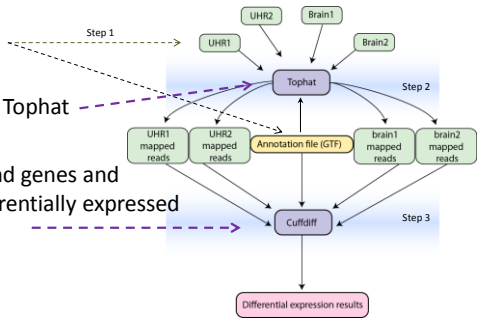
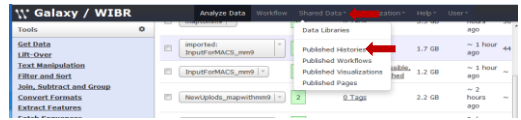
Hands-on 1
Quantification and assay for differential expression of reference annotation

1. Get input files
2. Map reads with Tophat
3. Quantify and find genes and transcripts differentially expressed with Cuffdiff

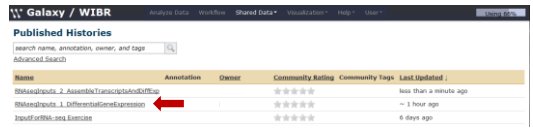


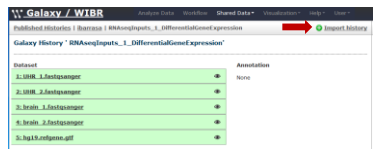
Step 1: Get input Files

- Click on the **Shared Data** Tab and select **Published Histories**



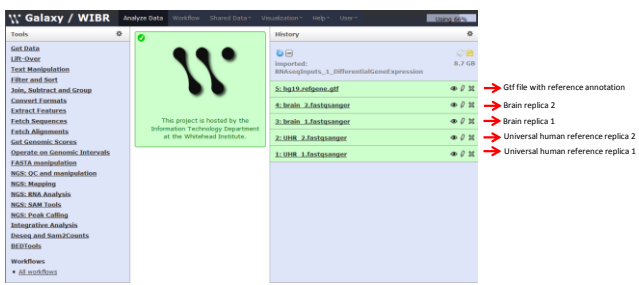
- Select *RNAseqInputs_1_DifferentialGeneExpression*





Click on **Import history** then click on **start using this history**

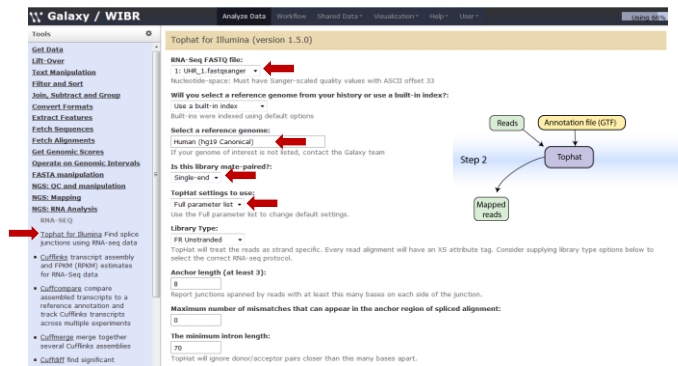
These are our input files



Reads are 35 nt long in fastqsanger format. You will have to **map each of the 4 fastqsanger input** files with tophat.
Reads were downloaded from GEO: http://www.ncbi.nlm.nih.gov/sra/SRX016368.

