

Public data resources

Stockholm, November 9 2018

Jakub Orzechowski Westholm

Long-term bioinformatics support

NBIS, SciLifeLab, Stockholm University

SciLifeLab

NBIS
NATIONAL BIOINFORMATICS
INFRASTRUCTURE SWEDEN

This lecture

- Big projects generating a lot of ChIP-seq data
 - ENCODE/modENCODE
 - Roadmap Epigenomics
- How to find public ChIP-seq data sets from smaller studies
 - Cistrome data browser
- Motif data bases

Public data can be very useful

- Good to have reference data to check if your experiment is ok
- Overlaps between your data and other TFs and chromatin marks
- Compare CHIP-seq data to your expression data



The ENCODE project

- **Enc**yclopedia **O**f **D**N **A** **E**lements: <https://www.encodeproject.org>
- Aim: Using different techniques to annotate the human genome
 - RNA-seq
 - ChIP-seq (around 5000 experiments, TFs, histones and histone marks)
 - DNase-seq/ATAC-seq
 - Hi-C
 - Bisulphite seq
- Mostly human cell lines. Now also some primary tissue, and mouse cell lines and primary cells.
 - modENCODE - a side project for model organisms: fly and worm
 - The ENCODE website also contains data from Roadmap Epigenomics
- Well defined pipelines and quality standards.

ENCODE Data Encyclopedia Materials & Methods Help

Search...

Matrix
Search
 Summary

Search by region
 Reference epigenomes
 Publications

ENCODE: Encyclopedia of DNA Elements

5C ChIA-PET Hi-C
 DNase-seq FAIRE-seq ATAC-seq
 ChIP-seq
 WGBS RRBS methyl array
 Computational predictions
 RNA-seq
 CLIP-seq RIP-seq

Long-range regulatory elements (enhancers, repressors/silencers, insulators)
 Promoters
 Transcripts

Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)

ENCODE Data Encyclopedia Materials & Methods Help

p300 homo sapiens

Showing 25 of 100 results

Data Type	Count
Dataset	100
Experiment	81
Series	18
ReferenceEpigenome	17
FileSet	1
Project	1
TreatmentTimeSeries	1

Audit category: ⚠️
 missing control alignments 46
 extremely low read depth

Audit category: 📄
 missing controlled_by 5
 insufficient replicate concordance
 control insufficient read depth
 insufficient read depth
 poor library complexity

Audit category: 🟡
 antibody characterized with exemption 46
 low read length 12
 partial reference epigenome 12
 low read depth 10
 mild to moderate bottleneaking 6
 inconsistent platforms 4
 borderline replicate concordance 3
 inconsistent target of control experiment 3
 inconsistent control read length 2
 moderate library complexity 2
 control low read depth 1

audit.NOT_COMPLIANT.category!
 poor library complexity
 insufficient read depth
 control insufficient read depth

