

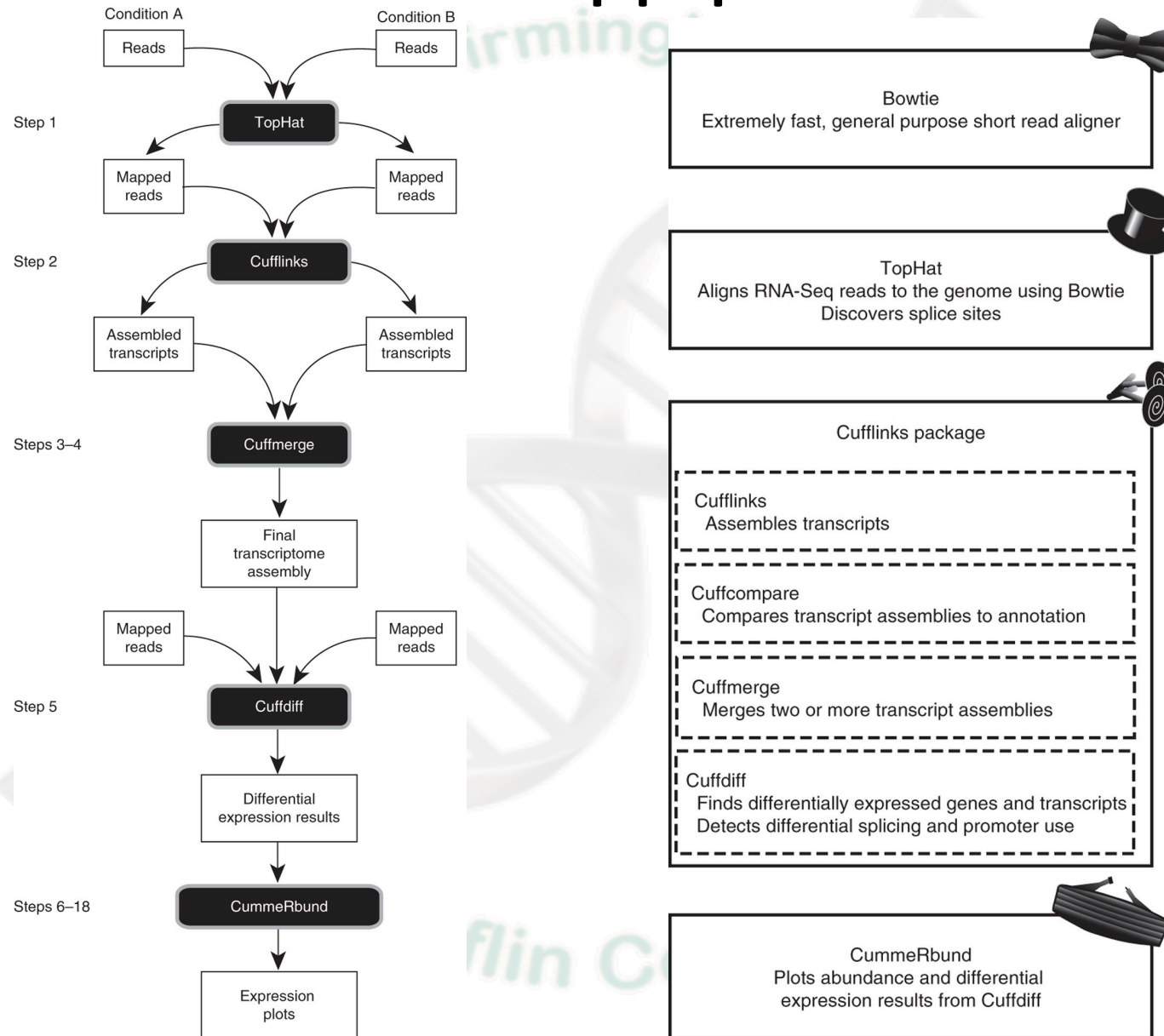


University of Alabama at Birmingham

Transcriptome Analysis

Heflin Center for Genomic Science

RNA-Seq pipeline



Upload/Import Data

Tools 1

- Get Data 2**
 - Upload File from your computer
 - UCSC Main table browser
 - UCSC Test table browser
 - UCSC Archaea table browser
 - BX main browser
 - Get Microbial Data
 - BioMart Central server
 - BioMart Test server
 - CBI Rice Mart rice mart
 - GrameneMart Central server
 - modENCODE fly server
 - Flymine server
 - Flymine test server
 - modENCODE modMine server
 - Ratmine server
 - YeastMine server
 - metabolicMine server
 - modENCODE worm server
 - WormBase server
 - Wormbase test server
 - EuPathDB server
 - EncodeDB at NHGRI
 - EpiGRAPH server
 - EpiGRAPH test server
 - HbVar Human Hemoglobin Variants and Thalassemias

Upload File (version 1.1.3)

File Format:

Auto-detect

3a

Which format? See help below

File:

No file chosen

3b-1

TIP: Due to browser limitations, uploading files larger than 2GB is guaranteed to fail. To upload large files, use the URL method (below) or FTP (if enabled by the site administrator).

URL/Text:

3b-2

Here you may specify a list of URLs (one per line) or paste the contents of a file.

Files uploaded via FTP:

File	Size	Date
<input type="checkbox"/> MF2_R1.fastqsanger	33.2 Mb	07/19/2012 07:26:42 AM
<input type="checkbox"/> MF2_R2.fastqsanger	33.2 Mb	07/19/2012 07:26:45 AM
<input type="checkbox"/> MF3_R1.fastqsanger	17.1 Mb	07/19/2012 07:26:47 AM
<input type="checkbox"/> MF3_R2.fastqsanger	17.1 Mb	07/19/2012 07:26:48 AM
<input type="checkbox"/> Treeshrew67 GeneScaffold_800_4487.gtf	17.3 Kb	07/19/2012 07:26:48 AM
<input type="checkbox"/> GeneScaffold_800_4487.fasta	251.2 Kb	07/19/2012 07:26:48 AM

3b-3

This Galaxy server allows you to upload files via FTP. To upload some files, log in to the FTP server at galaxy.uabgrid.uab.edu using your Galaxy credentials (email address and password).

Convert spaces to tabs:

Yes

Use this option if you are entering intervals by hand.

