

# Public data resources

Stockholm, November 9 2018

Jakub Orzechowski Westholm

Long-term bioinformatics support

NBIS, SciLifeLab, Stockholm University



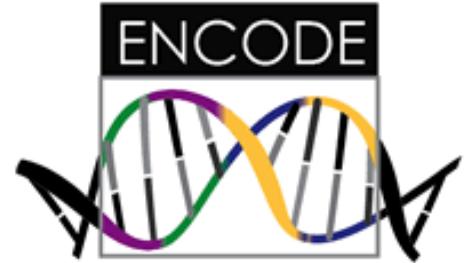
# This lecture

- 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

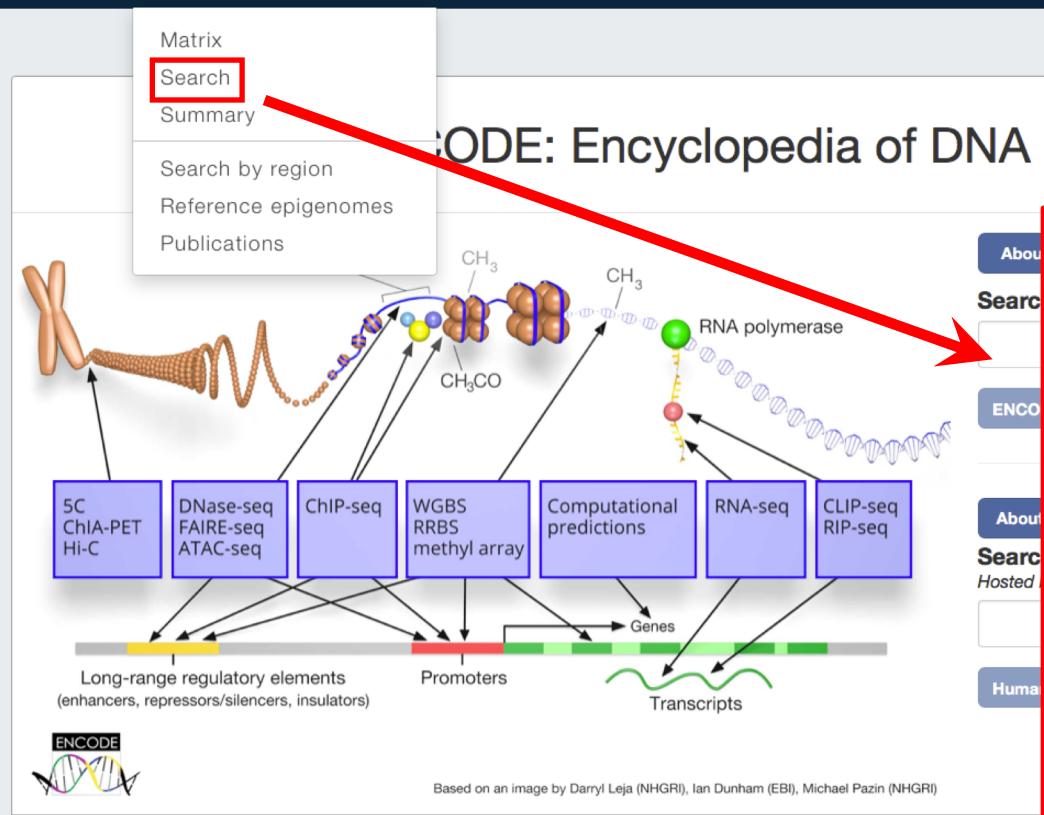
# Public data can be very useful

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

# The ENCODE project



- **Encyclopedia Of DNA Elements:** <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
  - ..
- 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**

p300 homo sapiens

**Showing 25 of 100 results**

**ChIP-seq of A549**  
*Homo sapiens A549 treated with 100 nM dexamethasone for 5 hours*  
Target: EP300  
Lab: Tim Reddy, Duke  
Project: GGR

**ChIP-seq of A549**  
*Homo sapiens A549 treated with 100 nM dexamethasone for 30 minutes*  
Target: EP300  
Lab: Tim Reddy, Duke  
Project: GGR

**ChIP-seq of A549**  
*Homo sapiens A549 treated with 100 nM dexamethasone for 15 minutes*  
Target: EP300  
Lab: Tim Reddy, Duke  
Project: GGR

