



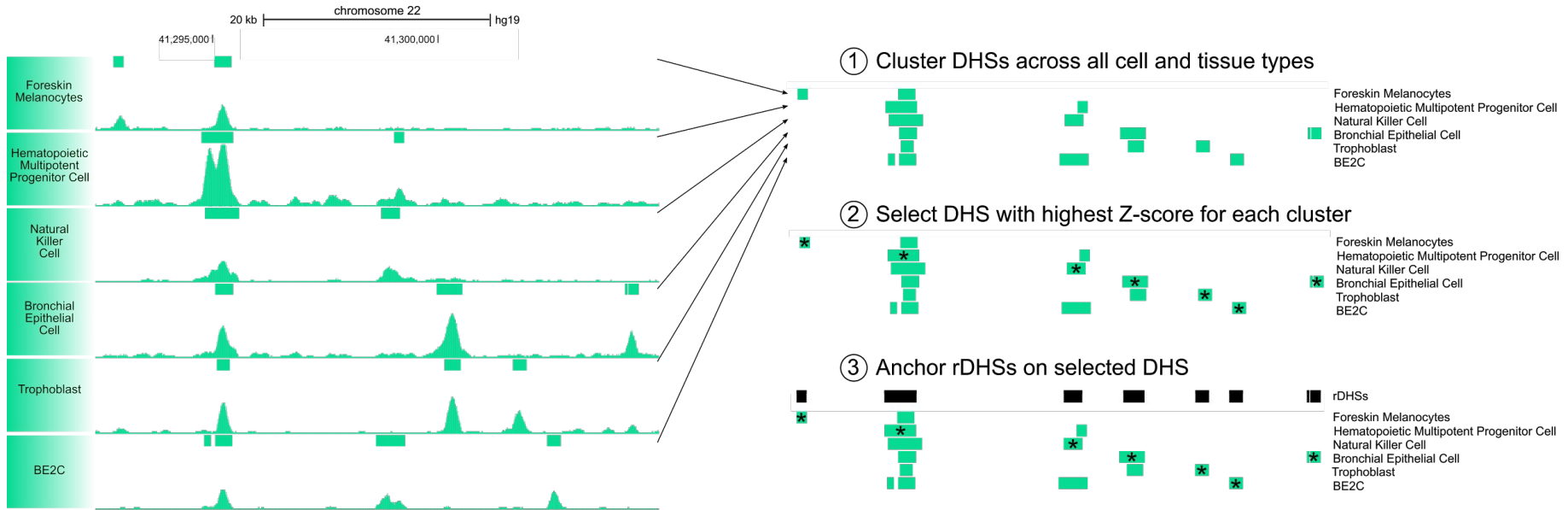
Overview of the Registry of cREs & SCREEN

Jill Moore, Michael Purcaro, Henry Pratt, and Zhiping Weng
University of Massachusetts Medical School