Genome:

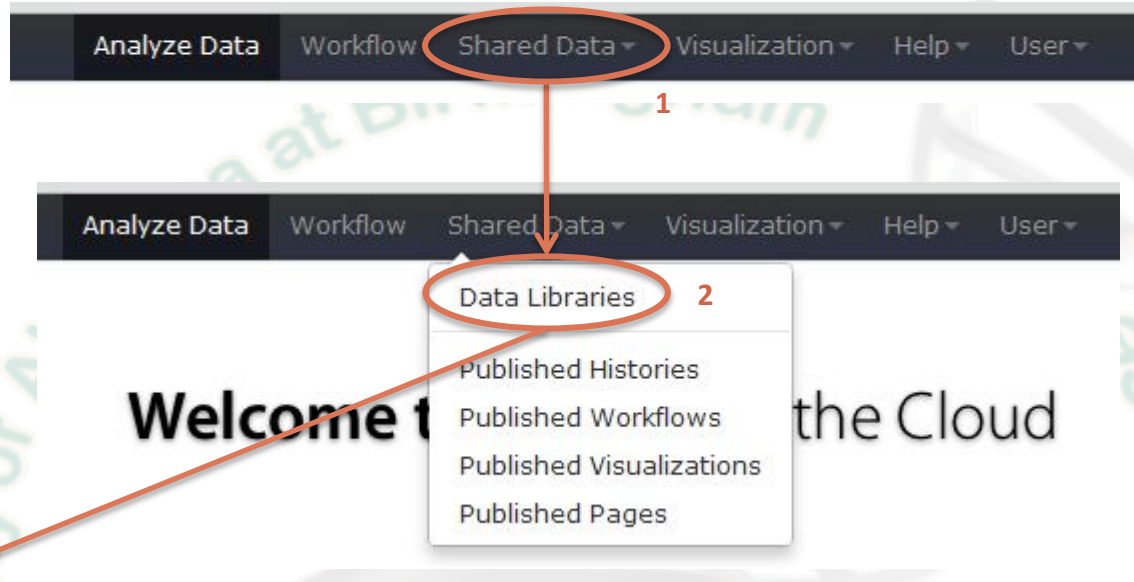
Click to Search or Select

3c

3d

1. Click "Get Data"
2. Click "Upload File"
3. Boxes to be aware of:
 - a) File Format
 - b) File to be uploaded:
 - 1) File from computer
 - 2) URL/text
 - 3) FTP
 - c) Genome
4. Click "Execute"

Shared Data



Data Library "GBS722-2014"

Raw data for various NGS demos that will be performed during class

<input type="checkbox"/> Name	Message	Data type	Date uploaded	File size
<input type="checkbox"/> RNA-Seq	Human data from a brain and adrenal sample			
<input type="checkbox"/> Adrenal_1	None	fastqsanger	2014-02-14	7.8 MB
<input type="checkbox"/> Adrenal_2	None	fastqsanger	2014-02-14	7.8 MB
<input type="checkbox"/> Brain_1	None	fastqsanger	2014-02-14	5.9 MB
<input type="checkbox"/> Brain_2	None	fastqsanger	2014-02-14	5.9 MB

For selected datasets:

1. Click on "Shared Data" (located on top toolbar)
2. Drop down box appears; click on "Data Libraries"
3. Will see this Data Library. Click on it to expand (as shown)

Import Shared Data to Current History

Data Library "GBS722-2014"

Raw data for various NGS demos that will be performed during class

<input type="checkbox"/>	Name	Message	Data type	Date uploaded	File size
<input checked="" type="checkbox"/>	RNA-Seq	Human data from a brain and adrenal sample			
1	<input checked="" type="checkbox"/> Adrenal_1	None	fastqsanger	2014-02-14	7.8 MB
	<input checked="" type="checkbox"/> Adrenal_2	None	fastqsanger	2014-02-14	7.8 MB
	<input checked="" type="checkbox"/> Brain_1	None	fastqsanger	2014-02-14	5.9 MB
	<input checked="" type="checkbox"/> Brain_2	None	fastqsanger	2014-02-14	5.9 MB

For selected datasets: 2

1. Check boxes of files you want to import
2. Choose "Import to current history" and then click "Go"

Quality Control of raw fastq reads

Tools

- NGS: QC and manipulation **1**
- FASTQC: FASTQ/SAM/BAM
- Fastqc: Fastqc QC using FastQC from Babraham **2**
- ILLUMINA FASTQ
- FASTQ Groomer convert between various FASTQ quality formats
- FASTQ splitter on joined paired end reads
- FASTQ joiner on paired end reads
- FASTQ Summary Statistics by column
- ROCHE-454 DATA
- Build base quality distribution
- Select high quality segments
- Combine FASTA and QUAL into FASTQ

FastQC:Read QC (version 0.51)

3a Short read data from your current history:
4: Brain_2

Title for the output file - to remind you what the job was for:
FastQC

Letters and numbers only please - other characters will be removed

Contaminant list:
Selection is Optional
tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Execute

FastQC:Read QC (version 0.51)

3b Short read data from your current history:
1: Adrenal_1 *

Title for the output file - to remind you what the job was for:
Adrenal_1 FastQC *

Letters and numbers only please - other characters will be removed

Contaminant list:
Selection is Optional
tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Execute **4**

1. Click on "NGS: QC and manipulation"
2. Click on "Fastqc: Fastqc QC"
3. Select options:
 - a) This is what the window looks like when first opened
 - b) Choose fastq file and give it a useful name
4. Click "Execute"
5. Do the exact same thing for the other 3 fastq files

