

Web applications for human epigenomics

EPIGEN2018

2018-03-14

Guillaume Devailly

@G_Devailly



Slides available online. Feel free to share.

Why web applications?

- easy to access
 - easy to use (?)
 - fancy and attracting
- **Allow users to do things they wouldn't have done otherwise**

Why web applications?

- easy to access
- easy to use (?)
- fancy and attracting

→ Allow users to do things they wouldn't have done otherwise

but...

- someone has to develop and maintain them...
- (almost) no flexibility
- apps appear and disappear with a high turnover

Types of web-apps for epigenomics

1. Data portals
2. Genome browsers
3. Compare experiments
4. Analyse

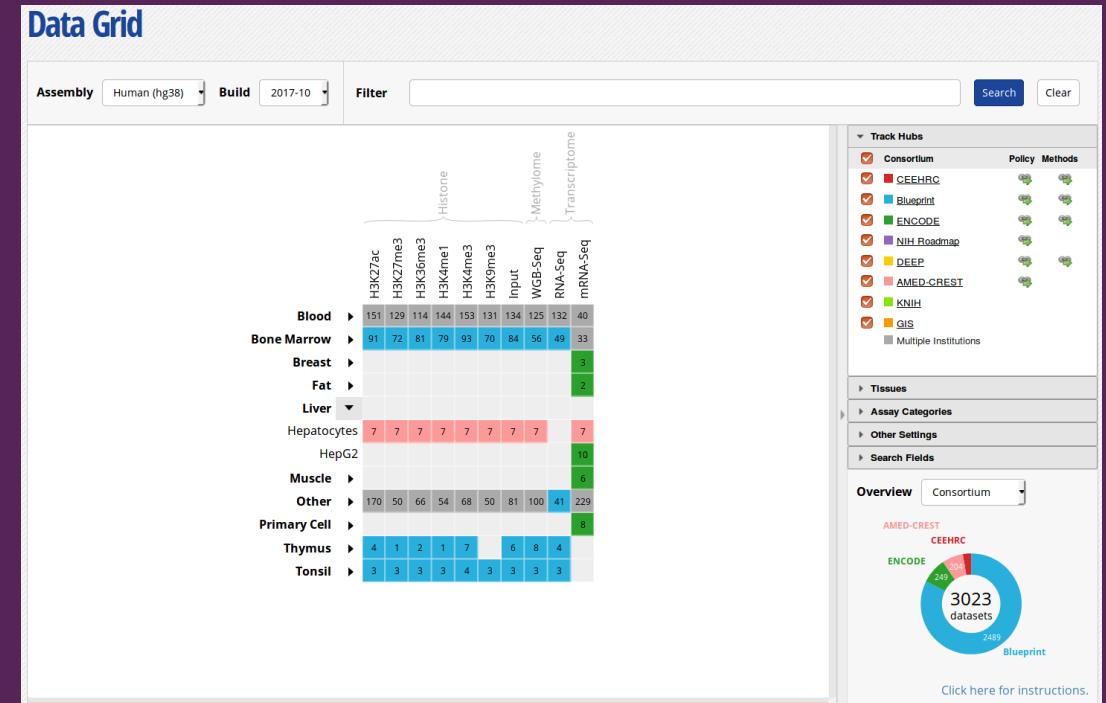
Types of web-apps for epigenomics

1. Data portals

2. Genome browsers

3. Compare experiments

4. Analyse



Data portals

- An eagle-eye view of available data.
- Easier to use than SRA / GEO searches

Data portals

A precursor: ENCODE @ UCSC

 Encyclopedia of DNA Elements at UCSC 2003 - 2012

Human Data at UCSC

- Downloads
- Experiment Matrix
- Search
- Genome Browser (hg19)
- Experiment List
- Cell Types

Mouse Data at UCSC

- Downloads
- Experiment Matrix
- Search
- Genome Browser (mm9)
- Experiment List
- Cell Types
- Metadata Terms
- Registered Variables
- Antibodies

About

The [Encyclopedia of DNA Elements \(ENCODE\)](#) Consortium is an international collaboration of research groups funded by the National Human Genome Research Institute ([NHGRI](#)). The goal of ENCODE is to build a comprehensive parts list of functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements that control cells and circumstances in which a gene is active.

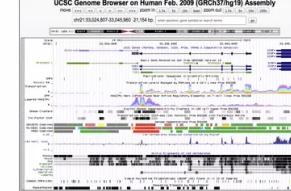
ENCODE results from 2007 and later are available from the ENCODE Project Portal, [encodeproject.org](#). This covers data generated during the two production phases 2007-2012 and 2013-present. The ENCODE Project Portal also hosts additional ENCODE access tools, and ENCODE project pages including up-to-date information about data releases, publications, and upcoming tutorials.

UCSC coordinated data for the ENCODE Consortium from its inception in 2003 (Pilot phase) to the end of the first 5 year phase of whole-genome data production in 2012. All data produced by ENCODE investigators and the results of ENCODE analysis projects from this period are hosted in the UCSC Genome browser and database. Explore ENCODE data using the image links below or via the left menu bar. **All ENCODE data at UCSC are freely available for download and analysis.**

Explore ENCODE data (2003 - 2012) at UCSC



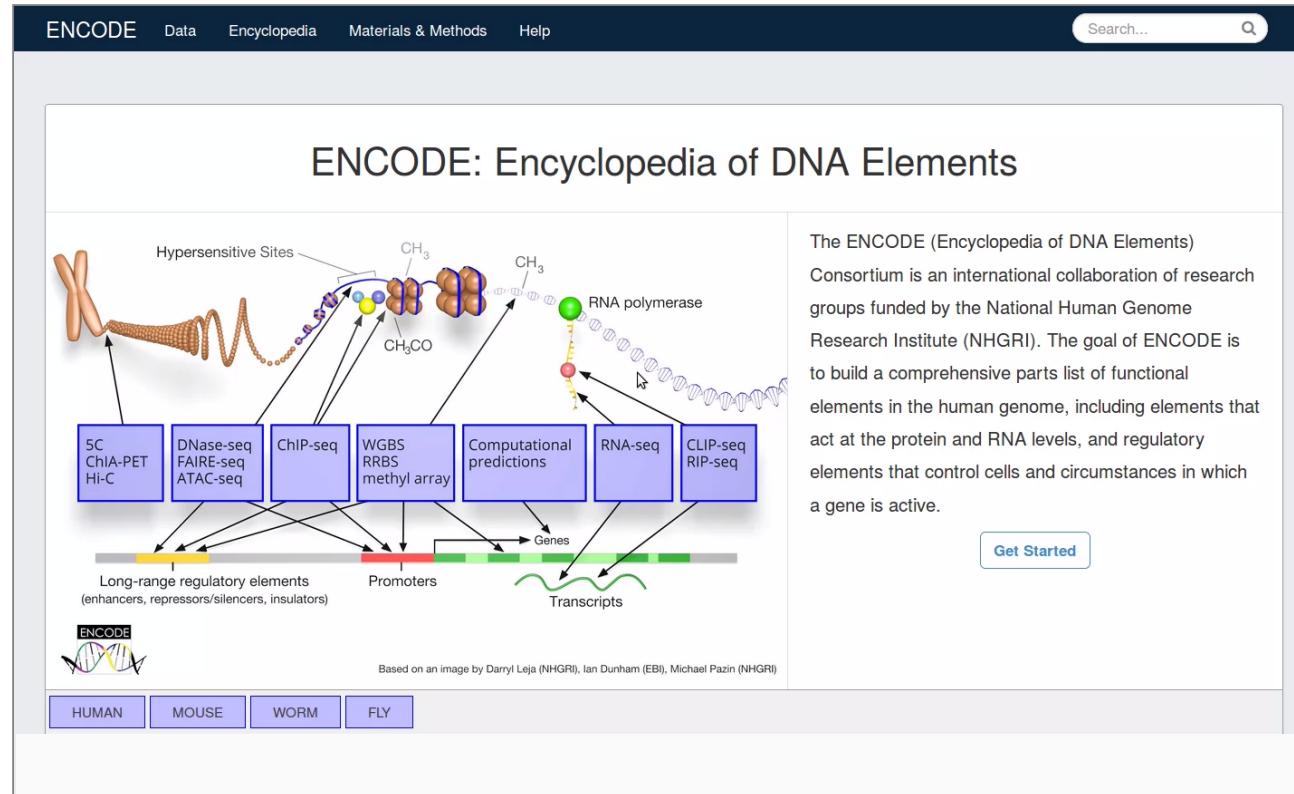
View ENCODE data (2003 - 2012) in the UCSC Genome Browser



