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...

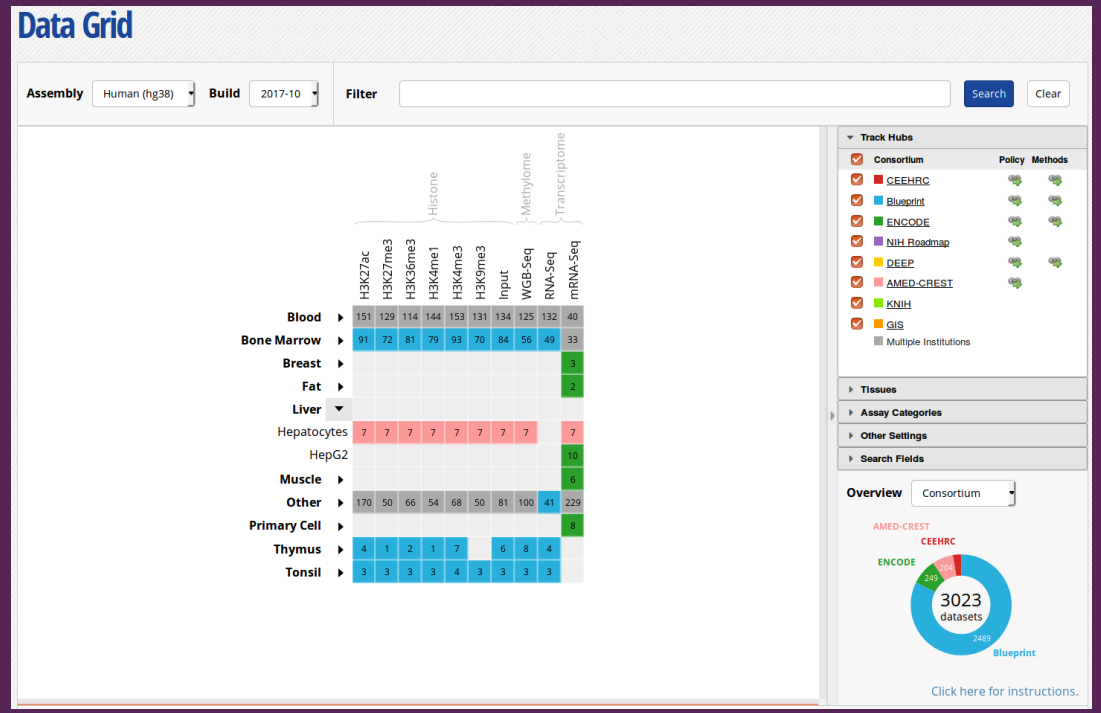
- someone has to develop and maintain them...
- (almost) no flexibility
- apps appear and disappear with a high turnover

