Visualization of Next Generation Sequencing Data using the Integrative Genomics Viewer (IGV v2.4+)
- TP -

Elodie Girard

Bioinformatics Engineer

Institut Curie – U900 Inserm – Mines ParisTech - PSL

elodie.girard@curie.fr

Table of Contents

Visualization of variants in DNA-seq data.....3

Visualization of RNA-seq data with the sashimi-plot view.....4

Visualization of the Transcription Factor GATA-3 BindingSites by ChIP-seq from ENCODE.....6

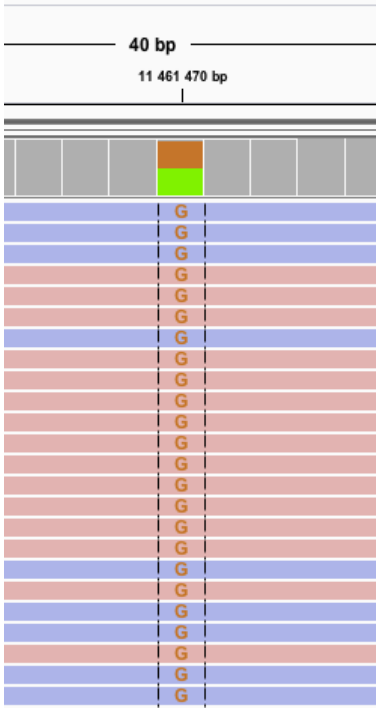
Training data is available on

https://zerkalo.curie.fr/partage/IGV/igv_data.zip

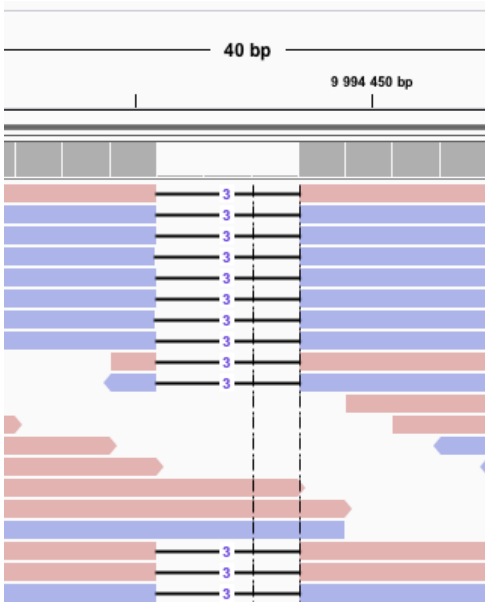
Visualization of variants in DNA-seq data

- In the top menu, select "**New session**"
- In the top menu, select "**Load from File**" then select the "**dnaseq.bam**" and the "**dnaseq.bed**" file

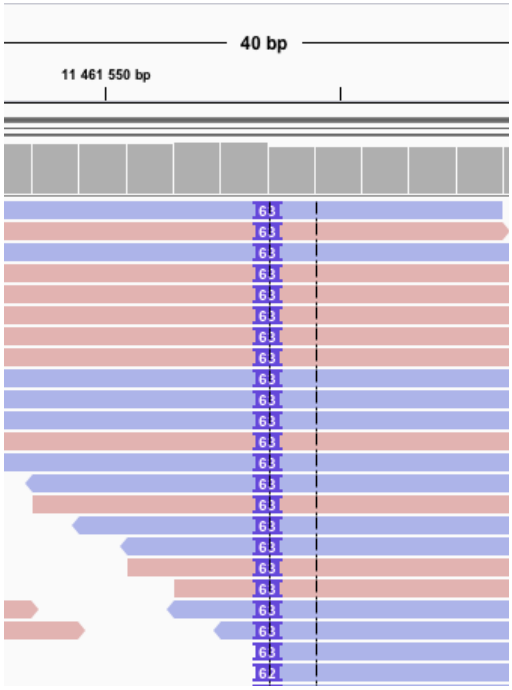
- Zoom in the ***"variant239"*** region and observe the heterozygous SNV at the position chr12:11,461,470, indicated by colored bases



- Zoom in the ***"variant230"*** region and observe the 3bp-deletion at the position chr2:9,994,446, indicated by a blank space crossed by a black line



- Zoom in the ***"variant240"*** region and observe the insertion at the position chr12:11,461,554 , indicated by a purple bar 



