

# Gene Set Analysis for Sequence Data: An Exploration of Tools and Methods

Shaurya Jauhari  
Mora Lab, A2-212.  
Email: [shauryajauhari@gzhmu.edu.cn](mailto:shauryajauhari@gzhmu.edu.cn)  
September 26<sup>th</sup> 2018



# Outline

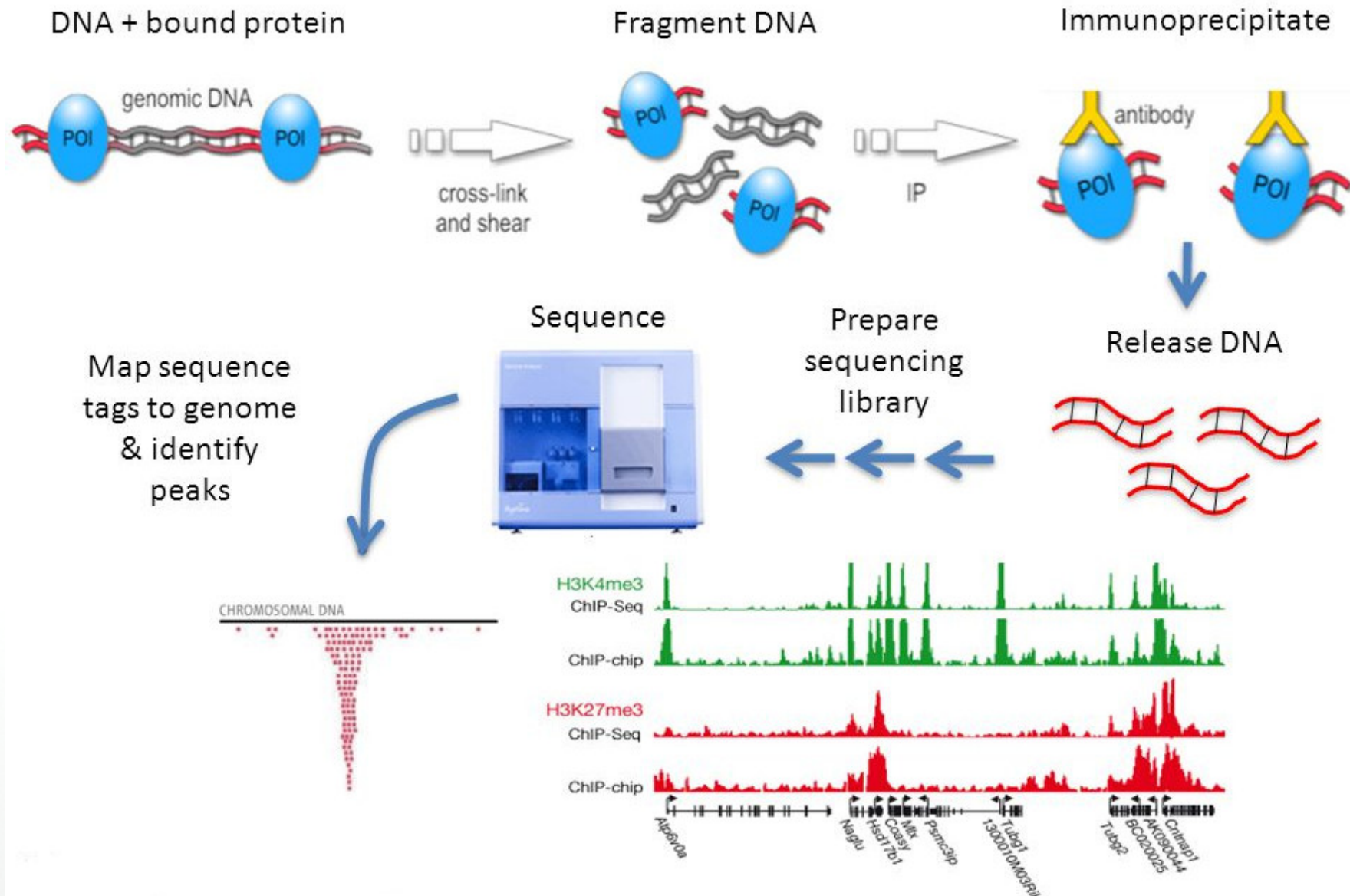
- Technology
  - ChIP Seq
- Tools
  - GREAT
  - ChIP- Enrich
  - Broad- Enrich
  - Seq2Pathway
  - Enrichr



# ChIP Seq

- **Chromatin Immunoprecipitation Sequencing**
- Technology to discern protein binding sites in the DNA
- These proteins *may* contribute to gene regulation
- Gene Regulation
  - Transcription Factor binding (narrow peaks)
  - Histone modification (broad domains)

# ChIP Seq Overview





# BED File

- Defines coordinates for genomic region
- Tab Delimited format

```
$ head ../data/Input_tags.bed
```

```
chr1 233604 233639 0 2
chr1 559767 559802 0 3
chr1 742600 742635 0 2
chr1 742600 742635 0 0
chr1 744231 744266 0 0
chr1 744307 744342 0 2
chr1 746885 746920 0 2
chr1 746958 746993 0 1
chr1 748226 748261 0 2
chr1 748357 748392 0 0
```

- → Non-coding strand  
(3' → 5')

+ → Coding strand (5' → 3')



# CENTRAL DOGMA

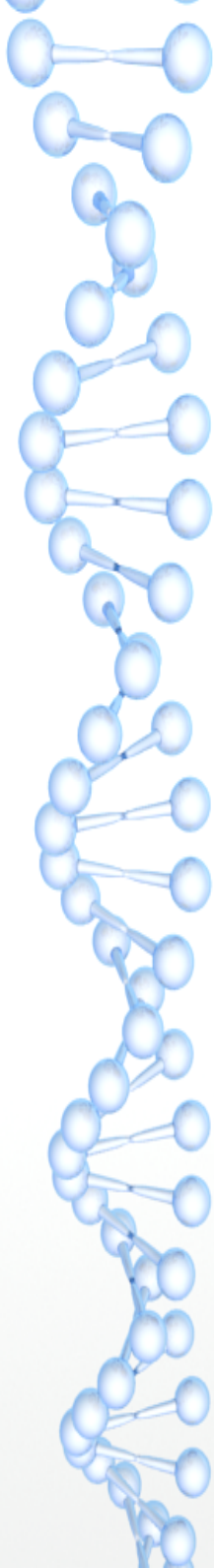
- Gene based enrichment considers coding regions only.
- ~ 99 % of the nucleotides in the human genome **do not** code for proteins.
  - Nearly half of the disease associated SNPs are located in the “junk DNA”.
- Pathway affiliation restricted to genes' definitions in databases.
- Henceforth, Sequence based enrichment is preferred.



