



Analysis of RNA-seq data : answers to questions





Céline Keime
keime@igbmc.fr



Question 1

■ Number of uniquely mapped reads

```
Started job on | Feb 04 09:10:08
Started mapping on | Feb 04 09:14:43
Finished on | Feb 04 09:14:54
Mapping speed, Million of reads per hour | 327.27




Number of input reads | 1000000
Average input read length | 50
UNIQUE READS:
Uniquely mapped reads number | 852434
Uniquely mapped reads % | 85.24%
Average mapped length | 49.84
Number of splices: Total | 137459
Number of splices: Annotated (sjdb) | 136335
Number of splices: GT/AG | 136060
Number of splices: GC/AG | 1157
Number of splices: AT/AC | 108
Number of splices: Non-canonical | 134
Mismatch rate per base, % | 0.15%
Deletion rate per base | 0.01%
Deletion average length | 1.60
Insertion rate per base | 0.00%
Insertion average length | 1.29
MULTI-MAPPING READS:
Number of reads mapped to multiple loci | 133958
% of reads mapped to multiple loci | 13.40%
Number of reads mapped to too many loci | 4067
% of reads mapped to too many loci | 0.41%
UNMAPPED READS:
Number of reads unmapped: too many mismatches | 0
% of reads unmapped: too many mismatches | 0.00%
Number of reads unmapped: too short | 7302
% of reads unmapped: too short | 0.73%
Number of reads unmapped: other | 2239
% of reads unmapped: other | 0.22%
CHIMERIC READS:
Number of chimeric reads | 0
% of chimeric reads | 0.00%
```



















History    

search datasets  

RNA-seq data analysis

16 shown





7.23 GB   



- 8: RNA STAR on data 5 and data 4: mapped.bam   
- 7: RNA STAR on data 5 and data 4: splice junctions.bed   
- 6: RNA STAR on data 5 and data 4: log   
- 5: Homo_sapiens.GRC.h38.105.chr.gtf.gz   
- 4: siLuc2_1000000.fas.tq.gz   
- 3: FastQC on data 1: RawData   

Question 1




Category	RNA STAR on data 5 and data 4: mapped.bam
__no_feature	67657
__ambiguous	32425
__too_low_aQual	0
__not_aligned	13608
__alignment_not_unique	450475

- No feature reads
 - Number
 - 67657
 - Proportion :
 - $67657 * 100 / 852434 = 7.94$
- Ambiguous reads
 - Number
 - 32425
 - Proportion
 - $32425 * 100 / 852434 = 3.80$




History    

search datasets  

RNA-seq data analysis

16 shown   

7.23 GB

16: htseq-count on data 5 and data 8 (no feature)   





Question 1



- Proportion of reads among uniquely aligned reads
 - Assigned : $100 - 7.94 - 3.80 = 88.26$ %
 - No feature : 7.94 %
 - Ambiguous : 3.80 %

Question 1

■ Number of assigned reads




Geneid	RNA STAR on data 5 and data 4: mapped.bam
ENSG00000000003	31
ENSG00000000005	0
ENSG00000000419	95
ENSG00000000457	18
ENSG00000000460	55
ENSG00000000938	0
ENSG00000000971	3
ENSG00000001036	66
ENSG00000001084	50
ENSG00000001167	43
ENSG00000001460	6
ENSG00000001461	18
ENSG00000001497	72
ENSG00000001561	2
ENSG00000001617	2
ENSG00000001626	0
ENSG00000001629	53
ENSG00000001630	5
ENSG00000001631	5
ENSG00000002016	7
ENSG00000002079	0
ENSG00000002330	27
ENSG00000002549	68
ENSG00000002586	123
ENSG00000002587	1
ENSG00000002726	0
ENSG00000002745	0




History    




search datasets  

RNA-seq data analysis

16 shown









7.23 GB   

16: htseq-count on data 5 and data 8 (no feature)   

15: htseq-count on data 5 and data 8   

61,487 lines
format: **tabular**, database: **GRCh38**

100000 GFF lines processed.
200000 GFF lines processed.
300000 GFF lines processed.
400000 GFF lines processed.
500000 GFF lines processed.
600000 GFF lines processed.
700000 GFF lines processed.
800000 GFF lines processed.
900000 GFF lines processed.
10

1. Geneid 2. RNA STAR on data 5 and data 4: mapped.bam

ENSG00000000003 31

Question 1

- Number of assigned reads
 - Open the downloaded file with excel
 - Calculate the total number of reads in the second column

	A	B	C	D
61477	ENSG00000289634	0		
61478	ENSG00000289635	0		
61479	ENSG00000289636	0		
61480	ENSG00000289637	0		
61481	ENSG00000289638	0		
61482	ENSG00000289639	0		
61483	ENSG00000289640	0		
61484	ENSG00000289641	0		
61485	ENSG00000289642	0		
61486	ENSG00000289643	0		
61487	ENSG00000289644	0		
61488		752352		

→ Number of assigned reads = 752352

→ Proportion of assigned reads = $752352 * 100 / 852434 = 88.26$

Number of assigned reads

= number of uniquely aligned reads – number of no feature reads – number of ambiguous reads

= 852434 – 67657 - 32425 = 752352