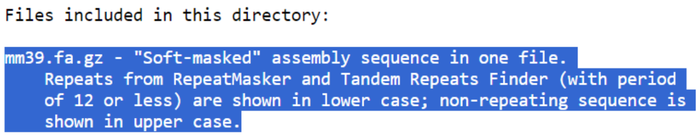
SAMPLE PROJECT ON RNA-SEQUENCING

Name- Manjushri Kalpande

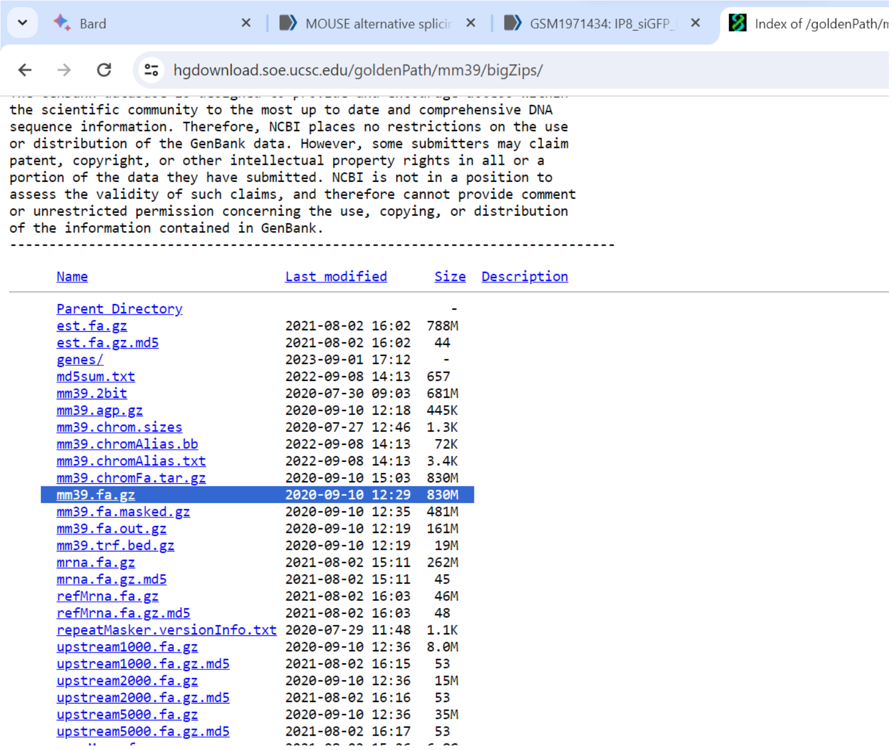
SRR504382 our read id - paired end (Study: Muscleblind-Like 2 mediated alternative splicing in the developing bain by mRNA sequencing)



```
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.
```

1) Downloaded a genome sequence of mouse



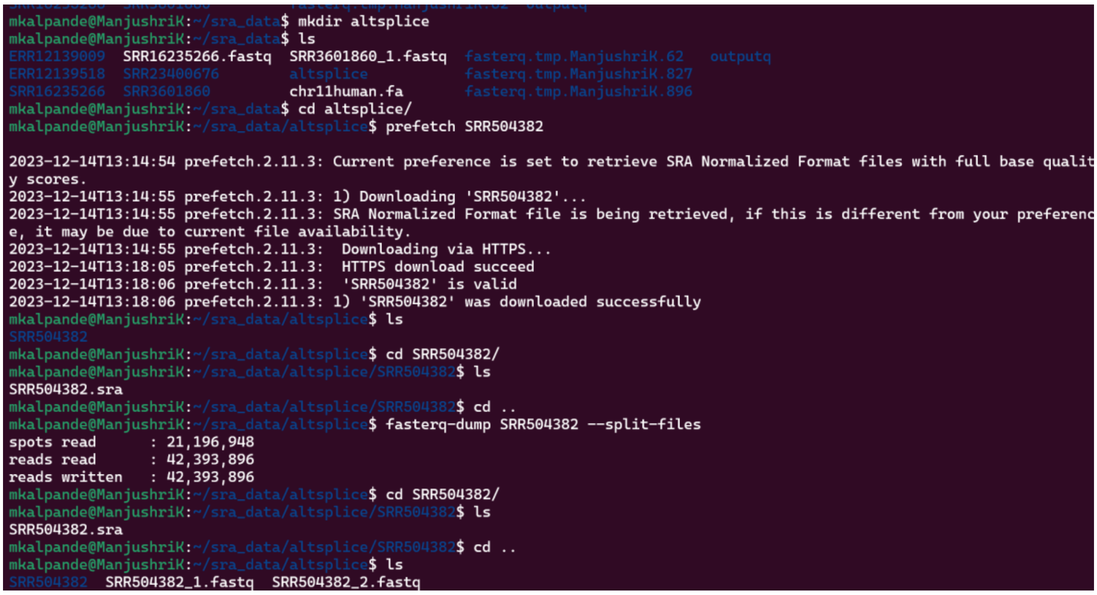
Bard
MOUSE alternative splicir
GSM1971434: IP8_siGFP_
Index of /goldenPath/n
hgdownload.soe.ucsc.edu/goldenPath/mm39/bigZips/

```
the scientific community to the most up to date and comprehensive DNA
sequence information. Therefore, NCBI places no restrictions on the use
or distribution of the GenBank data. However, some submitters may claim
patent, copyright, or other intellectual property rights in all or a
portion of the data they have submitted. NCBI is not in a position to
assess the validity of such claims, and therefore cannot provide comment
or unrestricted permission concerning the use, copying, or distribution
of the information contained in GenBank.
-----------------------------------------------------------------------------

      Name                         Last modified      Size  Description

      Parent Directory                                  -
      est.fa.gz                    2021-08-02 16:02  788M
      est.fa.gz.md5                2021-08-02 16:02   44
      genes/                       2023-09-01 17:12    -
      md5sum.txt                   2022-09-08 14:13  657
      mm39.2bit                    2020-07-30 09:03  681M
      mm39.agp.gz                  2020-09-10 12:18  445K
      mm39.chrom.sizes             2020-07-27 12:46  1.3K
      mm39.chromAlias.bb           2022-09-08 14:13   72K
      mm39.chromAlias.txt          2022-09-08 14:13  3.4K
      mm39.chromFa.tar.gz          2020-09-10 15:03  830M
      mm39.fa.gz                   2020-09-10 12:29  830M
      mm39.fa.masked.gz            2020-09-10 12:35  481M
      mm39.fa.out.gz               2020-09-10 12:19  161M
      mm39.trf.bed.gz              2020-09-10 12:19   19M
      mrna.fa.gz                   2021-08-02 15:11  262M
      mrna.fa.gz.md5               2021-08-02 15:11   45
      refMrna.fa.gz                2021-08-02 16:03   46M
      refMrna.fa.gz.md5            2021-08-02 16:03   48
      repeatMasker.versionInfo.txt 2020-07-29 11:48  1.1K
      upstream1000.fa.gz           2020-09-10 12:36  8.0M
      upstream1000.fa.gz.md5       2021-08-02 16:15   53
      upstream2000.fa.gz           2020-09-10 12:36   15M
      upstream2000.fa.gz.md5       2021-08-02 16:16   53
      upstream5000.fa.gz           2020-09-10 12:36   35M
      upstream5000.fa.gz.md5       2021-08-02 16:17   53
```