# GREAT

- Genomic Regions Enrichment of Annotations Tool
- Bejerano Lab, Stanford
- Online interface
- Highest cited tool
- Input: BED file from the Chip-Seq experiment
- Employs Binomial Distribution





# Snapshot

**GREAT** Overview News Use GREAT Demo Video How to Cite Help Forum

GREAT version 3.0.0 current (02/15/2015 to now)

**GREAT predicts functions of cis-regulatory regions.**

Many coding genes are well annotated with their biological functions. Non-coding regions typically lack such annotation. GREAT assigns biological meaning to a set of non-coding genomic regions by analyzing the annotations of the nearby genes. Thus, it is particularly useful in studying cis functions of sets of non-coding genomic regions. Cis-regulatory regions can be identified via both experimental methods (e.g. [ChIP-seq](#)) and by computational methods (e.g. [comparative genomics](#)). For more see our [Nature Biotech Paper](#).

### News

- October 23, 2017: GREAT is being serviced to eliminate proxy errors.
- June 22, 2017: GREAT hardware is being upgraded, GREAT will be down for 10-15 minutes
- December 7, 2016: GREAT will be down intermittently for hardware relocation.
- Feb 15, 2015: GREAT version 3.0 [switches to Ensembl genes](#), [adds the mouse mm10 assembly](#), and [adds new ontologies](#).
- Apr 3, 2012: GREAT version 2.0 [adds new annotations to human and mouse ontologies](#) and [visualization tools for data exploration](#).
- Feb 18, 2012: The [GREAT forums](#) are released, allowing increased user-to-user interaction

[More news items...](#)

### Species Assembly

☐ Human: GRCh37 ([UCSC hg19](#), Feb/2009)

☐ Mouse: NCBI build 37 ([UCSC mm9](#), Jul/2007)

☐ Mouse: NCBI build 38 ([UCSC mm10](#), Dec/2011)

☐ Zebrafish: Wellcome Trust Zv9 ([danRer7](#), Jul/2010) [Zebrafish CNE set](#)

[Can I use a different species or assembly?](#)

### Test regions

☒ BED file:  No file selected.

☐ BED data:

[What should my test regions file contain?](#)  
[How can I create a test set from a UCSC Genome Browser annotation track?](#)

### Background regions

☒ Whole genome

☐ BED file:  No file selected.

☐ BED data:

[When should I use a background set?](#)  
[What should my background regions file contain?](#)