**ChIP-seq of A549**  
*Homo sapiens A549 treated with 100 nM dexamethasone for 3 hours*  
Target: EP300  
Lab: Tim Reddy, Duke  
Project: GGR

**ChIP-seq of A549**  
*Homo sapiens A549 treated with 100 nM dexamethasone for 1 hour*  
Target: EP300  
Lab: Tim Reddy, Duke  
Project: GGR

**ChIP-seq of A549**  
*Homo sapiens A549*  
Target: EP300  
Lab: Tim Reddy, Duke  
Project: GGR

**Data Type**

|                     |     |
|---------------------|-----|
| Dataset             | 100 |
| Experiment          | 81  |
| Series              | 18  |
| ReferenceEpigenome  | 17  |
| FileSet             | 1   |
| Project             | 1   |
| TreatmentTimeSeries | 1   |

- See fewer

**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

+ See more...

**audit.NOT\_COMPLIANT.category!**

- poor library complexity
- insufficient read depth
- control insufficient read depth

**Experiment**  
ENCSR610RKF

**Experiment**  
ENCSR260WCE

**Experiment**  
ENCSR788VKG

**Experiment**  
ENCSR047EVQ

**Experiment**  
ENCSR358ELZ

**Experiment**  
ENCSR886OEO

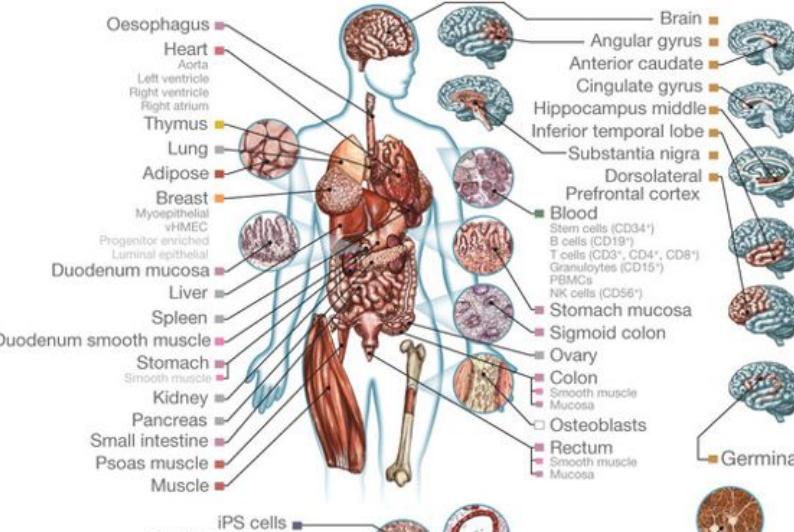
| Raw sequencing data           |                      |                               |                    |                  |                 |                 |            |              |                       |                       |
|-------------------------------|----------------------|-------------------------------|--------------------|------------------|-----------------|-----------------|------------|--------------|-----------------------|-----------------------|
| Isogenic replicate            | Library              | Accession                     | File type          | Run type         | Read            | Lab             | Date added | File size    | Audit status          | File status           |
| 1                             | ENCLB408ZKM          | <a href="#">ENCFF192EWP</a> ⓘ | fastq              | SE51nt           |                 | Tim Reddy, Duke | 2016-10-19 | 436 MB       | ✓                     | <span>released</span> |
| 2                             | ENCLB473RRB          | <a href="#">ENCFF685TOJ</a> ⓘ | fastq              | SE51nt           |                 | Tim Reddy, Duke | 2016-10-19 | 471 MB       | ✓                     | <span>released</span> |
| 3                             | ENCLB283CWG          | <a href="#">ENCFF860OTI</a> ⓘ | fastq              | SE51nt           |                 | Tim Reddy, Duke | 2016-10-19 | 368 MB       | ✓                     | <span>released</span> |
| Processed data                |                      |                               |                    |                  |                 |                 |            |              |                       |                       |
| Accession                     | File type            | Output type                   | Isogenic replicate | Mapping assembly | Lab             | Date added      | File size  | Audit status | File status           |                       |
| <a href="#">ENCFF141LII</a> ⓘ | bigBed<br>narrowPeak | peaks                         | 1                  | GRCh38           | Tim Reddy, Duke | 2017-10-19      | 507 kB     | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF507PNL</a> ⓘ | bam                  | alignments                    | 1                  | GRCh38           | Tim Reddy, Duke | 2016-10-19      | 423 MB     | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF891WYF</a> ⓘ | bed<br>narrowPeak    | peaks                         | 1                  | GRCh38           | Tim Reddy, Duke | 2016-10-20      | 139 kB     | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF028JFF</a> ⓘ | bigWig               | signal of unique reads        | 1                  | GRCh38           | Tim Reddy, Duke | 2016-10-19      | 99.6 MB    | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF598IVG</a> ⓘ | bed<br>narrowPeak    | peaks                         | 2                  | GRCh38           | Tim Reddy, Duke | 2016-10-20      | 164 kB     | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF992TSS</a> ⓘ | bigWig               | signal of unique reads        | 2                  | GRCh38           | Tim Reddy, Duke | 2016-10-19      | 122 MB     | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF595LRY</a> ⓘ | bigBed<br>narrowPeak | peaks                         | 2                  | GRCh38           | Tim Reddy, Duke | 2017-10-19      | 578 kB     | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF706FEI</a> ⓘ | bam                  | alignments                    | 2                  | GRCh38           | Tim Reddy, Duke | 2016-10-19      | 512 MB     | ✓            | <span>released</span> |                       |
| <a href="#">ENCFF852JEQ</a> ⓘ | bigWig               | signal of unique reads        | 3                  | GRCh38           | Tim Reddy, Duke | 2016-10-19      | 96.5 MB    | ✓            | <span>released</span> |                       |