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

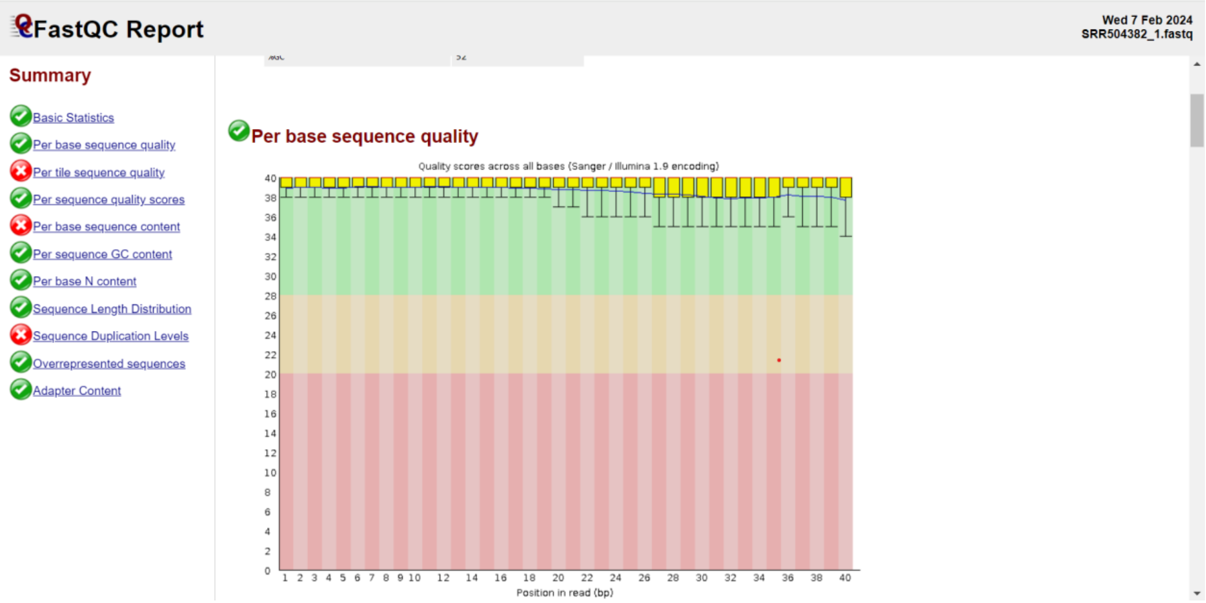


```
mkalpande@ManjushriK:~/sra_data$ mkdir altsplice
mkalpande@ManjushriK:~/sra_data$ ls
ERR12139009  SRR16235266.fastq  SRR3601860_1.fastq  fasterq.tmp.ManjushriK.62   outputq
ERR12139518  SRR23400676        altsplice           fasterq.tmp.ManjushriK.827
SRR16235266  SRR3601860         chr11human.fa       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) FASTQC output of both reads

1st read (good quality)



FastQC Report

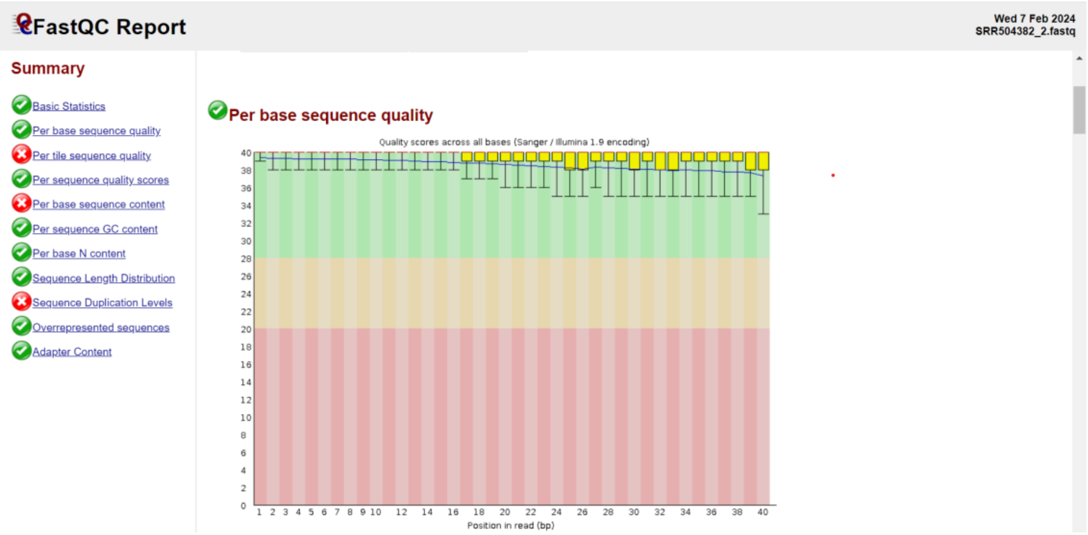
Wed 7 Feb 2024
SRR504382_1.fastq

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

Per base sequence quality

2nd read:: (ok but few place have low coverage)



FastQC Report

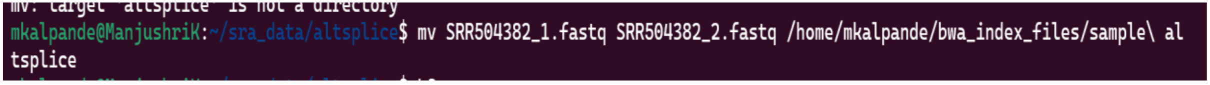
Wed 7 Feb 2024
SRR504382_2.fastq

Summary

- Basic Statistics
- Per base sequence quality
- Per tile sequence quality
- Per sequence quality scores
- Per base sequence content
- Per sequence GC content
- Per base N content
- Sequence Length Distribution
- Sequence Duplication Levels
- Overrepresented sequences
- Adapter Content

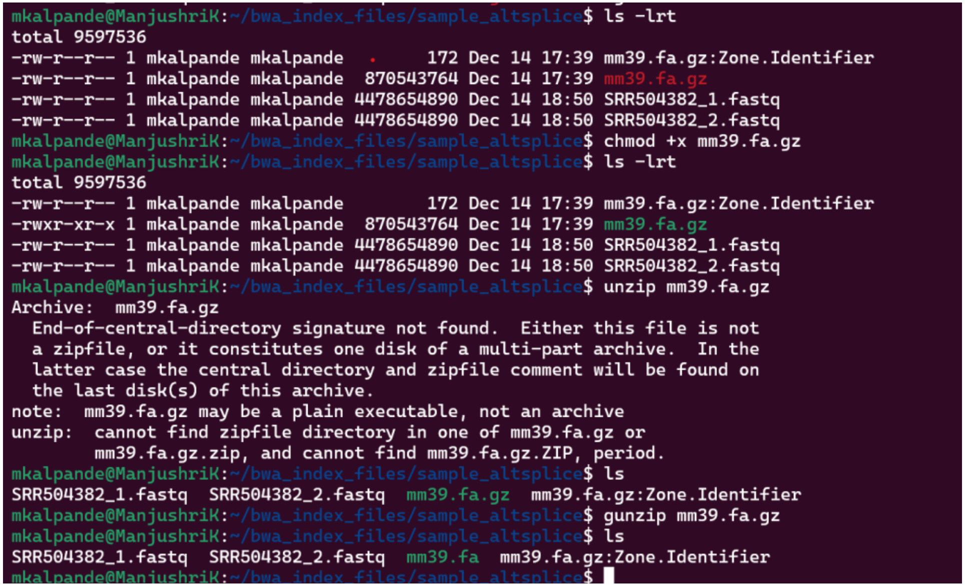
Per base sequence quality

4) Then we move the fastq files to place where we want to do BWA



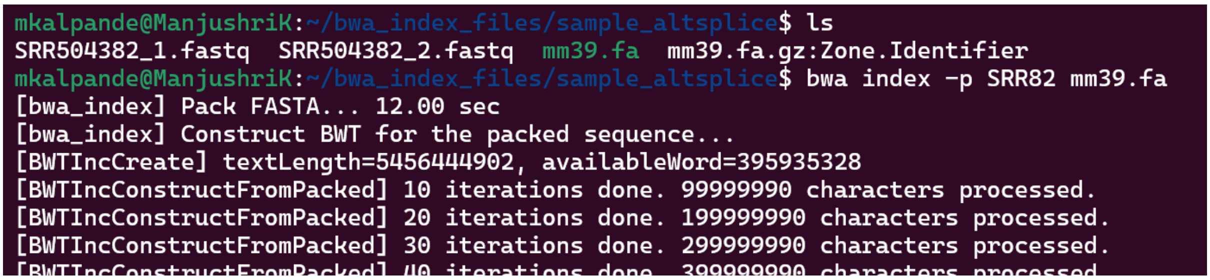
```
mv: target 'altsplice' is not a directory
mkalpande@ManjushriK:~/sra_data/altsplice$ mv SRR504382_1.fastq SRR504382_2.fastq /home/mkalpande/bwa_index_files/sample\ al
tsplice
```

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

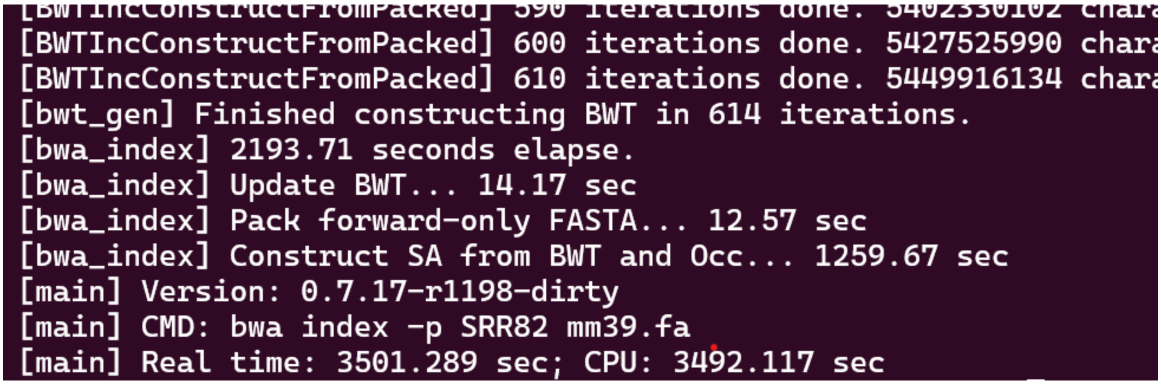


