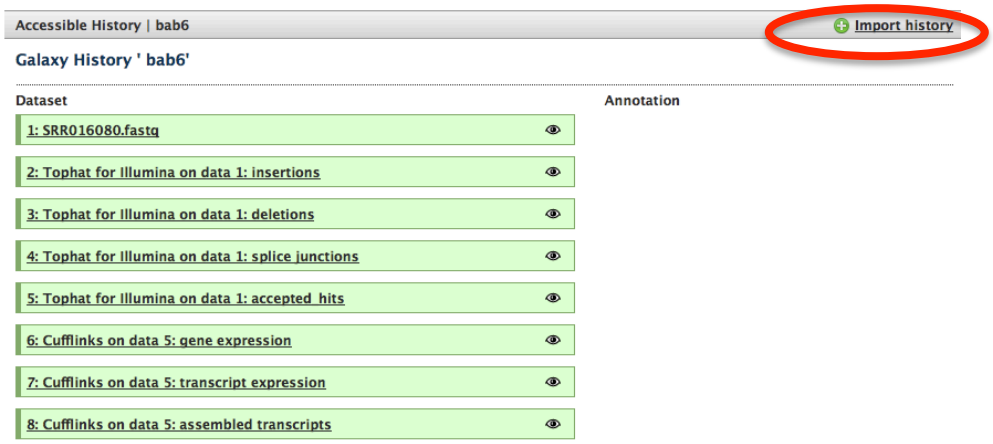
RNA sequence data analysis
(Part 2: viewing and analyzing your results -- continuation of exercise 3).
Exercise 10

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

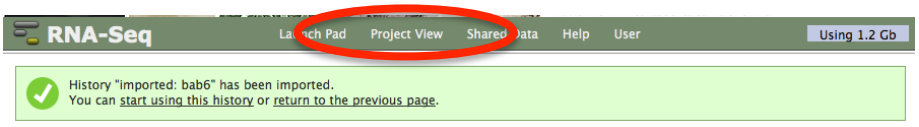
   Go to the following link:

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

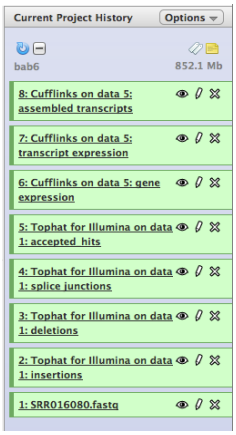
   Click on “Import history”