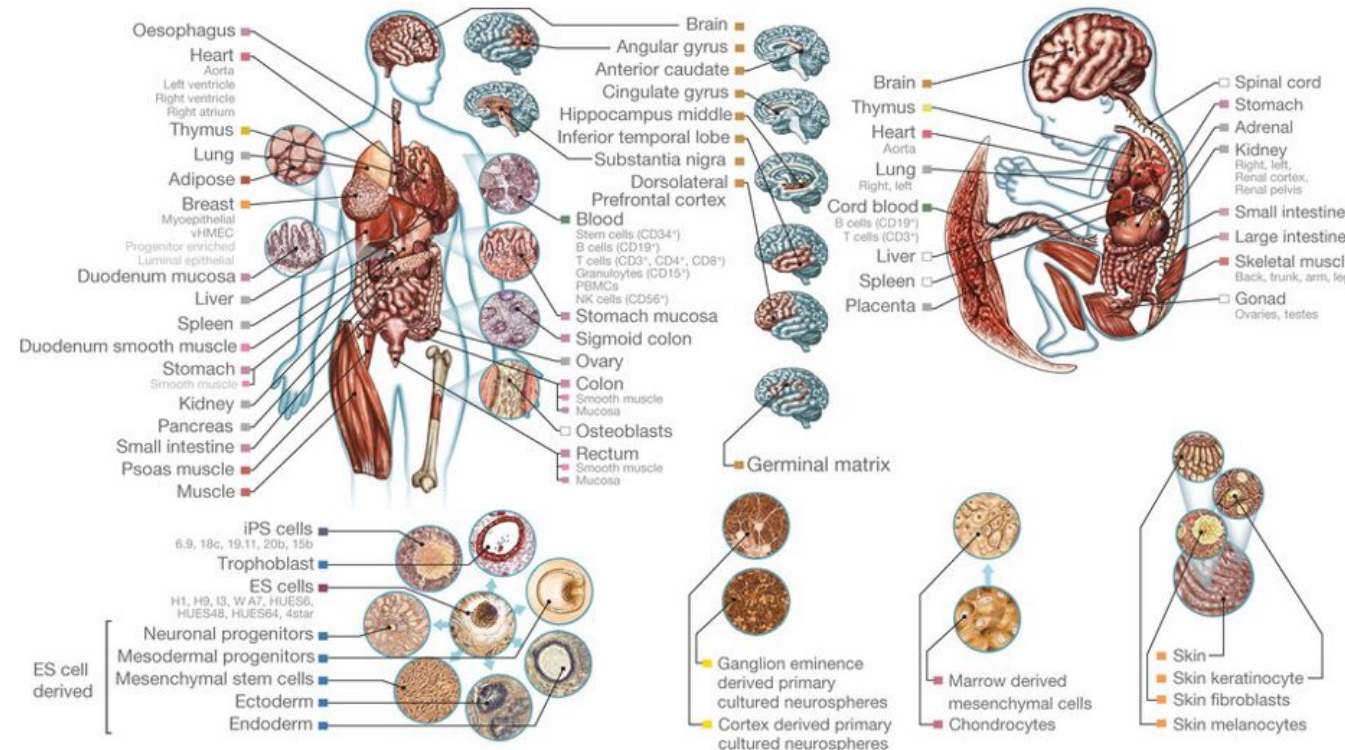
ChIP-seq of A549	Experiment
<i>Homo sapiens</i> A549 treated with 100 nM dexamethasone for 5 hours Target: EP300 Lab: Tim Reddy, Duke Project: GGR	ENCSR610RKF released
<i>Homo sapiens</i> A549 treated with 100 nM dexamethasone for 30 minutes Target: EP300 Lab: Tim Reddy, Duke Project: GGR	ENCSR260WCE released
<i>Homo sapiens</i> A549 treated with 100 nM dexamethasone for 15 minutes Target: EP300 Lab: Tim Reddy, Duke Project: GGR	ENCSR788VKG released
<i>Homo sapiens</i> A549 treated with 100 nM dexamethasone for 3 hours Target: EP300 Lab: Tim Reddy, Duke Project: GGR	ENCSR047EVQ released
<i>Homo sapiens</i> A549 treated with 100 nM dexamethasone for 1 hour Target: EP300 Lab: Tim Reddy, Duke Project: GGR	ENCSR358ELZ released
<i>Homo sapiens</i> A549 Target: EP300 Lab: Tim Reddy, Duke Project: GGR	ENCSR886OEO released

- Downloads:
 - Raw reads: fastq
 - Aligned reads: bam
 - Read coverage: bw
 - Peaks: MACS2

