



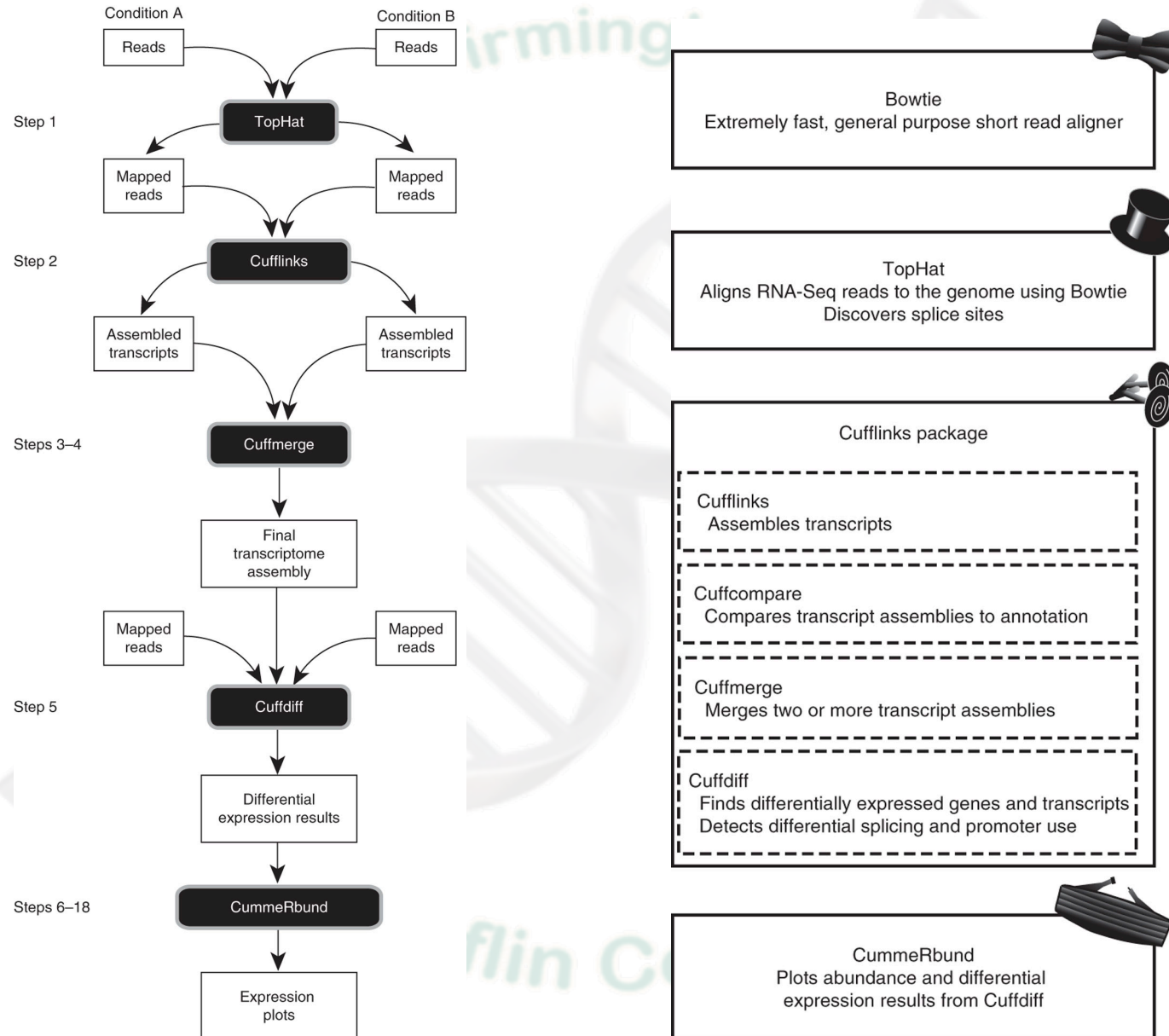
Transcriptome and Pathway Analysis

David Crossman, Ph.D.

UAB Heflin Center for Genomic Science

Immersion Course

RNA-Seq pipeline



Galaxy Splash Page

<https://www.uab.edu/galaxy>

<https://main.g2.bx.psu.edu/>

Galaxy / UAB

Analyze DataWorkflowShared DataVisualizationAdminHelpUser

Tools

search tools

Get Data

Send Data

Demo Tools

ENCODE Tools

Lift-Over

Text Manipulation

Filter and Sort

Join, Subtract and Group

Convert Formats

Extract Features

Fetch Sequences

Get Genomic Scores

Operate on Genomic Intervals

Statistics

Wavelet Analysis

Graph/Display Data

Regional Variation

Multiple regression

Multivariate Analysis

Evolution

Motif Tools

Multiple Alignments

Metagenomic analyses

FASTA manipulation

NCBI BLAST+

NGS TOOLBOX BETA

NGS: QC and manipulation

NGS: Assembly

NGS: Mapping

NGS: Indel Analysis

NGS: RNA Analysis

NGS: SAM Tools

NGS: HA_GSL_Tools

NGS: Peak Calling

SNP/WGA: Data; Filters

SNP/WGA: QC; LD; Plots

SNP/WGA: Statistical Models

Human Genome Variation

SnpEff tools

VCF Tools

DebugTools

EMBOSS

BEDTools

NGS: GATK

UCSC Tools

Phenotype Association

NGS: Analysis

Lastz

NGS: Picard

NGS: LEFSe

Workflows

All workflows

Galaxy is implementing a New Deletion Policy

To combat the growing problem of lack of disk space on UAB Galaxy, we will now be implementing an automated deletion of datasets that have not been updated in more than 6 months.
Deletion will happen on the third Monday of every month, with warning emails being issued the previous two Mondays.
Histories will not be affected by the auto-deletion only datasets.
For more information on UAB Galaxy's deletion policy please see: [Deletion Policy](#)

Welcome to UAB Galaxy!

Where all you need is a BlazerId and a web browser to run NGS analyses on the UAB Cheaha Cluster!

Local Resources

UAB Galaxy Wiki: [Overview](#), [Data Import](#)
UAB Mailing Lists
[UAB Galaxy—users](#) (search archive; [subscribe](#)) discuss with other UAB users
[UAB Galaxy—help](#) ask the UAB admins for help!
UAB Cheaha Computing Cluster
Cluster Hardware ([wiki](#))
Request a [cheaha account](#) (needed only for command-line access and bulk data upload)

Internet Resources

[Learn Galaxy](#) – tutorials
[Galaxy Project](#) user mailing list ([searchable archives](#); [subscribe](#); [post](#))
[Galaxy Toolshed](#) plug-ins for additional tools that you can request for installation at UAB
Public Galaxy Server at Penn State (PSU): [UseGalaxy.org](#) (more tools, but small disk quotas)

Brought to you by

UAB IT [Research Computing](#) under the Office of the Vice President for Information Technology at UAB
UAB [CCTS](#) (Center for Clinical and Translational Science under grant UL1 RR025777 from the NIH National Center for Research Resources)
The [Galaxy Platform](#) is developed by Penn State and Emory University

Live Quickies

History

search datasets

Unnamed history

0 bytes

This history is empty. You can [load your own data](#) or [get data from an external source](#)

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: 3a
Auto-detect
Which format? See help below

File: 3b-1
 No file chosen
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: 3b-3

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

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: 3c

Execute 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

1

2

3

Data Library “Immersion course prep”

Name	Message	Data type	Date uploaded	File size
<input type="checkbox"/> Control_rep1_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/> Control_rep1_r2.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/> Control_rep2_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/> Control_rep2_r2.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/> Treated_rep1_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/> Treated_rep1_r2.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/> Treated_rep2_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/> Treated_rep2_r2.fastq		fastqsanger	2012-08-06	14.4 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 "Immersion course prep"




1					
<input type="checkbox"/>	Name	Message	Data type	Date uploaded	File size
<input checked="" type="checkbox"/>	Control_rep1_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input checked="" type="checkbox"/>	Control_rep1_r2.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/>	Control_rep2_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/>	Control_rep2_r2.fastq		fastqsanger	2012-08-06	14.4 Mb
<input checked="" type="checkbox"/>	Treated_rep1_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input checked="" type="checkbox"/>	Treated_rep1_r2.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/>	Treated_rep2_r1.fastq		fastqsanger	2012-08-06	14.4 Mb
<input type="checkbox"/>	Treated_rep2_r2.fastq		fastqsanger	2012-08-06	14.4 Mb
For selected datasets: Import to current history <input type="button" value="Go"/> 2					