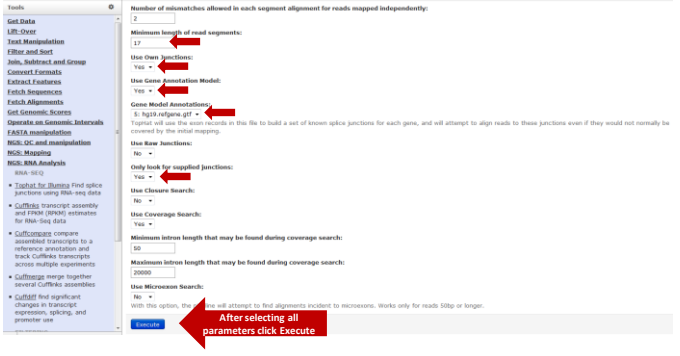
Note: when you do your own analysis, if your reads are not in fastqsanger format you have to convert them to fastqsanger using the **NGS: QC and manipulation -> FASTQ Groomer** tool.

Step 2: Map each of the four fastqsanger read with Tophat (v1.3.3)



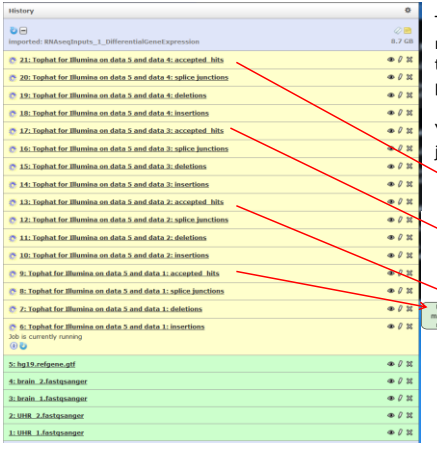
Red arrows show the parameters that you have to set, or settings I want you to notice. Leave all other settings as they are.
Check next slide for the rest of the parameters. This screenshot doesn't show the full interface.

Step 2_cont: Map each of the four fastqsanger read with Tophat (v1.3.3)



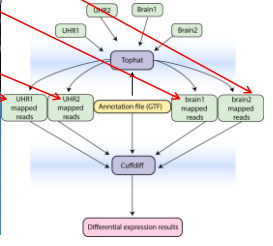
Hint: Once you have sent the first sample you can click on the rerun icon and change only the fastq input file to summit the other samples. All the other settings will be already selected.

Outputs from step 2 (Tophat)

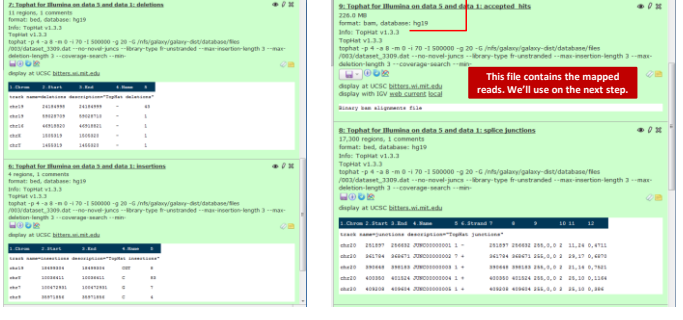


This is how your history looks now, after you have send four tophat jobs. Each tophat job produces 4 output files.

You don't have to wait for those jobs to finish to send **Cuffdiff**.



