

# Data release

Cecilia Klein

ERC project

23/02/2017



# Outline

Goal

Dataset ERC project

Fly

Human

Website examples

ENCODE portal

GRAPE pipeline

Flybase

SCREEN

Genome browser

Interactive plots

Biocore

## Website to release an ERC data:

- ▶ Overview of the data
  - ▶ dashboard matrix including organisms, sample types (RNASeq,ChIPSeq...), tissues and time points
  - ▶ (interactive) plots
  - ▶ genome browser (track hubs)
- ▶ Analysis - plots, tables, download
- ▶ Search tools
  - ▶ by chrmosome regions (e.g. tracks)
  - ▶ by gene expression profile
  - ▶ by expression silimarity
  - ▶ by regulatory elements
  - ▶ links between data types
  - ▶ ...

# Outline

Goal

Dataset ERC project

Fly

Human

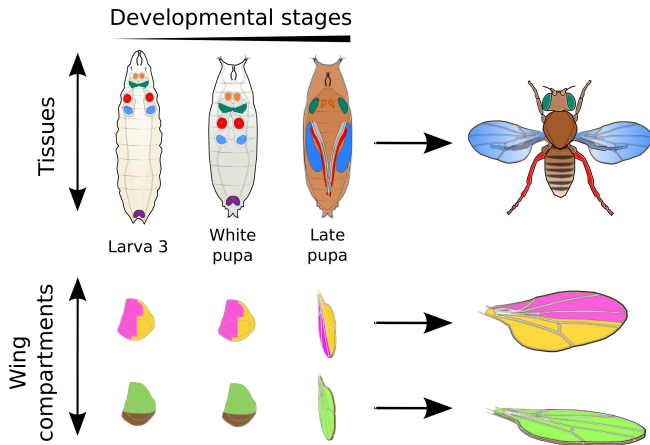
Website examples

Genome browser

Interactive plots

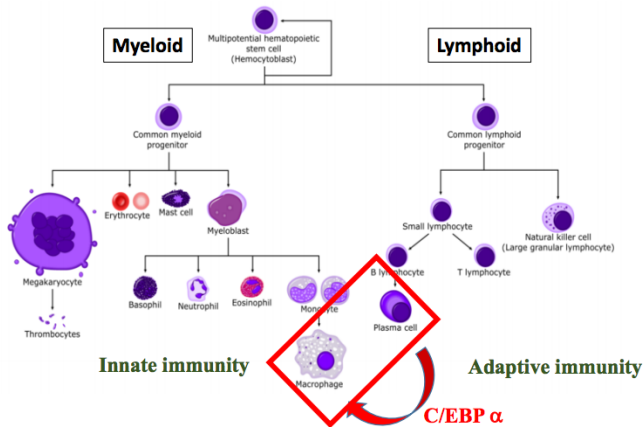
Biocore

# Fly development



# Transdifferentiation model - human

C/EBP  $\alpha$  induces transdifferentiation of B cells into macrophages



# Dataset

Datatype	Organism	Number of metadata attributes	Time points	Tissues/Fractions	Histone marks	Replicates	Total number of samples
RNASeq	Human	18	12	3	3	2	72
ChIPSeq	Human	25	12	-	9	2	216
Riboprofiling	Human	24	12	-	-	2	24
Proteomics	Human	9	12	-	-	2	24
RNASeq	Fly	33	3	14	-	2	84
ChIPSeq	Fly	33	3	4	9	2	216

File types: bam, bigwig, bed, quantifications

# Outline

Goal

Dataset ERC project

**Website examples**

ENCODE portal

GRAPE pipeline

Flybase

SCREEN

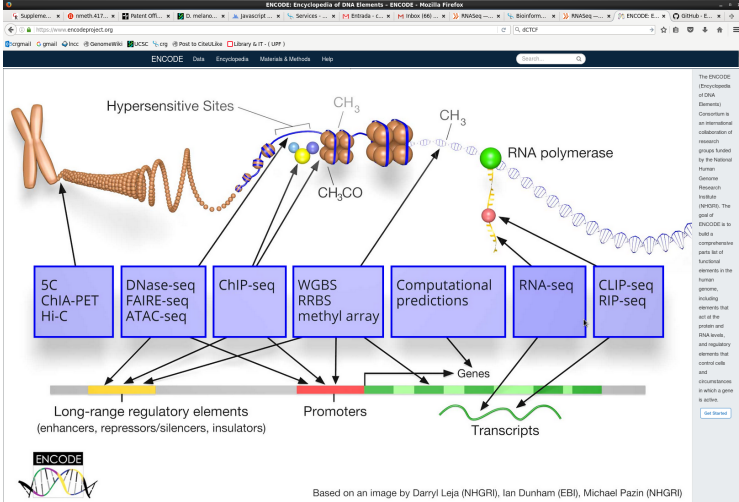
Genome browser

Interactive plots

Biocore

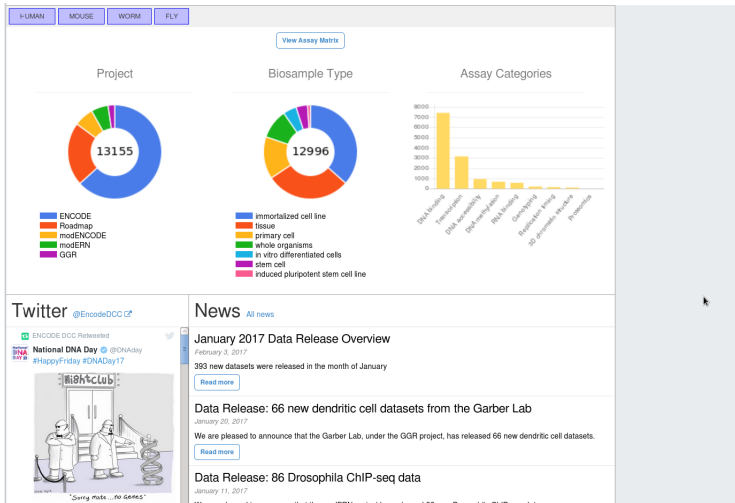


# ENCODE Portal



ENCODE portal

# ENCODE Portal



## Experiment Matrix

Click or enter search terms to filter the experiments included in the matrix.