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

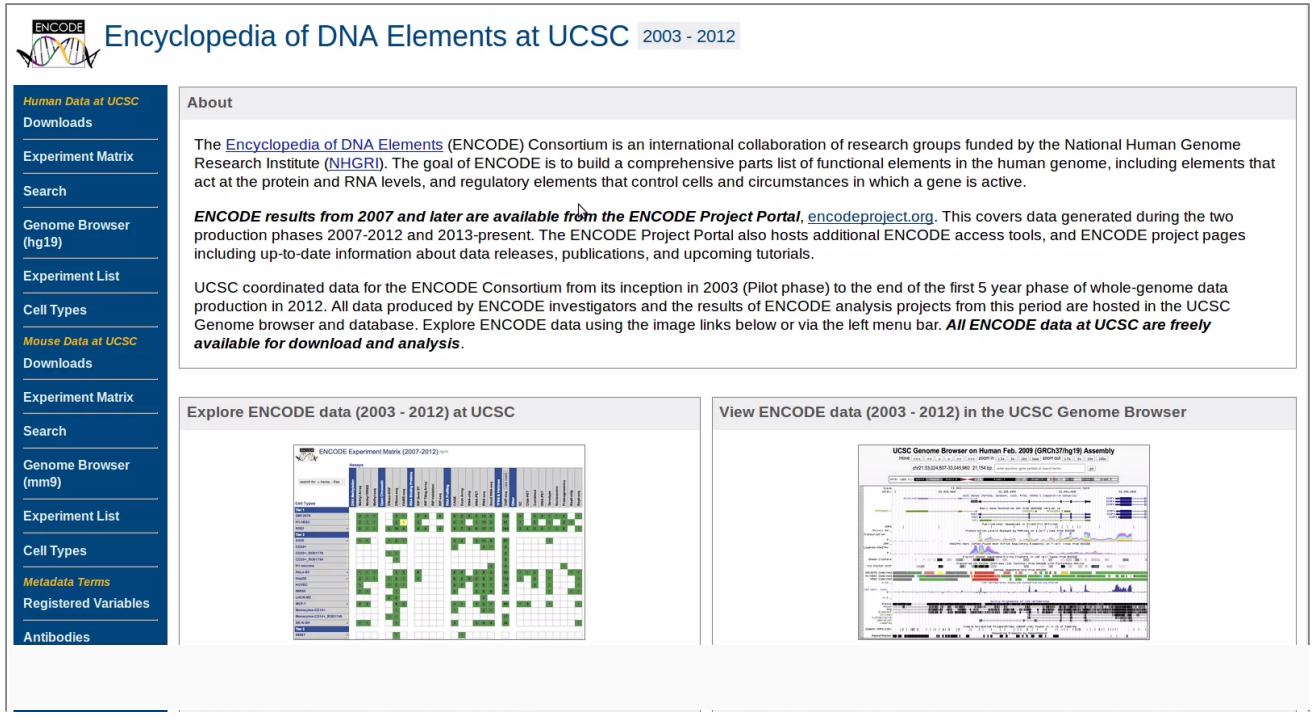


Data portals

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

Data portals

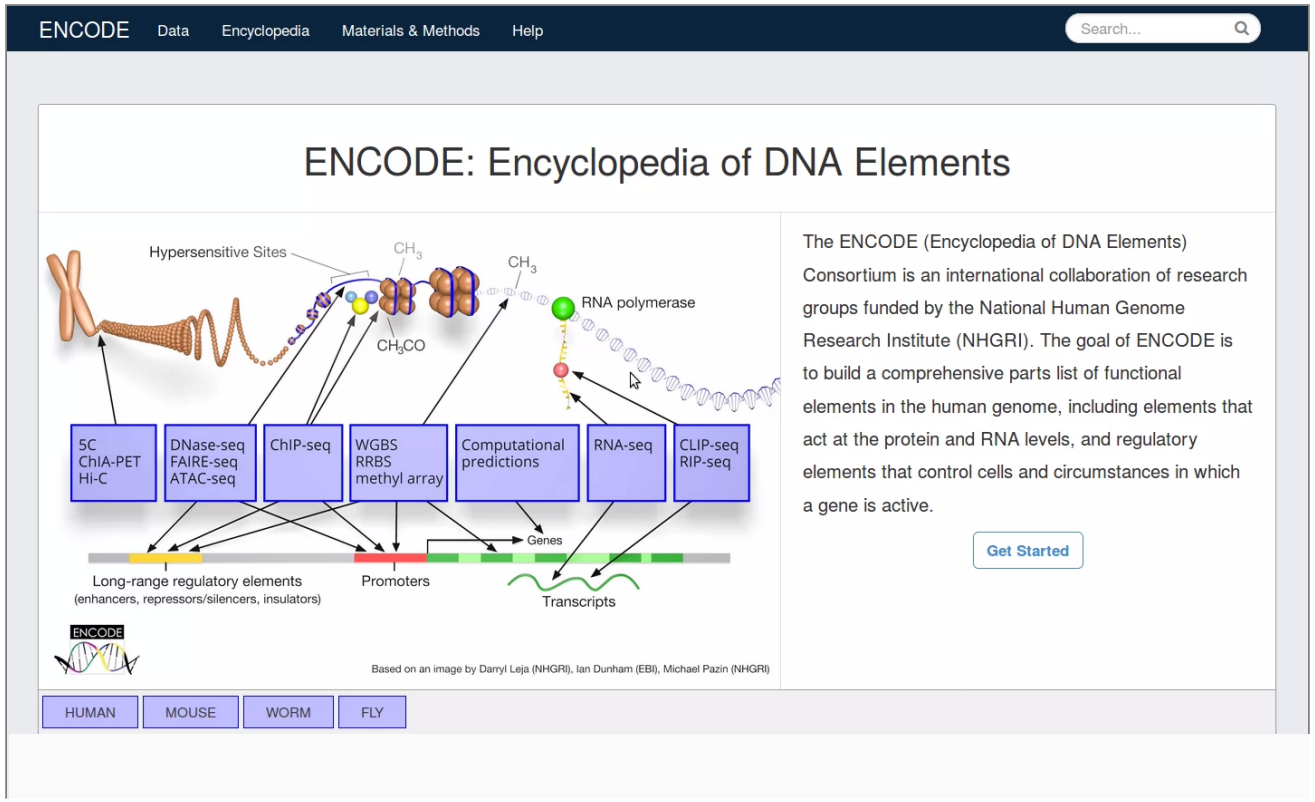
A precursor: ENCODE @ UCSC



unmaintained since 2012

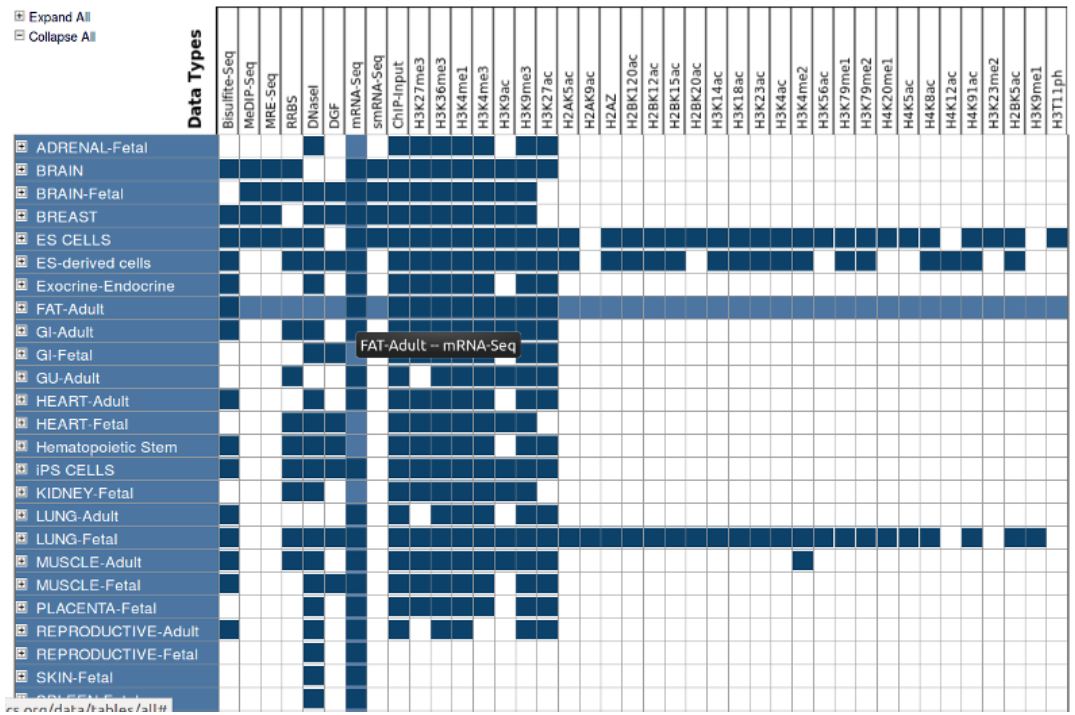
Data portals

Modern ENCODE data portal