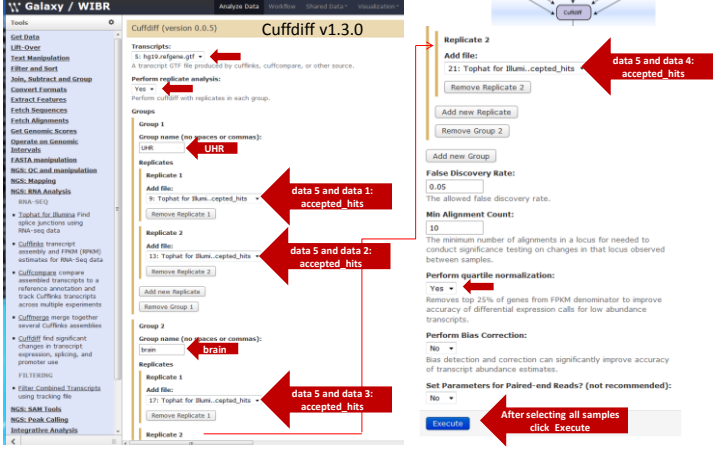
TopHat output example



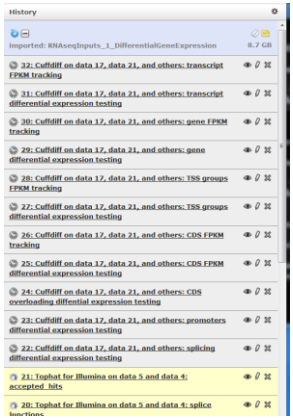
If you want to see the output from TopHat, import history *Hands-on 1 _ Output* from Published Histories.

Step 3: Run Cuffdiff to find genes differentially expressed

To get to this screen you have to click on **Add new Group** under Groups heading and also click on **Add new Replicate.**



Output from step 3 (Cuffdiff):



This is how your history looks now. Cuffdiff won't start until Tophat finishes.

If you want to see the output from Cuffdiff, import history *Hands-on 1 _ Output* from Published Histories. It contains these same jobs already run.

Cuffdiff output files

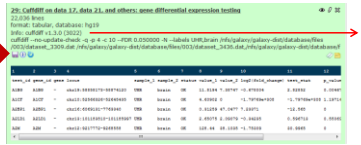
- FPKM tracking files for genes and transcripts
- Gene and transcript differential expression testing files
- The other files are empty because we didn't assemble transcripts for each sample

Sample of gene differential expression testing file

test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2
A1BG	A1BG	-	chr19:58858173-58874120	UHR	brain	OK	11.8194	7.38747
A1CF	A1CF	-	chr10:52566326-52645435	UHR	brain	OK	4.60902	0
A2BP1	A2BP1	-	chr16:6069131-7763340	UHR	brain	OK	0.31259	47.0477
A2LD1	A2LD1	-	chr13:101183810-101185997	UHR	brain	OK	2.65075	2.09079
A2M	A2M	-	chr12:9217772-9268558	UHR	brain	OK	128.44	38.1035
A2ML1	A2ML1	-	chr12:8975149-9029375	UHR	brain	NOTEST	0.0634409	0.550176

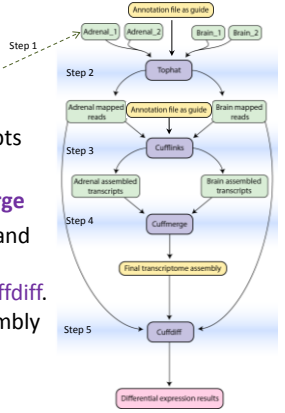
test_id	gene_id	log2(fold_change)	test_stat	p_value	q_value	significant
A1BG	A1BG	-0.678004	2.82852	0.00467631	0.0066993	yes
A1CF	A1CF	-1.79769e+308	-1.79769e+308	1.19714e-13	3.53612e-13	yes
A2BP1	A2BP1	7.23371	-12.568	0	0	yes
A2LD1	A2LD1	-0.34235	0.596713	0.550699	0.585973	no
A2M	A2M	-1.75309	30.9965	0	0	yes
A2ML1	A2ML1	3.11641	-2.3784	0.017388	1	no

Download data to your computer



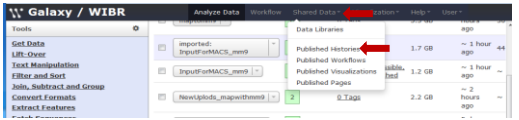
Hands-on 2: Assemble transcripts and assay for differential expression

1. Get input files: fastqsanger files
2. Map reads with Tophat
3. Assemble and quantify transcripts for each sample using **Cufflinks**
4. Merge assemblies with **Cuffmerge**
5. Quantify genes and transcripts and find genes and transcripts differentially expressed with Cuffdiff. Use the merged transcript assembly

