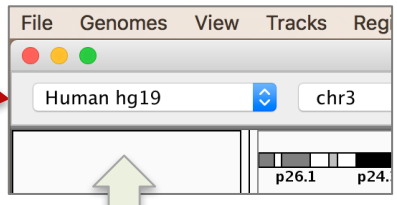
IGV Hands-on Exercise | IGV Basics

1. Launch IGV

2. Select reference genome.

- Click on *Human hg19* in the genome drop-down menu in the upper left corner.

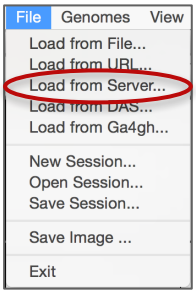
If you only see *Human hg18* in the menu, it's ok to select that instead



If this is the first time you run IGV, there may be **only one** entry in the menu. More about that later…

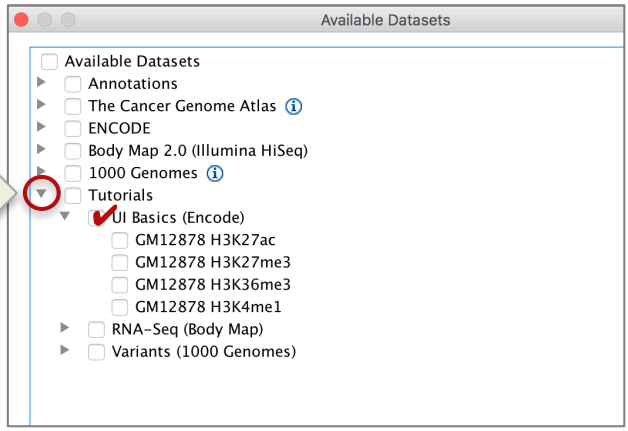
3. Load data from the IGV hosted server.

- Select *File > Load from Server…*

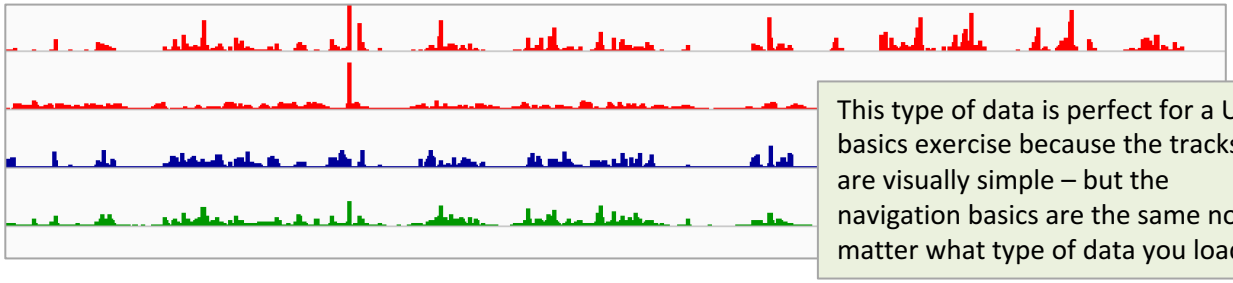


- Open the *Tutorials* menu (Use ▶ on Mac, and + on Windows) and click on the *UI Basics* checkbox.

Make sure you only **open** the *Tutorials* menu. Do **not** check the box next to *Tutorials.* That will select everything under *Tutorials,* but we only want *UI Basics* for this exercise.



Four tracks are loaded: ENCODE project ChIP-seq data representing histone modifications. Each track is displayed as a bar chart of signal intensities.

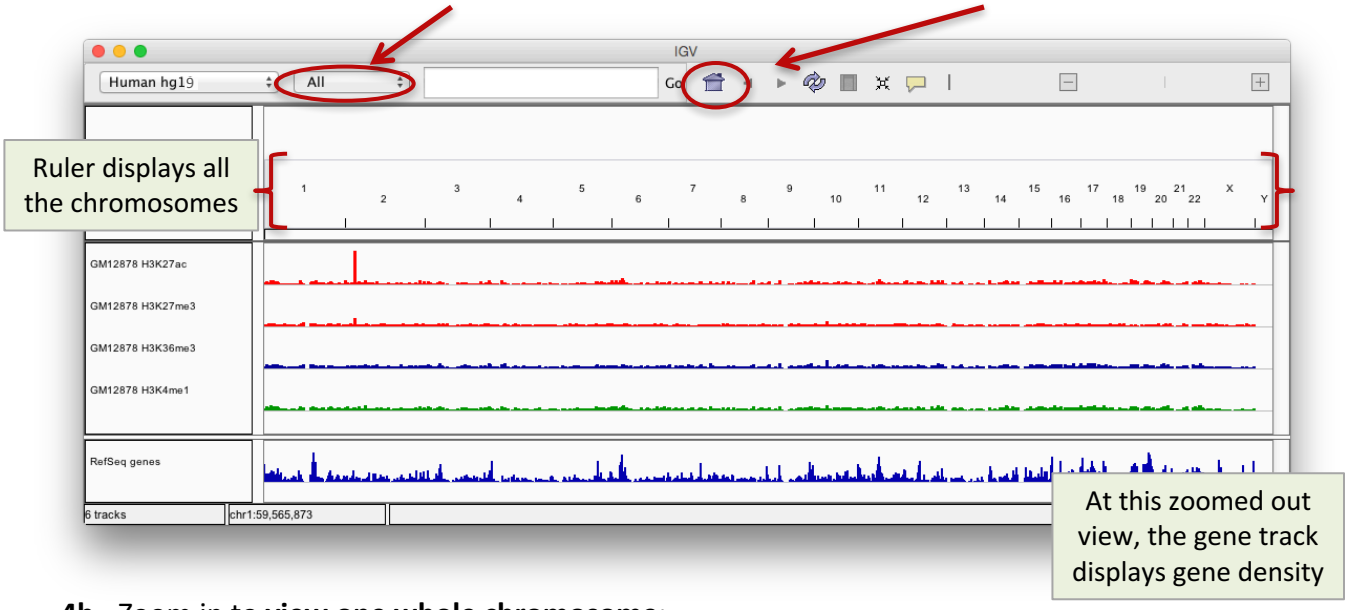


This type of data is perfect for a UI basics exercise because the tracks are visually simple – but the navigation basics are the same no matter what type of data you load.

4. **Navigate** across different genomic loci and at different zoom levels, from whole genome view and down to base-pair resolution.

