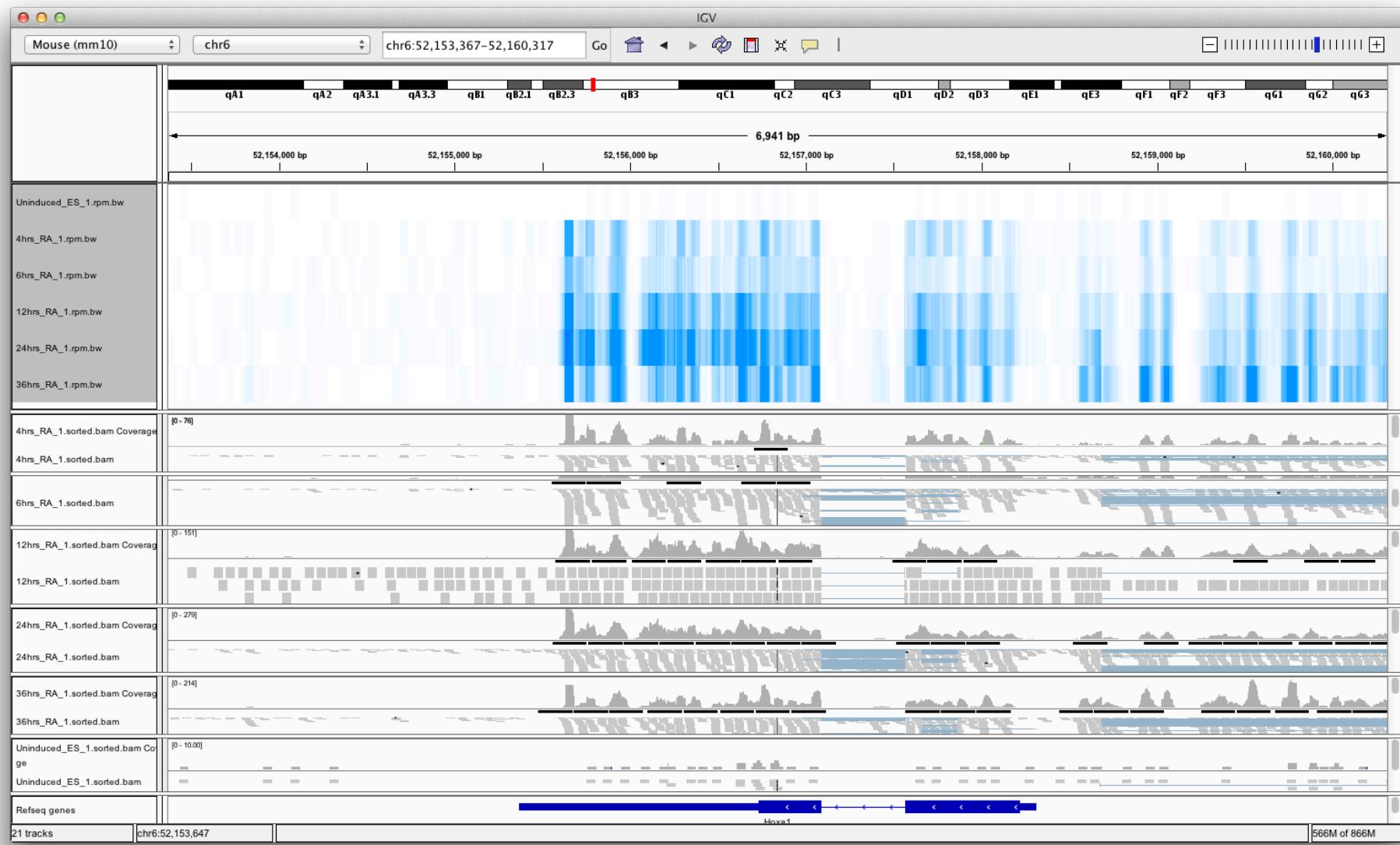
Shift to select all, right click



Select “Heatmap”



Heatmap view



Saving Images

IGV_2.3.60

File Genomes View Tracks Regions Tools GenomeSpace Help

- Load from File...
- Load from URL...
- Load from Server...
- Load from DAS...
- New Session...
- Open Session...
- Save Session...
- Save Image ...**
- Exit

/Volumes/core/Bioinformatics/analysis/Yu/cry4_cry6_ens63/apa_redo/revisiting_rnaseq_analysis/data/igv_session_mac.xml

Uninduced_ES_1.rpm.bw

4hrs_RA_1.rpm.bw

6hrs_RA_1.rpm.bw

12hrs_RA_1.rpm.bw

24hrs_RA_1.rpm.bw

36hrs_RA_1.rpm.bw

4hrs_RA_1.sorted.bam Coverage [0 - 76]

4hrs_RA_1.sorted.bam

6hrs_RA_1.sorted.bam

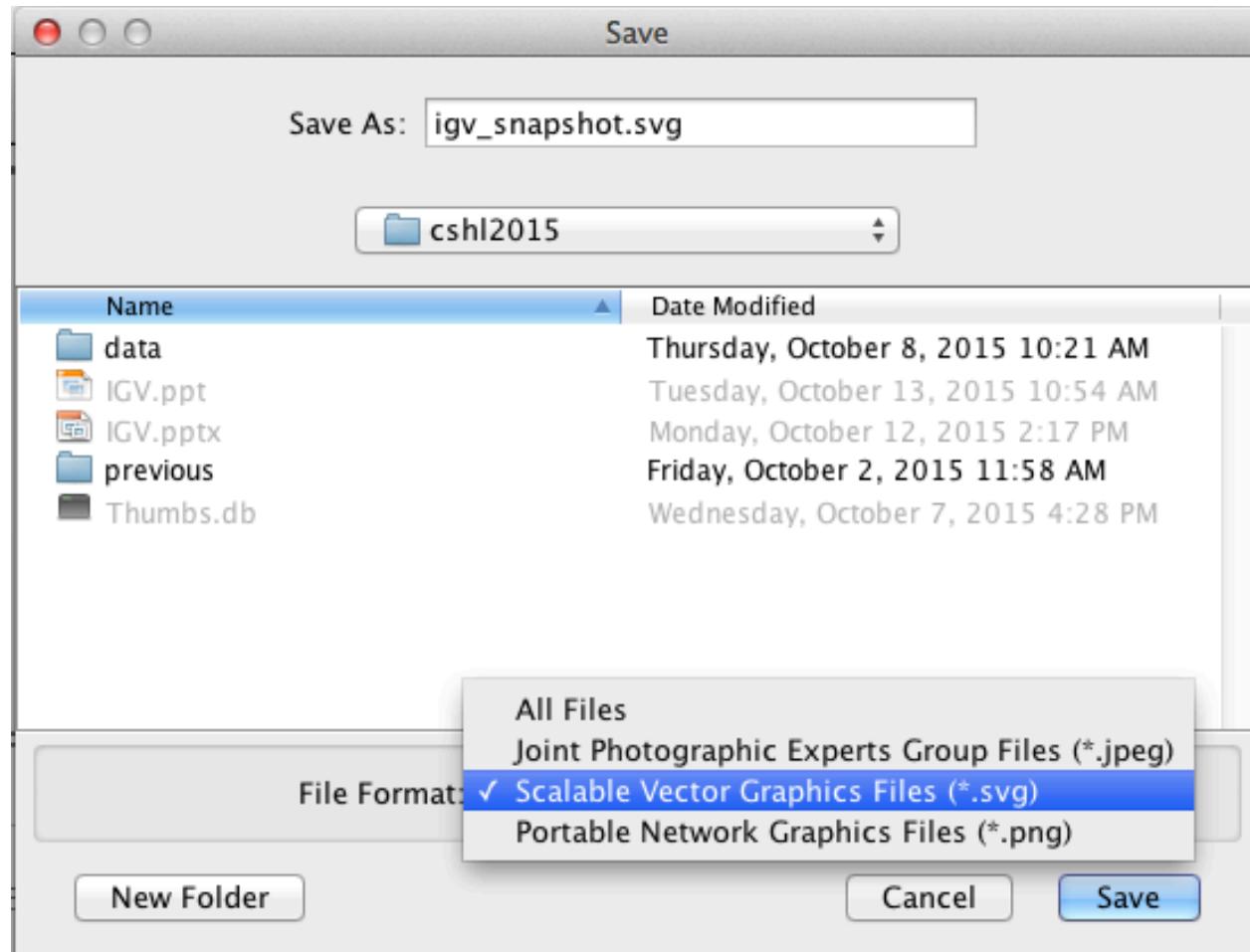
12hrs_RA_1.sorted.bam Coverage [0 - 151]