### Association rule settings

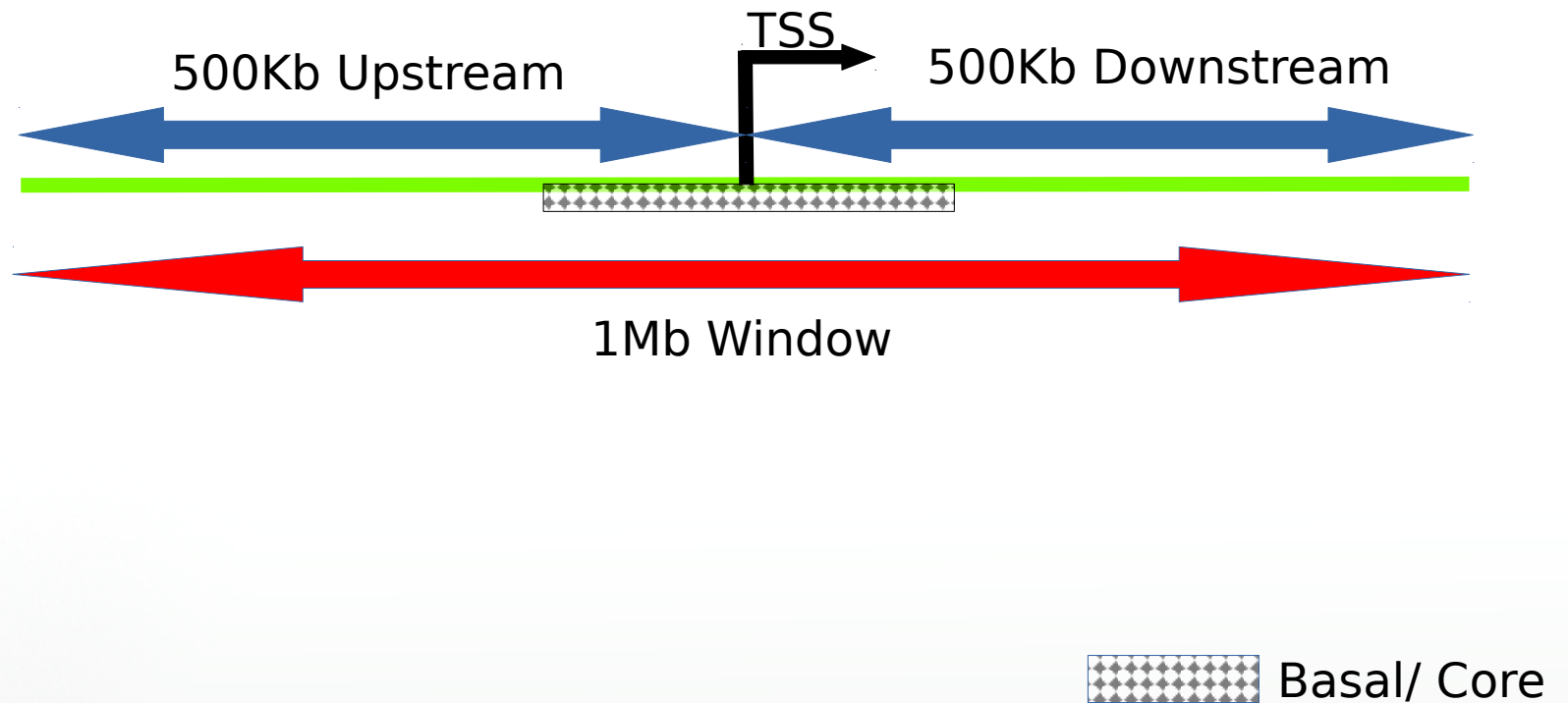
**STANFORD**  
SCHOOL OF MEDICINE

**Bejerano Lab**

**STANFORD**  
ENGINEERING



# Illustration of Regulatory Domain





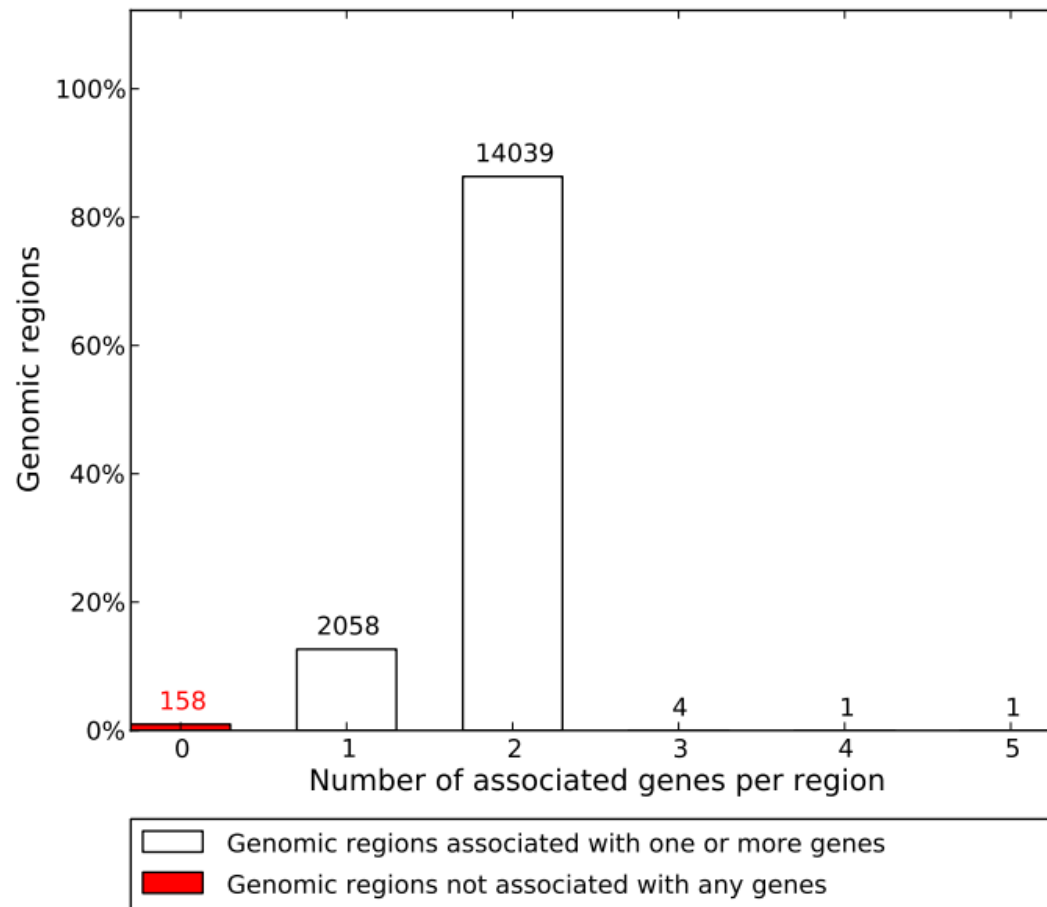
# Assumptions

- Invariable to the gene locus length, the 1Mb window (upstream and downstream TSS) outside 5Kb(upstream) and 1Kb(downstream) holds all vital non-coding elements.
- The above landscape, in unision, constitutes an ideal gene regulatory region.
- The basal gene length has been accounted for normalization.