FastQC Output Report

Adrenal_1 FastQC Report
FastQC Report
Sun 9 Feb 2014
Adrenal_1

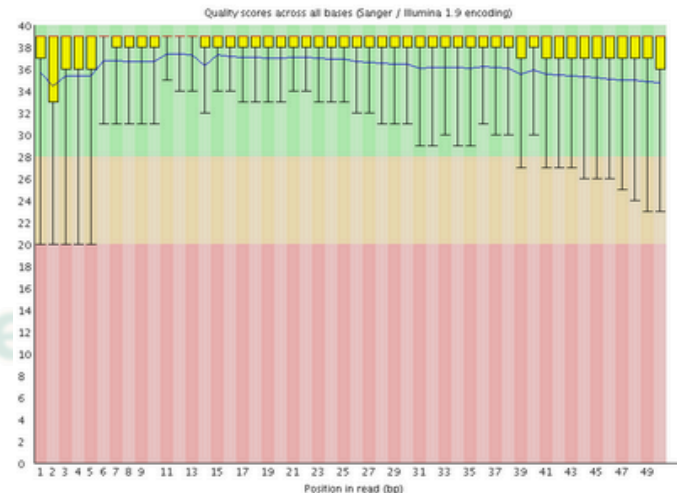
Summary




- ✔ [Basic Statistics](#)
- ✔ [Per base sequence quality](#)
- ✔ [Per sequence quality scores](#)
- ✘ [Per base sequence content](#)
- ✘ [Per base GC content](#)
- ! [Per sequence GC content](#)
- ✔ [Per base N content](#)
- ✔ [Sequence Length Distribution](#)
- ✘ [Sequence Duplication Levels](#)
- ! [Overrepresented sequences](#)
- ✘ [Kmer Content](#)




✔ Basic Statistics

Measure	Value
Filename	Adrenal_1
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	50121
Filtered Sequences	0
Sequence length	50
%GC	58

✔ Per base sequence quality



7:   
[Adrenal2FastQC Adrenal 2.html](#)

6:   
[Adrenal1FastQC Adrenal 1.html](#)

TopHat

NGS: RNA Analysis

1

RNA-SEQ

- [Tophat for Illumina](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(6hrs/6G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(12hrs/10G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(24hrs/16G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(48hrs/24G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(72hrs/36G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(96hrs/44G\)](#) Find splice junctions using RNA-seq data

2

Tophat for Illumina (48hrs/24G) (version 1.5.0)

3

RNA-Seq FASTQ file:

10: Brain_1

Nucleotide-space: Must have Sanger-scaled quality values with ASCII offset 33

Use a built in reference genome or own from your history:

Use a built-in genome

Built-ins genomes were created using default options

Select a reference genome:

Caenorhabditis elegans: ce10

If your genome of interest is not listed, contact the Galaxy team

Is this library mate-paired?:

Single-end

TopHat settings to use:

Default settings

Use the Full parameter list to change default settings.

Execute

1. Click on “NGS: RNA Analysis”
2. Click on “Tophat for Illumina (48hrs/24G)”
3. Default window with options appears

TopHat

Tophat for Illumina (version 1.5.0)

RNA-Seq FASTQ file:

1: Adrenal_1 ↕ 1

Nucleotide-space: Must have Sanger-scaled quality values with ASCII offset 33

Use a built in reference genome or own from your history:

Use a built-in genome ↕ 2a

Built-ins genomes were created using default options

Select a reference genome:

Human (Homo sapiens): hg19 Full ▾ 2b

If your genome of interest is not listed, contact the Galaxy team

Is this library mate-paired?:

Paired-end ↕ 3

RNA-Seq FASTQ file:

2: Adrenal_2 ↕ 4

Nucleotide-space: Must have Sanger-scaled quality values with ASCII offset 33

Mean Inner Distance between Mate Pairs:

110 5

TopHat settings to use:

Default settings ↕ 6

Use the Full parameter list to change default settings.




Execute 7




1. Select forward fastq read file
2. Select reference genome:
 - a) Choose "Use a built-in genome"
 - b) Select the reference genome
3. Select "Paired-end"
4. Select reverse fastq read file
5. Input "110" (ask sequencing center for this info)
6. Can choose "Commonly used" or "Full parameter list"
7. Click "Execute"
8. Do the exact same thing for the other sample




Note about FASTA files not already indexed in Galaxy




- If a FASTA is not indexed in Galaxy, then it is easy to upload the appropriate FASTA file into Galaxy. (Get Data -> Upload File)
- However, it can take up to 5 hours extra to run TopHat because Bowtie has to index your uploaded FASTA file (best to have your own instance of Galaxy) each time you run TopHat!
- Where do I go to get a non-model organism FASTA file?
 - NCBI: <http://www.ncbi.nlm.nih.gov/genome>
 - Ensembl: <http://useast.ensembl.org/info/data/ftp/index.html>
 - iGenome: <http://cufflinks.cbcb.umd.edu/igenomes.html>
 - Your favorite species website: <http://www...>

TopHat output files

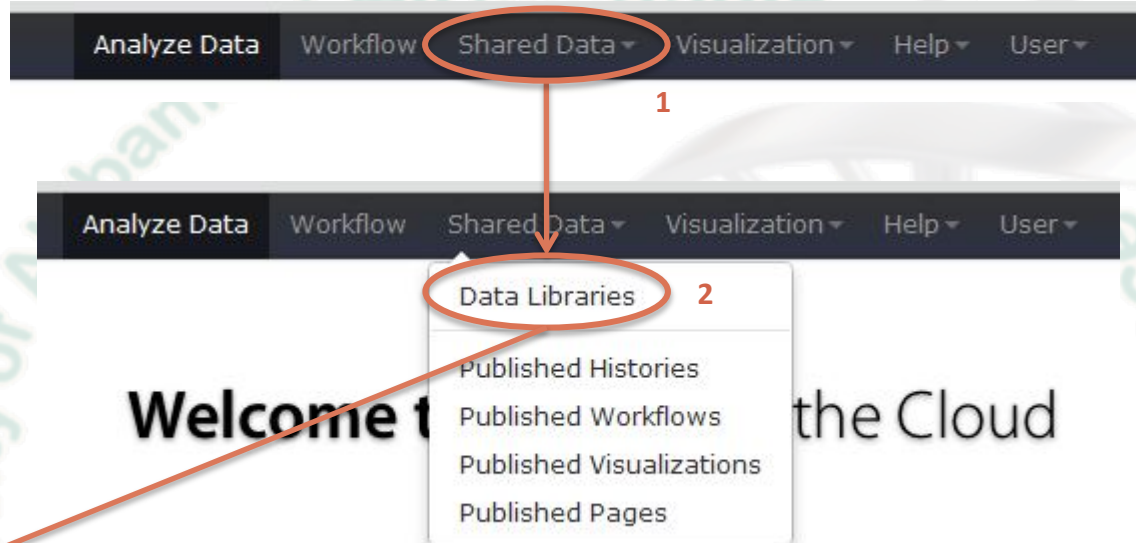
15: Adrenal Tophat for
Illumina on data 2 and data 1:
accepted hits   

