Brief User Manual of IGV

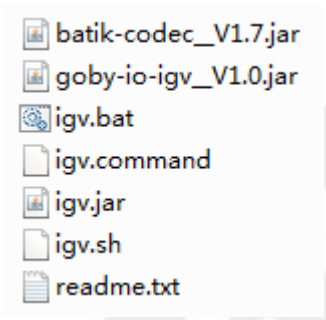
(for RNA-Seq Genome Alignment)

Contents

1. Preparing Softwares

1) IGV tool

We have downloaded a IGV package suitable for Windows, Linux, and Mac OS. Uncompress IGV_2.3.32.zip file under this folder and these following files will be got:



2) JAVA Environment

Operating system is required to install JRE(Java Runtime Environment) library before using IGV. Please visit this website to download one for your system if necessary: http://java.com/en/download/manual.jsp

2. Starting IGV

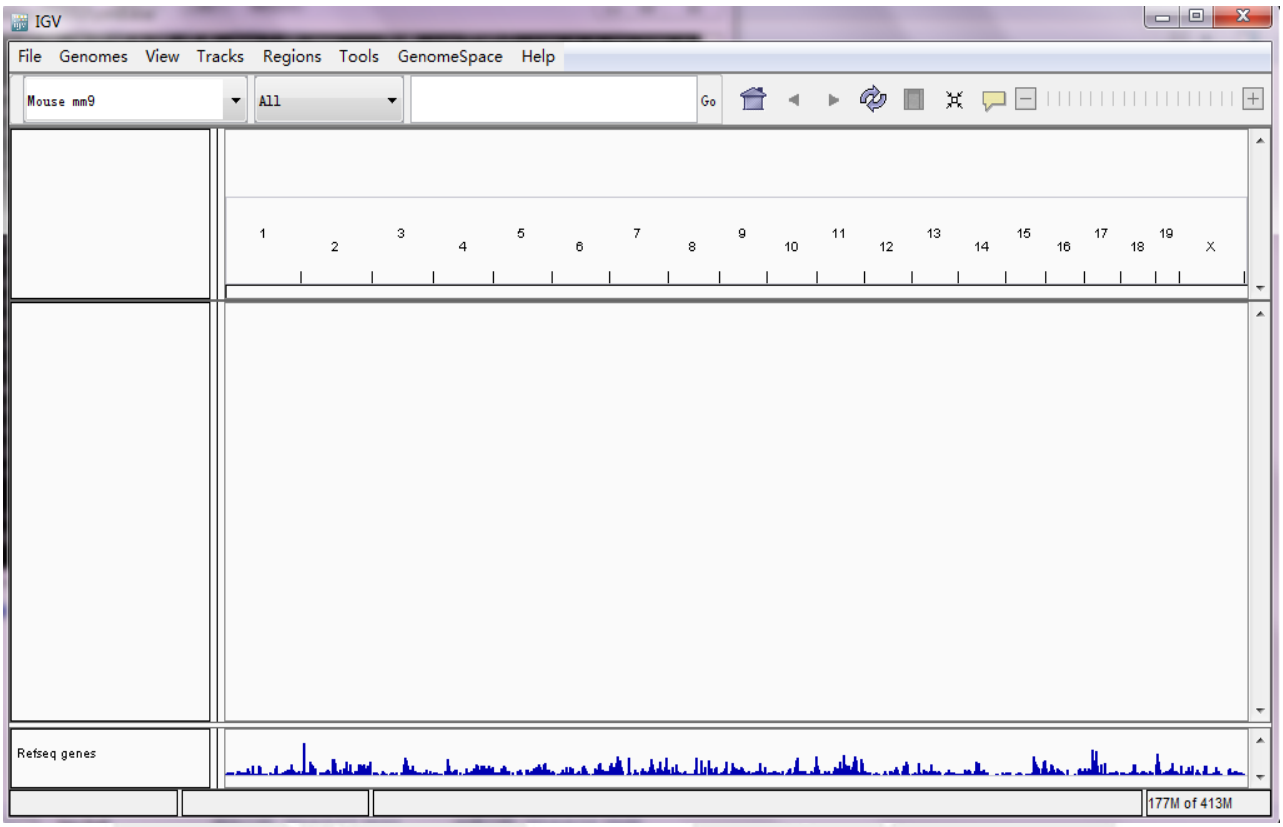
IGV package includes three types of commond file for different users , see as follow:

- Windows: igv.bat
- Linux: igv.sh
- Mac OS: igv.command

You can modify the config file using text editor and the parameter -Xmx1200m draws the main concern. The number 1200 presents memory size of system when running IGV. More memory is required if large file imported.