June 14, 2017

Creating the Registry of cREs

1. Defining representative DNase hypersensitivity sites (rDHSs)

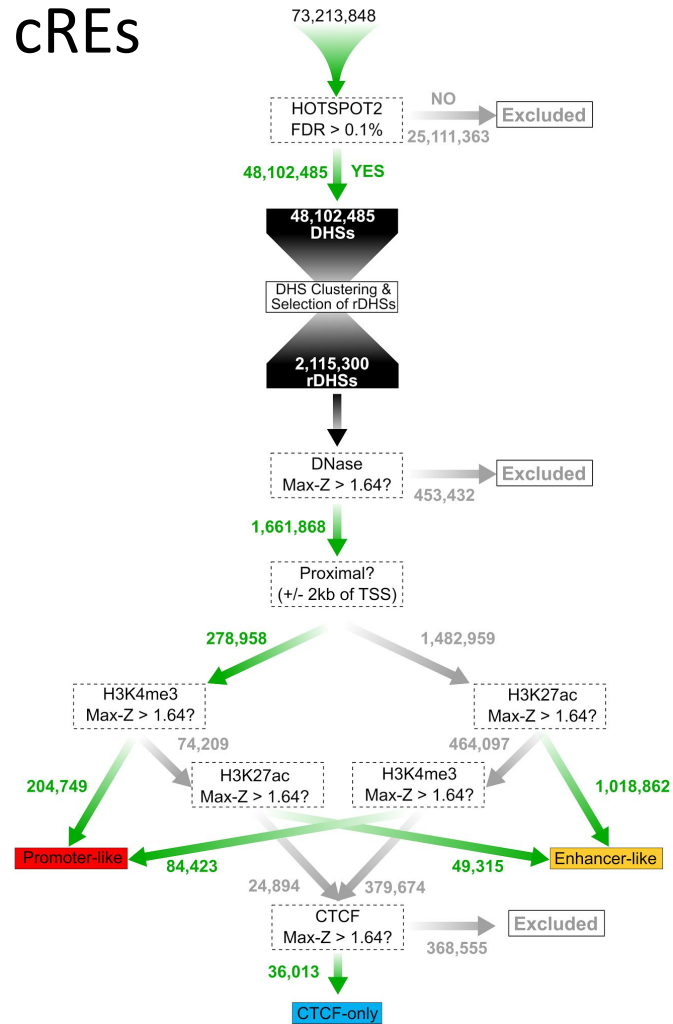


Mouse:

For this analysis, we adapted the DNase master peak script by Bob Thurman in John Stam's lab

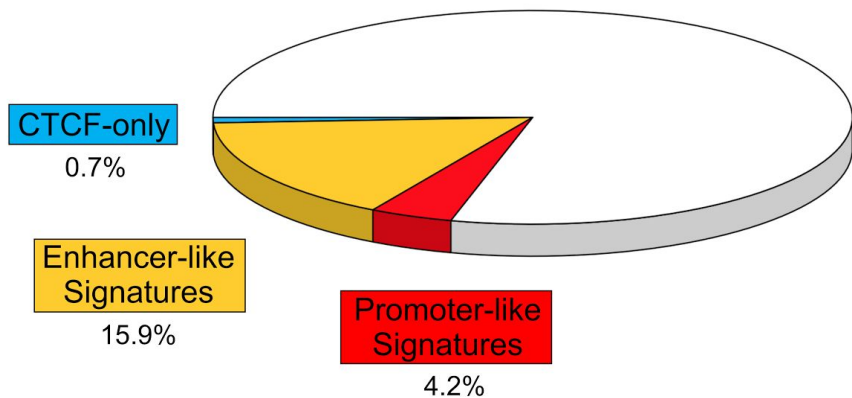
Creating the Registry of cREs

1. Defining representative DNase hypersensitivity sites (rDHSs)
2. Filter & classify rDHSs using DNase, H3K4me3, H3K27ac, and CTCF signals

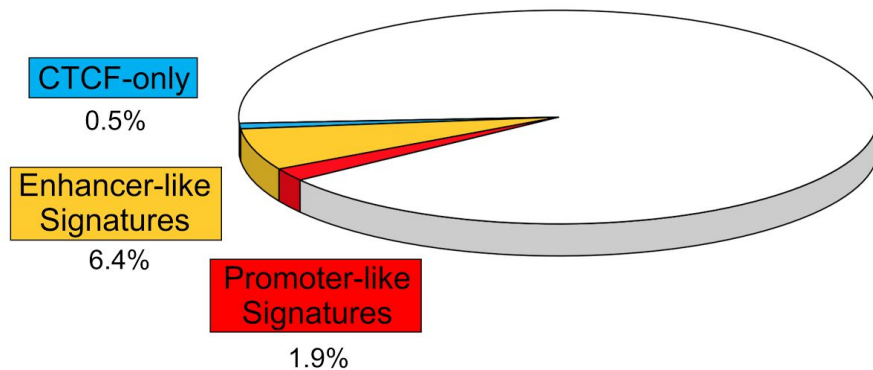


Genomic Coverage of cREs

Human
DNase Mappable Genome
(2.65 Billion Bases)

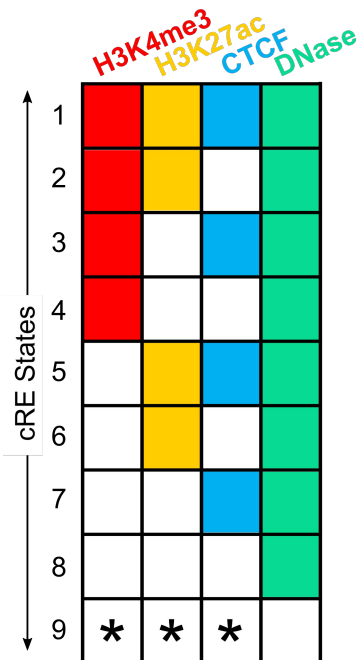


Mouse
DNase Mappable Genome
(2.29 Billion Bases)



