

Untitled

E1.1 Use the STAR aligner to map the RNA-seq reads to the reference genome. Compare the results with those from HISAT2, in terms of the numbers or percentages of reads uniquely aligned to the reference genome for each sequencing library. Note

STAR binaries for Linux and Mac can be directly downloaded from <https://github.com/alexdobin/STAR/releases>

Using Star:

Make an index:

```
mkdir star_genome_index STAR --runThreadN 4 --runMode genomeGenerate --genomeDir
star_genome_index
--genomeFastaFiles dmel-all-chromosome-r6.53.fasta
--sjdbGTFfile dmel-all-r6.53.gtf
```

Testis

```
ENCLB925FOQ: to generate sam file: STAR --runThreadN 4 --genomeDir /home/zheng/Documents/course/509B
--readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCL
/home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB925FOQ/E
--outFileNamePrefix ENCLB925FOQ
```

```
result: zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseq
STAR --runThreadN 4 --genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star
--readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCL
/home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB925FOQ/E
--outFileNamePrefix ENCLB925FOQ /usr/lib/rna-star/bin/STAR-avx2 --runThreadN 4 --
genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index
--readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCL
/home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB925FOQ/E
--outFileNamePrefix ENCLB925FOQ STAR version: 2.7.10a compiled: 2022-01-16T16:35:44+00:00
Oct 10 18:39:46 ..... started STAR run Oct 10 18:39:46 ..... loading genome Oct 10 18:39:48 .....
started mapping Oct 10 18:52:25 ..... finished mapping Oct 10 18:52:26 ..... finished successfully
zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/E
head ENCLB925FOQLog.final.out Started job on | Oct 10 18:39:46 Started mapping on | Oct
10 18:39:48 Finished on | Oct 10 18:52:26 Mapping speed, Million of reads per hour | 27.88
```

Number of input reads	5871282
Average input read length	200
UNIQUE READS:	
Uniquely mapped reads number	5333229
Uniquely mapped reads %	90.84%

ENCLB597ZOR: STAR --runThreadN 15 --genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB597ZOR/ENCLB597ZOR --readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB597ZOR/ENCLB597ZOR --outFileNamePrefix ENCLB597ZOR

result: zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB597ZOR/ENCLB597ZOR\$ STAR --runThreadN 15 --genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index --readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB597ZOR/ENCLB597ZOR --outFileNamePrefix ENCLB597ZOR /usr/lib/rna-star/bin/STAR-avx2 --runThreadN 15 --genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index --readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB597ZOR/ENCLB597ZOR --outFileNamePrefix ENCLB597ZOR STAR version: 2.7.10a compiled: 2022-01-16T16:35:44+00:00 Oct 10 23:22:23 started STAR run Oct 10 23:22:23 loading genome Oct 10 23:22:25 started mapping Oct 10 23:24:24 finished mapping Oct 10 23:24:24 finished successfully

zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/TestisRNAseqdata/ENCLB597ZOR\$ head ENCLB597ZORLog.final.out Started job on | Oct 10 23:22:23 Started mapping on | Oct 10 23:22:25 Finished on | Oct 10 23:24:24 Mapping speed, Million of reads per hour | 141.09

Number of input reads	4663928
Average input read length	200
UNIQUE READS:	
Uniquely mapped reads number	4401788
Uniquely mapped reads %	94.38%

Ovary : ENCLB117FKX: STAR --runThreadN 14 --genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/ENCLB117FKX --readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/ENCLB117FKX --outFileNamePrefix ENCLB117FKX

result: zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/ENCLB117FKX\$ STAR --runThreadN 14 --genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index --readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/ENCLB117FKX --outFileNamePrefix ENCLB117FKX /usr/lib/rna-star/bin/STAR-avx2 --runThreadN 14 --genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index --readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/ENCLB117FKX --outFileNamePrefix ENCLB117FKX

```

-outFileNamePrefix ENCLB117FKX /usr/lib/rna-star/bin/STAR-avx2 -runThreadN 14 -
genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index
-readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/
/home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/E
-outFileNamePrefix ENCLB117FKX STAR version: 2.7.10a compiled: 2022-01-16T16:35:44+00:00
Oct 10 23:43:21 ..... started STAR run Oct 10 23:43:21 ..... loading genome Oct 10 23:43:23 .....
started mapping Oct 10 23:46:33 ..... finished mapping Oct 10 23:46:33 ..... finished successfully
zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/I
head ENCLB117FKXLog.final.out Started job on | Oct 10 23:43:21 Started mapping on | Oct
10 23:43:23 Finished on | Oct 10 23:46:33 Mapping speed, Million of reads per hour | 158.05

```

Number of input reads	8341619
Average input read length	200
UNIQUE READS:	
Uniquely mapped reads number	7985350
Uniquely mapped reads %	95.73%

```

ENCLB117FKX: STAR -runThreadN 14 -genomeDir /home/zheng/Documents/course/509Bioinformation/Pro
-readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB117FKX/
/home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB129EAK/E
-outFileNamePrefix ENCLB129EAK

```

```

result: zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/I
STAR -runThreadN 14 -genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index
-readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB129EAK/
/home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB129EAK/E
-outFileNamePrefix ENCLB129EAK /usr/lib/rna-star/bin/STAR-avx2 -runThreadN 14 -
genomeDir /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/star_genome_index
-readFilesIn /home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB129EAK/
/home/zheng/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/ENCLB129EAK/E
-outFileNamePrefix ENCLB129EAK STAR version: 2.7.10a compiled: 2022-01-16T16:35:44+00:00
Oct 10 23:49:07 ..... started STAR run Oct 10 23:49:07 ..... loading genome Oct 10 23:49:09 .....
started mapping Oct 10 23:51:12 ..... finished mapping Oct 10 23:51:12 ..... finished successfully
zheng@zheng-Redmi-Book-14:~/Documents/course/509Bioinformation/Project/P1STAR/OvaryRNAseqdata/I
head ENCLB129EAKLog.final.out Started job on | Oct 10 23:49:07 Started mapping on | Oct
10 23:49:09 Finished on | Oct 10 23:51:12 Mapping speed, Million of reads per hour | 170.43

```

Number of input reads	5823083
Average input read length	200
UNIQUE READS:	
Uniquely mapped reads number	5596446
Uniquely mapped reads %	96.11%

E2.2 The STAR by defaults generates a tab-separated file (SJout.tab) that contains all unique splicing junctions and the number of reads that cross each splicing junction. Please summarize the file format using the STAR manual. Write R code to map the junctions to known genes if the start and end sites of a junction are entirely contained within a gene on the chromosome.