The following introduction in this user manual is based on Windows system. Double click igv.bat, then a cmd window will appear calling JAVA library automatically. When an IGV

interface as following comes out after a while, it means IGV has been started successfully.



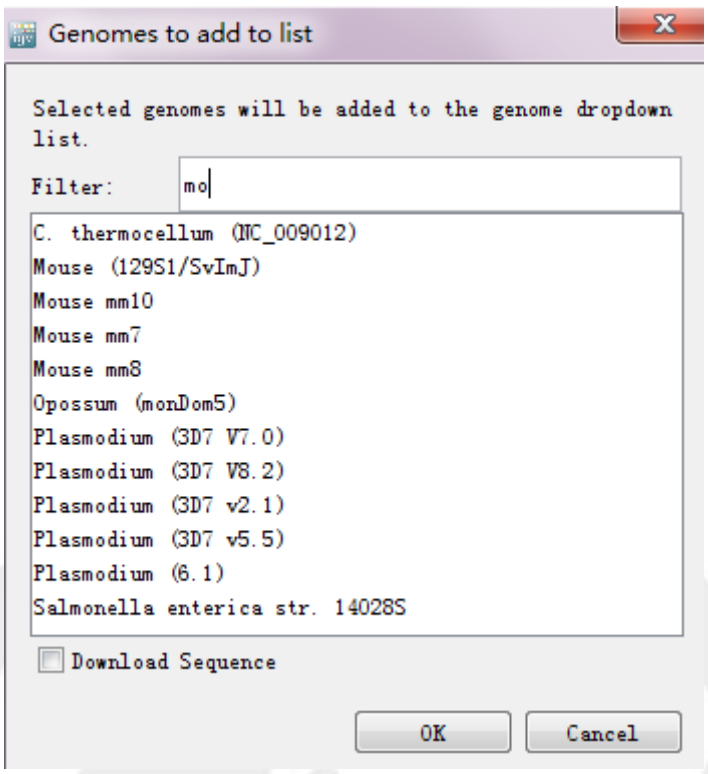
3. Importing Reference

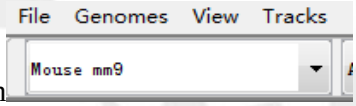
1. Loading local FASTA file

Click *Genome->Load Genome from File…* in menu bars, then choose the prepared local FASTA file through pop-up system window.

2. Loading from server

Make sure that your computer is able to visit Internet. Click *Genome->Load Genome from server…* in menu bars, then type key word of specie in the Filter blank through the pop-up window. You can quickly select it from matched list. You can check “Download Sequence” item to download reference from server in case disconnection with Internet in the future.





All imported references will be save in through these two second method. You can tap the saved list to cutover reference anytime. You can also click Genomes->Manage Genome List… in menu bars to add or delete references or change their orders. The second importing method is recommended since files on the server are usually exhaustive and gene annotation file will be helpful to view reads alignment.