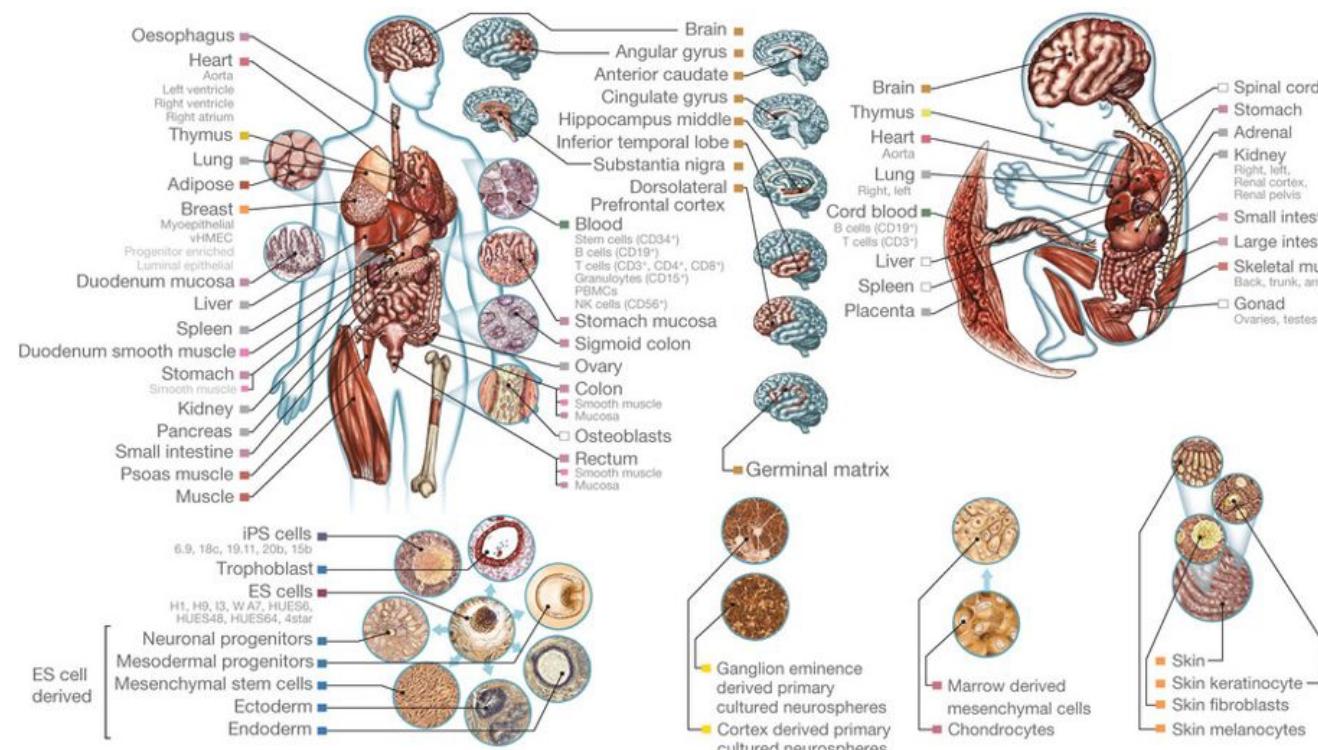
- Downloads:

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

# 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.A detailed anatomical diagram of the human body showing various tissues and organs. Labels include: Oesophagus, Heart, Aorta, Left ventricle, Right ventricle, Right atrium, Thymus, Lung, Adipose, Breast, Myoepithelial vHMEC, Progenitor enriched Luminol epithelial, Duodenum mucosa, Liver, Spleen, Duodenum smooth muscle, Stomach, Smooth muscle, Kidney, Pancreas, Small intestine, Psoas muscle, Muscle, Brain, Angular gyrus, Anterior caudate, Cingulate gyrus, Hippocampus middle, Inferior temporal lobe, Substantia nigra, Dorsolateral prefrontal cortex, Blood, Stem cells (CD34+), B cells (CD19+), T cells (CD3+, CD4+, CD8+), Granulocytes (CD15+), NK cells (CD56+), Stomach mucosa, Sigmoid colon, Ovary, Colon, Smooth muscle, Mucosa, Osteoblasts, Rectum, Smooth muscle, Mucosa, Germinal ma, iPS cells, 6.9, 18c, 19.11, 20b, 15b, Trophoblast, ES cells, H1, H9, 13, WA7, HUES6, HUES48, HUES64, 4star, Neuronal progenitors. Arrows point from the labels to specific tissue samples or detailed views of organs like the brain and heart.



## Index of /EdaccData/Release-9/sample-experiment

|  |                   |   |
|--|-------------------|---|
| <a href="#">..</a>   |                   |   |
| <a href="#">Adipose_Derived_Mesenchymal_Stem_Cell_Cultured_...</a> | 04-Jun-2013 16:04 | - |
| <a href="#">Adipose_Nuclei/</a>                                    | 04-Jun-2013 16:04 | - |
| <a href="#">Adipose_Tissue/</a>                                    | 04-Jun-2013 16:07 | - |
| <a href="#">Adrenal_Gland/</a>                                     | 04-Jun-2013 16:07 | - |
| <a href="#">Adult_Kidney/</a>                                      | 04-Jun-2013 16:07 | - |
| <a href="#">Adult_Liver/</a>                                       | 04-Jun-2013 16:08 | - |
| <a href="#">Aorta/</a>   | 04-Jun-2013 16:08 | - |
| <a href="#">Bladder/</a>   | 04-Jun-2013 16:08 | - |
| <a href="#">Bone_Marrow_Derived_Mesenchymal_Stem_Cell_Cultu...</a> | 04-Jun-2013 16:05 | - |
| <a href="#">Brain_Angular_Gyrus/</a>                               | 15-May-2013 19:03 | - |
| <a href="#">Brain_Anterior_Caudate/</a>                            | 04-Jun-2013 16:05 | - |
| <a href="#">Brain_Cerebellum/</a>                                  | 12-Jun-2013 16:29 | - |
| <a href="#">Brain_Cingulate_Gyrus/</a>                             | 04-Jun-2013 16:06 | - |
| <a href="#">Brain_Germinal_Matrix/</a>                             | 04-Jun-2013 16:08 | - |
| <a href="#">Brain_Hippocampus_Middle/</a>                          | 04-Jun-2013 16:06 | - |
| <a href="#">Brain_Inferior_Temporal_Lobe/</a>                      | 04-Jun-2013 16:06 | - |
| <a href="#">Brain_Mid_Frontal_Lobe/</a>                            | 04-Jun-2013 16:06 | - |
| <a href="#">Brain_Substantia_Nigra/</a>                            | 04-Jun-2013 16:07 | - |
| <a href="#">Breast_Fibroblast_Primary_Cells/</a>                   |                   |   |
| <a href="#">Breast_Luminal_Epithelial_Cells/</a>                   |                   |   |
| <a href="#">Breast_Myoepithelial_Cells/</a>                        |                   |   |
| <a href="#">Breast_Stem_Cells/</a>                                 |                   |   |
| <a href="#">Breast_vHMEC/</a>                                      |                   |   |
| <a href="#">CD14_Primary_Cells/</a>                                |                   |   |
| <a href="#">CD15_Primary_Cells/</a>                                |                   |   |
| <a href="#">CD19_Primary_Cells/</a>                                |                   |   |
| <a href="#">CD20_Primary_Cells/</a>                                |                   |   |
| <a href="#">CD34_Cultured_Cells/</a>                               |                   |   |
| <a href="#">CD34_Primary_Cells/</a>                                |                   |   |
| <a href="#">CD3_Primary_Cells/</a>                                 |                   |   |
| <a href="#">CD4+_CD25+_Treg_Primary_Cells/</a>                     |                   |   |
| <a href="#">CD4+_CD25-_Naive_Primary_Cells/</a>                    |                   |   |
| <a href="#">CD4+_CD25-_CD45RO+_Memory_Primary_Cells/</a>           |                   |   |
| <a href="#">CD4+_CD25-_IL17+_PMA-Ionomycin_stimulated_Tl...</a>    |                   |   |
| <a href="#">CD4+_CD25-_IL17-_PMA-Ionomycin_stimulated_M...</a>     |                   |   |
| <a href="#">CD4+_CD25-_Th_Primary_Cells/</a>                       | 12-Feb-2013 16:15 | - |
| <a href="#">CD4+_CD25int_CD127+_Tmem_Primary_Cells/</a>            | 12-Feb-2013 16:15 | - |
| <a href="#">CD4_Memory_Primary_Cells/</a>                          | 04-Jun-2013 16:08 | - |
| <a href="#">CD4_Naive_Primary_Cells/</a>                           | 04-Jun-2013 16:08 | - |

