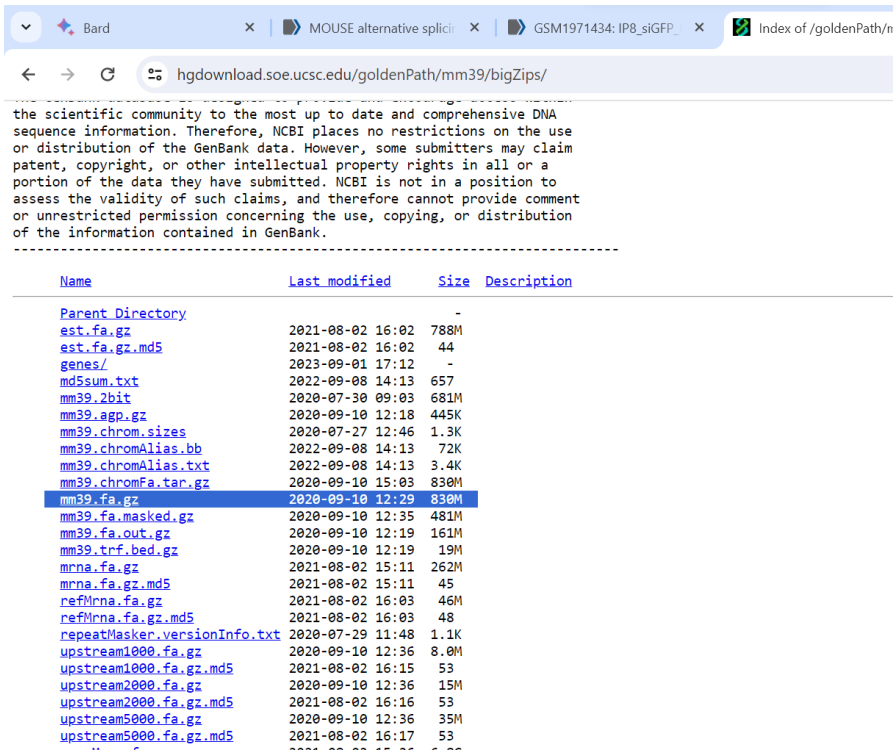
TAKING MOUSE AS REF SEQUENCE AND AN ALTERNATIVE SPLICE EVENT FROM SRA

SRR504382 our read id - paired end (Study: Muscleblind-Like 2 mediated alternative splicing in the developing bain by mRNA sequencing)

1)

Files included in this directory:

mm39.fa.gz - "Soft-masked" assembly sequence in one file.
Repeats from RepeatMasker and Tandem Repeats Finder (with period of 12 or less) are shown in lower case; non-repeating sequence is shown in upper case.



2) Next we created a new folder in sra data to store our prefetch our id and store fastq file

```
mkalpande@ManjushriK:~/sra_data$ mkdir altsplice
mkalpande@ManjushriK:~/sra_data$ ls
ERR12139009  SRR16235266.fastq  SRR3601860_1.fastq  fasterq.tmp.ManjushriK.62   outputq
ERR12139518  SRR23400676        altsplice           fasterq.tmp.ManjushriK.827
SRR16235266  SRR3601860         chr11human.fa       fasterq.tmp.ManjushriK.896
mkalpande@ManjushriK:~/sra_data$ cd altsplice/
mkalpande@ManjushriK:~/sra_data/altsplice$ prefetch SRR504382

2023-12-14T13:14:54 prefetch.2.11.3: Current preference is set to retrieve SRA Normalized Format files with full base qualit
y scores.
2023-12-14T13:14:55 prefetch.2.11.3: 1) Downloading 'SRR504382'...
2023-12-14T13:14:55 prefetch.2.11.3: SRA Normalized Format file is being retrieved, if this is different from your preferenc
e, it may be due to current file availability.
2023-12-14T13:14:55 prefetch.2.11.3:  Downloading via HTTPS...
2023-12-14T13:18:05 prefetch.2.11.3:  HTTPS download succeed
2023-12-14T13:18:06 prefetch.2.11.3:  'SRR504382' is valid
2023-12-14T13:18:06 prefetch.2.11.3: 1) 'SRR504382' was downloaded successfully
mkalpande@ManjushriK:~/sra_data/altsplice$ ls
SRR504382
mkalpande@ManjushriK:~/sra_data/altsplice$ cd SRR504382/
mkalpande@ManjushriK:~/sra_data/altsplice/SRR504382$ ls
SRR504382.sra
mkalpande@ManjushriK:~/sra_data/altsplice/SRR504382$ cd ..
mkalpande@ManjushriK:~/sra_data/altsplice$ fasterq-dump SRR504382 --split-files
spots read      : 21,196,948
reads read      : 42,393,896
reads written   : 42,393,896
mkalpande@ManjushriK:~/sra_data/altsplice$ cd SRR504382/
mkalpande@ManjushriK:~/sra_data/altsplice/SRR504382$ ls
SRR504382.sra
mkalpande@ManjushriK:~/sra_data/altsplice/SRR504382$ cd ..
mkalpande@ManjushriK:~/sra_data/altsplice$ ls
SRR504382  SRR504382_1.fastq  SRR504382_2.fastq
```

3) Then we move the fastq files to place where we want to do bwa

```
mv: target 'altsplice' is not a directory
mkalpande@ManjushriK:~/sra_data/altsplice$ mv SRR504382_1.fastq SRR504382_2.fastq /home/mkalpande/bwa_index_files/sample\ al
tsplice
```

4) Next we give permissions to file, gunzip it as follows:

```
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls -lrt
total 9597536
-rw-r--r-- 1 mkalpande mkalpande        172 Dec 14 17:39 mm39.fa.gz:Zone.Identifier
-rw-r--r-- 1 mkalpande mkalpande  870543764 Dec 14 17:39 mm39.fa.gz
-rw-r--r-- 1 mkalpande mkalpande 4478654890 Dec 14 18:50 SRR504382_1.fastq
-rw-r--r-- 1 mkalpande mkalpande 4478654890 Dec 14 18:50 SRR504382_2.fastq
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ chmod +x mm39.fa.gz
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls -lrt
total 9597536
-rw-r--r-- 1 mkalpande mkalpande        172 Dec 14 17:39 mm39.fa.gz:Zone.Identifier
-rwxr-xr-x 1 mkalpande mkalpande  870543764 Dec 14 17:39 mm39.fa.gz
-rw-r--r-- 1 mkalpande mkalpande 4478654890 Dec 14 18:50 SRR504382_1.fastq
-rw-r--r-- 1 mkalpande mkalpande 4478654890 Dec 14 18:50 SRR504382_2.fastq
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ unzip mm39.fa.gz
Archive:  mm39.fa.gz
  End-of-central-directory signature not found.  Either this file is not
  a zipfile, or it constitutes one disk of a multi-part archive.  In the
  latter case the central directory and zipfile comment will be found on
  the last disk(s) of this archive.
note:  mm39.fa.gz may be a plain executable, not an archive
unzip:  cannot find zipfile directory in one of mm39.fa.gz or
        mm39.fa.gz.zip, and cannot find mm39.fa.gz.ZIP, period.
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls
SRR504382_1.fastq  SRR504382_2.fastq  mm39.fa.gz  mm39.fa.gz:Zone.Identifier
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ gunzip mm39.fa.gz
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls
SRR504382_1.fastq  SRR504382_2.fastq  mm39.fa  mm39.fa.gz:Zone.Identifier
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$
```