12hrs_RA_1.sorted.bam

The screenshot shows the IGV (Integrating Genome Viewer) software interface version 2.3.60. The main window displays a genomic track with blue vertical bars representing data across a genomic region from 52,158,000 bp to 52,159,000 bp. Below this are several tracks showing coverage levels. On the left, a sidebar lists sample names: Uninduced_ES_1.rpm.bw, 4hrs_RA_1.rpm.bw, 6hrs_RA_1.rpm.bw, 12hrs_RA_1.rpm.bw, 24hrs_RA_1.rpm.bw, and 36hrs_RA_1.rpm.bw. Another section shows coverage tracks for 4hrs_RA_1.sorted.bam, 6hrs_RA_1.sorted.bam, and 12hrs_RA_1.sorted.bam, with coverage scales ranging from 0 to 76 and 0 to 151 respectively. A context menu is open at the top, with 'Save Image ...' selected. The full path to the session XML file is visible in the menu: /Volumes/core/Bioinformatics/analysis/Yu/cry4_cry6_ens63/apa_redo/revisiting_rnaseq_analysis/data/igv_session_mac.xml.

Image type options

jpeg
png



Editing SVG

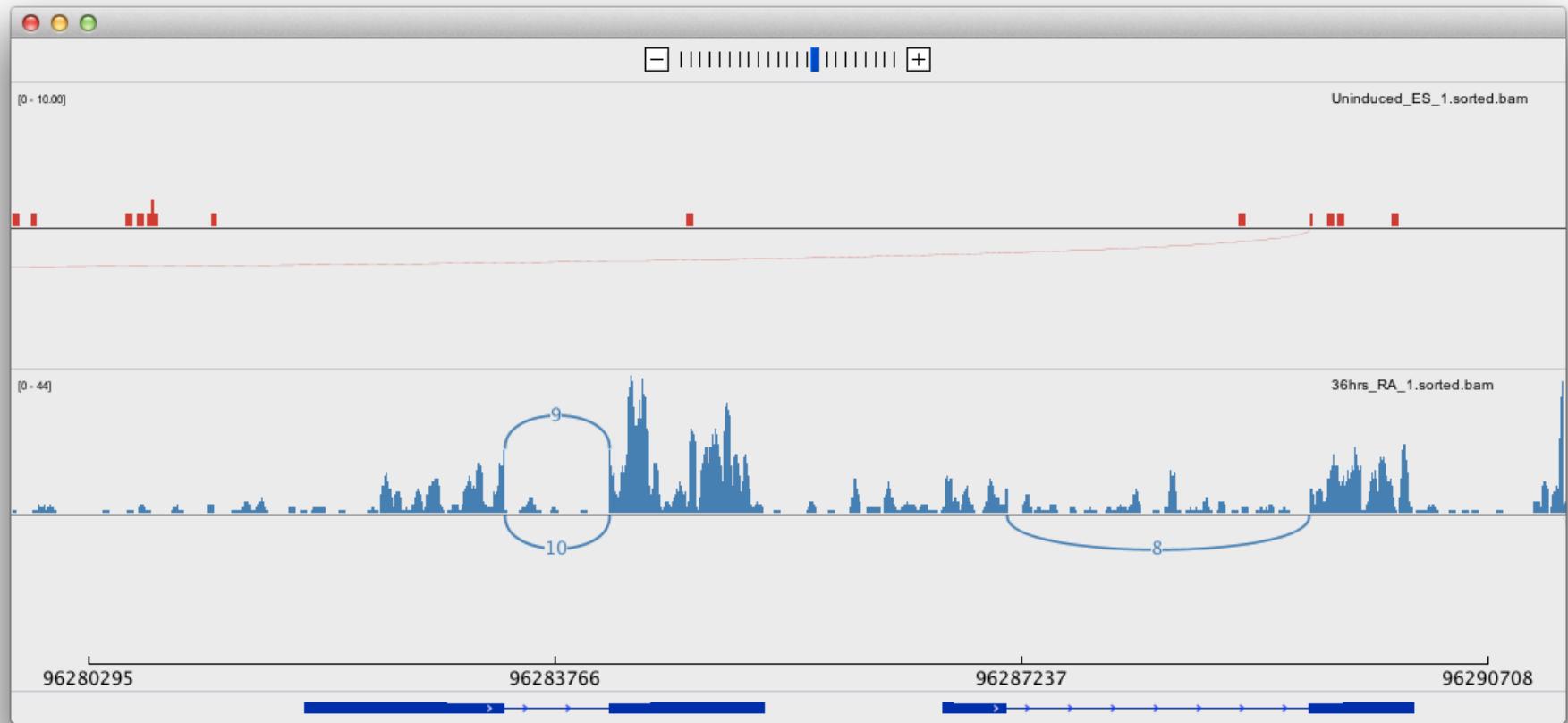
- Open in illustrator or inkscape (free, open source)
- Ungroup, edit individual elements



Sashimi Plot

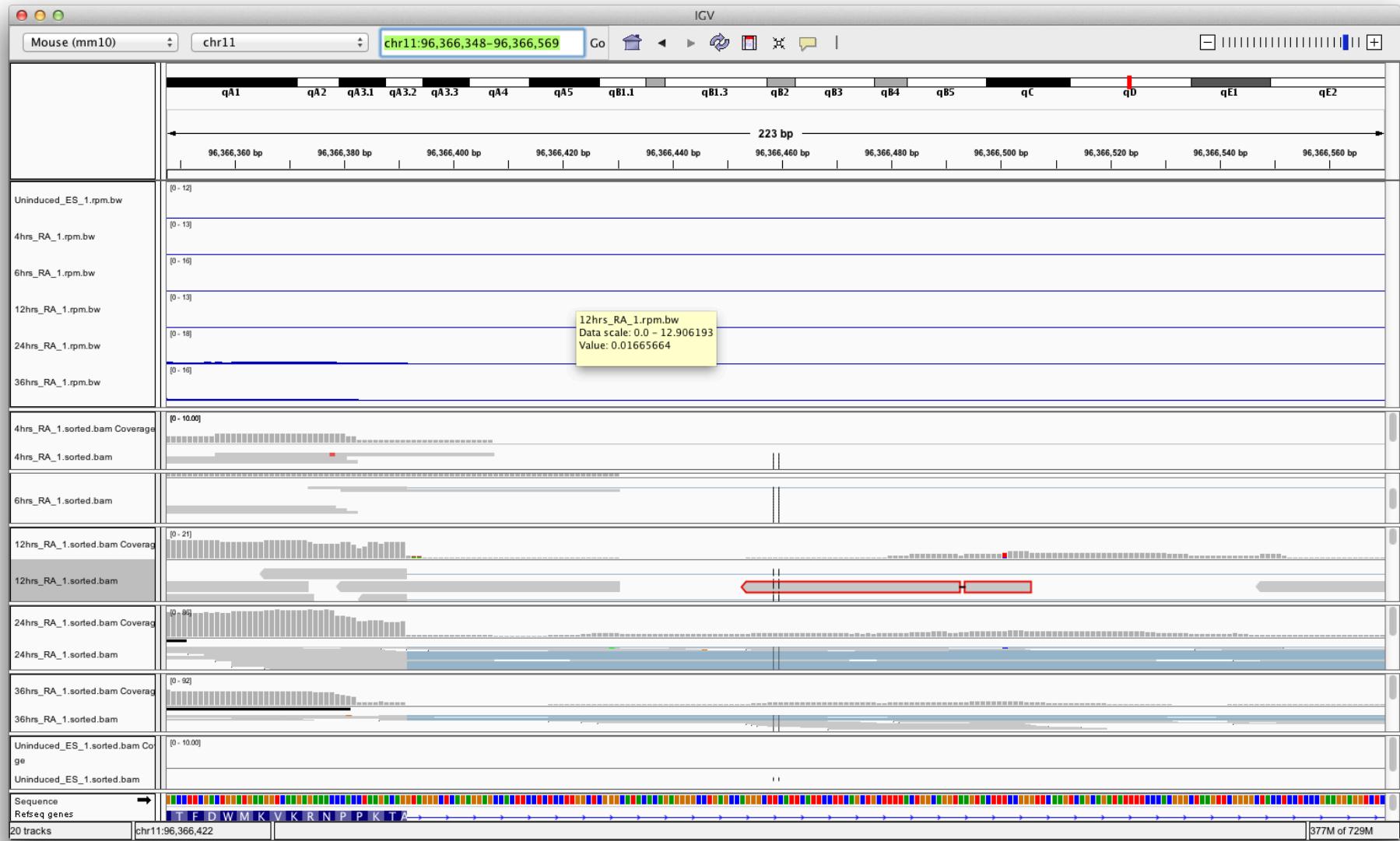
- Go to a gene, right click on bam data, “Sashimi plot”