ENCODE										
Data Encyclopedia Materials & Methods Help										
Search...										
- Raw sequencing data										
Isogenic replicate	Library	Accession	File type	Run type	Read	Lab	Date added	File size	Audit status	File status
1	ENCLB408ZKM	ENCFF192EWP ⓘ ⬇	fastq	SE51nt		Tim Reddy, Duke	2016-10-19	436 MB	✓	● released
2	ENCLB473RRB	ENCFF685TOJ ⓘ ⬇	fastq	SE51nt		Tim Reddy, Duke	2016-10-19	471 MB	✓	● released
3	ENCLB283CWG	ENCFF860OTI ⓘ ⬇	fastq	SE51nt		Tim Reddy, Duke	2016-10-19	368 MB	✓	● released
- Processed data										
Accession	File type	Output type	Isogenic replicate	Mapping assembly	Lab	Date added	File size	Audit status	File status	
ENCFF141LII ⓘ ⬇	bigBed narrowPeak	peaks	1	GRCh38	Tim Reddy, Duke	2017-10-19	507 kB	✓	● released	
ENCFF507PNL ⓘ ⬇	bam	alignments	1	GRCh38	Tim Reddy, Duke	2016-10-19	423 MB	✓	● released	
ENCFF891WYF ⓘ ⬇	bed narrowPeak	peaks	1	GRCh38	Tim Reddy, Duke	2016-10-20	139 kB	✓	● released	
ENCFF028JFF ⓘ ⬇	bigWig	signal of unique reads	1	GRCh38	Tim Reddy, Duke	2016-10-19	99.6 MB	✓	● released	
ENCFF598IVG ⓘ ⬇	bed narrowPeak	peaks	2	GRCh38	Tim Reddy, Duke	2016-10-20	164 kB	✓	● released	
ENCFF992TSS ⓘ ⬇	bigWig	signal of unique reads	2	GRCh38	Tim Reddy, Duke	2016-10-19	122 MB	✓	● released	
ENCFF595LRY ⓘ ⬇	bigBed narrowPeak	peaks	2	GRCh38	Tim Reddy, Duke	2017-10-19	578 kB	✓	● released	
ENCFF706FEI ⓘ ⬇	bam	alignments	2	GRCh38	Tim Reddy, Duke	2016-10-19	512 MB	✓	● released	
ENCFF852JEQ ⓘ ⬇	bigWig	signal of unique reads	3	GRCh38	Tim Reddy, Duke	2016-10-19	96.5 MB	✓	● released	

Roadmap epigenomics project

- <http://www.roadmapepigenomics.org>
- Aim: “producing a public resource of human epigenomic data to catalyze basic biology and disease-oriented research”
 - RNA-seq
 - ChIP-seq (mostly chromatin)
 - Bisulphite seq
 - .
- Primary cells, and stem cells
- No nice interface to download data → Better to use ENCODE website.



Index of /EdaccData/Release-9/sample-experim

../			
Adipose_Derived_Mesenchymal_Stem_Cell_Cultured_..>	04-Jun-2013	16:04	-
Adipose_Nuclei/	04-Jun-2013	16:04	
Adipose_Tissue/	04-Jun-2013	16:07	
Adrenal_Gland/	04-Jun-2013	16:07	
Adult_Kidney/	04-Jun-2013	16:04	
Adult_Liver/	04-Jun-2013	16:05	
Aorta/	04-Jun-2013	16:07	
Bladder/	15-May-2013	19:03	
Bone_Marrow_Derived_Mesenchymal_Stem_Cell_Cultu..>	04-Jun-2013	16:05	
Brain_Angular_Gyrus/	04-Jun-2013	16:05	
Brain_Anterior_Caudate/	04-Jun-2013	16:05	
Brain_Cerebellum/	12-Jun-2013	16:29	
Brain_Cingulate_Gyrus/	04-Jun-2013	16:06	
Brain_Germinal_Matrix/	04-Jun-2013	16:08	
Brain_Hippocampus_Middle/	04-Jun-2013	16:06	
Brain_Inferior_Temporal_Lobe/	04-Jun-2013	16:06	
Brain_Mid_Frontal_Lobe/	04-Jun-2013	16:06	
Brain_Substantia_Nigra/	04-Jun-2013	16:07	
Breast_Fibroblast_Primary_Cells/			
Breast_Luminal_Epithelial_Cells/			
Breast_Myoepithelial_Cells/			
Breast_Stem_Cells/			
Breast_vHMEC/			
CD14_Primary_Cells/			
CD15_Primary_Cells/			
CD19_Primary_Cells/			
CD20_Primary_Cells/			
CD34_Cultured_Cells/			
CD34_Primary_Cells/			
CD3_Primary_Cells/			
CD4+ CD25+ CD127- Treg_Primary_Cells/			
CD4+ CD25- CD45RA+ Naive_Primary_Cells/			
CD4+ CD25- CD45RO+ Memory_Primary_Cells/			
CD4+ CD25- IL17+ PMA-Ionomycin_stimulated_T			
CD4+ CD25- IL17- PMA-Ionomycin_stimulated_M			
CD4+ CD25- Th_Primary_Cells/	12-Feb-2013	16:15	-
CD4+ CD25int CD127+ Tmem_Primary_Cells/	12-Feb-2013	16:15	-
CD4_Memory_Primary_Cells/	04-Jun-2013	16:08	-
CD4_Naive_Primary_Cells/	04-Jun-2013	16:08	-