14: Adrenal Tophat for
Illumina on data 2 and data 1: splice
junctions   

13: Adrenal Tophat for
Illumina on data 2 and data 1:
deletions   

12: Adrenal Tophat for
Illumina on data 2 and data 1:
insertions   

GTF Annotation Files



3

Data Library "Patched GTF annotation files for Cufflinks"

RefGene annotation files patched for Cufflinks in GTF format

Name	Message	Data type	Date uploaded	File size
<input type="checkbox"/> hg19_RefGene_patched3.gtf	None	gtf	2011-07-22	92.7 Mb
<input type="checkbox"/> mm9_RefGene_patched3.gtf	None	gtf	2011-07-22	65.5 Mb
<input type="checkbox"/> m4_RefGene_patched3.gtf		gtf	2012-02-29	38.4 Mb
<input type="checkbox"/> Tupaia_belangeri.TREESHREW.63.sorted2.patched.gtf	Not sure if the tupBel1 is the same build as 63!	gtf	2011-08-03	70.4 Mb
<input type="checkbox"/> Zv9_refGene_patched3.gtf		gtf	2012-02-29	35.6 Mb

For selected datasets:

The image shows a 'History' panel with a gear icon in the top right corner. Below the title bar, there are icons for a refresh and a minus sign on the left, and a plus and a document icon on the right. The text 'Unnamed history' is displayed on the left, and '186.8 Mb' is displayed on the right. Below this, a file entry is shown with a green background. The file name is '13: hg19 RefGene patched3.gtf'. To the right of the file name are icons for an eye, a pencil, and a cross.

Cufflinks

NGS: RNA Analysis

RNA-SEQ

- [Tophat for Illumina](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(6hrs/6G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(12hrs/10G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(24hrs/16G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(48hrs/24G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(72hrs/36G\)](#) Find splice junctions using RNA-seq data
- [Tophat for Illumina \(96hrs/44G\)](#) Find splice junctions using RNA-seq data
- [Cufflinks](#) transcript assembly and FPKM (RPKM) estimates for RNA-Seq data

Cufflinks (version 0.0.5)

SAM or BAM file of aligned RNA-Seq reads:

19: Brain Tophat for Illumina on data 4 and data 10: accepted_hits

Max Intron Length:

300000

Min Isoform Fraction:

0.1

Pre mRNA Fraction:

0.15

Perform quartile normalization:

No

Removes top 25% of genes from FPKM denominator to improve accuracy of differential expression calls for low abundance transcripts.

Use Reference Annotation:

No

Perform Bias Correction:

No

Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

Use multi-read correct:

No

Tells Cufflinks to do an initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome.

Execute

1. Click on “NGS: RNA Analysis”
2. Click on “Cufflinks”
3. Default window with options appears

Cufflinks

Cufflinks (version 0.0.5)

SAM or BAM file of aligned RNA-Seq reads:

15: Adrenal Tophat for Illumina on data 2 and data 1: accepted_hits ↕ 1

Max Intron Length:

300000

Min Isoform Fraction:

0.1

Pre mRNA Fraction:

0.15

Perform quartile normalization:

Yes ↕ 2

Removes top 25% of genes from FPKM denominator to improve accuracy of differential expression calls for low abundance transcripts.

Use Reference Annotation:

Use reference annotation as guide ↕ 3a

Reference Annotation:

5: iGenomes UCSC hg19, chr19 gene annotation ↕ 3b

Gene annotation dataset in GTF or GFF3 format.

Perform Bias Correction:

Yes ↕ 4

Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

Reference sequence data:

Locally cached ↕

Set Parameters for Paired-end Reads? (not recommended):

No ↕

Execute 5

1. Choose TopHat accepted hits file
2. Perform quartile normalization (for this demo choose "No")
3. Reference Annotation:
 - a) For genomes in scaffolds, choose "Use reference annotation as guide"
 - b) Choose GTF file from history
4. Perform Bias Correction (for this demo choose "No")
5. Click "Execute"
6. Do the exact same thing for the other TopHat accepted hits file

Note about GTF files for Cuff*

- If you use a GTF file from Ensembl, then you need to convert the chromosome column (column 1) to include 'chr' in front of the chromosome #. You can do this by:
 - Using Jeremy Goecks' published workflow "Make Ensembl GTF compatible with Cufflinks" in Galaxy:
<https://main.g2.bx.psu.edu/u/jeremy/w/make-ensembl-gtf-compatible-with-cufflinks>
 - Use 'awk' to add 'chr' to column 1 (if using Mac or Linux)
- Where do I go to get a GTF file?
 - NCBI: <http://www.ncbi.nlm.nih.gov/genome>
 - Ensembl: <http://useast.ensembl.org/info/data/ftp/index.html>
 - iGenome: <http://cufflinks.cbcb.umd.edu/igenomes.html>
 - Your favorite species website: <http://www...>

Some Cufflinks options to be aware of

-I/--max-intron-length <int>

The maximum intron length. Cufflinks will not report transcripts with introns longer than this, and will ignore SAM alignments with REF_SKIP CIGAR operations longer than this. The default is 300,000.




-F/--min-isoform-fraction <0.0-1.0>




After calculating isoform abundance for a gene, Cufflinks filters out transcripts that it believes are very low abundance, because isoforms expressed at extremely low levels often cannot reliably be assembled, and may even be artifacts of incompletely spliced precursors of processed transcripts. This parameter is also used to filter out introns that have far fewer spliced alignments supporting them. The default is 0.1, or 10% of the most abundant isoform (the major isoform) of the gene.




-j/--pre-mrna-fraction <0.0-1.0>

Some RNA-Seq protocols produce a significant amount of reads that originate from incompletely spliced transcripts, and these reads can confound the assembly of fully spliced mRNAs. Cufflinks uses this parameter to filter out alignments that lie within the intronic intervals implied by the spliced alignments. The minimum depth of coverage in the intronic region covered by the alignment is divided by the number of spliced reads, and if the result is lower than this parameter value, the intronic alignments are ignored. The default is 15%.