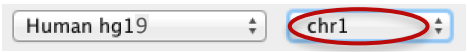
4a. Start at **whole genome view**:

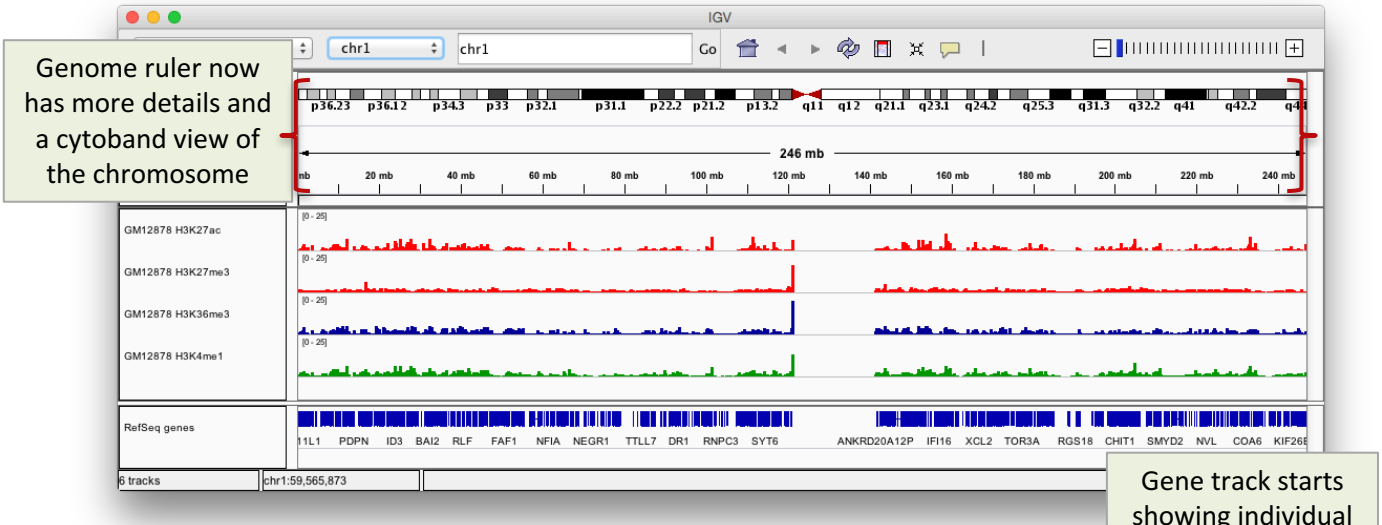
- Select *All* from the chromosome drop-down menu –OR– Click the *Home* button.



4b. Zoom in to **view one whole chromosome**:

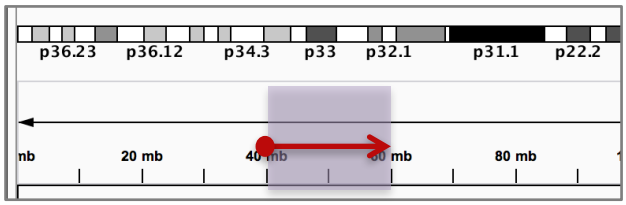
- Select *Chr1* from the chromosome drop-down menu –OR– Click the *1* in the genome ruler.



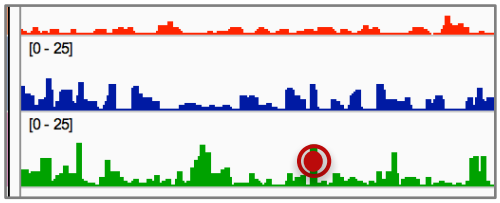


4c. Zoom in further:

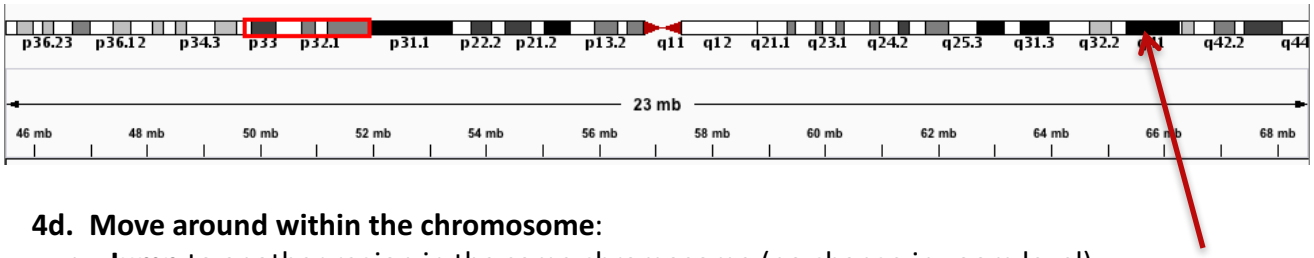
- Click and drag to zoom in on a region swept out in the ruler



- Double-click in the data track to zoom in on a point of interest. [Alt-click to zoom out]

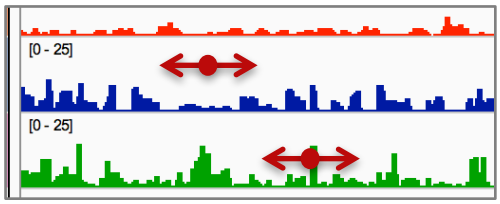


Ruler measurements and a red box on the cytoband diagram show where you are in the chromosome



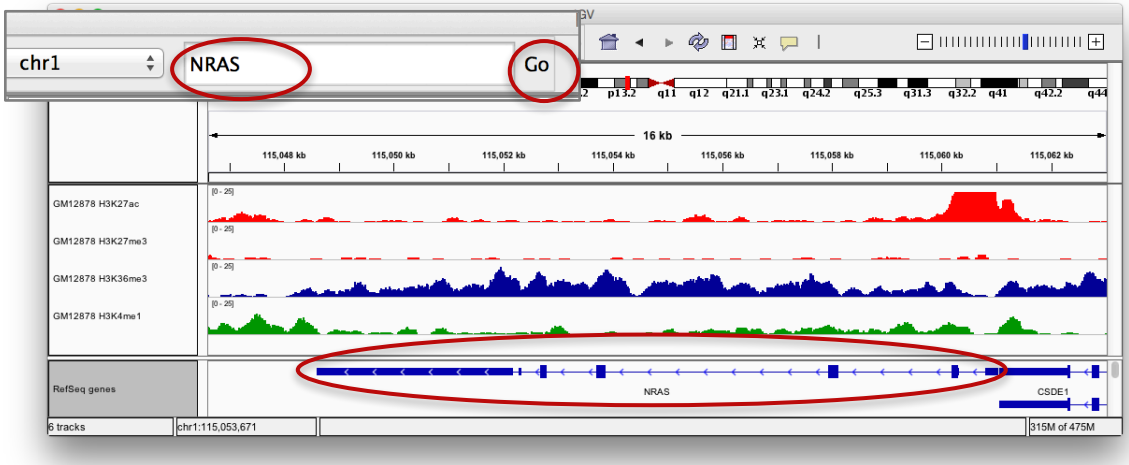
4d. Move around within the chromosome:

- **Jump** to another region in the same chromosome (no change in zoom level):
  Click anywhere in the cytoband diagram.
- **Scroll** across genome coordinates: Click anywhere in the data panel and drag left & right.