### Organism

Homo sapiens	9825
Mus musculus	1710
Drosophila melanogaster	986
Caenorhabditis elegans	647
Drosophila pseudoobscura	10

[+ See more...](#)

### Biosample type

immortalized cell line	4839
tissue	3815
primary cell	1923
whole organisms	1324
in vitro differentiated cells	626

[+ See more...](#)

### Organ

brain	649
skin of body	367
muscle organ	339
heart	288
lung	279

[+ See more...](#)

### Project

ENCODE	8467
Roadmap	2936
modENCODE	803
modERN	775
GGR	299

### Genome assembly (visualization)

hg19	5063
GRCh38	4280
mm10	1413
dm3	602
dm6	496

[+ See more...](#)

Audit category: ▲

Assay	Assay category	Target of assay	Date released	Available data
ChIP-seq 7564	DNA binding 7564	histone 3119	July, 2013 2934	fastq 11010
DNase-seq 811	Transcription 3142	histone modification 3119	March, 2014 887	bam 9236
polyA mRNA RNA-seq 704	DNA accessibility 963	transcription factor 2983	July, 2016 614	bigWig 8445
shRNA RNA-seq 526	DNA methylation 687	control 2137	May, 2016 569	bed narrowPeak 5020
RNA-seq 498	RNA binding 593	broad histone mark 1758	October, 2016 485	bigBed narrowPeak 4478

[+ See more...](#) [+ See more...](#) [+ See more...](#) [+ See more...](#) [+ See more...](#)

BIOSAMPLE	ASSAY												...and 23 more					
	ChIP-seq	DNase-seq	polyA mRNA RNA-seq	shRNA RNA-seq	scRNA-seq	RNA microarray	WGBS	small RNA-seq	microRNA-seq	RNA bind-n-seq	RNAi-seq	CRISPRi		CRISPRa	single cell RNA-seq	bulk RNA-seq	Hi-C	
<b>13360 results</b>																		
<b>Immortalized cell line</b>																		
K562	560	7	18	272	13	178	12	1	1	8	1	1	2	9	1	50	2	6
HepG2	290	3	11	254	6	140	7	1	2	3	2	6	1	2	6	2	6	
A549	330	14	21			2	1	9			2	3	1					
GM12878	216	2	10	7	8	2	2	6	1	1	2	6	1	13	2	6		
HEK293	200					1	2				2				1			
...and 177 more																		
<b>tissue</b>																		
liver	155	5	14	11		10	1	1	6	2	3	7	2	1				
heart	100	21	8	11	10	9	1	1	7	2		8	1					
stomach	94	19	11	9	8	1	4	4	5	1	4							
lung	80	16	8	5	10	7	3	1	4	1		4		2				
kidney	69	16	9	4	2	4	4	4			4			2				
...and 154 more																		
<b>primary cell</b>																		
IMR-90	67	3	3	2	1	3	1	9	1		3	3		1	6			
common myeloid progenitor, CD34-positive	67	15	1			13	1	9										
CD4-positive helper T cell	77	8	3			1	3	1	1		1							
foreskin fibroblast	31	6	4			3	1	2	1	3	2	1		2	6			
endothelial cell of umbilical vein	36	2	5			2		1		1	5	1		1	6			
...and 137 more																		
<b>whole organisms</b>																		
multi-cellular organism	126		73	50										15				
carcass	12		4											4				
<b>In vitro differentiated cells</b>																		
mesenchymal stem cell	61	2	4					2	2									
dendritic cell	11			25														
mesendoderm	29	2	4					2										
trophoblast cell	36	2																
chondrocyte	32								3									
...and 40 more																		

### Assay category

DNA binding	7564
Transcription	3142
DNA accessibility	963
DNA methylation	687
RNA binding	593

[+ See more...](#)

### Assay

ChIP-seq	7564
DNase-seq	811
polyA mRNA RNA-seq	704
shRNA RNA-seq	526
RNA-seq	496

[+ See more...](#)

### Project

ENCODE	8467
Roadmap	2936
modENCODE	883
modERN	775
GGR	299

### RFA

ENCODE3	5106
Roadmap	2936
ENCODE2	2773
modENCODE	883
modERN	775

[+ See more...](#)

### Experiment status

released	13155
archived	126
revoked	79

### Genome assembly (visualization)

hg19	5063
GRCh38	4280
mm10	1413
dm3	602
dm6	496

[+ See more...](#)

