

Epigenetic marks and the human transcript diversity



GenEpi@Genphyse

Guillaume Devailly - 2018/05/14

Introduction

- Abundant epigenomic data for model organisms.
- Under-utilised?
- Epigenomic data will soon be abundant for agricultural species too.

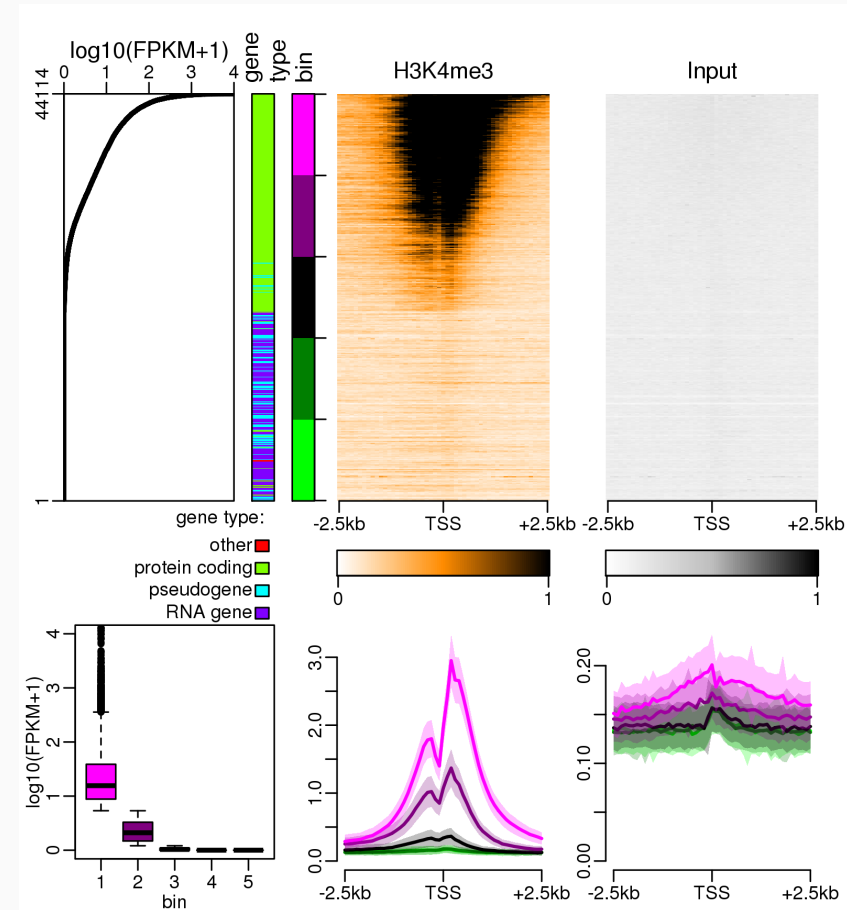
Dataset:



- RNA-seq
- DNase1
- WGBS, 10 different histone methylations, 17 different histone acetylations
- 33 cell lines & tissues

Objectives

- Systemic correlation of marks and **gene** transcription level:
 - at TSS
 - at TTS
- Systemic correlation of marks and **exon**:
 - transcription level
 - inclusion ratio
- For **all** genes/exons in **each** cell type.
- For **each** gene/exon in **all** cell types.



Methodology

Processed data taken from RoadmapEpigenomics:


- gene and exon FPKM
- CpG methylation ratios
- DNase-seq and ChIP-seq coverage

Reprocessing of fastq for exon inclusion ratio (aka %spliced, or PSI):

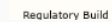
- pseudo-mapping and quantification with **Salmon**
- exon Psi for exon **e** of gene **g** using *tpm* quantification of transcripts **t**:

$$PSI_{ij} = \frac{\sum_{t \in e} tpm_t}{\sum_{t \in g} tpm_t}$$

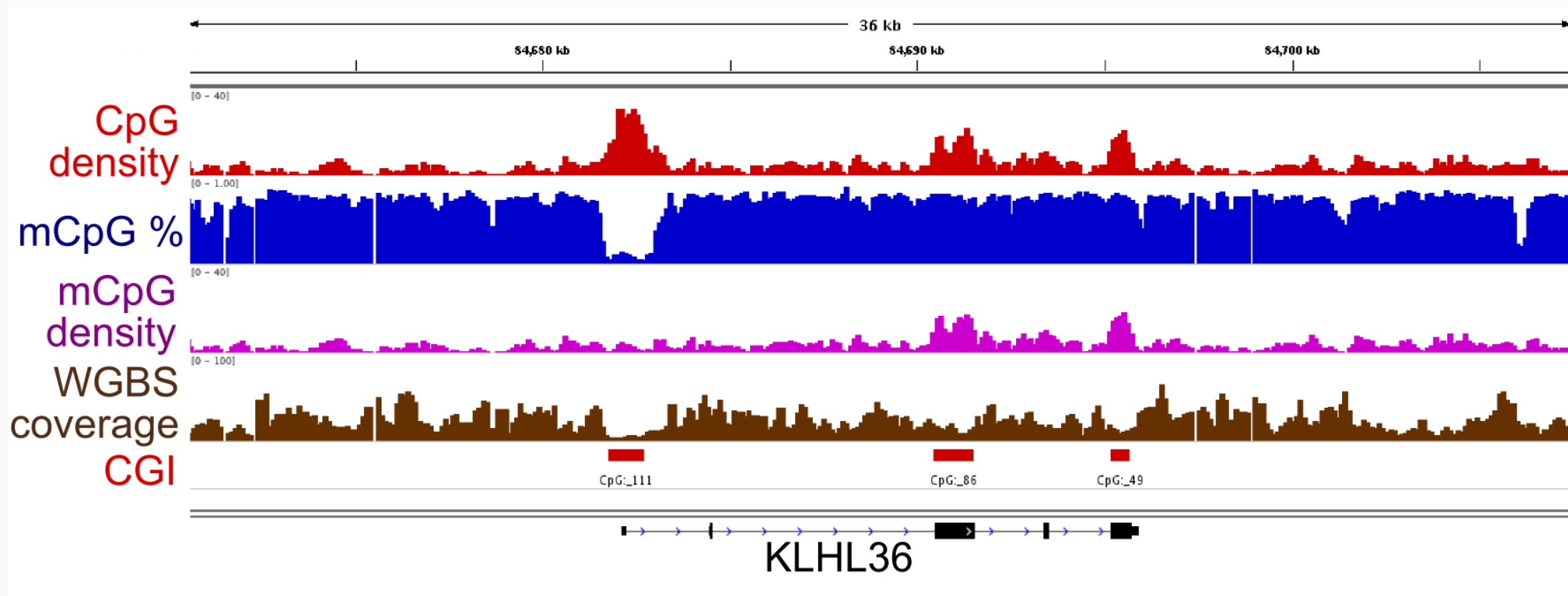
What is a gene?