3



History





Unnamed history 0 bytes

4: Treated rep1 r2.fastq   

3: Treated rep1 r1.fastq   

2: Control rep1 r2.fastq   

1: Control rep1 r1.fastq   

1. Check boxes of files you want to import
2. Choose "Import to current history" and then click "Go"
3. Will see the files in the right-hand pane of the Galaxy window

Quality Control of raw fastq reads

The screenshot displays the Babraham FASTQC web interface. On the left, a 'Tools' sidebar lists various bioinformatics tools. Two items are circled in red: 'NGS: QC and manipulation' (labeled with a red '1') and 'Fastqc: Fastqc QC using FastQC from Babraham' (labeled with a red '2'). An arrow points from the second item to the main panel. The main panel is divided into two sections, 3a and 3b, both titled 'Fastqc: Fastqc QC (version 0.4)'. Section 3a shows the initial state with 'Short read data from your current history' set to '4: Treated_rep1_r2.fastq', an empty 'Title for the output file' field, and a 'Contaminant list' dropdown set to 'Selection is Optional'. A red arrow points down from the 'Execute' button in section 3a to section 3b. Section 3b shows the state after clicking 'Execute': the 'Short read data' is now '1: Control_rep1_r1.fastq' (marked with a red asterisk), the 'Title for the output file' is 'Control rep1 r1 FastQC' (also marked with a red asterisk), and the 'Execute' button is now disabled (labeled with a red '4').

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

3a Fastqc: Fastqc QC (version 0.4)

Short read data from your current history:
4: Treated_rep1_r2.fastq

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

Contaminant list:
Selection is Optional

tab delimited file with 2 columns: name and sequence. For example: Illumina Small RNA RT Primer CAAGCAGAAGACGGCATACGA

Execute

3b Fastqc: Fastqc QC (version 0.4)

Short read data from your current history:
1: Control_rep1_r1.fastq *

Title for the output file - to remind you what the job was for:
Control rep1 r1 FastQC *

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

This data looks awful because this is filtered data from a much larger fastq file. Better results when using entire file!

History	
Unnamed history	1.3 Mb
8: Treated rep1 r2 FastQC_data 4.html	  
7: Treated rep1 r1 FastQC_data 3.html	  
6: Control rep1 r2 FastQC_data 2.html	  
5: Control rep1 r1 FastQC_data 1.html	  
4: Treated rep1 r2.fastq	  
3: Treated rep1 r1.fastq	  
2: Control rep1 r2.fastq	  
1: Control rep1 r1.fastq	  

Control_rep1_r1.fastq FastQC Report

FastQC Report
Mon 6 Aug 2012
Control_rep1_r1.fastq

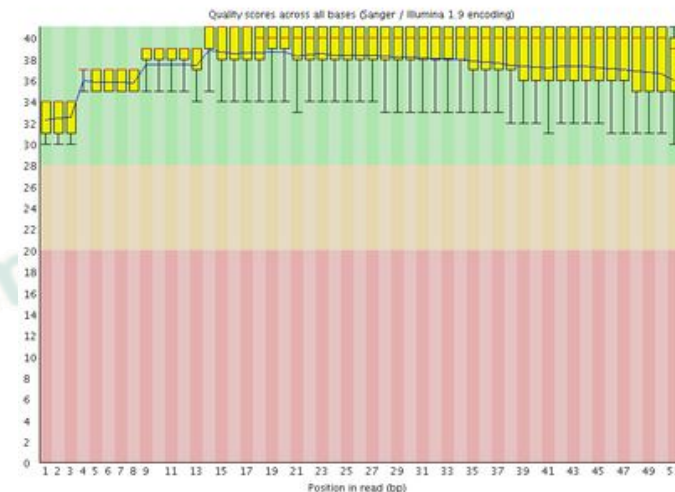
Summary

- ✓ 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	Control_rep1_r1.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	100000
Filtered Sequences	0
Sequence length	51
%GC	46

Per base sequence quality



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

3

Tophat for Illumina (6hrs/6G) (version 1.5.0)

RNA-Seq FASTQ file:

4: Treated_rep1_r2.fastq ▼

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

Will you select a reference genome from your history or use a built-in index?:

Use a built-in index ▼

Built-ins were indexed using default options

Select a reference genome:

A. thaliana Feb. 2011 (arabidopsis.org/tair) ▼

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

Is this library mate-paired?:

Single-end ▼

TopHat settings to use:

Use Defaults ▼

You can use the default settings or set custom values for any of Tophat's parameters.

Execute

1. Click on "NGS: RNA Analysis"
2. Click on "Tophat for Illumina"
3. Default window with options appears

TopHat

Tophat for Illumina (6hrs/6G) (version 1.5.0)

RNA-Seq FASTQ file:

1: Control_rep1_r1.fastq ▼ 1

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

Will you select a reference genome from your history or use a built-in index?:

Use a built-in index ▼ 2a

Built-ins were indexed using default options

Select a reference genome:

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: Control_rep1_r2.fastq ▼ 4

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

Mean Inner Distance between Mate Pairs:

150 5

TopHat settings to use:

Commonly used ▼ 6

For most mapping needs use Commonly used settings. If you want full control use Full parameter list

Execute 7

1. Select forward fastq read file
2. Select reference genome:
 - a) Choose "Use a built-in index"
 - b) Select the reference genome
3. Select "Paired-end"
4. Select reverse fastq read file
5. Input "150" (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

✓ The following job has been successfully added to the queue:

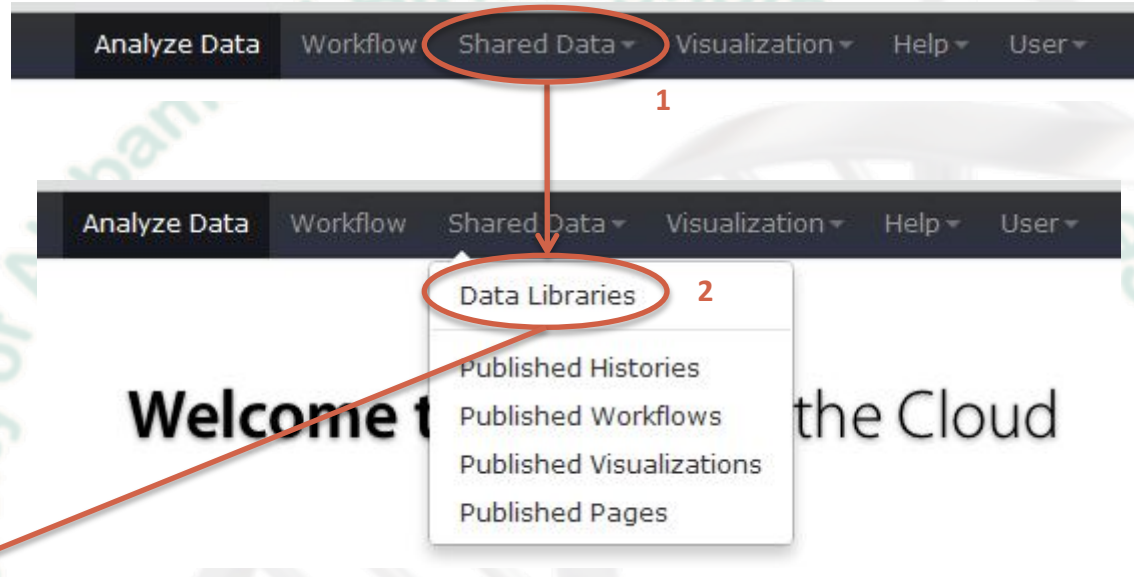
13: Tophat for Illumina (6hrs/6G) on data 2 and data 1: splice junctions

14: Tophat for Illumina (6hrs/6G) on data 2 and data 1: accepted_hits

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History		⚙
 		 
Unnamed history		94.1 Mb
<u>12: Treated Tophat for Illumina (6hrs/6G) on data 4 and data 3: accepted_hits</u>		  
<u>11: Treated Tophat for Illumina (6hrs/6G) on data 4 and data 3: splice junctions</u>		  
<u>10: Control Tophat for Illumina (6hrs/6G) on data 2 and data 1: accepted_hits</u>		  
<u>9: Control Tophat for Illumina (6hrs/6G) on data 2 and data 1: splice junctions</u>		  

GTF Annotation Files



Data Library "Patched GTF annotation files for Cufflinks"

RefGene annotation files patched for Cufflinks in GTF format

<input type="checkbox"/> 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"/> rn4_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:



Cufflinks

NGS: RNA Analysis

RNA-SEQ

1

- [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

2

3

Cufflinks (version 0.0.5)

SAM or BAM file of aligned RNA-Seq reads:

12: Treated Tophat fo...cepted_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.