```
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls -lrt
total 9597536
-rw-r--r-- 1 mkalpande mkalpande        172 Dec 14 17:39 mm39.fa.gz:Zone.Identifier
-rw-r--r-- 1 mkalpande mkalpande  870543764 Dec 14 17:39 mm39.fa.gz
-rw-r--r-- 1 mkalpande mkalpande 4478654890 Dec 14 18:50 SRR504382_1.fastq
-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$
```

6) Now will do indexing using bwa index -p SRR82 file.fa
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.

7) 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

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

where- samtools view: Starts the conversion process.

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

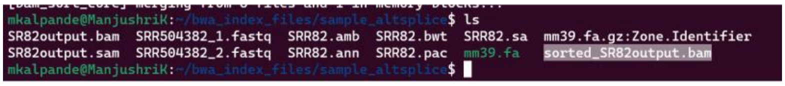
-b: Converts the output to BAM format.

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

9) Next we will do sorting of bam file 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.



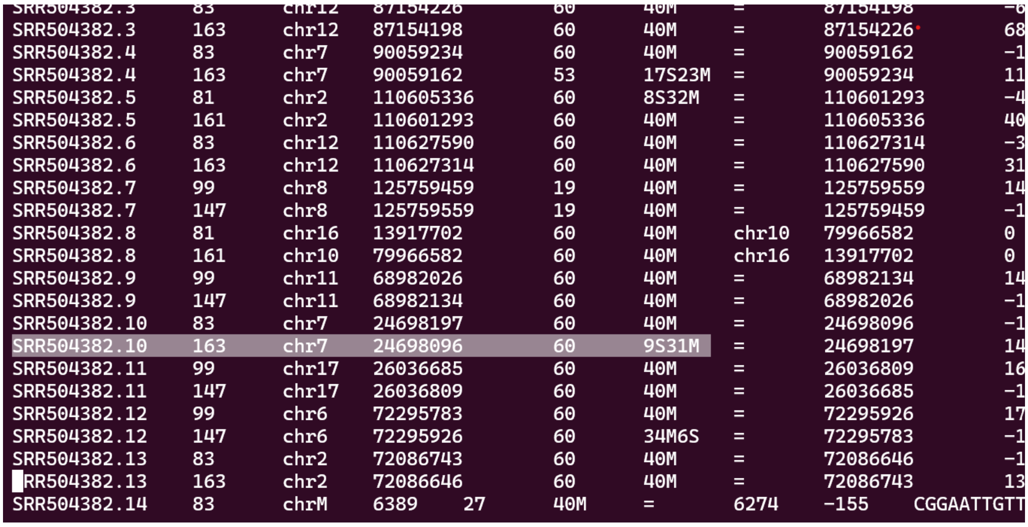
```
[bam_sort_core] merging from 0 files and 1 in-memory blocks...
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$ ls
SR82output.bam  SRR504382_1.fastq  SRR82.amb  SRR82.bwt  SRR82.sa  mm39.fa.gz:Zone.Identifier
SR82output.sam  SRR504382_2.fastq  SRR82.ann  SRR82.pac  mm39.fa   sorted_SR82output.bam
mkalpande@ManjushriK:~/bwa_index_files/sample_altsplice$
```

10) Next we will do indexing of sorted bam file