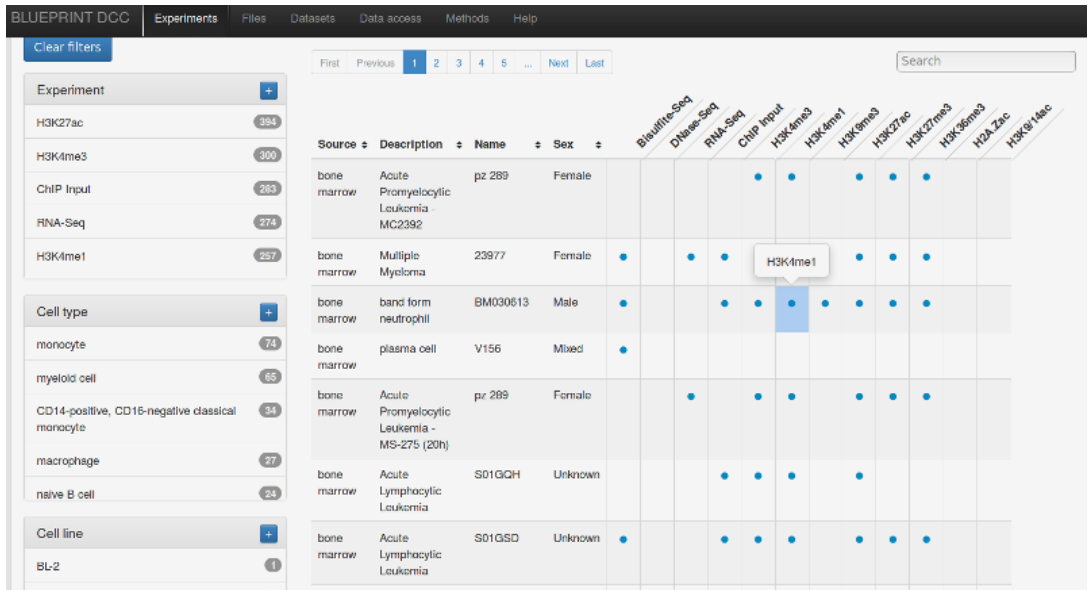


Other consortia

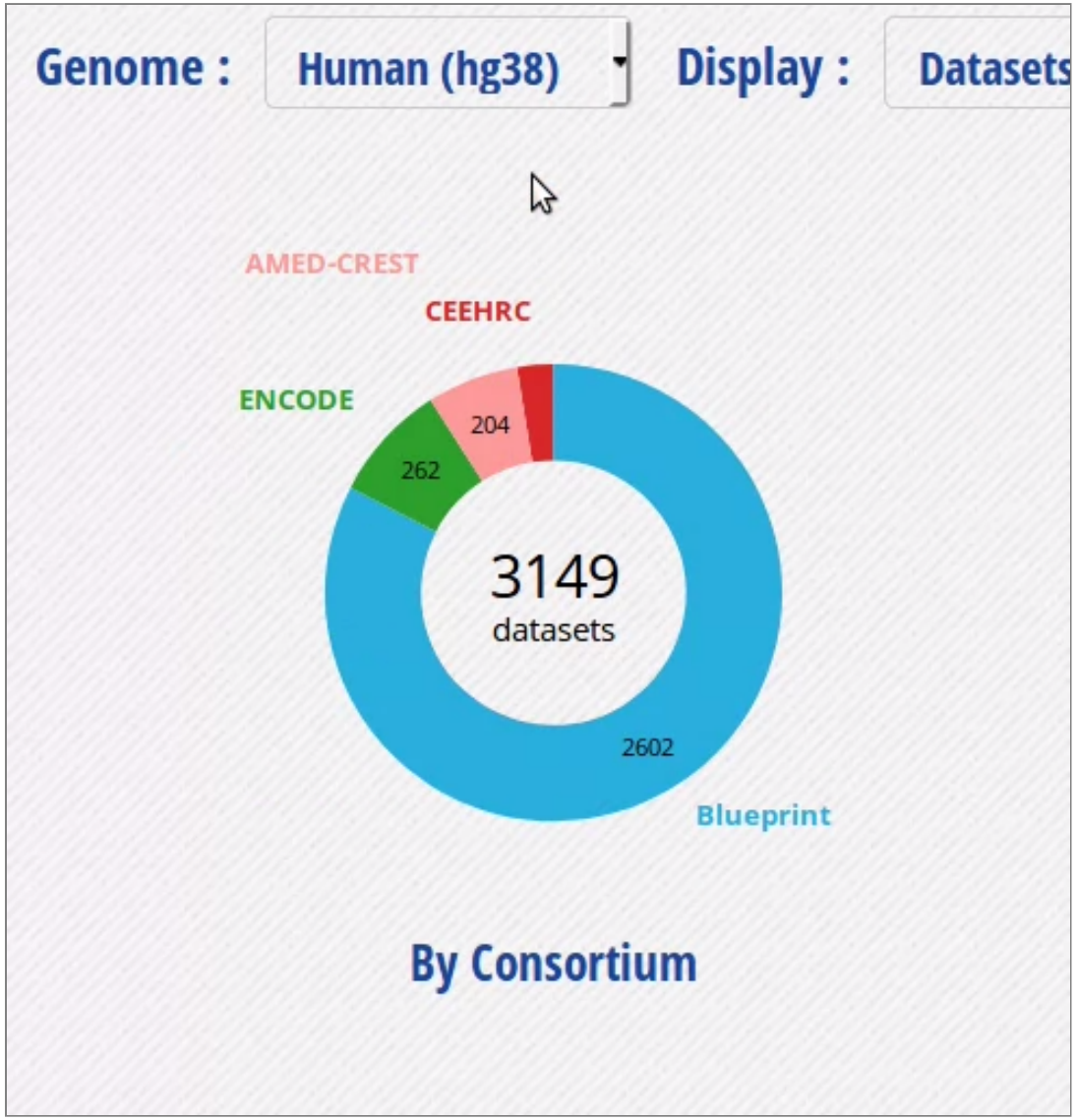
ROADMAP epigenomics PROJECT

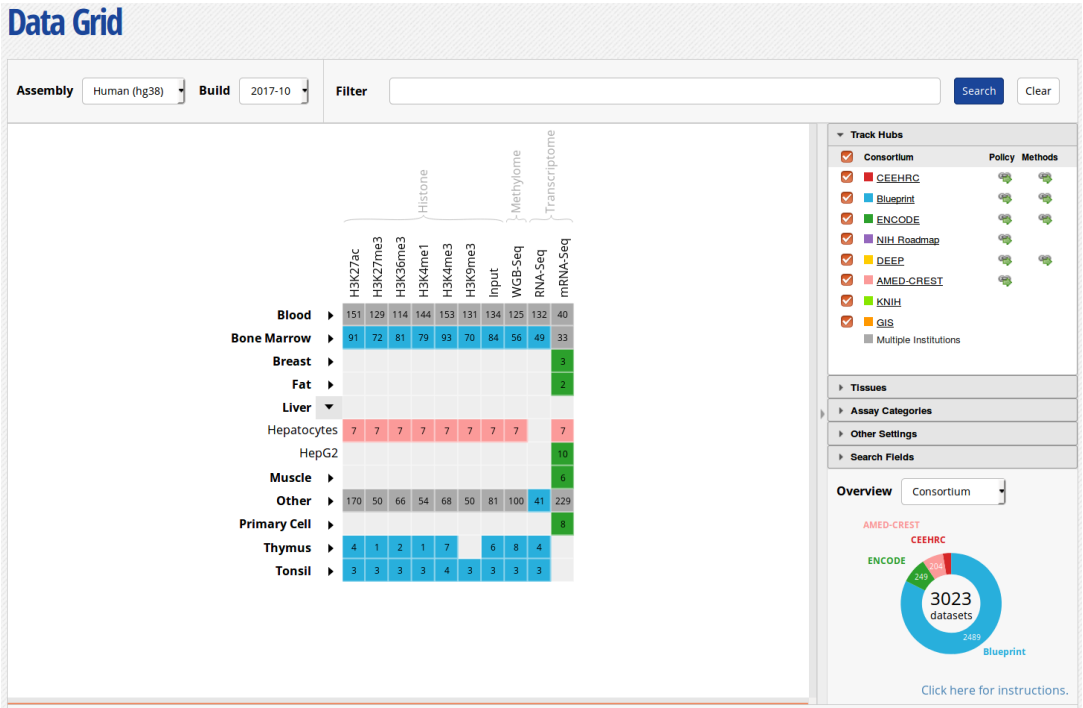


BLUEPRINT epigenome



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

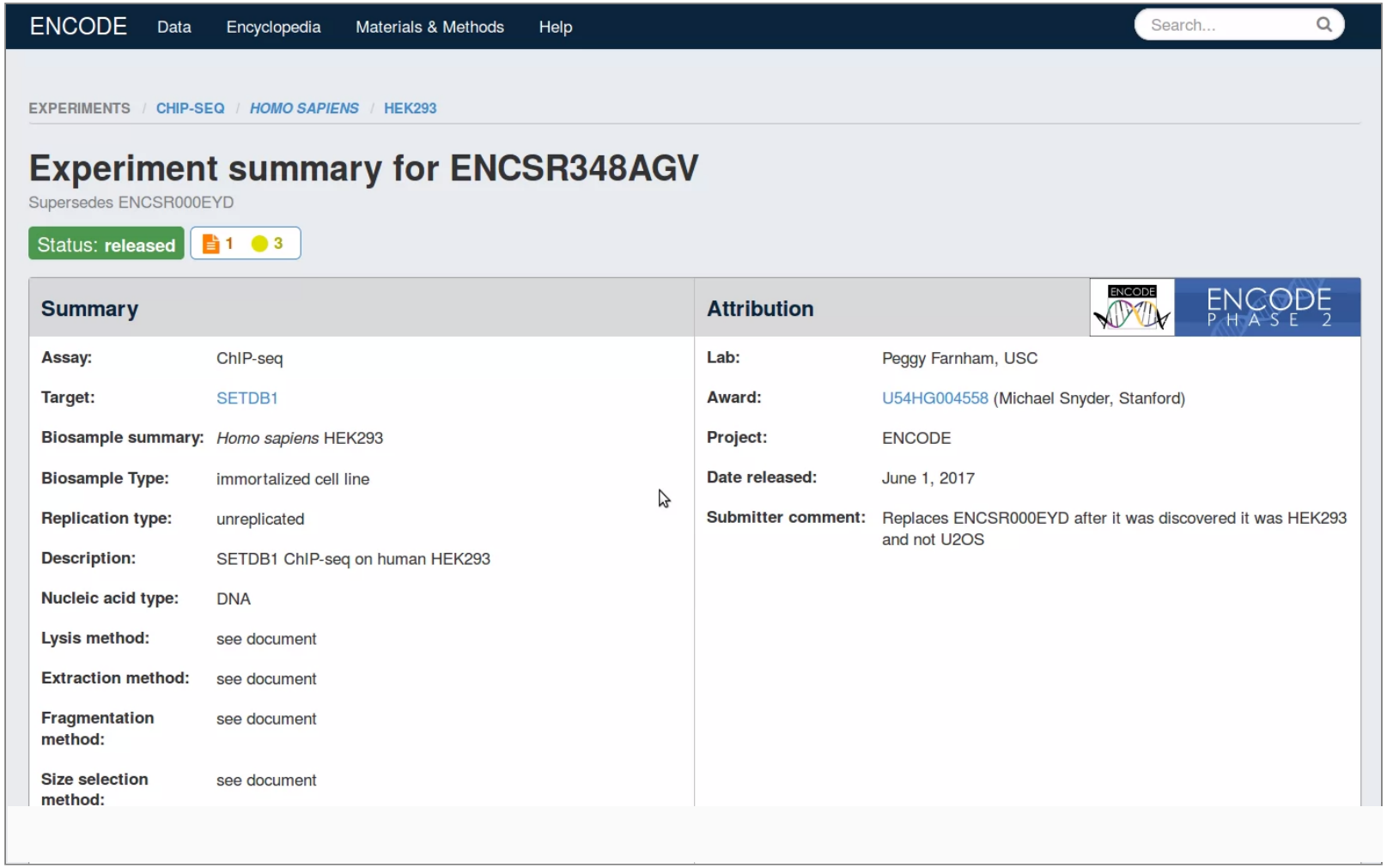
Types of web-apps for epigenomics

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