5) Now will do indexing using <u>bwa index -p SRR82 file.fa</u>

Here, we are using -p as prefix as we can use same ref for other reads also.

```
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls
SRR504382_1.fastq  SRR504382_2.fastq  mm39.fa  mm39.fa.gz:Zone.Identifier
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ bwa index -p SRR82 mm39.fa
[bwa_index] Pack FASTA... 12.00 sec
[bwa_index] Construct BWT for the packed sequence...
[BWTIncCreate] textLength=5456444902, availableWord=395935328
[BWTIncConstructFromPacked] 10 iterations done. 99999990 characters processed.
[BWTIncConstructFromPacked] 20 iterations done. 199999990 characters processed.
[BWTIncConstructFromPacked] 30 iterations done. 299999990 characters processed.
[BWTIncConstructFromPacked] 40 iterations done. 399999990 characters processed.
```

```
[BWTIncConstructFromPacked] 590 iterations done. 5402330102 chara
[BWTIncConstructFromPacked] 600 iterations done. 5427525990 chara
[BWTIncConstructFromPacked] 610 iterations done. 5449916134 chara
[bwt_gen] Finished constructing BWT in 614 iterations.
[bwa_index] 2193.71 seconds elapse.
[bwa_index] Update BWT... 14.17 sec
[bwa_index] Pack forward-only FASTA... 12.57 sec
[bwa_index] Construct SA from BWT and Occ... 1259.67 sec
[main] Version: 0.7.17-r1198-dirty
[main] CMD: bwa index -p SRR82 mm39.fa
[main] Real time: 3501.289 sec; CPU: 3492.117 sec
```

Its done now.

6) Now we will do bwa mem for mapping of fastq files and ref sequence.

bwa mem SRR82 SRR504382_1.fastq SRR504382_2.fastq > SR82output.sam :we use

*** HALF SAM FILE IS GENERATED

```
[M::mem_pestat] skip orientation FF
[M::mem_pestat] skip orientation RF
[M::mem_pestat] skip orientation RR
[M::mem_process_seqs] Processed 250000 reads in 22.328 CPU sec, 22.136 real sec
[M::process] read 250000 sequences (10000000 bp)...
Killed
```

7) For sam to bam we will use following command:
samtools view -1 -bS SR82output.sam > SR82output.bam

samtools view: Starts the conversion process.

-1: Specifies the input file is in SAM format.

-b: Converts the output to BAM format.

-S: Sorts the alignments by reference coordinates before converting to BAM. This is crucial for efficient downstream analyses.

Got an error –

```
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ samtools view -1 -bS SR82output.sam > SR82output.bam
[W::sam_read1_sam] Parse error at line 20500063
samtools view: error reading file "SR82output.sam"
```

8) Next we will do sorting of bam fie followed by indexing..
samtools sort -T temp -o sorted_SR82output.bam SR82output.bam

- **samtools sort** orders the alignments by chromosomal coordinates, meaning reads from the beginning of the first chromosome will appear first in the file, followed by reads from the beginning of the second chromosome….
- **-T** This command tells Samtools to use the prefix **temp** for the temporary files and write the final sorted BAM file
- **-o** Specifies the path and filename for the output BAM file containing the sorted alignments.

```
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls
SR82output.bam  SRR504382_1.fastq  SRR82.amb  SRR82.bwt  SRR82.sa  mm39.fa.gz:Zone.Identifier
SR82output.sam  SRR504382_2.fastq  SRR82.ann  SRR82.pac  mm39.fa   sorted_SR82output.bam
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$
```

9) Next we will do indexing of sorted bam file
samtools index sorted_SR82output.bam

- **Samtools index**-The index file acts as a map, allowing tools to quickly find specific regions of the genome within the BAM file, improves the performance of many downstream analysis tasks, such as variant calling, read counting, and gene expression analysis.
- **.bai is index file**

```
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls
SR82output.bam      SRR504382_2.fastq  SRR82.bwt  mm39.fa                    sorted_SR82output.bam.bai
SR82output.sam      SRR82.amb          SRR82.pac  mm39.fa.gz:Zone.Identifier
SRR504382_1.fastq   SRR82.ann          SRR82.sa   sorted_SR82output.bam
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$
```

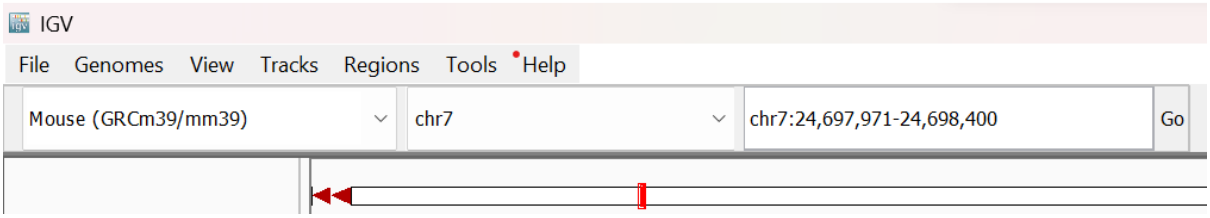
Q) How sorting and indexing helps?
ANS: Sorted BAM files are also required for using some analysis tools. Indexing further speeds up searching and retrieving specific reads within the sorted BAM file.