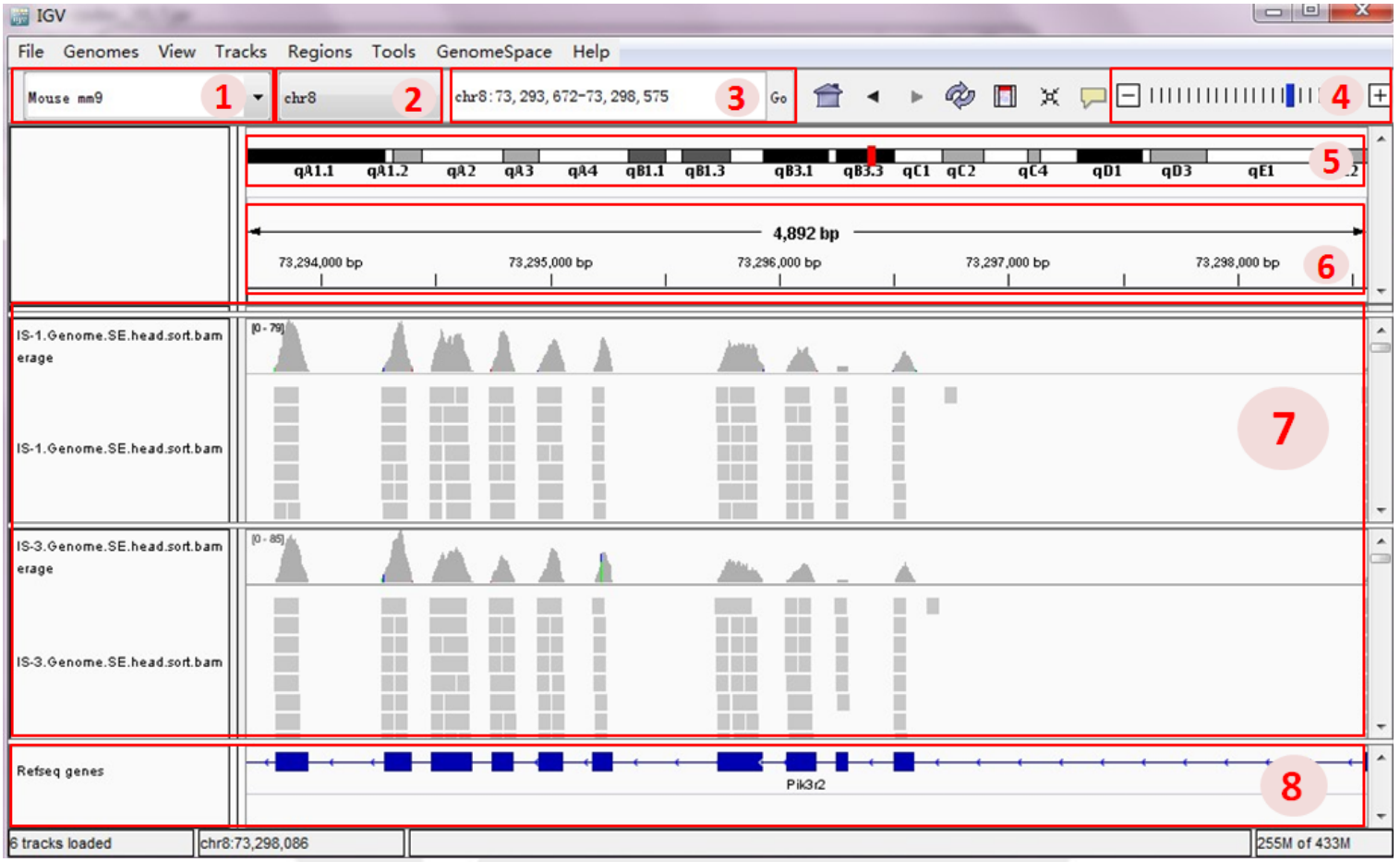
4. Importing BAM File

It is required that bam file and its corresponding bai file should be in the same folder. We provide bam and bai for all samples in project result.

Click *File->Load from File…*, choose one bam file from the pop-up system window. You can repeat this step to import multiple samples.

5. Viewing Alignment Information

It is able to view alignment information after sucessfully importing reference and BAM. Areas in the following complete interface will be introduced one by one.



① Managing references

See the third part “Importing Reference” of this manual.

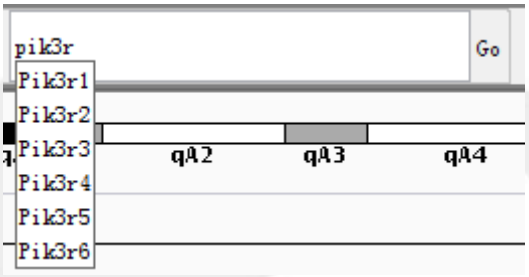
② Choosing chromosome

This dropdown menu will automatically display all chromosome’s name from imported reference. Choose any one that you want to view.

③ Searching for locus or gene elements

You can input following format to search alignment locus:

- A chromosome's name , e.g. chr1
- A loci on certain chromosome, e.g. chr1:73298086
- A region on certain chromosome, e.g. chr1:73293672-73298575
- If refGene is imported, gene elements search is supported here, e.g. Pik3r2. IGV will automatically match it in database and appear related list for quick selection.



After typing locus, click GO button and all the corresponding display region will changeaccording the new typed locus(It takes a while to run this step).

The button ◄ ► beside GO button is shortcut for locus that you have searched for. Left direction button goes to the previous one and right direction button goes to the next one.

④ Zooming in or out chromosome region

This function can be performed only in the situation that a chromosome has been located (if you choose All in ② area, it can't work.). Minus(-) zooms out current displayed region to a larger one ; plus(+) oppositely zooms it in to a smaller one. You can also click the vertical line between (–) and (+) to do quick zooming.