Online Genome browsers

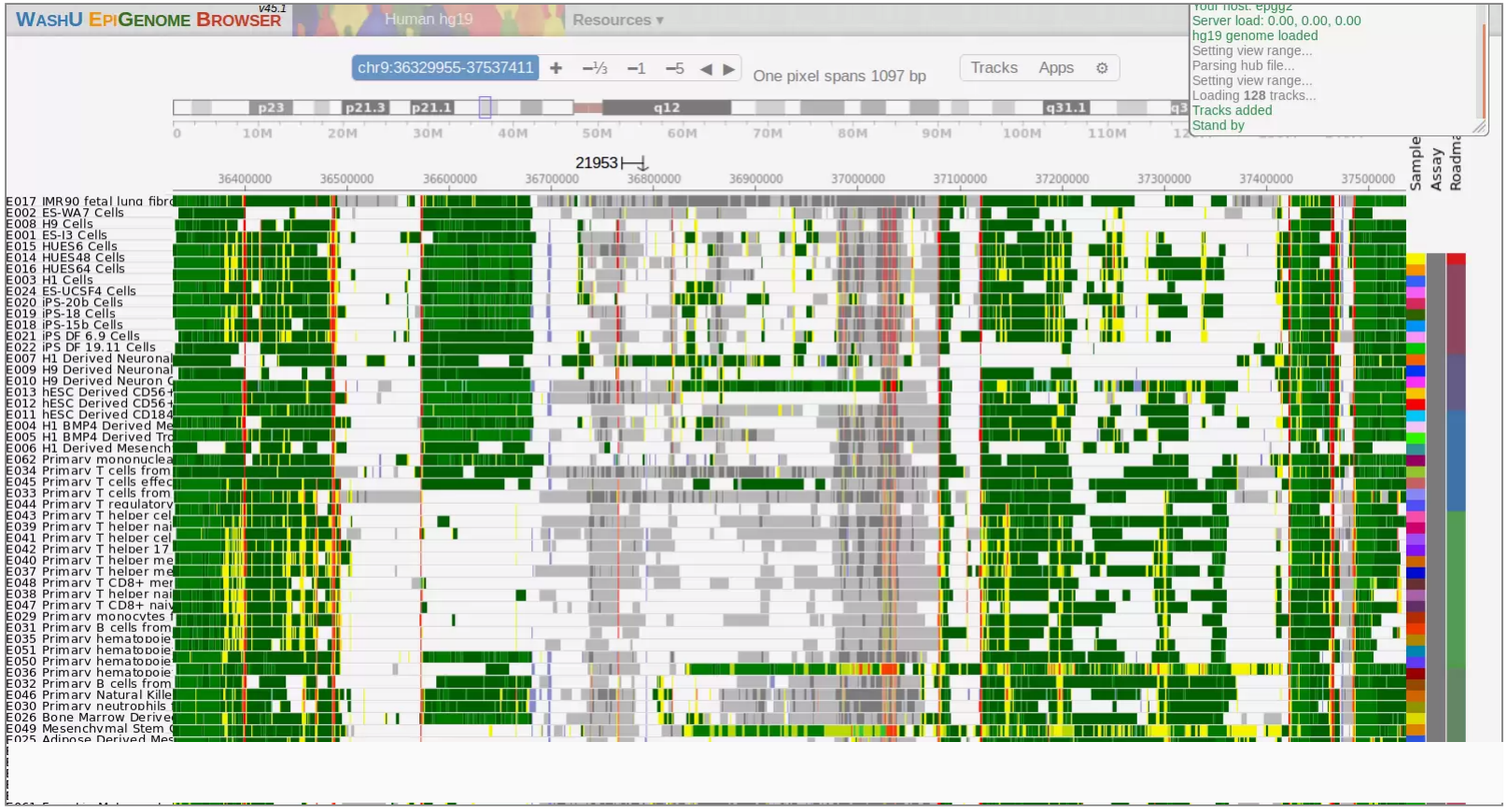
- quick look at the data
- looking at a few regions manually
- Sometimes ♥ *integrated to the Data portal* ♥

ENCODE quick view genome browser



Roadmap *Chromatin states* browser

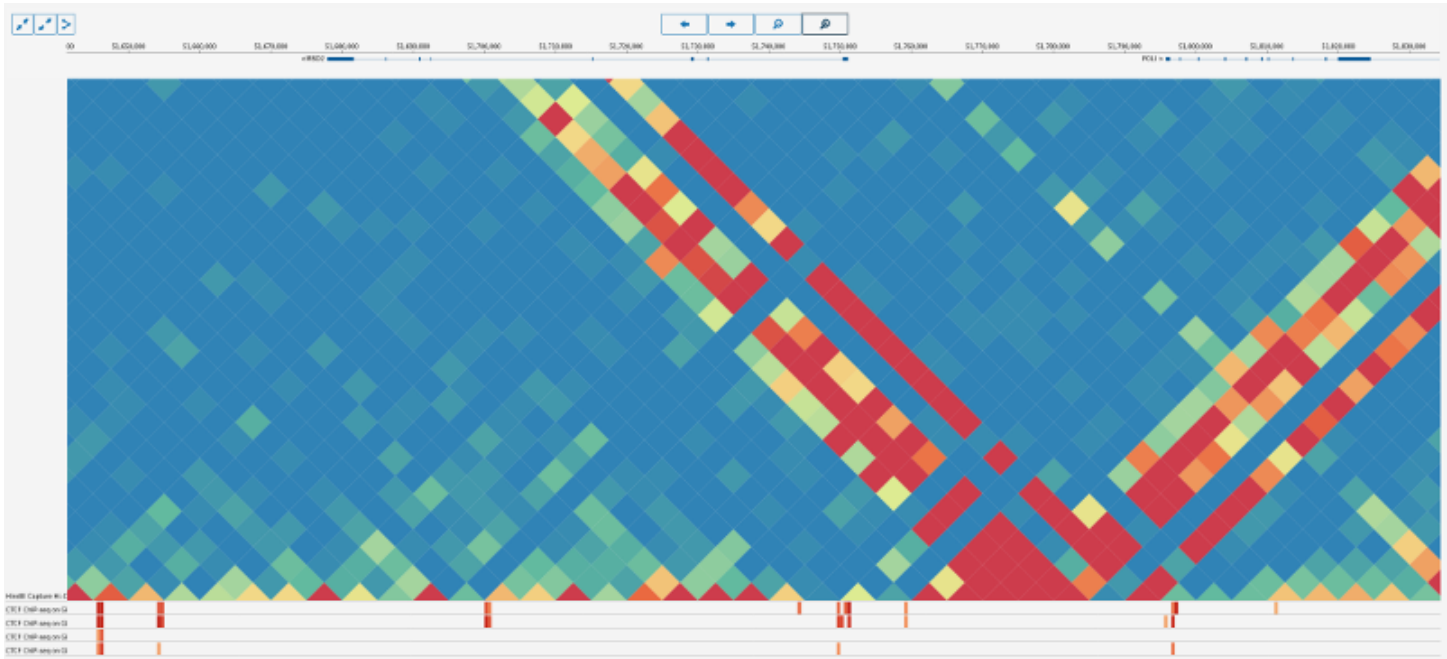
egg2.wustl.edu/roadmap/web_portal/chr_state_learning.html