samtools index sorted_SR82output.bam

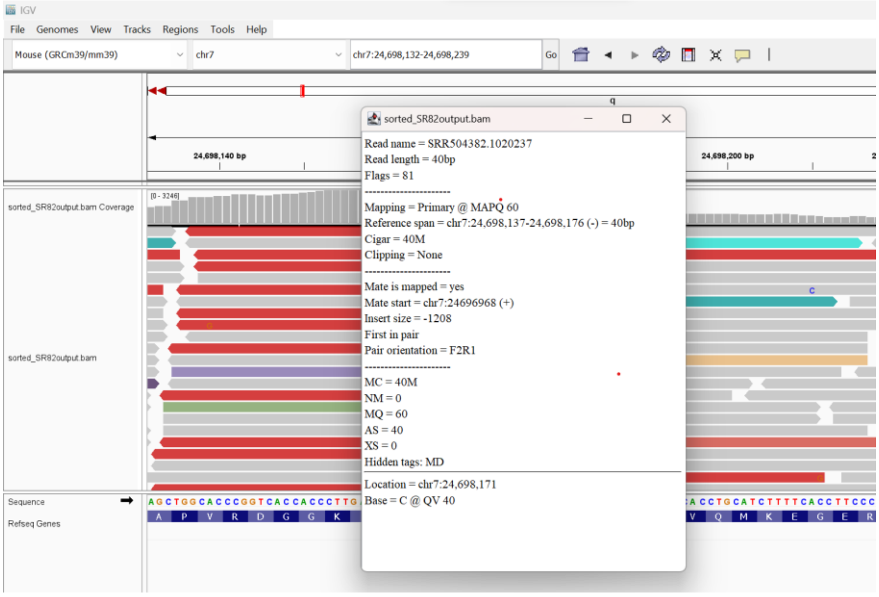
- **Samtools index**-The index file acts as a map, allowing tools to quickly find specific regions of the genome within the BAM file, improves the performance of many downstream analysis tasks, such as variant calling, read counting, and gene expression analysis.
- **.bai is index file**
- Next we will 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
RR504382.13     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



IGV
File Genomes View Tracks Regions Tools Help
Mouse (GRCm39/mm39)
chr7
chr7:24,698,132-24,698,239
Go
24,698,140 bp
24,698,200 bp
sorted_SR82output.bam Coverage
sorted_SR82output.bam
Sequence
Refseq Genes

sorted_SR82output.bam

Read name = SRR504382.1020237
Read length = 40bp
Flags = 81

Mapping = Primary @ MAPQ 60
Reference span = chr7:24,698,137-24,698,176 (-) = 40bp
Cigar = 40M
Clipping = None

Mate is mapped = yes
Mate start = chr7:24696968 (+)
Insert size = -1208
First in pair
Pair orientation = F2R1

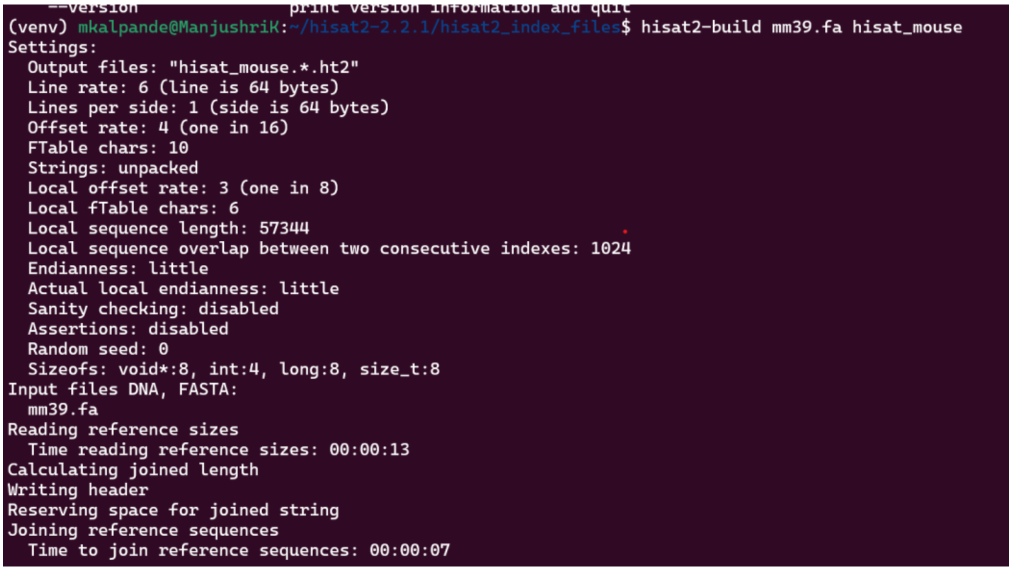
MC = 40M
NM = 0
MQ = 60
AS = 40
XS = 0
Hidden tags: MD

Location = chr7:24,698,171
Base = C @ QV 40

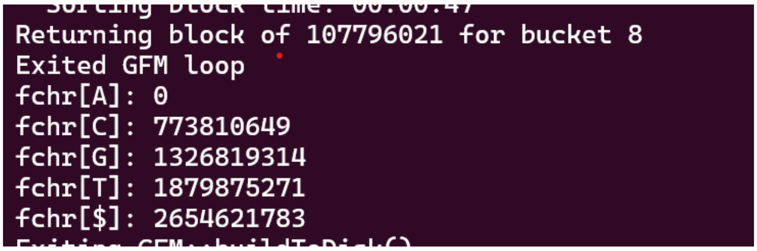
---- we perform alignment using HISAT2 as well----

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

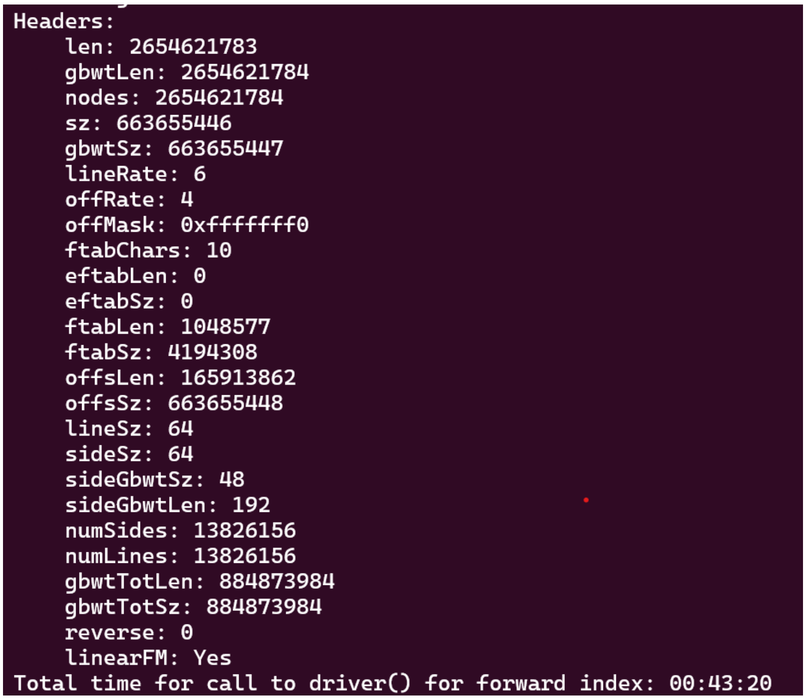
hisat2-build mm39.fa hisat_mouse