Cufflinks output files

[22: Adrenal Cufflinks on data 15 and data 5: assembled transcripts](#)   

[21: Adrenal Cufflinks on data 15 and data 5: transcript expression](#)   

[20: Adrenal Cufflinks on data 15 and data 5: gene expression](#)   

Cuffmerge

1

NGS: RNA Analysis

RNA-SEQ

2

- Cuffmerge merge together several Cufflinks assemblies

Cuffmerge (version 0.0.5)

GTF file produced by Cufflinks:

28: Cuffmerge on data 22, data 5, and data 26: merged transcripts

Additional GTF Input Files

Add new Additional GTF Input Files

Use Reference Annotation:

No

Use Sequence Data:

No

Use sequence data for some optional classification functions, including the addition of the p_id attribute required by Cuffdiff.

Execute

1. Click on “NGS: RNA Analysis”
2. Click on “Cuffmerge”
3. Default window with options appears

Cuffmerge

Cuffmerge (version 0.0.5)

GTF file produced by Cufflinks:

22: Adrenal Cufflinks on data 15 and data 5: assembled transcripts 1

Additional GTF Input Files

Additional GTF Input Files 1

GTF file produced by Cufflinks:

26: Brain Cufflinks on data 19 and data 5: assembled transcripts 2b

Remove Additional GTF Input Files 1

Add new Additional GTF Input Files 2a

Use Reference Annotation:

Yes 3a

Reference Annotation:

5: iGenomes UCSC hg19, chr19 gene annotation 3b

Make sure your annotation file is in GTF format and that Galaxy knows that your file is GTF--not GFF.

Use Sequence Data:

Yes 4a

Use sequence data for some optional classification functions, including the addition of the p_id attribute required by Cuffdiff.




Choose the source for the reference list:

Locally cached 4b

Execute 5

1. Choose GTF file produced by Cufflinks
2. Additional GTF Input Files:
 - a) Click on "Add new Additional GTF Input Files"
 - b) Choose other GTF file produced by Cufflinks
3. Reference Annotation:
 - a) Select "Yes" to Use Reference Annotation
 - b) Choose GTF Reference Annotation file from history
4. Sequence Data:
 - a) Select "Yes" to Use Sequence Data
 - b) Choose "Locally cached"
5. Click "Execute"

Cuffmerge output files

28: Cuffmerge on data 22, data 5, and data 26: merged transcripts   

Cuffdiff

1

NGS: RNA Analysis

RNA-SEQ

2

Cuffdiff find significant changes in transcript expression, splicing, and promoter use

1. Click on “NGS: RNA Analysis”
2. Click on “Cuffdiff”
3. Default window with options appears

Cuffdiff (version 0.0.5)

Transcripts:

28: Cuffmerge on data 22, data 5, and data 26: merged transcripts

A transcript GFF3 or GTF file produced by cufflinks, cuffcompare, or other source.

Perform replicate analysis:

No

Perform cuffdiff with replicates in each group.

SAM or BAM file of aligned RNA-Seq reads:

19: Brain Tophat for Illumina on data 4 and data 10: accepted_hits

SAM or BAM file of aligned RNA-Seq reads:

19: Brain Tophat for Illumina on data 4 and data 10: accepted_hits

Library normalization method:

geometric

Dispersion estimation method:

pooled

If using only one sample per condition, you must use 'blind.'

False Discovery Rate:

0.05

The allowed false discovery rate.

Min Alignment Count:

10

The minimum number of alignments in a locus for needed to conduct significance testing on changes in that locus observed between samples.

Perform quartile normalization:

No

Removes top 25% of genes from FPKM denominator to improve accuracy of differential expression calls for low abundance transcripts.

Use multi-read correct:

No

Tells Cufflinks to do an initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome.

Perform Bias Correction:

No

Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

Set Additional Parameters? (not recommended):

No

Execute

Cuffdiff

Cuffdiff (version 0.0.5)

Transcripts:

28: Cuffmerge on data 22, data 5, and data 26: merged transcripts 1
A transcript GTF file produced by cufflinks, cuffcompare, or other source.

Perform replicate analysis:

2a
Perform cuffdiff with replicates in each group.

Groups

Group 1
Group name (no spaces or commas):

2c

Replicates

Replicate 1

Add file:
 2d

2e

Group 2

Group name (no spaces or commas):

2g

Replicates

Replicate 1

Add file:
 2h

2i

2b, 2f, 2j

False Discovery Rate:

3
The allowed false discovery rate.

Min Alignment Count:

4
The minimum number of alignments in a locus for needed to conduct significance testing on changes in that locus observed between samples.

Perform quartile normalization:

5
Removes top 25% of genes from FPKM denominator to improve accuracy of differential expression calls for low abundance transcripts.

Perform Bias Correction:

6
Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

Reference sequence data:

Set Parameters for Paired-end Reads? (not recommended):

7

1. Choose GTF transcript file from either Cuffmerge or Cuffcompare
2. Perform replicate analysis:
 - a) Choose "Yes"
 - b) Click "Add new Group"
 - c) Select a name to give the Group
 - d) Choose TopHat accepted hits file associated with this Group
 - e) If you have more than one TopHat accepted hits file associated with this Group, then click "Add new Replicate"
 - f) Click "Add new Group"
 - g) Select a name to give the Group
 - h) Choose TopHat accepted hits file associated with this Group
 - i) If you have more than one TopHat accepted hits file associated with this Group, then click "Add new Replicate"
 - j) Click "Add new Group" if you have another Group you want to add
3. Select a False Discovery Rate cutoff
4. Select the minimum # of reads that will align to a locus in order to perform significant testing
5. Perform quartile normalization (for this demo choose "No")
6. Perform bias correction (for this demo choose "No")
7. Click "Execute"

Cuffdiff output files

39: Cuffdiff on data 19, data 15, and data 28: transcript FPKM tracking

38: Cuffdiff on data 19, data 15, and data 28: transcript differential expression testing

37: Cuffdiff on data 19, data 15, and data 28: gene FPKM tracking

36: Cuffdiff on data 19, data 15, and data 28: gene differential expression testing

35: Cuffdiff on data 19, data 15, and data 28: TSS groups FPKM tracking

34: Cuffdiff on data 19, data 15, and data 28: TSS groups differential expression testing

33: Cuffdiff on data 19, data 15, and data 28: CDS FPKM tracking

32: Cuffdiff on data 19, data 15, and data 28: CDS FPKM differential expression testing

31: Cuffdiff on data 19, data 15, and data 28: CDS overloading differential expression testing