Uses the great WahsU EpiGenome Browser

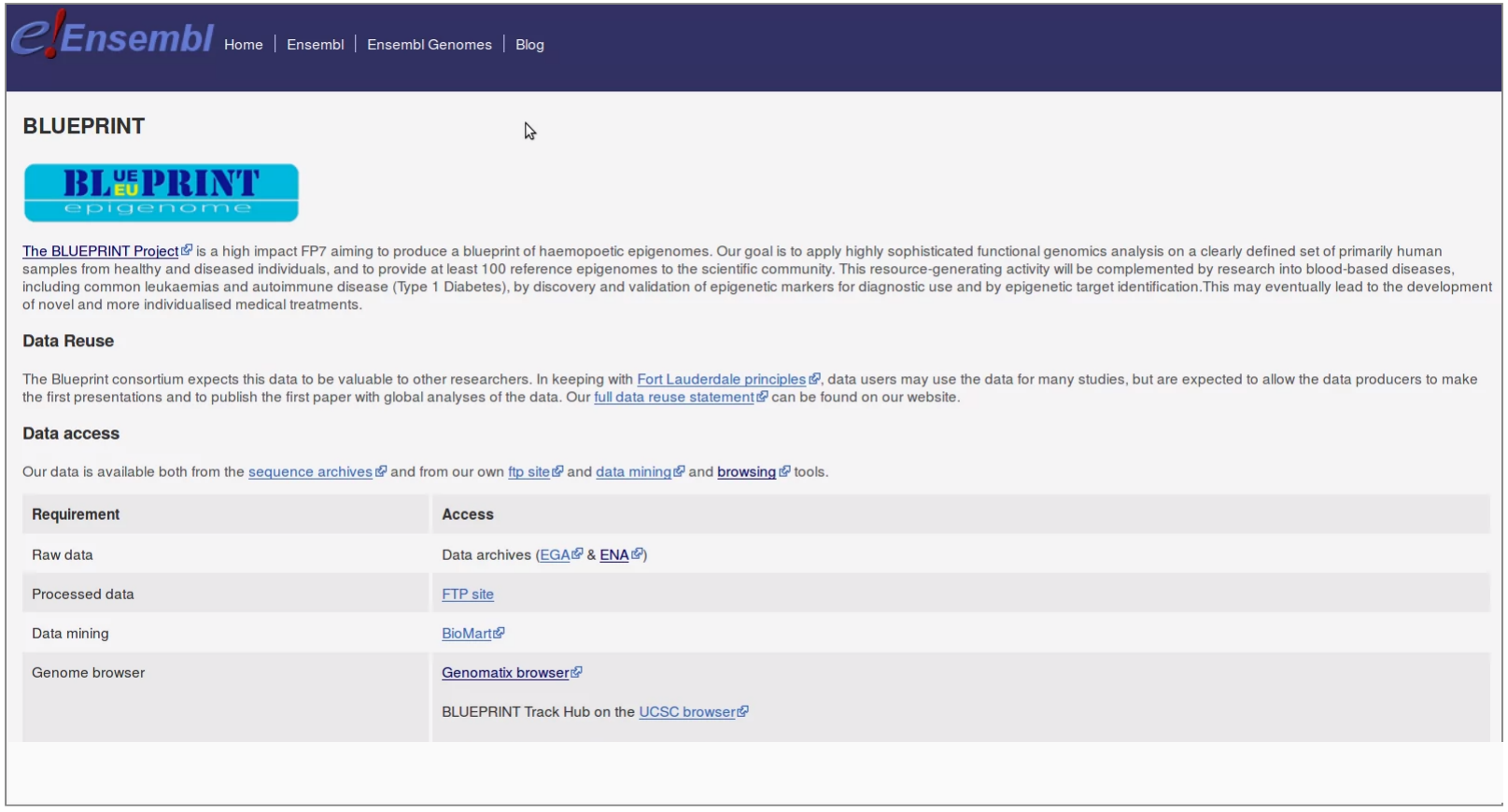
Other great options

- **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 ?



e!Ensembl Home | Ensembl | Ensembl Genomes | Blog

BLUEPRINT

BLUEPRINT epigenome

The BLUEPRINT Project is a high impact FP7 aiming to produce a blueprint of haemopoetic 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.

Data Reuse

The Blueprint consortium expects this data to be valuable to other researchers. In keeping with Fort Lauderdale principles, data users may use the data for many studies, but are expected to allow the data producers to make the first presentations and to publish the first paper with global analyses of the data. Our full data reuse statement can be found on our website.

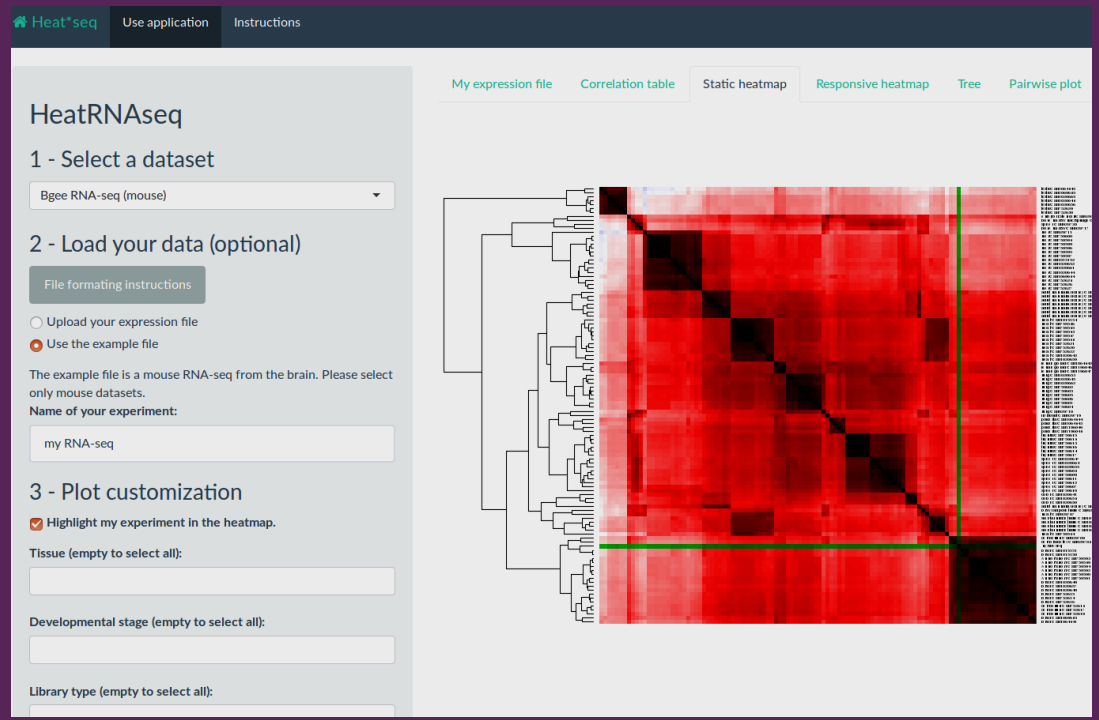
Data access

Our data is available both from the sequence archives and from our own ftp site and data mining and browsing tools.