⑤ Displaying locus on chromosome

This area shows a complete composition of certain selected chromosome. Red rectangle indicates the location of current displayed region on chromosome. You can drag it to a new locus.

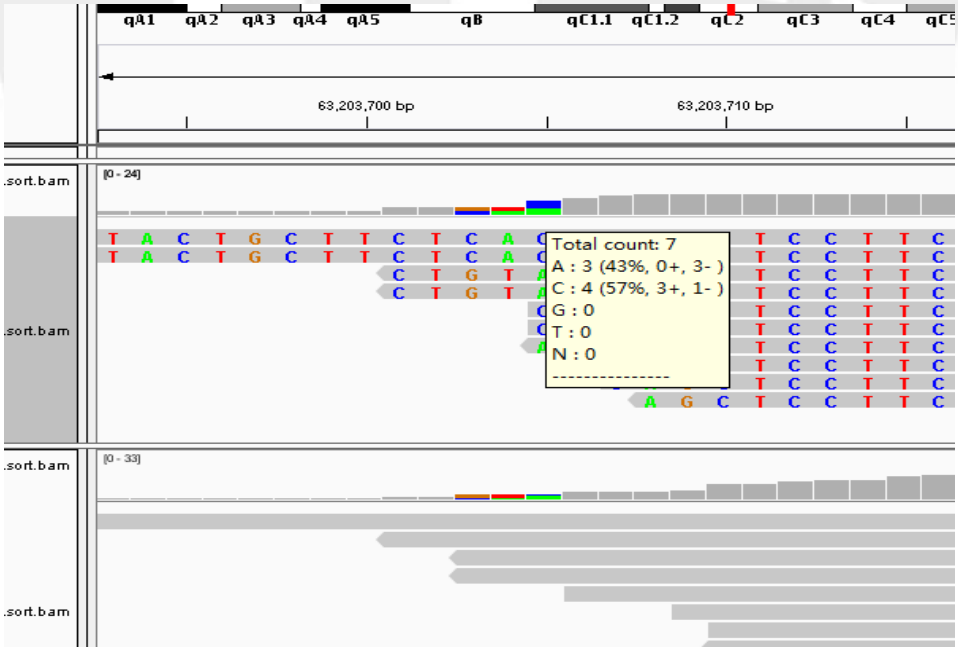
⑥ Scale of chromosome locus

This area only works when a chromosome is selected. The number between two arrows is length for current displayed region. The scale in bottom shows concrete coordinate on chromosome with three level(bp, kb, mb) , which can be adjusted automatically.

⑦ Displaying BAM alignment

This area display key alignment information of BAM file. IGV supports importing multiple samples, making users to do comparison on the same dimension conveniently.

Each sample has two tracks. One is coverage track, showing the coverage and depth situation; the other one is visualization of alignment. In each visualization alignment area, you can drag the vertical scroll bar or press direction keys(up, down) on keyboard to view all alignment reads. But horizontally, since all samples share the same scale, only direction keys(left, right) can work to go forword or go back for all samples at the same time.



These two pictures above, shows an example that how reads align to a small reference region for two samples. Four base types are in different colors(if one base has low quality, the color will be shallow).If there is only one base type aligned to certain loci, the corresponding coverage track is in gray; but if there are different types, the coverage track will count base numbers and display corresponding colors based on ratios.

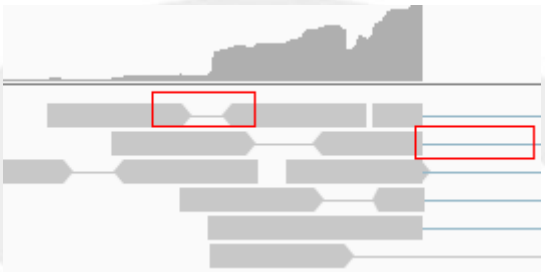
A yellow slight window as showing in the first above picture, will appear if you suspend mouse on a site in coverage track. It includes some useful alignment information(counts of each base type, ratio, counts from strand or anti-strand).

Another yellow slight window as showing in the second above picture, will also appear if you suspend mouse on a base in one read sequence. Useful information is collected and the text is easy to read.

Note that the first sample's alignment track shows all bases but the second one doesn't. You can right click alignment track and choose or cancel *Show all bases* to do cutover. If

you want to remove sample, right click and chose *Remove Track*.

For PE reads, you can right click and chose *View as pairs…* to look over pair-end alignment situation. Suspend mouse on read then two yellow boxes with reads information will appear. As show in the following picture, the single line in the left red box connects two pair-end reads. But note that the single line in the right red box shows junction alignment of one read.