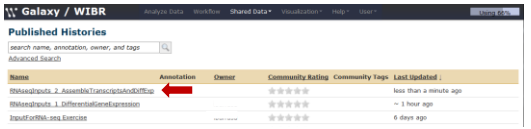


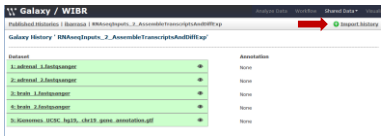
Step 1: Get input Files

- Click on the **Shared Data** Tab and select **Published Histories**



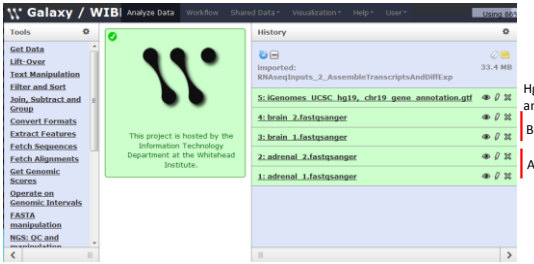
- Select *RNAseqInputs_2_AssembleTranscriptsAndDiffExp*





Click on **Import history**
then click on
start using this history

These are our input files



Hg19 Illumina gene annotation for chr19

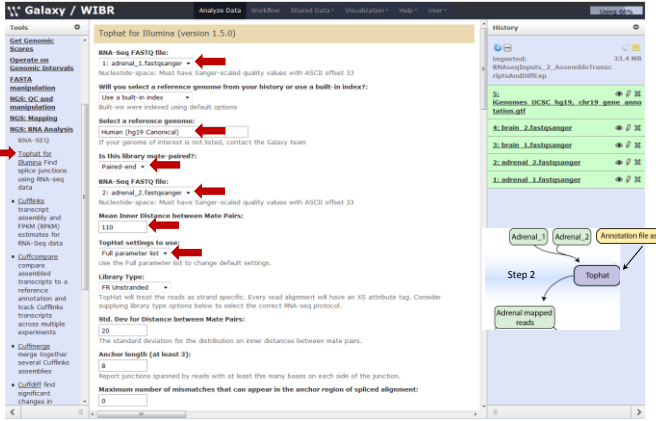
Brain paired end reads chr19 only

Adrenal paired end reads chr19 only

Reads are 50 nt long **paired end reads** in fastqsanger format. Only reads mapping to chr19 are included. You will have to map each of the 2 sets of reads.

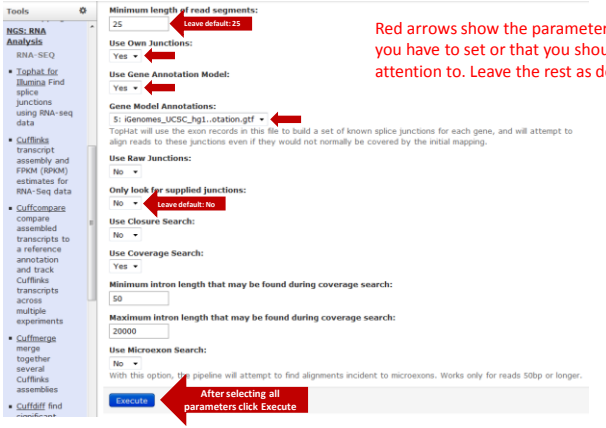
Annotation file for chr 19 is coming from Illumina igenomes
http://cufflinks.cbcb.umd.edu/igenomes.html
These input files were taken from the published RNA-seq Analysis Exercise
https://main.g2.bx.psu.edu/u/jeremy/p/galaxy-rna-seq-analysis-exercise

Step 2: Map reads with Tophat (v1.3.3)



Check next slide for the rest of the parameters. This screenshot doesn't show the full interface.