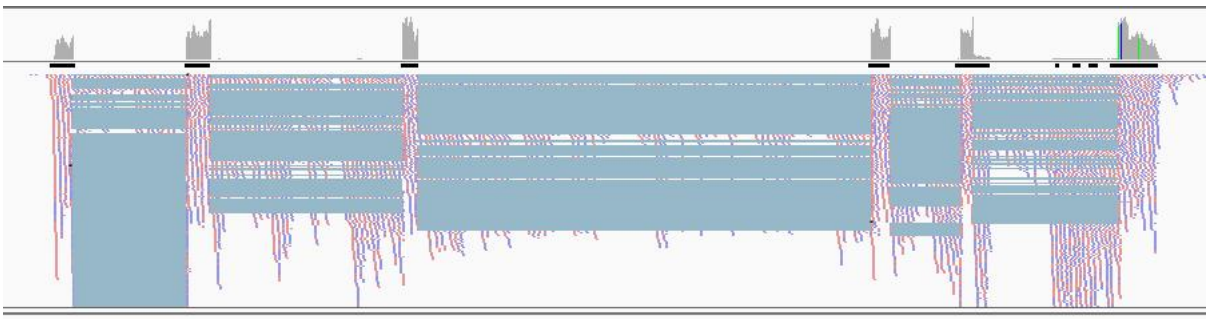
Tips:

- Always **sort reads by base** at a variant position and **color them by strand**
- Base counts indicated by hovering on the coverage track at the variant position can help assess the variant allelic ratio. The decomposition on the forward/reverse strands can help determine a **potential strand bias**: when a base is covered by the same amount of forward and reverse alignments (blue and pink) and the variant is supported by a high proportion of one type of strand (ie: 90%), it might be an artifact.

Visualization of an alternative isoform in RNA-seq data

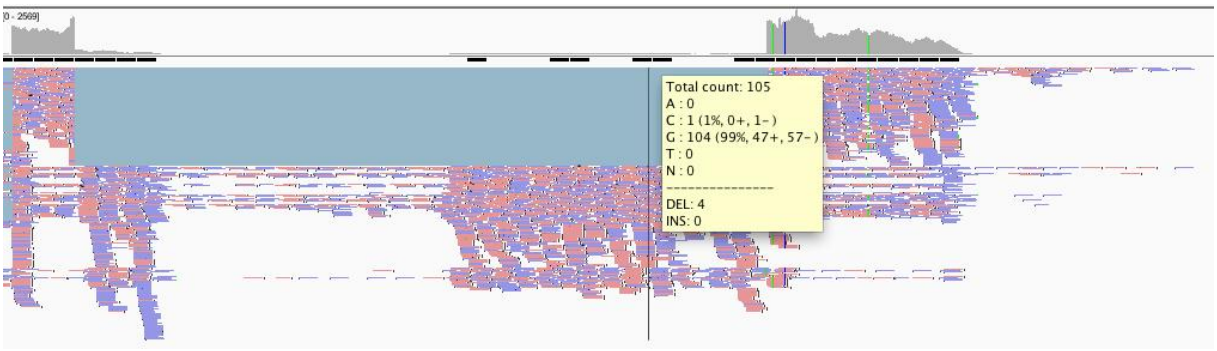
A variant impacting the splicing site of the exon 6 of the ***OAS1*** gene (chr12:113,357,193 ; Pickrell *et al*, 2012) is producing an alternative isoform that contains the retention of a part of the following intron.

- In the top menu, select "**New session**"
- In the top menu, select "**Load from File"** then select the "**rnaseq.bam**" file
- Zoom in the ***OAS1*** region. If you see the "*Zoom in to see alignments'* message, zoom in until you see alignments.