### Showing results 1 to 25 of 13360


 Columns

ID	Accession	Assay Type	Assay Nickname	Target label	Target gene	Biosample summary	Biosample	Description
<a href="#">ENCSR697YIN</a>	ENCSR697YIN	ChIP-seq	ChIP-seq	CTCF	CTCF	breast epithelium male adult (54 years)	breast epithelium	CTCF ChIP-seq or human breast epithelium
<a href="#">ENCSR96YXU</a>	ENCSR96YXU	ChIP-seq	ChIP-seq	Control		breast epithelium male adult (54 years)	breast epithelium	ChIP-seq or human breast epithelium
<a href="#">ENCSR699BEK</a>	ENCSR699BEK	ChIP-seq	ChIP-seq	CTCF	CTCF	tibial artery male adult (37 years)	tibial artery	CTCF ChIP-seq or human tibia artery
<a href="#">ENCSR405DOJ</a>	ENCSR405DOJ	ChIP-seq	ChIP-seq	Control		tibial artery male adult (37 years)	tibial artery	ChIP-seq or human tibia artery
<a href="#">ENCSR066GBX</a>	ENCSR066GBX	ChIP-seq	ChIP-seq	CTCF	CTCF	right atrium auricular region female adult (53 years)	right atrium auricular region	CTCF ChIP-seq or human right atrium auricular region
<a href="#">ENCSR457JBY</a>	ENCSR457JBY	ChIP-seq	ChIP-seq	Control		right atrium auricular region female adult (53 years)	right atrium auricular region	ChIP-seq or human right atrium auricular region
<a href="#">ENCSR821YOI</a>	ENCSR821YOI	ChIP-seq	ChIP-seq	Control		HEK293T	HEK293T	ChIP-seq or human HEK293T
<a href="#">ENCSR305PNK</a>	ENCSR305PNK	ChIP-seq	ChIP-seq	Control		HCT116	HCT116	ChIP-seq or human

ENCODE [Data](#) [Encyclopedia](#) [Materials & Methods](#) [Help](#)

## Region search

Enter any one of human Gene name, Symbol, Synonyms, Gene ID, HGNC ID, coordinates, rsid, Ensemble ID

hg19 ▾

No annotations found Search

# Outline

Goal

Dataset ERC project

## Website examples

ENCODE portal

**GRAPE pipeline**

Flybase

SCREEN

Genome browser

Interactive plots

Biocore

# GRAPE pipeline



## RNASeq Analysis of transcriptome data

Track what's new in your expression studies.

### Encyclopedia of DNA Elements

#### **ENCODE**

Functional elements in the human genome

### Human

#### **Breast**

Breast Cancer

#### **Down**

Down Syndrome

#### **CLL**

Chronic Lymphocytic Leukemia

#### **HBM**

Human Body Map

### Model Organisms

#### **Dm3**

Drosophila Melanogaster

© 2010 - 2012 CRG. All rights reserved.

<http://rnaseq.crg.cat/>

# GRAPE pipeline

 RNASeq Analysis of transcriptome data

Projects ▾ Project: ENCODE

## Experiment: CALTECH-GM12878-CELL-LONGPOLYA-GENCODEv3c-75-1

Replicates

WebGM

Experiments Overview Read Mapping Expression Splicing Discovery

EXPERIMENTS [View](#) [Edit](#)

Experiment	# Replicates	Lab	Cell type	Localization	RNA extract	Annotation Version	Read length	Paired
<a href="#">CALTECH-GM12878-CELL-LONGPOLYA-GENCODEv3c-75-1</a>	1	CALTECH	GM12878	CELL	LONGPOLYA	GENCODEv3c	75	1

RNA DASHBOARD RESULTS [View](#) [Edit](#)

Please observe the guidelines of the ["ENCODE Consortium Data Release Policy Summary"](#) when using ENCODE data.

The end of the moratorium is indicated in the first column of the table. In case the moratorium has already ended, the table cells are empty.

Consult the [ENCODE RNA Dashboard](#) page for a summary of transcriptome data production in the ENCODE project.

Restricted until	File Type	File View	File Lab	File URL	File Size	File *
	BAM	Alignments	CALTECH	<a href="http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75AlignRep1V2.bam">http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75AlignRep1V2.bam</a>	994M	
	BIGWIG	Raw signal, (-) strand	CALTECH	<a href="http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75Th1014MinusRawRep1V4.bigWig">http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75Th1014MinusRawRep1V4.bigWig</a>	31M	
	BIGWIG	Raw signal, (+) strand	CALTECH	<a href="http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75Th1014PlusRawRep1V4.bigWig">http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75Th1014PlusRawRep1V4.bigWig</a>	31M	
	FASTQ	Raw data	CALTECH	<a href="http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75FastRep1.fastq.gz">http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R1x75FastRep1.fastq.gz</a>	1.5G	
	BAM	Splice sites	CALTECH	<a href="http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqGm12878R1x75SplicesRep1V2.bam">http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqGm12878R1x75SplicesRep1V2.bam</a>	197M	
	BAM	Alignments	CALTECH	<a href="http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R2x75I200AlignRep1V2.bam">http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R2x75I200AlignRep1V2.bam</a>	6.0G	
	FASTQ	Fastq read1	CALTECH	<a href="http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R2x75I200FastR1Rep1.fastq.gz">http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/AwEncodeCaltechRnaSeqReleaseLatest/AwEncodeCaltechRnaSeqGm12878R2x75I200FastR1Rep1.fastq.gz</a>	4.8G	



# GRAPE pipeline

 RNASEq Analysis of transcriptome data

Projects > Project: ENCODE

## Experiment: CALTECH-GM12878-CELL-LONGPOLYA-GENCODEv3c-75-1

Replicates

WGS/SM

Experiments Overview **Read** Mapping Expression Splicing Discovery

### READ SUMMARY

Overview of the reads that were mapped in this experiment.

- Unique reads: Number of unique sequences in the dataset, this reflects the redundancy of the data.
- Unambiguous reads: Reads for which every base was resolved by the sequencer.
- Ambiguous reads: Reads for which at least one nucleotide was not resolved by the sequencer and was assigned an N.

One lane corresponds to one read set in single-end data.  
One lane corresponds to two read sets in paired-end data.

Average over 4 sets of reads	Total	Percent
Unique Reads	66,128,639	62.3%
Unambiguous Reads	99,796,370	94.0%
Ambiguous Reads	6,023,954	5.7%

### MAPPING SUMMARY

Overview of the results after mapping.

- Uniquely mapped reads: Reads for which we find a unique alignment that is better (with a lower number of mismatches) than any of the other possible alignments.
- Multi-mapped reads: Reads for which we find 2 or more equally good best alignments.
- Unmapped reads: Read for which we were unable to find an alignment within the parameters used in the mapping as specified in mapping information.