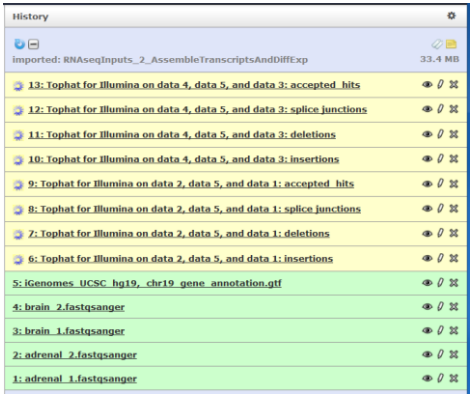
Step 2: Map reads with Tophat cont.



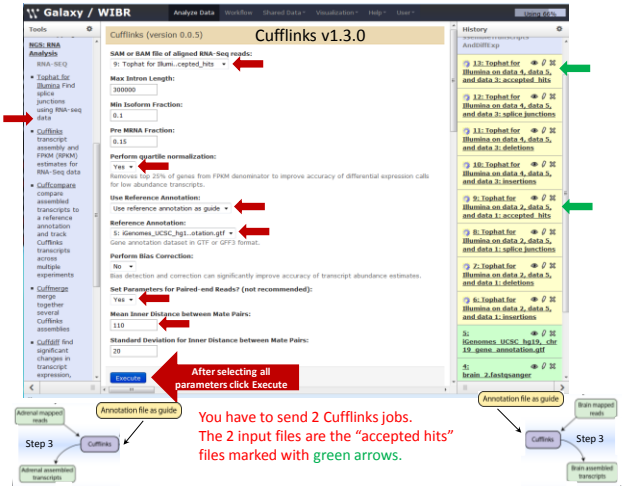
Red arrows show the parameters that you have to set or that you should pay attention to. Leave the rest as default.

Tophat output

This is how your history should look after sending the 2 tophat jobs



Step 3: Assemble and quantify transcripts for each sample with Cufflinks



You have to send 2 Cufflinks jobs.
The 2 input files are the "accepted hits" files marked with green arrows.

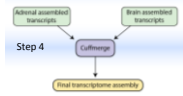
Cufflinks Output

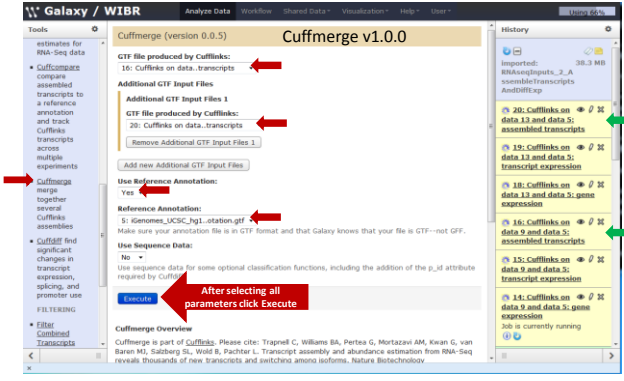
Assembled transcripts file. Will be the input file for Cuffmerge and Cuffcompare

Transcript expression: FPKM per transcript

Gene expression: FPKM per gene

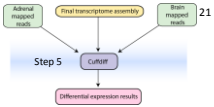
Step 4: Merge assemblies with Cuffmerge (v1.0.0)