	Version 22 (October 2014 freeze, GRCh38) - Ensembl 79, 80	
	General stats	
	Total No of Genes	60483

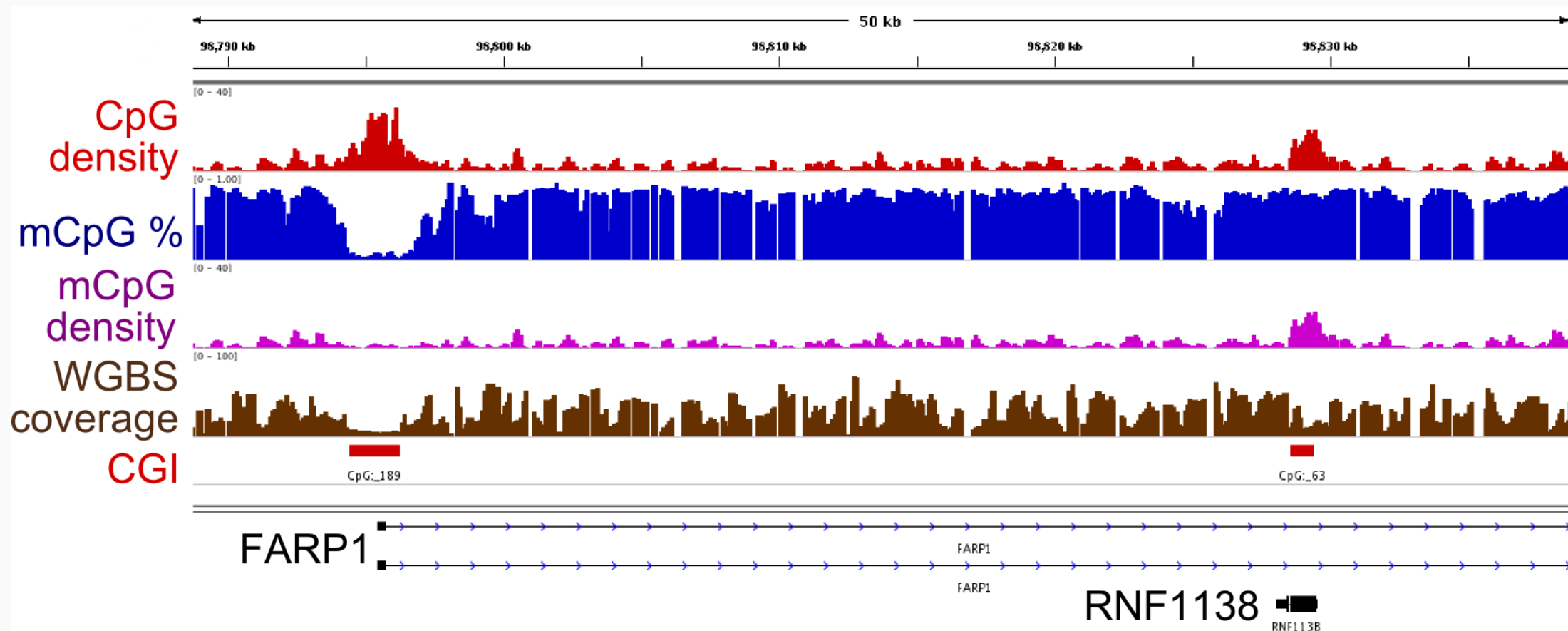
Genes
(Comprehensive set...)



