Exercise 3

Instructions –

- Read the exercise instructions carefully and answer the questions below

- In this exercise we will use the IGV application. Install IGV according to the instructions in the lecture (https://software.broadinstitute.org/software/igv/download). If you are working in a Feinberg Class B computer, IGV is already installed.

- This exercise is part of part of a Broad workshop from April 2017: https://www.igv.org/workshops/BroadApril2017/IGV_Exercises.pdf

- The exercise includes ChiP-seq and RNA-seq data. Chip-seq is a method used to analyze protein interactions with DNA. Chip-seq combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins. RNA-seq can be used to determine mRNA expression levels.

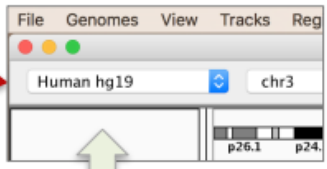
IGV Hands-on Exercise	IGV Basics

1. Launch IGV

2. Select reference genome.

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

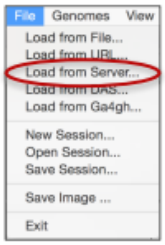
If you only see *Human hg18* in the menu, it's ok to select that instead



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

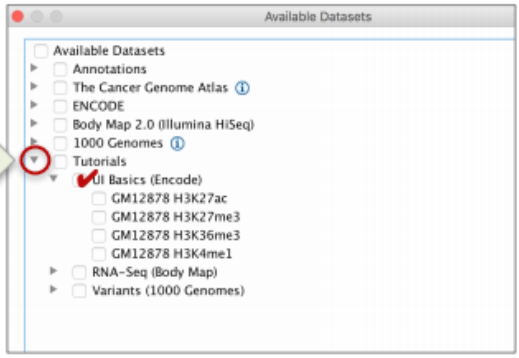
3. Load data from the IGV hosted server.

- Select *File > Load from Server...*



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

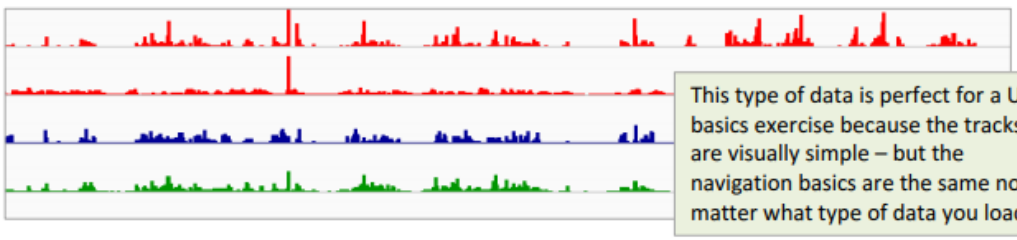
Make sure you only **open** the *Tutorials* menu. Do **not** check the box next to *Tutorials*. That will select everything under *Tutorials*, but we only want *UI Basics* for this exercise.



The loading from the server takes time. Be patient!

Click OK after selecting the UI Basics tutorial.