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

No

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:

10: Control Tophat fo...cepted_hits **1**

Max Intron Length:

300000

Min Isoform Fraction:

0.1

Pre MRNA Fraction:

0.15

Perform quartile normalization:

No **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:

13: hg19_RefGene_patched3.gtf **3b**

Gene annotation dataset in GTF or GFF3 format.

Perform Bias Correction:

No **4**

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

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

No

Execute **5**

1. Choose TopHat accepted hits file
2. Perform quartile normalization (for this demo sample, 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



The following job has been successfully added to the queue:























14: Cufflinks on data 10 and data 13: gene expression

15: Cufflinks on data 10 and data 13: transcript expression

16: Cufflinks on data 10 and data 13: assembled transcripts

17: Cufflinks on data 10 and data 13: total map mass

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History	
 	 
Unnamed history	361.6 Mb
<u>20: Treated Cufflinks on data 12 and data 13: assembled transcripts</u>	  
<u>19: Treated Cufflinks on data 12 and data 13: transcript expression</u>	  
<u>18: Treated Cufflinks on data 12 and data 13: gene expression</u>	  
<u>16: Control Cufflinks on data 10 and data 13: assembled transcripts</u>	  
<u>15: Control Cufflinks on data 10 and data 13: transcript expression</u>	  
<u>14: Control Cufflinks on data 10 and data 13: gene expression</u>	  

Cuffmerge

1

NGS: RNA Analysis

RNA-SEQ

2

- Cuffmerge merge together several Cufflinks assemblies

Cuffmerge (version 0.0.5)

GTF file produced by Cufflinks:

20: Treated Cufflinks..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:

16: Control Cufflinks..transcripts ▾ 1

Additional GTF Input Files

Additional GTF Input Files 1

GTF file produced by Cufflinks:

20: Treated Cufflinks..transcripts ▾ 2b

Remove Additional GTF Input Files 1

Add new Additional GTF Input Files 2a

Use Reference Annotation:

Yes ▾ 3a

Reference Annotation:

13: hg19_RefGene_patched3.gtf ▾ 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

✓ The following job has been successfully added to the queue:




22: Cuffmerge on data 16, data 13, and data 20: merged transcripts

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History



Unnamed history456.3 Mb

22: Cuffmerge on data 16, data 13, and data 20: merged transcripts

Cuffdiff

1

NGS: RNA Analysis

RNA-SEQ

2

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

Cuffdiff (version 0.0.5)

Transcripts:

22: Cuffmerge on data..transcripts ▼

A transcript 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:

12: Treated Tophat fo..cepted_hits ▼

SAM or BAM file of aligned RNA-Seq reads:

12: Treated Tophat fo..cepted_hits ▼

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.

Perform Bias Correction:

No ▼

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

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

No ▼

Execute

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

Cuffdiff

Cuffdiff (version 0.0.5)

Transcripts:
22: Cuffmerge on data..transcripts
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):
Control **2c**

Replicates

Replicate 1

Add file:
10: Control Tophat fo..cepted_hits

2e

Group 2

Group name (no spaces or commas):
Treated **2g**

Replicates

Replicate 1

Add file:
12: Treated Tophat fo..cepted_hits

2i

2b, 2f, 2j

False Discovery Rate:
0.05 **3**
The allowed false discovery rate.

Min Alignment Count:
10 **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.

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



The following job has been successfully added to the queue:

- 23: Cuffdiff on data 12, data 10, and data 22: splicing differential expression testing
- 24: Cuffdiff on data 12, data 10, and data 22: promoters differential expression testing
- 25: Cuffdiff on data 12, data 10, and data 22: CDS overloading differential expression testing
- 26: Cuffdiff on data 12, data 10, and data 22: CDS FPKM differential expression testing
- 27: Cuffdiff on data 12, data 10, and data 22: CDS FPKM tracking
- 28: Cuffdiff on data 12, data 10, and data 22: TSS groups differential expression testing
- 29: Cuffdiff on data 12, data 10, and data 22: TSS groups FPKM tracking
- 30: Cuffdiff on data 12, data 10, and data 22: gene differential expression testing
- 31: Cuffdiff on data 12, data 10, and data 22: gene FPKM tracking
- 32: Cuffdiff on data 12, data 10, and data 22: transcript differential expression testing
- 33: Cuffdiff on data 12, data 10, and data 22: transcript FPKM tracking

You can check the status of queued jobs and view the resulting data by refreshing the **History** pane. When the job has been run the status will change from 'running' to 'finished' if completed successfully or 'error' if problems were encountered.

History	
Unnamed history 482.8 Mb	
33: Cuffdiff on data 12, data 10, and data 22: transcript FPKM tracking	🔄 ⌂
32: Cuffdiff on data 12, data 10, and data 22: transcript differential expression testing	🔄 ⌂
31: Cuffdiff on data 12, data 10, and data 22: gene FPKM tracking	🔄 ⌂
30: Cuffdiff on data 12, data 10, and data 22: gene differential expression testing	🔄 ⌂
29: Cuffdiff on data 12, data 10, and data 22: TSS groups FPKM tracking	🔄 ⌂
28: Cuffdiff on data 12, data 10, and data 22: TSS groups differential expression testing	🔄 ⌂
27: Cuffdiff on data 12, data 10, and data 22: CDS FPKM tracking	🔄 ⌂
26: Cuffdiff on data 12, data 10, and data 22: CDS FPKM differential expression testing	🔄 ⌂
25: Cuffdiff on data 12, data 10, and data 22: CDS overloading differential expression testing	🔄 ⌂
24: Cuffdiff on data 12, data 10, and data 22: promoters differential expression testing	🔄 ⌂
23: Cuffdiff on data 12, data 10, and data 22: 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