Index of /EdaccData/Release-9/sample-experim

../			
Bisulfite-Seq/	04-Jun-2013	16:06	-
ChIP-Seq_Input/	04-Jun-2013	16:06	-
Histone_H3K27ac/	04-Jun-2013	16:06	-
Histone_H3K27me3/	04-Jun-2013	16:06	-
Histone_H3K36me3/	04-Jun-2013	16:06	-
Histone_H3K4me1/	04-Jun-2013	16:06	-
Histone_H3K4me3/	04-Jun-2013	16:06	-
Histone_H3K9ac/	04-Jun-2013	16:06	-
Histone_H3K9me3/	04-Jun-2013	16:06	-
mRNA_Seq/	04-Jun-2013	16:06	-

Index of /EdaccData/Release-9/sample-experiment/Brain_Hippocampus_Middle/Histone_H3K27ac/

../			
BI.Brain_Hippocampus_Middle.H3K27ac.112.bed.gz	21-Dec-2012	16:20	197M
BI.Brain_Hippocampus_Middle.H3K27ac.112.wig.gz	21-Dec-2012	16:44	27M
BI.Brain_Hippocampus_Middle.H3K27ac.149.bed.gz	28-Dec-2011	18:05	332M
BI.Brain_Hippocampus_Middle.H3K27ac.149.wig.gz	21-Dec-2011	21:12	36M
BI.Brain_Hippocampus_Middle.H3K27ac.150.bed.gz	09-Mar-2012	22:27	312M
BI.Brain_Hippocampus_Middle.H3K27ac.150.wig.gz	08-Mar-2012	22:30	33M

BI.Brain_Hippocampus_Middle.H3K27ac.112.bed.gz	21-Dec-2012	16:20	197M
BI.Brain_Hippocampus_Middle.H3K27ac.112.wig.gz	21-Dec-2012	16:44	27M
BI.Brain_Hippocampus_Middle.H3K27ac.149.bed.gz	28-Dec-2011	18:05	332M
BI.Brain_Hippocampus_Middle.H3K27ac.149.wig.gz	21-Dec-2011	21:12	36M
BI.Brain_Hippocampus_Middle.H3K27ac.150.bed.gz	09-Mar-2012	22:27	312M
BI.Brain_Hippocampus_Middle.H3K27ac.150.wig.gz	08-Mar-2012	22:30	33M

Cistrome data browser

- An interface for accessing many ChIP-seq data sets.
<http://cistrome.org/db/>
- All data have been re-processed using the same pipeline.
- 47000 experiments, about 50-50 from human and mouse
- Data from many smaller studies collected



Cistrome Data Browser

Containing word(s):