## Index of /EdaccData/Release-9/sample-experiment

|                                   |                   |   |
|-----------------------------------|-------------------|---|
| <a href="#">..</a>                | 04-Jun-2013 16:06 | - |
| <a href="#">Bisulfite-Seq/</a>    | 04-Jun-2013 16:06 | - |
| <a href="#">ChIP-Seq_Input/</a>   | 04-Jun-2013 16:06 | - |
| <a href="#">Histone_H3K27ac/</a>  | 04-Jun-2013 16:06 | - |
| <a href="#">Histone_H3K27me3/</a> | 04-Jun-2013 16:06 | - |
| <a href="#">Histone_H3K36me3/</a> | 04-Jun-2013 16:06 | - |
| <a href="#">Histone_H3K4me1/</a>  | 04-Jun-2013 16:06 | - |
| <a href="#">Histone_H3K4me3/</a>  | 04-Jun-2013 16:06 | - |
| <a href="#">Histone_H3K9ac/</a>   | 04-Jun-2013 16:06 | - |
| <a href="#">Histone_H3K9me3/</a>  | 04-Jun-2013 16:06 | - |
| <a href="#">-DNA_Seq/</a>         | 04-Jun-2013 16:06 | - |

## Index of /EdaccData/Release-9/sample-experiment/Brain\_Hippocampus\_Middle/Histone\_H3K27ac/

|  |                   |      |
|--|-------------------|------|
| <a href="#">..</a>   |                   |      |
| <a href="#">BI.Brain_Hippocampus_Middle.H3K27ac.112.bed.gz</a> | 21-Dec-2012 16:20 | 197M |
| <a href="#">BI.Brain_Hippocampus_Middle.H3K27ac.112.wig.gz</a> | 21-Dec-2012 16:44 | 27M  |
| <a href="#">BI.Brain_Hippocampus_Middle.H3K27ac.149.bed.gz</a> | 28-Dec-2011 18:05 | 332M |
| <a href="#">BI.Brain_Hippocampus_Middle.H3K27ac.149.wig.gz</a> | 21-Dec-2011 21:12 | 36M  |
| <a href="#">BI.Brain_Hippocampus_Middle.H3K27ac.150.bed.gz</a> | 09-Mar-2012 22:27 | 312M |
| <a href="#">BI.Brain_Hippocampus_Middle.H3K27ac.150.wig.gz</a> | 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):

P300

Search

Options ▾

## Species

All

Homo sapiens

Mus musculus

## Biological Sources

All

786-O

A549

Acute myeloid leukemia

B cell lymphoma

B Lymphocyte

## Factors

All

EP300

## Results

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

&lt;&lt; Prev

Page

1



of 11



Next &gt;