One lane corresponds to one read set in single-end data.  
One lane corresponds to two read sets in paired-end data.

Average over 4 sets of reads	Total	Percent
Uniquely Mapped Reads	61,090,503	57.6%
Multi-Mapped Reads	21,705,149	20.5%
Unmapped Reads	23,317,238	22.0%

### EXPRESSION SUMMARY

Overview of the annotated features detected in this experiment.

The detection threshold is set to a minimum of one read completely included in the feature.

Average over 1 replicates	Total	Detected	Percent
Genes	46,985	28,449	60.5%
Transcript	131,107	70,656	53.9%
Exons	418,332	242,844	58.1%

### SPLICING SUMMARY

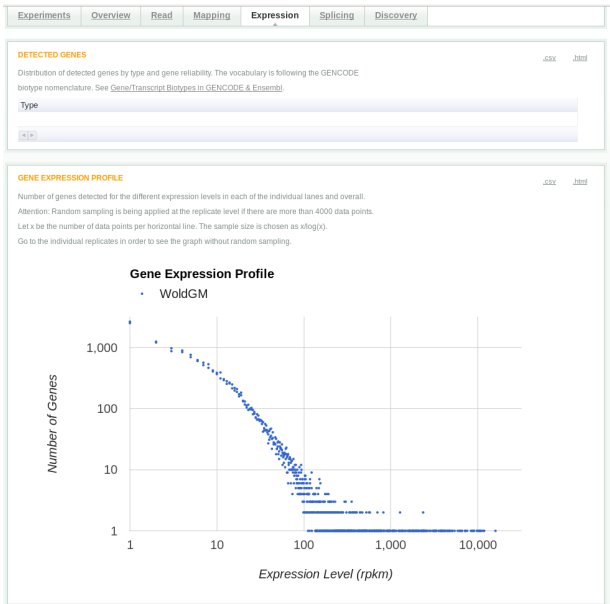
Overview of the splicing events that were detected.

- Known junctions: Junctions present in the annotation as part of an existing transcript.
- Novel junctions from Annotated exons: Junctions formed by two annotated exons from the same gene that are part of any annotated transcript.
- Novel junctions from unannotated exons: Junctions where at least one of the exons involved in the event is not present in the annotation.

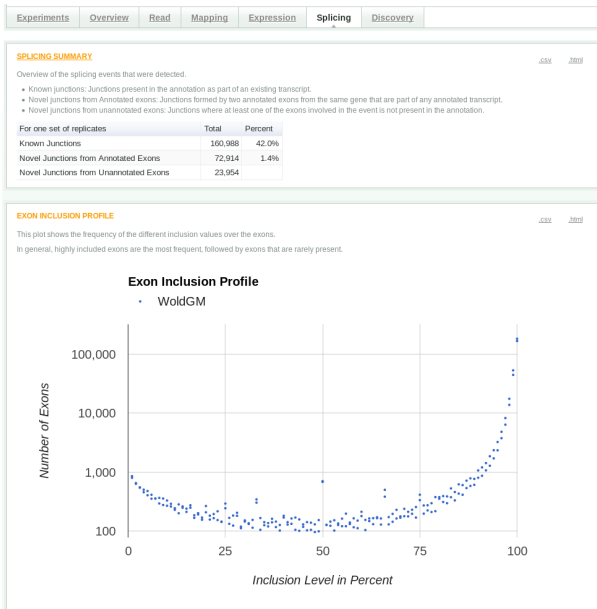
For one set of replicates	Total	Percent
Known Junctions	160,988	42.0%
Novel Junctions from Annotated Exons	72,914	1.4%
Novel Junctions from Unannotated Exons	23,954	



# GRAPE pipeline



# GRAPE pipeline



# Outline

Goal

Dataset ERC project

## Website examples

ENCODE portal

GRAPE pipeline

**Flybase**

SCREEN

Genome browser

Interactive plots

Biocore

FB2017\_01, released February 14, 2017

## FlyBase

A Database of *Drosophila* Genes & Genomes

Home Tools Downloads Links Community Species About Help Archives [Jump to Gene](#) Go

**BLAST** **GBrowse** **Resources** **RNA-Seq** **Vocabularies** **ImageBrowse** **Batch Download**

**FAST-TRACK YOUR PAPER** **FLY BOARD**  
**NEWS** **COMMUNITY**  
**MEETINGS** **COURSES**  
**COMMUNITY ADVOCACY**

### QuickSearch

Human Disease Expression GO Phenotype References  
Simple Orthologs Protein Domains Gene Groups Data Class

Author(s) : e.g. 'Smith NOT Johnson'

Year(s) : e.g. '2004-2008', '>2001', etc.

Title/Abstract : search Title and Abstract text

Journal : e.g. 'Dev. Biol.'

QuickSearch autocomplete: Note: Wild cards (\*) can be added to your search term

### Commentary

[See all commentaries](#)


#### Enhanced Orthology data in FlyBase

Mar 30, 2016 As of the FB2016\_02 release, FlyBase has incorporated orthology data from the DRSC Integrative Orthology Prediction Tool (DIOPT). The DIOPT dataset integrates orthology predictions for eight model organisms from ten individual tools. This approach provides a streamlined method for comparison of orthology predictions originating from different algorithms based on sequence homology, phylogenetic trees, and functional similarity... (More)

FlyBase is supported by a grant from the National Human Genome Research Institute at the U.S. National Institutes of Health U41HG000739. Support is also provided by the British Medical Research Council, the Indiana Genomics Initiative, and the National Science Foundation through XSEDE resources provided by Indiana University. Copyright Statement.

version FB2017\_01, released February 14, 2017  
[Contact FlyBase](#) [Cite FlyBase](#)

FB2017\_01, released February 14, 2017



## FlyBase

Search RNA-Seq data by expression profile

[Home](#) [Tools](#) [Downloads](#) [Links](#) [Community](#) [Species](#) [About](#) [Help](#) [Archives](#) [Jump to Gene](#) [Go](#)

---

The forms below can be used to query FlyBase records using the modENCODE high-throughput RNA-seq data published in [Graveley et al., 2010](#). Results show genes for which the RNA-seq data match a user-selected expression profile.