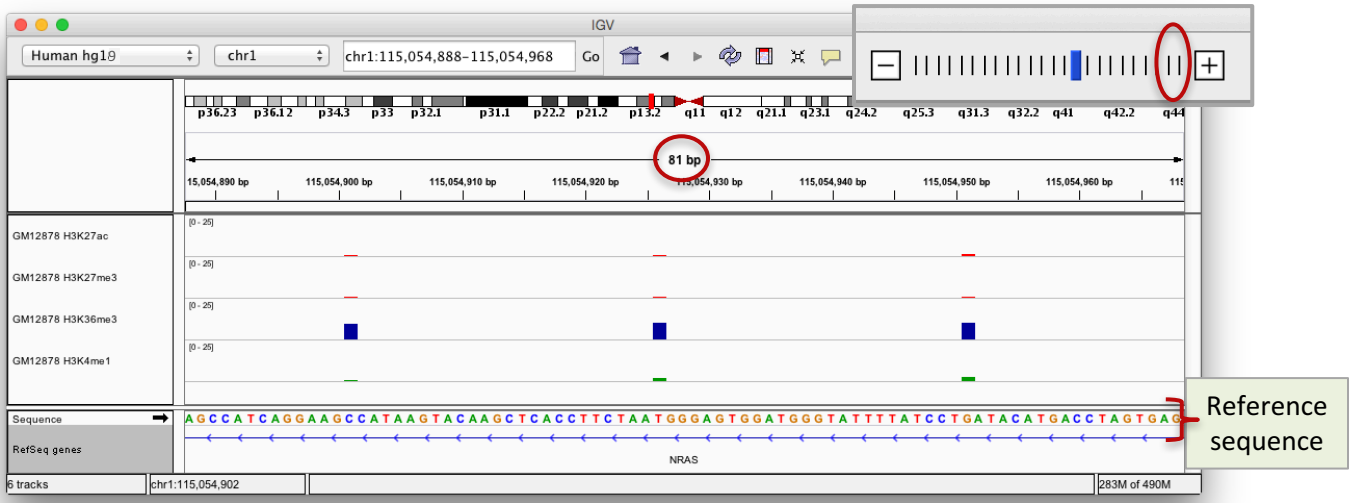
4e. Navigate to specific locus or gene on any chromosome

- Type into the search box in the IGV toolbar and click *Go:*
  either a locus in **genomic coordinates** (e.g. chr1:144,874-969,268) or a **gene name** (e.g. NRAS)



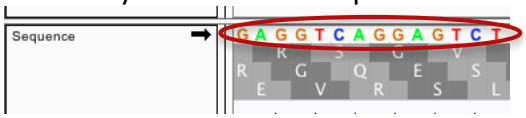
4f. Zoom in to base-pair resolution:

- Keep zooming in as before, or click on one of the rightmost ticks on the "railroad track" zoom widget in the upper right corner.

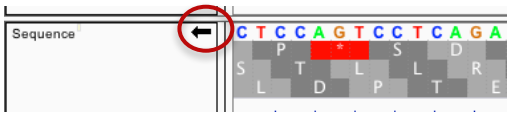


5. Options for viewing the reference sequence track

- Click anywhere on the sequence to show/hide a 3-frame translation

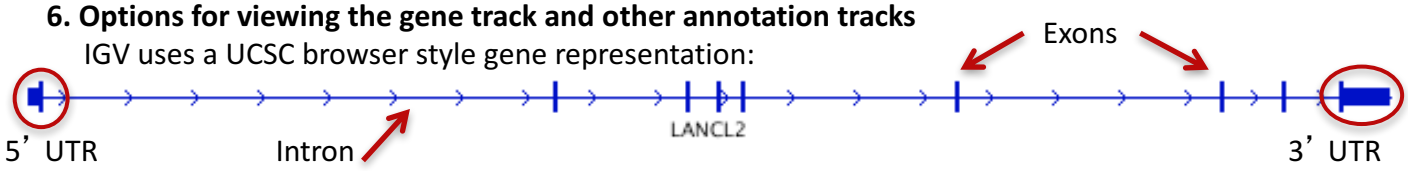


- By default, the sequence for the forward strand is shown. Click on the arrow to reverse the strand.

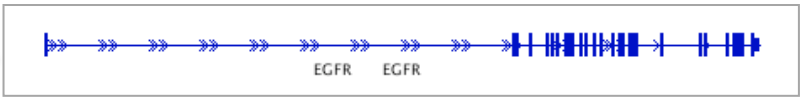


6. Options for viewing the gene track and other annotation tracks

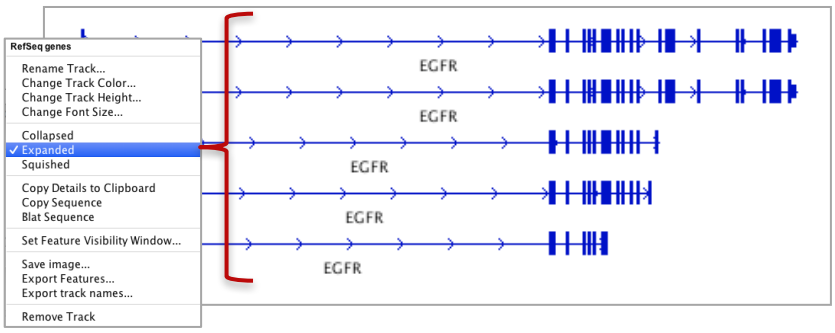
IGV uses a UCSC browser style gene representation:



Features are drawn in a single line, by default



- *Expand* the track using the right-click popup menu



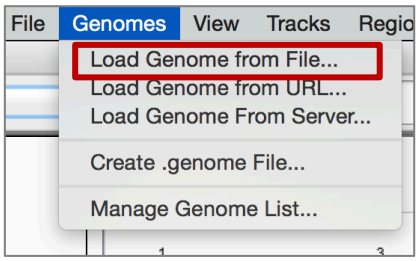
Use *Squished* for an even more compact view

END OF EXERCISE

IGV Hands-on Exercise | Load genome from file

1. Clear out the data from the previous exercise: Select *File* > *New Session*

2. Select ***Genomes*** > *Load Genome from File*



Note: this is the **Genome** menu, not the File menu

and browse to the **igvData** workshop folder, and then to the **genome** subfolder and open ***chr1.fasta***

The *chr1.fasta* file contains chromosome 1 from Human hg18.

3. Note in the IGV window: there is no gene track, and no cytoband ideogram in the genome ruler.

IGV hosted genomes package everything together, but you loaded only the FASTA file with the sequence. You can zoom in and out as before, and enter the numeric value of a locus, but you cannot find a gene locus by entering the name in the search box.

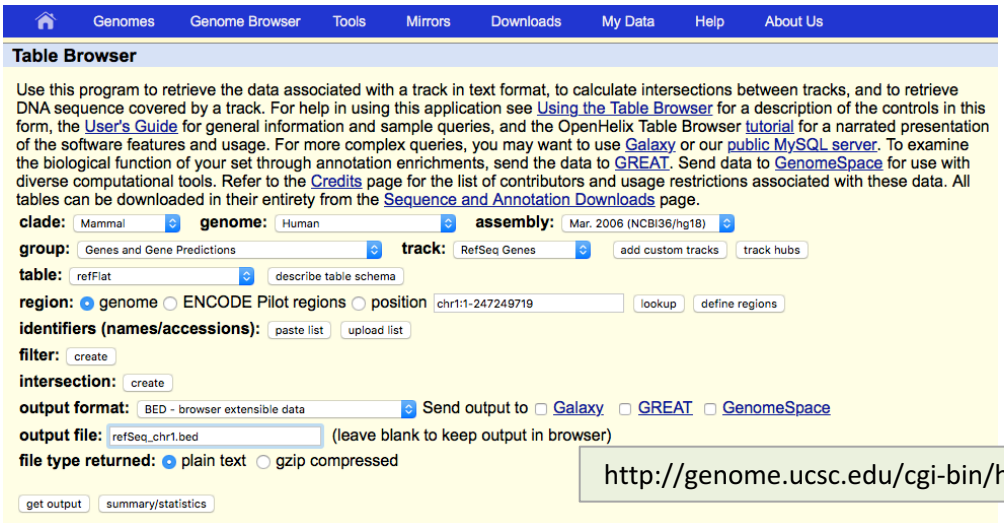
4. Load a gene annotation track
Select ***File*** > *Load from File*
and open ***refSeq_chr1.bed*** from the igvData / genome folder.