# Results 1

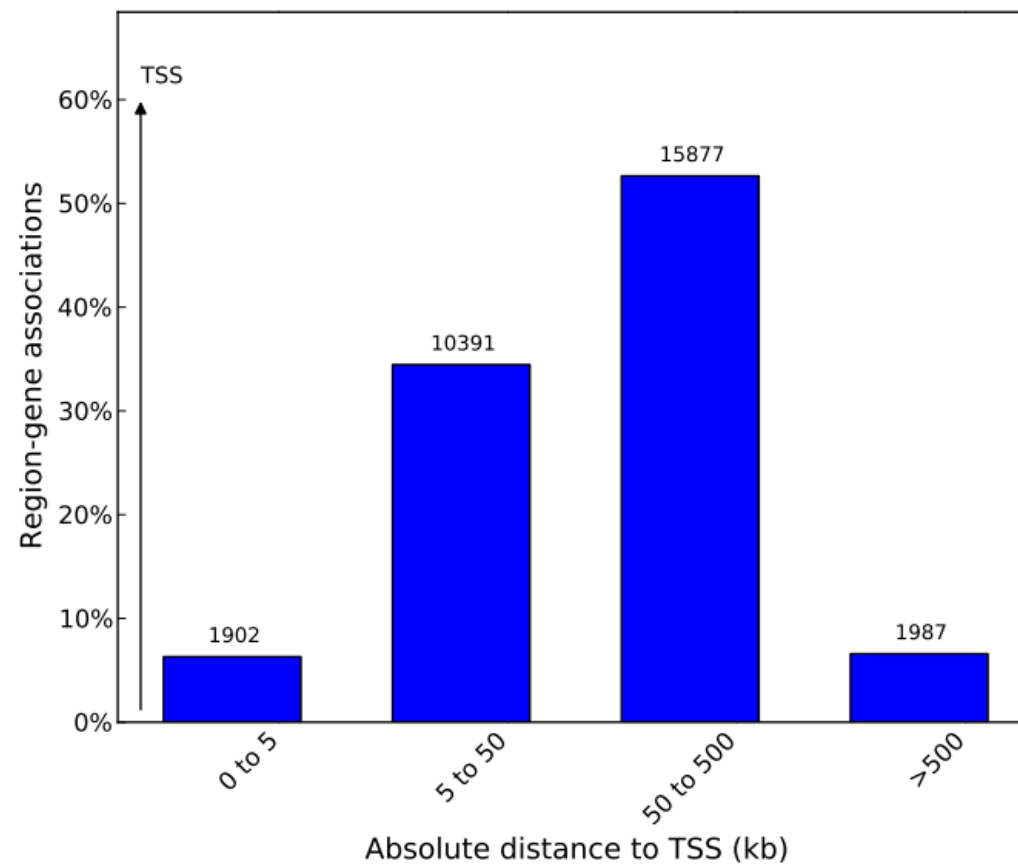
Job ID: 20181008-public-3.0.0-c1gamQ

Display Name: ChIP+Seq+BRCA1+HeLa+Basic+Sample.bed



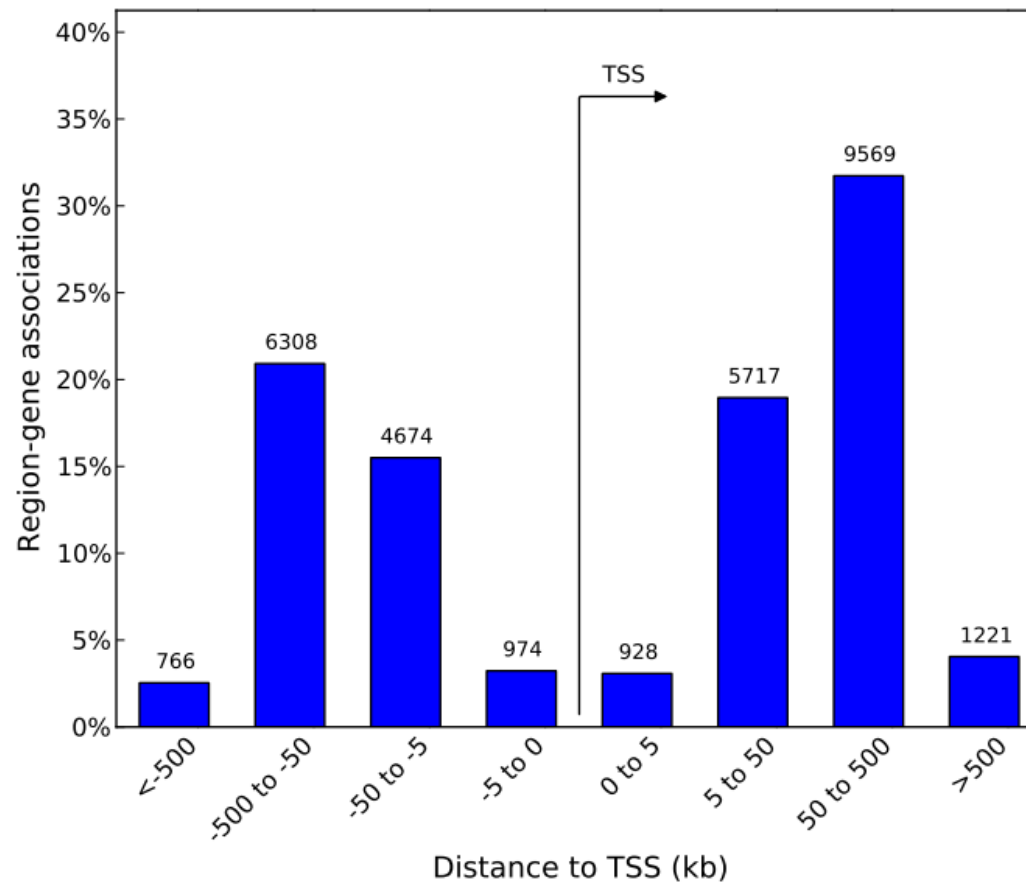
# Results 2

Job ID: 20181008-public-3.0.0-c1gamQ  
Display Name: ChIP Seq BRCA1 HeLa Basic Sample.bed



# Results 3

Job ID: 20181008-public-3.0.0-c1gamQ  
Display Name: ChIP Seq BRCA1 HeLa Basic Sample.bed



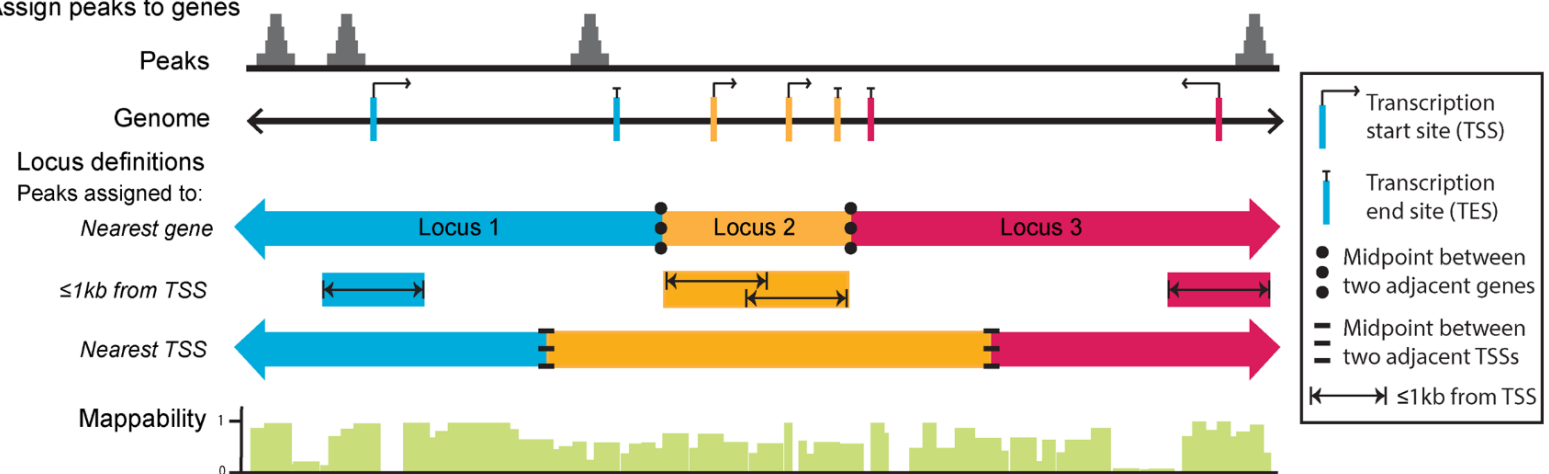


# ChIP- Enrich

- ChIP -Seq data enrichment tool
- University of Michigan
- Available as “chipenrich” R package and also as an online tool: <http://chip-enrich.med.umich.edu>
- Better handles Type I error rate as compared to GREAT
- Employs logistic regression approach
- ChIP-Enrich is designed for use with 1,000s or 10,000s of narrow peaks which results in fewer gene loci containing a peak overall. For example, ChIP-seq experiments for transcription factors

# Core Model

## 1. Assign peaks to genes



## 2. Determine presence of peaks in genes

Gene	Locus length	Presence of peak
ACP1	11,541	0
CHL1	447,985	1
HES4	23,485	0
ITPR1	24,602	1
MYT1L	500,221	1
SAMD11	266,255	0
...	...	...

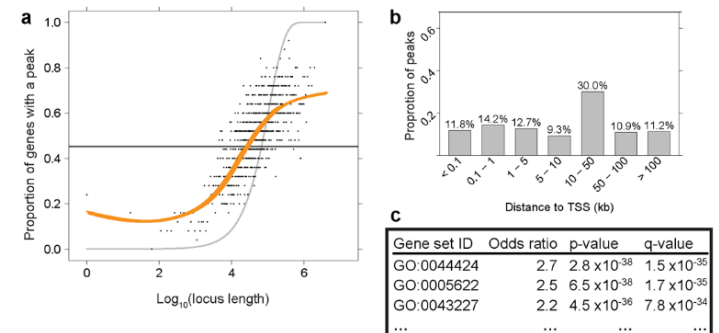
## 3. Test for gene set enrichment

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 g + f(\log_{10}(mL + 1))$$

Logistic regression model

- Adjust for (mappable,  $m$ ) locus length ( $L$ )
- Estimate gene set ( $g$ ) effect size ( $\beta_1$ )

## 4. Summarize data and enrichment results







# Web Interface

- Gene locus definitions/  
available choices

## Locus Definition

- ☐ < 1kb  
*(only use peaks within 1kb of a transcription start site)*
- ☐ < 5kb  
*(only use peaks within 5kb of a transcription start site)*
- ☐ < 10kb  
*(only use peaks within 10kb of a transcription start site)*
- ☐ > 10kb upstream  
*(only use peaks greater than 10kb upstream of a transcription start site)*
- ☐ Exon  
*(only use peaks that fall within an annotated exon)*
- ☐ Intron  
*(only use peaks that fall within an annotated intron)*
- ☐ Nearest Gene  
*(use all peaks; assign peaks to the nearest gene defined by transcription start and end sites)*
- ☒ Nearest TSS  
*(use all peaks; assign peaks to the gene with the closest TSS)*
- ☐ User Defined  
*(user can input their own locus definition)*

# Web Interface Contd...

Supported  
Genomes

Select Genome

Annotation  
Databases

Functional Annotations

- ☐ [Biocarta Pathway](#)
- ☐ [EHMN metabolic pathways](#)
- ☐ [GO](#)
  - ☐ [GO Biological Process](#)
  - ☐ [GO Cellular Component](#)
  - ☐ [GO Molecular Function](#)

- ☐ [KEGG Pathway](#)
- ☐ [Panther Pathway](#)
- ☐ [pFAM](#)
- ☐ [Reactome](#)

Literature Derived

- ☐ [MeSH](#)

MSigDB Derived

- ☐ [Hallmark](#)
- ☐ [Immunologic](#)
- ☐ [Oncogenic](#)

Targets

- ☐ [Comparative Toxicogenomics Database \(CTD\)](#)
- ☐ [Drug Bank](#)
- ☐ [MicroRNA](#)
- ☐ [Transcription Factors](#)

Interaction

- ☐ [Protein Interaction BioGRID](#)

Other

- ☐ [Metabolite](#)
- ☐ [Cytoband](#)

Custom

- ☐ [Custom](#)  No file selected.