Species

- All
- Homo sapiens**
- Mus musculus

Biological Sources

- All
- 786-O
- A549
- Acute myeloid leukemia
- B cell lymphoma
- B Lymphocyte

Factors

- All
- EP300

Batch	Species	Biological Source	Factor	Publication	Status	Quality Control
<input type="checkbox"/>	Homo sapiens	LP-1	EP300	Conery AR, et al. Elife 2016	completed	
<input type="checkbox"/>	Homo sapiens	Endothelial Cell; Umbilical Vein	EP300	Wong BW, et al. Nature 2016	completed	
<input type="checkbox"/>	Homo sapiens	H1; Embryonic Stem Cell; Embryo	EP300	Lister R, et al. Nature 2009	completed	
<input type="checkbox"/>	Homo sapiens	GM12878; B Lymphocyte; Blood	EP300	Pope BD, et al. Nature 2014	completed	
<input type="checkbox"/>	Homo sapiens	GM12878; B Lymphocyte; Blood	EP300	Gaspar WC, et al. Sci Rep 2014	completed	
<input type="checkbox"/>	Homo sapiens	T47D; Epithelium; Mammary Gland	EP300	Mohammed H, et al. Nature 2015	completed	
<input type="checkbox"/>	Homo sapiens	MCF-7; Epithelium; Breast	EP300	Mohammed H, et al. Nature 2015	completed	
<input type="checkbox"/>	Homo sapiens	MCF-7; Epithelium; Breast	EP300	Mohammed H, et al. Nature 2015	completed	
<input type="checkbox"/>	Homo sapiens	A549; Epithelium; Lung	EP300	Davis CA, et al. Nucleic Acids Res.	completed	
<input type="checkbox"/>	Homo sapiens	T Lymphocyte; Blood	EP300		completed	
<input type="checkbox"/>	Homo sapiens	Cortex	EP300	Visel A, et al. Cell 2013	completed	
<input type="checkbox"/>	Homo sapiens	SK-N-SH; Neuroblastoma cell; Brain	EP300	Pope BD, et al. Nature 2014	completed	
<input type="checkbox"/>	Homo sapiens	MCF-7; Epithelium; Breast	EP300	Liu Z, et al. Cell 2014	completed	
<input type="checkbox"/>	Homo sapiens	Kasumi-1; Myeloblast; Blood	EP300	Trombly DJ, et al. BMC Genomics 2015	completed	
<input type="checkbox"/>	Homo sapiens	PC-3; Epithelium; Prostate	EP300		completed	
<input type="checkbox"/>	Homo sapiens	A549; Epithelium; Lung	EP300	Davis CA, et al. Nucleic Acids Res.	completed	
<input type="checkbox"/>	Homo sapiens	Endothelial Cell; Umbilical Vein	EP300	Zhang B, et al. Genome Res. 2013	completed	
<input type="checkbox"/>	Homo sapiens	CUTLL1; T Lymphocyte; Blood	EP300	Wang H, et al. Proc. Natl. Acad. Sci. U.S.A. 2014	completed	
<input type="checkbox"/>	Homo sapiens	GM12878; B Lymphocyte; Blood	EP300	Gaspar WC, et al. Sci Rep 2014	completed	
<input type="checkbox"/>	Homo sapiens	GM12878; B Lymphocyte; Blood	EP300	Gaspar WC, et al. Sci Rep 2014	completed	

Inspector

Title: Treatment
• LP1_DMSO_P300

ID: GEO or ENCODE: GSM1847311
CistromeDB: 67841

Species: Homo sapiens

Citation: Conery AR, et al. Bromodomain inhibition of the transcriptional coactivators CBP/EP300 as a therapeutic strategy to target the IRF4 network in multiple myeloma. Elife 2016
PMID: 26731516

Factor: EP300

Biological Source: Cell Line: LP-1

Quality Control

Visualize

Download

Tools

QC	Sample
Raw sequence median quality score	38
% Reads uniquely mapped	87.9%
PCR bottleneck coefficient (PBC)	99.7%
Number of merged Total/Fold 10/Fold 20 peaks	3750 / 29 / 1
Fraction of reads in peaks (FRIP)	0.4%
% Peaks in promoter/exon/intron/intergenic	7.4% / 3.6% / 38.9% / 50.1%
% Top 5k peaks overlapping with union DHS	86.1%