Sashimi Plot

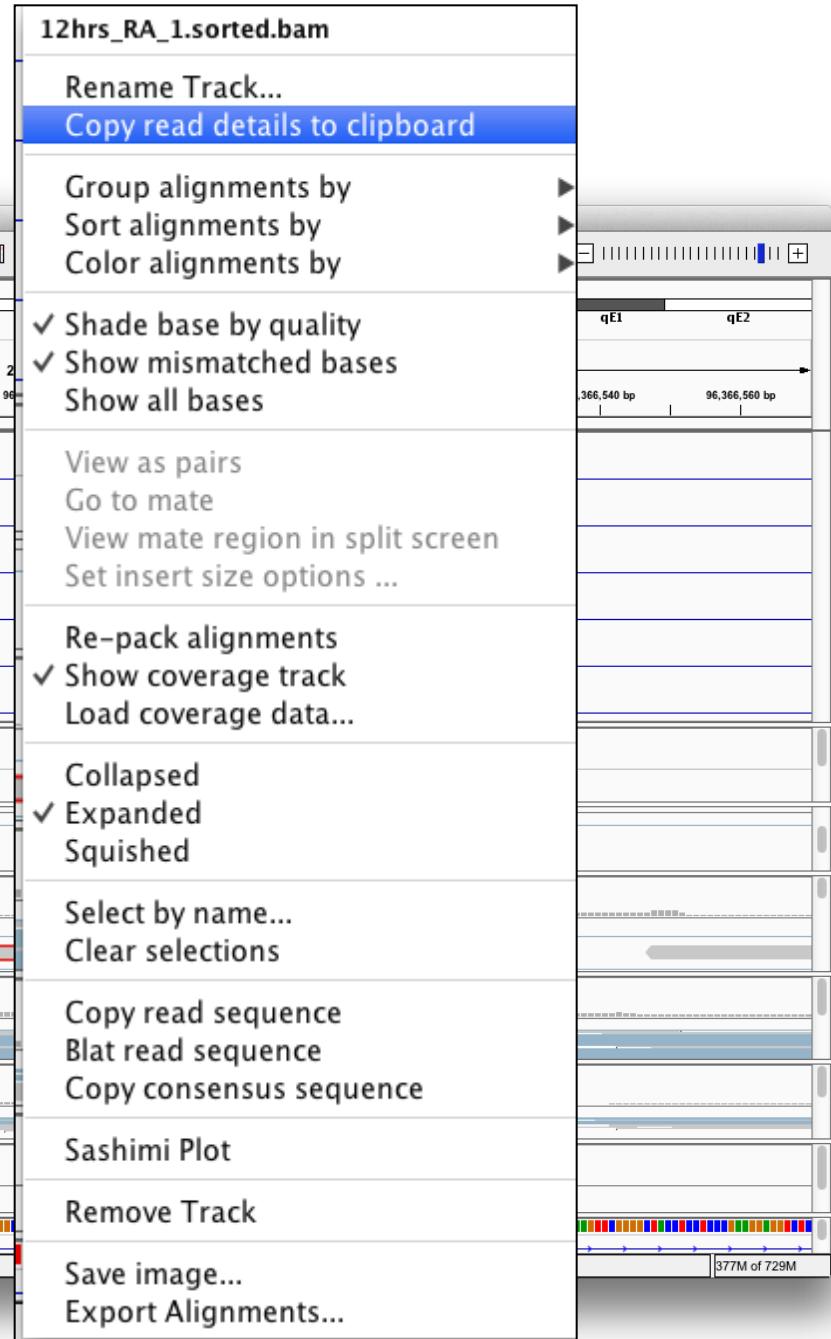
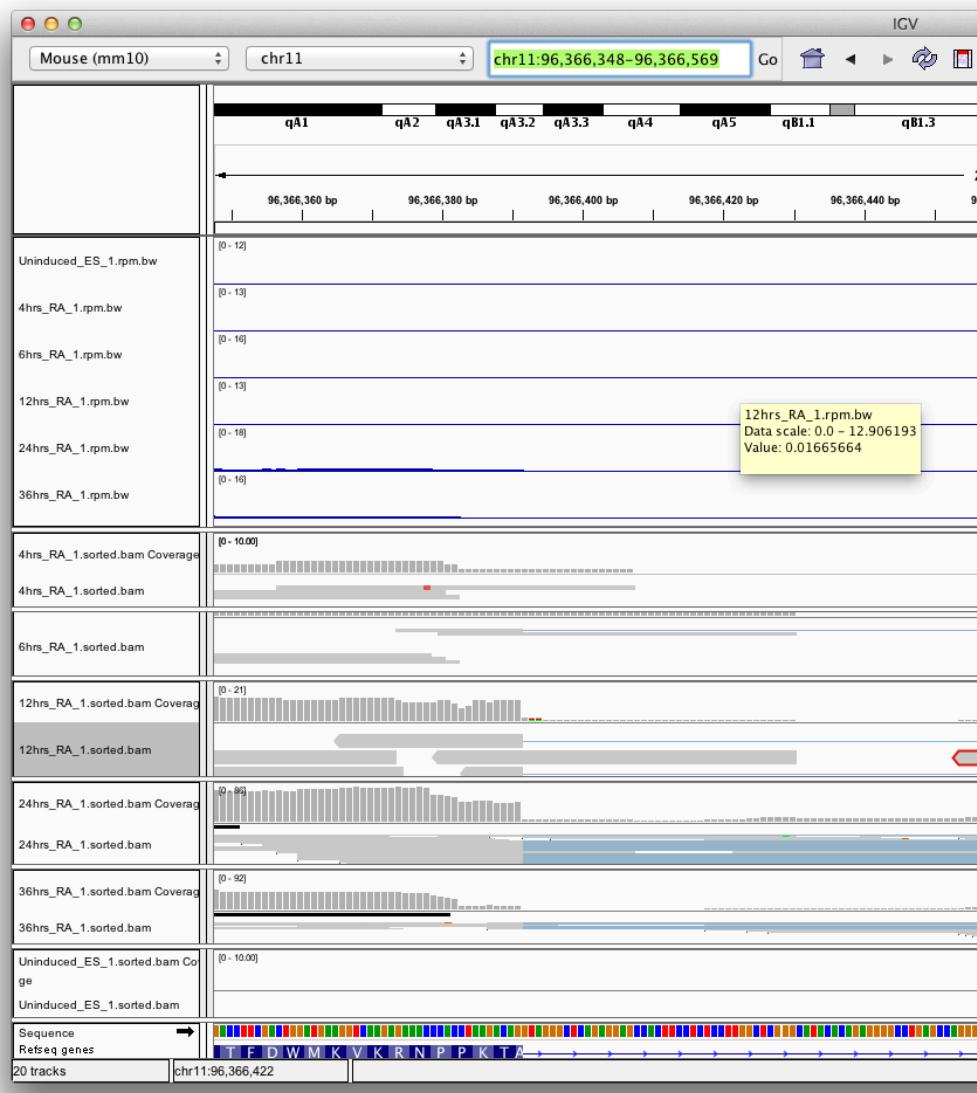