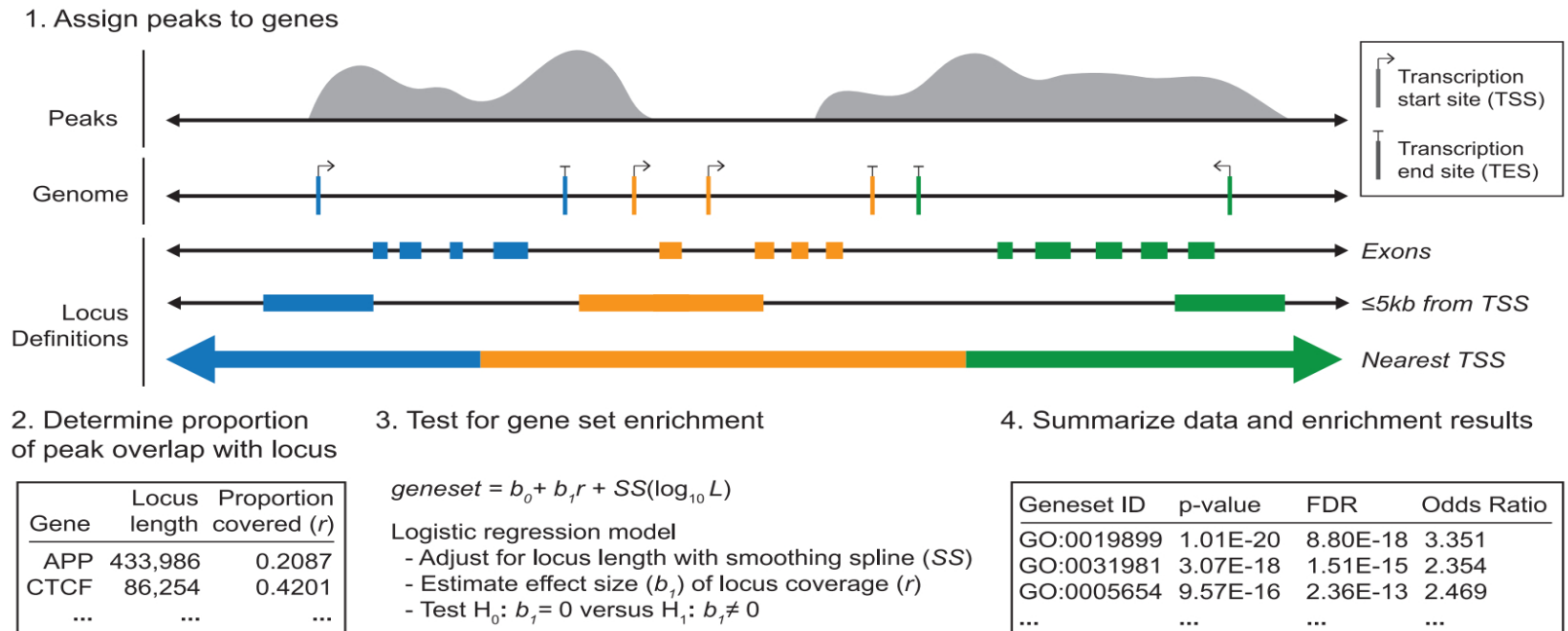
To test custom gene sets, file should be defined in tab-delimited text file with the first column geneset ID or name, and the Entrez IDs belonging to the geneset. An example is provided [here](#).

Select All Databases

- ☐ [SelectAll](#)

Selecting multiple, or a large, annotation database may require several minutes of computation time. For approximate Chip-Enrich running times against different databases view [this](#) table.

# Broad- Enrich



- Broad-Enrich is designed for use with broad peaks that may intersect multiple gene loci, and cumulatively cover greater than 5% of the genome. For example, ChIP-seq experiments for histone modifications
- HISTONE MODIFICATIONS SPAN POLYGENIC DOMAINS
- Greater loci/ coverage than Chip Enrich (TF sites)



# Broad-Enrich: Model

- Agenda is to examine whether a gene with locus length  $L$  and coverage proportion  $r$  claims a gene set membership
- Variable *geneset* is a binary vector
- Logistic regression model applied

$$\log \frac{\pi}{1 - \pi} = b_0 + b_1 r + SS(\log_{10} L)$$



Dependent variable/ odds

# Web Interface



**Broad-Enrich: gene set enrichment testing for sets of broad genomic regions**

## Overview

Broad-Enrich tests sets of broad genomic regions (e.g., from ChIP-seq data for histone modifications or copy number variations) for enriched biological pathways, Gene Ontology terms, or other gene sets. The pre-defined gene sets are the same as used in LRpath, and can be browsed [here](#). Using an input `.bed`, `.narrowPeak` or `.broadPeak` file, Broad-Enrich determines the proportion of each gene locus covered by a peak, using a chosen "gene locus definition". The "locus" of a gene is the region from which the gene is predicted to be regulated. Broad-Enrich uses a logistic regression model to test for association between the proportion of each gene locus covered by a peak and gene set membership. It empirically adjusts for the bias due to locus length using a binomial cubic smoothing spline within the logistic model. Detailed methods are provided [here](#). Output includes summary plots, peak to gene assignments, and enrichment (and depletion) results including odds ratio, p-value, and FDR for each gene set. [more](#)

Broad-Enrich is also available as part of the Chip-Enrich R package: [Broad-Enrich.zip](#)  
Vignette: [pdf](#)

Select input file

Browse...

No file selected.