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

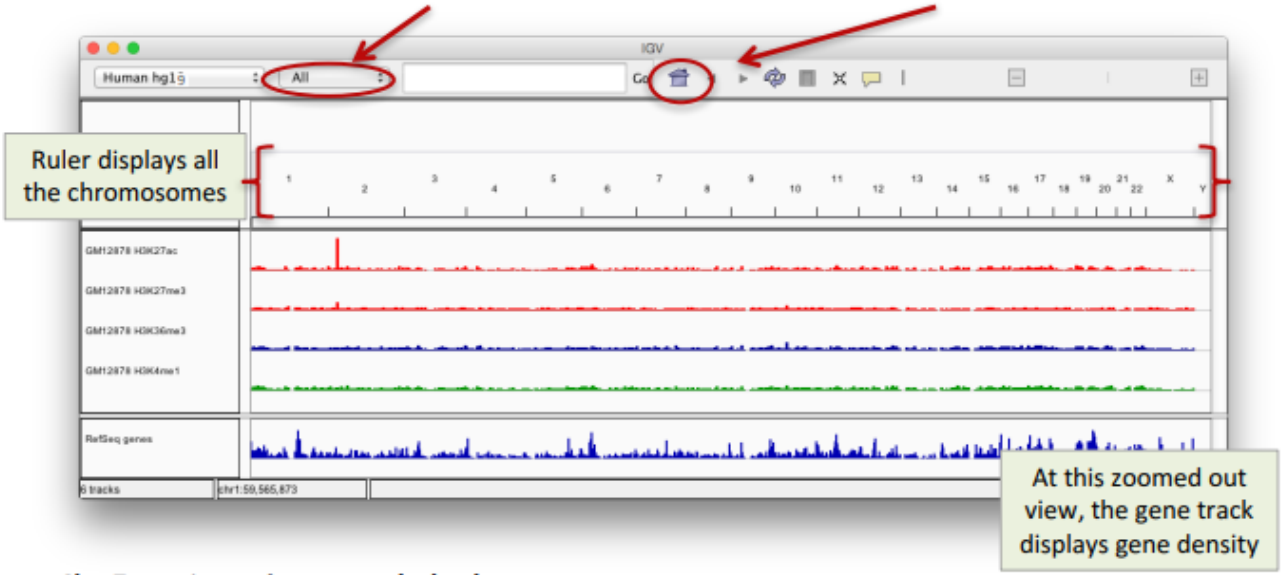


This type of data is perfect for a UI basics exercise because the tracks are visually simple – but the navigation basics are the same no matter what type of data you load.

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

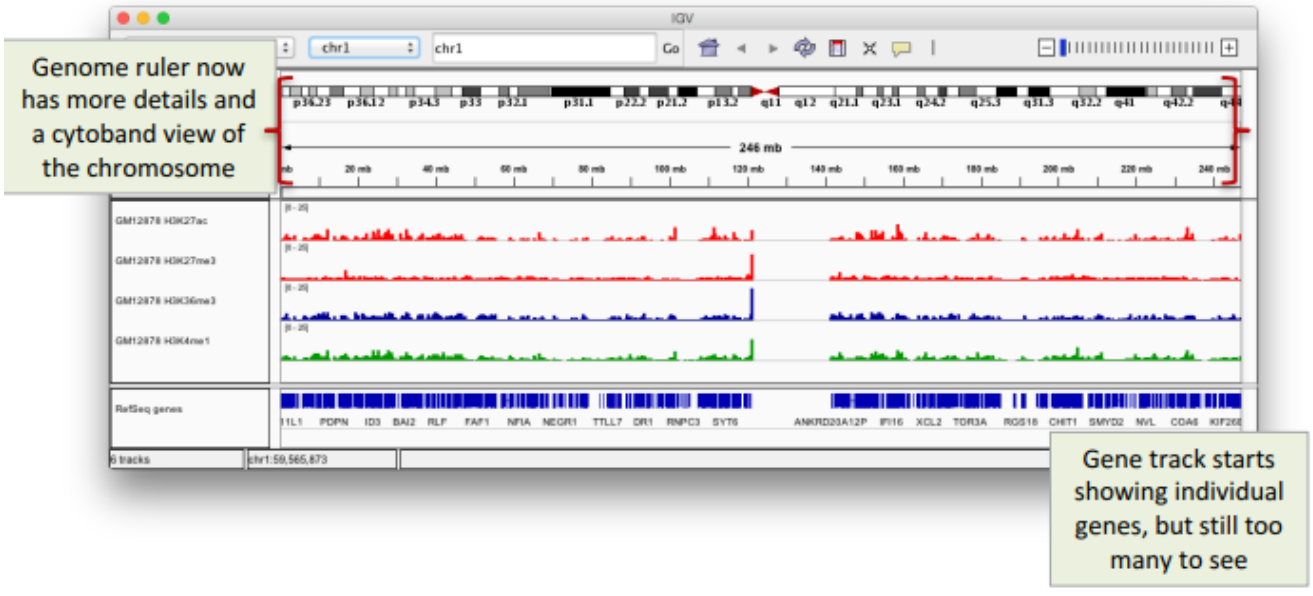
4a. Start at **whole genome view**:

- 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.



