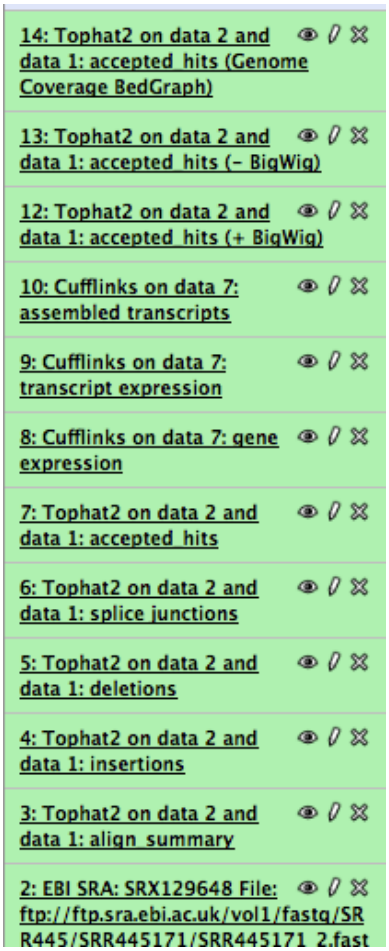
Interpreting RNAseq Mapping results
(Part 2: Loading data generated by the pathogen portal's RNAseq pipeline in the Genome Browser)
Exercise 7

For this exercise we will be using:
http://pathogenportal.org
http://giardiadb.org



1. Explore the results of the RNA-sequence pipeline. What files were generated? To view contents of any of the results, click on the eye icon () next to the file name.

!!! important note – do not click on the icon next to the file called "Tophat2 on data 1 and data 3: accepted_hits" – this file is huge and will not display but rather will download the contents to your computer.

a. TopHat in RNArocket generates five files:
   Align_summary: this includes a summary of how the alignment went (ie. the number of reads that were aligned).
b. *Insertions*: reported insertions.
c. *Deletions:* reported deletions.
d. *Splice junctions*: reported junctions. Each junction consists of two connected BED blocks, where each block is as long as the maximal overhang of any read spanning the junction. The score is the number of alignments spanning the junction.
e. *Accepted hits*: BAM file (binary alignment map). Note that many alignment programs will generate a file called a SAM file (sequence alignment map) which is a table including text of the alignment and mapping. However, for viewing results in a sequence browser like GBrowse, the file needs to be converted into the binary formatted (BAM) – you do not have to worry about this for this exercise.

Cufflinks generates three files:
gene expression, transcript expression and assembled transcripts. The gene

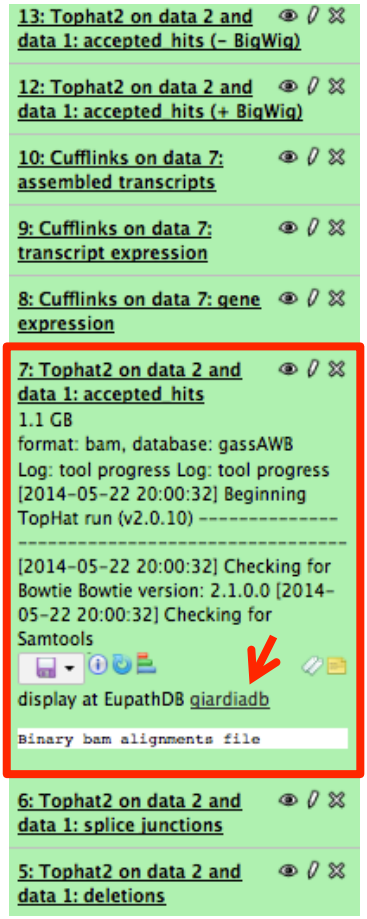
expression and transcript expression files for our purposes should be identical since EuPathDB genomes do not have separate genes and transcripts. These files include the FPKM values (Fragments Per Kilobase of transcript per Million mapped reads) for each gene in the genome analyzed – in this case *Giardia* assemblages.
Additional files include files of the format BigWig and BedGraph. You can read more about these file formats here:
http://genome.ucsc.edu/goldenPath/help/bigWig.html
In a nutshell, these are file formats created from large binary files like BAM files and makes it possible to load these data in a genome browser.