unmaintained since 2012

Data portals

Modern ENCODE data portal

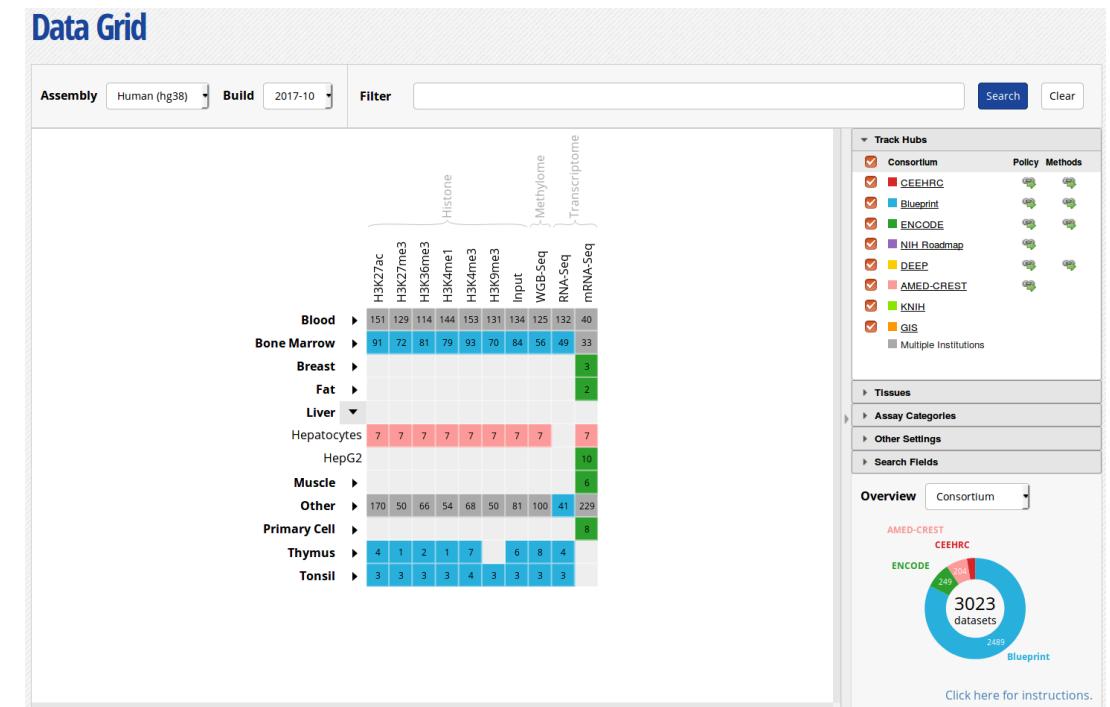
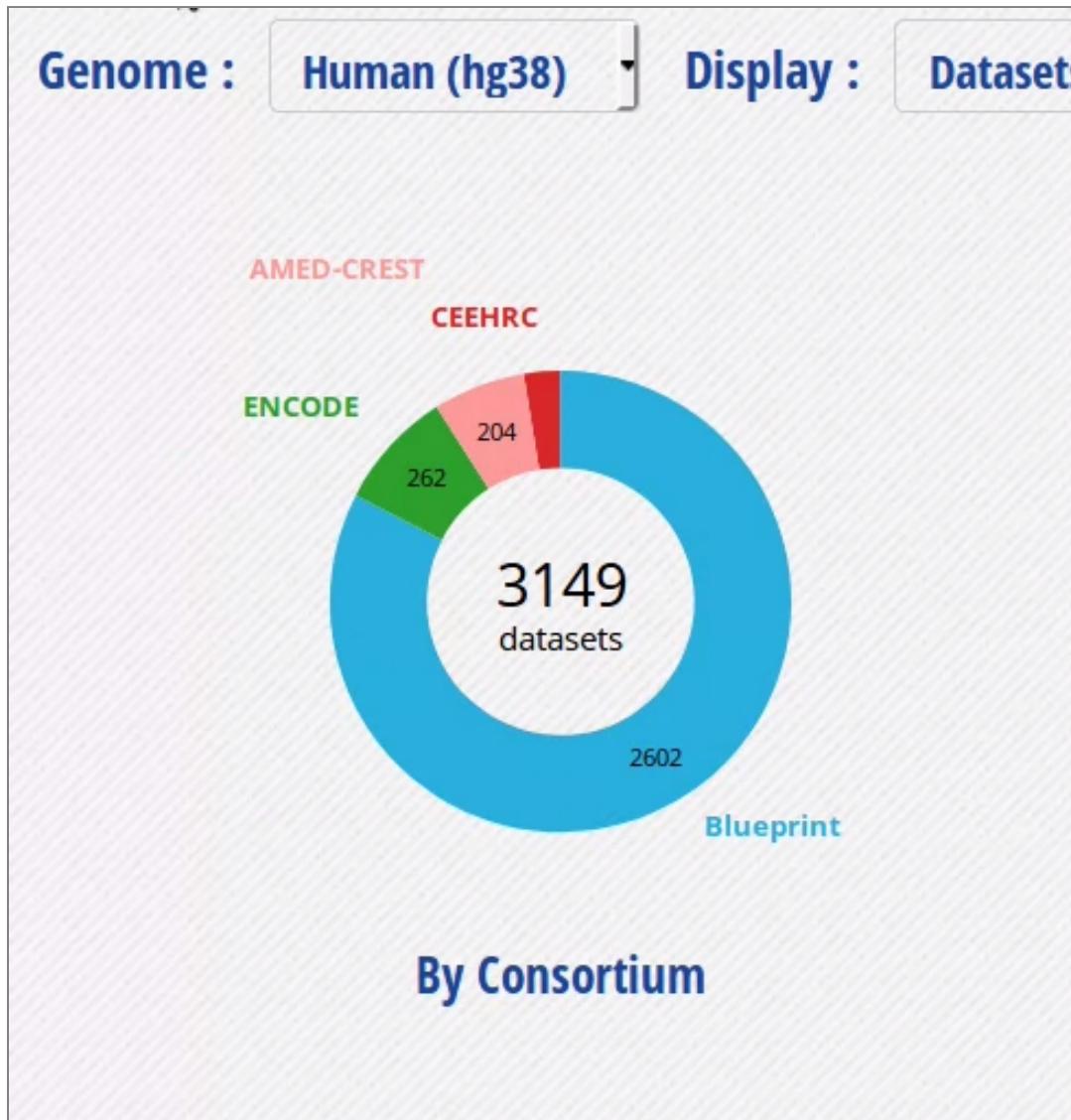