```
  --version                    print version information and quit
(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
```

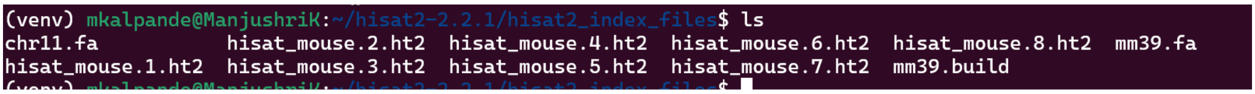


```
  Sorting block time: 00:00:47
Returning block of 107796021 for bucket 8
Exited GFM loop
fchr[A]: 0
fchr[C]: 773810649
fchr[G]: 1326819314
fchr[T]: 1879875271
fchr[$]: 2654621783
Exiting GFM::buildToDisk()
```



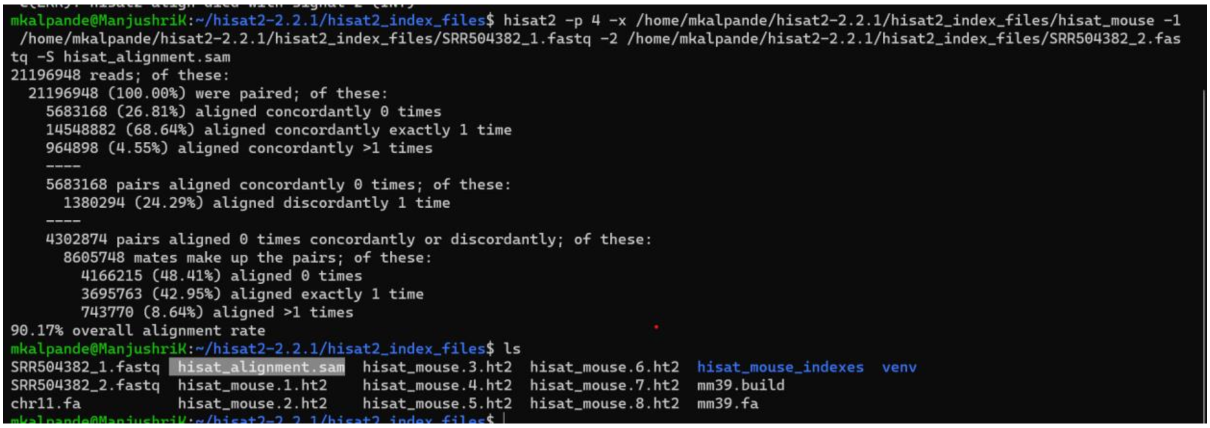
```
Headers:
    len: 2654621783
    gbwtLen: 2654621784
    nodes: 2654621784
    sz: 663655446
    gbwtSz: 663655447
    lineRate: 6
    offRate: 4
    offMask: 0xfffffff0
    ftabChars: 10
    eftabLen: 0
    eftabSz: 0
    ftabLen: 1048577
    ftabSz: 4194308
    offsLen: 165913862
    offsSz: 663655448
    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-



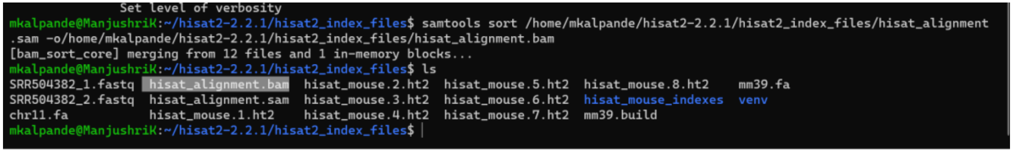
```
(venv) mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$ ls
chr11.fa          hisat_mouse.2.ht2  hisat_mouse.4.ht2  hisat_mouse.6.ht2  hisat_mouse.8.ht2  mm39.fa
hisat_mouse.1.ht2  hisat_mouse.3.ht2  hisat_mouse.5.ht2  hisat_mouse.7.ht2  mm39.build
(venv) mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$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

2) Running hisat2 so need to use the command-



```
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
mkalpande@ManjushriK:~/hisat2-2.2.1/hisat2_index_files$
```

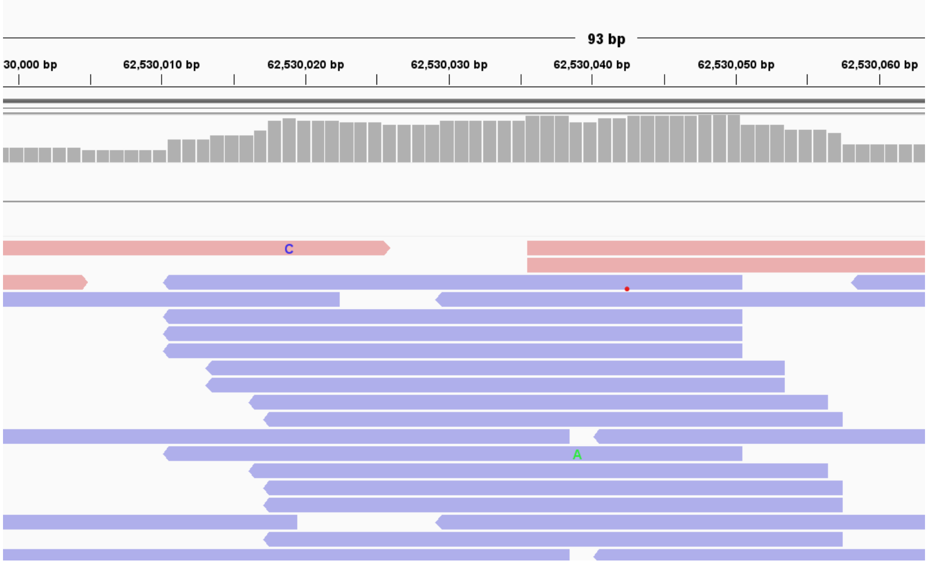
3) We converted to **BAM** file and then further sort the file .