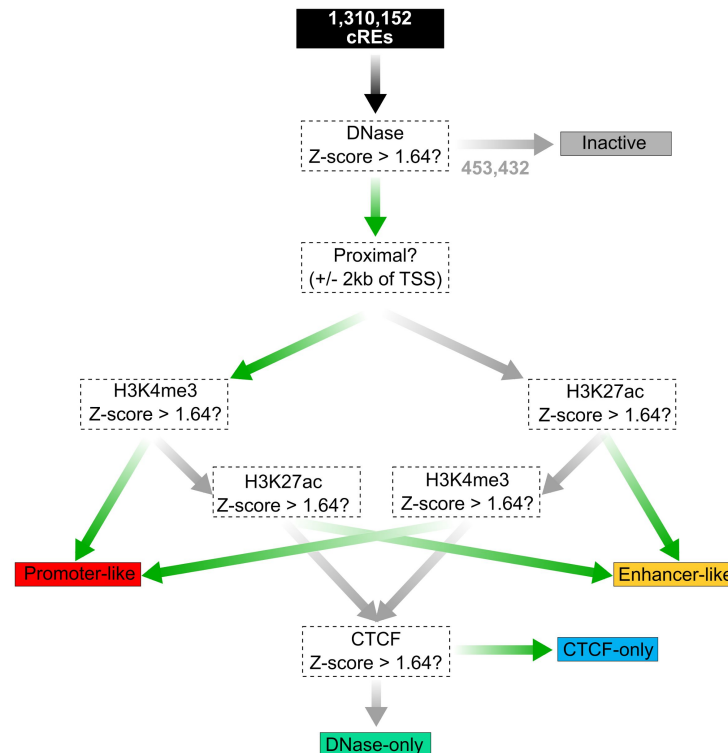
Cell Type Specific Classification: Two Models

9 States



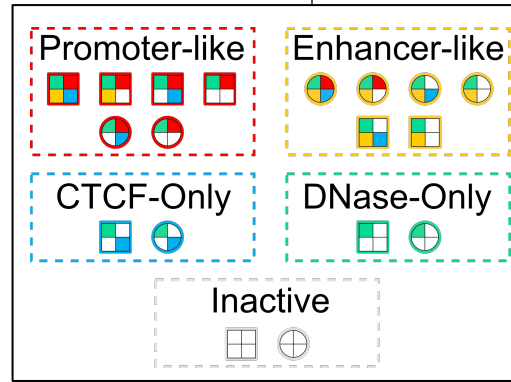
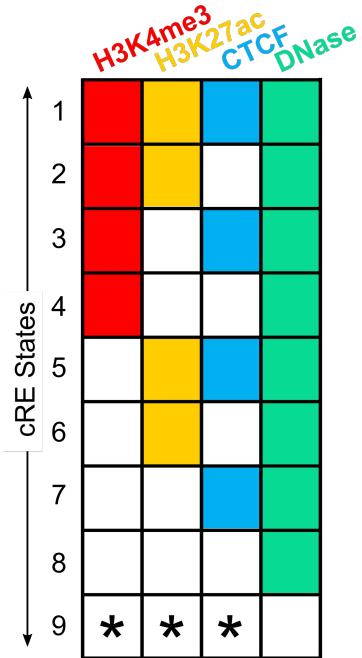
- For each cell type, we classify cREs with DNase Z-score < 1.64 as inactive (if DNase is available)
- The remaining cREs fall into 8 states based on if the Z-scores for H3K4me3, H3K27ac, or CTCF are > 1.64
- For example, a cRE with the grey-yellow-grey-green combination of colors would have H3K27ac & DNase Z-scores > 1.64 and H3K4me3 & CTCF Z-scores \leq 1.64