Other consortia



Clear filters		Experiments	Files	Datasets	Data access	Methods	Help
		First Previous 1 2 3 4 5 ... Next Last					
		<input type="text" value="Search"/>					
Experiment	+/-						
H3K27ac		394					
H3K4me3		300					
ChIP Input		283					
RNA-Seq		274					
H3K4me1		257					
Cell type	+/-						
monocyte		74					
myeloid cell		65					
CD14-positive, CD16-negative classical monocyte		34					
macrophage		27					
naive B cell		24					
Cell line	+/-						
BL-2		1					

International Human Epigenome Consortium (IHEC)



Data portals

Portals browsing SRA/ENA public datasets:

- codex.stemcells.cam.ac.uk: TF & HisMod ChIP-seq, RNA-seq, DNase-seq in Haematopoietic Cells, or Embryonic Stem Cells
- cistrome.org: TF & HisMod & DNase/ATAC, all cell types & tissues
- ngs-qc.org: Almost everything (TF, HisMod, RNA, MeDP, ...), all cell types & tissues, assess dataset quality

→ Finding data is easy

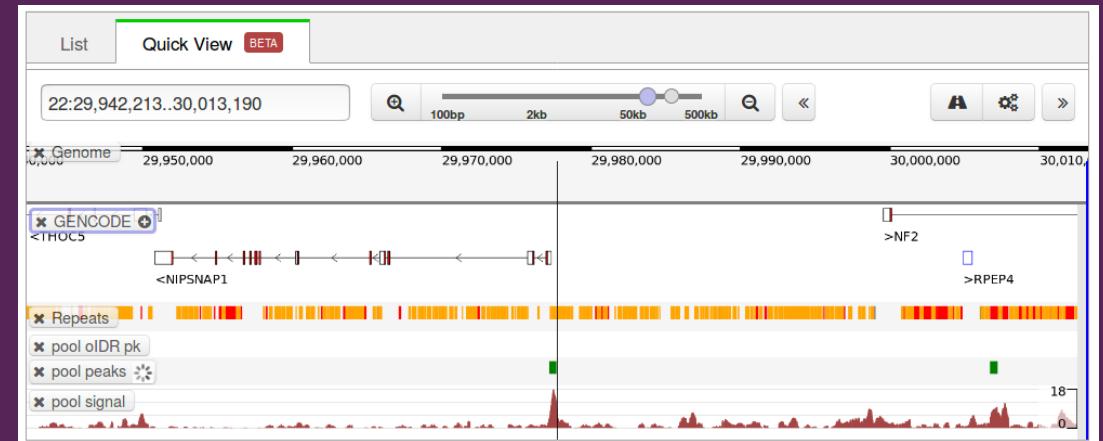
Types of web-apps for epigenomics

1. Data portals

2. Genome browsers

3. Compare experiments

4. Analyse



Online Genome browsers

- quick look at the data
- looking at a few regions manually
- Sometimes ♥ *integrated to the Data portal* ♥

ENCODE quick view genome browser

ENCODE Data Encyclopedia Materials & Methods Help Search... 

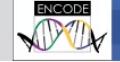
EXPERIMENTS / CHIP-SEQ / HOMO SAPIENS / HEK293