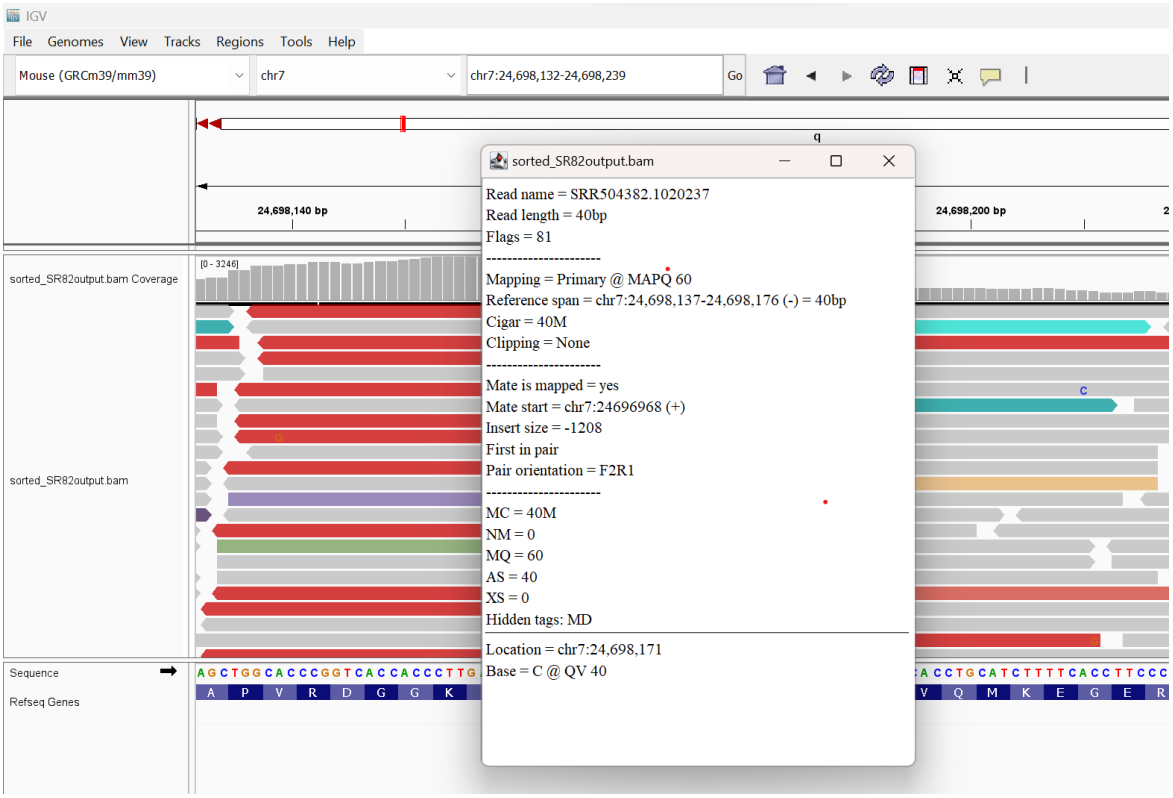
10) Next we will view our data in IGV

For input in IGV we selected file that is (sorted_SR82output.bam)

SRR504382.3	83	chr12	87154226	60	40M	=	87154198	-6
SRR504382.3	163	chr12	87154198	60	40M	=	87154226	68
SRR504382.4	83	chr7	90059234	60	40M	=	90059162	-1
SRR504382.4	163	chr7	90059162	53	17S23M	=	90059234	11
SRR504382.5	81	chr2	110605336	60	8S32M	=	110601293	-4
SRR504382.5	161	chr2	110601293	60	40M	=	110605336	40
SRR504382.6	83	chr12	110627590	60	40M	=	110627314	-3
SRR504382.6	163	chr12	110627314	60	40M	=	110627590	31
SRR504382.7	99	chr8	125759459	19	40M	=	125759559	14
SRR504382.7	147	chr8	125759559	19	40M	=	125759459	-1
SRR504382.8	81	chr16	13917702	60	40M	chr10	79966582	0
SRR504382.8	161	chr10	79966582	60	40M	chr16	13917702	0
SRR504382.9	99	chr11	68982026	60	40M	=	68982134	14
SRR504382.9	147	chr11	68982134	60	40M	=	68982026	-1
SRR504382.10	83	chr7	24698197	60	40M	=	24698096	-1
SRR504382.10	163	chr7	24698096	60	9S31M	=	24698197	14
SRR504382.11	99	chr17	26036685	60	40M	=	26036809	16
SRR504382.11	147	chr17	26036809	60	40M	=	26036685	-1
SRR504382.12	99	chr6	72295783	60	40M	=	72295926	17
SRR504382.12	147	chr6	72295926	60	34M6S	=	72295783	-1
SRR504382.13	83	chr2	72086743	60	40M	=	72086646	-1
SRR504382.13	163	chr2	72086646	60	40M	=	72086743	13
SRR504382.14	83	chrM	6389	27	40M	=	6274	-155	CGGAATTGTT

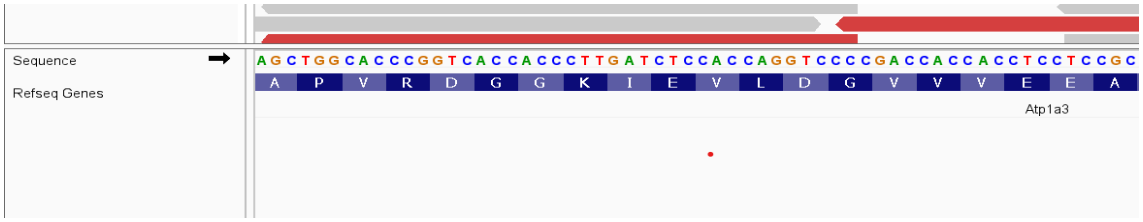
We selected this and enter the nucleotide number along with chr number in igv as cigar string is showing different



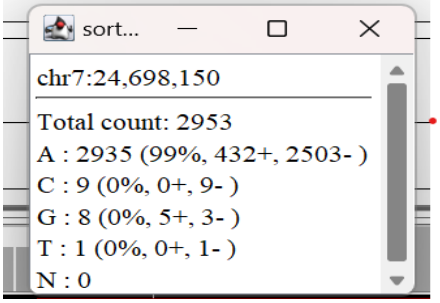


** Ideal MAPQ should be above 30.

#below are protein and nucleotide sequence:

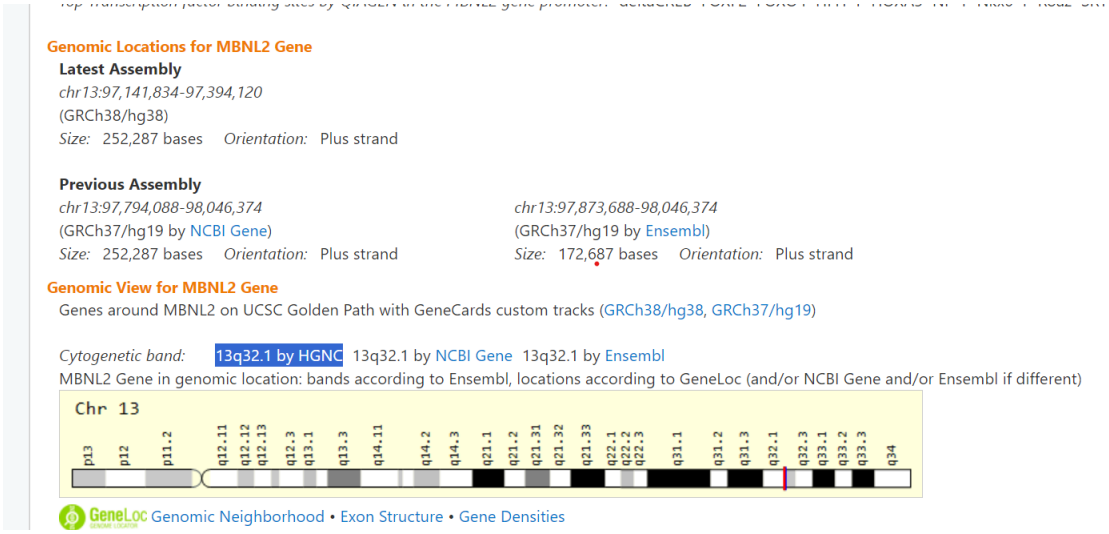


nucleotide number count – ideally the numbers should be 50% in other bases.