DNA methylation: ratio *and* density



DNA methylation: ratio *and* density



Promoter DNA methylation

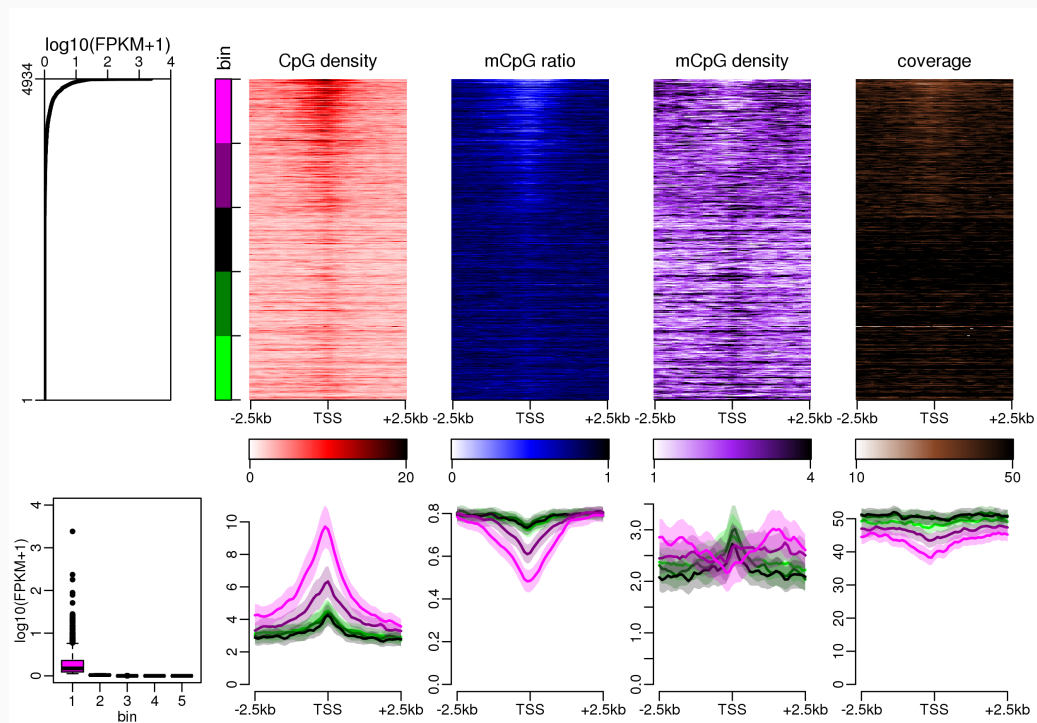
Lung, all genes

Promoter DNA methylation

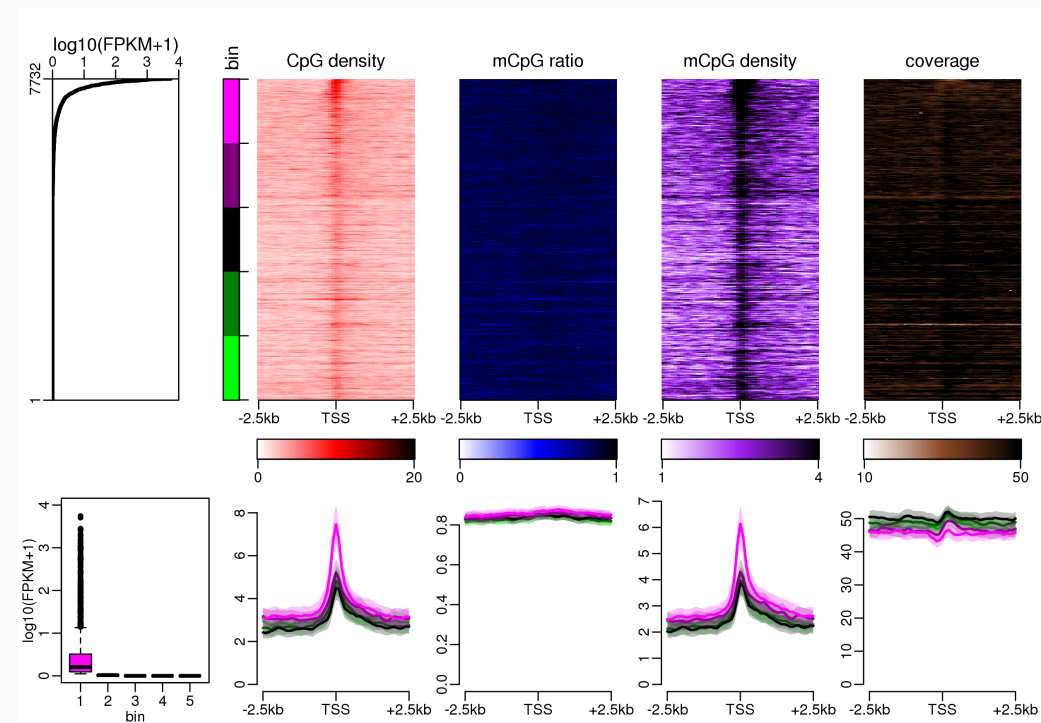
Lung, protein coding genes

Promoter DNA methylation

Lung, lincRNA genes