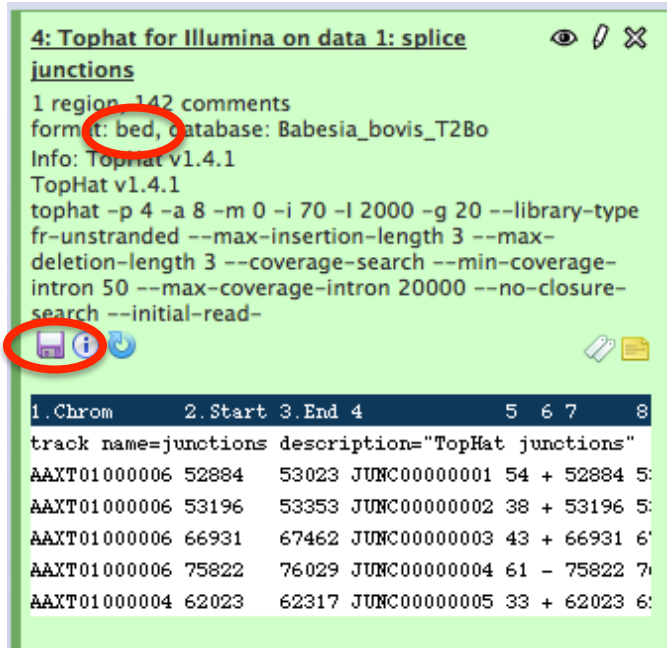
Next click on Project View



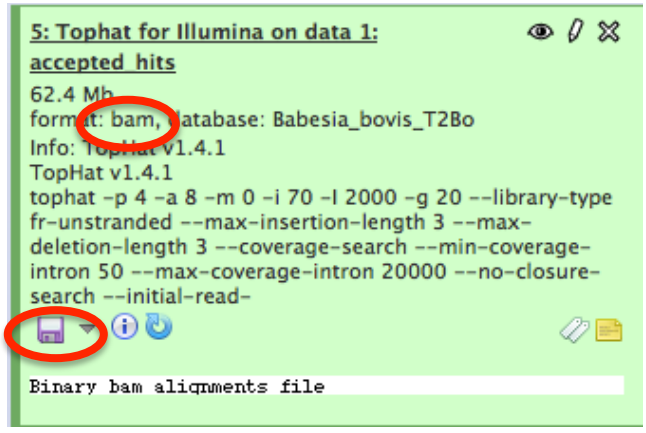
In project view you should see the imported project with the workflow components.



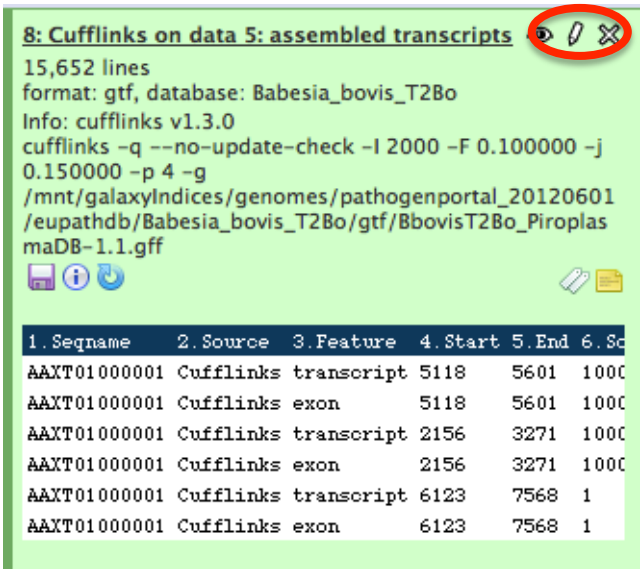
Click on “4: Tophat for Illumina on data 1: splice junctions”. Notice the file format is Bed. This format can be used in gbrowse. Click on the download icon to download these data.



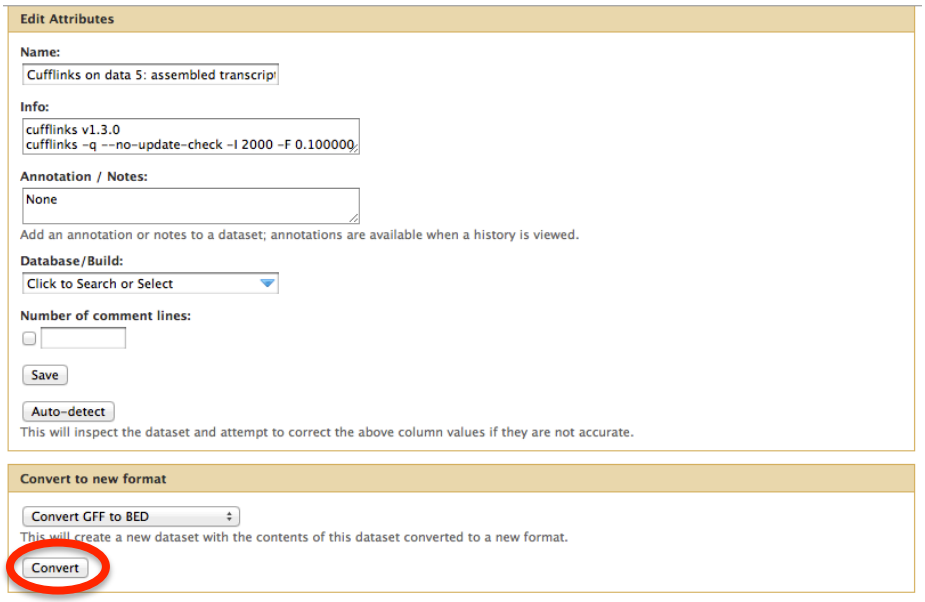
Click on “5: Tophat for Illumina on data 1: accepted_hits”. Notice the file format is BAM. This format can be used in gbrowse. Click on the download icon to download these data.