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

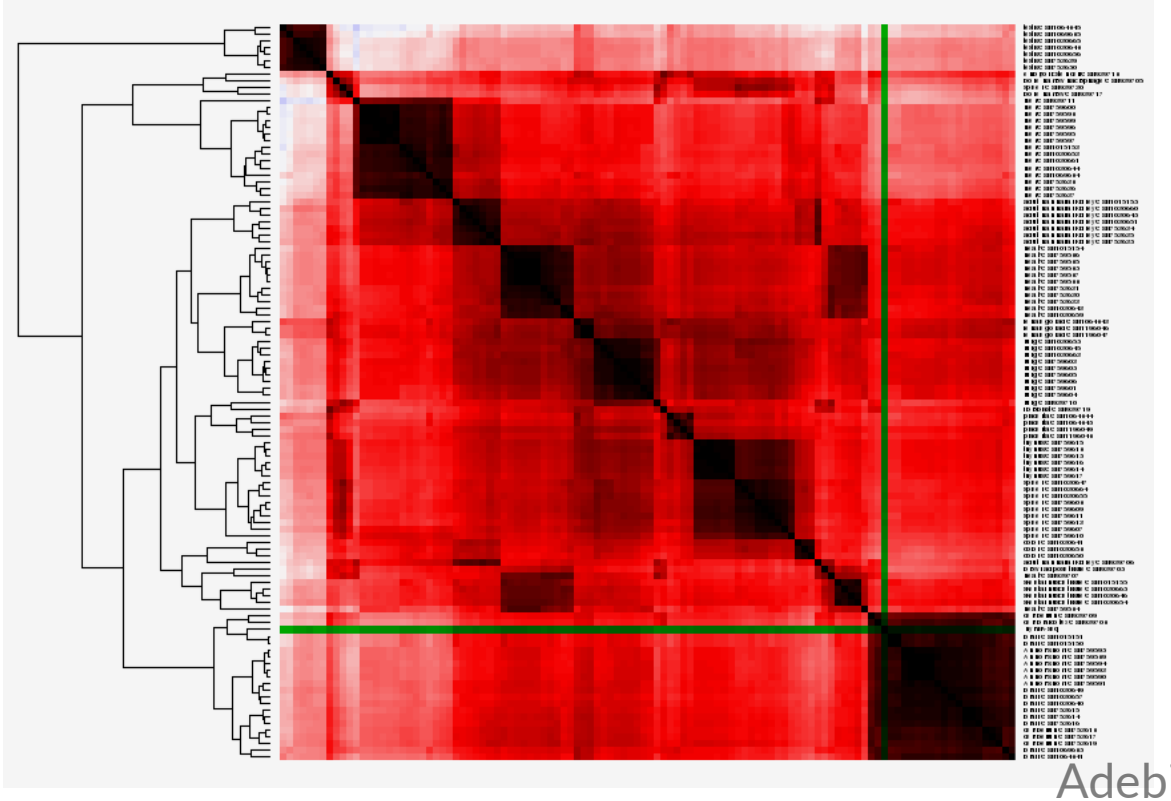
Types of web-apps for epigenomics

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



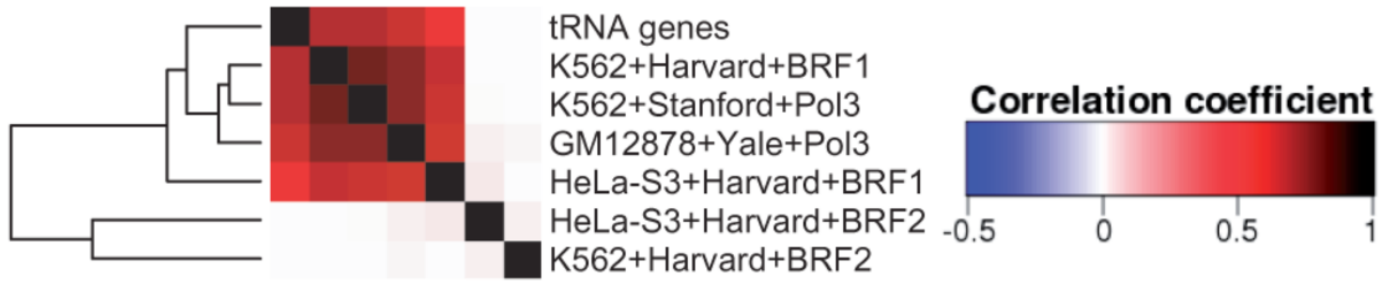
Heat*seq: genome-wide comparison of NGS experiments

- 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



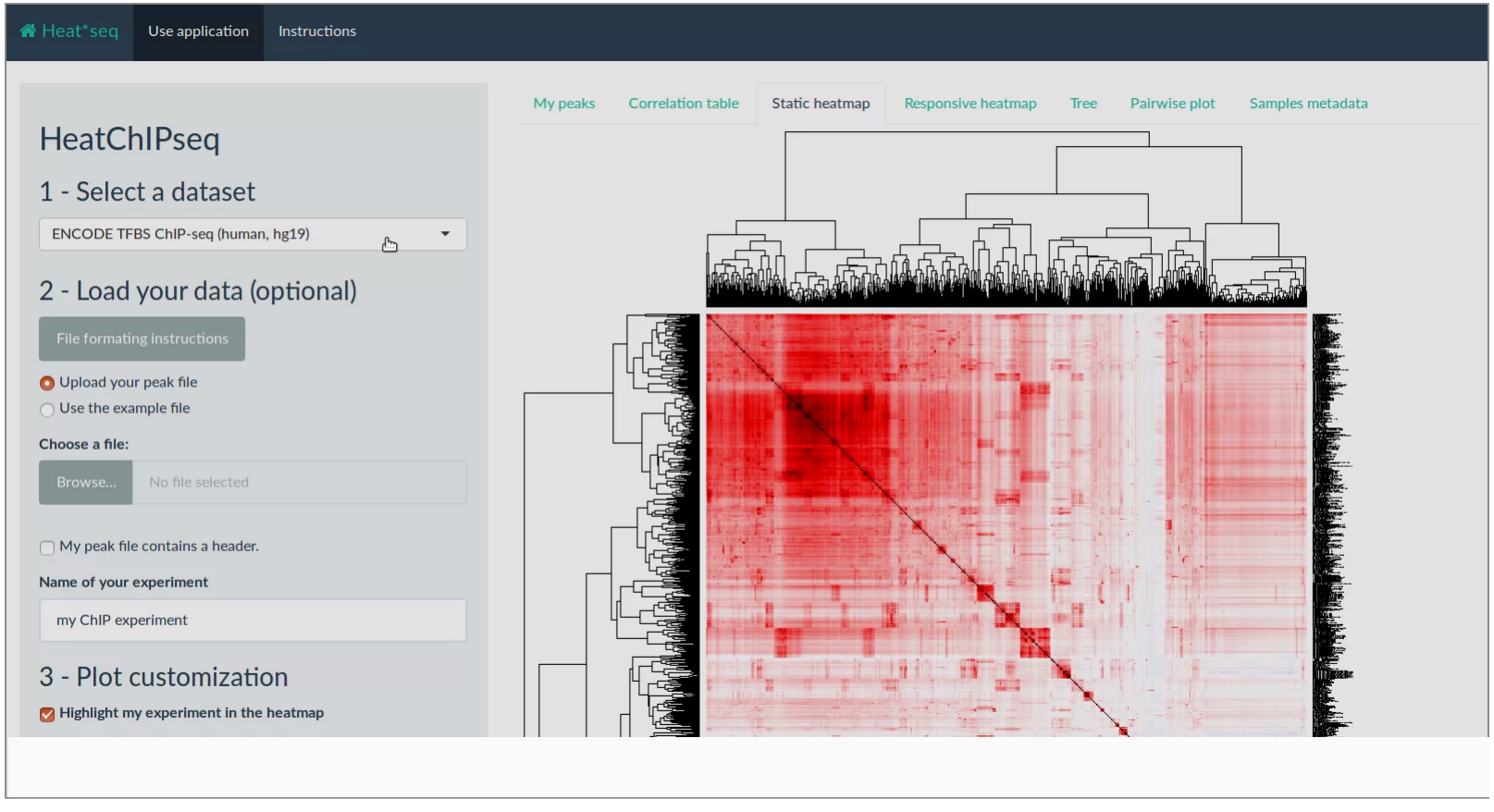
Clustered correlation heatmaps:

- 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

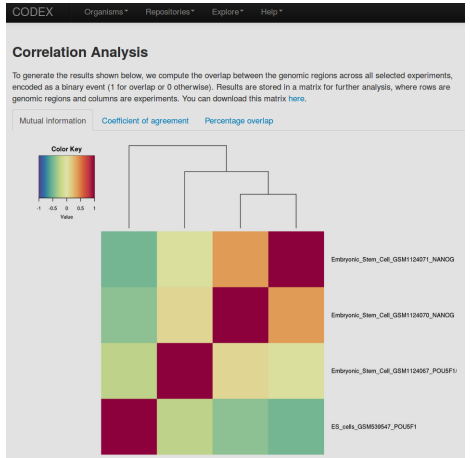
Example: ESR1 TF ChIP-seq from GEO, compared against all TF ChIP-seq from ENCODE



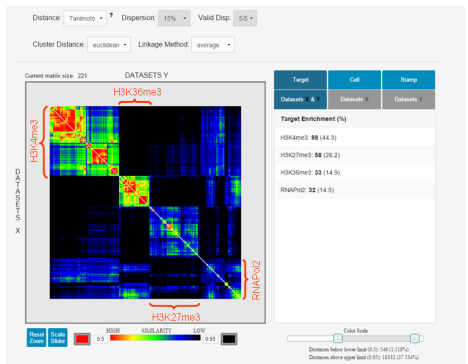
Input file: peak list, *.bed* format.