- *gene to search for
- **Mbnl2**, This gene encodes a C3H-type zinc finger protein that modulates alternative splicing of pre-mRNAs
- Research paper link - https://pubmed.ncbi.nlm.nih.gov/22884328/
- https://www.cell.com/neuron/fulltext/S0896-6273(12)00525-9?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0896627312005259%3Fshowall%3Dtrue
- Myotonic dystrophy is a rare progressive disorder that universally presents with weakness. In addition to musculoskeletal weakness, cardiac conduction defects and early cataracts are common.
- Importantly, the majority of the Mbnl2-regulated exons examined were similarly misregulated in DM. We propose that major pathological features of the DM brain result from disruption of the MBNL2-mediated developmental splicing program.
- DM can stand for Diabetes Myopathy or **Myotonic Dystrophy**
- Few highlights of Mbnl2
- Muscleblind-like 2 regulates alternative splicing in the brain
- Developmental regulation of splicing is disrupted in Mbnl2 knockout mice
- RNAs targeted by MBNL2 are misspliced in the myotonic dystrophy brain
- **C(C)UG** expression refers to the presence and levels of this expanded repeat in an individual's RNA. Studying C(C)UG expression helps understand DM2 severity and diagnose the disease.

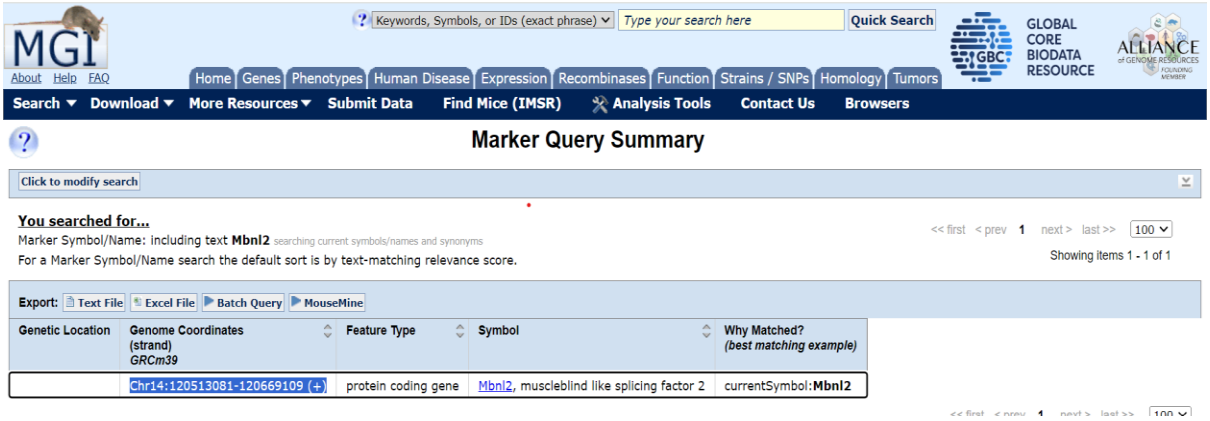
but this is for human FROM GENECARDS



##

The Mbnl2 gene in Mus musculus, the laboratory mouse, is located on **chromosome 11, band D1**. It occupies a specific region on this chromosome designated as q13.42.

searched on MGI (MOUSE GENOME INFORMATICS)



USING hisat2 as alignment tool-

Trying to set python path using venv

```
mkalpande@ManjushriK:~/hisat2-2.2.1$ python3 -m venv .venv
The virtual environment was not created successfully because ensurepip is not
available.  On Debian/Ubuntu systems, you need to install the python3-venv
package using the following command.

    apt install python3.10-venv

You may need to use sudo with that command.  After installing the python3-venv
package, recreate your virtual environment.

Failing command: /home/mkalpande/hisat2-2.2.1/.venv/bin/python3

mkalpande@ManjushriK:~/hisat2-2.2.1$ apt install python3.10-venv
E: Could not open lock file /var/lib/dpkg/lock-frontend - open (13: Permission denied)
E: Unable to acquire the dpkg frontend lock (/var/lib/dpkg/lock-frontend), are you root?
mkalpande@ManjushriK:~/hisat2-2.2.1$ sudo apt install python3.10-venv
[sudo] password for mkalpande:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  python3-pip-whl python3-setuptools-whl
The following NEW packages will be installed:
  python3-pip-whl python3-setuptools-whl python3.10-venv
0 upgraded, 3 newly installed, 0 to remove and 0 not upgraded.
Need to get 2473 kB of archives.
After this operation, 2884 kB of additional disk space will be used.
Do you want to continue? [Y/n] Y
Get:1 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 python3-pip-whl all 22.0.2+dfsg-1ubuntu0.4 [1680 kB]Get:2 http://
archive.ubuntu.com/ubuntu jammy-updates/universe amd64 python3-setuptools-whl all 59.6.0-1.2ubuntu0.22.04.1 [788 kB]
```

Used the following commands-

```
mkalpande@ManjushriK:~/hisat2-2.2.1$ python3.10 -m venv hisat2-env
mkalpande@ManjushriK:~/hisat2-2.2.1$ source hisat2-env/bin/activate
(hisat2-env) mkalpande@ManjushriK:~/hisat2-2.2.1$
```

```
(hisat2-env) mkalpande@ManjushriK:~/hisat2-2.2.1$ python -m pip install --upgrade pip
Requirement already satisfied: pip in ./hisat2-env/lib/python3.10/site-packages (22.0.2)
Collecting pip
  Downloading pip-23.3.2-py3-none-any.whl (2.1 MB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.1/2.1 MB 8.2 MB/s eta 0:00:00
Installing collected packages: pip
  Attempting uninstall: pip
    Found existing installation: pip 22.0.2
    Uninstalling pip-22.0.2:
      Successfully uninstalled pip-22.0.2
Successfully installed pip-23.3.2
```

So finally we used virtual environment---

```
mkalpande@ManjushriK:~/hisat2-2.2.1$ sudo apt install python3-virtualenv
[sudo] password for mkalpande:
Reading package lists... Done
Building dependency tree... Done
Reading state information... Done
The following additional packages will be installed:
  python3-distlib python3-filelock python3-platformdirs python3-wheel-whl
Suggested packages:
  python2-pip-whl python2-setuptools-whl
The following NEW packages will be installed:
  python3-distlib python3-filelock python3-platformdirs python3-virtualenv python3-wheel-whl
0 upgraded, 5 newly installed, 0 to remove and 0 not upgraded.
Need to get 411 kB of archives.
After this operation, 1791 kB of additional disk space will be used.
Do you want to continue? [Y/n] Y
Get:1 http://archive.ubuntu.com/ubuntu jammy/universe amd64 python3-distlib all 0.3.4-1 [269 kB]
Get:2 http://archive.ubuntu.com/ubuntu jammy/universe amd64 python3-filelock all 3.6.0-1 [8788 B]
Get:3 http://archive.ubuntu.com/ubuntu jammy/universe amd64 python3-platformdirs all 2.5.1-1 [14.2 kB]
Get:4 http://archive.ubuntu.com/ubuntu jammy-updates/universe amd64 python3-wheel-whl all 0.37.1-2ubuntu0.22.04.1 [38.0 kB]
Get:5 http://archive.ubuntu.com/ubuntu jammy/universe amd64 python3-virtualenv all 20.13.0+ds-2 [80.3 kB]
Fetched 411 kB in 3s (133 kB/s)
Selecting previously unselected package python3-distlib.
(Reading database ... 32598 files and directories currently installed.)
```