Questions

1. Go to gene BRCA2. How many isoforms has it (Be sure the RefSeq track is expanded)?

   Answer: only 1 isoform

2. Go to gene N4BP2L1.
   A. How many isoforms has it?
   B. Is isoform 2 translated? (Hint: look at the height of the exons).

   C. What are the differences between isoforms 14 and 15 (from top to bottom) as they appear in the IGV, in exons and UTRs?

   Use exon numbers to answer the following questions on isoform 14. Note that exon 1 is the closest to the 5' of the transcript.

   D. Which amino acids are translated from the first exon (from 5' to 3')? Write in the answer the first 5 and the last 5.

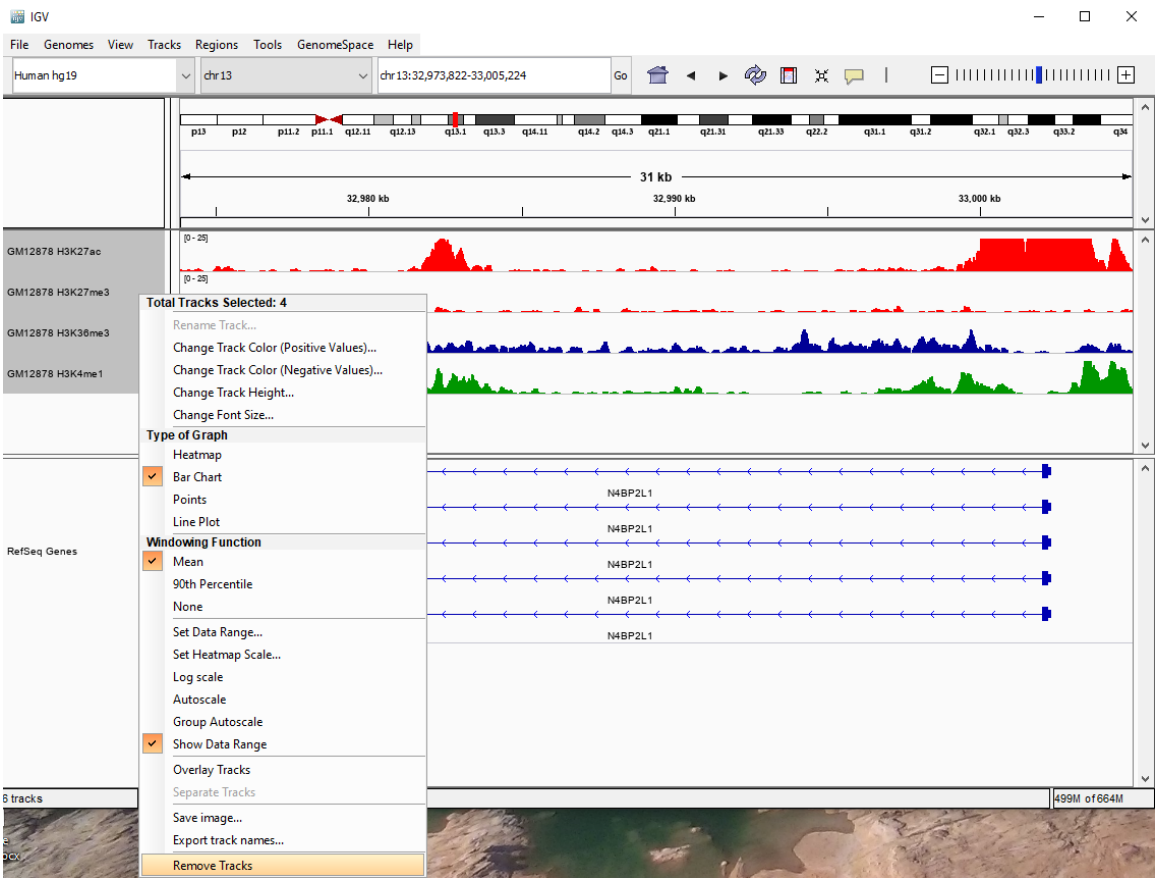
   E. What is the size of the first intron?