Note: this is the **File** menu, not the Genome menu

Now you can jump to a locus by entering the name of a gene on chr1 in the search box, e.g. CAP9

Not part of this exercise:

*You can use the **UCSC Table Browser** to get a file of gene annotations.*



http://genome.ucsc.edu/cgi-bin/hgTables

5. The cytoband cannot be loaded separately into the genome ruler.

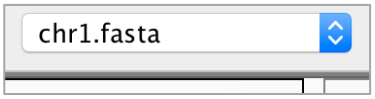
END OF EXERCISE

IGV Hands-on Exercise | Viewing SNPs

1. Reference Genome

First make sure the reference genome is set to the one you loaded in the previous exercise:

Select ***chr1.fasta*** from the genome dropdown menu



2. Load data

Click *File > Load from File*

Navigate to the workshop igvData folder,
and then the **snps** subfolder. Open the following files

igvData / snps / ***snp_calls.bed***
igvData / snps / ***NA12878.SLX.sample.bam***

Note: Do **not** load the .bai file

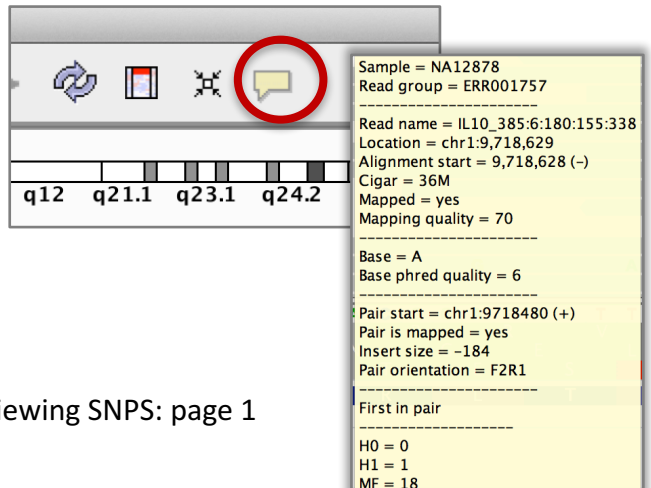
IGV automatically finds the index file – as long as it is named correctly and is in the same folder as the .bam file

3. Navigate to first putative SNP locus

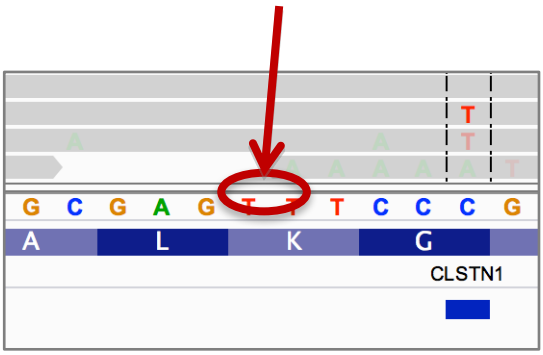
Type ***snp1*** in the search box and click **Go**

4. Optional

a) Modify the info popup behavior if you don't want the yellow info window visible all the time ➔ Click on the yellow balloon icon, and select ***Show details on click***



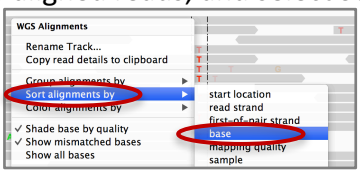
b) For a larger data panel, click & drag the window divider

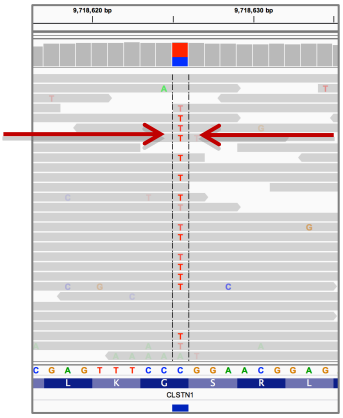


5. Sort the mismatched aligned reads by base

First, click and drag to position the mismatched bases between the center guidelines

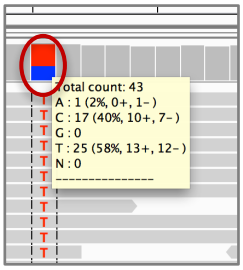
Right-click (on Mac: control-click) anywhere in the aligned reads, and select *Sort alignments by > base*





6. See the allele counts and frequencies

Mouse over the red/blue bar in the coverage track
(Or *click* on the bar, if you changed the information popup behavior to display on click only)



Observe the distribution of mismatches at that locus.
Observe the lack other mismatches in the region.

➔ This appears to be a heterozygous SNP.

---

7. Go to the locus of the second putative SNP

Type ***snp2*** in the search box and click *Go*

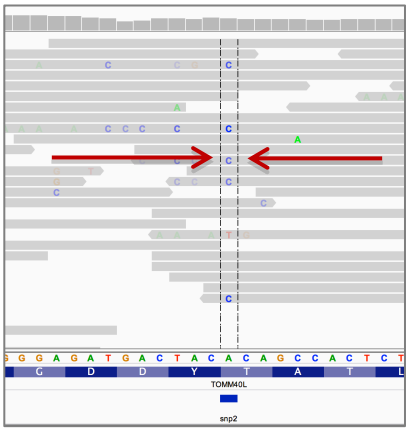
Observe the mismatched bases and their apparent low qualities.
(Mismatched bases are drawn in a fainter color if the base call is of low quality)

8. Disable shading by quality

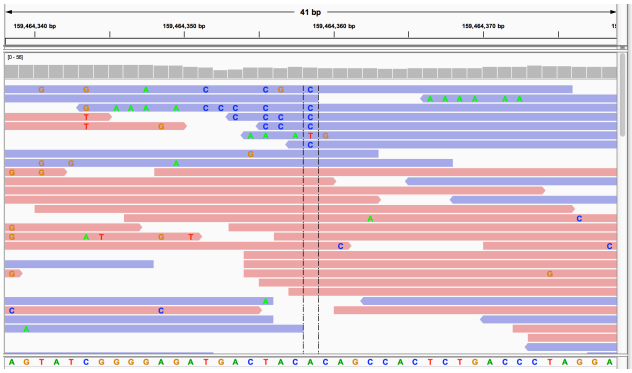
Click & drag to position the ***snp2*** locus
(with the 5 blue Cs) between the vertical center guidelines.