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

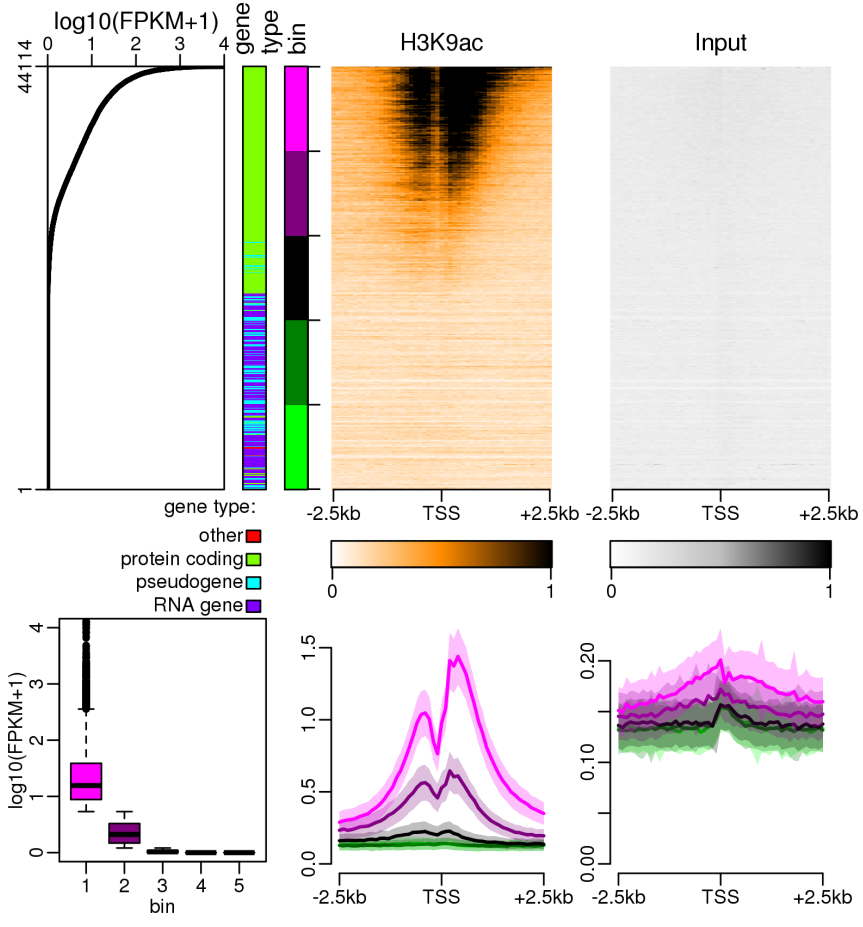
Profile Explorer for Roadmap Epigenomics

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

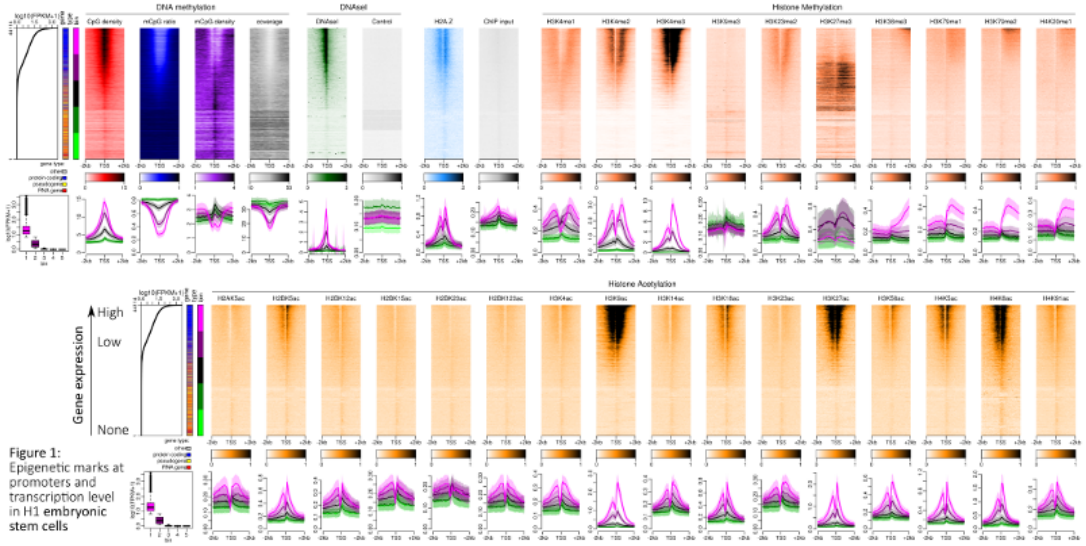


Figure 1: Epigenetic marks at promoters and transcription level in H1 embryonic stem cells