   F. In which track do you see the highest peak near the promoter?

   Answer:
   A. 22 isoforms.
   B. No.
   C. Exons 5 and 6 differ (the two at the 3'). Exon 5 is shorter in transcript 14 and it is fully coding region. In transcript 15, exon 5 is longer, part coding and part UTR. Exon 6 is part coding and part UTR in transcript 14 region and only UTR in transcript 15.
   D. MEDSFLQSFGRLSLQPQQQQQRQRPPRPPPRGTPPRRHSFRKHLYLLRGL PGSGKTTLAR
   E. Exon 1 coordinates: 33002041-33002267
      Exon 2 coordinates: 32981782-32981909
      Therefore intron 1 = 33002041 – 32981909 = 20132 bases
   F. GM12878 H3K27ac

Instructions (continuation) –

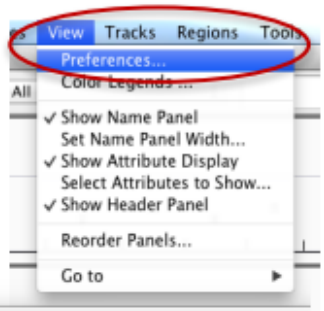
- Before continuing with the tutorial, remove the loaded tracks.
  Select the four loaded tracks using Shift and the mouse. Right click on the selection and choose Remove Tracks.



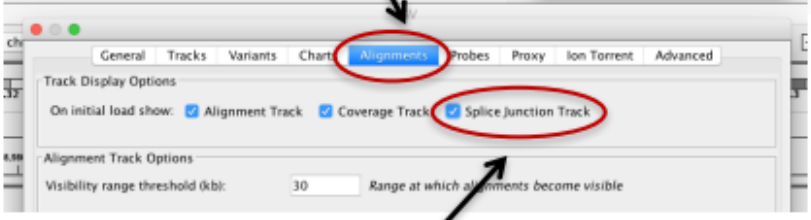
IGV Hands-on Exercise	Viewing RNA-Seq Data

1. Set preferences for viewing RNA-seq data

Click *View > Preferences*



Select *Alignments* tab

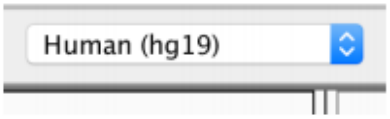


Check *Splice Junction Track*

In the newer version instead of "*Splice Junction Track*", it is written "*Show junction track*".

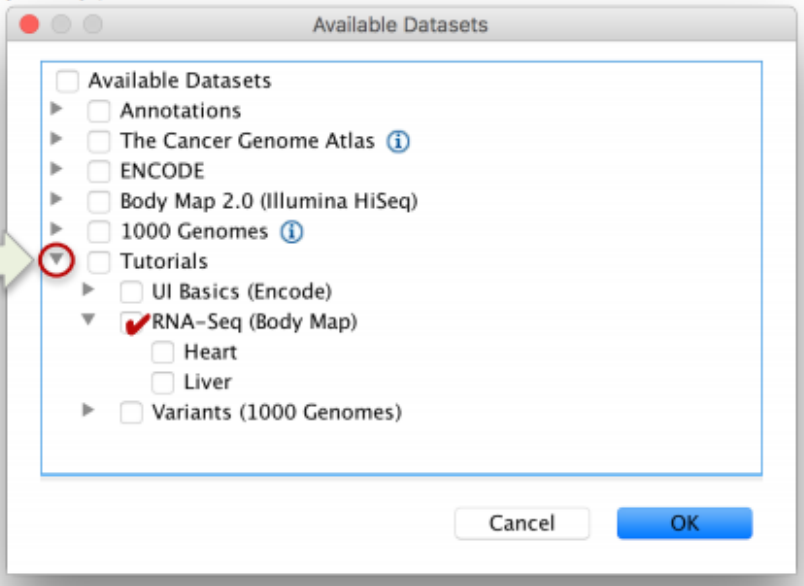
2. Load data

Select *Human hg19* from the genome dropdown menu



Click *File > Load from Server*

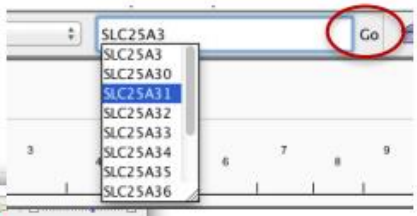
Open the *Tutorials* menu (Use ▶ on Mac, and + on Windows) and click on *RNA-Seq (Body Map)* and then click on *OK*