5 Groups

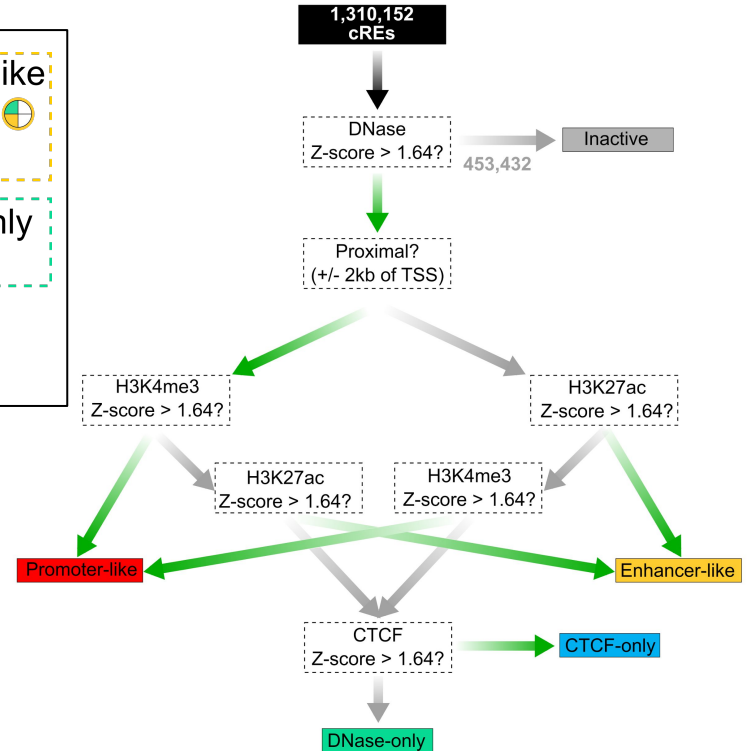


Cell Type Specific Classification: Relating Models

9 States

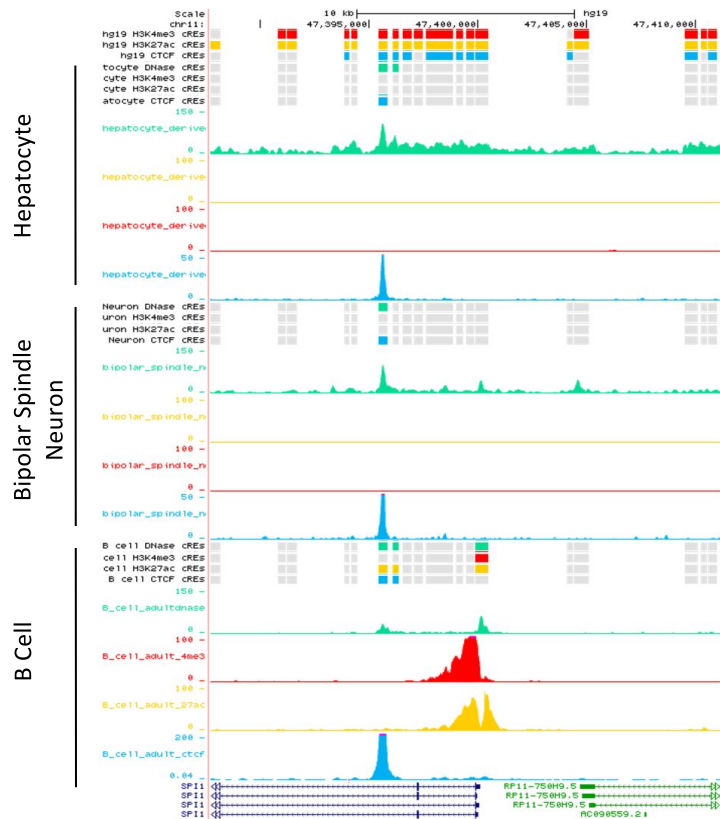


5 Groups



Cell Type Specific Classification: Visualization

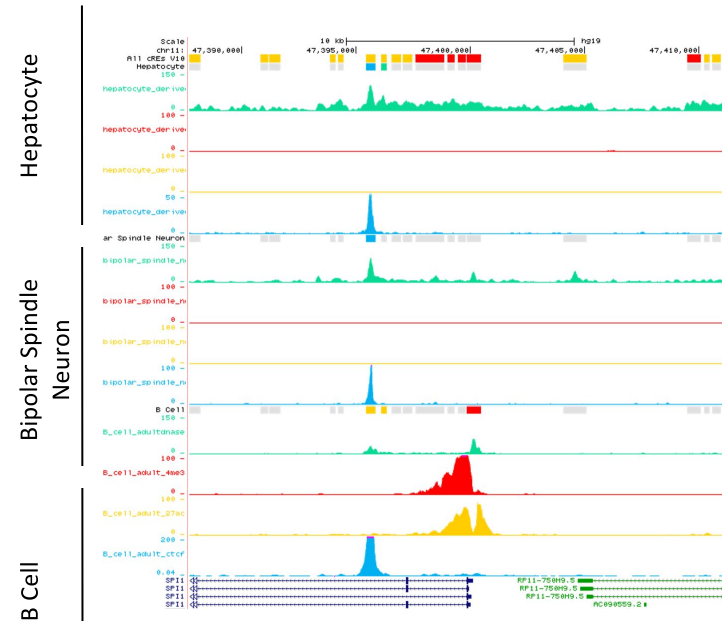
9 States



[Link to browser:](#)

[chr11:47,387,721-47,411,156](#)

5 Groups



[Link to browser:](#)

[chr11:47,387,721-47,411,156](#)

Additional Analysis Using cREs (From Manuscript)

- Enrichment of cREs in complementary data (i.e. cREs with promoter-like signatures are enriched in POL2 signal)
- Cell type clustering
- Use Cases:
 - Correlation of expression & cRE activity in Limb at *Ogn* locus during development
 - Fine mapping multiple sclerosis variant and identifying a new potential target gene
 - Fine mapping schizophrenia variants using orthologous human and mouse cREs

SCREEN overview (recap): three perspectives

cRE / locus-centric Search View

1) enter search keywords