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

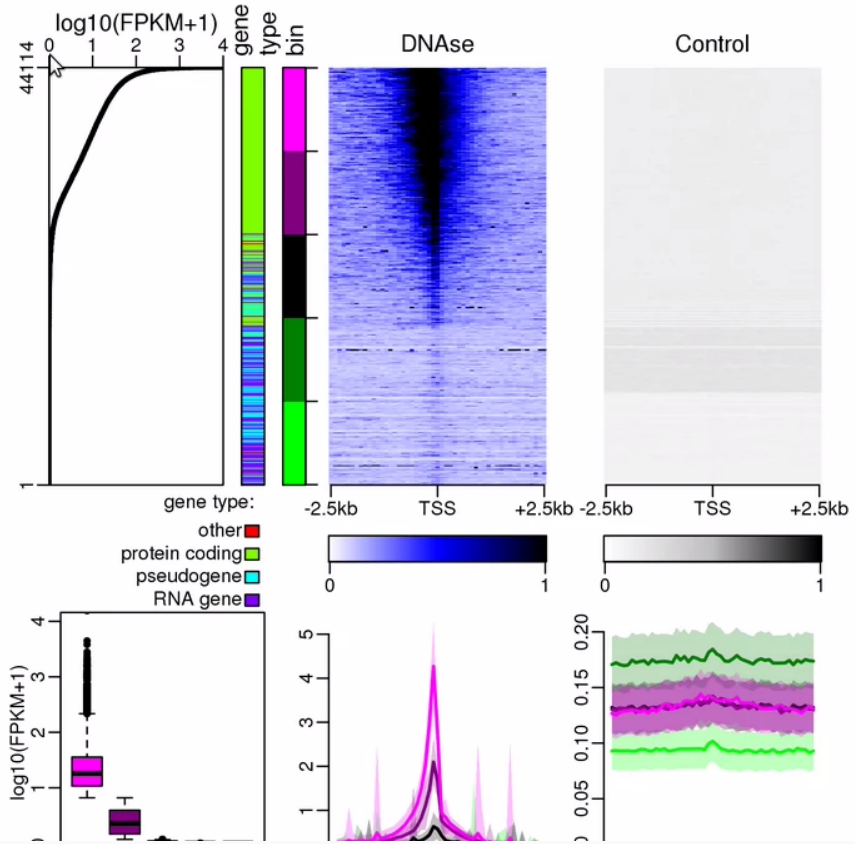
3- Choose an assay: DNAse 1

4- Choose a cell type: H1_Cell_Line

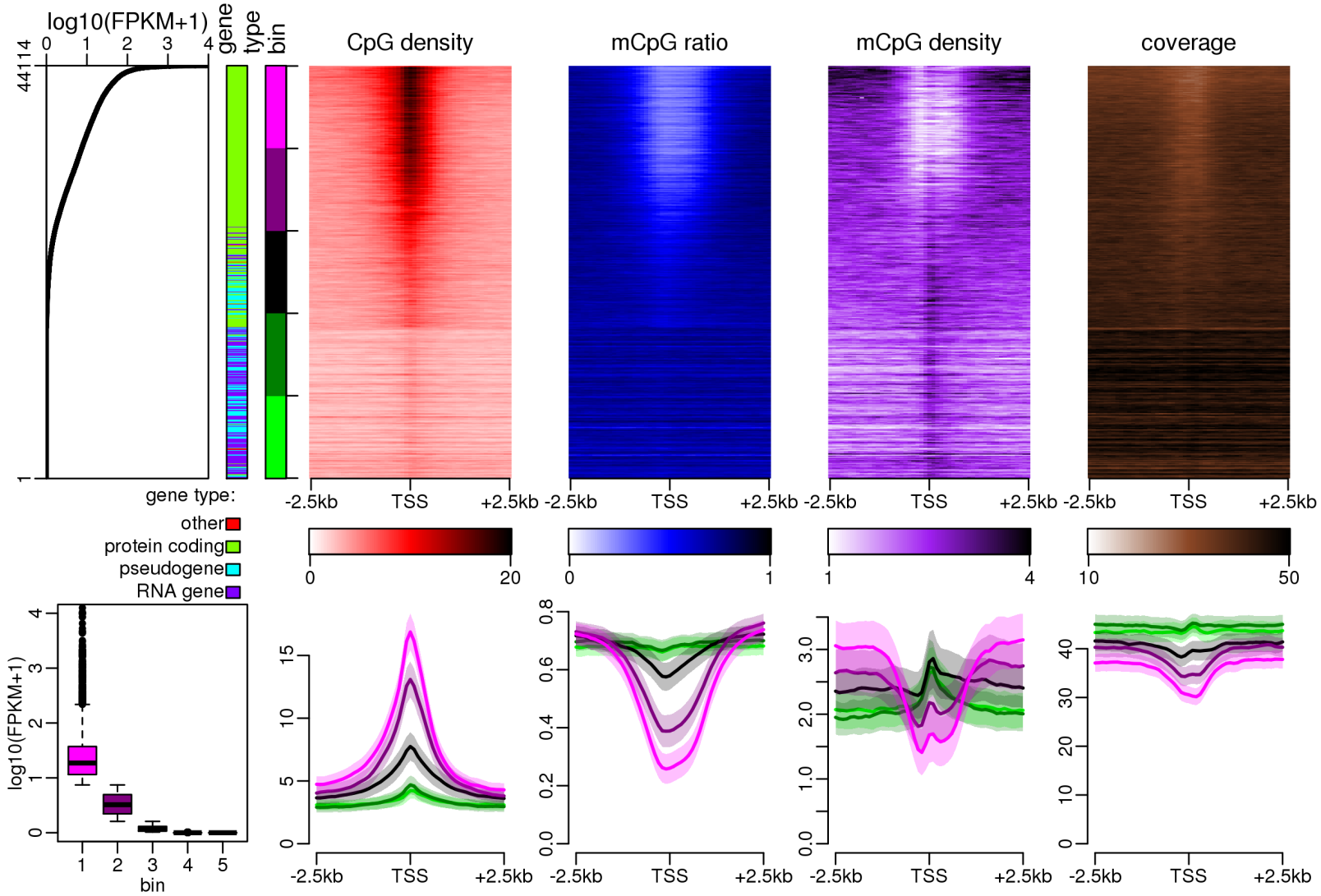
5- Choose a gene category: all

Download image

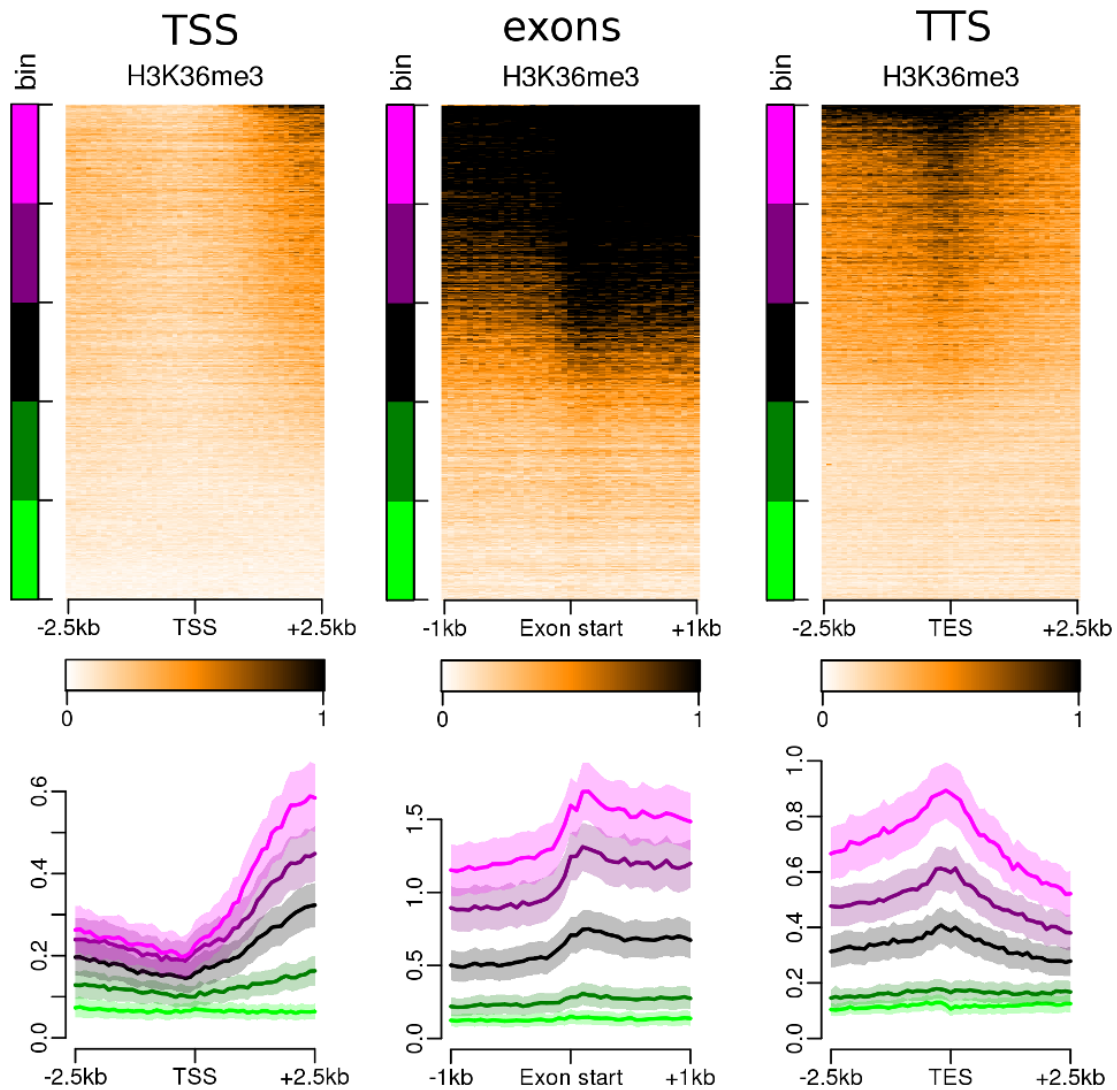
DNAse 1 at TSS in H1_Cell_Line for all genes



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

ROSLIN

THE UNIVERSITY of EDINBURGH

BBSRC

CASCADE-FELLOWS

GenPhySE

Génétique, Physiologie et Systèmes d'Elevage

INRA

SCIENCE & IMPACT