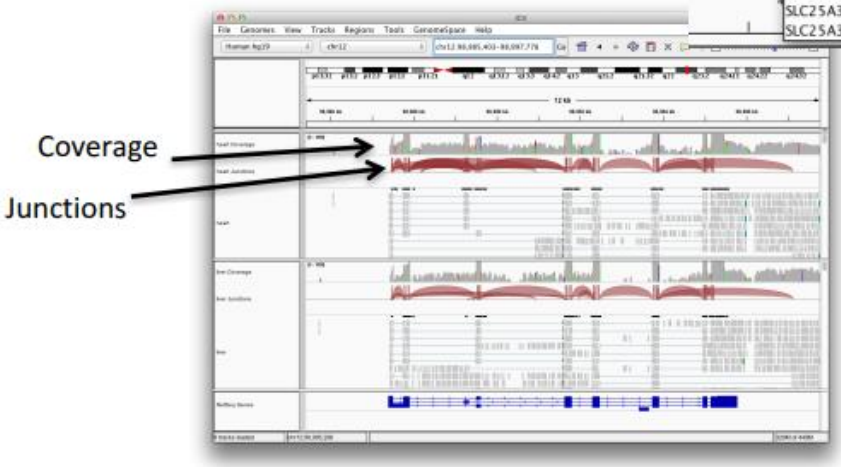


Make sure you only **open** the *Tutorials* menu. Do **not** check the box next to *Tutorials*. That will select everything under *Tutorials*, but we only want *RNA-Seq* for this exercise.

3. Jump to gene SLC25A3

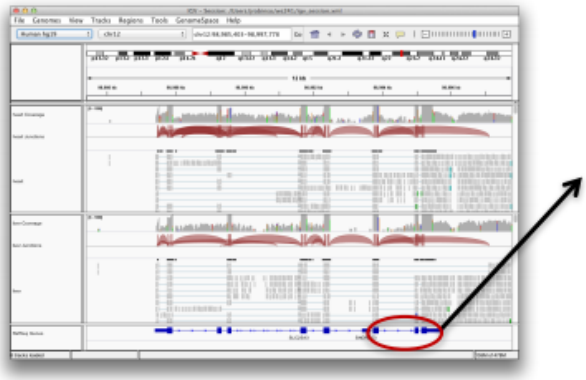
Type *SLC25A3* in the search box and click ***Go***