(genes, coordinates, SNPs...)

rs3
rs3
rs30566880
rs30617825
rs31471951
rs32412663
rs32681267
rs367896724
rs369606208
rs372841554
rs274020747

2) use facets to filter search results; click result for details

Cell types **i**

Search: HeLa

cell type **⇅** tissue **⇅**

- HeLa-S3 cervix
- HeLa-S3 G1b phase cervix

< 1 >

Chromosome

chr13

Coordinates: chr13:32444841-32448842 **i**

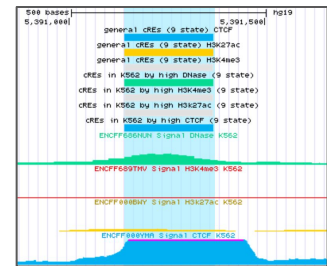
accession i	K562 i	H3K4me3 Z ⇅	H3K27ac Z ⇅	CTCF Z ⇅	chr ⇅	start ⇅	length ⇅	nearest genes: protein-coding / all
EH37E1091086 ★P		3.65	0.36	0.08	chr11	13,298,849	607	pc: ARNTL, BTBD10, PTH all: ARNTL, RN7SKP151, CTC-497E21.5
EH37E1090291 ★P		5.84	4.24	1.32	chr11	6,502,226	750	pc: ARFIP2, TIMM10B, TIMM10B all: ARFIP2, TIMM10B, TIMM10B

Genes within TAD **i**

symbol
AC025300.1
AC084859.1
ARNTL
BTBD10

TFs that bind this cRE **i**

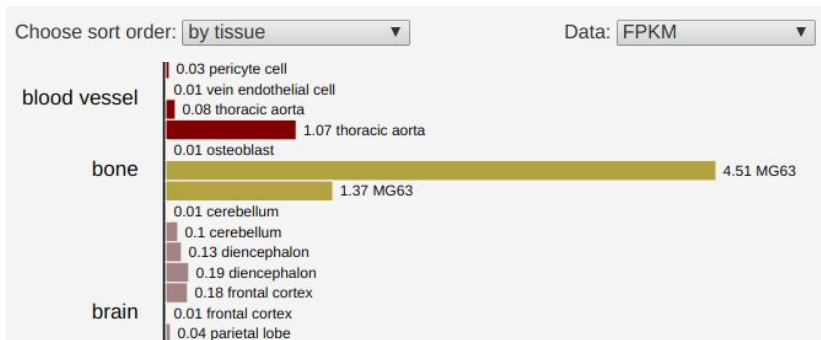
factor	# of experiments that support TF binding
POLR2A	36
EZH2	12
SP1	7
FOS	7



