BIOL 343
Applied Bioinformatics I
Post-alignment QC

Dr. Nic Wheeler

Alignment is the most important step in RNA-seq analysis

Counting (also important) and DEG ID relies on high-confidence mapping

- Recall the goal of our RNA-seq experiments…
  - Treatment vs Control
  - Mutant vs Wild type
  - ***Identify differentially expressed genes (DEGs)***
- DEGs will be identified using statistical tests comparing ***expression values*** of transcripts/genes
- Expression values will be calculated based on the number of reads that ***align/map*** to a given genomic locus
- ***Filtering*** out spurious or redundant alignments is an important QC step after mapping

Browsing RNA-seq alignments

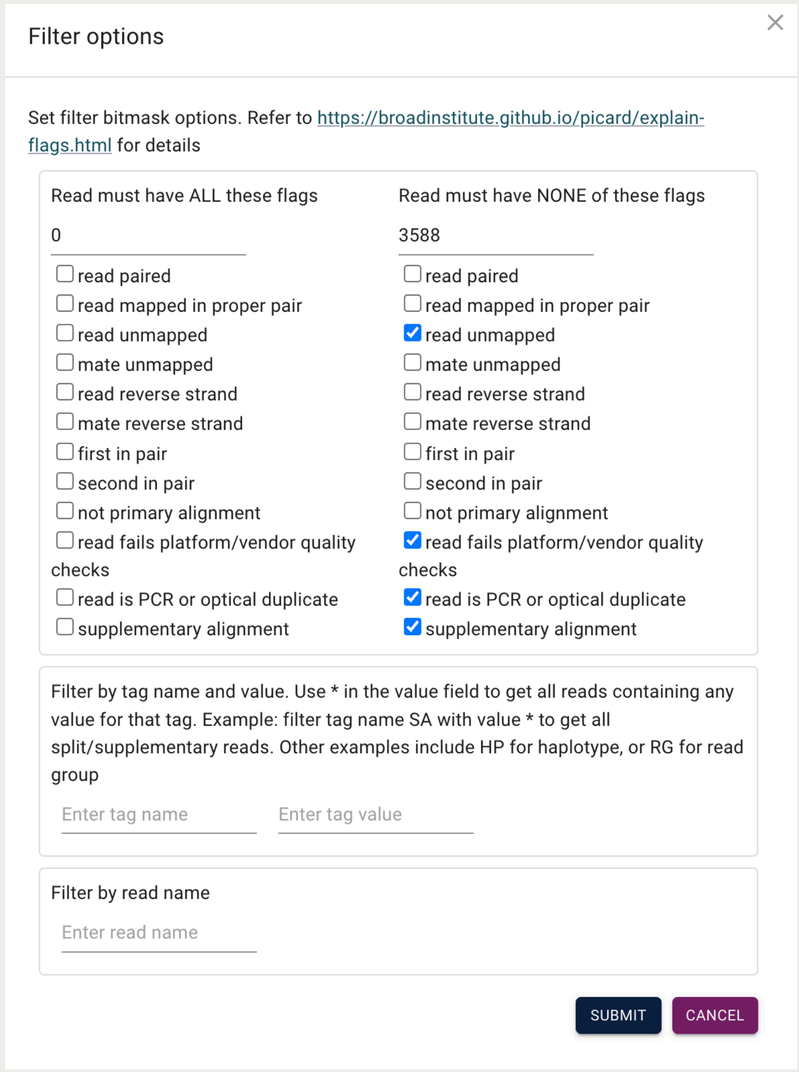
JBrowse2

1. Open JBrowse2
2. Download the genome files (.fa, .fai, .gtf)
3. Download the alignments (merged.bam and Aligned.sortedByCoord.out.bam) and indexes
4. Load the genome (.fa and .fai)
5. Add tracks:
    1. GTF
    2. Both alignments

Browsing RNA-seq alignments

JBrowse2

1. Navigate to SM_V10_3:12,442,794..12,489,015
2. Three dots on RNA-seq track…
    1. Pileup settings > Filter by…
    2. Pileup settings > Color by > Strand
    3. SNPCoverage settings > Draw arcs (deselect)