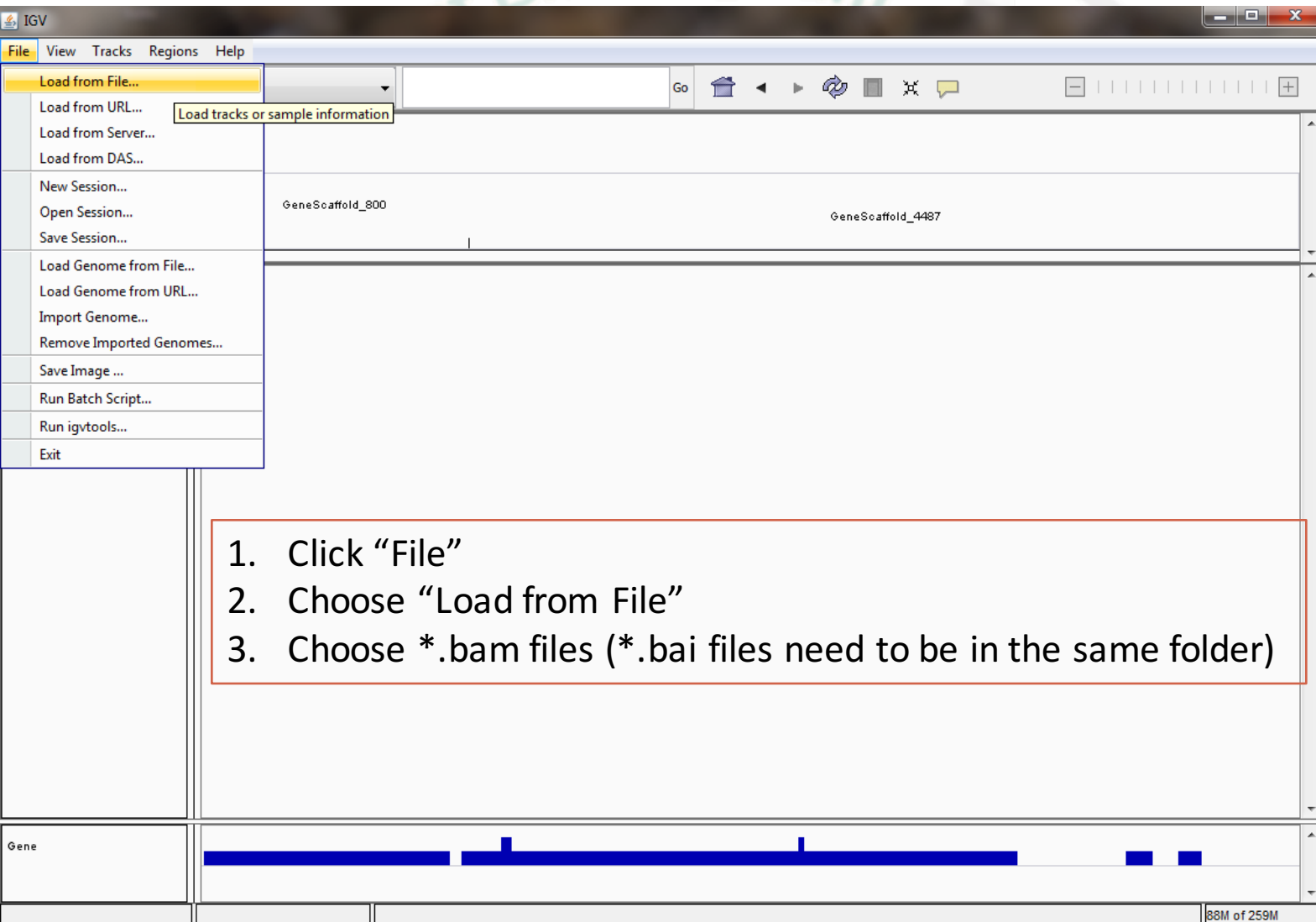
- <http://www.broadinstitute.org/igv/>
- Several ways to view the accepted_hits.bam file from TopHat:
 - Download the bam file to your computer (don't forget to download the bam_index file (*.bai) and then load into IGV
 - View them directly from Galaxy (no downloading required)

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

Load aligned BAM files into IGV

1

2



1. Click "File"

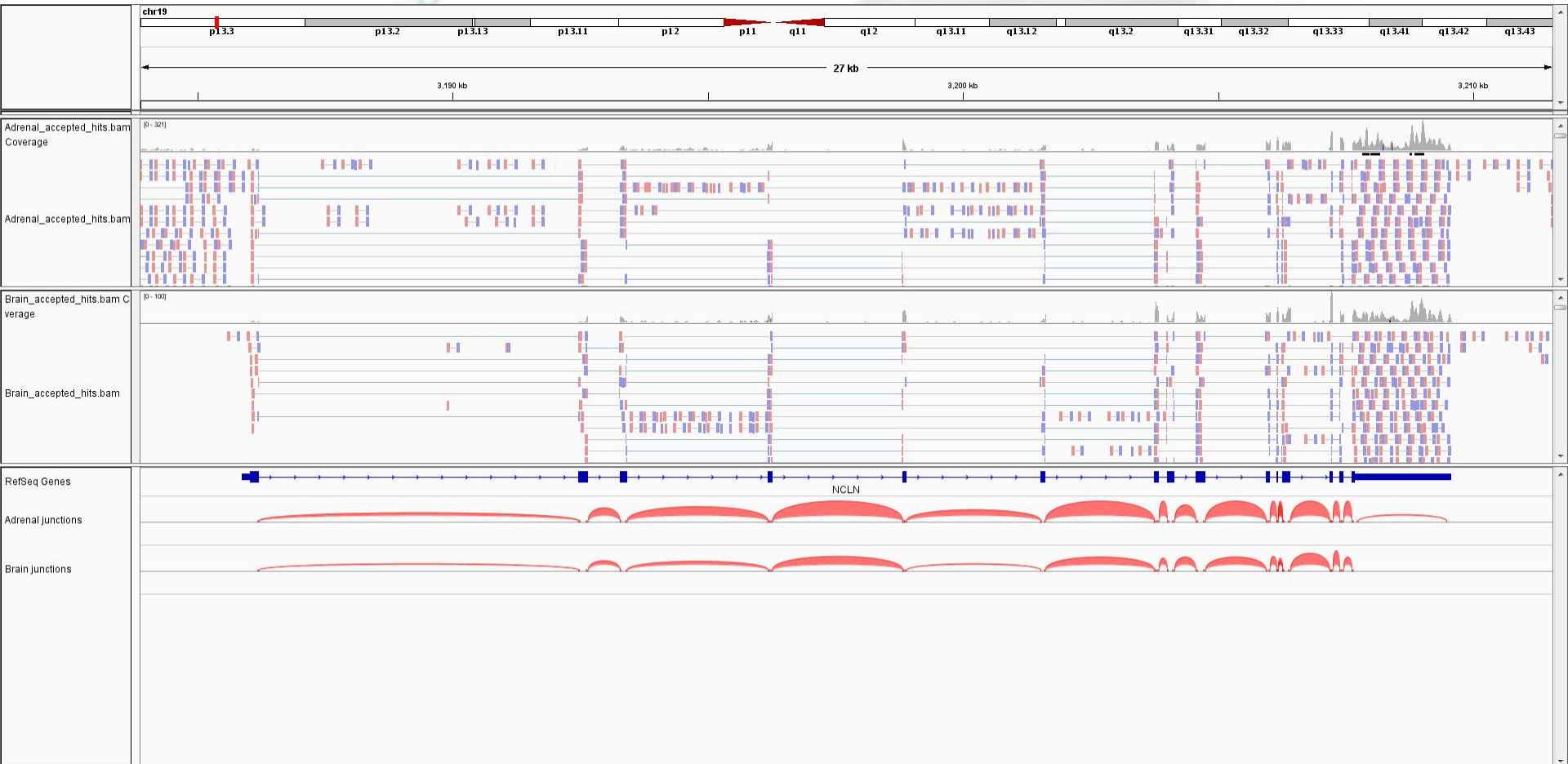
2. Choose "Load from File"