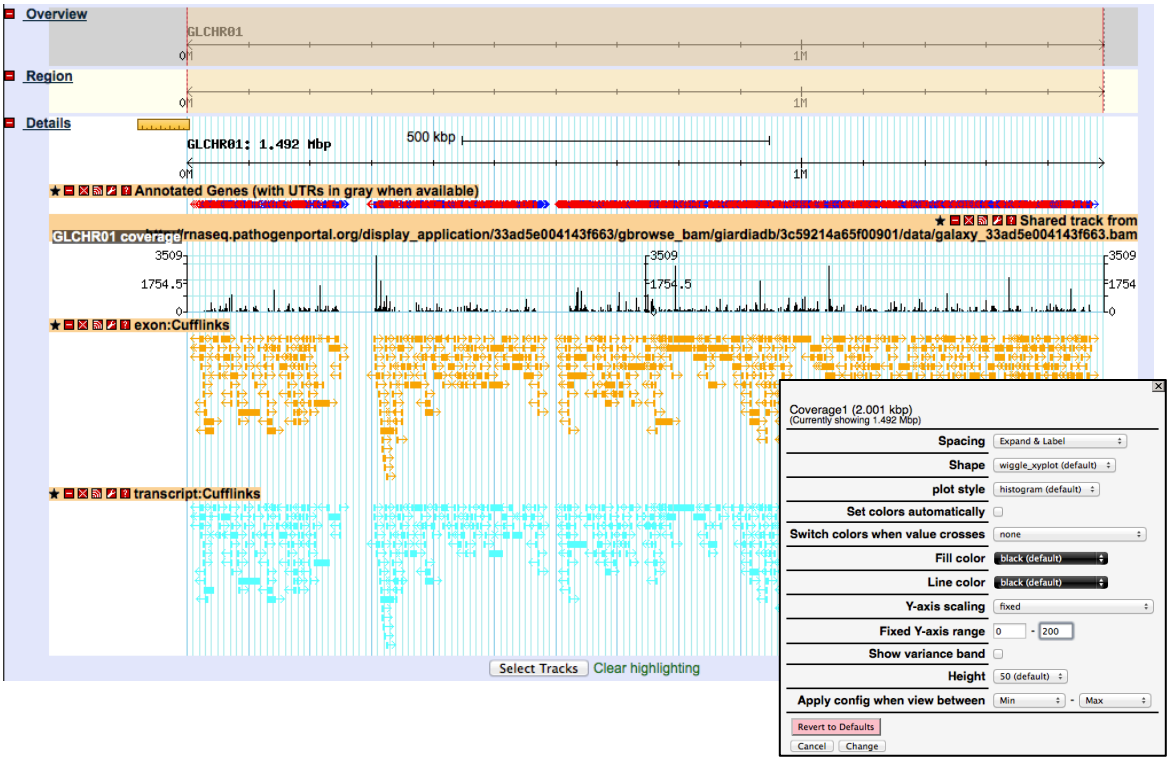
2. Load your BAM data (accepted hits) into GBrowse. Click on your "Tophat2 on data 2 and data 1: accepted_hits" in your project history panel. This will show you information about the file including a link to display data in GiardiaDB – click on the link.

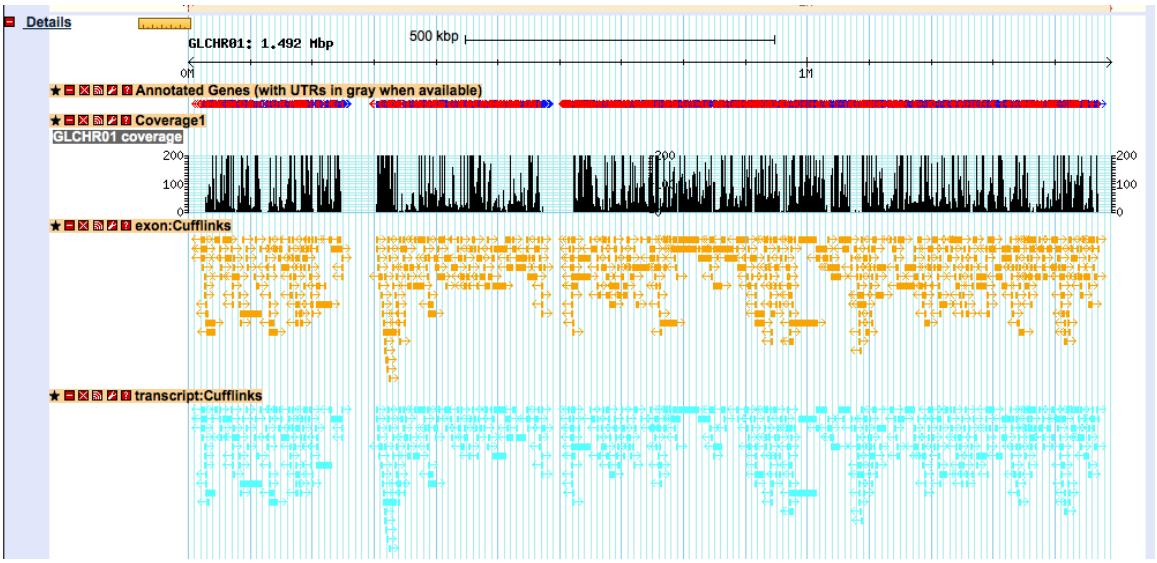
3. Load the assembled transcript data. Essentially use a similar procedure as above.

4. Wait a couple of minutes for GBrowse to load your data.

5. Once data has been loaded, you can configure the track display settings. For example, you can adjust the Y-axis scaling to a fixed axis.




Overview
GLCHR01
0M
1M
Region
Details
GLCHR01: 1.492 Mbp
500 kbp
Annotated Genes (with UTRs in gray when available)
Shared track from
GLCHR01 coverage
rnaseq.pathogenportal.org/display_application/33ad5e004143f663/gbrowse_bam/giardiadb/3c59214a65f00901/data/galaxy_33ad5e004143f663.bam
3509
1754.5
0
exon:Cufflinks
transcript:Cufflinks
Select Tracks
Clear highlighting
Coverage1 (2.001 kbp)
(Currently showing 1.492 Mbp)
Spacing
Expand & Label
Shape
wiggle_xyplot (default)
plot style
histogram (default)
Set colors automatically
Switch colors when value crosses
none
Fill color
black (default)
Line color
black (default)
Y-axis scaling
fixed
Fixed Y-axis range
0
200
Show variance band
Height
50 (default)
Apply config when view between
Min
Max
Revert to Defaults
Cancel
Change


Details
GLCHR01: 1.492 Mbp
500 kbp
0M
1M
Annotated Genes (with UTRs in gray when available)
Coverage1
GLCHR01 coverage
200
100
0
exon:Cufflinks
transcript:Cufflinks