Read Details

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



Copy read details

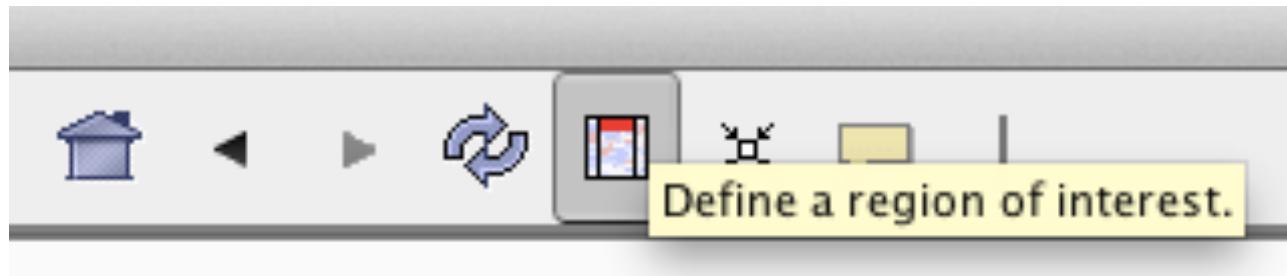


“Read Details”

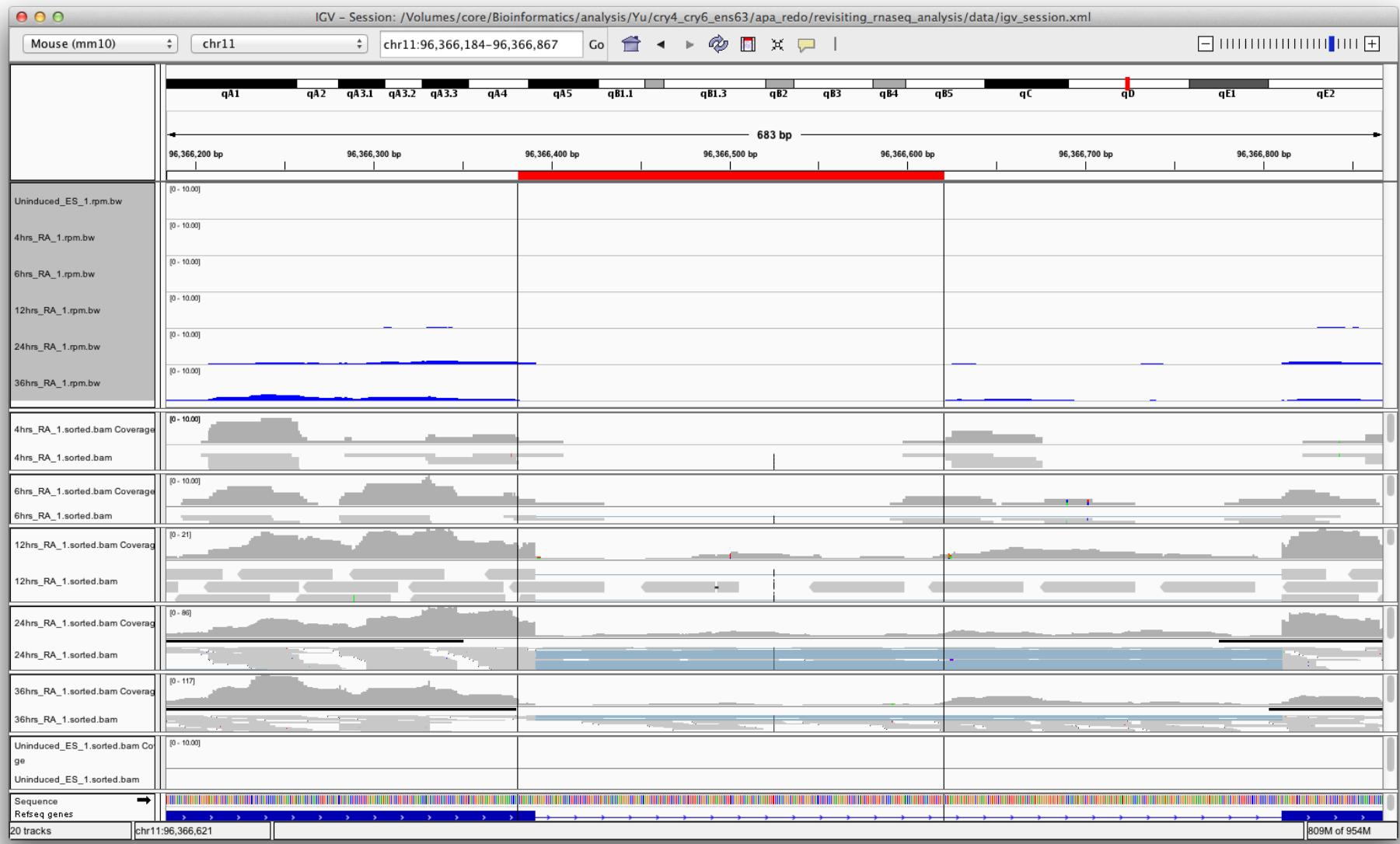
sometimes useful for digging deeply into alignments or troubleshooting

```
Read name = V00D00575:72:C6A47ANXX:5:1201:7321:83390
-----
Location = chr11:96,366,467
Alignment start = 96,366,454 (-)
Cigar = 39M1D12M
Mapped = yes
Mapping quality = 50
Secondary = no
Supplementary = no
Duplicate = no
Failed QC = no
-----
Base = T
Base phred quality = 37
-----
MD = 39^A12
XG = 1
NH = 1
NM = 1
XM = 0
XN = 0
X0 = 1
AS = -8
XS = +
YT = UU
-----
Alignment start position = chr11:96366454
GGGCCTTCATTCAATTCCCTTCAGCTGCTTTACTGGCAGGTAGTACTTTCCGG|
~  
~  
~
```