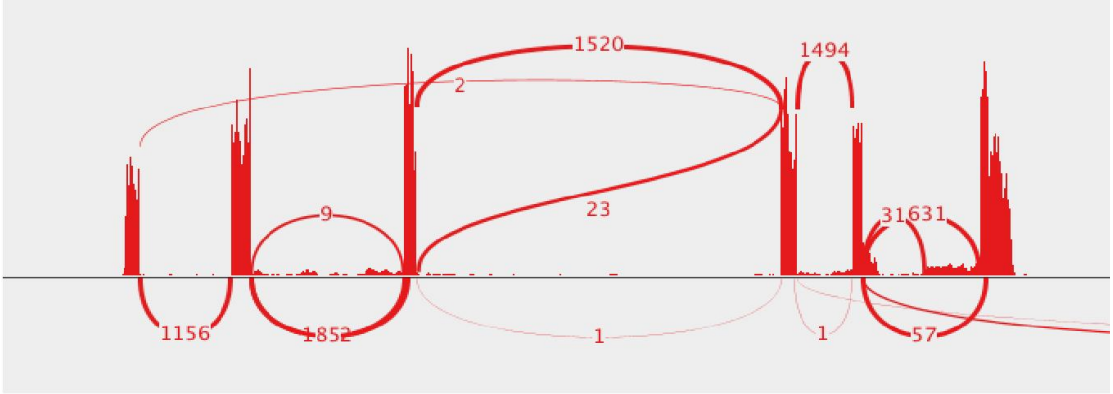


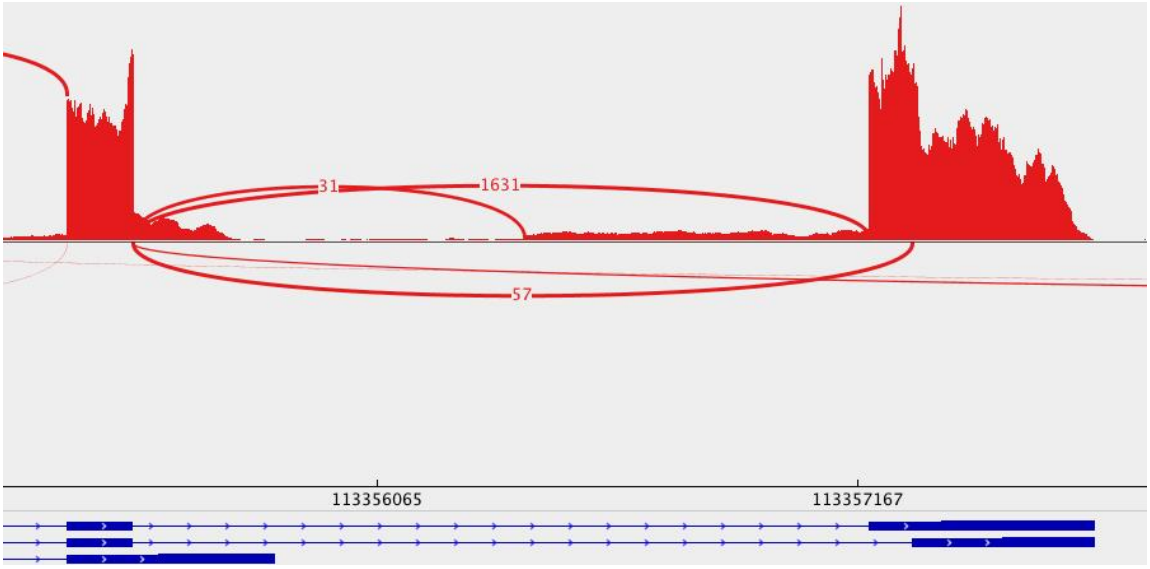
- Set the track in its **collapsed mode** by right clicking on the alignment track

The blue lines link the different parts of a spanning read that, by definition, map on several exons. Zoom in on the two last exons of *OAS1* then sort the alignments by base just before the last exon. You can see alignments outside of the known exons of this gene.



- Right click on the alignment track, select "**sashimi-plot**". You'll see the exons coverage and junctions lines with a number specifying the number of spanning reads for this junction. Zoom in the last 2 exons, move the track and click on an exon to only see junctions involving this exon (click on an intron to see everything).





- 31 reads span the junction of exon 5 and the cryptic 3' splice site upstream of the mutation
- Some options are available by right clicking on the sashimi:

➢ **Set color**: to distinguish between different tracks

➢ **Save image**: to save your sashimi (svg format is recommended for high resolution picture and can be modified using illustrator or inkscape)

➢ **Set min junction coverage**: alignment data are noisy and there are a lot of junctions with a low number of spanning reads. Put a higher number of minimal junction coverage to only see the higher represented junctions.