30: Cuffdiff on data 19, data 15, and data 28: promoters differential expression testing

29: Cuffdiff on data 19, data 15, and data 28: splicing differential expression testing

Transcript differential expression testing output

test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
TCONS_00000001	XLOC_000001	OR4F5	chr1:69090-70008	Control	Treated	NOTEST	0	7.91888	1.79769e+308	1.79769e+308	0.369441	1	no
TCONS_00000002	XLOC_000002	LOC100132062	chr1:323891-328581	Control	Treated	OK	6512.86	50.1428	-7.0211	4.36714	1.25886e-05	0.000667762	yes
TCONS_00000003	XLOC_000002	LOC100133331	chr1:323891-328581	Control	Treated	OK	40727.9	1208.59	-5.07462	3.12382	0.00178519	0.0157435	yes
TCONS_00000004	XLOC_000003	OR4F29	chr1:367658-368597	Control	Treated	NOTEST	120.192	11.5757	-3.37617	0.827381	0.408021	1	no
TCONS_00000005	XLOC_000004	LOC643837	chr1:763015-791316	Control	Treated	OK	0	1136.01	1.79769e+308	1.79769e+308	0.0959697	0.130354	no
TCONS_00000006	XLOC_000004	LOC643837	chr1:763015-791316	Control	Treated	LOWDATA	0	0	-1.79769e+308	0	1	1	no
TCONS_00000007	XLOC_000005	SAMD11	chr1:861120-894687	Control	Treated	NOTEST	0	165.375	1.79769e+308	1.79769e+308	0.0784572	1	no
TCONS_00000008	XLOC_000006	KLHL17	chr1:895863-901099	Control	Treated	OK	0	935.161	1.79769e+308	1.79769e+308	0.0958257	0.130354	no
TCONS_00000009	XLOC_000006	KLHL17	chr1:895863-901099	Control	Treated	OK	0	1552.38	1.79769e+308	1.79769e+308	0.098175	0.130354	no
TCONS_00000010	XLOC_000006	KLHL17	chr1:895863-901099	Control	Treated	OK	0	653.036	1.79769e+308	1.79769e+308	0.0842346	0.130354	no
TCONS_00000011	XLOC_000007	PLEKHN1	chr1:901876-917473	Control	Treated	OK	0	259.895	1.79769e+308	1.79769e+308	0.0782193	0.130354	no
TCONS_00000012	XLOC_000007	PLEKHN1	chr1:901876-917473	Control	Treated	NOTEST	0	0	0	0	1	1	no
TCONS_00000013	XLOC_000007	PLEKHN1	chr1:901876-917473	Control	Treated	OK	0	366.221	1.79769e+308	1.79769e+308	0.077757	0.130354	no
TCONS_00000014	XLOC_000007	PLEKHN1	chr1:901876-917473	Control	Treated	NOTEST	0	0	0	0	1	1	no
TCONS_00000015	XLOC_000008	ISG15	chr1:948846-949919	Control	Treated	OK	0	6611.59	1.79769e+308	1.79769e+308	0.0677355	0.130354	no
TCONS_00000016	XLOC_000009	AGRN	chr1:955502-991492	Control	Treated	OK	0	27000.8	1.79769e+308	1.79769e+308	0.215057	0.219233	no
TCONS_00000017	XLOC_000010	LOC254099	chr1:1072396-1079434	Control	Treated	NOTEST	0	0	0	0	1	1	no
TCONS_00000018	XLOC_000011	MIR200B	chr1:1102483-1102578	Control	Treated	NOTEST	0	0	0	0	1	1	no
TCONS_00000019	XLOC_000012	MIR200A	chr1:1103242-1103332	Control	Treated	NOTEST	0	0	0	0	1	1	no
TCONS_00000020	XLOC_000013	MIR429	chr1:1104384-1104467	Control	Treated	NOTEST	0	0	0	0	1	1	no
TCONS_00000021	XLOC_000014	TTL10	chr1:1109285-1133313	Control	Treated	NOTEST	0	0	0	0	1	1	no
TCONS_00000022	XLOC_000014	TTL10	chr1:1109285-1133313	Control	Treated	NOTEST	0	0	0	0	1	1	no