Right-click (on Mac: control-click) anywhere in the aligned reads, and click ***Shade base by quality***

Observe the mismatched bases.



9. Sort and color the aligned reads by read strand



Right-click (on Mac: control-click) anywhere in the aligned reads, and select ***Sort alignments by > read strand***

Right-click (on Mac: control-click) anywhere in the aligned reads, and select ***Color alignments by > read strand***

Observe where the mismatches are.

Note: We know that this sequencing was not with a strand-preserving library, so the expected strand distribution is 50-50.

➔ This is likely a false positive.

END OF EXERCISE

IGV Hands-on Exercise | Pre-compute coverage track

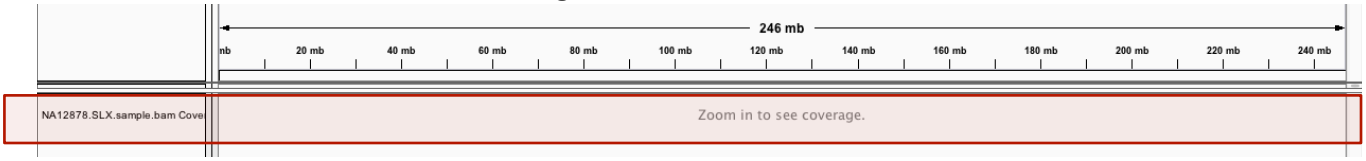
This exercise assumes the following have already been loaded in the previous exercises:

a) Reference genome chr1.fasta
(from exercise "Load genome from file")

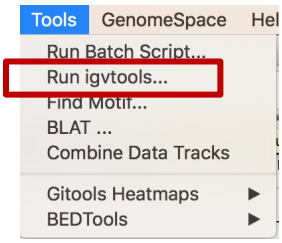
b) BAM file from workshop folder: igvData / snps / *NA12878.SLX.sample.bam*
(from exercise "Viewing SNPs")

1. First, zoom all the way out by clicking on the leftmost tick on the railroad track in the zoom tool.

Observe there is no data in the coverage track

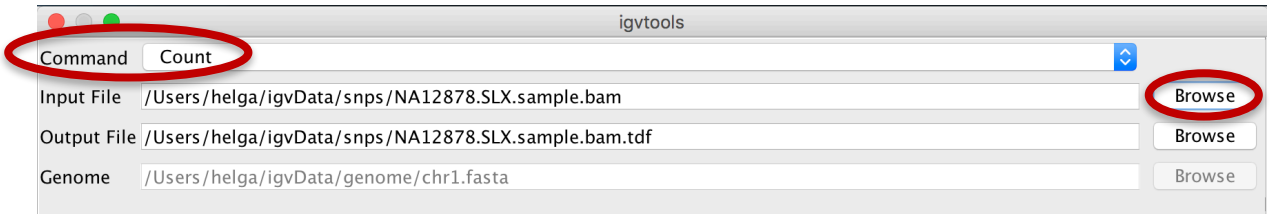


2. Launch igvtools: click ***Tools > Run igvtools***



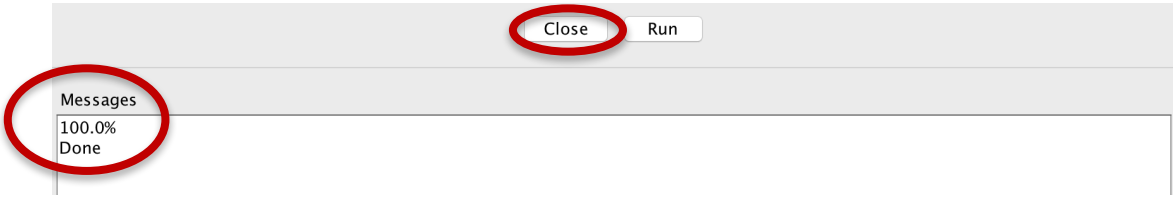
3. Run the Count tool

> Select ***Count*** from the *Command* dropdown menu
> Set the *Input File* to the ***NA12878.SLX.sample.bam*** file in the workshop folder *igvData / snps.*
The *Output File* will automatically be set to the same folder, and same name + *.tdf* suffix



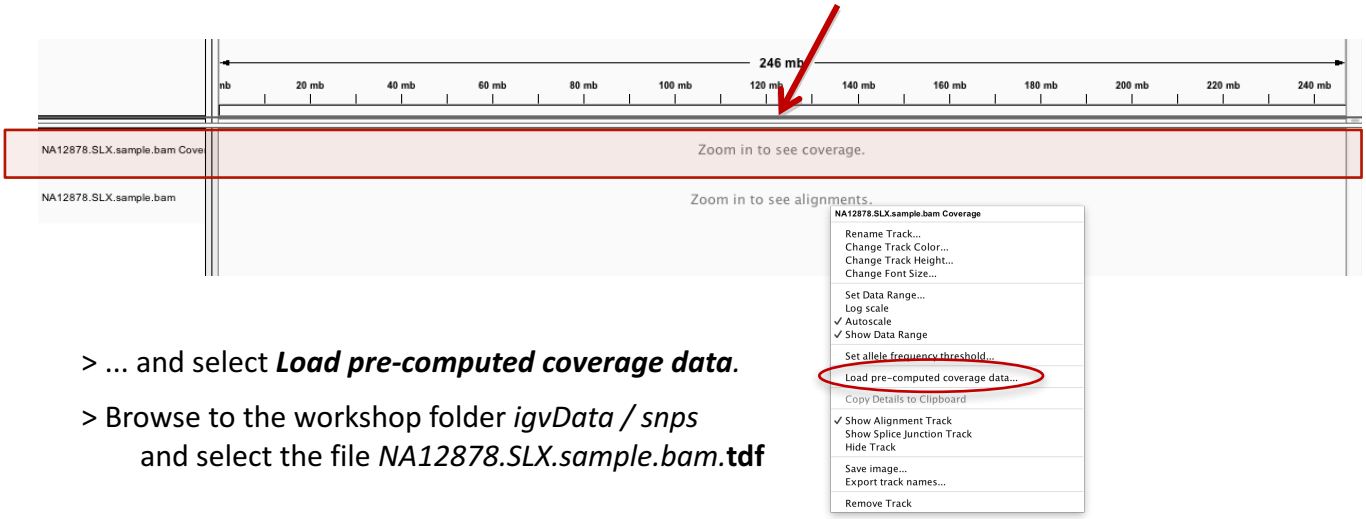
> Keep the defaults for all other values.
> Click ***Run***

Wait until you see "Done" in the *Messages* area, and click ***Close***.