*Input file should be a standard `.bed`, `.narrowPeak` or `.broadPeak` file containing at least three columns: (1) chromosome (of the form "chr3") (2) start position, and (3) end position. Additional columns will be ignored.*

Analysis Name

*Please provide a meaningful name for this analysis (used to name output files).*

Email

*Please provide your email address to be notified when the analysis is complete.*

Supported  
Genomes

Annotation  
Databases

Select Genome

Select Genome

Human (hg19)

Mouse (mm9)

Rat (rn4)

Mouse(mm10)

GO

[Pathway](#)

[Metabolic Pathways](#)

# Web Interface

## Annotation Databases

- Functional Annotations
  - ☐ [Biocarta Pathway](#)
  - ☐ [EHMN Metabolic Pathways](#)
  - ☐ [GO](#)
    - ☐ [GO Biological Process](#)
    - ☐ [GO Cellular Component](#)
    - ☐ [GO Molecular Function](#)
  - ☐ [KEGG Pathway](#)
  - ☐ [Panther Pathway](#)
  - ☐ [pFAM](#)
- Literature Derived
  - ☐ [MeSH](#)
- Targets
  - ☐ [Drug Bank](#)
  - ☐ [miRBase](#)
  - ☐ [Transcription Factors](#)
- Interaction
  - ☐ [Protein Interaction \(MiMI\)](#)
- Other
  - ☐ [Metabolite](#)
  - ☐ [Cytoband](#)
- Select All Databases
  - ☐ [SelectAll](#)

Selecting multiple, or a large, annotation database may require several minutes of computation time. For approximate Broad-Enrich running times against different databases view [this](#) table.

**Filter** Only test gene sets with less than the following number of genes:

2000

Filter value should be numeric and greater than 30. It can be used to remove large, vague gene sets such as "binding".

## Locus Definition

- ☐ 1kb  
(only use peaks within 1kb of a transcription start site)
- ☐ 5kb  
(only use peaks within 5kb of a transcription start site)
- ☐ Exon  
(only use peaks that fall within an annotated exon)
- ☐ Nearest Gene  
(use all peaks; assign peaks to the nearest gene defined by transcription start and end sites)
- ☒ Nearest TSS  
(use all peaks; assign peaks to the gene with the closest TSS)
- ☐ User Defined  
(user can input their own locus definition)

Adjust for the mappability of the gene locus regions

- ☐ True
- ☐ False

Submit



# Seq2Pathway

- Only available as a R package
- *Seq2gene*
  - Many-many mapping
  - Links coding and non-coding regions to coding genes
- *Gene2pathway*
  - Takes into account the quantity of significance for gene members within a pathway compared to those outside pathway (*Competitive hypothesis*)



# Mapping in Seq2Pathway

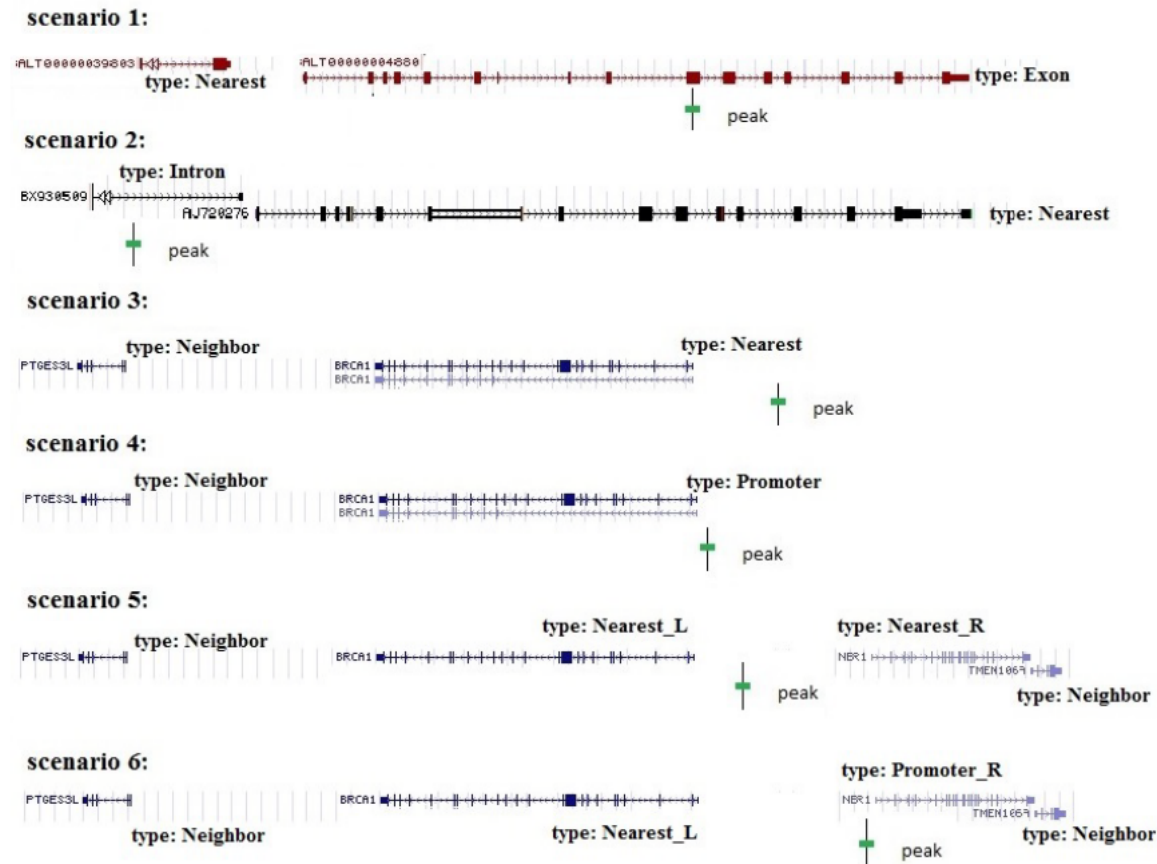


Figure 6: **Six output type values in several scenarios.** In each scenario, we map the genomic region of interest in green to the following types of a coding gene: exon (1), intron (2), the nearest (3), promoter (4), Nearest\_L and Nearest\_R (5), or Promoter\_R (6).



# Application

```
• biocLite("seq2pathway.data")
• biocLite("seq2pathway")
• biocLite("chipenrich")
•
• ### Calling functional libraries ###
•
• library(seq2pathway.data)
• library(seq2pathway)
• library(chipenrich.data)
• library(chipenrich)
•
•
• ### Testing Seq2Pathway with a sample file ###
•
• inputfile <- read.table("/home/postdoc1/Postdoc@GMU/Data/ChIP Seq BRCA1 HeLa Seq2Pathway.bed", sep = "\t", header = FALSE)
• result_seq2gene <- runseq2gene(inputfile, search_radius = 150000, promoter_radius = 200, promoter_radius2 = 100, genome = "hg38",
•     adjacent = FALSE, SNP = FALSE, PromoterStop = FALSE, NearestTwoDirection = TRUE, UTR3 = FALSE)
•
• ### Extracting gene list from the result of seq2gene algorithm ###
•
• seq2gene_gene_list <- result_seq2gene$seq2gene_CodingGeneOnlyResult[, "gene_name"]
• seq2gene_gene_list <- as.data.frame(seq2gene_gene_list)
•
• ### Executing gene list for enrichment ###
• data(MsigDB_C5, package="seq2pathway.data")
• result_gene2pathway <- gene2pathway_test(dat= seq2gene_gene_list, FisherTest=TRUE, EmpiricalTest=TRUE,
•     method="FAIME", alpha=5, logCheck= FALSE, na.rm=FALSE)
```



# Enrichr

- <http://amp.pharm.mssm.edu/Enrichr>
- Freely available online tool
- Released in 2013 with updates in 2014 and 2016.
- HTML5, web-based application with mobile support for iPhone, Android, and Blackberry too.