```
               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$ |
```

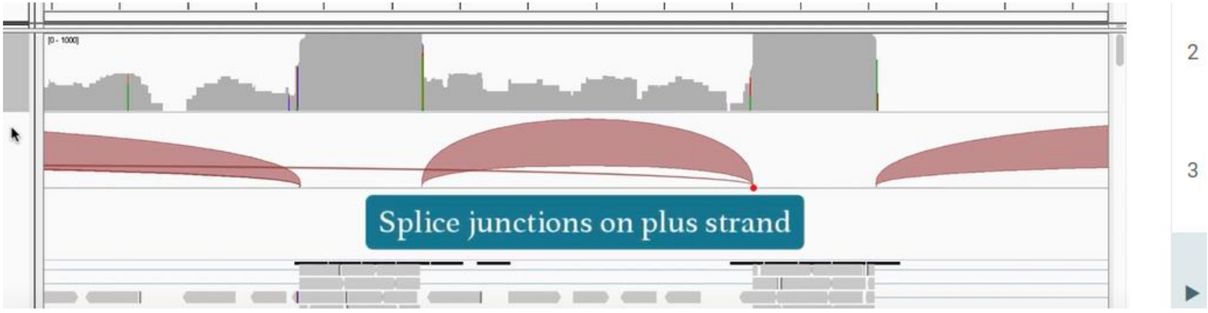
4) Next we will be using IGV to analyse the alignment

- Colour alignments by read strand

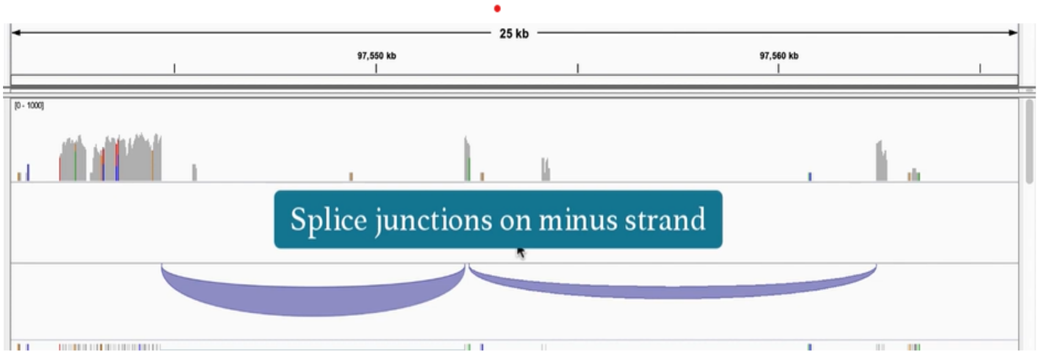


Red reads are forward strands & blue reads are reverse strand

- Brown colour uplifts are splice junctions

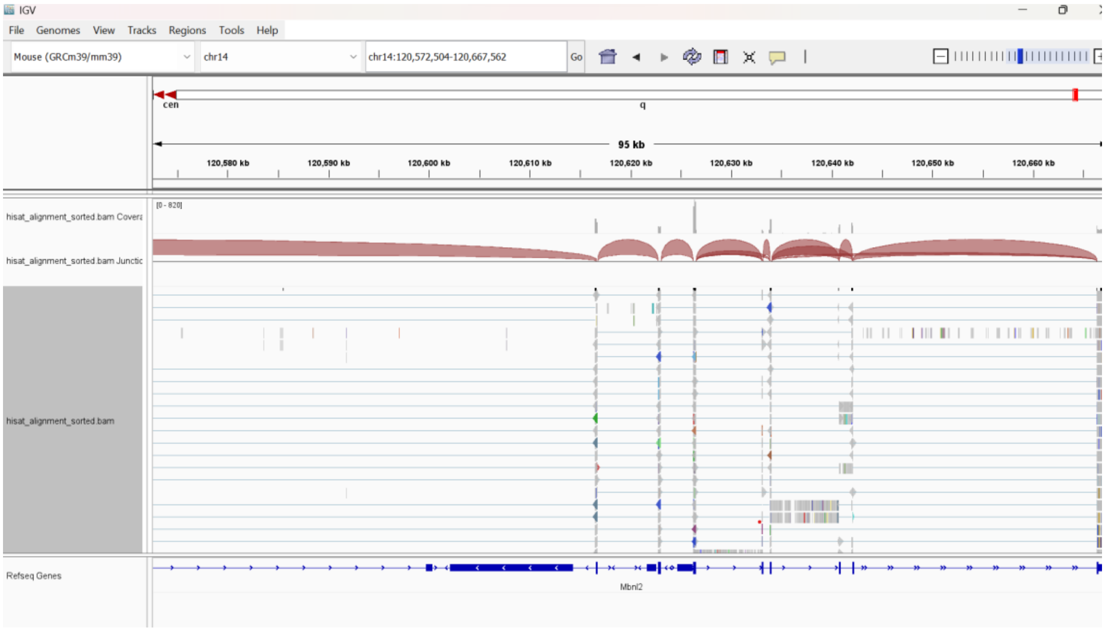


Splice junctions on plus strand

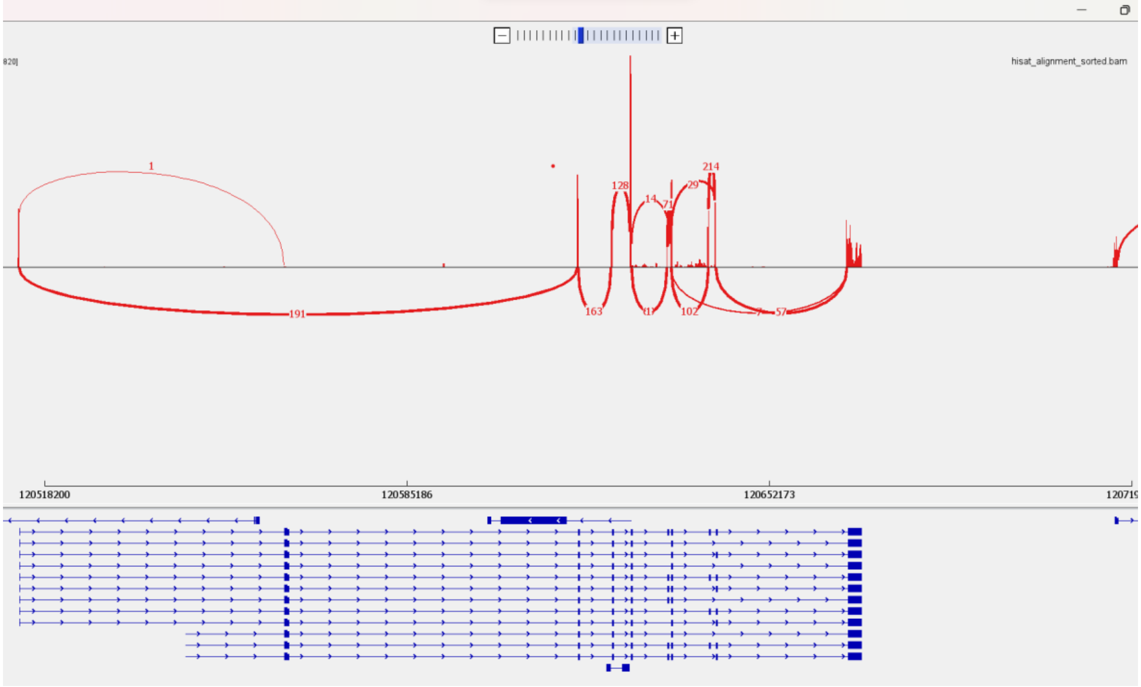


Splice junctions on minus strand

** SHASHIMI PLOT IS USED TO STUDY SPLICE JUNCTIONS

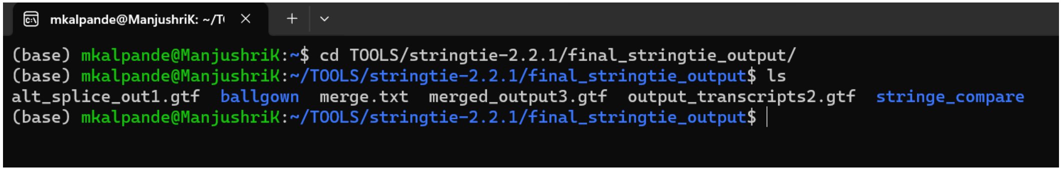