4. Associate the new .tdf file with the coverage track

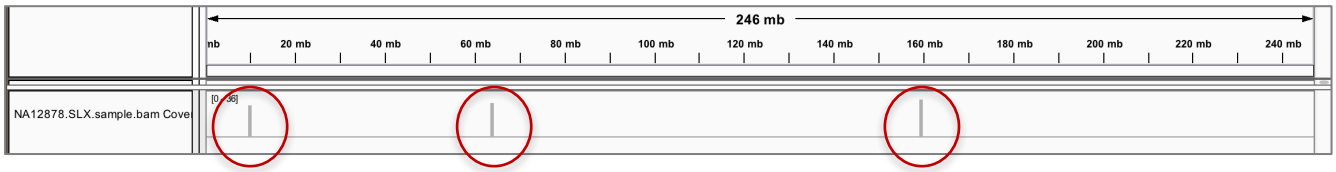
> Right-click (command-click on Mac) anywhere on the **coverage track** in the IGV window ...



> ... and select ***Load pre-computed coverage data***.

> Browse to the workshop folder *igvData / snps*
and select the file *NA12878.SLX.sample.bam.***tdf**

2. Observe the 3 spikes in the coverage track.



The .bam file for the exercise was stripped down and only has data in these 3 regions.

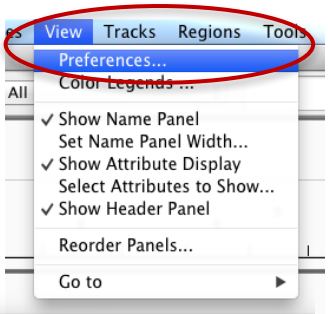
If you zoom in on the leftmost spike, you will see the *snp1* locus from the *Viewing SNPs* exercise.

END OF EXERCISE

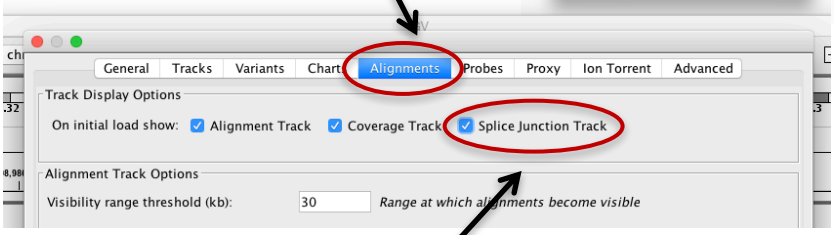
IGV Hands-on Exercise | Viewing RNA-Seq Data

1. Set preferences for viewing RNA-seq data

Click *View > Preferences*



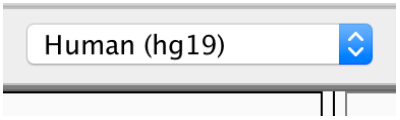
Select *Alignments* tab



Check *Splice Junction Track*

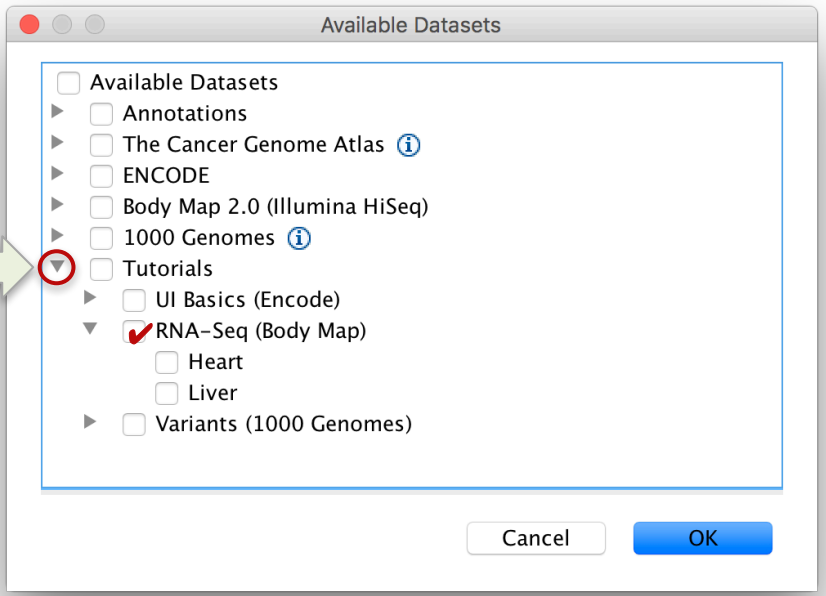
2. Load data

Select *Human hg19* from the genome dropdown menu



Click *File > Load from Server*

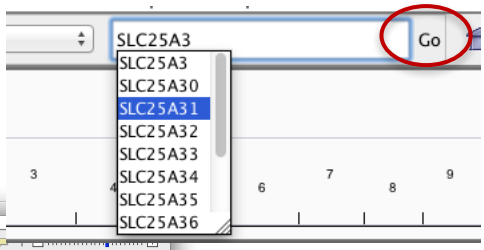
Open the *Tutorials* menu (Use ▶ on Mac, and + on Windows)
and click on *RNA-Seq (Body Map)*
and then click on *OK*

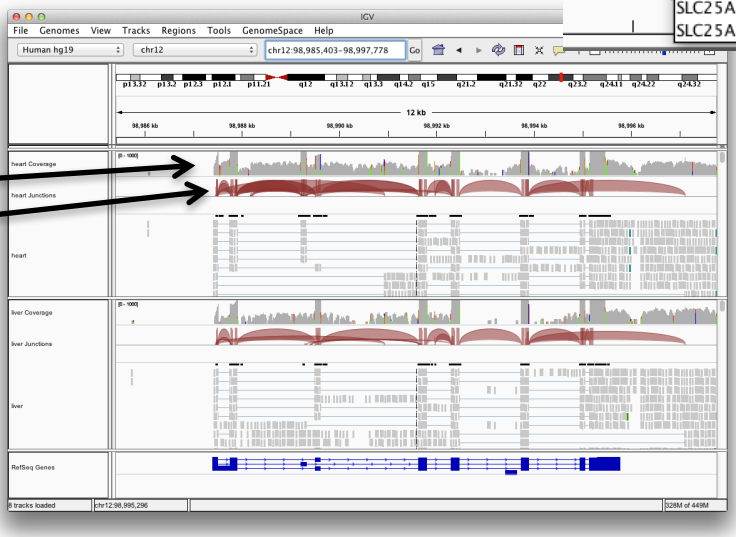


Make sure you only **open** the *Tutorials* menu. Do **not** check the box next to *Tutorials.* That will select everything under *Tutorials,* but we only want *RNA-Seq* for this exercise.