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).



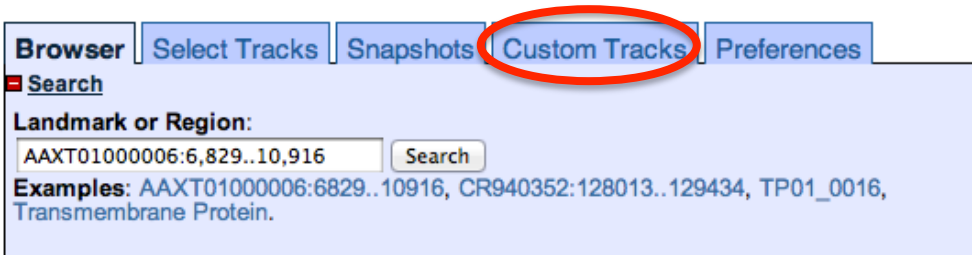
Next click on the “Convert” button. This will convert the file to the BED format which can be loaded into gbrowse.



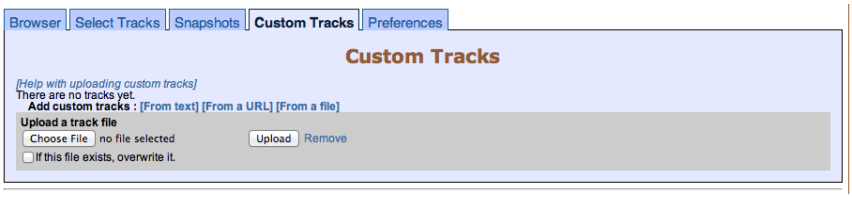
You will notice a new step in the workflow appearing to the left. This should go quickly – when it is green, you are ready to download this data. Click on “10: Convert GFF to BED on data 8” and download this data as before.

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

Go to the genome browser and click on Custom Tracks.



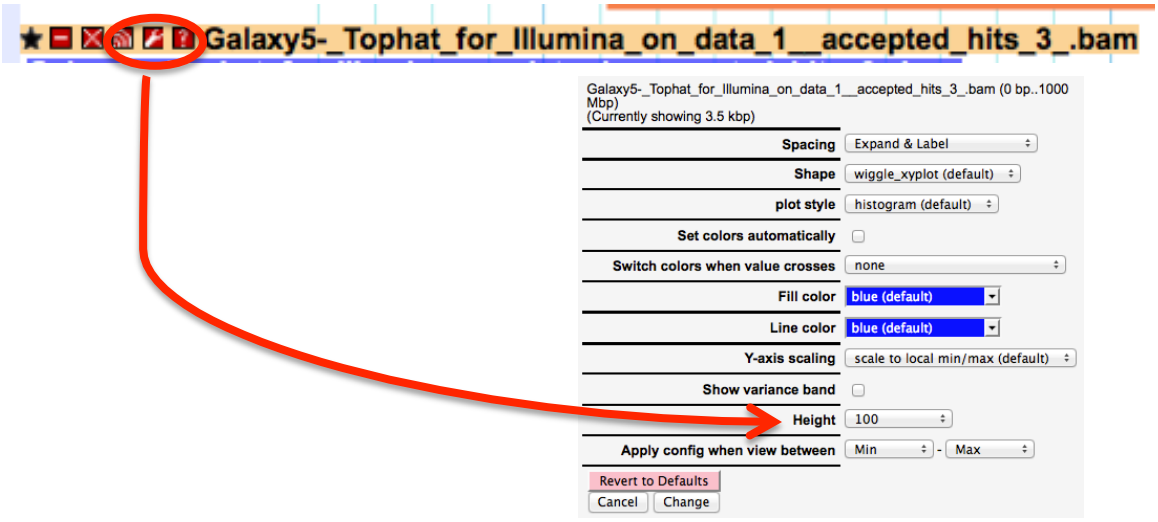
Next, click on “From File”, then choose the file and upload. You will have to do this three times to upload the three files you downloaded.