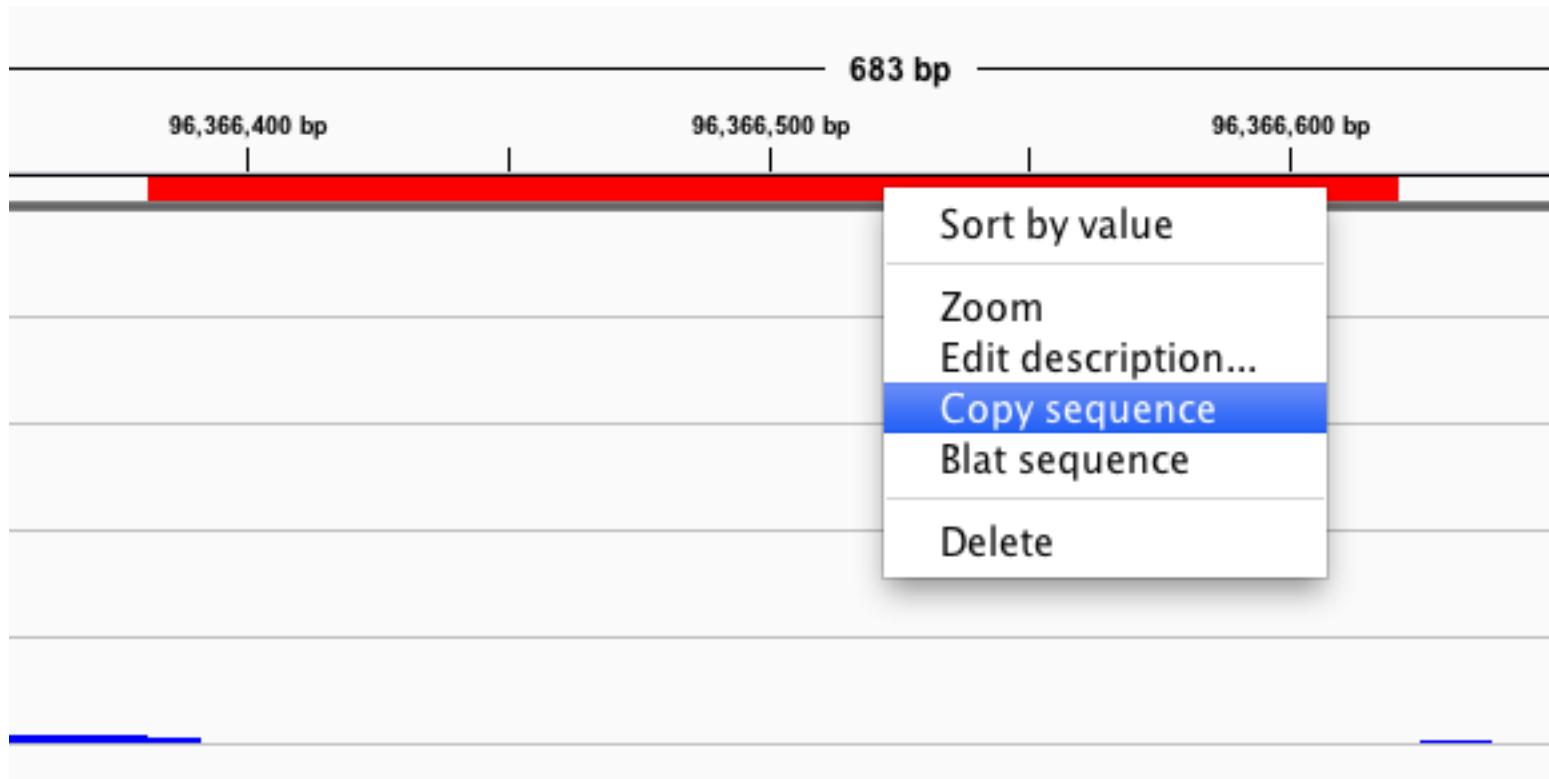
Exporting sequence for a region



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

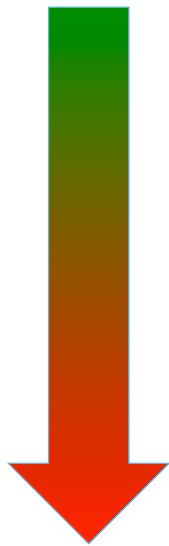


Right click on the red rectangle



Viewing multiple regions at once

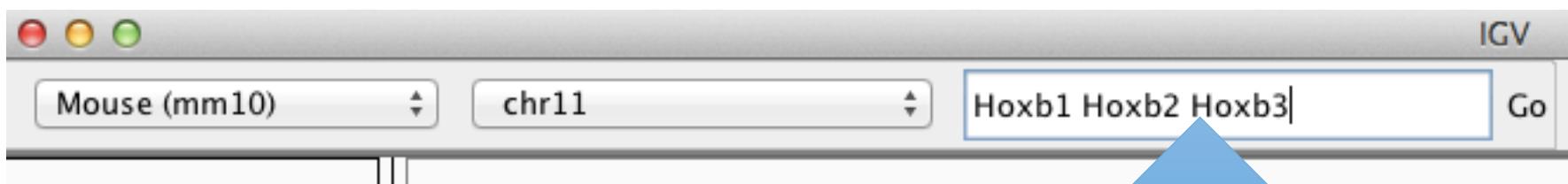
Quick & Easy, Less Powerful



Search box
Constructing links
Batch scripting

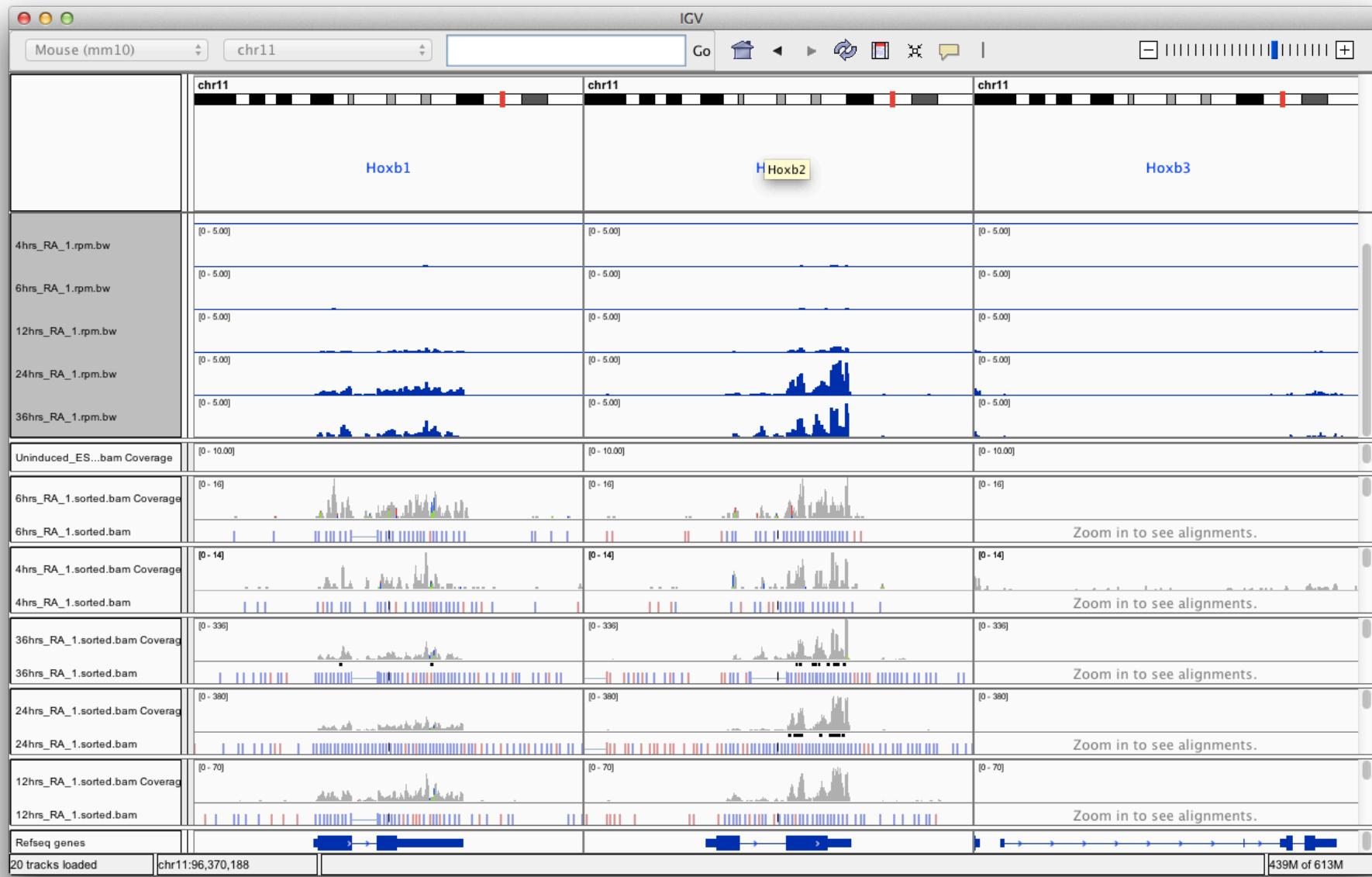
Harder, More Powerful

The search box will take multiple genes
(or chromosomal locations)



You can type a few
genes or locations
here

Genes are displayed side by side in panels



You try it!

Looking at Hoxa genes...

Search for Hoxa1, Hoxa2, Hoxa3...

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/ControllIGV>

Constructing links to genes (in Excel)

The screenshot shows a Microsoft Excel spreadsheet titled "Workbook1". The ribbon menu is visible at the top, with the "Home" tab selected. The "Font" group is active, showing "Calibri (Body)" and "12" selected. The "Cells" group is also visible. The worksheet contains a single row of data:

	A	B	C	D	E	F	G	H	I	J	K
1	Gene										
2	Hoxa1										
3	Meis2										
4	Cyp26a1										
5	Dleu7										
6											
7											
8											
9											
10											
11											
12											
13											
14											

The cell B1 is currently selected. The status bar at the bottom indicates "Normal View" and "Ready".