Experiment summary for ENCSR348AGV

Supersedes ENCSR000EYD

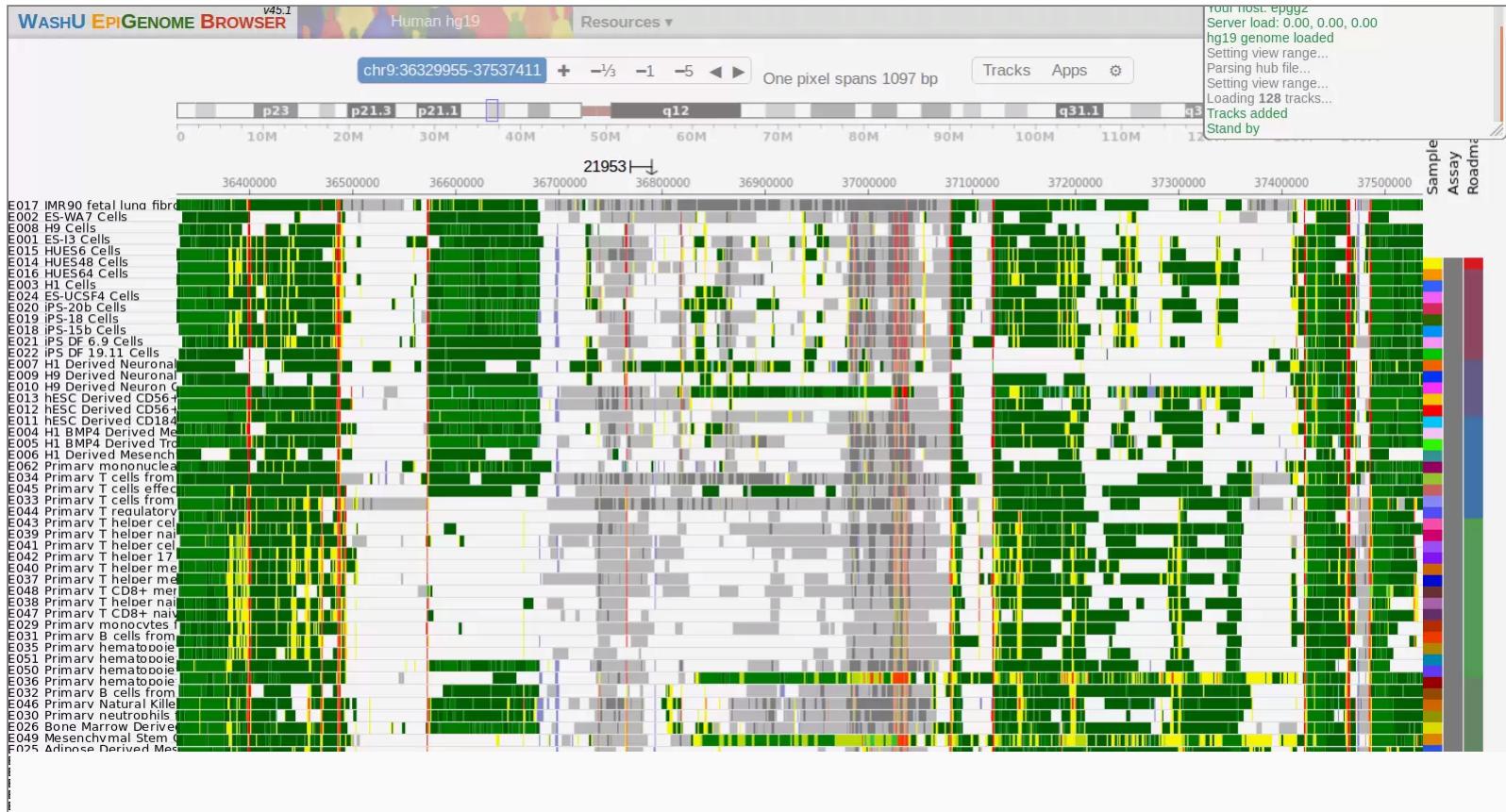
Status: released 

Summary		Attribution
Assay:	ChIP-seq	Lab: Peggy Farnham, USC
Target:	SETDB1	Award: U54HG004558 (Michael Snyder, Stanford)
Biosample summary:	<i>Homo sapiens</i> HEK293	Project: ENCODE
Biosample Type:	immortalized cell line	Date released: June 1, 2017
Replication type:	unreplicated	Submitter comment: Replaces ENCSR000EYD after it was discovered it was HEK293 and not U2OS
Description:	SETDB1 ChIP-seq on human HEK293	
Nucleic acid type:	DNA	
Lysis method:	see document	
Extraction method:	see document	
Fragmentation method:	see document	
Size selection method:	see document	

 ENCODE PHASE 2

Roadmap Chromatin states browser

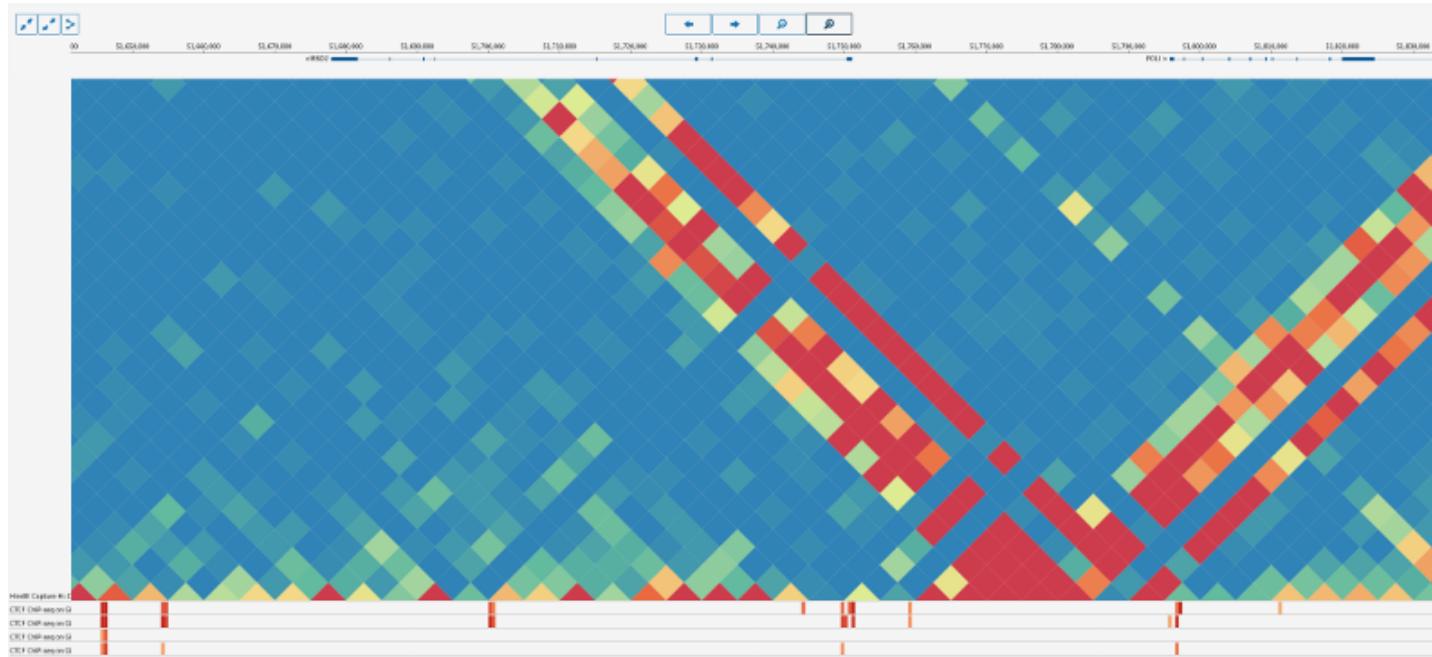
egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html



Uses the great WahsU EpiGenome Browser

Other great options

- Ensembl regulatory build
- Cistrome send any tracks to the WashU EpiBrowser in a single click
- QC Genomics is developing NAVi



Most of the time...

Old & unfriendly or New & buggy ?

The screenshot shows the Ensembl BLUEPRINT page. At the top, there's a dark blue header with the Ensembl logo and navigation links for Home, Ensembl, Ensembl Genomes, and Blog. Below the header, the word "BLUEPRINT" is prominently displayed in a large, bold, black font. Underneath it is a logo for "BL ue PR INT epigenome". A brief description follows: "The BLUEPRINT Project is a high impact FP7 aiming to produce a blueprint of haemopoietic epigenomes. Our goal is to apply highly sophisticated functional genomics analysis on a clearly defined set of primarily human samples from healthy and diseased individuals, and to provide at least 100 reference epigenomes to the scientific community. This resource-generating activity will be complemented by research into blood-based diseases, including common leukaemias and autoimmune disease (Type 1 Diabetes), by discovery and validation of epigenetic markers for diagnostic use and by epigenetic target identification. This may eventually lead to the development of novel and more individualised medical treatments." Below this, sections for "Data Reuse" and "Data access" are present. The "Data access" section includes a table:

Requirement	Access
Raw data	Data archives (EGA & ENA)
Processed data	FTP site
Data mining	BioMart
Genome browser	Genomatix browser BLUEPRINT Track Hub on the UCSC browser