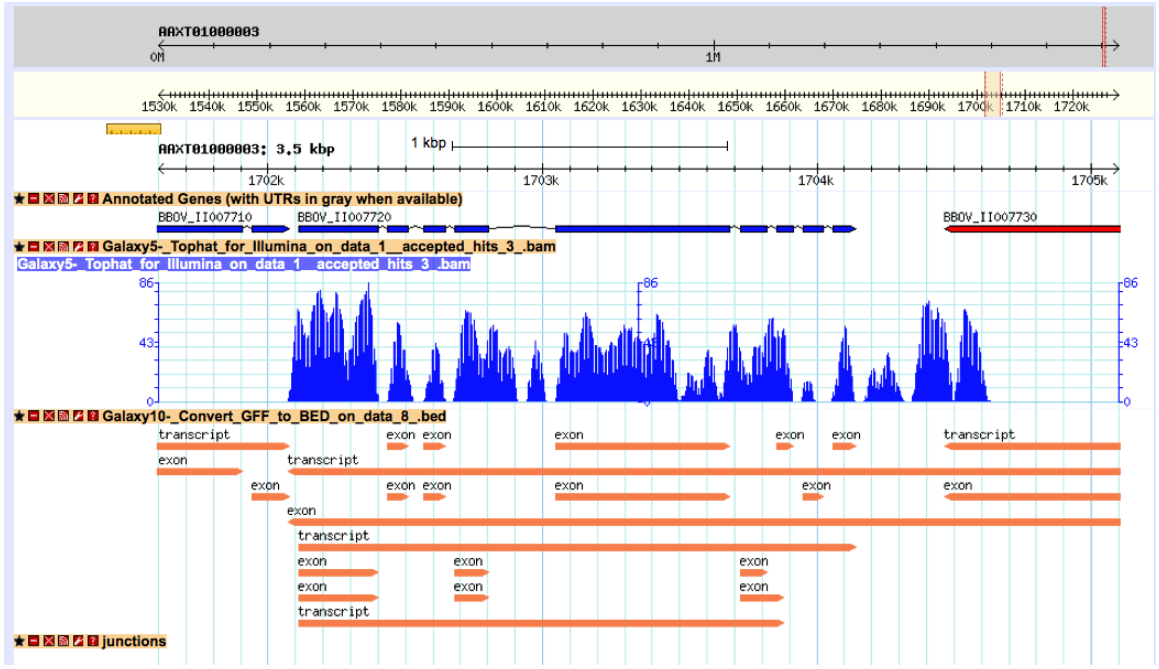


Once your files have been uploaded into gbrowse you can start exploring the data.

We will look at this the gene BBOV_II007720 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).




AAXT01000003
0M
1M
1530k 1540k 1550k 1560k 1570k 1580k 1590k 1600k 1610k 1620k 1630k 1640k 1650k 1660k 1670k 1680k 1690k 1700k 1710k 1720k
AAXT01000003: 3.5 kbp
1 kbp
1702k
1703k
1704k
1705k
Annotated Genes (with UTRs in gray when available)
BBOV_II007710
BBOV_II007720
BBOV_II007730
Galaxy5-_Tophat_for_Illumina_on_data_1__accepted_hits_3_.bam
Galaxy5-_Tophat_for_Illumina_on_data_1__accepted_hits_3_.bam
86
43
0
Galaxy10-_Convert_GFF_to_BED_on_data_8_.bed
transcript
exon
junctions