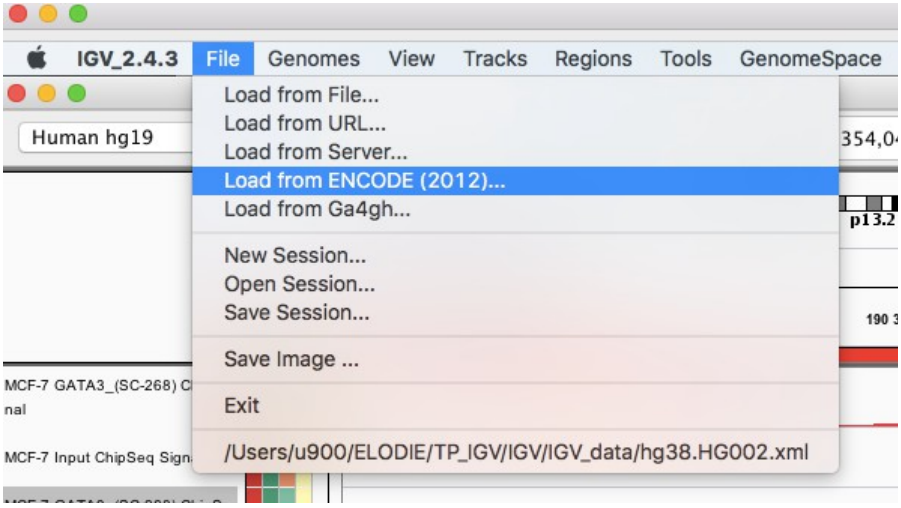
/!\ Sashimi plots are very useful to get a descriptive view of RNA-seq data but cannot replace a proper analysis : it's only a visualization tool. /!

Visualization of the Transcription Factor GATA-3 BindingSites by ChIP-seq from ENCODE

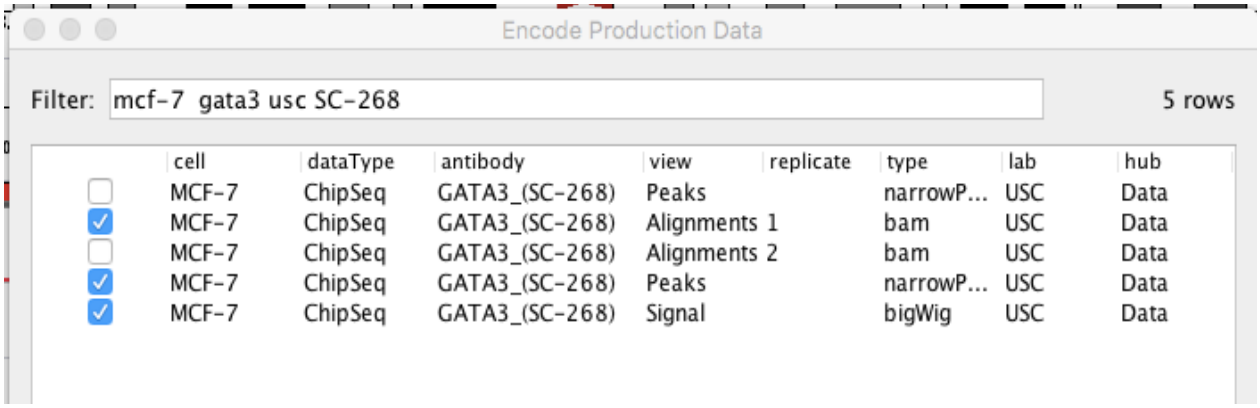
ChIP-seq data from the ENCODE project are used in this part in order to observe at the same time a BAM file containing the reads alignments, the normalized signal in a BIGWIG file and a BED file containing the enriched regions of high read density (peaks) identified by the bioinformatics analysis. These peaks correspond to the predicted binding sites of the studied transcription factor, GATA-3 in the Mcf7 cell line.