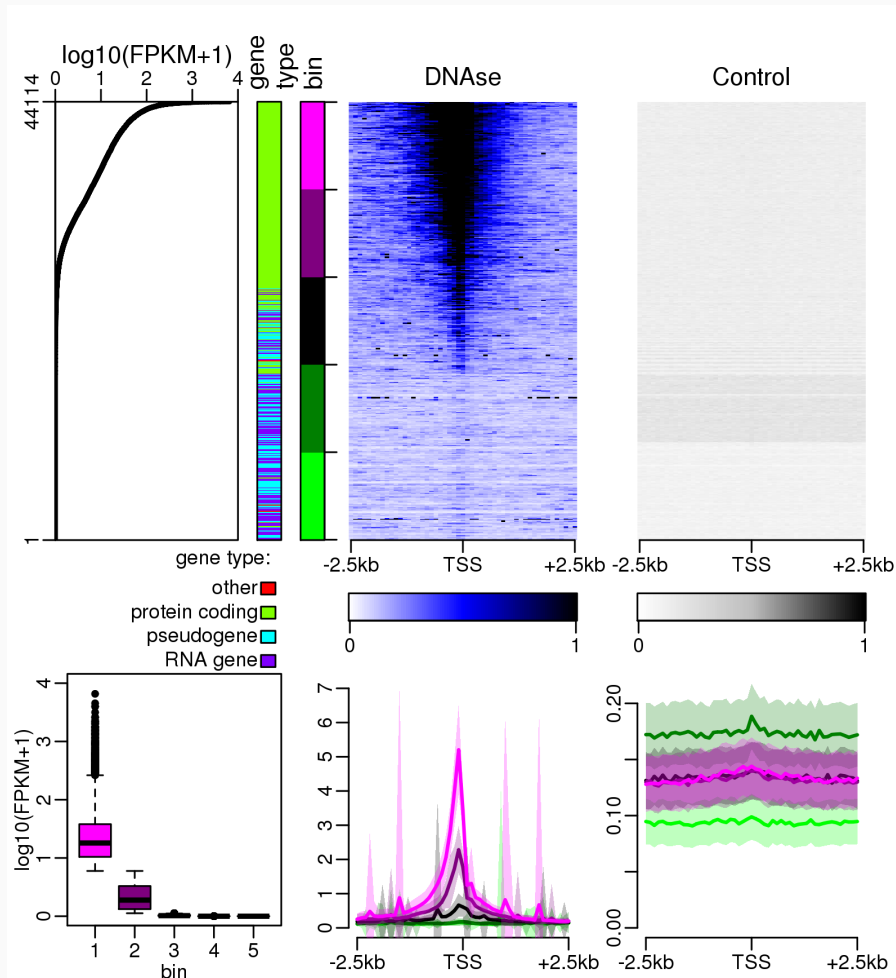


Lung, processed pseudogenes

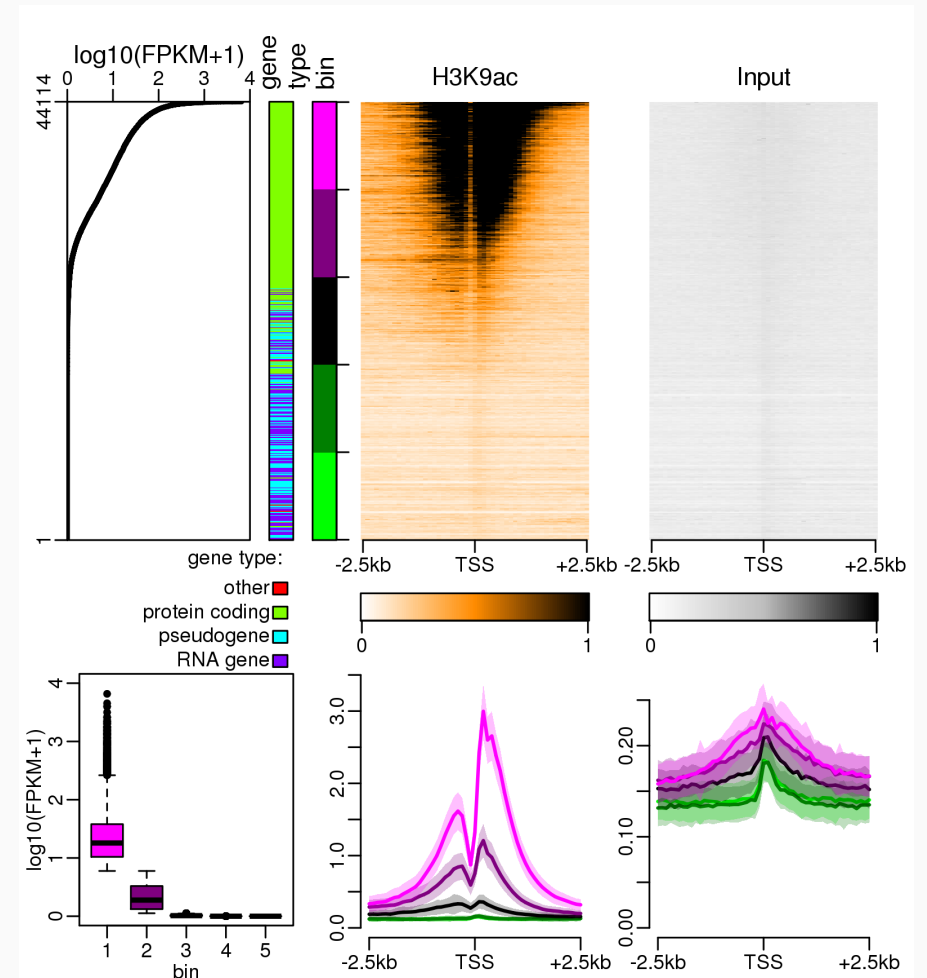


Activating marks

DNase-seq, all genes, H1-derived mesoderm

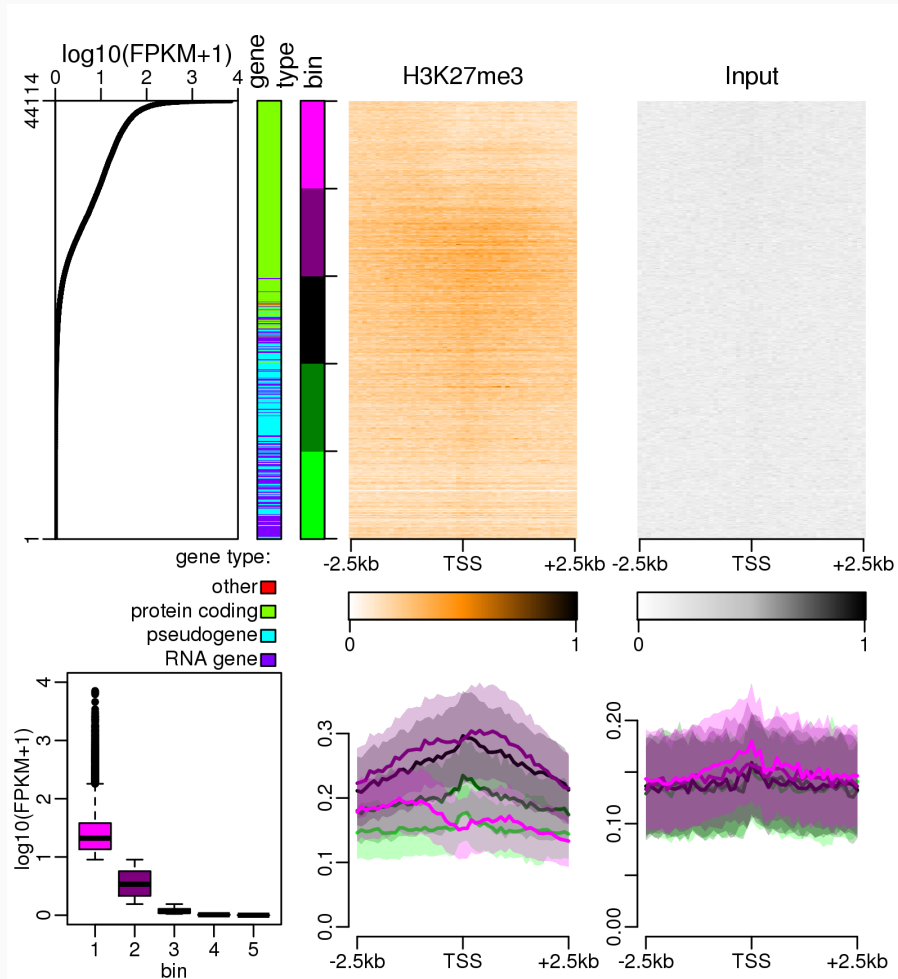


H3K9ac ChIP-seq, all genes, H1-derived mesoderm



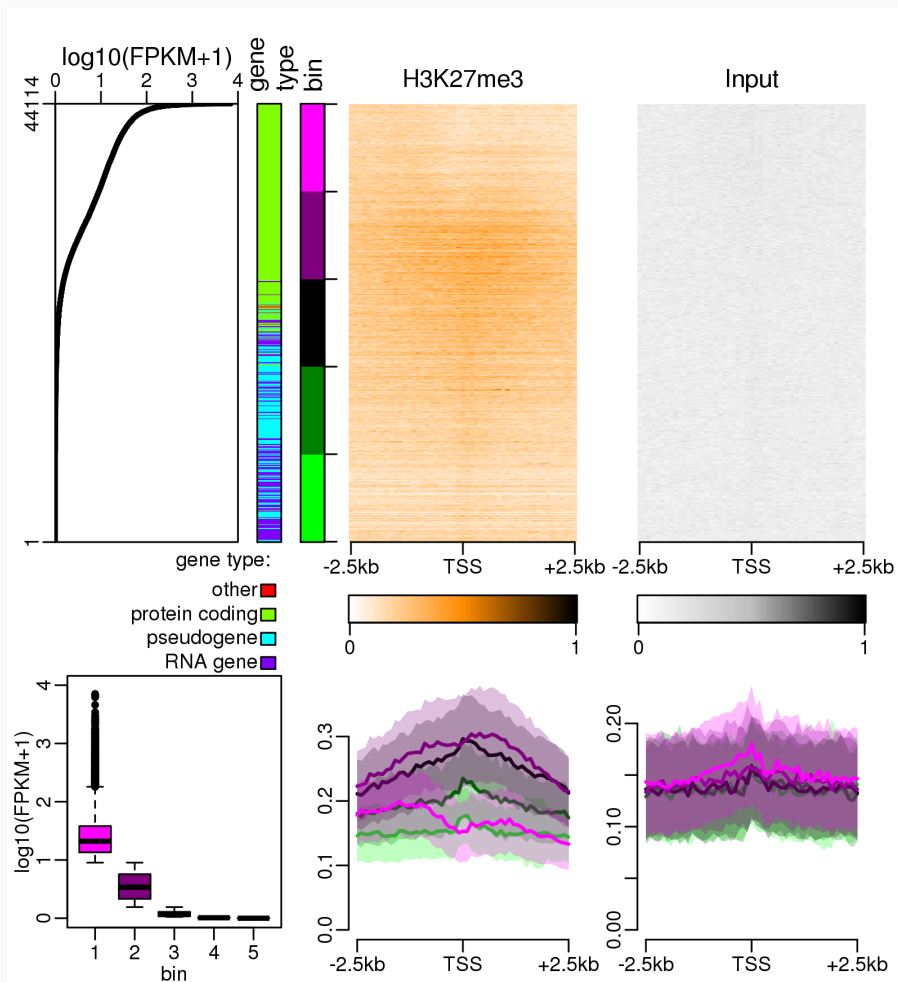