Excel demo

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

The screenshot shows a Microsoft Excel spreadsheet titled "Workbook1". The ribbon menu is visible at the top, with the "Home" tab selected. The formula bar shows "A7" as the active cell. The table below contains the following data:

	A	B	C	D	E	F
1	Gene	Link	hyperlink			
2	Hoxa1	http://localhost:60151/goto?locus=Hoxa1	Hoxa1			
3	Meis2	http://localhost:60151/goto?locus=Meis2	Meis2			
4	Cyp26a1	http://localhost:60151/goto?locus=Cyp26a1	Cyp26a1			
5	Dleu7	http://localhost:60151/goto?locus=Dleu7	Dleu7			
6						
7						
8						

Excel demo

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

The screenshot shows a Microsoft Excel spreadsheet titled "Workbook1". The ribbon menu is visible at the top, and the formula bar shows "D7". The table structure is as follows:

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

Batch Scripting IGV Demo

IGV has it's 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>

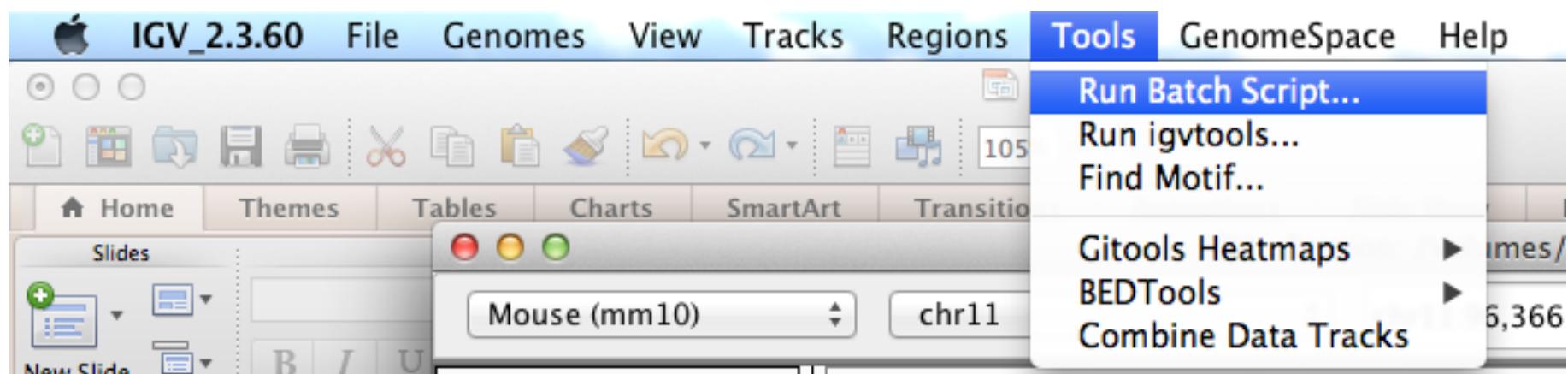
Let's drop to the terminal...

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



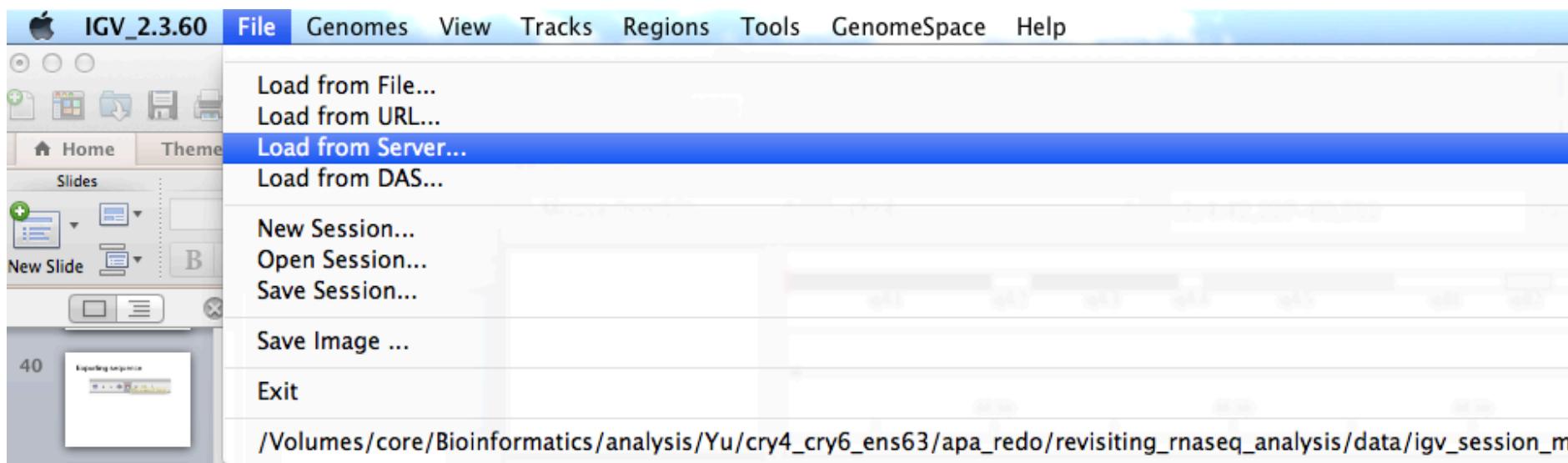
<code>chr2_74,668,309_74,671,599.png</code>	<code>chr2_74,675,012_74,677,705.png</code>	<code>chr2_74,679,557_74,687,016.png</code>	<code>chr2_74,682,322_74,687,016.png</code>	<code>chr2_74,691,890_74,695,106.png</code>	<code>chr2_74,697,762_74,700,208.png</code>
<code>chr2_74,704,614_74,707,932.png</code>	<code>chr2_74,705,488_74,707,932.png</code>	<code>chr2_74,710,043_74,716,130.png</code>	<code>chr2_74,711,992_74,748,271.png</code>	<code>chr2_74,721,977_74,729,159.png</code>	<code>chr2_74,762,979_74,765,142.png</code>
<code>chr6_52,155,366_52,158,317.png</code>	<code>chr6_52,162,416_52,164,831.png</code>	<code>chr6_52,165,673_52,169,451.png</code>	<code>chr6_52,169,061_52,213,067.png</code>	<code>chr6_52,189,686_52,191,703.png</code>	<code>chr6_52,201,123_52,213,597.png</code>
<code>chr6_52,201,753_52,204,587.png</code>	<code>chr6_52,206,364_52,208,624.png</code>	<code>chr6_52,215,623_52,218,573.png</code>	<code>chr6_52,223,096_52,227,370.png</code>	<code>chr6_52,231,196_52,234,939.png</code>	<code>chr6_52,231,196_52,240,854.png</code>
<code>chr6_52,242,104_52,245,810.png</code>	<code>chr6_52,245,242_52,249,769.png</code>	<code>chr6_52,258,852_52,260,880.png</code>	<code>chr11_96,194,360_96,196,599.png</code>	<code>chr11_96,271,329_96,276,593.png</code>	<code>chr11_96,281,904_96,285,325.png</code>
<code>chr11_96,286,645_96,290,163.png</code>	<code>chr11_96,299,170_96,301,569.png</code>	<code>chr11_96,303,511_96,306,121.png</code>	<code>chr11_96,318,266_96,321,638.png</code>	<code>chr11_96,323,125_96,347,930.png</code>	<code>chr11_96,343,768_96,347,930.png</code>
<code>chr11_96,351,631_96,354,014.png</code>	<code>chr11_96,365,757_96,368,253.png</code>	<code>chr15_102,921,130_102,9...14.png</code>	<code>chr15_102,936,852_102,9...17.png</code>	<code>chr15_102,954,525_102,9...01.png</code>	<code>chr15_102,966,795_102,9...97.png</code>
<code>chr15_102,977,031_102,9...44.png</code>	<code>chr15_102,990,538_102,9...54.png</code>	<code>chr15_103,009,561_103,0...81.png</code>	<code>chr15_103,014,007_103,0...29.png</code>	<code>chr15_103,034,394_103,0...52.png</code>	

By default, image will be a png with the locus name. Can also specify filename with extension like “.svg”

Viewing SNPs in IGV

First, switch genome to hg19...

... Then we'll load some 1000 genomes data.