Then created virtual env & then **source venv/bin/activate** use this command to activate environment & to close the venv use **deactivate** command

```
mkalpande@ManjushriK:~/hisat2-2.2.1$ virtualenv venv
created virtual environment CPython3.10.12.final.0-64 in 230ms
  creator CPython3Posix(dest=/home/mkalpande/hisat2-2.2.1/venv, clear=False, no_vcs_ignore=False, global=False)
  seeder FromAppData(download=False, pip=bundle, setuptools=bundle, wheel=bundle, via=copy, app_data_dir=/home/mkalpande/.local/share/virtualenv)
    added seed packages: pip==22.0.2, setuptools==59.6.0, wheel==0.37.1
  activators BashActivator,CShellActivator,FishActivator,NushellActivator,PowerShellActivator,PythonActivator
```

1) Build reference genome- **hisat_mouse** -is the index file name

hisat2-build mm39.fa hisat_mouse

```
(venv) mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$ hisat2-build mm39.fa hisat_mouse
Settings:
  Output files: "hisat_mouse.*.ht2"
  Line rate: 6 (line is 64 bytes)
  Lines per side: 1 (side is 64 bytes)
  Offset rate: 4 (one in 16)
  FTable chars: 10
  Strings: unpacked
  Local offset rate: 3 (one in 8)
  Local fTable chars: 6
  Local sequence length: 57344
  Local sequence overlap between two consecutive indexes: 1024
  Endianness: little
  Actual local endianness: little
  Sanity checking: disabled
  Assertions: disabled
  Random seed: 0
  Sizeofs: void*:8, int:4, long:8, size_t:8
Input files DNA, FASTA:
  mm39.fa
Reading reference sizes
  Time reading reference sizes: 00:00:13
Calculating joined length
Writing header
Reserving space for joined string
Joining reference sequences
  Time to join reference sequences: 00:00:07
```

```
Returning block of 107796021 for bucket 8
Exited GFM loop
fchr[A]: 0
fchr[C]: 773810649
fchr[G]: 1326819314
fchr[T]: 1879875271
fchr[$]: 2654621783
```

```
Headers:
    len: 2654621783
    gbwtLen: 2654621784
    nodes: 2654621784
    sz: 663655446
    gbwtSz: 663655447
    lineRate: 6
    offRate: 4
    offMask: 0xfffffff0
    ftabChars: 10
    eftabLen: 0
    eftabSz: 0
    ftabLen: 1048577
    ftabSz: 4194308
    offsLen: 165913862
    offsSz: 663655448
    lineSz: 64
    sideSz: 64
    sideGbwtSz: 48
    sideGbwtLen: 192
    numSides: 13826156
    numLines: 13826156
    gbwtTotLen: 884873984
    gbwtTotSz: 884873984
    reverse: 0
    linearFM: Yes
Total time for call to driver() for forward index: 00:43:20
```

Files generated during indexing is-

```
(venv) mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$ ls
chr11.fa           hisat_mouse.2.ht2  hisat_mouse.4.ht2  hisat_mouse.6.ht2  hisat_mouse.8.ht2  mm39.fa
hisat_mouse.1.ht2  hisat_mouse.3.ht2  hisat_mouse.5.ht2  hisat_mouse.7.ht2  mm39.build
```

2) Running hisat2 so need to use the command-

hisat2 -p 4 -x /home/mkalpande/hisat2-2.2.1/hisat2_index_files/hisat_mouse -1 /home/mkalpande/hisat2-2.2.1/hisat2_index_files/SRR504382_1.fastq -2 /home/mkalpande/hisat2-2.2.1/hisat2_index_files/SRR504382_2.fastq -S hisat_alignment.sam

- -p 4/8 is number of threads which increases the speed of alignment of big genomes (need good system for this good RAM, cpu core)
- -x path to index file
- hisat_mouse is the index file name
- -1,2 represents read1, read2
- hisat_alignmnet.sam is output file name
- -S represents sam file

output is SAM file

```
mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$ hisat2 -p 4 -x /home/mkalpande/hisat2-2.2.1/hisat2_index_files/hisat_mouse -1
 /home/mkalpande/hisat2-2.2.1/hisat2_index_files/SRR504382_1.fastq -2 /home/mkalpande/hisat2-2.2.1/hisat2_index_files/SRR504382_2.fas
tq -S hisat_alignment.sam
21196948 reads; of these:
  21196948 (100.00%) were paired; of these:
    5683168 (26.81%) aligned concordantly 0 times
    14548882 (68.64%) aligned concordantly exactly 1 time
    964898 (4.55%) aligned concordantly >1 times
    ----
    5683168 pairs aligned concordantly 0 times; of these:
      1380294 (24.29%) aligned discordantly 1 time
    ----
    4302874 pairs aligned 0 times concordantly or discordantly; of these:
      8605748 mates make up the pairs; of these:
        4166215 (48.41%) aligned 0 times
        3695763 (42.95%) aligned exactly 1 time
        743770 (8.64%) aligned >1 times
90.17% overall alignment rate
mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$ ls
SRR504382_1.fastq  hisat_alignment.sam  hisat_mouse.3.ht2  hisat_mouse.6.ht2  hisat_mouse_indexes  venv
SRR504382_2.fastq  hisat_mouse.1.ht2    hisat_mouse.4.ht2  hisat_mouse.7.ht2  mm39.build
chr11.fa           hisat_mouse.2.ht2    hisat_mouse.5.ht2  hisat_mouse.8.ht2  mm39.fa
```

For bam file we use command-
samtools sort /home/mkalpande/hisat2-2.2.1/hisat2_index_files/hisat_alignment.sam -o/home/mkalpande/hisat2-2.2.1/hisat2_index_files/hisat_alignment.bam
OR or (samtools view -bS eg2.sam > eg2.bam)