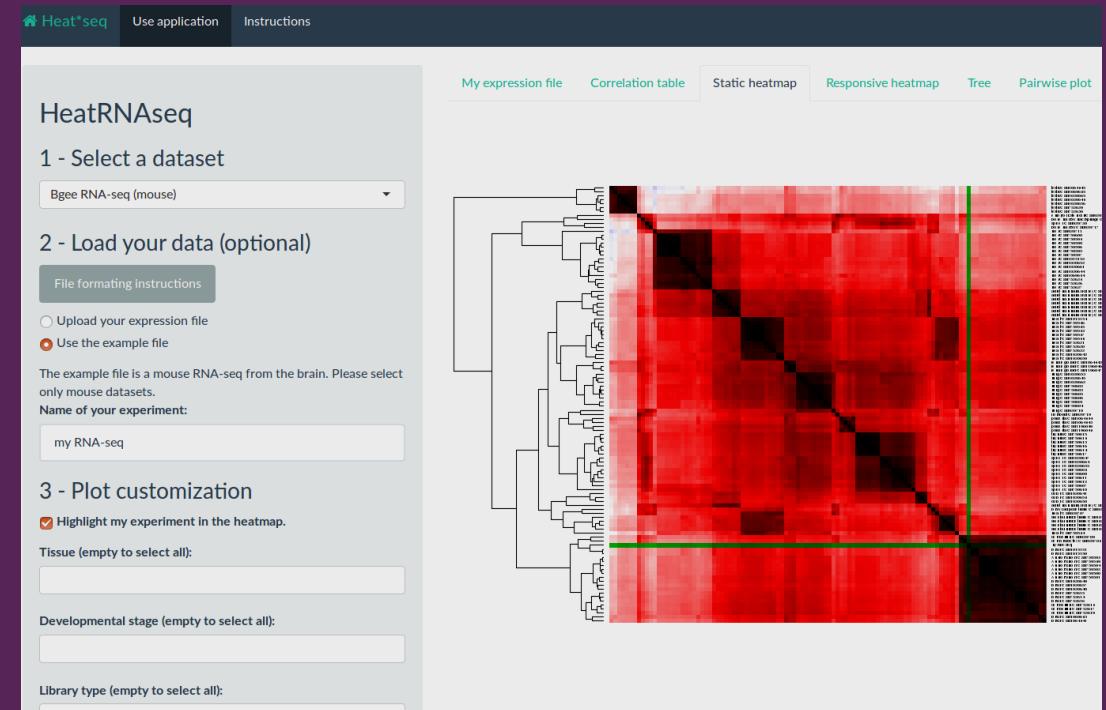
Types of web-apps for epigenomics

1. Data portals

2. Genome browsers

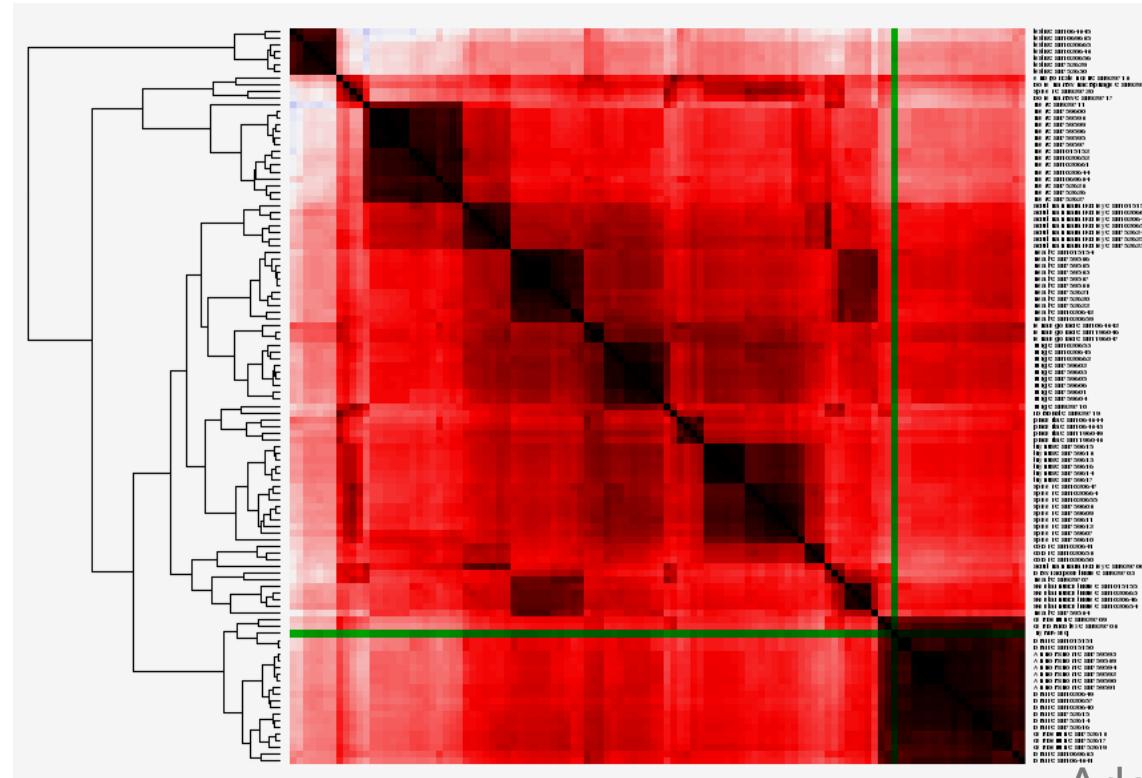
3. Compare experiments

4. Analyse



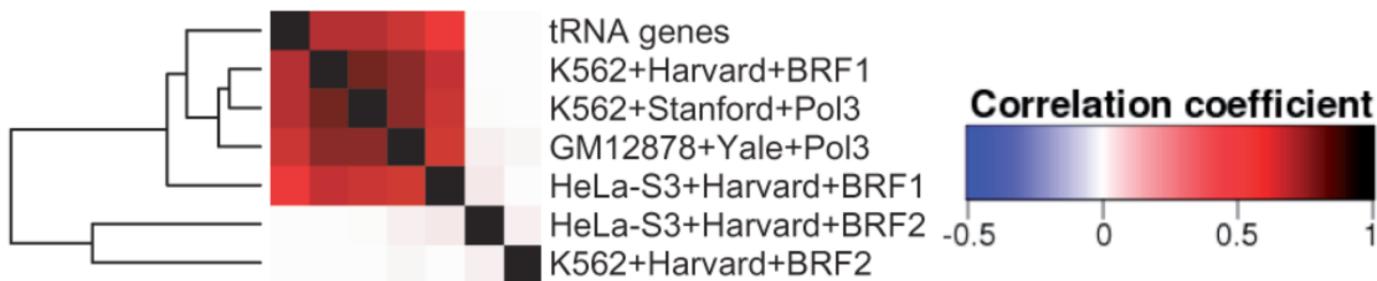
Heat*seq: genome-wide comparison of NGS experiments

- RNA-seq, TF-ChIP-seq, CAGE
- Compare a processed file you upload to all experiments in a dataset
- available at www.heatstarseq.roslin.ed.ac.uk