3. Choose *.bam files (*.bai files need to be in the same folder)

chr19



NCLN



What to do with your list of genes

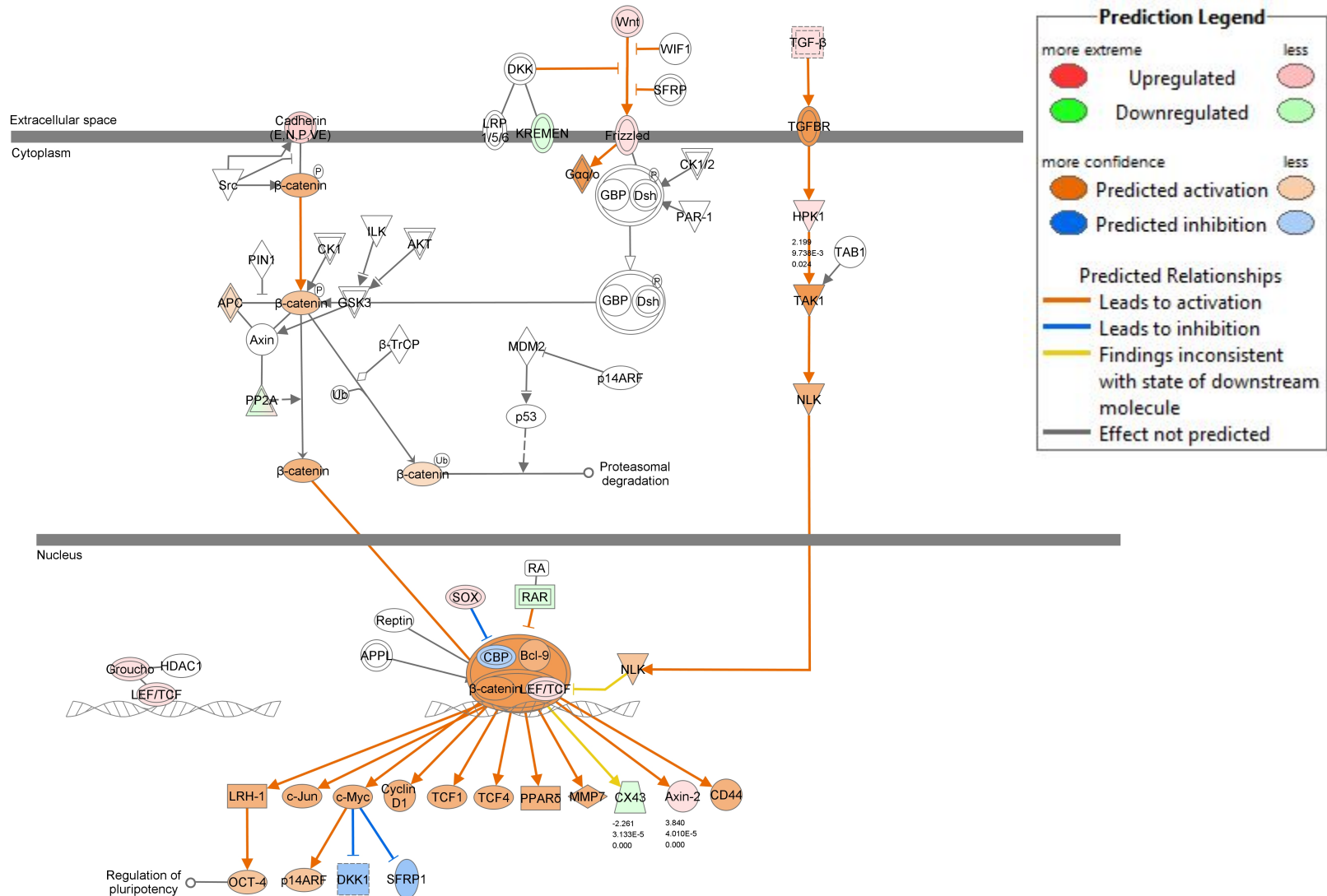
- Apply a Systems Biology approach to data mine and analyze your data
- Tools and databases available (some free, others \$\$) to define the underlying biology behind different –omics data
- These tools and databases will identify and prioritize the most relevant pathways, networks and cellular processes affected by your dataset.

Pathways & Ontology Analysis Tools

Tool	Link	Price
Reactome	http://www.reactome.org	Free
IPA	http://ingenuity.com/	\$\$\$
GeneGo Metacore	http://www.genego.com/	\$\$\$
Cytoscape	http://www.cytoscape.org/	Free
GenMAPP	http://www.genmapp.org/	Free
InterMine	http://intermine.org/	Free
KEGG	http://www.genome.jp/kegg/	Free
GO	http://www.geneontology.org/	Free
Panther	http://www.pantherdb.org/	Free
DAVID	http://david.abcc.ncifcrf.gov/	Free
And many, many more!!		

Network Example

Wnt/ β -catenin Signaling

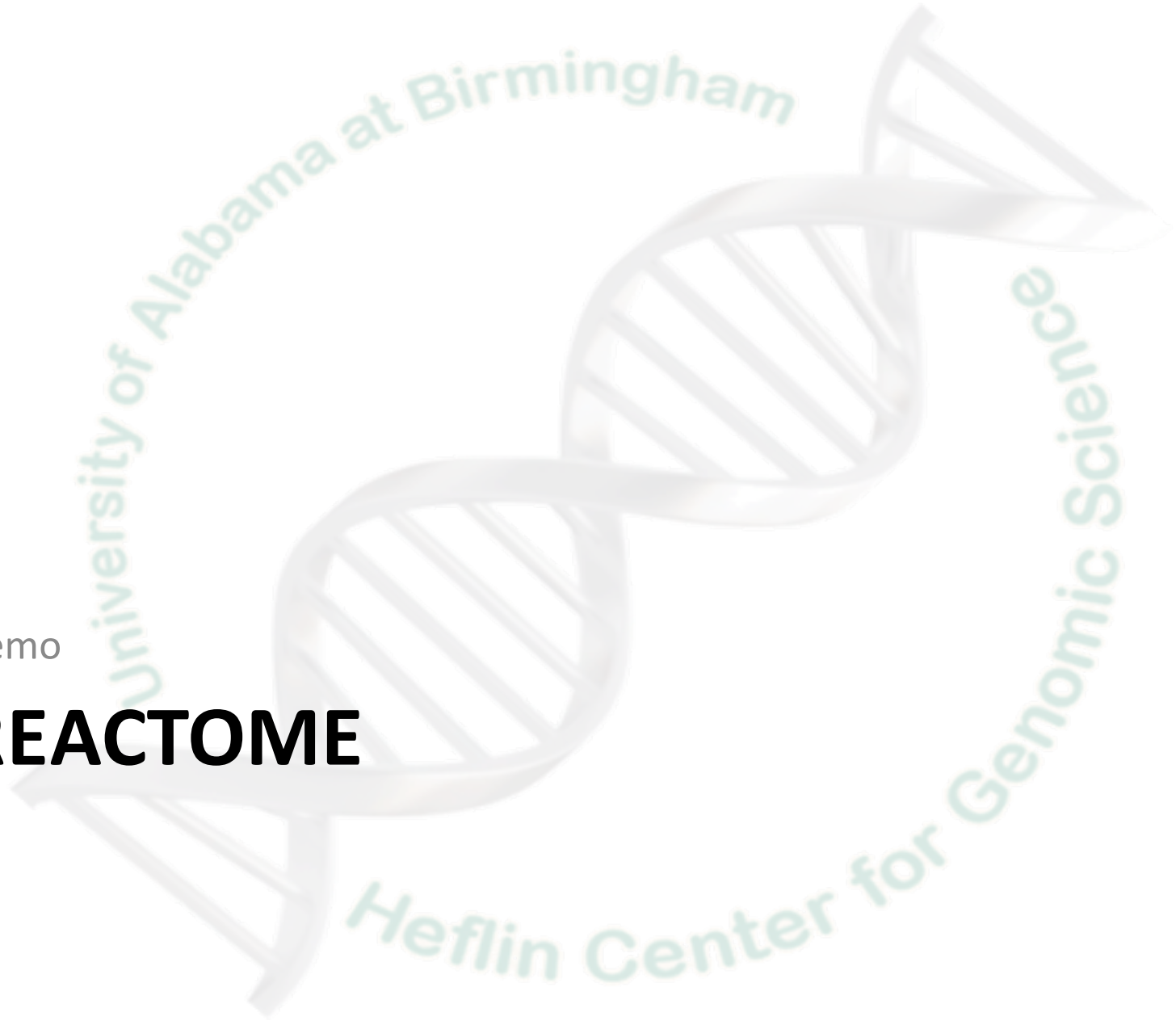


Variant Pathway Tools

Tool	Link	Price
Ingenuity Variant Analysis	http://ingenuity.com/	\$\$\$
Clinical Genomics Toolkit	http://lsresearch.thomsonreuters.com/pages/solutions/21/clinical-genomics-toolkit	\$\$\$
Cartagenia	http://www.cartagenia.com/	\$\$\$
Tute Genomics	http://tutegenomics.com/	\$\$\$
VariantStudio	http://www.illumina.com/informatics/research/biological-data-interpretation/variantstudio.html	\$\$\$

Demo

REACTOME



Reactome

- Open-source, open access.
- Manually curated.
- Peer-reviewed pathway database (pathway annotations are authored by “expert” biologists).
- Some of the tools they have:
 - Browse pathways
 - Map IDs to pathways
 - Overrepresentation analysis
 - Compare species
 - Analyze expression data

Quick links
to tools
most
commonly
used.