**Inspector**

|                    |   |
|--------------------|---|
| Title:             | Treatment<br>• LP1_DMSO_P300  |
| ID:                | GEO or ENCODE: GSM1847311<br>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<br>PMID: 26731516 |
| Factor:            | EP300   |
| Biological Source: | Cell Line: LP-1   |

**Comment Selection**

**Tools**

|            |           |                          |                         |
|------------|-----------|--------------------------|-------------------------|
| QC reports | QC motifs | Get top putative targets | Check a putative target |
|------------|-----------|--------------------------|-------------------------|

**QC**

|  |                             |
|--|-----------------------------|
| 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%                       |

**Sample**

**% Top 5k peaks Phastcon Conservation Profiles**

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

# R interfaces

 Bioconductor  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About Search:

Home » Bioconductor 3.8 » Software Packages » ENCODEExplorer

## ENCODEExplorer

platforms all rank 599 / 1649 posts 0 in Bioc 3.5 years  
build ok updated before release

DOI: [10.18129/B9.bioc.ENCODEExplorer](https://doi.org/10.18129/B9.bioc.ENCODEExplorer) [f](#) [t](#)

### A compilation of ENCODE metadata

Bioconductor version: Release (3.8)

This package allows user to quickly access ENCODE project files metadata and give access to helper functions to query the ENCODE rest api, download ENCODE datasets and save the database in SQLite format.

Author: Charles Joly Beauparlant <charles.joly-beauparlant at crchul.ulaval.ca>, Audrey Lemaçon <lemacon.audrey at ulaval.ca>, Louis Gendron <louisg.212 at gmail.com> Astrid-Louise Deschenes, <astrid-louise.deschenes at crchudequebec.ulaval.ca> and Arnaud Droit <arnaud.droit at crchudequebec.ulaval.ca>

Maintainer: Charles Joly Beauparlant <charles.joly-beauparlant at crchul.ulaval.ca>

Citation (from within R, enter `citation("ENCODEExplorer")`):

Beauparlant CJ, Lemaçon A, Deschenes LGA, , Droit A (2018). *ENCODEExplorer: A compilation of ENCODE metadata*. R package version 2.8.0.

### Installation

To install this package, start R and enter:

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

### Documentation

To view documentation for the version of this package installed in your system, start R and enter:

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

Documentation »

Bioconductor

- Package [vignettes](#) and manuals.
- [Workflows](#) for learning and use.
- [Course and conference](#) material.
- [Videos](#).
- Community [resources](#) and [tutorials](#).

R / [CRAN](#) packages and [documentation](#)

Support »

Please read the [posting guide](#). Post questions about Bioconductor to one of the following locations:

- [Support site](#) - for questions about Bioconductor packages
- [Bioc-devel](#) mailing list - for package developers

 Bioconductor  
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

Home Install Help Developers About Search:

Home » Bioconductor 3.8 » Software Packages » DeepBlueR

## DeepBlueR

platforms all rank 724 / 1649 posts 2 / 1 / 1 / 0 in Bioc 2 years  
build error updated before release

DOI: [10.18129/B9.bioc.DeepBlueR](https://doi.org/10.18129/B9.bioc.DeepBlueR) [f](#) [t](#)

### DeepBlueR

Bioconductor version: Release (3.8)

Accessing the DeepBlue Epigenetics Data Server through R.

Author: Felipe Albrecht, Markus List

Maintainer: Felipe Albrecht <felipe.albrecht at mpi-inf.mpg.de>, Markus List <markus.list at mpi-inf.mpg.de>

Citation (from within R, enter `citation("DeepBlueR")`):

Albrecht F, List M, Bock C, Lengauer T (2017). "DeepBlueR: large-scale epigenomic analysis in R." *Bioinformatics*. doi: [10.1093/bioinformatics/btx099](https://doi.org/10.1093/bioinformatics/btx099), [https://academic.oup.com/bioinformatics/article/3045024/DeepBlueR..pdf](https://academic.oup.com/bioinformatics/article-pdf/doi/10.1093/bioinformatics/btx099/10931981/btx099.pdf), <https://academic.oup.com/bioinformatics/article/3045024/DeepBlueR..>

### Installation

To install this package, start R and enter:

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

### Documentation

To view documentation for the version of this package installed in your system, start R and enter:

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

HTML R Script The DeepBlue epigenomic data server - R package  
PDF Reference Manual

Documentation »

Bioconductor

- Package [vignettes](#) and manuals.
- [Workflows](#) for learning and use.
- [Course and conference](#) material.
- [Videos](#).
- Community [resources](#) and [tutorials](#).