% Top 5k peaks Phastcon Conservation Profiles

- Downloads:
 - Read coverage: bw
 - Peaks: bed

R interfaces

The screenshot shows the Bioconductor website for the ENCODEExplorer package. The top navigation bar includes Home, Install, Help, Developers, and About. The breadcrumb trail is Home » Bioconductor 3.8 » Software Packages » ENCODEExplorer. The package name ENCODEExplorer is displayed in green. Metadata includes platforms (all), rank (599 / 1649), posts (0), and in Bioc (3.5 years). The build status is 'ok' and it is updated before release. The DOI is 10.18129/B9.bioc.ENCODEExplorer. The description is 'A compilation of ENCODE metadata'. The Bioconductor version is Release (3.8). The author list includes Charles Joly Beauparlant, Audrey Lemacon, Louis Gendron, Astrid-Louise Deschenes, and Arnaud Droit. The maintainer is Charles Joly Beauparlant. The citation is provided in R code format. The installation instructions are:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("ENCODEExplorer", version = "3.8")
```

 The documentation section includes a link to browse vignettes:

```
browseVignettes("ENCODEExplorer")
```

The screenshot shows the Bioconductor website for the DeepBlueR package. The top navigation bar includes Home, Install, Help, Developers, and About. The breadcrumb trail is Home » Bioconductor 3.8 » Software Packages » DeepBlueR. The package name DeepBlueR is displayed in green. Metadata includes platforms (all), rank (724 / 1649), posts (2 / 1 / 1 / 0), and in Bioc (2 years). The build status is 'error' and it is updated before release. The DOI is 10.18129/B9.bioc.DeepBlueR. The description is 'Accessing the DeepBlue Epigenetics Data Server through R'. The author is Felipe Albrecht, Markus List. The maintainer is Felipe Albrecht and Markus List. The citation is provided in R code format. The installation instructions are:

```
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("DeepBlueR", version = "3.8")
```

 The documentation section includes a link to browse vignettes:

```
browseVignettes("DeepBlueR")
```

Databases with TF binding site motifs

- JASPAR (<http://jaspar.genereg.net>). Good, curated, free, data base with around 1500 motifs from all kinds of species.
- Transfac (<http://genexplain.com/transfac/>, <http://gene-regulation.com/pub/databases.html>). Good, curated, not free, data base with around 5000 motifs from all kinds of species.
 - Old version with 400 motifs is free for academic use.
- Other databases
 - CHIPBase <http://rna.sysu.edu.cn/chipbase/>
 - HOCOMOCO (human only) <http://hocomoco11.autosome.ru>
 - footprintDB (combining several databases) <http://floresta.eead.csic.es/footprintdb/index.php>

The JASPAR database

The screenshot displays the JASPAR 2018 website interface. At the top left, the logo "JASPAR²⁰¹⁸" is visible. The top right corner contains a "Cart" icon with a "0" and a link to the "JASPAR Blog". A navigation menu on the left lists: Home, About, Search, Browse JASPAR CORE, Browse Collections, Tools, RESTful API, Download Data, Matrix Clusters, and Genome Tracks. The main content area features a search bar with "Sox2" entered and a "Search" button. Below the search bar, it lists "Examples: SPI1, P17676, CHIP-seq, Homo sapiens" and an "Advanced Options" link. A section titled "Browse JASPAR CORE for six different taxonomic groups" displays six categories with representative images: Vertebrata (fish), Nematoda (red worm), Insecta (fly), Plantae (fern), Fungi (mushrooms), and Urochordata (peacock butterfly). To the right, a large banner reads "The high-quality transcription factor binding profile database" with a "Read more about JASPAR" link and a "JASPAR interactive tour" button. At the bottom, there is a "Citing JASPAR 2018" section with links to PubMed, NAR, and PDF, followed by the citation: "Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. 2018; 46:D260–D266, doi: 10.1093/nar/gkx1126". A final section titled "JASPAR CORE & when should it be used?" includes a link to "Info about other collections" and a paragraph explaining that the JASPAR CORE contains a curated, non-redundant set of profiles derived from published and experimentally defined transcription factor binding sites for eukaryotes, used for seeking models for specific factors or structural classes.

- Home
- About
- Search
- Browse JASPAR CORE
- Browse Collections
- Tools
- RESTful API
- Download Data
- Matrix Clusters
- Genome Tracks

Detailed information of matrix profile MA0142.1

Home > Matrix > MA0142.1

Profile summary [Add](#)

Name:	Pou5f1::Sox2
Matrix ID:	MA0142.1
Class:	Homeo domain factors::High-mobility group (HMG) domain factors
Family:	POU domain factors::SOX-related factors
Collection:	CORE
Taxon:	Vertebrates
Species:	Mus musculus
Data Type:	ChIP-seq
Validation:	18555785
Uniprot ID:	P20263 P48432
Pazar TF:	
TFBSshape ID:	135
TFencyclopedia IDs:	810
Source:	

Sequence logo [Download SVG](#)



Frequency matrix

- [JASPAR](#)
- [TRANSFAC](#)
- [MEME](#)
- [RAW PFM](#)
- [Reverse comp.](#)











A	[63 579 11 47 118 415 206 1235 17 10 14 1050 887 11
C	[846 58 50 12 363 18 365 30 5 16 1029 16 68 31
G	[66 28 36 79 825 29 186 24 15 1239 129 30 315 71
T	[389 701 1270 1231 63 907 612 80 1332 104 196 269 92 51

Binding sites information [-](#)

- [HTML file](#)
- [FASTA file](#)







TFBS profiles [-](#)



TFBSshape

-  Home
-  About <
-  Search
-  Browse JASPAR CORE
-  Browse Collections <
-  Tools <
-  RESTful API
-  Download Data
-  Matrix Clusters
-  Genome Tracks

Download JASPAR data

Home > Downloads

-  CORE PFMs
-  TFFMs
-  JASPAR collections (PFMs)
-  Matrix clustering
-  SQL dump
-  Other data

 Download JASPAR CORE *Position Frequency Matrix (PFM)* (non-redundant and redundant) in JASPAR, MEME and TRANSFAC matrix format for six taxonomic groups. More information about matrix formats can be found [here](#). 

Taxonomic group	PFMs (non-redundant)	PFMs (redundant)
Vertebrates	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC]
Plants	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC]
Insects	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC]
Nematodes	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC]
Fungi	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC]
Urochordates	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC]
All CORE profiles	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC] 	<ul style="list-style-type: none"> • Individual PFMs (zip) [JASPAR MEME TRANSFAC] • Single batch file (txt) [JASPAR MEME TRANSFAC]

Downloading the free TRANSFAC database

The screenshot displays the CIS-BP Database website interface. At the top, a blue header reads "CIS-BP Database: Catalog of Inferred Sequence Binding Preferences". Below this is a navigation menu on the left with links for Home, Tools, View cart, Bulk downloads, Database stats, Contact us, Help, Update Log, FAQ, Links, and How to cite. The main content area features the CIS-BP logo and a welcome message: "Welcome to CIS-BP, the online library of transcription factors and their DNA binding motifs." There are two main search sections: "Search for a TF" with a "By Identifier" search box (examples: Gata*, YEL009C, I\$FTZ_01) and "Browse TFs / Restrict Search for TFs" with several dropdown filters: "By Model Organism", "By Any Species", "By Domain Type", "By Motif Evidence", "By Evidence Type", and "By Study". A dropdown menu for "By Evidence Type" is open, listing various methods: B1H, ChIP-chip, ChIP-seq, COMPILED, DeBoer11, DNaseI footprinting, EMSA, High-throughput SELEX SAGE, HocoMoco, PBM, PBM, CSA and or DIP-chip, SELEX, and Transfac (which is highlighted in blue). At the bottom, a footer indicates the database was last updated on Apr 5th, 2015 (Build 1.02) and provides current content statistics: 6559 motifs, 59998 TFs with at least one experiment, and a total of 167081 TFs from 263 families in 340 experiments.

<http://cisbp.cabr.utoronto.ca>

Today's exercise

- Search the ENCODE website, and download data
- Search the Cistrome website, and download data
- (Search JASPAR)