Manual &
tutorials

Other
useful
tools



The header of the Reactome website features a dark blue background with a stylized illustration of a metabolic pathway on the left and a 3D molecular model on the right. The word "REACTOME" is prominently displayed in large, white, serif capital letters, with the tagline "A CURATED PATHWAY DATABASE" in smaller white text below it. A navigation bar contains links: "About", "Content", "Documentation", "Tools", "Community", "Download", and "Contact". A search bar on the right contains the text "e.g. O95631, NTN1, signalir" and a "Search" button. Red boxes highlight the "Documentation" and "Tools" links, and a red arrow points from the "Tools" link to the "Browse Pathways" button in the main menu.



A grid of six buttons with icons and text labels, all enclosed in a red border. The buttons are: "Browse Pathways" (hierarchy icon), "Analyze Data" (magnifying glass over a bar chart with a "New" star), "Reactome FI Network" (network graph icon), "User Guide" (document with person icon), "Data Download" (downward arrow icon), and "Contact Us" (envelope icon).



A section titled "Tweets" with a dark blue header. Below the header, it displays "Current Version: Reactome V51" in blue text, followed by "Tweets by @reactome" in a smaller blue font.

About Reactome

Reactome is a free, open-source, curated and peer reviewed pathway database. Our goal is to provide intuitive bioinformatics tools for the visualization, interpretation and analysis of pathway knowledge to support basic research, genome analysis, modeling, systems biology and education. The current version (v51) of Reactome was released on December 8, 2014.



The development of Reactome is supported by a grant from the US National Institutes of Health (P41 HG003751), Ontario Research Fund, and the European Molecular Biology Laboratory.

About

- About Reactome
- News
- Reactome Team
- Scientific Advisory Board
- Other Reactomes
- License Agreement
- Reactome Disclaimer

Contact

Content

- Table of Contents
- DOIs
- Data Schema
- Editorial Calendar
- Statistics
- ORCID Integration Project

Documentation

- User Guide
- Data Model
- Orthology Prediction
- Object/Relational Mapping
- Wiki
- Linking to Reactome
- Referencing Reactome

Tools

- Pathway Browser
- Analyze Data
- Species Comparison
- Reactome FI Network
- Advanced Search
- Author/Reviewer Search
- Small Molecule Search
- Analysis Service

Community

- Reactome Outreach
- Reactome Events
- Reactome Training
- Reactome Publications
- Papers Citing Reactome
- Resources Guide
- Mailing List

Download



Browse Pathways

REACTOME

Pathways for: Homo sapiens

Tour this pathway browser? Hide

Protein Small molecule Complex

Event Hierarchy:

- Binding and Uptake of Ligands by S
- Cell Cycle
- Cell-Cell communication
- Cellular responses to stress
- Chromatin organization
- Circadian Clock
- Developmental Biology
- Disease
- DNA Repair
- DNA Replication
- Extracellular matrix organization
- Gene Expression
- Hemostasis
- Immune System
- Membrane Trafficking
- Metabolism
- Metabolism of proteins
- Muscle contraction
- Neuronal System
- Organelle biogenesis and mainte
- Programmed Cell Death
- Reproduction
- Signal Transduction
- Transmembrane transport of sma

Welcome to the Reactome Pathway Browser, a tool for visualizing and interacting with pathways.

To view a Pathway Diagram either:

1. Click on a pathway name in the hierarchical list on the left.
2. Return to the Homepage (button top left) and text search using a pathway, protein or compound name, or an accession number (e.g. Q9HCN6). In the resulting list, click on pathway name to open it in the Pathway Browser.

To see a key to pathway diagrams click the Diagram Key link top right

For a detailed explanation see Section 3 of the [Reactome User Guide](#)

For help please contact us (help@reactome.org)

The Pathway Browser includes tools for several types of analysis, detailed in Section 5 of the [Reactome User Guide](#).

Overview Molecules Structures Expression Analysis Processes Downloads

Will display context-sensitive, general information for the item you've clicked in the diagram above or the Pathway hierarchy on the left-hand side.

Helpful tutorial

This beginning screen says it all on how to browse pathways

Analyze Data

Analysis Tools

This tool merges pathway identifier mapping, overrepresentation and expression analysis into a single tabbed data analysis portal, with integrated visualization and summary features.

Select a file from your computer and click on the "Analyze" button to perform the analysis.

Select data file for analysis **1b** No file chosen

☒ Project to human

2

▼ [Click here to paste your data or try example data sets...](#)

Paste the data to analyse

1a

1. Upload data to Reactome:
 - a. Paste, or
 - b. Choose file from computer (tab-delimited), or
 - c. For demo purposes, click "Example"
2. Click "Analyze"