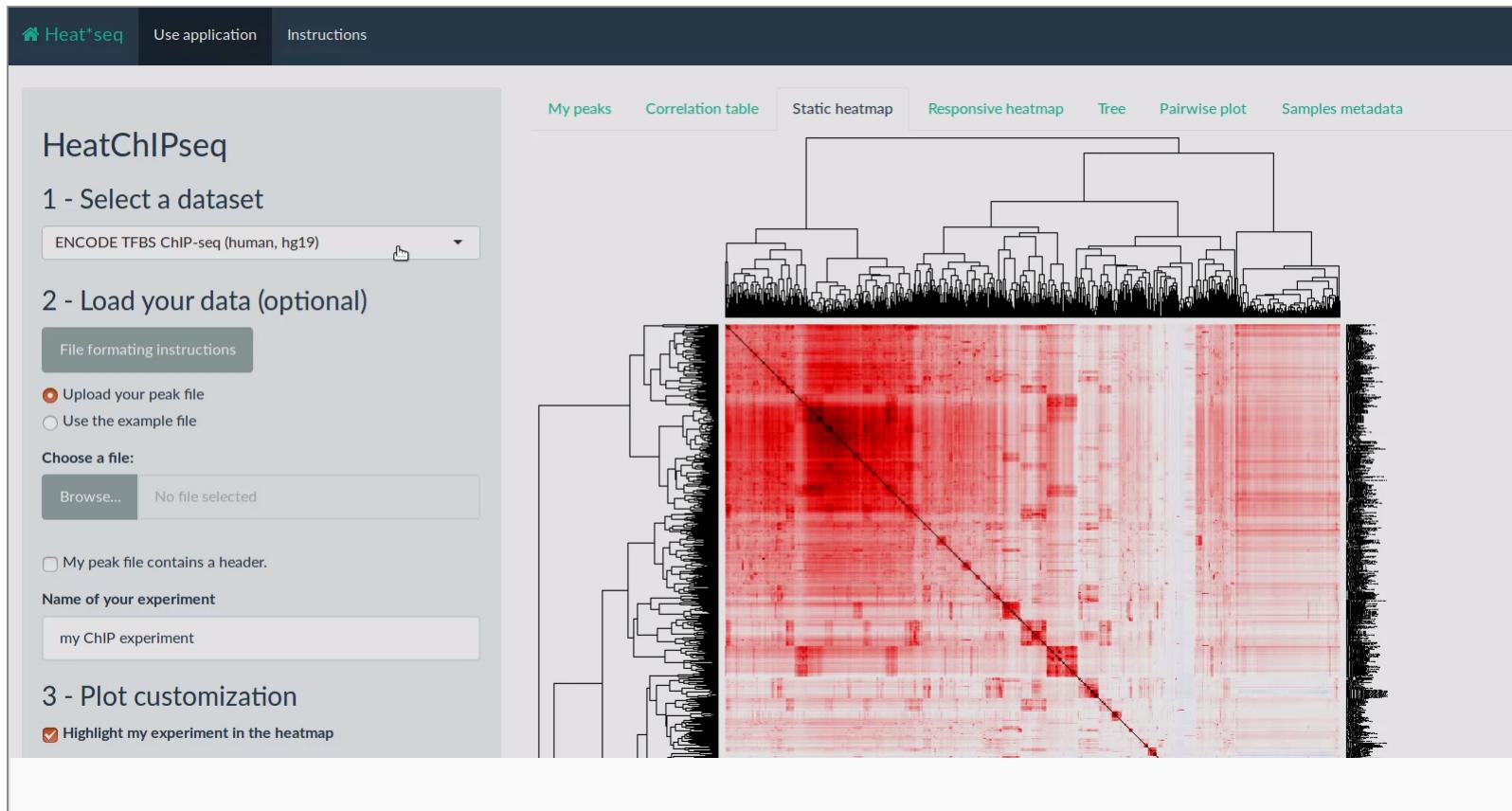
Clustered correlation heatmaps:

- Correlation of every pairs of experiment
- Clustering of the correlation matrix
- Two similar experiment will be close on the clustered correlation heatmap



Heat*seq: genome-wide comparison of NGS experiments

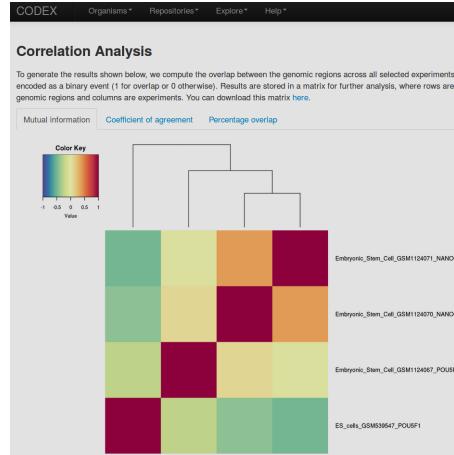
Example: ESR1 TF ChIP-seq from GEO, compared against all TF ChIP-seq from ENCODE



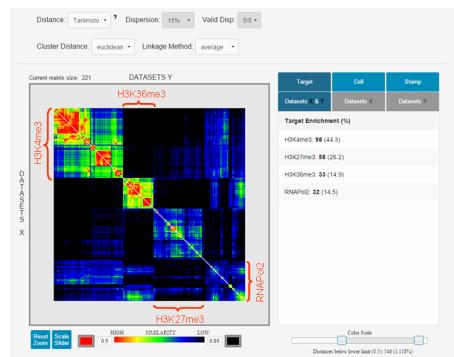
Input file: peak list, .bedformat.

Alternatives:

CODEX Correlation analysis



QC Genomics: Comparator ngs-qc.org/qcgenomics/



Types of web-apps for epigenomics

1. Data portals

2. Genome browsers

3. Compare experiments

4. Analyse



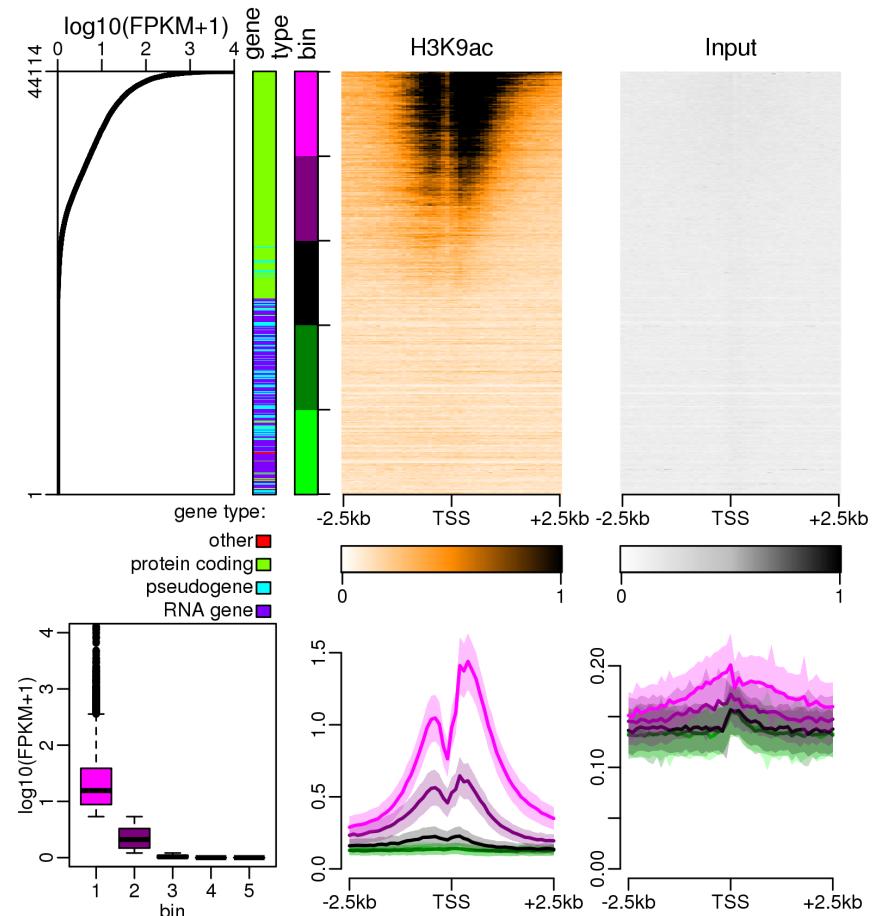
Profile Explorer for Roadmap Epigenomics

www.perepigenomics.roslin.ed.ac.uk

Using Roadmap Epigenomics data:

- Stack profiles
- Centered on a feature (here, TSS)
- Sorted according to expression (RNA-seq)

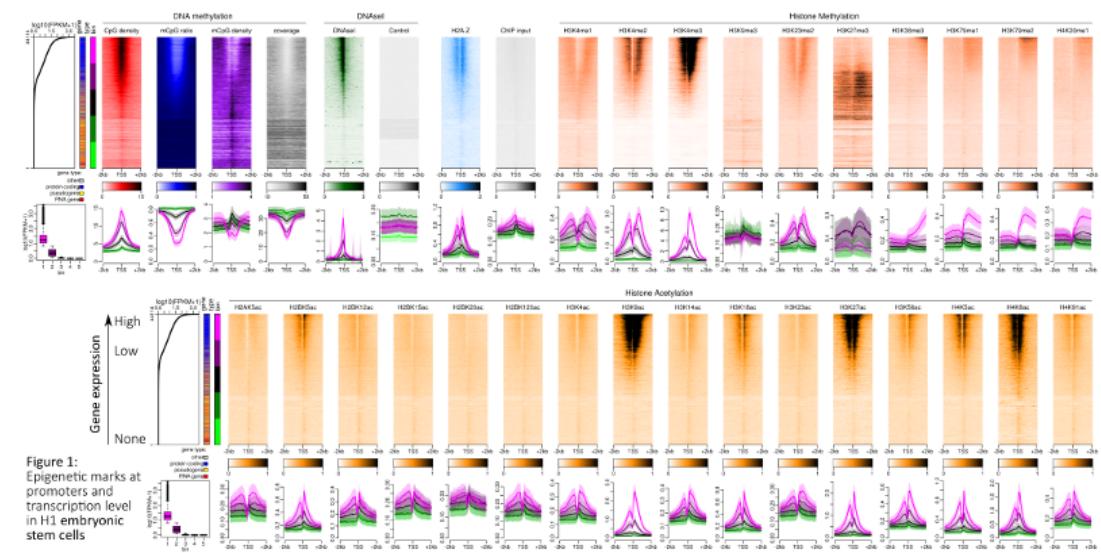
H3K9Ac, adult liver



Profile Explorer for Roadmap Epigenomics

Pre-generated *stack profiles*:

- TSS, TES, middle exons
 - By gene types (all, protein coding, lincRNA, pseudogenes, ...)
 - WGBS, DNase1, H2AZ, 10 histone methylations, 16 Histone acetylations
 - 30 cell types (cell lines + tissues)
- 9 921 plots so far



Profile Explorer for Roadmap Epigenomics

PEREpigenomics

Explore

Compare

Correlate

About

Available profiles

Select a plot:

1- Order by

- Epigenetic assay first
- Cell type first

2- Focus on:

TSS

3- Choose an assay:

DNase 1

4- Choose a cell type:

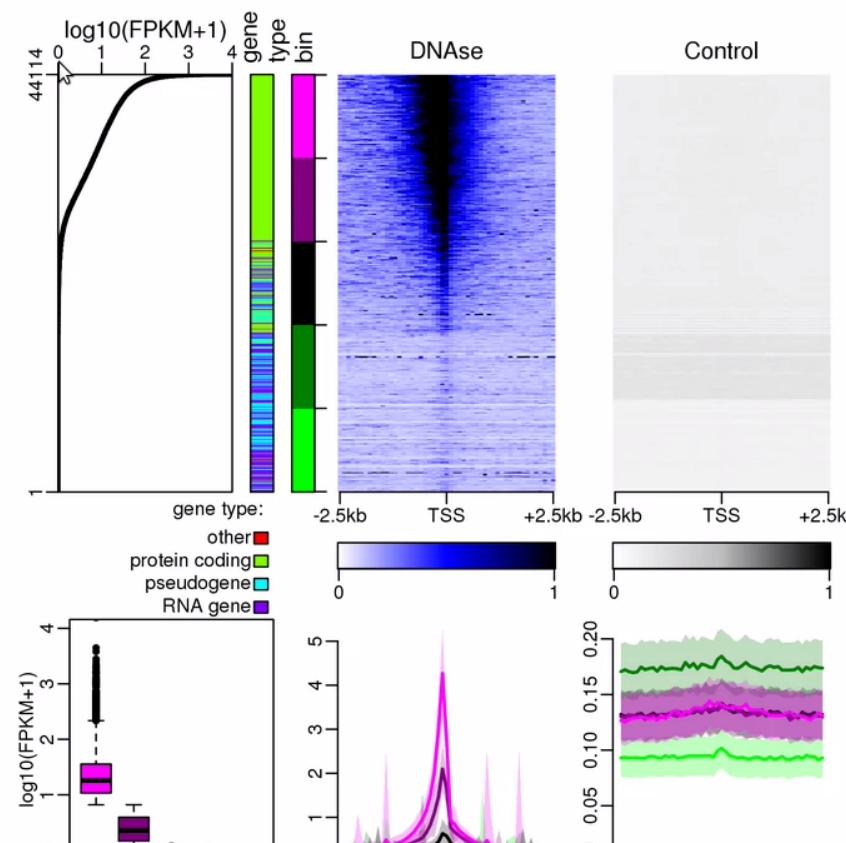
H1_Cell_Line

5- Choose a gene category:

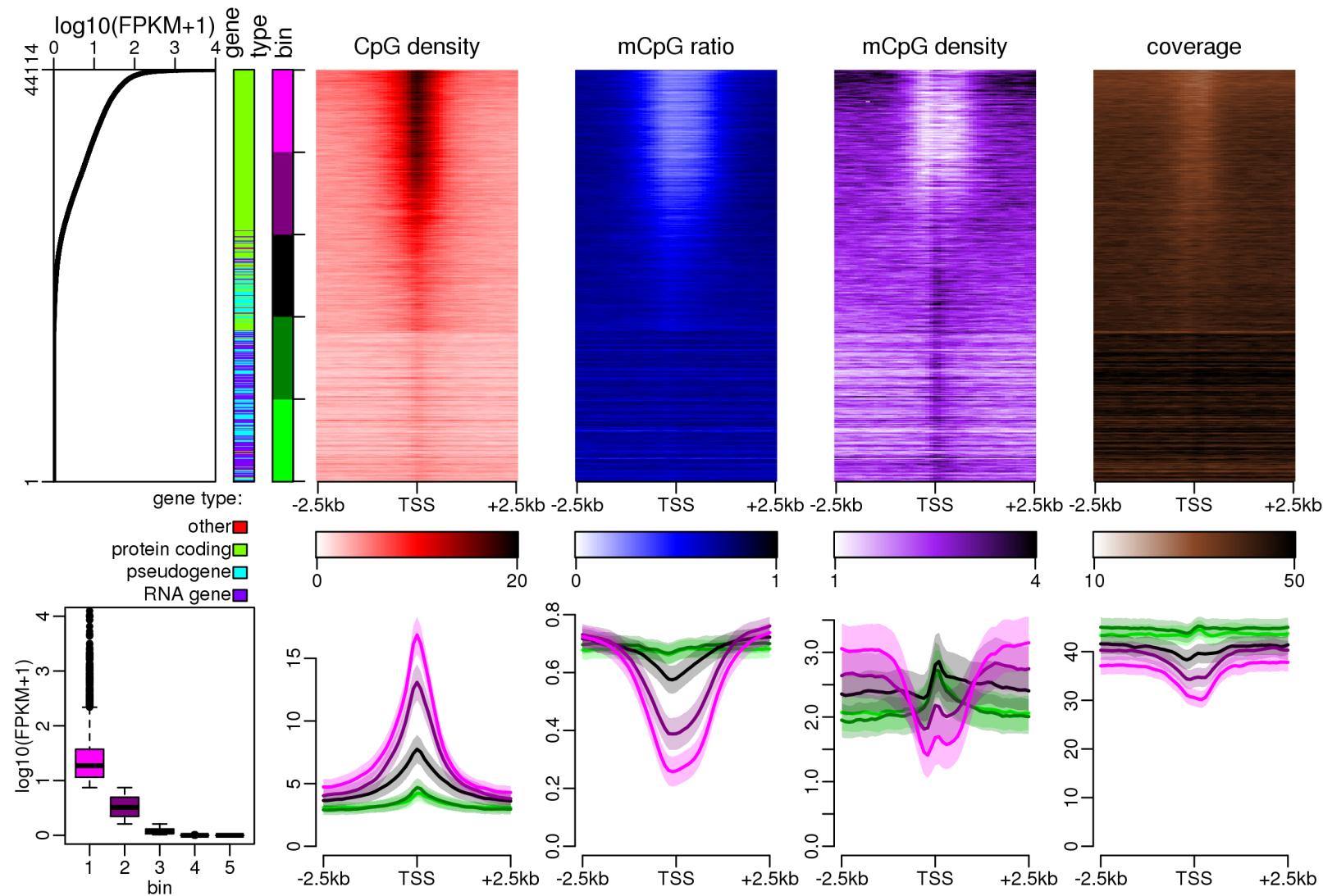
all

 Download image

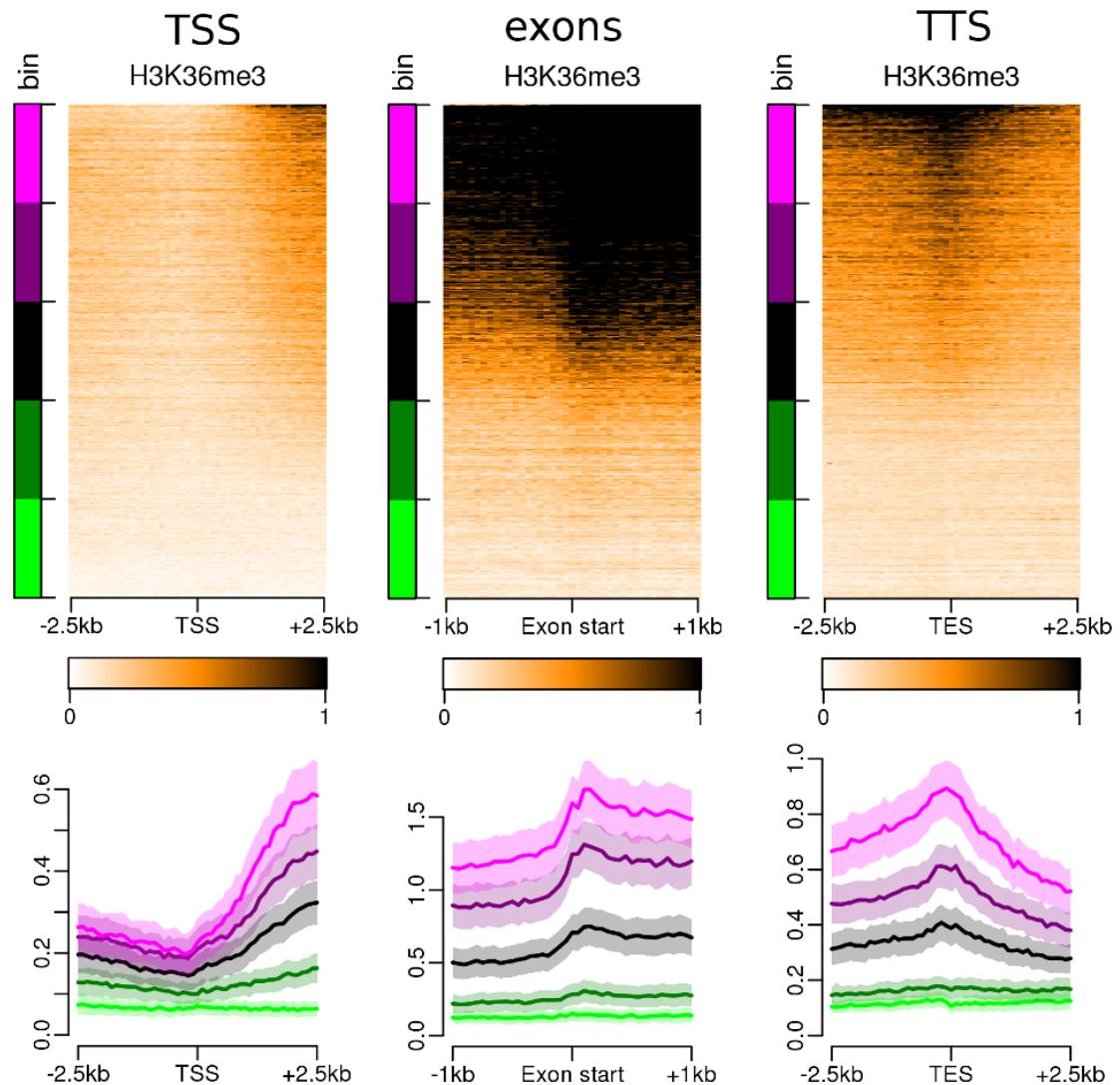
DNase 1 at TSS in H1_Cell_Line for all genes



WGBS, pancreas



H3K36me3, fetal large intestine



Alternatives

Tools from *Blueprint Epigenome*:

- *Blueprint* data analysis portal: blueprint-data.bsc.es/release_2016-08
- *DeepBlue* analysis server: deepblue.mpi-inf.mpg.de
- Dive (in early development): dive.mpi-inf.mpg.de

QC Genomics: ChromStater ngs-qc.org/qcgenomics/

Galaxy servers, notably:

- *deepTools*: deeptools.ie-freiburg.mpg.de
- *Cistrome*: cistrome.org/ap/root

Thanks

Anagha Joshi

Anna Mantsoki

Deepti Vipin