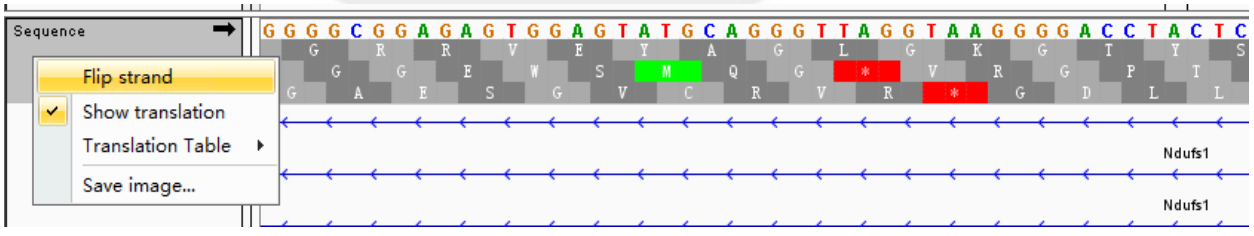


To do more operations of alignment, please right click and chose corresponding options.

⑧ Displaying reference

This area has two tracks. One is sequence of reference genome and the other one is transcripts of refseq genes.

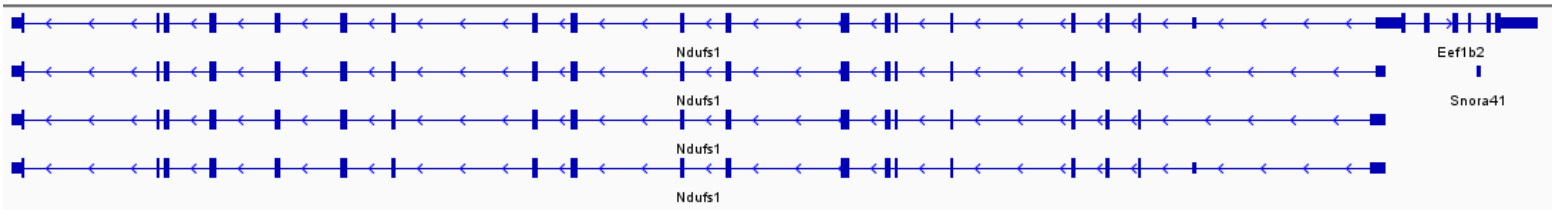
1. When zoom in sequence track to some degree, the sequences and black arrow will appear. Then you can right click and choose *Flip strand* to show anti-sequence; choose *Show translation* to show ORF(pen read frame) in three types. See examples in the following screenshot.



2. Refseq genes track shows transcript structure in three formats. These allow you to display overlapping features, such as different transcripts of a gene, on one line or multiple lines.
   - Collapsed, display transcripts on one line

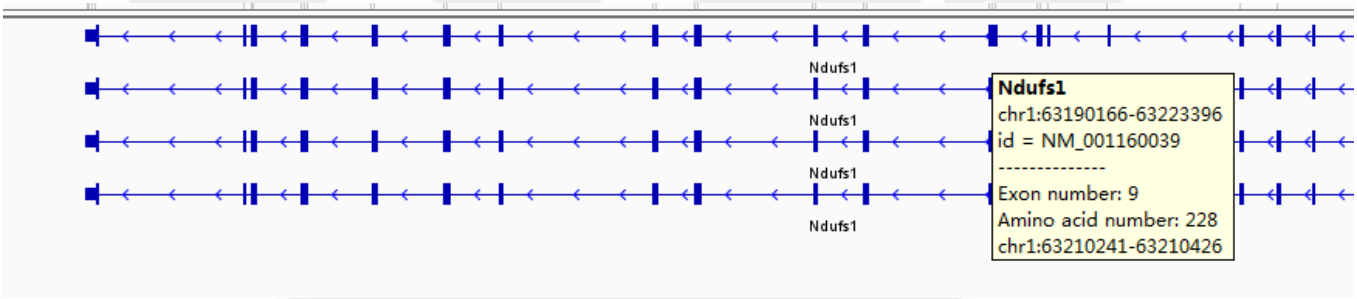
- Expanded，display transcripts on multiple lines
- Squished，similar with Expand

Right click in the track and choose one. The following picture is an example of Expanded format.We can clearly find that how homologous transcripts different from each other. Click one transcript line to go to NCBI website for more information.