☒ Project to human

2

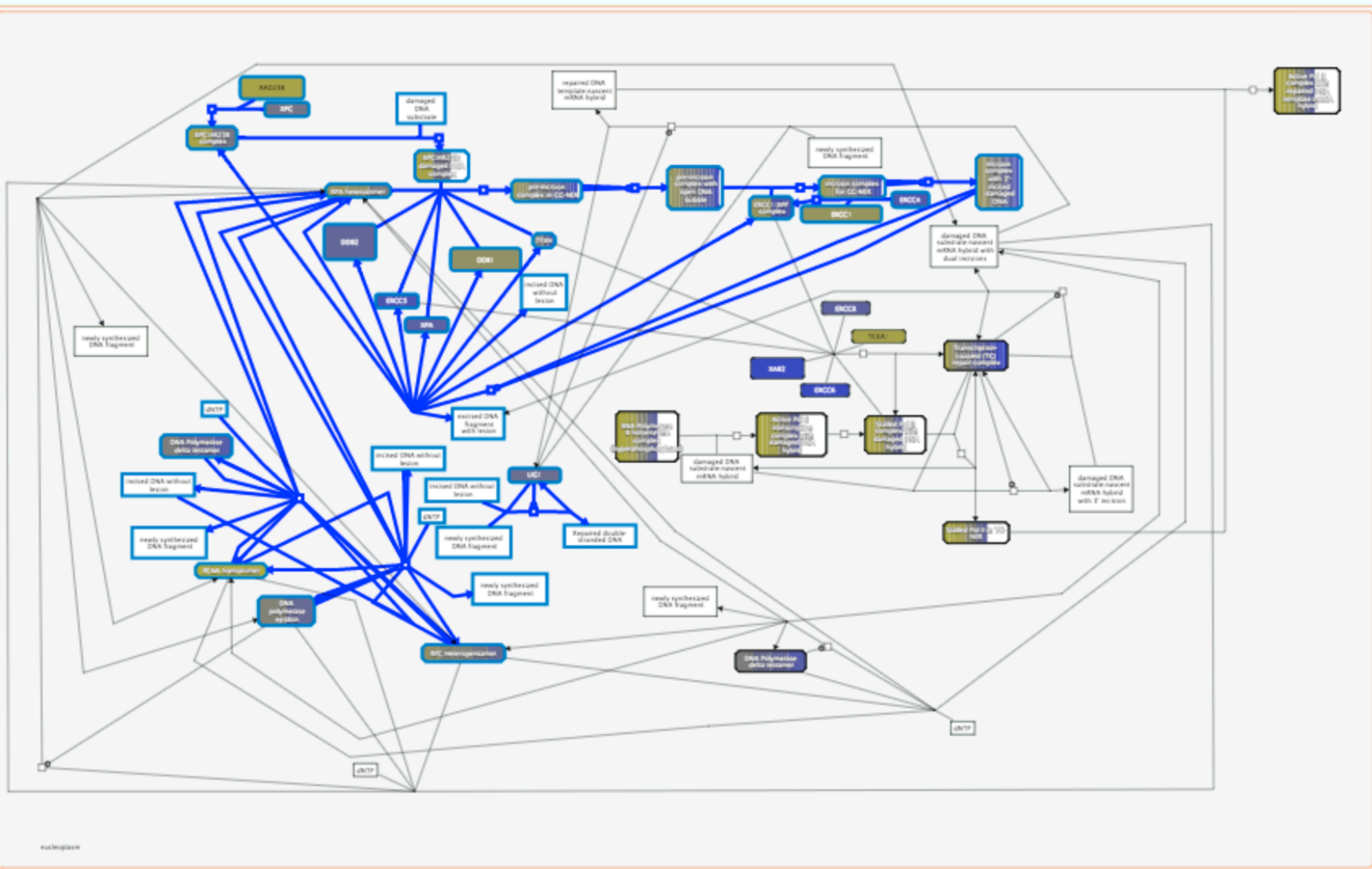
Some examples: **1c**

Analyze Data Results

Overview	Molecules	Structures	Expression	Analysis (973)	Processes	Downloads		
Results for: UNIPROT (973) ▾		Type: Expression		[Data: Probeset]		Results	Identifiers not found: 29	
Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	
Global Genomic NER (GG-NER)	<u>34</u>	34	0.004	1.11E-16	1.33E-15	11	11	
Nucleotide Excision Repair	<u>46</u>	50	0.006	1.11E-16	1.33E-15	21	21	
Transcription-coupled NER (TC-NER)	<u>41</u>	45	0.006	1.11E-16	1.33E-15	10	10	
RNA Polymerase II Transcription Elongation	<u>37</u>	42	0.005	1.11E-16	1.33E-15	8	8	
Formation of HIV elongation complex in the absence of HIV Tat	<u>37</u>	42	0.005	1.11E-16	1.33E-15	2	2	
Formation of RNA Pol II elongation complex	<u>37</u>	42	0.005	1.11E-16	1.33E-15	2	2	
Formation of the HIV-1 Early Elongation Complex	<u>28</u>	32	0.004	1.11E-16	1.33E-15	5	5	
Formation of the Early Elongation Complex	<u>28</u>	32	0.004	1.11E-16	1.33E-15	3	3	
Activation of the pre-replicative complex	<u>26</u>	30	0.004	1.11E-16	1.33E-15	8	8	
HIV Transcription Elongation	<u>36</u>	42	0.005	1.11E-16	1.33E-15	15	15	
Tat-mediated elongation of the HIV-1 transcript	<u>26</u>	42	0.005	1.11E-16	1.33E-15	8	8	
Formation of HIV-1 elongation complex containing HIV-1 Tat	<u>36</u>	42	0.005	1.11E-16	1.33E-15	5	5	
RNA Polymerase II Pre-transcription Events	<u>53</u>	62	0.008	1.11E-16	1.33E-15	17	17	
Transcription of the HIV genome	<u>55</u>	65	0.008	1.11E-16	1.33E-15	45	45	
RNA Polymerase II Transcription Initiation And Promoter Clearance	<u>35</u>	43	0.005	1.11E-16	1.33E-15	9	9	
RNA Polymerase II HIV Promoter Escape	<u>35</u>	43	0.005	1.11E-16	1.33E-15	7	7	
RNA Polymerase II Transcription Pre-Initiation And Promoter Opening	<u>35</u>	43	0.005	1.11E-16	1.33E-15	5	5	
RNA Polymerase II Promoter Escape	<u>35</u>	43	0.005	1.11E-16	1.33E-15	5	5	
HIV Transcription Initiation	<u>35</u>	43	0.005	1.11E-16	1.33E-15	4	4	
RNA Polymerase II Transcription Initiation	<u>35</u>	43	0.005	1.11E-16	1.33E-15	3	3	

Table can be downloaded in various formats

Clicking any of the pathway names will show the respective pathway



Compare species

Species Comparison

This tool allows you to compare human pathways with those in any of the other species inferred from Reactome by orthology.

Use the species selector to choose the other species and click on the "Compare" button to perform the comparison.

Compare Homo sapiens with Select a species...





Compare

This tool will allow you to compare human pathways to any other species pathways they have in their database

Compare species

Overview	Molecules	Structures	Expression	Analysis (1,610)	Processes	Downloads				
Results for: TOTAL (1610) ▾							Type: Species Comparison		Results	
Pathway name				Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total
Signal Transduction				1827	2,372	0.245	3.06E-12	1.16E-8	1,388	1,408
Hemostasis				474	561	0.058	2.06E-7	2.8E-4	299	305
Cell Cycle, Mitotic				425	496	0.051	2.21E-7	2.8E-4	272	272
Signalling by NGF				288	325	0.034	1.72E-6	1.63E-3	170	170
Separation of Sister Chromatids				171	179	0.018	4.66E-6	3.54E-3	8	8
Cell Cycle				470	574	0.059	5.69E-6	3.6E-3	345	345
Mitotic Metaphase and Anaphase				180	194	0.02	1.35E-5	6.87E-3	12	12
Mitotic Anaphase				179	193	0.02	1.45E-5	6.87E-3	11	11
Platelet activation, signaling and aggregation				216	248	0.026	8.47E-5	3.05E-2	106	109
Signaling by Rho GTPases				120	125	0.013	9.21E-5	3.05E-2	5	5
Rho GTPase cycle				120	125	0.013	9.21E-5	3.05E-2	5	5
Assembly of the primary cilium				173	193	0.02	1.08E-4	3.05E-2	50	50
NGF signalling via TRKA from the plasma membrane				208	239	0.025	1.18E-4	3.05E-2	116	116
Cell Cycle Checkpoints				116	121	0.012	1.25E-4	3.05E-2	38	38
M Phase				258	306	0.032	1.37E-4	3.05E-2	63	63
Peptide ligand-binding receptors				173	194	0.02	1.38E-4	3.05E-2	66	68
Axon guidance				298	360	0.037	1.45E-4	3.05E-2	239	239
Downstream signaling events of B Cell Receptor (BCR)				173	195	0.02	1.75E-4	3.49E-2	37	37
Mitotic G1-G1/S phases				129	139	0.014	2.03E-4	3.83E-2	53	53
mRNA Splicing				112	119	0.012	3.12E-4	5.36E-2	14	14

 Result

 Mapping

⏮

⏪

1-20 of 1,610

⏩

⏭

Here, I chose mouse and as before, clicking on the pathway name will open the pathway map, and the table/results can be downloaded.

Browse pathways

Event Hierarchy:

- + Binding and Uptake of Ligands by S
- + Cell Cycle
- + Cell-Cell communication
- + Cellular responses to stress
- + Chromatin organization
- + Circadian Clock
- + Developmental Biology
- + Disease
- + DNA Repair
- + DNA Replication
- + Extracellular matrix organization
- + Gene Expression
- + Hemostasis
- + Immune System
- + Membrane Trafficking
- + Metabolism
- + Metabolism of proteins
- + Muscle contraction
- + Neuronal System
- + Organelle biogenesis and mainter
- + Programmed Cell Death
- + Reproduction
- + Signal Transduction
- + Transmembrane transport of sma

Welcome to the Reactome Pathway Browser, a tool for visualizing and interacting with pathways.

To view a Pathway Diagram either:

1. Click on a pathway name in the hierarchical list on the left
2. Return to the Homepage (button top left) and text search using a pathway, protein or compound name, or an accession number (e.g. Q9HCN6). In the resulting list, click on pathway name to open it in the Pathway Browser.

To see a key to pathway diagrams click the Diagram Key link top right

For a detailed explanation see Section 3 of the [Reactome User Guide](#)

For help please contact us (help@reactome.org)

The Pathway Browser includes tools for several types of analysis, detailed in Section 5 of the [Reactome User Guide](#).

Search for your favorite
pathways by following their
instructions.

Demo

PANTHER



Panther

<http://pantherdb.org/>

- Tools and data on the PANTHER site can be used to:
 - Get information about a gene of interest
 - Explore protein families, molecular functions, biological processes, cellular components and pathways
 - Generate lists of genes that belong to a given protein family or subfamily, have a given molecular function or participate in a given biological process or pathway, e.g. generate a candidate gene list for a disease
 - Analyze lists of genes, proteins or transcripts according to categories based on family, molecular function, biological process, cellular component or pathway, e.g. analyze mRNA microarray data

Now includes comprehensive GO annotations directly imported from the GO database

Search

All 

Quick links

[Whole genome function views](#)

[Genome statistics](#)

[How to cite PANTHER](#)

NEW! [Recent publication describing PANTHER](#)

News

PANTHER gene analysis tools now support comprehensive GO annotations.

[Click](#) for additional info.

Newsletter subscription

Enter your Email:

Gene List Analysis

Browse

Sequence Search

cSNP Scoring

Keyword Search

Please refer to our article in [Nature Protocols](#) for detailed instructions on how to use this page.

Help Tips

Steps:

- ❖ 1. Select list and list type to analyze
- ❖ 2. Select Organism
- ❖ 3. Select operation

1.