Enrichr

Analyze

What's New?

Libraries

Find a Gene

About

Help

## Input data

Choose an Input file to upload. Either in BED format or a list of genes. For a quantitative set, add a comma and the level of membership of that gene. The membership level is a number between 0.0 and 1.0 to represent a weight for each gene, where the weight of 0.0 will completely discard the gene from the enrichment analysis and the weight of 1.0 is the maximum.

Or paste in a list of gene symbols optionally followed by a comma and levels of membership. Try two examples: [crisp set example](#), [fuzzy set example](#)

Try an example [BED file](#).

Browse...

No file selected.

0 gene(s) entered

Enter a brief description for the list in case you want to share it. (Optional)

Submit

☐ Contribute

Please acknowledge Enrichr in your publications by citing the following references:

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;128(14).

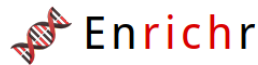
Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016; gkw377.



# Gene- Set Libraries

- *Transcription*
  - ChEA, PWMs(TRANSFAC, JASPAR, UCSC genome browser), TF targets from ENCODE, histone modifications from Roadmap Epigenomics Project, microRNA targets from TargetScan.
- *Pathways*
  - KEGG, BioCarta, Reactome, etc.
- *Ontologies*
  - Gene Ontology trees, Knockout Mouse Phenotypes Ontology from MGI-MP browser, List2Networks.
- *Diseases/ Drugs*
  - Connectivity Map database, GeneSigDB, MSigDB, OMIM, VirusMINT.
- *Cell Types*
  - Mouse and Human Gene Atlases, Cancer Cell Line Encyclopedia (CCLE), NCI-60.
- *Miscellaneous*
  - Chromosomal locations from MSigDB, Metabolite library from HMDB, structural domains library from PFAM and Interpro.

# Enrichr: Input Screen, Transcription



Analyze What's New? Libraries Find a Gene About Help

Login | Register

15,124,298 lists analyzed  
247,533 terms  
133 libraries

## Input data

Choose an input file to upload. Either in BED format or a list of genes. For a quantitative set, add a comma and the level of membership of that gene. The membership level is a number between 0.0 and 1.0 to represent a weight for each gene, where the weight of 0.0 will completely discard the gene from the enrichment analysis and the weight of 1.0 is the maximum.

Try an example [BED file](#).

[Browse...](#) ChIP Seq BRCA1 HeLa Basic Sample.bed

Select parameters for bed file to gene list conversion.

Species:

Human hg18

Max number of genes:

1000

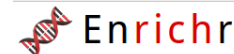
0 gene(s) entered

Enter a brief description for the list in case you want to share it. (Optional)

☐ Contribute

Please acknowledge Enrichr in your publications by citing the following references:  
Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR, Ma'ayan A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;128(14).

Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, Koplev S, Jenkins SL, Jagodnik KM, Lachmann A, McDermott MG, Monteiro CD, Gundersen GW, Ma'ayan A. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*. 2016; gkw377.



Login | Register

Transcription Pathways Ontologies Disease/Drugs Cell Types Misc Legacy Crowd

Description Hello\_World (2000 genes)

### ChEA 2016

GAD6G\_21074721\_Chip-Seq\_CACO-2\_Mouse  
GAD6G\_21074721\_Chip-Seq\_CACO-2\_Human  
S41\_27219007\_Chip-Seq\_Bcells\_Human  
CDK2\_21074721\_Chip-Seq\_CACO-2\_Mouse  
E2F4\_21247883\_Chip-Seq\_LYMPHOBLASTO

### ENCODE and ChEA Consensus TFs from

SMC3\_ENCODE  
E2F1\_CHEA  
JRF8\_CHEA  
SRF\_ENCODE  
FOXO2\_ENCODE

### ARCHS4 TFs Coexp

SP6\_human\_tf\_ARCHS4\_coexpression  
BARX2\_human\_tf\_ARCHS4\_coexpression  
MST1R\_human\_tf\_ARCHS4\_coexpression  
FOXP1\_human\_tf\_ARCHS4\_coexpression  
DOP\_human\_tf\_ARCHS4\_coexpression

### TF Perturbations Followed by

MB03\_KD\_MOUSE\_GSE31008\_CREEDSID\_GE  
FOXO3\_OE\_SW780\_HUMAN\_GSE81157\_RNA  
RARA\_KO\_MOUSE\_GSE31280\_CREEDSID\_GEN  
MECP2\_KO\_MOUSE\_GSE8720\_CREEDSID\_GE  
S6K17\_OE\_HUMAN\_GSE10809\_CREEDSID\_GI

### Enrichr Submissions TF-Gene Cooccurrence

THAP3  
ELF3  
ZBTB48  
MST1R  
DOP

### TRANSFAC and JASPAR PWMs

NR1H2 (human)  
IKZF1 (human)  
RARA (human)  
CRX (human)  
TBX5 (human)

### Epigenomics Roadmap HM ChIP-

H3K36me3 Skeletal Muscle  
H3K36me3 CD3 Primary Cells  
H3K36me3 CD4+ CD25+ CD127- Treg Prim  
H3K36me3 Stomach Mucosa  
H3K9me3 Stomach Mucosa

### TargetScan microRNA 2017

hsa-miR-1249  
mmu-miR-151-5p  
hsa-miR-4706  
hsa-miR-4749-5p  
hsa-miR-3177-3p

### miRTarBase 2017

hsa-miR-642b-5p  
hsa-miR-4534  
hsa-miR-8082  
hsa-miR-6800-5p  
hsa-miR-3167

### ENCODE TF ChIP-seq 2015

FOXL2\_MCF-7\_hg19  
SMARCC1\_Hela-S3\_hg19  
SLP120H\_Hela-S3\_hg19  
NR3C1\_HepG2\_hg19  
SMARCC2\_Hela-S3\_hg19

### TF-LOF Expression from GEO

zif263\_19887448\_helas3\_lof\_human\_gpi68  
myb\_16205643\_mcf7\_gof\_human\_gpi96\_gs  
klf9\_17379758\_hgjunum\_lof\_mouse\_gpi339  
hntf1b\_16297991\_hnk293\_embryonic\_gof\_n  
gls3\_19805515\_embryonic\_pancreas\_lof\_n

### ENCODE Histone Modifications 2015

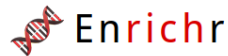
H3K4me1\_NT2-D1\_hg19  
H3K27me3\_GM06990\_hg19  
H3K27me3\_splenic B cell\_mmj9  
H3K36me3\_bronchial epithelial cell\_hg19  
H3K36me3\_BJ\_hg19