Suspend mouse on certain exon, then a slight yellow window as following will appear including some useful information:

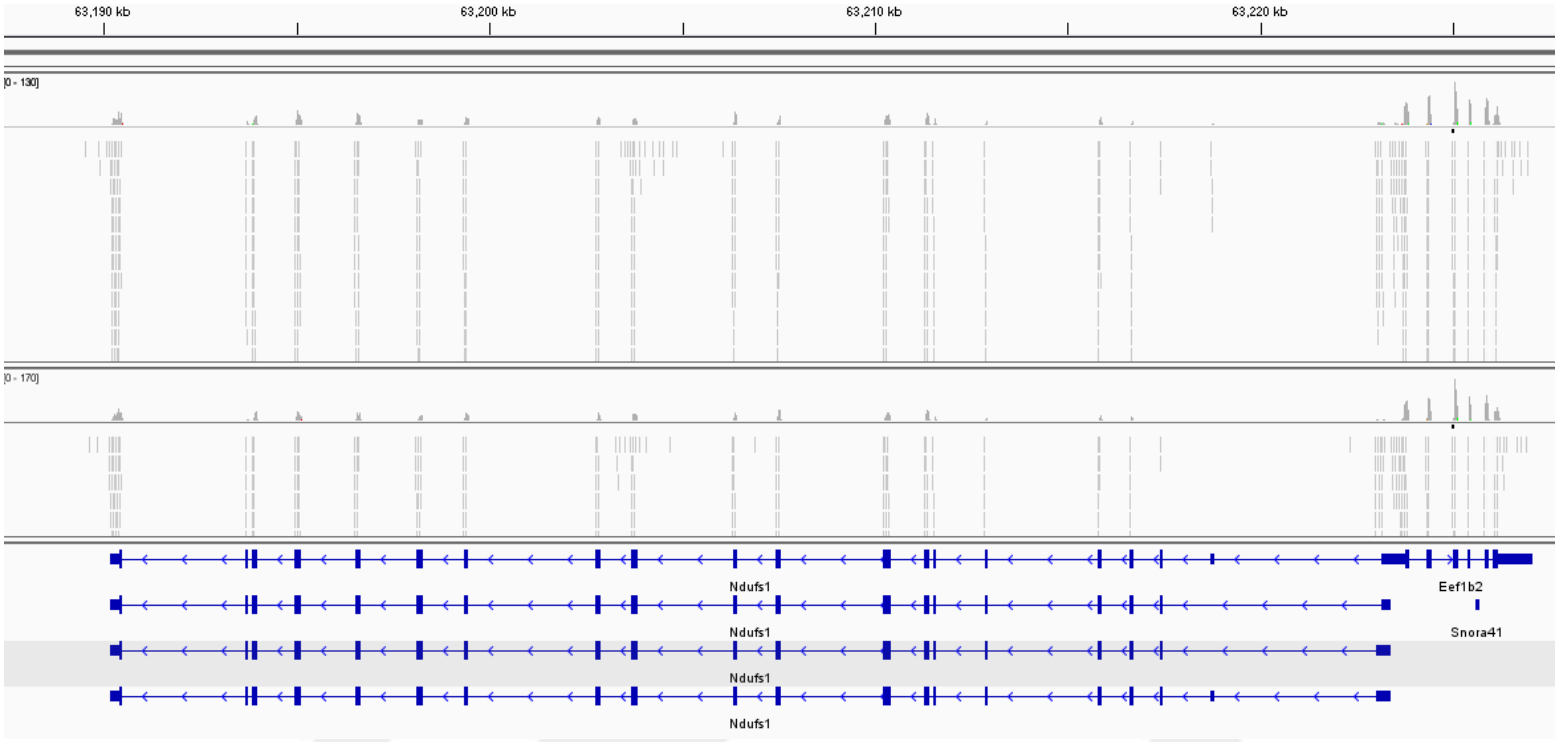
- gene name
- gene's location on chromosome
- transcript ID
- exon number
- amino acid number
- exon's location on chromosome



Pressing ctrl+b shortcut goes to the previous gene and ctrl+f goes to next one.

6. Others

IGV can do visualization of alignment data in different scale level, and have a friendly and intuitional interface. From the following screenshot, we clearly find out thow reads distribute to reference in a concrete and complete gene level between two samples.



This manual mainly does instruction for genome alignment visualization in RNA-Seq project. More information about how to use IGV is available in this website:

http://www.broadinstitute.org/software/igv/UserGuide