- -o used to mention output file

We converted to **BAM** file

```
              Set level of verbosity
mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$ samtools sort /home/mkalpande/hisat2-2.2.1/hisat2_index_files/hisat_alignment
.sam -o/home/mkalpande/hisat2-2.2.1/hisat2_index_files/hisat_alignment.bam
[bam_sort_core] merging from 12 files and 1 in-memory blocks...
mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$ ls
SRR504382_1.fastq  hisat_alignment.bam  hisat_mouse.2.ht2  hisat_mouse.5.ht2  hisat_mouse.8.ht2    mm39.fa
SRR504382_2.fastq  hisat_alignment.sam  hisat_mouse.3.ht2  hisat_mouse.6.ht2  hisat_mouse_indexes  venv
chr11.fa           hisat_mouse.1.ht2    hisat_mouse.4.ht2  hisat_mouse.7.ht2  mm39.build
mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$
```

**** REMEMBER as we are using venv so we have to use that environment only

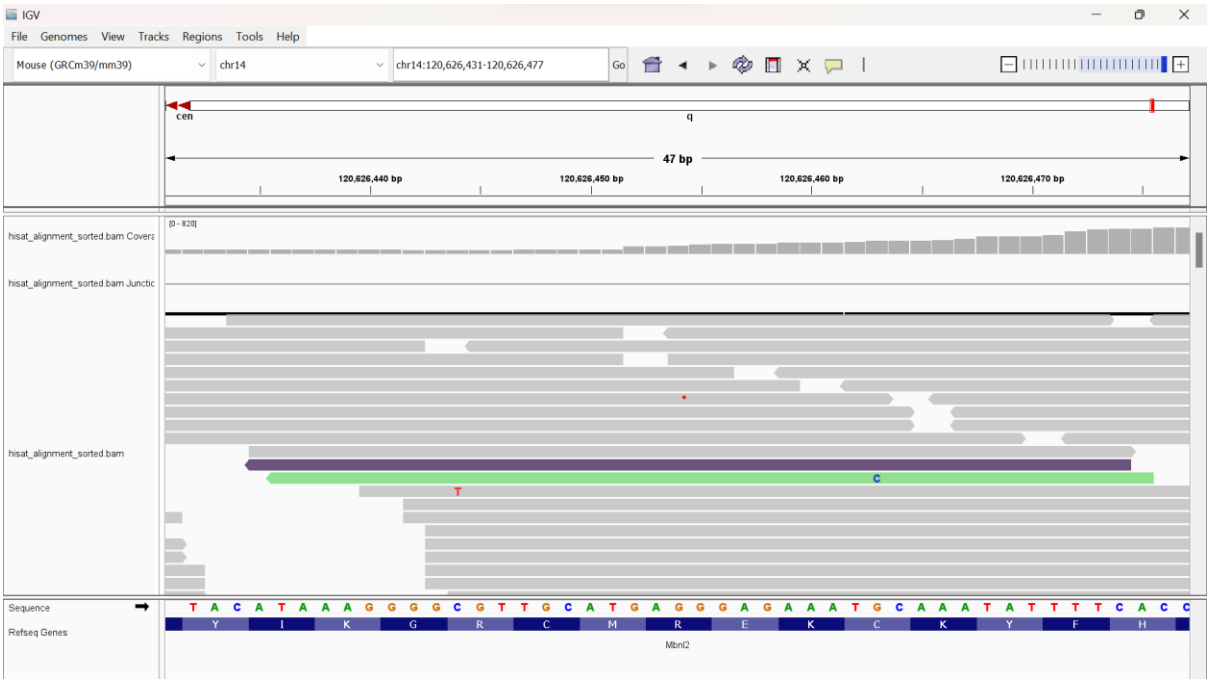
3) Creating a sorted bam file using command

samtools sort hisat_alignment.bam -o hisat_alignment_sorted.bam

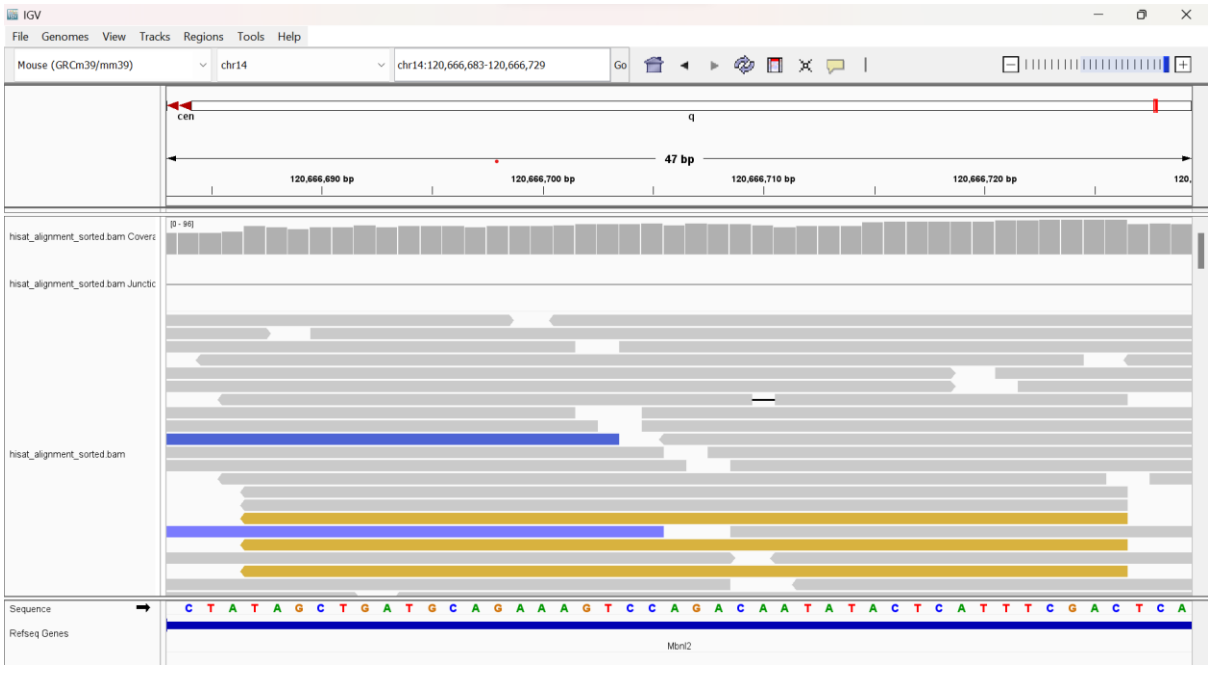
4) creating a index file

samtools index hisat_alignment_sorted.bam

5) Next we will be using IGV to analyse the alignment



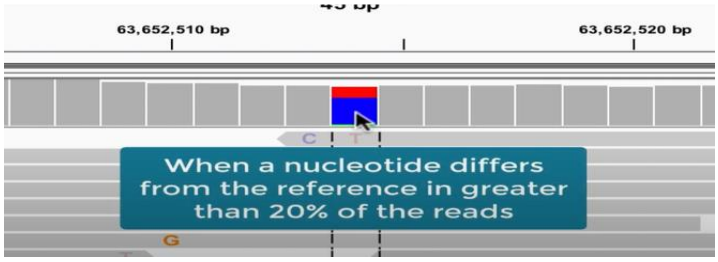
Whts the black line?