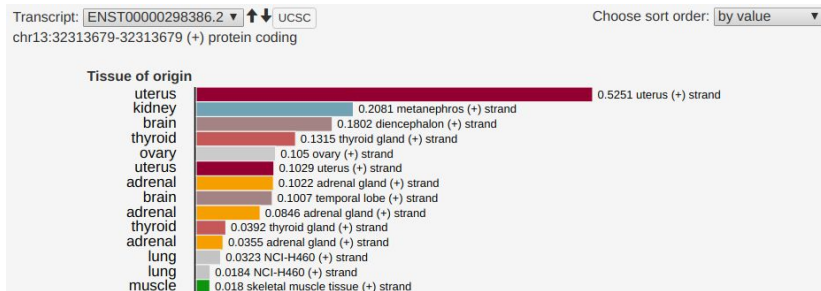
SCREEN overview (recap): three perspectives

Gene-centric View

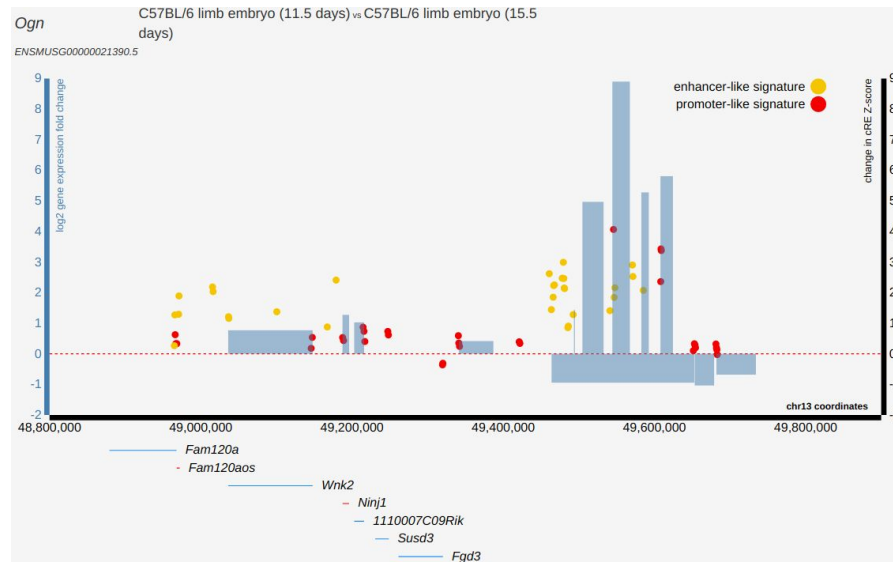
- 1) RNA-seq: human and mouse
gene expression app, cRE details



- 2) RAMPAGE: human only
cRE details



- 3) Differential Expression: mouse only
self-contained app



SCREEN overview (recap): three perspectives

SNP-centric GWAS View

1) select study

GWAS Studies

Studies

CSV

Search:

Study	Author	Pubmed
Inflammatory bowel disease	Jostins	23128233
Inflammatory bowel disease	Liu	26192919
Inflammatory skin disease	Baurecht	25574825

< 1 >

Total LD blocks # of LD blocks overlapping cREs # of overlapping cREs

109	106 (97%)	1205
-----	-----------	------

2) browse cREs intersecting LD blocks from study by cell type

Cell Type	p	FDR
T-cell male adult (37 years)	3.1e-17	4.2e-15
CD4-positive helper T cell	2.2e-16	1.5e-14
CD4-positive helper T cell	2.7e-15	1.2e-13
peripheral blood mononuclear cell	5.2e-15	1.8e-13

accession	T-cell male adult (37 years)	H3K4me3 Z	H3K27ac Z	DNase Z	SNPs	gene
EH37E1089569 ★P		6.81	3.62	1.934	rs11041476	LSP1
EH37E1250115 ★P		6.22	4.54	3.772	rs2617434	LNPEP
EH37E1081006 ★P		6.15	4.37	3.401	rs306587	MAP3K8
EH37E1250117 ★P		5.98	3.41	2.545	rs3842058	CTD-2260A17.2

New since last update: UCSC browser configuration

cRE Search Results Configure Genome Browser

EH37E0204972 chr11:5,391,140-5,391,370 ★ D

Open in UCSC 9 state 5 group

Selected biosamples
Use the handles at left to drag items and change the order in which they will display in the browser.
Note: For best UCSC performance, choose <10 cell types.