Repressing marks

H3K27me3, all genes, Thymus

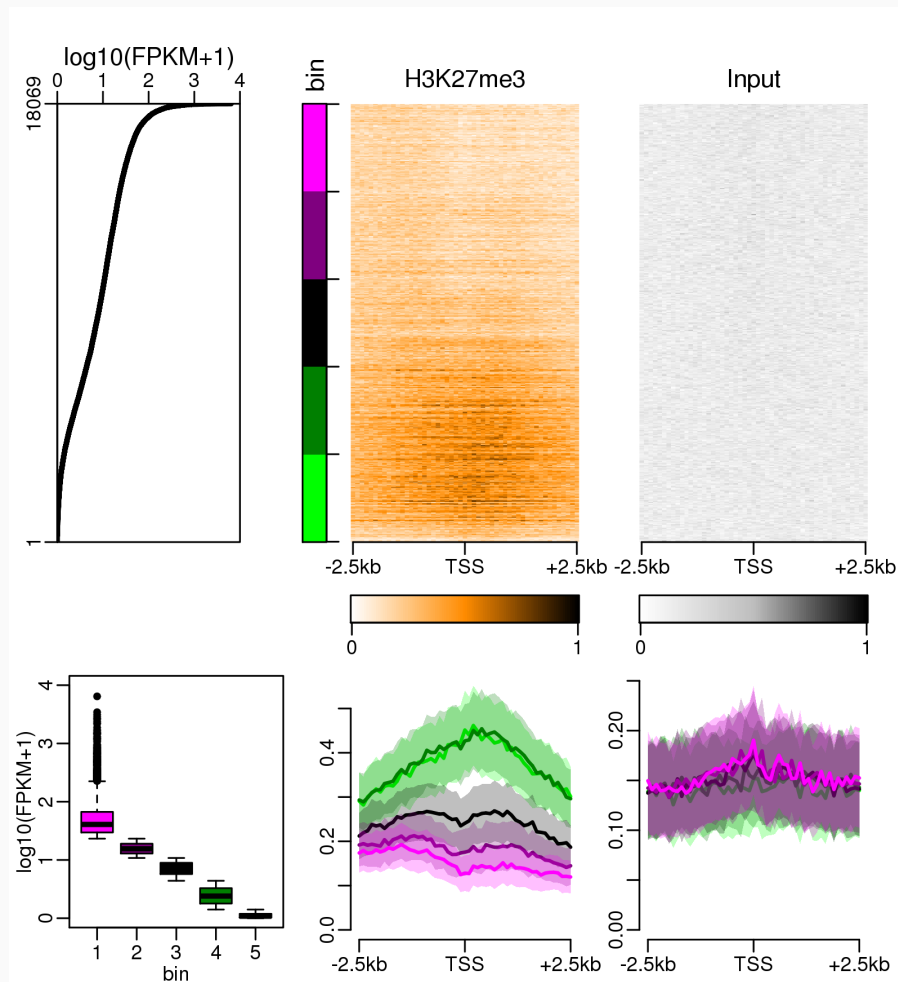


Repressing marks

H3K27me3, all genes, Thymus



H3K27me3, protein coding genes, Thymus



Assymetric marks

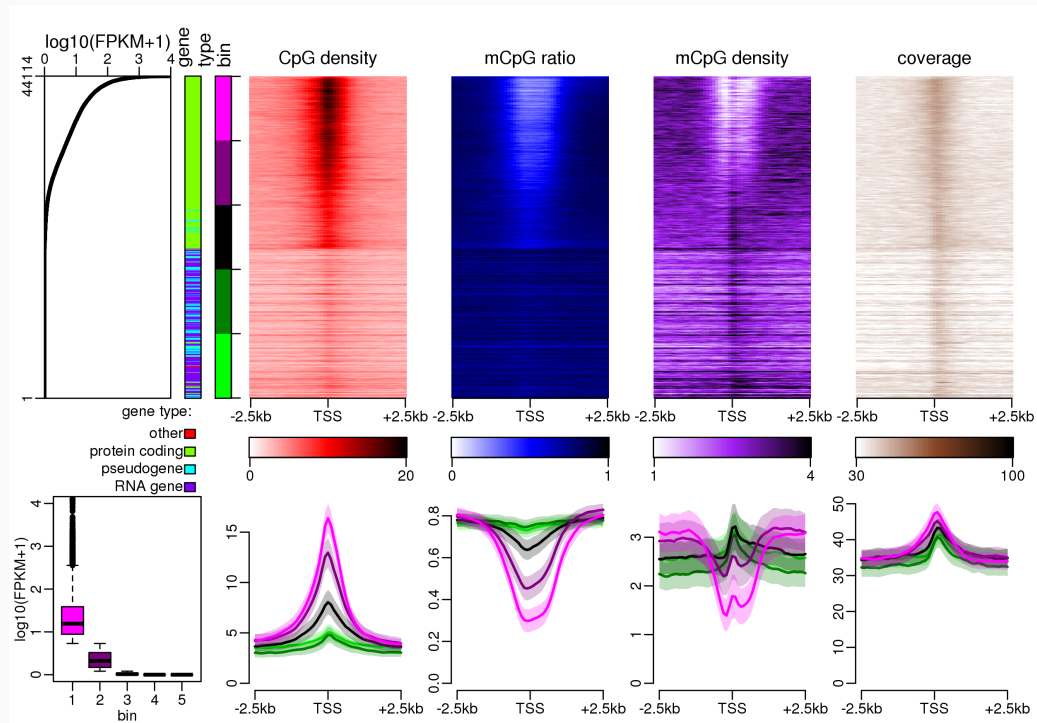
H3K79me2, protein coding genes, H1-derived mesoderm