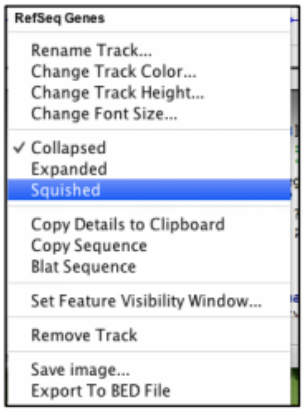




4. Expand gene track to see isoforms

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



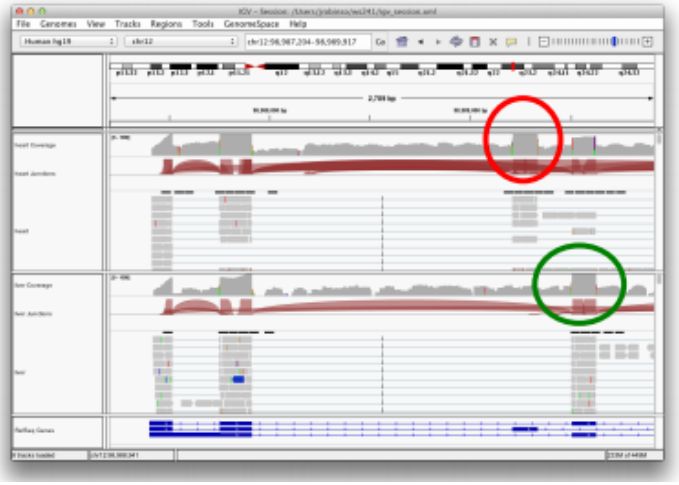


5. Zoom in on first 3 exons

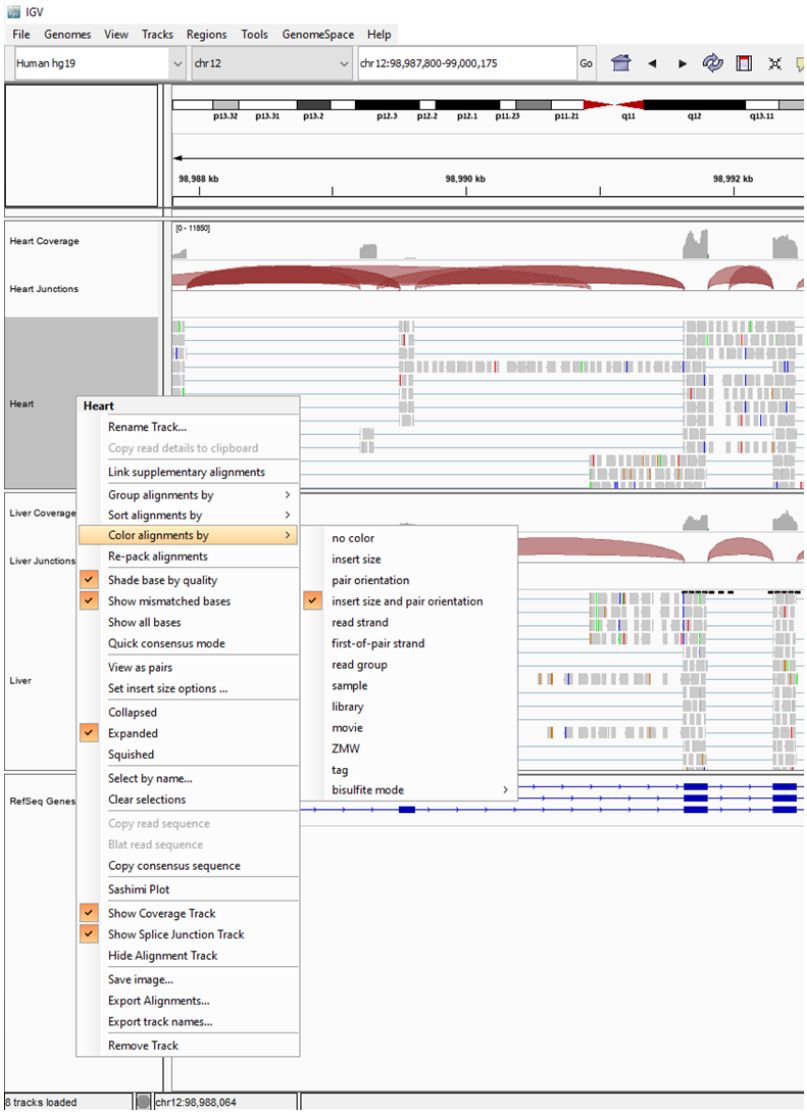
Click and drag in ruler region over area shown

6. Note evidence of alternative splicing.

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



Be sure that the Heart and Liver track alignments are colored by insert size and orientation according to the picture below (right click on the track).



Questions

3. What is the most probable format for the files loaded in this part of the tutorial?

   Answer: We are loading alignments and therefore bam files.

4. What are the gray rectangles you see in the browser?

   Answer: Each rectangle represent a read

5. Which isoforms of the SLC25A3 gene are most likely to be expressed in each tissue according to the data you see in IGV? Count the isoforms from top to bottom.

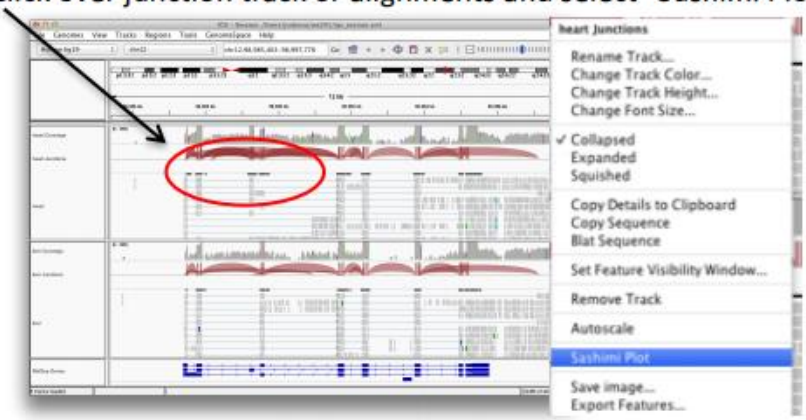
   Answer: In the heart isoform 1 is the most highly expressed, there might also be some low expression of the other two. In the liver only isoforms 2 and/or 3 are expressed and not 1.