IGV
File Genomes View Tracks Regions Tools Help
Mouse (GRCm39/mm39)
chr14
chr14:120,572,504-120,667,562
Go
cen
q
95 kb
120,580 kb
120,590 kb
120,600 kb
120,610 kb
120,620 kb
120,630 kb
120,640 kb
120,650 kb
120,660 kb
[0 - 820]
hisat_alignment_sorted.bam Covera
hisat_alignment_sorted.bam Junctic
hisat_alignment_sorted.bam
Refseq Genes
Mbnl2

- 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.



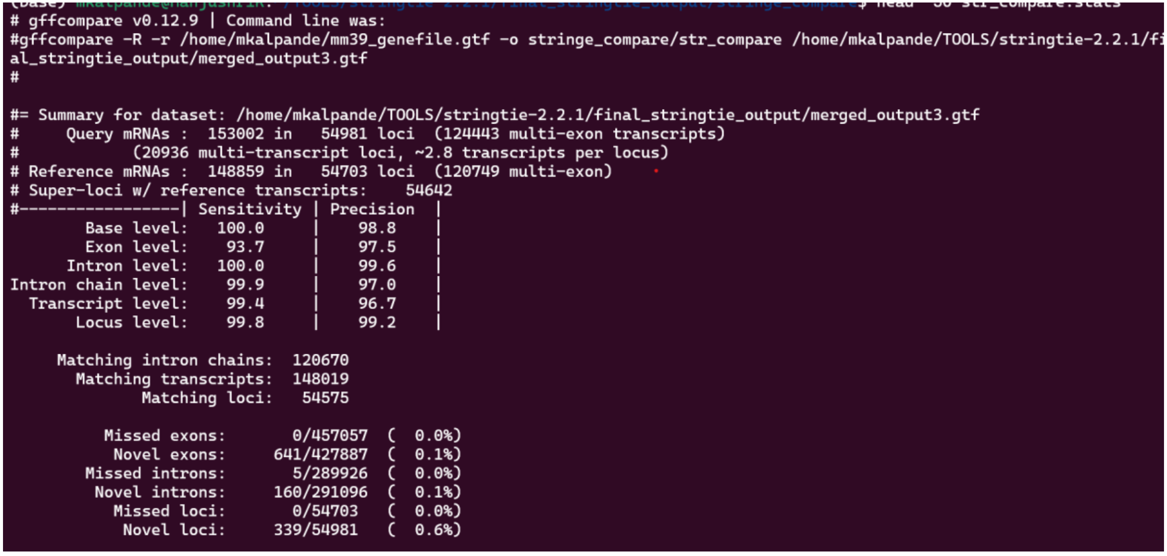
- The bottom blue lines from above image shows genes isoforms

Using Stringtie for transcript quantification:



```
(base) mkalpande@ManjushriK:~$ cd TOOLS/stringtie-2.2.1/final_stringtie_output/
(base) mkalpande@ManjushriK:~/TOOLS/stringtie-2.2.1/final_stringtie_output$ ls
alt_splice_out1.gtf  ballgown  merge.txt  merged_output3.gtf  output_transcripts2.gtf  stringe_compare
(base) mkalpande@ManjushriK:~/TOOLS/stringtie-2.2.1/final_stringtie_output$ 
```

- Using gffcompare generated a statistical summary file:



```
(base) mkalpande@hanjushrik: ~/TOOLS/stringtie-2.2.1/final_stringtie_output/stringe_compare$ head -30 str_compare.stats
# gffcompare v0.12.9 | Command line was:
#gffcompare -R -r /home/mkalpande/mm39_genefile.gtf -o stringe_compare/str_compare /home/mkalpande/TOOLS/stringtie-2.2.1/fi
al_stringtie_output/merged_output3.gtf
#

#= Summary for dataset: /home/mkalpande/TOOLS/stringtie-2.2.1/final_stringtie_output/merged_output3.gtf
#     Query mRNAs :  153002 in   54981 loci  (124443 multi-exon transcripts)
#            (20936 multi-transcript loci, ~2.8 transcripts per locus)
# Reference mRNAs :  148859 in   54703 loci  (120749 multi-exon)
# Super-loci w/ reference transcripts:    54642
#-----------------| Sensitivity | Precision  |
        Base level:   100.0     |    98.8    |
        Exon level:    93.7     |    97.5    |
      Intron level:   100.0     |    99.6    |
Intron chain level:    99.9     |    97.0    |
  Transcript level:    99.4     |    96.7    |
       Locus level:    99.8     |    99.2    |

     Matching intron chains:  120670
       Matching transcripts:  148019
              Matching loci:   54575

          Missed exons:       0/457057  (  0.0%)
           Novel exons:     641/427887  (  0.1%)
        Missed introns:       5/289926  (  0.0%)
         Novel introns:     160/291096  (  0.1%)
           Missed loci:       0/54703   (  0.0%)
            Novel loci:     339/54981   (  0.6%)
```