Exonic marks

H3K36me3, fetal large intestine

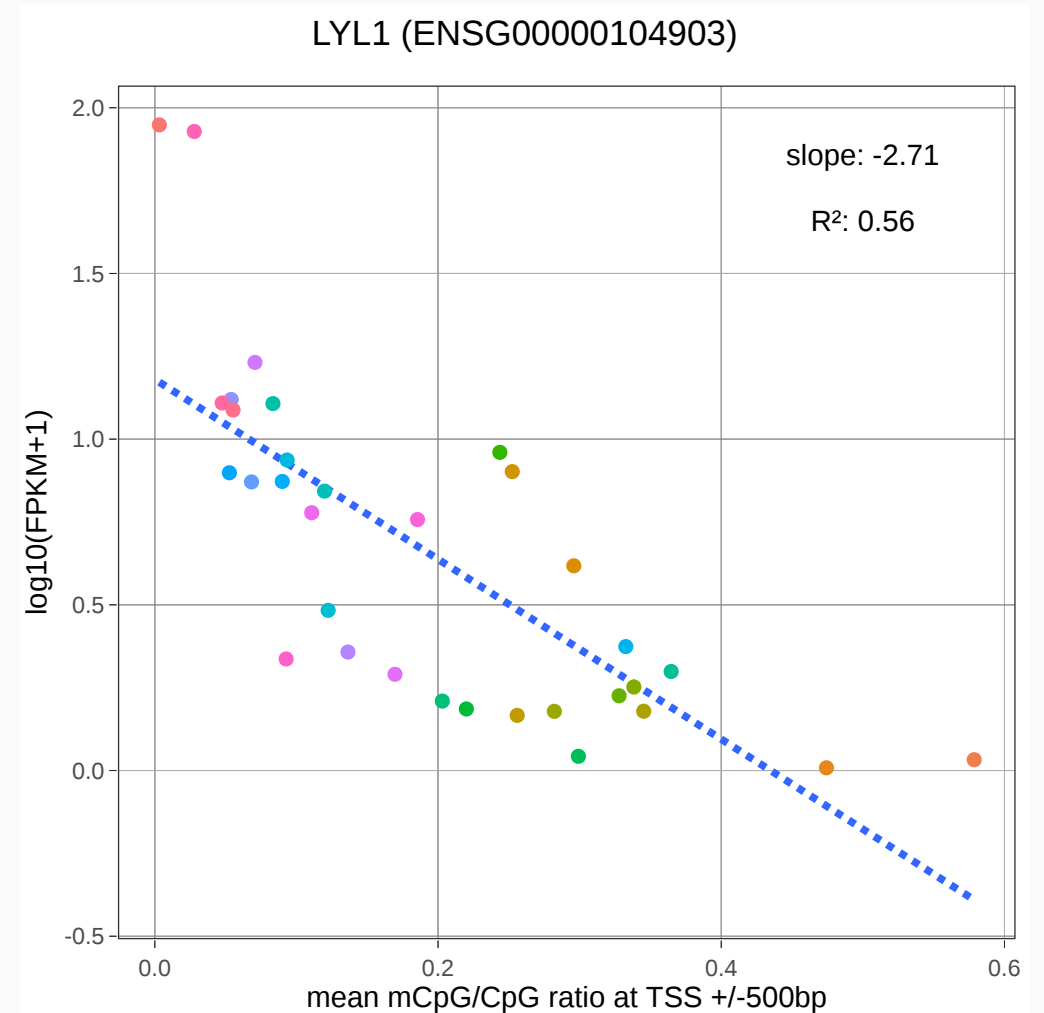
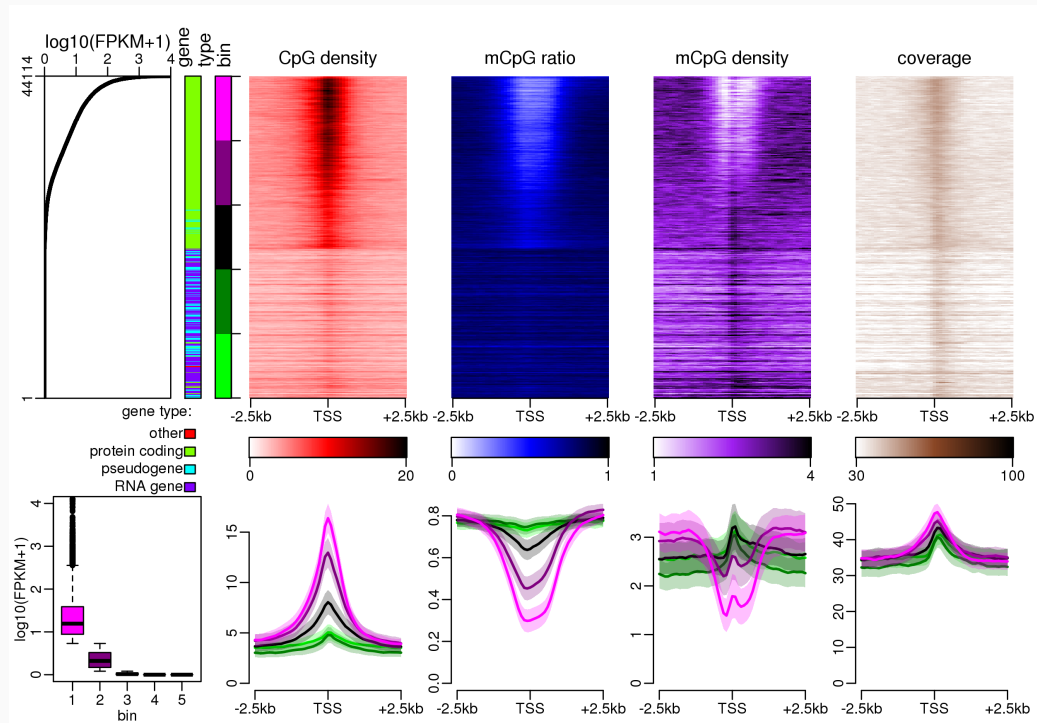
Cell by cell vs gene by gene