A video tutorial for this tool can be viewed [here](#).

**search using several modENCODE expression datasets in conjunction**

Join selections in the following forms for the search:  stage  tissue  treatment  cell line

---

**modENCODE expression by stage data**

\*Expression off\* means a peak expression level<sup>†</sup> not more than  expression

*Guide to modENCODE expression RPM level bins*	
Not/Extremely low	0 - 0
Very low	1 - 3
Low	4 - 10
Moderate	11 - 25
Moderately high	26 - 50
High	51 - 100
Very high	101 - 1000
Extremely high	>1000

\*Gelbart and Estimer, 2013

Expression Off	Stage	Expression On				
<input type="checkbox"/>	<input type="checkbox"/>	embryo, 0-2 hr	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	embryo, 2-4 hr	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	embryo, 4-6 hr	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 6-8 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 8-10 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 10-12 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 12-14 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 14-16 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 16-18 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 18-20 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 20-22 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input checked="" type="checkbox"/>	embryo, 22-24 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	larva, L1	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	larva, L2	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	larva, L3 12 hrs old	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	larva, L3 puff stage 1-2	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	larva, L3 puff stage 3-6	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	larva, L3 puff stage 7-9	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	white prepupae, now	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	white prepupae, 12 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	white prepupae, 24 hr	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	pupa, 2 days post-WPP	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	pupa, 3 days post-WPP	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	pupa, 4 days post-WPP	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	adult male, 1 day	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	adult male, 5 days	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	adult male, 30 days	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	adult female, 1 day	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	adult female, 5 days	<input type="checkbox"/>	<input type="checkbox"/>		
<input type="checkbox"/>	<input type="checkbox"/>	adult female, 30 days	<input type="checkbox"/>	<input type="checkbox"/>		

\*Expression on\* means a peak expression level<sup>†</sup> not less than  expression

FB2007\_01, released February 14, 2017



## Genes

Home
Tools
Downloads
Links
Community
Species
About
Help
Archives
Jump to Gene
Go

215 matches

Convert to


Results Analysis Refinement

HitList Conversion Tools

<input checked="" type="checkbox"/>	#Symbol	Name	Annotation ID	Cytology	Alleles	Stacks	Clones
<input checked="" type="checkbox"/>	1 <a href="#">Acoy3D-2</a>	acy1 Coenzyme A oxidase at 57D distal	CG9709	57E1-57E1	5	4	99
<input checked="" type="checkbox"/>	2 <a href="#">ADCS</a>	Argonaute 3	CG4300	80F-80F1	31	5	104
<input checked="" type="checkbox"/>	3 <a href="#">AG10</a>	Alpha 2 glycoxytransferrase	CG2379	65A9-65A9	4	4	25
<input checked="" type="checkbox"/>	4 <a href="#">Abl2</a>	Amyotrophic lateral sclerosis 2	CG1758	78E3-78E3	9	5	12
<input checked="" type="checkbox"/>	5 <a href="#">ance</a>	aberrant MD neurons and olfactory sensilla	CG10393	30F6-30F6	19	143	2
<input checked="" type="checkbox"/>	6 <a href="#">anad</a>	anastical spindle 2	CG8262	44F3-44F3	18	4	25
<input checked="" type="checkbox"/>	7 <a href="#">Anad33</a>	Anastical 33D	CG8979	18C1-18C1	9	10	126
<input checked="" type="checkbox"/>	8 <a href="#">APC7</a>	Anaphase Promoting Complex subunit 7	CG14444	9C1-6C1	10	8	22
<input checked="" type="checkbox"/>	9 <a href="#">APPBP1</a>	$\beta$ -Amyloid precursor protein binding protein 1	CG7828	65A9-65A9	11	5	12
<input checked="" type="checkbox"/>	10 <a href="#">ars</a>	arrestage	CG11513	63E1-43E1	30	13	27
<input checked="" type="checkbox"/>	11 <a href="#">ATPases</a>	ATP synthase, $\alpha$ subunit-Bio	CG1477	69F12-69F13	4	2	4
<input checked="" type="checkbox"/>	12 <a href="#">Atxn1</a>	Ataxin 1	CG4947	6C3-6C3	4	6	5
<input checked="" type="checkbox"/>	13 <a href="#">aut</a>	autophagy	CG4137	32C1-32C1	42	15	140
<input checked="" type="checkbox"/>	14 <a href="#">bax</a>	bag of marbles	CG10422	99C7-99C8	42	9	15
<input checked="" type="checkbox"/>	15 <a href="#">Bicc</a>	Bicoid C	CG4824	35E2-35E2	26	10	248
<input checked="" type="checkbox"/>	16 <a href="#">Bim</a>	Bloom syndrome helicase	CG8325	98E12-98E17	18	13	24
<input checked="" type="checkbox"/>	17 <a href="#">bra</a>	brachyotect	CG1480	100B5-100B5	5	1	7
<input checked="" type="checkbox"/>	18 <a href="#">Bro</a>	brother	CG7965	65A9-65A9	14	10	7
<input checked="" type="checkbox"/>	19 <a href="#">brs1</a>	brms 1	CG11762	33D3-33D5	37	33	140
<input checked="" type="checkbox"/>	20 <a href="#">Brg1A</a>	Brachyotect-specific gene 25A	CG12269	25A3-25A3	3	2	20
<input checked="" type="checkbox"/>	21 <a href="#">C11.1</a>	C11.1	CG12132	8C4-6C5	7	6	32
<input checked="" type="checkbox"/>	22 <a href="#">CG1289</a>	-	CG1289	83C1-83C1	9	8	25
<input checked="" type="checkbox"/>	23 <a href="#">CG1812</a>	-	CG1812	19D2-19D2	6	5	45
<input checked="" type="checkbox"/>	24 <a href="#">CG2861</a>	-	CG2861	10A9-10A9	5	8	17
<input checked="" type="checkbox"/>	25 <a href="#">CG2120</a>	-	CG2129	7D15-7D16	4	2	8
<input checked="" type="checkbox"/>	26 <a href="#">CG2604</a>	-	CG2604	82F6-82F6	2	1	100
<input checked="" type="checkbox"/>	27 <a href="#">CG2662</a>	-	CG2662	3B2-3B2	3	2	4
<input checked="" type="checkbox"/>	28 <a href="#">CG2894</a>	-	CG2894	3B3-3B3	4	5	24
<input checked="" type="checkbox"/>	29 <a href="#">CG3837</a>	-	CG3832	67B8-67B8	8	6	13
<input checked="" type="checkbox"/>	30 <a href="#">CG3902</a>	-	CG3902	59D7-59D7	8	8	20
<input checked="" type="checkbox"/>	31 <a href="#">CG4238</a>	-	CG4238	23B9-23C1	5	4	9
<input checked="" type="checkbox"/>	32 <a href="#">CG4836</a>	-	CG4854	21D5-21D1	8	8	96
<input checked="" type="checkbox"/>	33 <a href="#">CG4876</a>	-	CG4876	67A3-67A3	15	6	52
<input checked="" type="checkbox"/>	34 <a href="#">CG4986</a>	-	CG4985	6E4-6E4	4	4	123
<input checked="" type="checkbox"/>	35 <a href="#">CG4984</a>	-	CG4884	98B9-98B6	2	2	18
<input checked="" type="checkbox"/>	36 <a href="#">CG4993</a>	-	CG4993	31B5-31B1	3	2	41
<input checked="" type="checkbox"/>	37 <a href="#">CG4991</a>	-	CG4991	15E3-15E3	5	6	119
<input checked="" type="checkbox"/>	38 <a href="#">CG5194</a>	-	CG5194	69F1-69F1	2	1	31