Available Datasets

- Available Datasets
 - ▶ Annotations
 - ▶ ENCODE
 - ▶ Body Map 2.0 (Illumina HiSeq)
 - ▼ 1000 Genomes [\(i\)](#)
 - ▼ Alignments
 - ▼ ACB
 - ▼ exome
 - HG01879 exome
 - HG01880 exome
 - HG01882 exome
 - HG01883 exome
 - HG01885 exome
 - HG01886 exome
 - HG01889 exome
 - HG01890 exome
 - HG01894 exome
 - HG01896 exome
 - HG01912 exome
 - HG01914 exome
 - HG01915 exome
 - HG01956 exome

Cancel

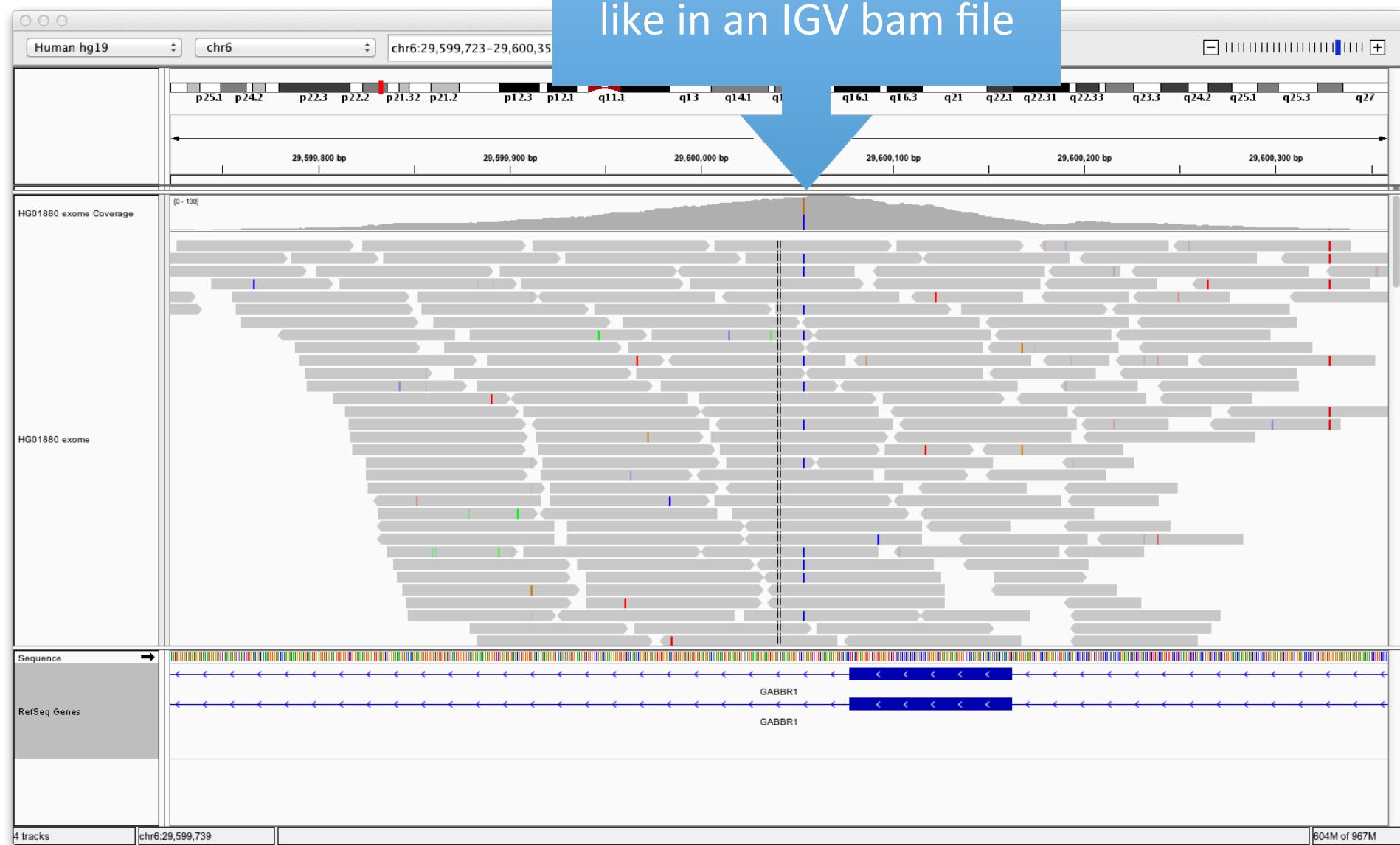
OK

You try it!