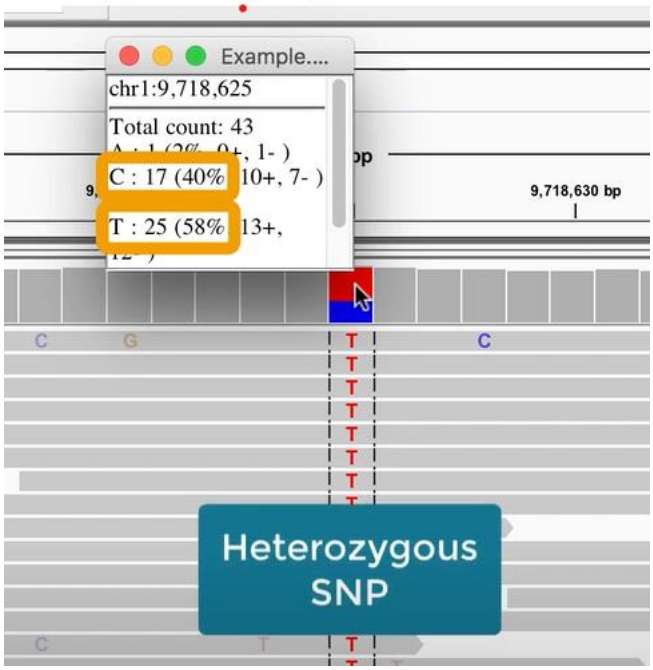
*Points of IGV

1. Bolder colours are higher quality and more likely to be real mismatches @QV
2. Lighter colours are lower quality and less likely to be real mismatches @QV
3. Filled grey reads have mapping quality different
4. Hollow reads have @MQ different = 0

5. Purple colour I represents insertion
6. When multiple bases are deleted then it is shown by **black bold** line as in image above.
7. Such major colour pattern occurs:



8. We can sort the alignments by base and many other options.
9. This fairly even split of bases shows heterozygous snp

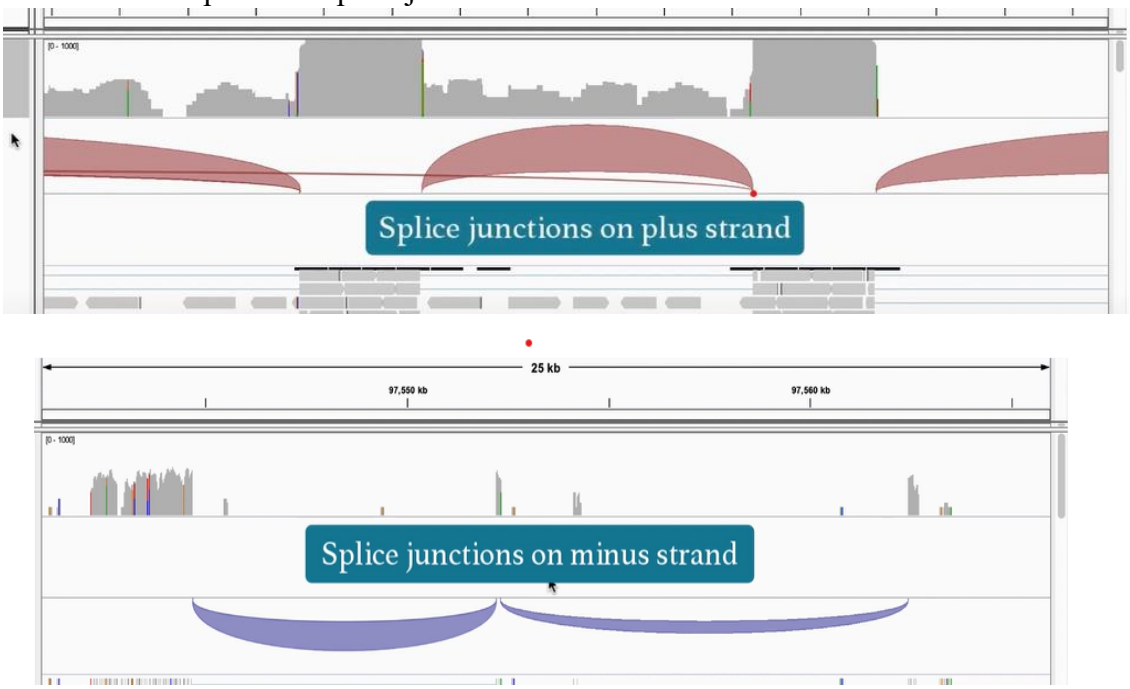


10. Colour alignments by read strand

Red reads are forward strands & blue reads are reverse strand

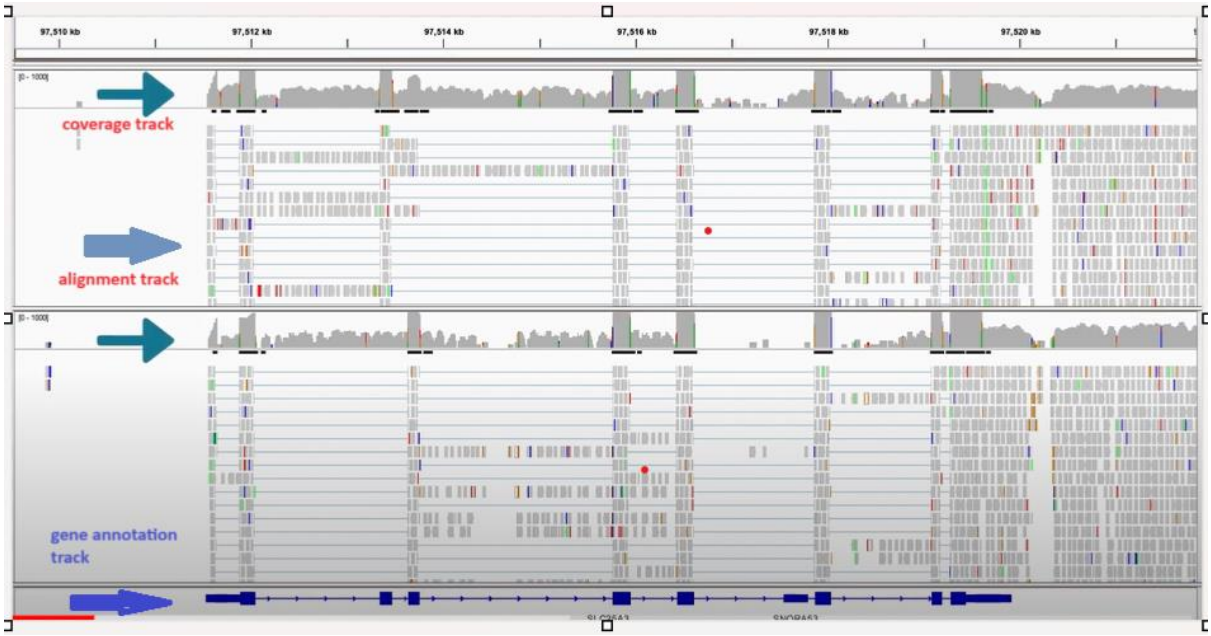
strand bias??????