3. Jump to gene SLC25A3

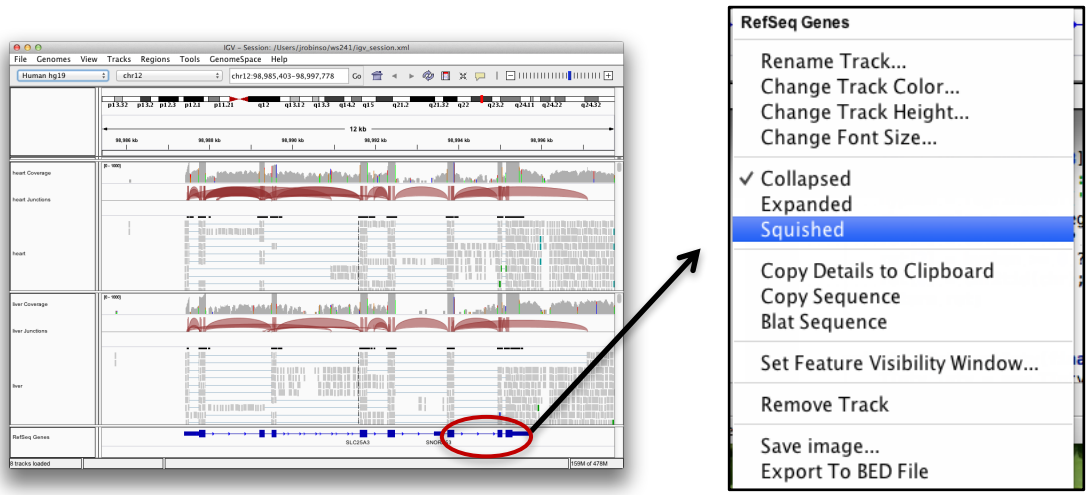
Type *SLC25A3* in the search box and click ***Go***





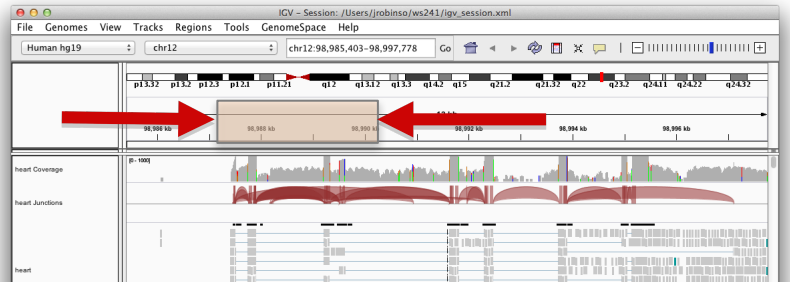
4. Expand gene track to see isoforms

Right-click over the *RefSeq Genes* track, and select ***Squished***



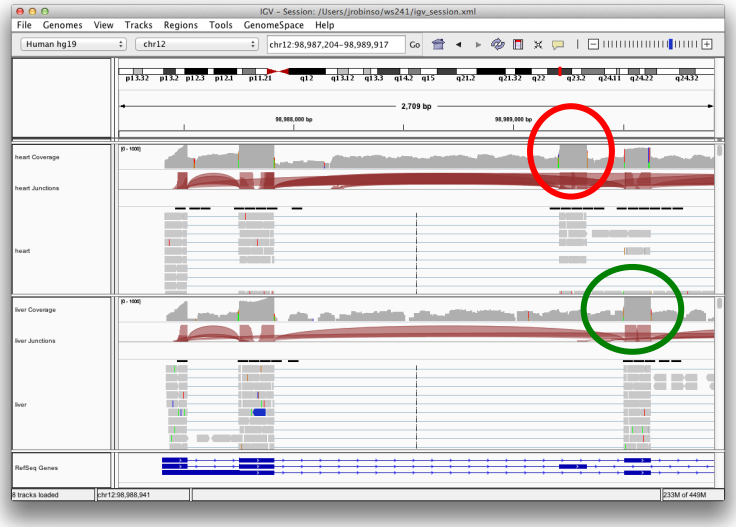
5. Zoom in on first 3 exons

Click and drag in ruler region over area shown



6. Note evidence of alternative splicing.

Observe which isoforms in the RefSeq track are expressed in each tissue.

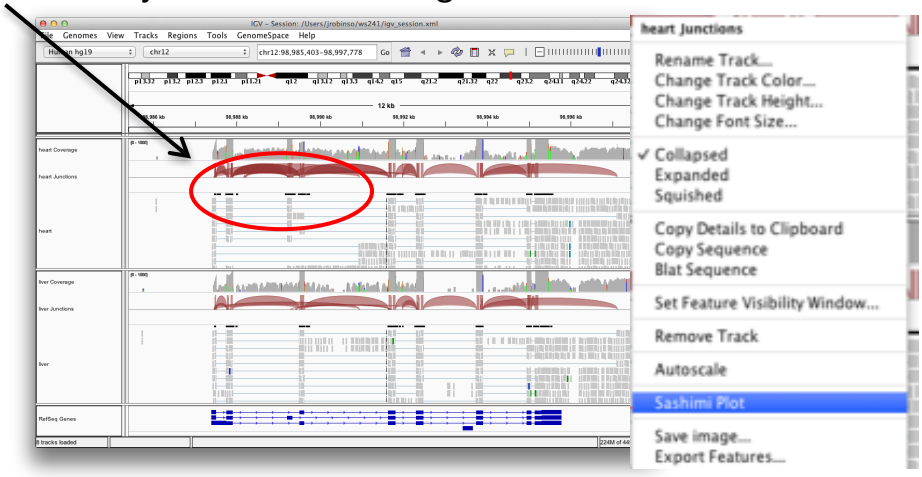


7. Zoom back out to view whole gene

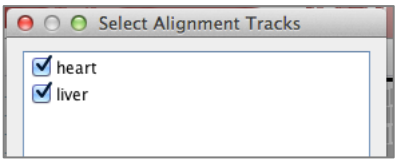
Click the back button in the command bar to zoom out to previous view

8. Open Sashimi plot

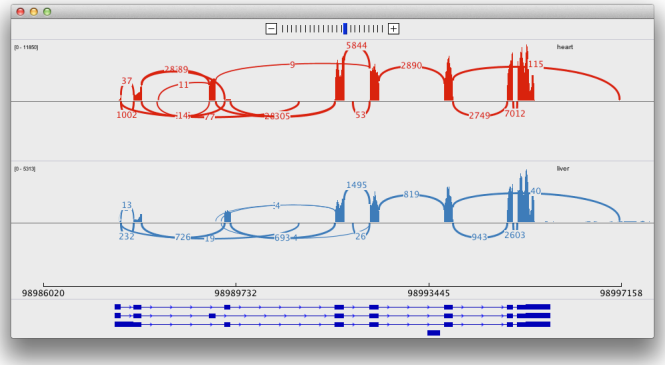
Right-click over junction track or alignments and select *"Sashimi Plot"*



Verify both *heart* and *liver* are checked, and click *OK*



9. Examine Sashimi plot



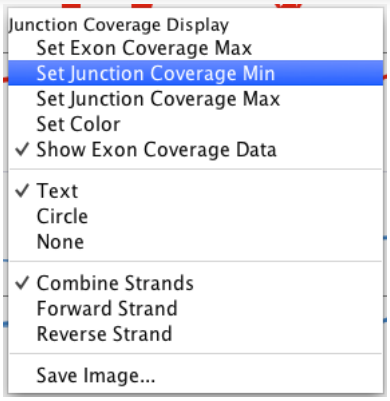
Note:

- Arcs represent reads spanning exon junctions
- Peaks represent exon coverage

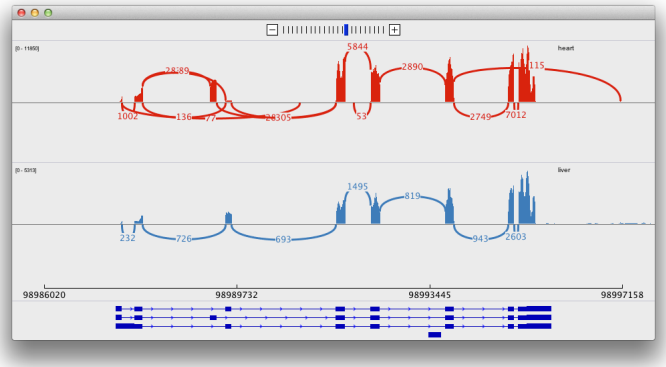
10. Filter out low-count splicing events

Right click over red (heart) track and select ***Set Junction Coverage Min***. Enter **50** and click ***OK***.