Gene differential expression testing output

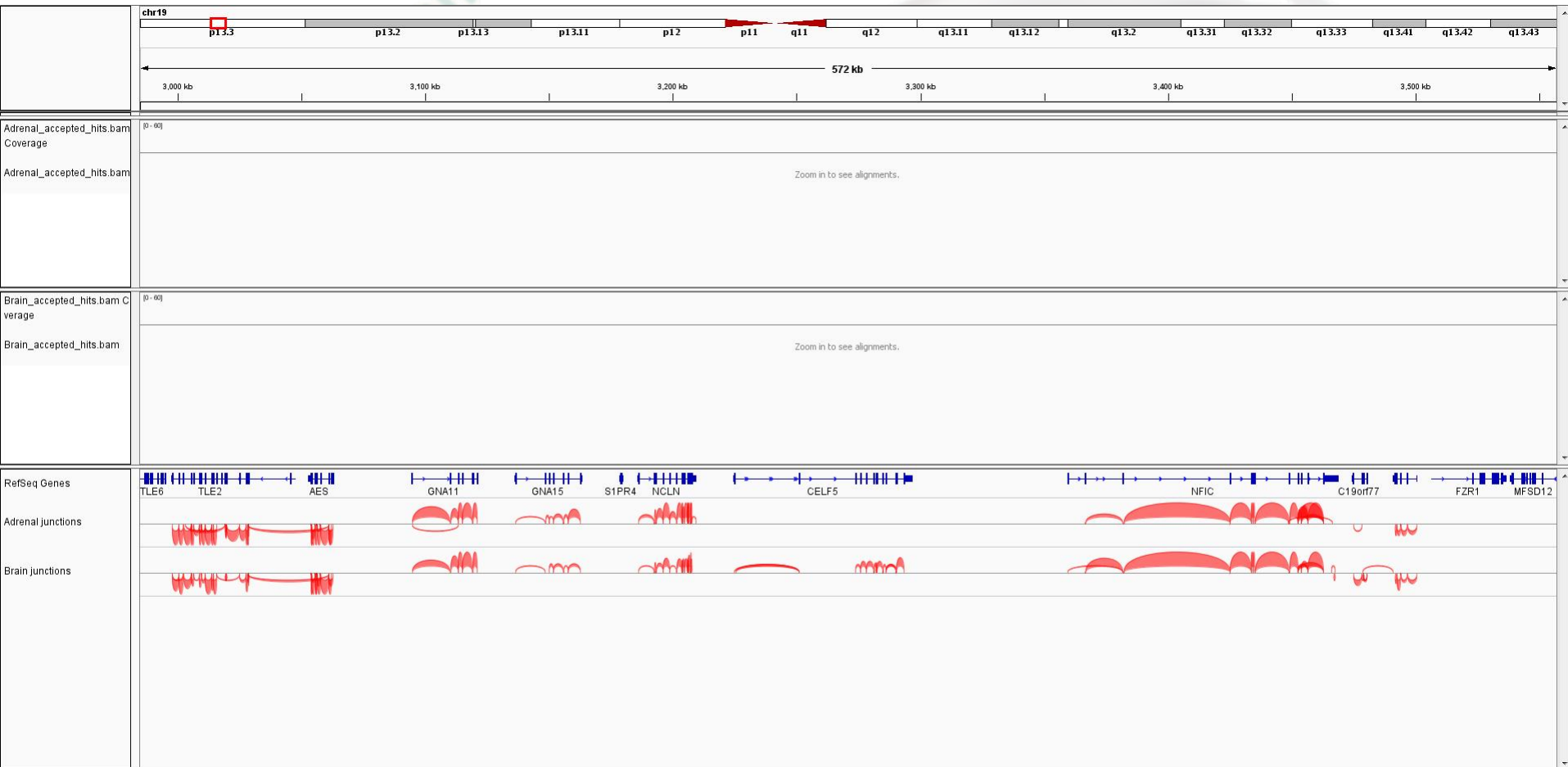
test_id	gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_change)	test_stat	p_value	q_value	significant
XLOC_000001	XLOC_000001	OR4F5	chr1:69090-70008	Control	Treated	NOTEST	0	7.91888	1.79769e+308	1.79769e+308	0.369441	1	no
XLOC_000002	XLOC_000002	LOC100132062,LOC100133331	chr1:323891-328581	Control	Treated	OK	47240.8	1258.73	-5.22999	3.58623	0.00033549	0.00357856	yes
XLOC_000003	XLOC_000003	OR4F29	chr1:367658-368597	Control	Treated	NOTEST	120.192	11.5757	-3.37617	0.827381	0.408021	1	no
XLOC_000004	XLOC_000004	LOC643837	chr1:763015-791316	Control	Treated	OK	0	1968.53	1.79769e+308	1.79769e+308	0.0161068	0.0355459	yes
XLOC_000005	XLOC_000005	SAMD11	chr1:861120-894687	Control	Treated	NOTEST	0	165.375	1.79769e+308	1.79769e+308	0.0784572	1	no
XLOC_000006	XLOC_000006	KLHL17	chr1:895863-901099	Control	Treated	OK	0	3140.58	1.79769e+308	1.79769e+308	0.00733214	0.0213299	yes
XLOC_000007	XLOC_000007	PLEKHN1	chr1:901876-917473	Control	Treated	OK	0	626.115	1.79769e+308	1.79769e+308	0.0132232	0.0313439	yes
XLOC_000008	XLOC_000008	ISG15	chr1:948846-949919	Control	Treated	OK	0	6611.59	1.79769e+308	1.79769e+308	0.0677355	0.0852164	no
XLOC_000009	XLOC_000009	AGRN	chr1:955502-991492	Control	Treated	OK	0	27000.8	1.79769e+308	1.79769e+308	0.215057	0.218471	no
XLOC_000010	XLOC_000010	LOC254099	chr1:1072396-1079434	Control	Treated	NOTEST	0	0	0	0	1	1	no
XLOC_000011	XLOC_000011	MIR200B	chr1:1102483-1102578	Control	Treated	NOTEST	0	0	0	0	1	1	no
XLOC_000012	XLOC_000012	MIR200A	chr1:1103242-1103332	Control	Treated	NOTEST	0	0	0	0	1	1	no
XLOC_000013	XLOC_000013	MIR429	chr1:1104384-1104467	Control	Treated	NOTEST	0	0	0	0	1	1	no
XLOC_000014	XLOC_000014	TTL10	chr1:1109285-1133313	Control	Treated	NOTEST	0	0	0	0	1	1	no
XLOC_000015	XLOC_000015	B3GALT6	chr1:1167628-1170420	Control	Treated	OK	0	1211.76	1.79769e+308	1.79769e+308	0.0668946	0.0852164	no
XLOC_000016	XLOC_000016	SCNN1D	chr1:1215815-1227409	Control	Treated	NOTEST	0	74.5236	1.79769e+308	1.79769e+308	0.0721728	1	no
XLOC_000017	XLOC_000017	PUS1	chr1:1243993-1260046	Control	Treated	OK	0	2317.82	1.79769e+308	1.79769e+308	0.0649866	0.0852164	no
XLOC_000018	XLOC_000018	GLTPD1	chr1:1260142-1264276	Control	Treated	OK	0	1597.74	1.79769e+308	1.79769e+308	0.0669804	0.0852164	no
XLOC_000019	XLOC_000019	TAS1R3	chr1:1266725-1269844	Control	Treated	NOTEST	0	31.2299	1.79769e+308	1.79769e+308	0.0912112	1	no
XLOC_000020	XLOC_000020	LOC148413	chr1:1334909-1342693	Control	Treated	OK	0	2591.73	1.79769e+308	1.79769e+308	0.101067	0.109708	no
XLOC_000021	XLOC_000021	TMEM88B	chr1:1361507-1363167	Control	Treated	NOTEST	0	0	0	0	1	1	no
XLOC_000022	XLOC_000022	VWA1	chr1:1370902-1378262	Control	Treated	NOTEST	0	4.59925	1.79769e+308	1.79769e+308	0.230105	1	no
XLOC_000023	XLOC_000023	ATAD3C	chr1:1385068-1405538	Control	Treated	OK	0	270.979	1.79769e+308	1.79769e+308	0.0615518	0.0852164	no
XLOC_000024	XLOC_000024	ATAD3B	chr1:1407163-1431582	Control	Treated	OK	0	9725.9	1.79769e+308	1.79769e+308	0.0932631	0.106586	no
XLOC_000025	XLOC_000025	ATAD3A	chr1:1447522-1470067	Control	Treated	OK	0	15128.3	1.79769e+308	1.79769e+308	0.125562	0.131737	no
XLOC_000026	XLOC_000026	MIB2	chr1:1550794-1565990	Control	Treated	OK	0	1139.11	1.79769e+308	1.79769e+308	0.00159396	0.00822516	yes

