

# RNA sequence data analysis

## (Part 2: viewing and analyzing your results -- continuation of exercise 3).

### Exercise 10

1. If you are starting from PathogenPortal then you will need to download three files: Splice junctions, accepted hits, assembled transcripts.

Go to the following link:

<http://rnaseq.pathogenportal.org/u/omarharb/h/bab6>

Click on “Import history”

Accessible History | bab6 + Import history

Galaxy History ' bab6'

Dataset	Annotation
1: SRR016080.fastq	
2: Tophat for Illumina on data 1: insertions	
3: Tophat for Illumina on data 1: deletions	
4: Tophat for Illumina on data 1: splice junctions	
5: Tophat for Illumina on data 1: accepted hits	
6: Cufflinks on data 5: gene expression	
7: Cufflinks on data 5: transcript expression	
8: Cufflinks on data 5: assembled transcripts	

Next click on Project View

RNA-Seq Launch Pad Project View Shared Data Help User Using 1.2 Gb

✓ History "imported: bab6" has been imported.  
You can start using this history or return to the previous page.




In project view you should see the imported project with the workflow components.

Current Project History Options

bab6 852.1 Mb






- 8: Cufflinks on data 5: assembled transcripts
- 7: Cufflinks on data 5: transcript expression
- 6: Cufflinks on data 5: gene expression
- 5: Tophat for Illumina on data 1: accepted hits
- 4: Tophat for Illumina on data 1: splice junctions
- 3: Tophat for Illumina on data 1: deletions
- 2: Tophat for Illumina on data 1: insertions
- 1: SRR016080.fastq

Click on “4: Tophat for Illumina on data 1: splice junctions”. Notice the file format is Bed. This format can be used in gbrowse. Click on the download icon to download these data.

**4: Tophat for Illumina on data 1: splice junctions**   




splice junctions

1 region, 142 comments  
format: **bed**, database: Babesia\_bovis\_T2Bo  
Info: Tophat v1.4.1  
TopHat v1.4.1  
tophat -p 4 -a 8 -m 0 -i 70 -l 2000 -g 20 --library-type fr-unstranded --max-insertion-length 3 --max-deletion-length 3 --coverage-search --min-coverage-intron 50 --max-coverage-intron 20000 --no-closure-search --initial-read-






1. Chrom	2. Start	3. End	4	5	6	7	8
track name=junctions description="TopHat junctions"							
AAXT01000006	52884	53023	JUNC00000001	54	+	52884	5:
AAXT01000006	53196	53353	JUNC00000002	38	+	53196	5:
AAXT01000006	66931	67462	JUNC00000003	43	+	66931	6:
AAXT01000006	75822	76029	JUNC00000004	61	-	75822	7:
AAXT01000004	62023	62317	JUNC00000005	33	+	62023	6:

Click on “5: Tophat for Illumina on data 1: accepted\_hits”. Notice the file format is BAM. This format can be used in gbrowse. Click on the download icon to download these data.

**5: Tophat for Illumina on data 1:**   

accepted hits

62.4 Mb  
format: **bam**, database: Babesia\_bovis\_T2Bo  
Info: Tophat v1.4.1  
TopHat v1.4.1  
tophat -p 4 -a 8 -m 0 -i 70 -l 2000 -g 20 --library-type fr-unstranded --max-insertion-length 3 --max-deletion-length 3 --coverage-search --min-coverage-intron 50 --max-coverage-intron 20000 --no-closure-search --initial-read-

Binary bam alignments file

Click on “5: Tophat for Illumina on data 1: accepted\_hits”. Notice the file format is GTF. This format **cannot** be used in gbrowse. To use it you first have to convert it to an acceptable format. Click on the edit attributes icon to edit this file (pencil icon).

1. Seqname	2. Source	3. Feature	4. Start	5. End	6. Score
AAXT01000001	Cufflinks	transcript	5118	5601	1000
AAXT01000001	Cufflinks	exon	5118	5601	1000
AAXT01000001	Cufflinks	transcript	2156	3271	1000
AAXT01000001	Cufflinks	exon	2156	3271	1000
AAXT01000001	Cufflinks	transcript	6123	7568	1
AAXT01000001	Cufflinks	exon	6123	7568	1

Next click on the “Convert” button. This will convert the file to the BED format which can be loaded into gbrowse.

**Edit Attributes**

Name: Cufflinks on data 5: assembled transcript

Info: cufflinks v1.3.0  
cufflinks -q --no-update-check -l 2000 -F 0.100000

Annotation / Notes: None

Add an annotation or notes to a dataset; annotations are available when a history is viewed.

Database/Build: Click to Search or Select

Number of comment lines: 0

Save

Auto-detect

This will inspect the dataset and attempt to correct the above column values if they are not accurate.

**Convert to new format**

Convert GFF to BED

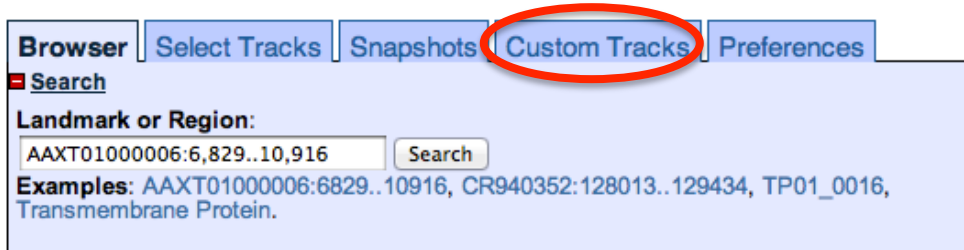
This will create a new dataset with the contents of this dataset converted to a new format.

Convert

You will notice a new step in the workflow appearing to the left. This should go quickly – when it is green, you are ready to download this data. Click on “10: Convert GFF to BED on data 8” and download this data as before.

Once you have download the files go to <http://piroplasmadb.org>

Go to the genome browser and click on Custom Tracks.



Next, click on “From File”, then choose the file and upload. You will have to do this three times to upload the three files you downloaded.



Once your files have been uploaded into gbrowse you can start exploring the data.

We will look at this the gene BBOV\_I1007720 together. Feel free to navigate to this gene. (hint: paste the ID in the Landmark box in gbrowse and click on Search).

One thing that would useful to do is change the height of the Y-axis on the BAM file track. Change this to the maximum value (100).