Repeat for blue (liver) track.



11. Compare with non-filtered view



12. Zoom in on 5' end

Click "+" button 2 times

Click-and-drag tracks to the right to bring the first 3 exons in view.

13. Observe the alternative splicing of the 3rd exon

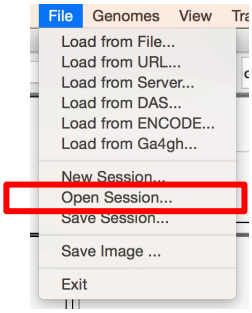
END OF EXERCISE

IGV Hands-on Exercise | Viewing variants (VCF file)

1. Load data

Click *File* > ***Open Session***

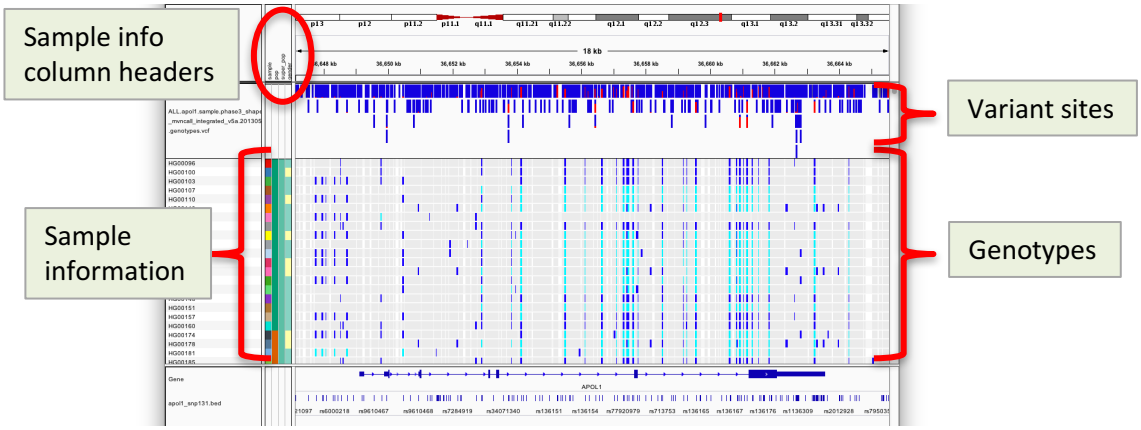
Then navigate to the workshop folder **/ igvData / vcf** and open ***vcf_session.xml***



2. Observe the different data panels

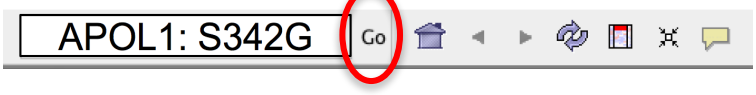
Hover over the **variant sites** and the **genotypes** to see the details.
(If you changed the popup behavior with the tool, you may have to *click* to see the details)

Observe how the same values in the **sample information** panel are assigned the same color.
Try clicking on the sample information column headers to sort by attribute.

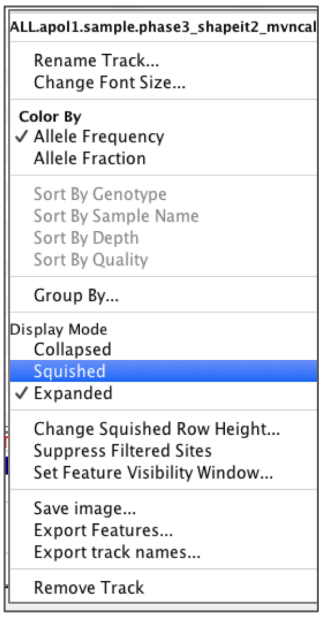


3. Highlight interesting event in gene APOL1

– Type APOL1:S342G in the search box and click ***Go***

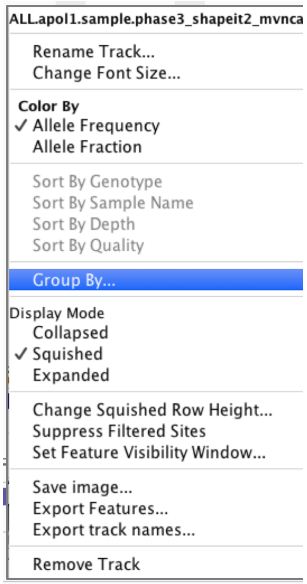


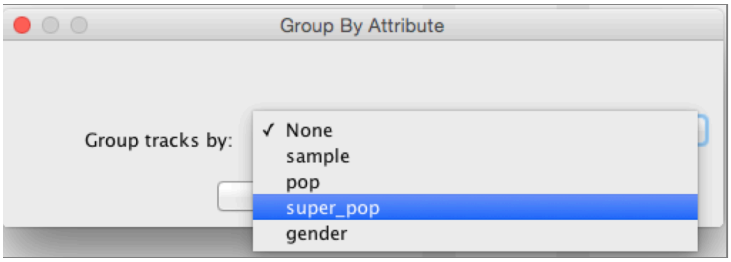
– Right-click over genotypes and select ***Display Mode: Squished***



3. (continued) Highlight interesting event in gene APOL1

− Right-click over genotypes and select ***Group By,*** then select attribute **super_pop**.

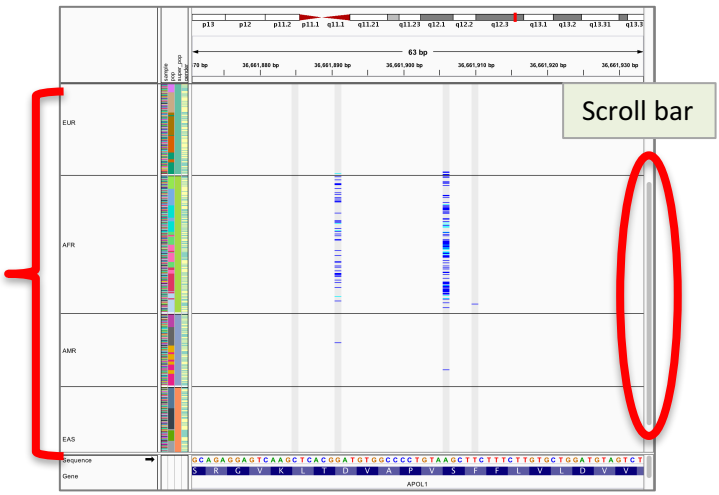




4. Observe differences between groups

Use the scrollbar to scroll down and see all the groups.

Note that the variants at this locus are not present in some population groups and are prevalent in others.



END OF EXERCISE