Enter ids and or select file for batch upload. Else enter ids or select file or list from workspace for comparing to a reference list.

Enter IDs:
[Supported IDs](#)

separate IDs by a space or comma

Upload IDs:
[File format](#)

No file chosen

Please [login](#) to be able to select lists from your workspace.

Select List Type:

- ☒ ID List
- ☐ Previously exported text search results
- ☐ Workspace list
- ☐ PANTHER Generic Mapping File

2.

Select organism.

Homo sapiens
Mus musculus
Rattus norvegicus
Gallus gallus
Danio rerio

3. Select Analysis.

- ☒ Functional classification viewed in gene list
- ☐ Functional classification viewed in pie chart
- ☐ Statistical overrepresentation test ☐ Use default settings
- ☐ Statistical enrichment test ☐ Use default settings

Upload multiple gene IDs

Gene List Analysis

Browse

Sequence Search

cSNP Scoring

Keyword Search

Please refer to our article in [Nature Protocols](#) for detailed instructions on how to use this page.

Help Tips

Steps:

1. Select list and list type to analyze
2. Select Organism
3. Select operation

1.

Enter ids and or select file for batch upload. Else enter ids or select file or list from workspace for comparing to a reference list.

Enter IDs:

[Supported IDs](#)

1a

separate IDs by a space or comma

Upload IDs:

[File format](#)

Choose File

No file chosen

1b

Please [login](#) to be able to select lists from your workspace.

Select List Type:



ID List



Previously exported text search results



Workspace list



PANTHER Generic Mapping File

2.

Select organism.

Homo sapiens
Mus musculus
Rattus norvegicus
Gallus gallus
Danio rerio

2

3. Select Analysis.



Functional classification viewed in gene list

3



Functional classification viewed in pie chart



Statistical overrepresentation test



Use default settings



Statistical enrichment test



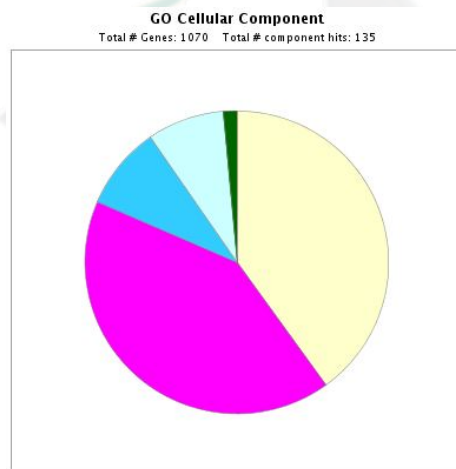
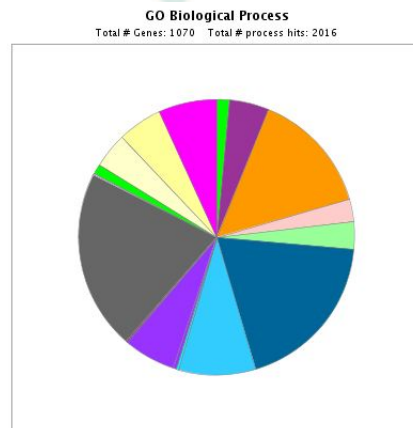
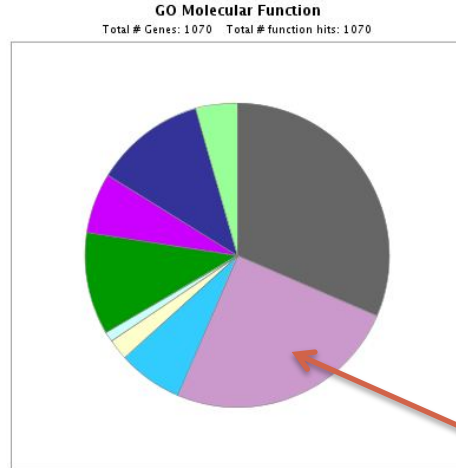
Use default settings

submit

4

1. Upload data to Panther:
 - a. Paste your IDs, or
 - b. Choose a file (tab-delimited)
2. Select Genome(s)
3. Select Analysis type
4. Click "submit"

- Table listing all the genes in your dataset (scroll to the right to see all the columns)
- Table can be downloaded by the “Send list to:” dropdown box
- Click the pie chart icon to view the GO categories in pie charts



- Pie charts of the 3 GO categories:
 - Molecular Function
 - Biological Process
 - Cellular Component
- Each “wedge” can be clicked on to drill further down into the category
- Clicking on any category name link will list a table of molecules from your dataset found in that particular category

Demo

DAVID



DAVID

<http://david.abcc.ncifcrf.gov/>

- **Database for Annotation, Visualization and Integrated Discovery (DAVID)**
- Provides a comprehensive set of functional annotation tools to better understand the biological meaning behind large datasets.
- Tools:
 - GO
 - Pathways
 - Gene-disease associations
 - Protein functional domains and motifs
 - And much, much more!
- [Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources](#). Nature Protocols 4, 44-57 (2009).

Shortcut to DAVID Tools

Functional Annotation

Gene-annotation enrichment analysis, functional annotation clustering , BioCarta & KEGG pathway mapping, gene-disease association, homologue match, ID translation, literature match and [more](#)

Gene Functional Classification

Provide a rapid means to reduce large lists of genes into functionally related groups of genes to help unravel the biological content captured by high throughput technologies. [More](#)

Gene ID Conversion

Convert list of gene ID/accessions to others of your choice with the most comprehensive gene ID mapping repository. The ambiguous accessions in the list can also be determined semi-automatically. [More](#)

Gene Name Batch Viewer

Display gene names for a given gene list; Search functionally related genes within your list or not in your list; Deep links to enriched detailed information. [More](#)

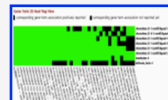
Recommending: A [paper](#) published in *Nature Protocols* describes step-by-step procedure to use DAVID!

Welcome to DAVID 6.7

2003 - 2014

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an [update to the sixth version](#) of our original web-accessible programs. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- ☒ Identify enriched biological themes, particularly GO terms
- ☒ Discover enriched functional-related gene groups
- ☒ Cluster redundant annotation terms
- ☒ Visualize genes on BioCarta & KEGG pathway maps
- ☒ Display related many-genes-to-many-terms on 2-D view.
- ☒ Search for other functionally related genes not in the list
- ☒ List interacting proteins
- ☒ Explore gene names in batch
- ☒ Link gene-disease associations
- ☒ Highlight protein functional domains and motifs
- ☒ Redirect to related literatures
- ☒ Convert gene identifiers from one type to another.
- ☒ And more



Screen Shot 1



Screen Shot 2

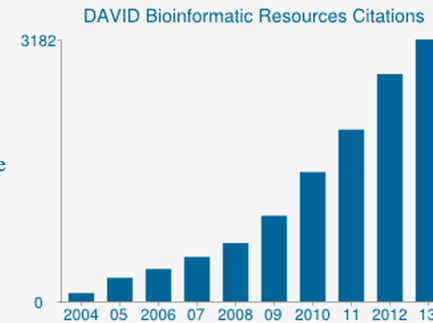


Screen Shot 3

What's Important in DAVID?

- [Current \(v 6.7\) release note](#)
- [New requirement to cite DAVID](#)
- [IDs of Affy Exon and Gene arrays supported](#)
- [Novel Classification Algorithms](#)
- [Pre-built Affymetrix and Illumina backgrounds](#)
- [User's customized gene background](#)
- [Enhanced calculating speed](#)

Statistics of DAVID



- [> 10,000 Citations](#)
- Daily Usage: ~1200 gene lists/sublists from ~400 unique researchers.
- Total Usage: ~800,000 gene lists/sublists from >5,000 research institutes world-wide

Please cite [Nature Protocols 2009; 4\(1\):44](#) & [Nucleic Acids Res. 2009;37\(1\):1](#) within any publication that makes use of any methods inspired by DAVID.

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/>

Thanks! Questions?

Contact info:

David K. Crossman, Ph.D.

Bioinformatics Director

Heflin Center for Genomic Science

University of Alabama at Birmingham

<http://www.heflingenetics.uab.edu>

dkcrossm@uab.edu