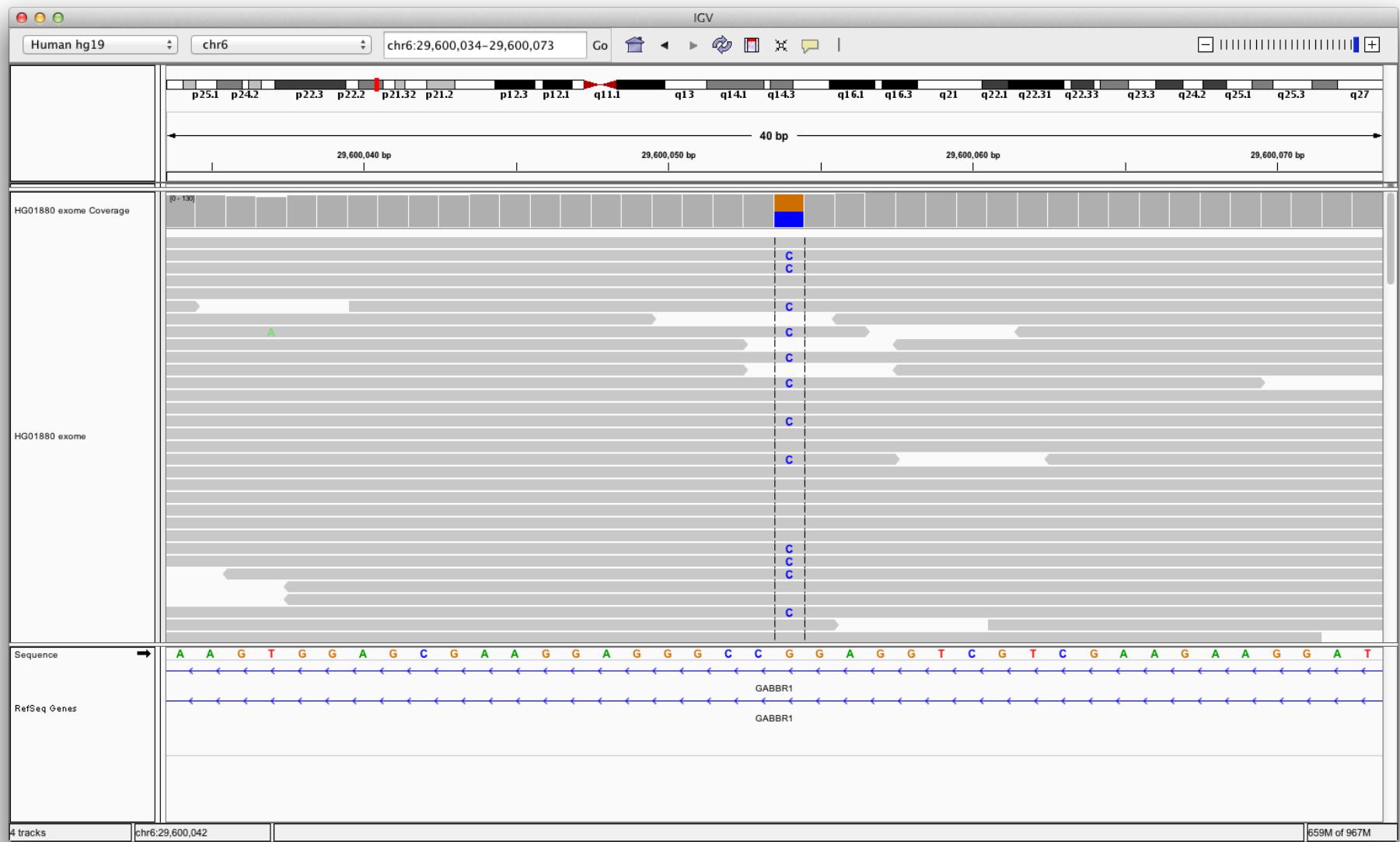
Looking at a SNP

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

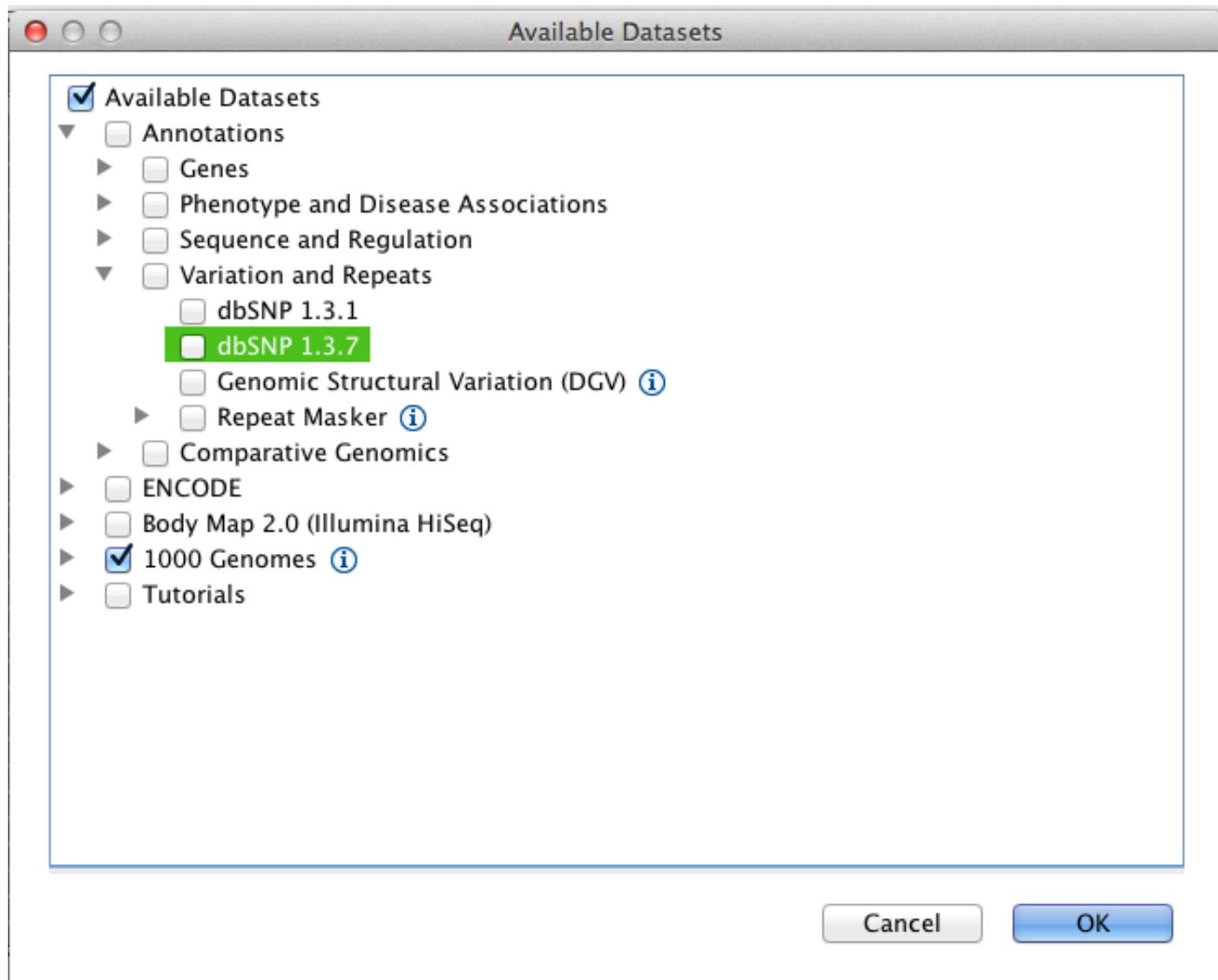
This is what a SNP looks like in an IGV bam file



SNP, up close



Load dbSNP annotation

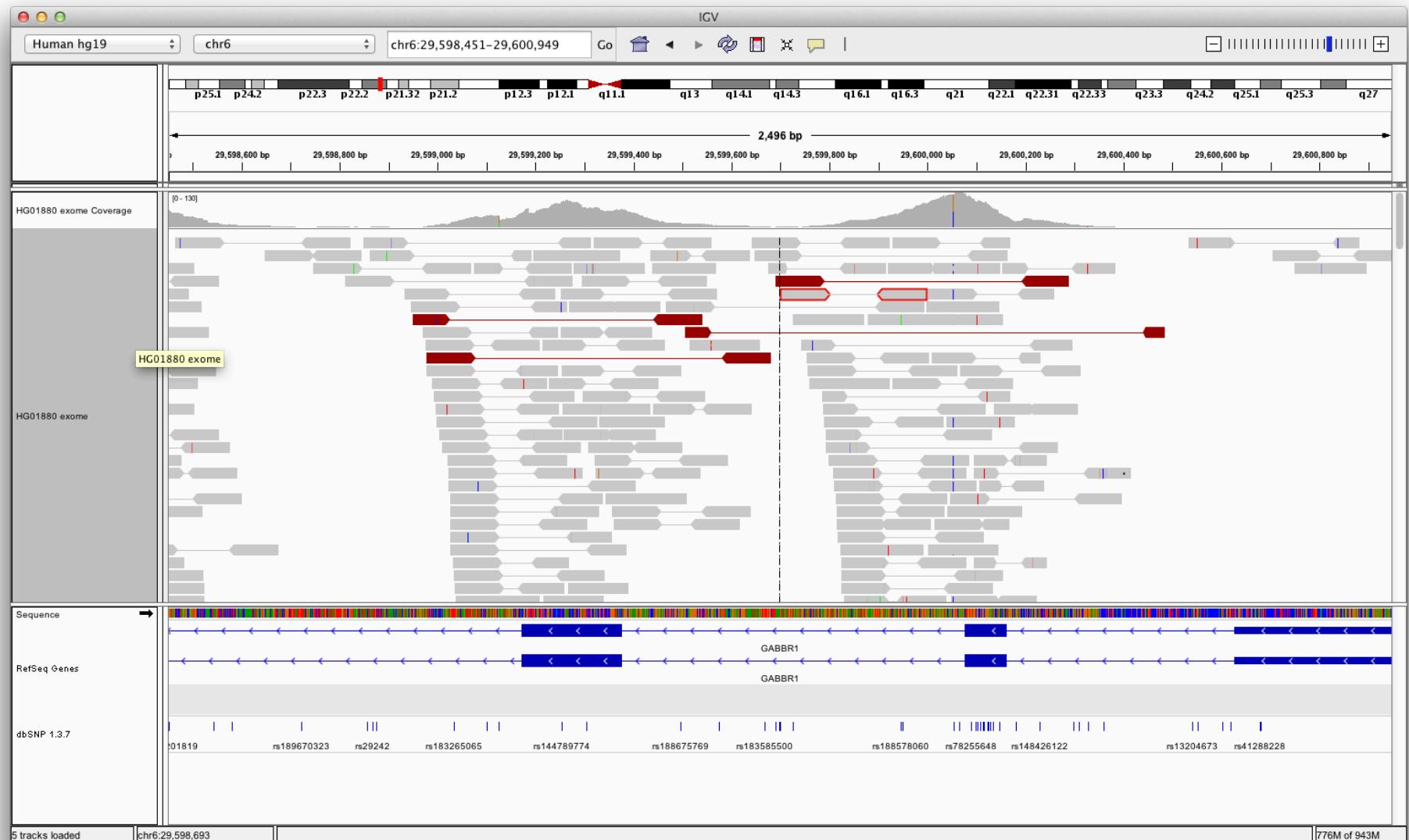


While we have some paired-end data to look at...

Right-click, View as Pairs



Right click left-hand side, color by insert size...



Red – larger insert than expected, Blue – smaller than expected
other colors = pair on another chromosome

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)