11. Brown colour uplifts are splice junctions



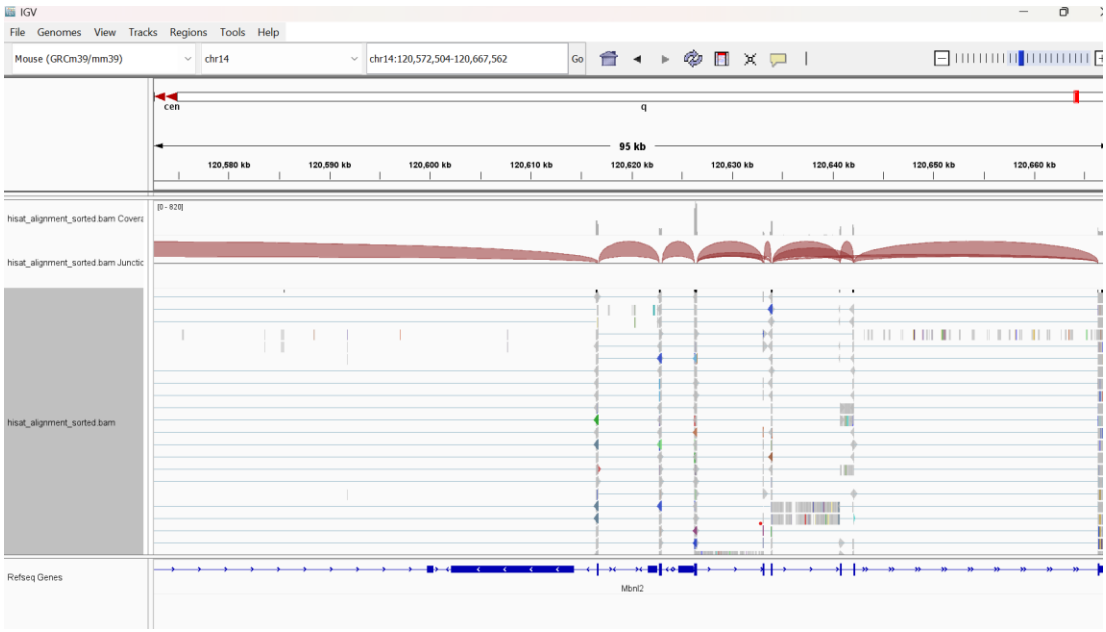
Thicker splice junction have more reads than other.

12. #sample pic

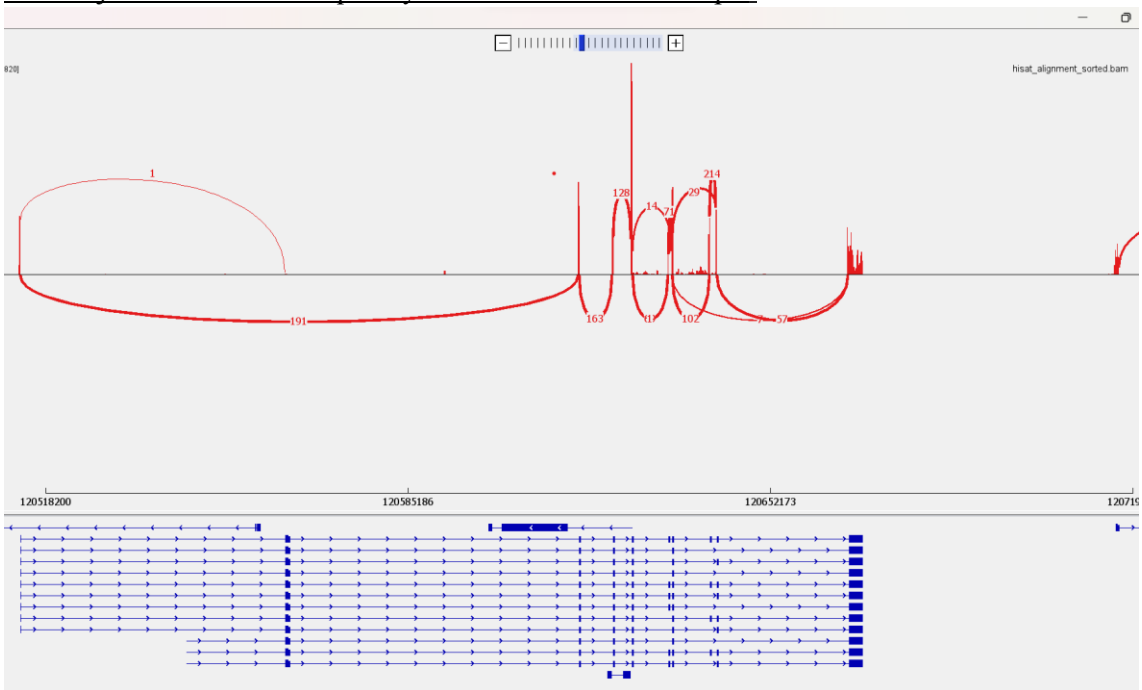


** SHASHIMI PLOT IS USED TO STUDY SPLICE JUNCTIONS

- Present in IGV only
- Right click on screen and select then it opens
- Our Splice junctions(brown in colour)

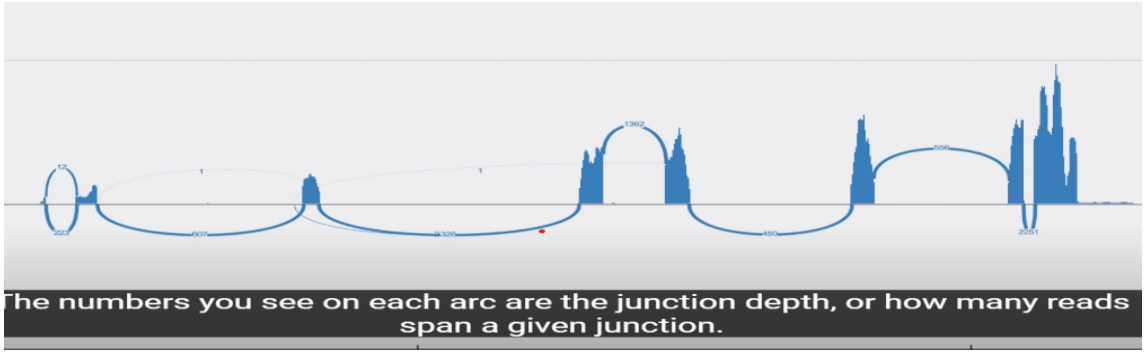


- Our shashimi plot result
- Junction depth refers to the number of reads that map across a specific splice junction. It is a measure of the abundance of that junction in the sample. A higher junction depth indicates that the junction is more frequently used in the RNA transcripts.



- The bottom blue lines from above image shows genes isoforms

#SAMPLE


12
1
1
1362
556
223
607
326
150
2251
The numbers you see on each arc are the junction depth, or how many reads span a given junction.