FB2017\_01, released February 14, 2017



## FlyBase

### RNA-Seq Expression Similarity Search

HomeToolsDownloadsLinksCommunitySpeciesAboutHelpArchivesJump to GeneGo

Sample gene:Search

Experiment:

Categories to use:  

- embryos0-2hr
- embryos2-4hr
- embryos4-6hr
- embryos6-8hr
- embryos8-10hr
- embryos10-12hr
- embryos12-14hr
- embryos14-16hr
- embryos16-18hr
- embryos18-20hr
- embryos20-22hr
- embryos22-24hr
- L1larvae
- L2larvae
- L3larvae\_12hr\_post\_molt
- L3larvaeP5\_1-2

This tool finds genes with expression patterns that are similar to that of a given gene. Enter your query gene symbol in the box, and choose to search for genes with similar expression by developmental stage, tissue, treatments, or cell lines. You can also specify a subset of experimental samples (categories) within a set of RNA-Seq expression data. Hold down the shift key to select multiple categories.

For more help with this tool, see [this help page](#).







To watch a video tutorial for this tool, go [here](#).

version FB2017\_01, released February 14, 2017  
[Contact FlyBase](#)    [Cite FlyBase](#)



RNA-Seq similarity search

Dataset: *modENCODE\_Stage5* ( [compared subsets](#) ) [Export hits into genes HiList](#)

Gene	Profile (selected subsets only)	Correlation (%)	Molecular function	Biological process
<i>Gapdh1</i>		100.00	NAD binding NADPH binding glyceraldehyde-3-phosphate dehydrogenase (NAD+) (phosphorylating) activity	glucose metabolic process glycolytic process myoblast fusion oxidation-reduction process somatic muscle development
<i>Pgym70</i>		96.52	phosphoglycerate mutase activity	glycolytic process myoblast fusion regulation of glucose metabolic process somatic muscle development
<i>Pyk</i>		95.62	magnesium ion binding potassium ion binding pyruvate kinase activity	glycolytic process myoblast fusion pyruvate metabolic process regulation of glucose metabolic process response to sucrose somatic muscle development
<i>CG32485</i>		95.49	Rab GTPase binding	
<i>GlyP</i>		95.17	glycogen phosphorylase activity protein hemodimerization activity pyridoxal phosphate binding	defense response to Gram-negative bacterium determination of adult lifespan fight glycogen catabolic process negative regulation of innate immune response positive regulation of glycogen catabolic process
<i>CG1927</i>		94.85		

# Outline

Goal

Dataset ERC project

**Website examples**

ENCODE portal

GRAPE pipeline

Flybase

**SCREEN**

Genome browser

Interactive plots

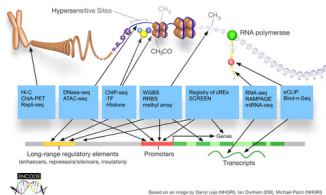
Biocore

## SCREEN: Search Candidate Regulatory Elements by ENCODE

Overview

About

Tutorial



SCREEN is a web interface for searching and visualizing the Registry of candidate Regulatory Elements (cREs) derived from [ENCODE data](#). The Registry contains 2.67M human cREs in hg19 and 1.67M mouse cREs in mm10, with orthologous cREs cross-referenced. SCREEN presents the data that support biochemical activities of the cREs and the expression of nearby genes in specific cell and tissue types.