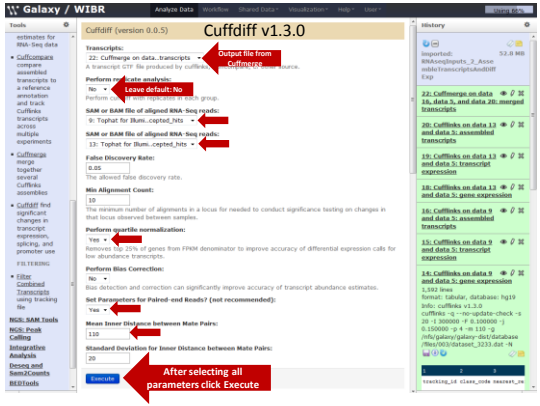




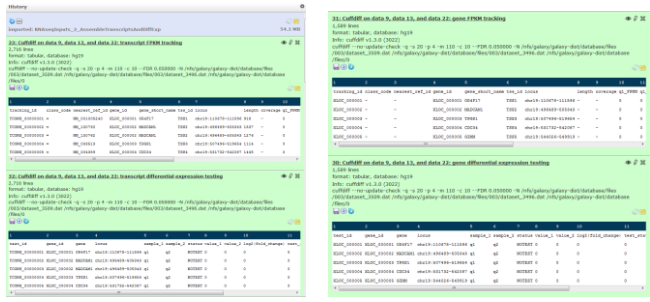
The input files are the assembled transcripts files generated by Cufflinks (green arrows).

Step 5: Find genes with Cuffdiff



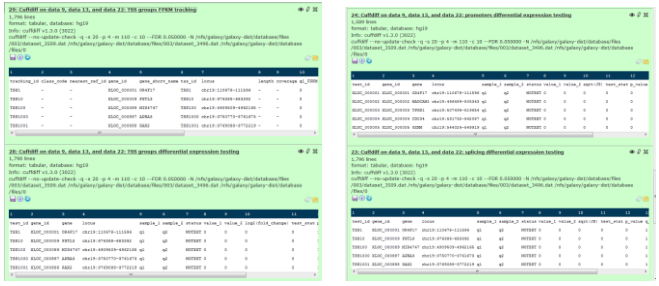


Step 5 output: Cuffdiff output Gene and transcript files



To access the outputs from Hands-on 2, import history *Hands-on 2 and 3 _ Output* from Published Histories.
Note that because we are working with a small sample size none of the differences found are significant.

Step 5 output: Cuffdiff ouput TSS, promoters and splicing files

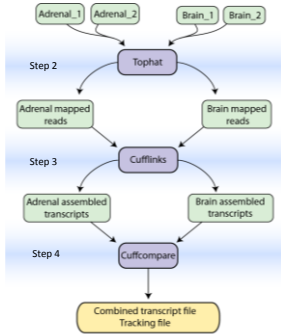


Note that because we are working with a small sample size none of the differences found are significant.

Hands-on 3
Assemble transcripts and compare assemblies

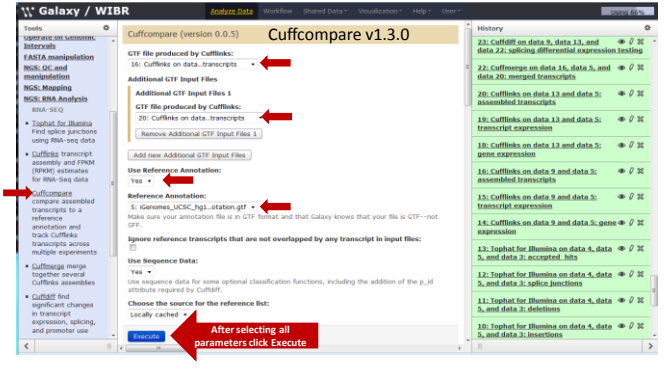
1. Get input files
2. Map reads with tophat
3. Assemble and quantify transcripts for each sample using Cufflinks
4. Compare assemblies with **Cuffcompare**

We have already done steps 1-3 in Hands-on 2. We only need to run Cuffcompare.



For simplicity the "annotation file" input has been removed
"Annotation file" is an optional input

Step 4: Compare assemblies with Cuffcompare



Cuffcompare output files

- Cuffcompare gives a gtf file of combined transcripts that you can use in other analysis
- Cuffcompare also gives tracking, refmap and tmap files that point to transcripts that fully or partially match between samples
- For more detailed description of output files see http://cufflinks.cbcb.umd.edu/manual.html#cuffcomp_output
- To access the outputs from Hands on 3, import history *Hands on 2 and 3 _ Output* from Published Histories