7. Zoom back out to view whole gene

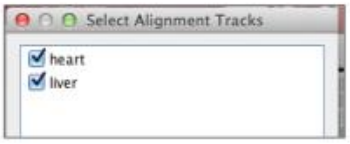
Click the back button in the command bar to zoom out to previous view

8. Open Sashimi plot

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

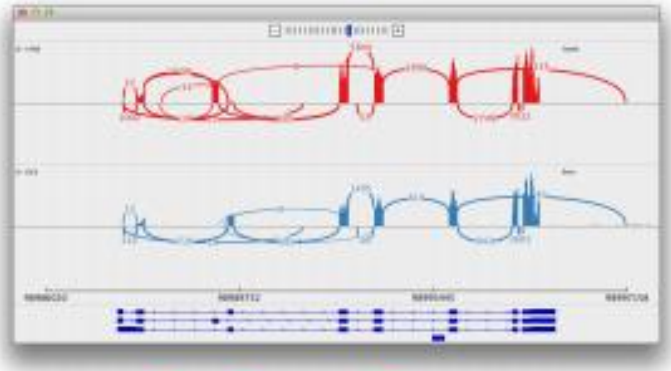


Verify both *heart* and *liver* are checked, and click *OK*



Sashimi Plot. Sashimi plots visualize splice junctions for multiple samples from their alignment data alongside genomic coordinates and a user-specified annotation track. IGV displays the Sashimi plot in a separate window and allows for more manipulations of the plots than the junctions track.

9. Examine Sashimi plot



Note:

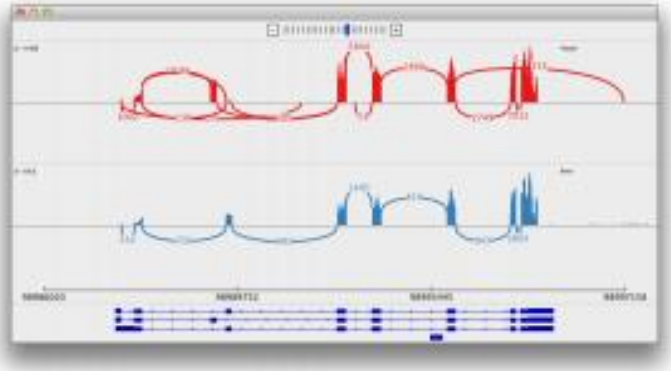
- Arcs represent reads spanning exon junctions
- Peaks represent exon coverage

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

Close the Sashimi plot.

Questions

6. A. In the liver sample, in the top read of the first exon, there is a green column, what does it mean?

   B. What is the reference genome base at the same location (coordinate)?

Answer:

A. There is a base mismatch (A) in the read according to the reference gene (C).
B. C

7. Go to chr21:33,979,695-34,159,621.
   A. Expand the RefSeq Genes track. Find 2 genes that are located head to head (5' UTR of both genes very close to each other). Zoom into this region.
      Which 2 genes are these? Do you think their promoters overlap?
   B. Which 2 genes have overlapping sequences in different orientations?
   C. Are there aligned reads in this region?

Answer:

A. SYNJ1 and PAXBP1-AS1. Yes.
   Additional option is PAXBP1 and NR_024622. Probably yes, the distance is ~250 bp.
B. PAXBP1-AS1 and PAXBP1
C. No

8. Go to chr12:99,041,218-99,051,312. Look at the blue and green reads in the intron of APAF1 in the liver sample. Explain their meaning.

   Answer: The green and blue reads indicate a structural variation, in this case it can be an inversion (blue reads) and a tandem duplication (green reads.

9. Look at the Coverage tracks. Are the reads evenly distributed along the exons?

Answer: No.

THE END