R / [CRAN](#) packages and [documentation](#)

### Support

Please read the [posting guide](#). Post questions about Bioconductor to one of the following locations:

- [Support site](#) - for questions about Bioconductor packages
- [Bioc-devel](#) mailing list - for package developers

# 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

JASPAR 2018

Home

About

Search

Browse JASPAR CORE

Browse Collections

Tools

RESTful API

Download Data

Matrix Clusters

Genome Tracks

Sox2

Examples: SPI1, P17676, ChIP-seq, Homo sapiens

Search Q Advanced Options

Browse JASPAR CORE for six different taxonomic groups

Vertebrata Nematoda

Insecta Plantae

Fungi Urochordata

JASPAR CORE & when should it be used? Info about other collections

The JASPAR CORE contains a curated, non-redundant set of profiles, derived from published and experimentally defined transcription factor binding sites for eukaryotes. It should be used, when seeking models for specific factors or structural classes, or if experimental evidence is paramount.

The high-quality transcription factor binding profile database

Read more about JASPAR

JASPAR interactive tour

Citing JASPAR 2018

PubMed | NAR | PDF

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

[Home](#)[About](#)[Search](#)[Browse JASPAR CORE](#)[Browse Collections](#)[Tools](#)[RESTful API](#)[Download Data](#)[Matrix Clusters](#)[Genome Tracks](#)

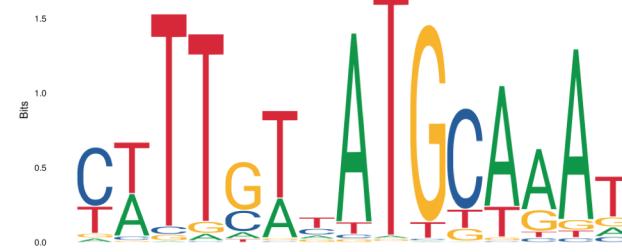
## Detailed information of matrix profile MA0142.1

[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 | 33 |
|-----|-----|----|----|----|-----|----|-----|----|---|----|------|----|----|----|

|     |    |    |    |    |     |    |     |    |    |      |     |    |     |    |
|-----|----|----|----|----|-----|----|-----|----|----|------|-----|----|-----|----|
| 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 | 53 |
|-----|-----|-----|------|------|----|-----|-----|----|------|-----|-----|-----|----|----|

### 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

[CORE PFM](#)s[TFFMs](#)[JASPAR collections \(PFMs\)](#)[Matrix clustering](#)[SQL dump](#)[Other data](#)

**i** 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)   |
|--------------------------|--|--|
| <b>Vertebrates</b>       | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> |
| <b>Plants</b>            | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> |
| <b>Insects</b>           | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> |
| <b>Nematodes</b>         | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> |
| <b>Fungi</b>             | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> |
| <b>Urochordates</b>      | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> |
| <b>All CORE profiles</b> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> | <ul style="list-style-type: none"><li>Individual PFM (zip) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li><li>Single batch file (txt) [<a href="#">JASPAR</a>   <a href="#">MEME</a>   <a href="#">TRANSFAC</a>]</li></ul> |

# Downloading the free TRANSFAC database

CIS-BP Database: Catalog of Inferred Sequence Binding Preferences

**CIS-BP**

Welcome to CIS-BP, the online library of transcription factors and their DNA binding motifs.

Search for a TF

By Identifier   
(e.g. Gata\*, YEL009C, I\$FTZ\_01)

Browse TFs / Restrict Search for TFs

By Model Organism  By Any Species  By Domain Type  By Motif Evidence  By Evidence Type  By Study  Database Build

B1H  
ChIP-chip  
ChIP-seq  
COMPILED  
DeBoer11  
DNaseI footprinting  
EMSA  
High-throughput SELEX SAGE  
HocoMoco  
PBM  
PBM, CSA and or DIP-chip  
SELEX  
Transfac

available!

Last updated: Apr 5th, 2015 Database Build 1.02

Current content: 6559 motifs. 59998 TFs with at least one motif available (from 167081 TFs from 263 families in 340 experiments), out of a total of 167081 TFs from 263 families in 340

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

# Todays exercise

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