Sm V10

SM_V10_1 20M 40M 60M 80M SM_V10_2 20M 40M 60M 80M SM_V10_3 20M 40M SM_V10_4 20M 40M SM_V10_2 20M 40M SM_V10_6 20M SM_V10_5 20M SM_V10_7 SM_V1

SM_V10_3:12,463,563

SM_V10_3:12,463,210..12,473,290 10.1Kbp

SM_V10_3 12,464,000 12,465,000 12,466,000 12,467,000 12,468,000 12,469,000 12,470,000 12,471,000 12,472,000 12,473,000

Annotations

Smp_346880.1

Smp_121800.1

Smp_121800.2

STAR Alignment

1000
500
0

HISAT Alignment

1000
800
600
400
200
0

Browsing RNA-seq alignments

JBrowse2

1. Close STAR Alignment
2. HISAT Alignment three dots…
    1. Pileup settings > Group by…
        1. Tag name = RG (read group)
3. Close all the im alignments
4. Navigate to SM_V10_1:7098303..7098851
5. Adjust heights so that they’re evenly spaced and all you can see is the histogram

FILE ADD TOOLS HELP

New Session 10/10/2024, 11:23:31 AM

JBrowse

Sm V10

SM_V10_1 5M 10M 15M 20M 25M 30M 35M 40M 45M 50M 55M 60M 65M 70M 75M 80M 85M

SM_V10_1:7,097,513..7,099,708 2.2Kbp

SM_V10_1 7,097,600 7,097,800 7,098,000 7,098,200 7,098,400 7,098,600 7,098,800 7,099,000 7,099,200 7,099,400 7,099,600

Annotations

Smp_245390.1

HISAT Alignment RG:INT_ma1

HISAT Alignment RG:INT_ma2

HISAT Alignment RG:INT_ma3

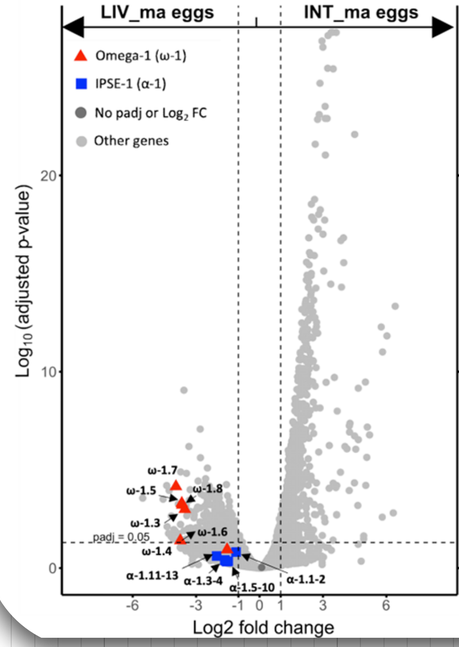
HISAT Alignment RG:LIV_ma1

HISAT Alignment RG:LIV_ma2

HISAT Alignment RG:LIV_ma3

Higher expression in LIV_ma than INT_ma…do we see that?

Immunomodulatory molecules



LIV_ma	INT_ma		
129	32	α-1.2	(Smp_245390)
17	3	α-1.11	(Smp_335430)
17	3	α-1.12	(Smp_335440)
17	3	α-1.13	(Smp_335450)
13	2	α-1.8	(Smp_334690)
13	2	α-1.1	(Smp_112110)
13	2	α-1.10	(Smp_335420)
12	2	α-1.7	(Smp_334660)
11	2	α-1.4	(Smp_334350)
8	2	α-1.9	(Smp_335410)
6	1	α-1.5	(Smp_334520)
6	1	α-1.6	(Smp_334590)
84	2	ω-1.7	(Smp_334170)
69	2	ω-1.5	(Smp_333930)
68	2	ω-1.8	(Smp_334240)
62	2	ω-1.3	(Smp_333220)
61	11	ω-1.9	(Smp_345790)
17	0	ω-1.4	(Smp_333870)

0 50 100 150
RPM

Genes of interest n = 22
Log2 fold change >|1| = 19