You may launch SCREEN using the search box below or browse a curated list of SNPs from the NHGRI-EBI GWAS catalog to annotate genetic variants using cREs.

[Browse GWAS](#)

Search hg19


Search mm10

Examples: "K562 chr11:5226493-5403124", "SOX4 TSS", "rs4846913"

© 2017 Weng Lab @ UMass Med, ENCODE Data Analysis Center

<http://screen.umassmed.edu/>

# SCREEN

SCREEN hg19 chr11:18455021-32289180 K562 immortalized cell line Search 

Cell types **K562**

Chromosome **chr11**









































Coordinates **18455021 - 32289180**

KEGG **DNase Z-score** **H3K4me3 Z-score** **H3K27ac Z-score** **CTCF Z-score**

cRE Search Results **cRE Details**

Candidate Regulatory Elements (cREs) that meet your search criteria:



Search:

accession	DNase Z	H3K4me3 Z	H3K27ac Z	CTCF-bound Z	chr	start	length	nearest genes	nearest protein-coding genes	cart	genome browsers
<a href="#">EH37E0374008</a>	3.5	3.6	1.8	0.5	chr11	31,832,686	405	PAX6, FCN1, ZK3307.3	PAX6, FCN1, ELP4	 	 
<a href="#">EH37E0364615</a>	3.5	5.3	3.7	-0.2	chr11	19,262,403	323	RP11-428C19.4, EZF8, CSRP3	EZF8, CSRP3, ZDHHC13	 	 
<a href="#">EH37E0371471</a>	3.4	5.3	4.2	1.2	chr11	28,129,410	721	KIF18A, METTL15, RP11-797J4.1	KIF18A, METTL15, BONF	 	 
<a href="#">EH37E0379751</a>	3.4	5.9	4.1	1.0	chr11	31,530,833	880	IMMP1L, ELP4, AC108456.1	IMMP1L, ELP4, DNAJC24	 	 
<a href="#">EH37E0365662</a>	3.4	0.7	2.8	0.6	chr11	20,113,324	342	NAV2, NAV2-AS1, NAV2-AS2	NAV2, DBX1, HTATIP2	 	 
<a href="#">EH37E0364947</a>	3.3	1.2	3.8	1.3	chr11	19,463,575	640	NAV2, RNASSP35, NAV2-IT1	NAV2, EZF8, CSRP3	 	 
<a href="#">EH37E0365507</a>	3.3	1.0	3.0	0.3	chr11	19,996,496	395	NAV2, NAV2-AS3, NAV2-AS2	NAV2, DBX1, HTATIP2	 	 
<a href="#">EH37E0365648</a>	3.3	2.3	2.6	-0.0	chr11	20,101,629	478	NAV2, NAV2-AS2, NAV2-AS1	NAV2, DBX1, HTATIP2	 	 
<a href="#">EH37E0370844</a>	3.2	1.0	3.2	-0.1	chr11	27,442,258	279	LGR4, RP11-159H22.2, CCDC34	LGR4, CCDC34, LINC7	 	 
<a href="#">EH37E0363939</a>	3.2	6.2	4.3	1.2	chr11	18,548,043	880	TSG101, UEVLD, RP11-813F22.6	TSG101, UEVLD, LDHAL6A	 	 

[< 1](#) [2](#) [3](#) [4](#) [5](#) [...](#) [30 >](#)

[Download bed](#) [Download JSON](#) found 295 results

# SCREEN

SCREEN hg19 chr11:18455021-32389180 K562 immortalized cell line Search  

**Cell types** 1

K562


---

**Chromosome**

chr11

---

**Coordinates** 1



18455021 - 32389180

---

**K562** 1

**DNase Z-score**

1.64 - 10

**H3K4me3 Z-score**

-10 - 10

**H3K27ac Z-score**

-10 - 10

**CTCF Z-score**

-10 - 10

cRE Search Results cRE Details

## EH37E0374008

Top Tissues
Neighb. Genomic Features
TF and Hi-mod Intersection
Associated Gene Expression
Orthologous cREs in mm10
Signal Profile

**H3K4me3 Z-scores**

cell type	H3K4me3 and DNase	H3K4me3 only
neural progenitor cell derived from H9	--	5.24
neuronal stem in vitro differentiated cells H9 derived	--	5.17
neural cell	--	4.81
retinal pigment epithelial primary cell	4.09	4.80
neuron in vitro differentiated cells	--	4.70

**H3K27ac Z-scores**

cell type	H3K27ac and DNase	H3K27ac only
neuronal stem in vitro differentiated cells H1 derived	4.34	4.21
radial glial cell NCNE and derived from H9, stably expressing fuson protein	--	3.16
neuroepithelial stem cell stably expressing fuson protein	--	3.10
neural cell	--	3.09
endocrine pancreas tissue male adult 45 years	--	2.71

**CTCF Z-scores**

cell type	CTCF and DNase	CTCF only
GM19239 immortalized cell line	--	1.77
GM19238 immortalized cell line	--	1.69
GM19240 immortalized cell line	--	1.60
GM12892 immortalized cell line	--	1.58
Fibroblast of upper leg skin primary cell male fetal 12 weeks	1.90	1.44

**DNase Z-scores**

cell type	Z-score
neuronal stem in vitro differentiated cells H1 derived	4.46
CD14-positive monocyte female	4.03
astrocyte of the cerebellum primary cell	4.00
common myeloid progenitor CD34-positive male adult 42 years	3.91
common myeloid progenitor CD34-positive male adult 49 years	3.85

# SCREEN

SCREEN hg19 chr11:1845021-32389180 K562 immortalized cell line

Search

HOME

## Cell types

K562

## Chromosome

chr11

## Coordinates

chr11:1845021-32389180

1845021

32389180

## H502

### DNase Z-score

1.64

10

### HK4me3 Z-score

-10

10

### HK27ac Z-score

-10

10

### CTCF Z-score

-10

10

## cRE Search Results

## cRE Details

EH37E0374008

Top Tissues

Nearby Genomic Features

TF and Hi-mod Intersection

Associated Gene Expression

Orthologous cREs in mm10

Signal Profile

1 2 3 4 5 ... 65 >

Search:

DNase	HK27ac	HK4me3	Tissue of origin	Cell Type	Biosample
149			kidney	immortalized cell line	G401
148.9			blood	immortalized cell line	KBM-7
146.1			eye	tissue	eye female fetal (76 days)
146.1			muscle	immortalized cell line	A673
144.9			eye	tissue	eye fetal (56 days) and male fetal (76 days)
144.9			kidney	primary cell	glomerular visceral epithelial cell chki (3 years)
143.8	▲ 21.47 ▲ 25.19		brain	primary cell	astrocyte
143.1			brain	immortalized cell line	Dasy
142			bone	immortalized cell line	MG63
140.4			spinal cord	tissue	spinal cord female fetal (113 days)

# Outline

Goal

Dataset ERC project

Website examples

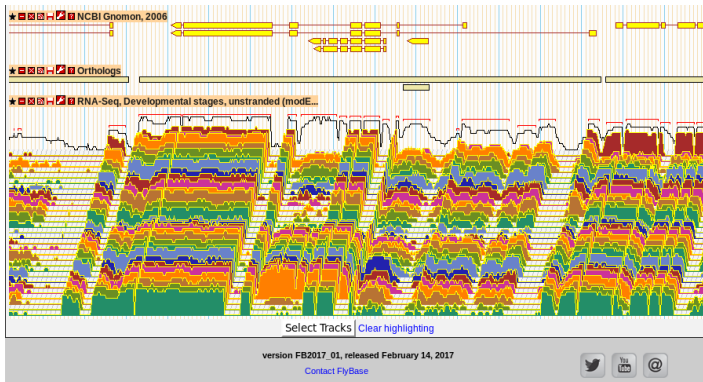
**Genome browser**

Interactive plots

Biocore

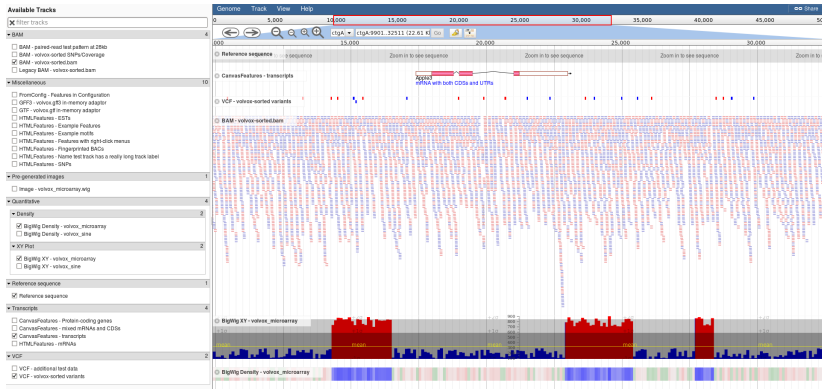


# Gbrowse



FlyBase Gbrowse

# JBrowse



JBrowse demonstration: *Volvox myxocarpus* example

# Outline

Goal

Dataset ERC project

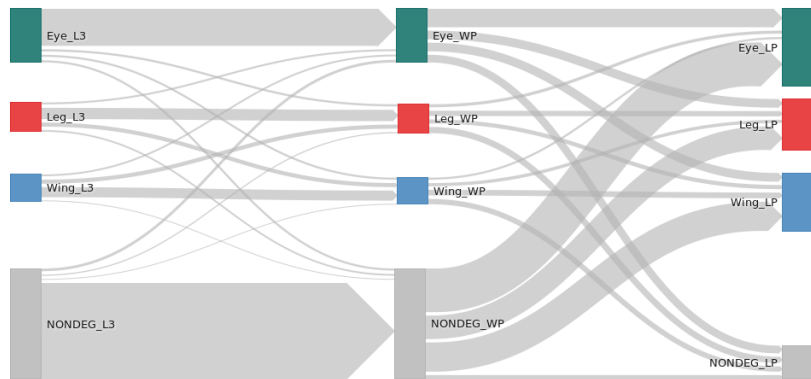
Website examples

Genome browser

**Interactive plots**

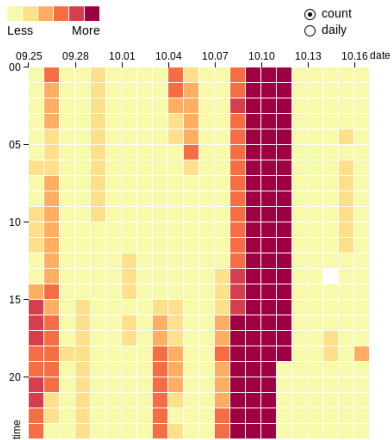
Biocore

# Differentially Expressed Genes



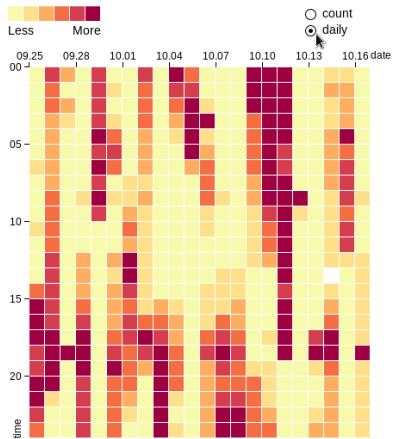
Interactive

# Heatmap



Interactive

# Heatmap



Interactive

- ▶ set up a database and a web server with a dashboard to publish the data of the project
- ▶ install a default release of a genome browser
- ▶ include summary plots produced by us as images