Transcription Factor

Genome Browser



# Enrichr: Pathways, Ontologies



Login | Register

Transcription Pathways Ontologies Disease/Drugs Cell Types Misc Legacy Crowd

Description Hello\_World (2000 genes)

## KEGG 2016

Other glycan degradation\_Homo sapiens\_h  
DNA replication\_Homo sapiens\_hsa03030  
Hematopoietic cell lineage\_Homo sapiens\_h  
mRNA surveillance pathway\_Homo sapiens  
Peroxisome\_Homo sapiens\_hsa04146

## WikiPathways 2016

Estrogen Receptor Pathway\_Homo sapiens  
Fatty Acid Biosynthesis\_Homo sapiens\_WP  
Fatty Acid Biosynthesis\_Mus musculus\_WP  
Complement and Coagulation Cascades\_Mi  
Complement and Coagulation Cascades\_Ho

## ARCHS4 Kinases Coexp

MST1R\_human\_kinase\_ARCHS4\_coexpressi  
CDC42BP8\_human\_kinase\_ARCHS4\_coexp  
PKDCC\_human\_kinase\_ARCHS4\_coexpressi  
RIPK4\_human\_kinase\_ARCHS4\_coexpressi  
CDC42BP6\_human\_kinase\_ARCHS4\_coexp

## Reactome 2016

Removal of the Flap Intermediate\_Homo sa  
Lagging Strand Synthesis\_Homo sapiens\_R-I  
Processive synthesis on the lagging strand  
Organic anion transport\_Homo sapiens\_R-I  
Extension of Homers\_Homo sapiens\_R-H

## BioCarta 2016

Induction of apoptosis through DR3 and DI  
Caspase Cascade in Apoptosis\_Homo sapie  
Control of Gene Expression by Vitamin D R  
Granzyme A mediated Apoptosis Pathway\_h  
TNFR2 Signalling Pathway\_Homo sapiens\_h

## HumanCyc 2016

protein citrullination\_Homo sapiens\_PWY-4  
coenzyme A biosynthesis\_Homo sapiens\_Ci  
protein ubiquitylation\_Homo sapiens\_PWY-  
leukotriene biosynthesis\_Homo sapiens\_PV  
terminal O-glycans residues modification\_h

## NCI-Nature 2016

EPHA2 forward signaling\_Homo sapiens\_h  
HIV-1 Nef: Negative effector of Fas and TNF-  
IL23-mediated signaling events\_Homo sapi  
IL6-mediated signaling events\_Homo sapien  
Notch-mediated HES/HEY network\_Homo sa

## Panther 2016

Alzheimer disease-presenilin pathway\_Hom  
Pyruvate metabolism\_Homo sapiens\_P0277  
Huntington disease\_Homo sapiens\_P00029  
Coenzyme A biosynthesis\_Homo sapiens\_P  
EGF receptor signaling pathway\_Homo sapi

## BioPlex 2017

TRIML2  
ZMPSTE24  
MED24  
MED14  
MED4

## huMAP

CYP51A1  
MAGOH  
MYO1E  
MED30  
RPL35A

## PPI Hub Proteins

NCOR1  
KAT5  
MTOR  
CSNK1E  
IL7R

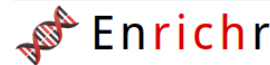
## KEA 2015

UHMK1  
BUB1  
PTK2  
IRAK1  
MAPKAPK2

## LINCS L1000 Kinase

## LINCS L1000 Kinase

## Kinase Perturbations



Login | Register

Transcription Pathways Ontologies Disease/Drugs Cell Types Misc Legacy Crowd

Description Hello\_World (2000 genes)

## GO Biological Process 2018

negative regulation of complement activati  
positive regulation of monocyte differenti  
ATP synthesis coupled electron transport (C  
urate transport (GO:0015747)  
photoperiodism (GO:0009648)

## GO Molecular Function 2018

RNA polymerase I core binding (GO:00010  
single-stranded DNA exodeoxyribonuclease  
phosphate ion binding (GO:0042301)  
5'-3' DNA helicase activity (GO:0043139)  
urate transmembrane transporter activity (C

## GO Cellular Component 2018

multivesicular body membrane (GO:00325  
ricolin-1-rich granule membrane (GO:01011  
AP-1 adaptor complex (GO:0030121)  
ricolin-1-rich granule (GO:0101002)  
platelet alpha granule (GO:0031091)

## MGI Mammalian Phenotype 2017

MP:0002796\_impaired\_skin\_barrier\_functi  
MP:0009932\_skin\_fibrosis  
MP:0001243\_abnormal\_dermal\_layer\_morp  
MP:0001222\_epidermal\_hyperplasia  
MP:0001861\_lung\_inflammation

## Human Phenotype Ontology

Mitral stenosis (HP:0001718)  
Generalized amyotrophy (HP:0003700)  
Facial shape deformation (HP:0011334)  
Potter facies (HP:0002009)  
Flattened epiphyses (HP:0003071)

## Jensen TISSUES

Phagocyte  
Pit-K1\_cell  
Excretory\_canal  
Stratified\_epithelium  
Mantle

## Jensen COMPARTMENTS

Platelet\_alpha\_granule  
Viral\_assembly\_compartment  
glycosylphosphatidylinositol-mannosyltran  
Keratin\_filament  
Phagolysosome\_membrane

## Jensen DISEASES

Amyotrophic\_lateral\_sclerosis\_type\_8  
Intermittent\_claudication  
Peripheral\_artery\_disease  
Sialadenitis  
Uveitis



# Deployment

Methods	R	Web Interface
GREAT		✓
SEQ2PATHWAY	✓	
CHIP ENRICH	✓	✓
BROAD ENRICH	✓	✓
ENRICH		✓





# References

- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., ... Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics, 14. <https://doi.org/10.1186/1471-2105-14-128>
- Welch, R. P., Lee, C., Imbriano, P. M., Patil, S., Weymouth, T. E., Smith, R. A., ... Sartor, M. A. (2014). ChIP-Enrich: Gene set enrichment testing for ChIP-seq data. Nucleic Acids Research, 42(13), 1–13. <https://doi.org/10.1093/nar/gku463>
- Cavalcante, R. G., Lee, C., Welch, R. P., Patil, S., Weymouth, T., Scott, L. J., & Sartor, M. A. (2014). Broad-Enrich: Functional interpretation of large sets of broad genomic regions. Bioinformatics, 30(17), 393–400. <https://doi.org/10.1093/bioinformatics/btu444>
- McLean, C. Y., Bristor, D., Hiller, M., Clarke, S. L., Schaar, B. T., Lowe, C. B., ... Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nature Biotechnology, 28(5), 495–501. <https://doi.org/10.1038/nbt.1630>
- Wang, B., Cunningham, J. M., & Yang, X. (2015). Seq2pathway: An R/Bioconductor package for pathway analysis of next-generation sequencing data. Bioinformatics, 31(18), 3043–3045. <https://doi.org/10.1093/bioinformatics/btv289>