These data can be access directly via IGV (or load the saved session “chipseq.xml”) :

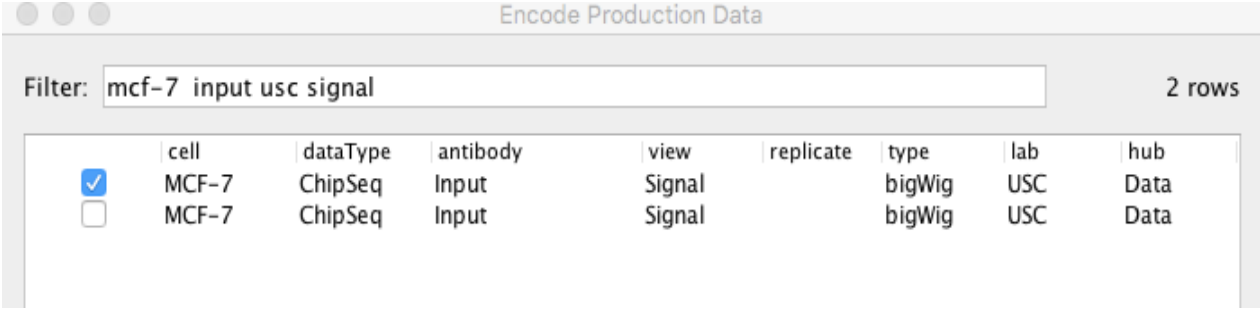
- In the top menu, select “**New session**”
- In the top menu, select “**File**” then “**Load from ENCODE (2012)**”



- Write the following keywords: “**mcf-7 gata3 usc SC-268**” and select the following tracks then click on “**Load**”



- Repeat the operation and write the following keywords to import the INPUT signal: "**mcf-7 input usc signal**" and select the following track then click on "**Load**"



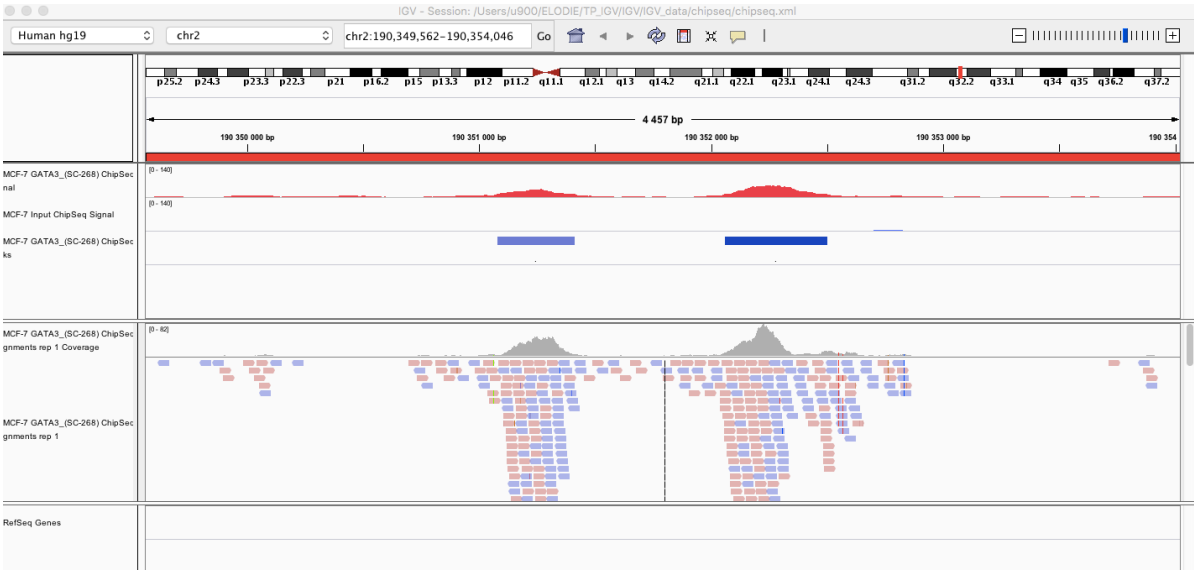
Visualization of two peaks :

- Zoom in this specific region : **chr2:190,349,562-190,354,046**

- Right click on the name of the "**MCF-7 GATA3**" IP Signal and click on "**Change track Color (positive values)**" then select the color **red**

- Repeat the operation for the "**MCF-7 Input**" Signal and select the color **light blue**

- Select both Signal tracks by maintaining the "**ctrl"** key and clicking on the names then **right-click** on one of them and click on "**Group autoscale"** to adapt the scale of both tracks



Visualization of a region that is not enriched in the IP :

- Remove the alignment track by **right clicking on the name** then select "**Remove track**"

- Zoom in this specific region : **chr20:55,741,582-55,790,445**

- Select both Signal tracks by maintaining the "**ctrl"** key and clicking on the names then **right-click** on one of them and click on "**Overlay"** to display both signals on the same track