Using IGV to view the data

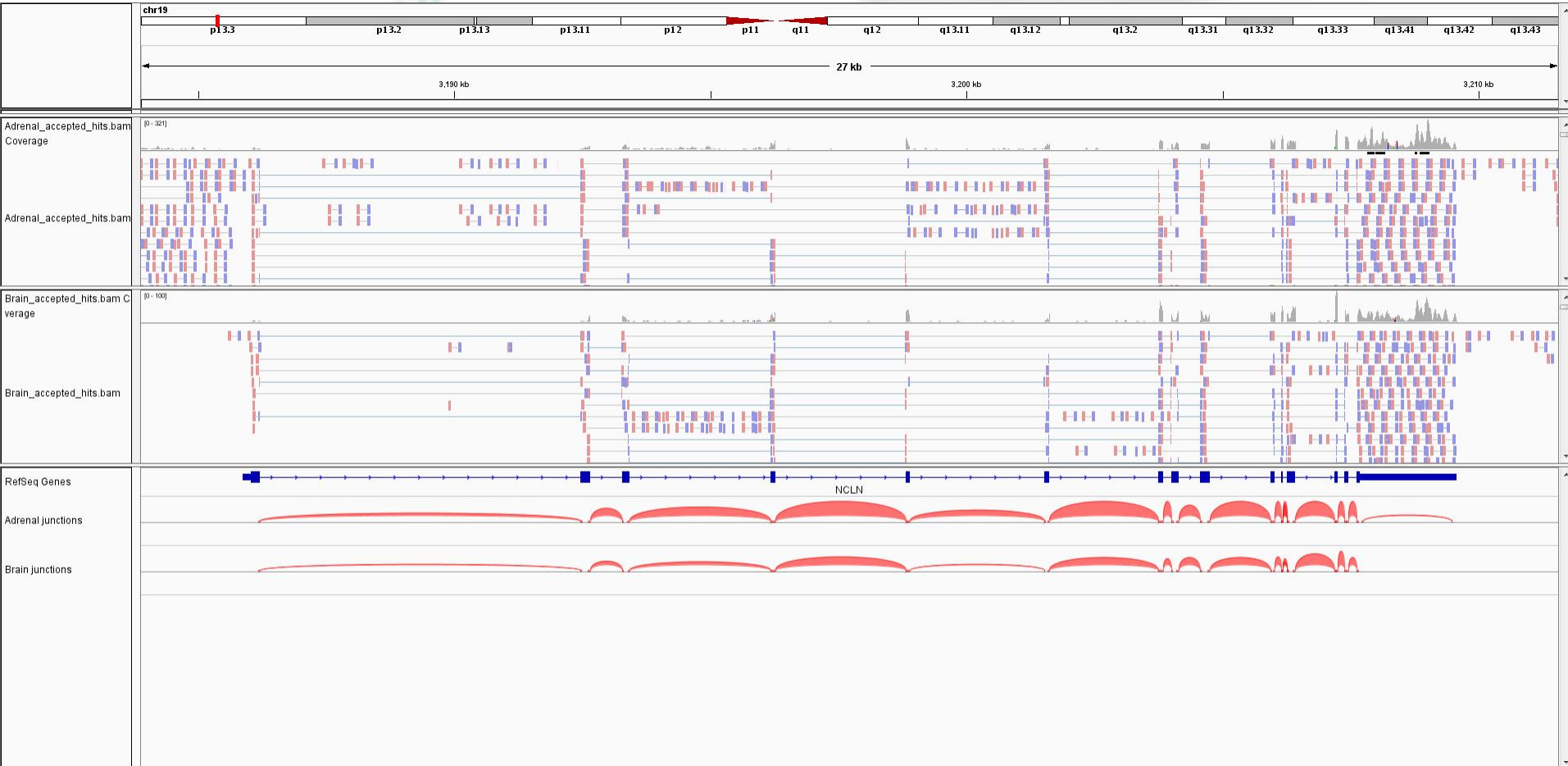
- <http://www.broadinstitute.org/igv/>
- Several ways to view the accepted_hits.bam file from TopHat:
 - Download the bam file to your computer (don't forget to download the bam_index file (*.bai) and then load into IGV
 - View them directly from Galaxy (no downloading required)

display at Ensembl Current
display with IGV web current local
display in IGB Local Web

chr19



NCLN



References and web links

- Galaxy
 - Public website: <https://main.g2.bx.psu.edu/>
 - UAB: <https://www.uab.edu/galaxy>
- TopHat
 - Trapnell C, Pachter L, Salzberg SL. [TopHat: discovering splice junctions with RNA-Seq](#). *Bioinformatics* doi:10.1093/bioinformatics/btp120
 - <http://tophat.cbcb.umd.edu/>
- Bowtie
 - Langmead B, Trapnell C, Pop M, Salzberg SL. [Ultrafast and memory-efficient alignment of short DNA sequences to the human genome](#). *Genome Biol* 10:R25.
 - <http://bowtie-bio.sourceforge.net/index.shtml>
- Cufflinks
 - Trapnell C, Williams BA, Pertea G, Mortazavi AM, Kwan G, van Baren MJ, Salzberg SL, Wold B, Pachter L. [Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation](#) *Nature Biotechnology* doi:10.1038/nbt.1621
 - Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. [Improving RNA-Seq expression estimates by correcting for fragment bias](#) *Genome Biology* doi:10.1186/gb-2011-12-3-r22
 - Roberts A, Pimentel H, Trapnell C, Pachter L. [Identification of novel transcripts in annotated genomes using RNA-Seq](#) *Bioinformatics* doi:10.1093/bioinformatics/btr355
 - <http://cufflinks.cbcb.umd.edu/>
- TopHat and Cufflinks protocol
 - Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. [Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks](#) *Nature Protocols* 7, 562-578 (2012) doi:10.1038/nprot.2012.016
- IGV
 - <http://www.broadinstitute.org/igv/>