WGBS, adult liver



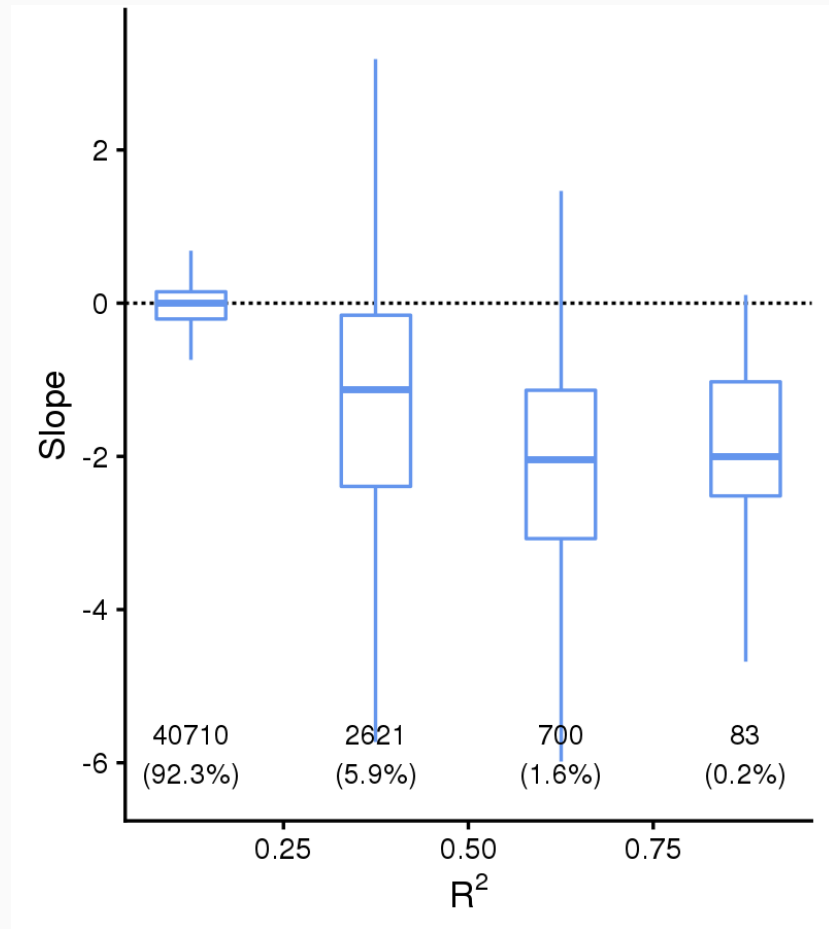
Cell by cell vs gene by gene

WGBS, adult liver



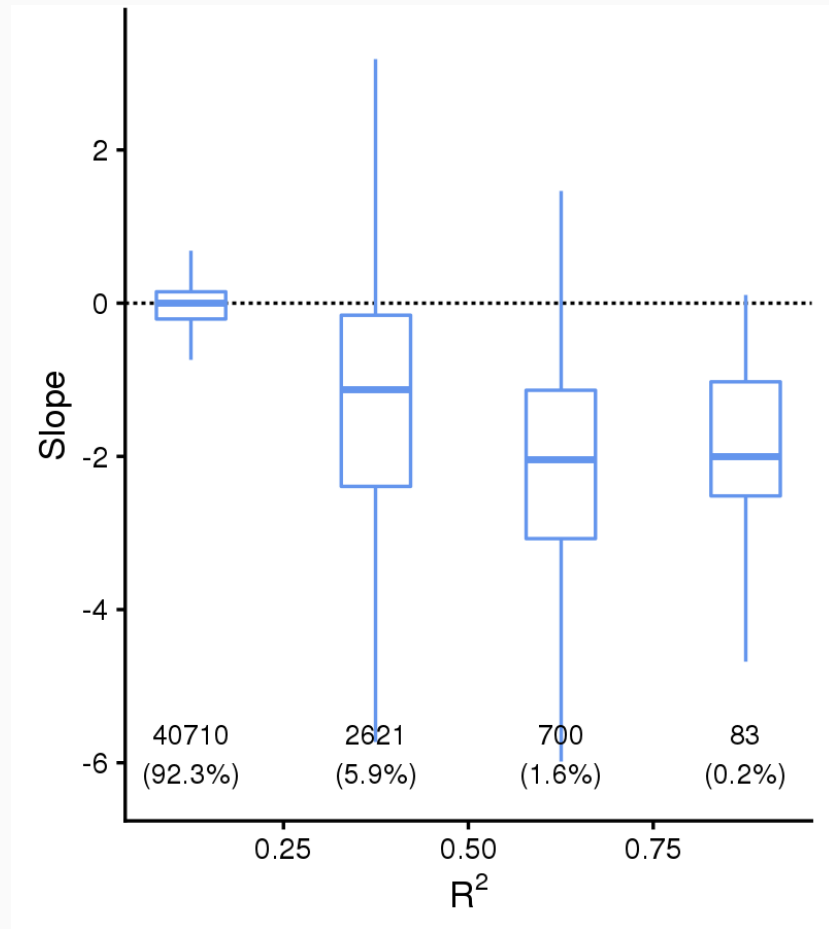
Repressing and activating marks

mCpG ratio

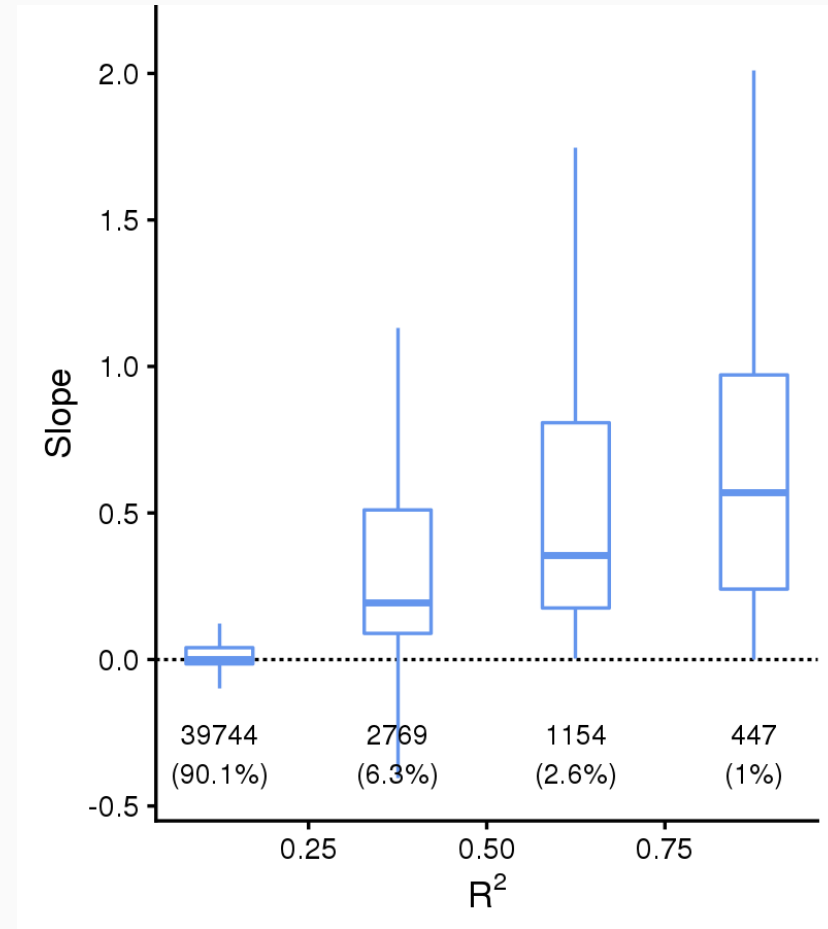


Repressing and activating marks

mCpG ratio



H3K4me3



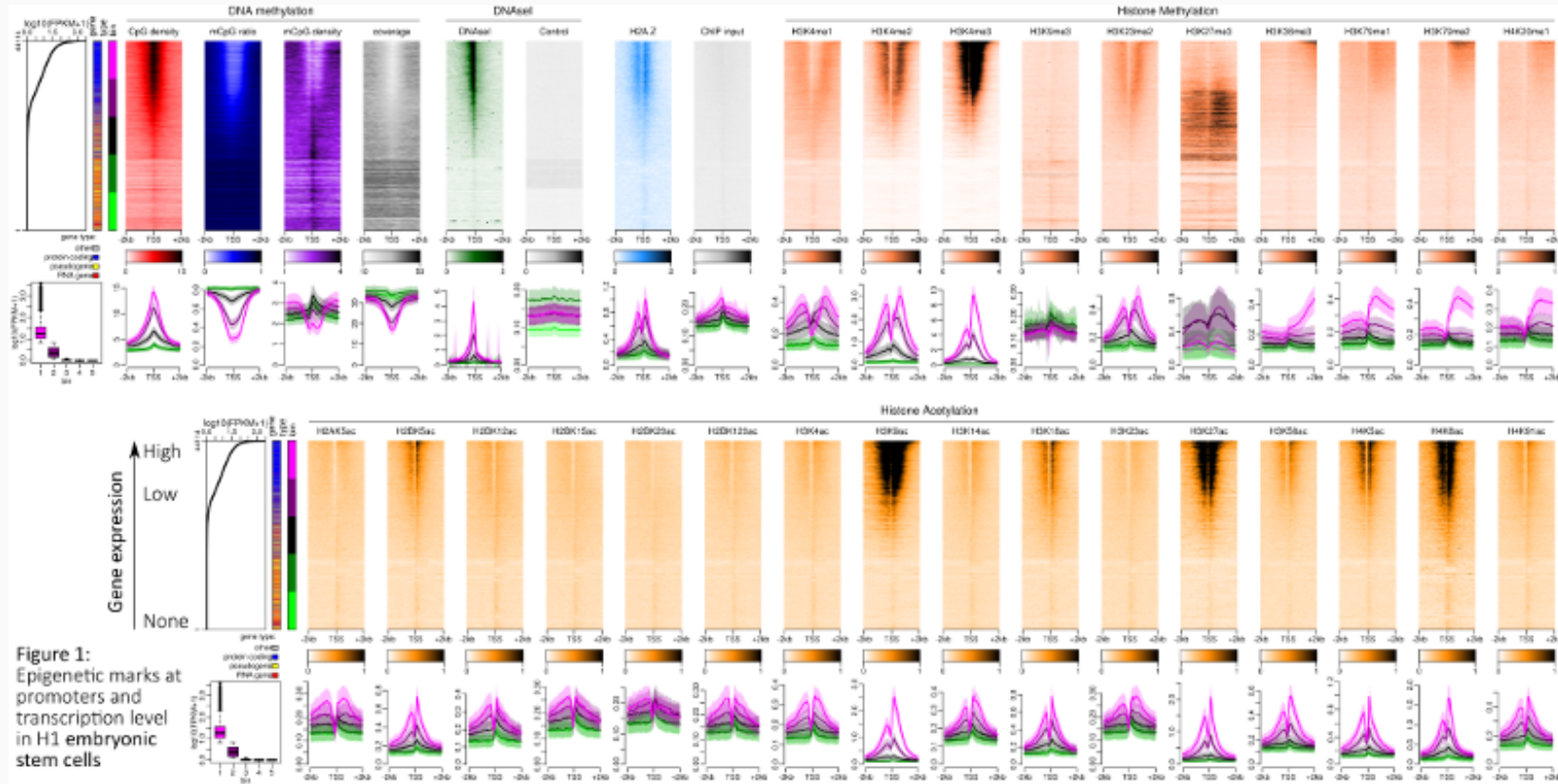
Summary of results

mark	cell type by cell type	gene by gene	center on TSS
WGBS	negative	negative	
DNAse	positive	positive	
H2A.Z	positive	balanced	
H3K4me1	positive	positive	
H3K4me2	positive	positive	
H3K4me3	positive	positive	
H3K9me3	unclear	unclear	no
H3K23me2	positive	NA	
H3K27me3	negative – variable	negative	
H3K36me3	positive	positive	no
H3K79me1	positive	positive	no
H3K79me2	positive	balanced	no

mark	cell type by cell type	gene by gene	center on TSS
H2AK5ac	positive	positive	
H2BK120ac	positive	positive	