- A172
- A549
- A549 treated with dexamethasone

Available biosamples

Search:

cell type	tissue	
<input checked="" type="checkbox"/> A172	brain	<input checked="" type="checkbox"/>
<input checked="" type="checkbox"/> A549	lung	<input type="checkbox"/>
<input checked="" type="checkbox"/> A549 treated with dexamethasone	lung	<input type="checkbox"/>
<input type="checkbox"/> A549 treated with ethanol	lung	<input type="checkbox"/>

tab opens after clicking a UCSC button

toggle 9-state or 5-group cRE tracks

click and drag selected cell types with handles to change order of tracks in the browser

click rows to toggle selected cell types

New since last update: UCSC buttons for transcripts

RNA-seq

OR51B6

UCSC

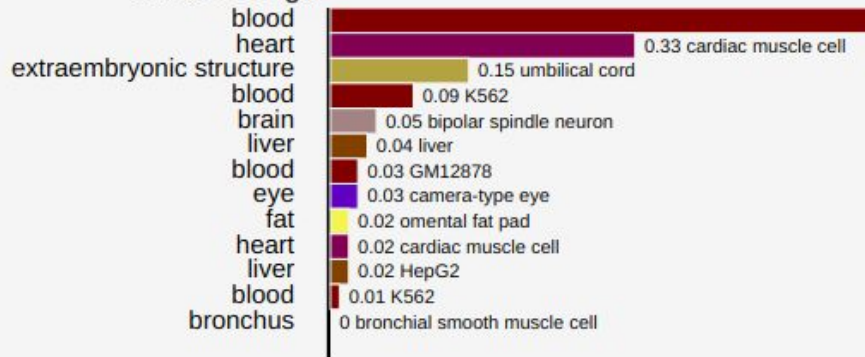
ENSG00000176239.7

This plot is displaying cell-wide expression of *OR51B6*. To view expression in d

Choose sort order:

Data:

Tissue of origin



RAMPAGE

OR51B6

ENSG00000176239.7 (18,402 bases from cRE)

Transcript: UCSC

Choos

chr11:5372738-5372738 (+) protein coding

Tissue of origin



New since last update: summary symbols

Available data types by cell type

A172	brain	
A549	lung	
A549 treated with dexamethasone	lung	
A549 treated with ethanol	lung	
A673	muscle	
ACC112	salivary glands	

search results table: P/D (proximal/distal), concordant support star, Z-scores by data type

EH37E0204972		
EH37E0204974		
EH37E0204971		
EH37E0204940		
EH37E1090146		

High H3K4me3	High H3K27ac	High CTCF	High DNase	Z-score < 1.64	No data
P/D Proximal/Distal to a Transcription Start Site		★ High DNase and High H3K4me3, H3K27ac, or CTCF in the same cell type			

key below search results table

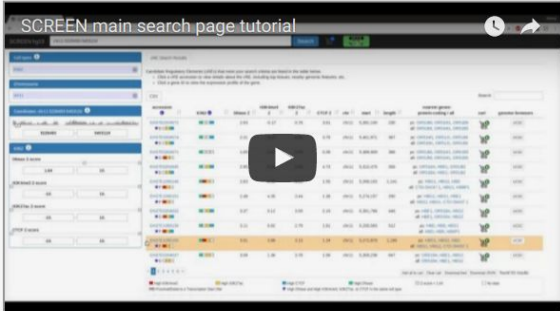
New since last update: new tutorials on homepage

SCREEN: Search Candidate Regulatory Elements by ENCODE

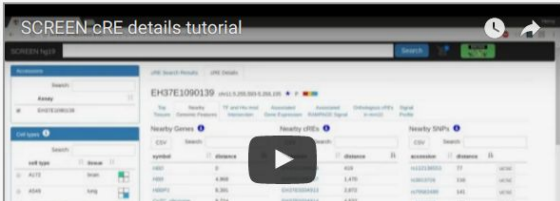
Overview About Tutorial

Tutorials

Main Search



cRE details



- six tutorials
 - main search
 - results table, cart, browser
 - cRE details
 - gene expression
 - differential expression
 - GWAS
- hosted on SCREEN YouTube channel
 - https://www.youtube.com/channel/UCFX89WFjvf-1NCj_4mVObRg