H2BK12ac	positive	positive	
H2BK15ac	positive	balanced	
H2BK20ac	neutral	NA	
H2BK5ac	positive	positive	
H3K4ac	positive	positive	
H3K9ac	positive	positive	
H3K14ac	positive	positive	
H3K18ac	positive	positive	
H3K23ac	positive	positive	
H3K27ac	positive	positive	
H3K56ac	positive	NA	
H4K8ac	positive	positive	
H4K12ac	positive	NA	
H4K91ac	positive	positive	

Lots of results

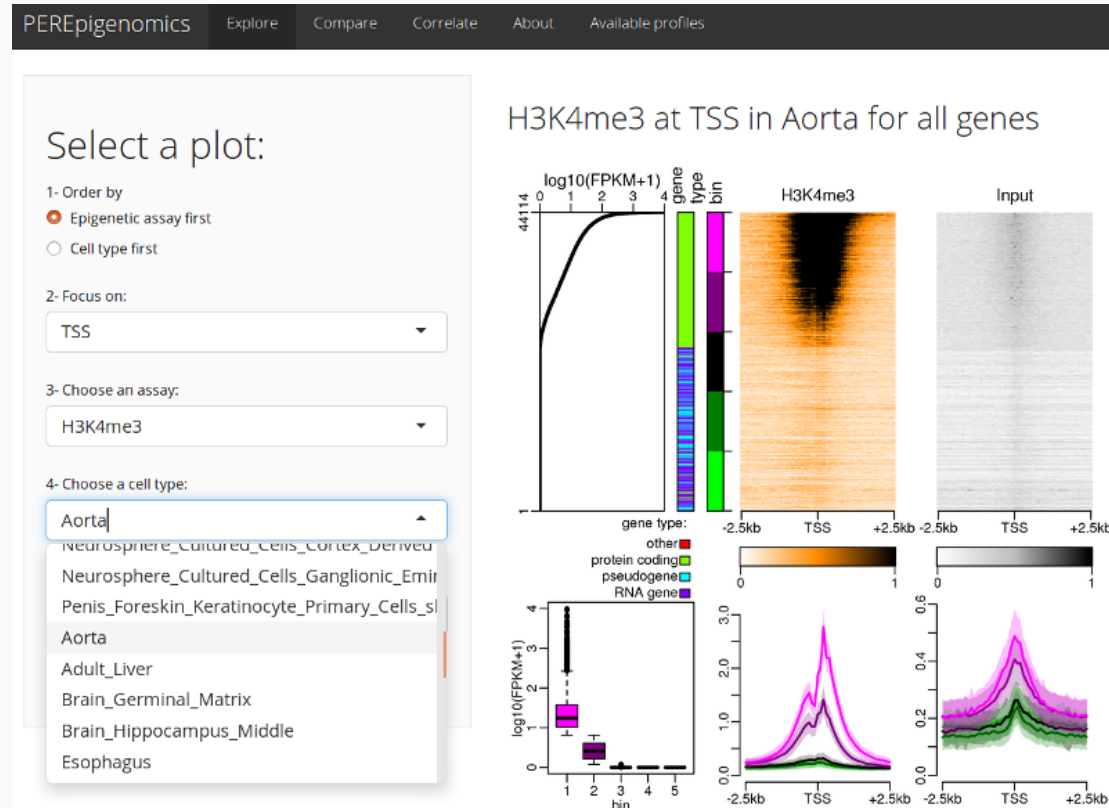
Combinatorial explosion: > 9000 plots!



PEREpigenomics

Profile Explorer of Roadmap *Epigenomic* data

www.perepigenomics.roslin.ed.ac.uk



Perspective

- Finish PSI analysis.
- Refine features (short vs long genes, mono-, bi-, tri- exonic genes, gene families, etc.).
- Other marks.
- Causality analysis using perturbation experiments.
- FAANG data.

Thanks

Anagha Joshi

Anna Mantsoki

Deepti Vipin



THE UNIVERSITY
of EDINBURGH

