



Valorisations of epigenomic and transcriptomic data

Guillaume Devailly

► To cite this version:

Guillaume Devailly. Valorisations of epigenomic and transcriptomic data. Animation scientifique d'Unité, Dec 2017, Toulouse, France. hal-01800387

HAL Id: hal-01800387

<https://hal.science/hal-01800387>

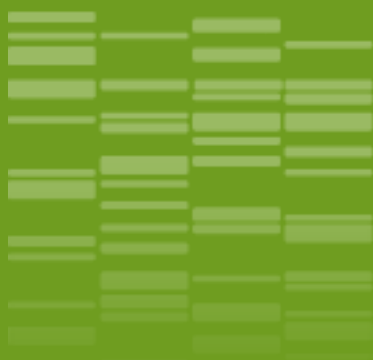
Submitted on 4 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



GenEpi@GenPhySE



Valorisations of epigenomic and transcriptomic data

Guillaume Devailly



@G_Devailly



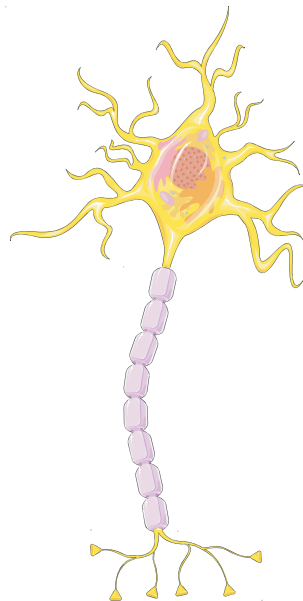
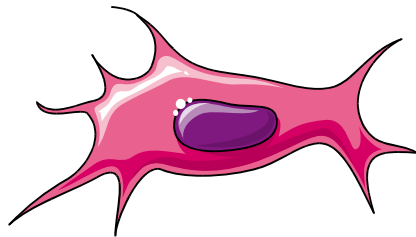
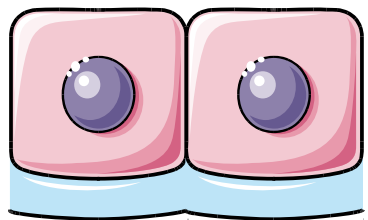
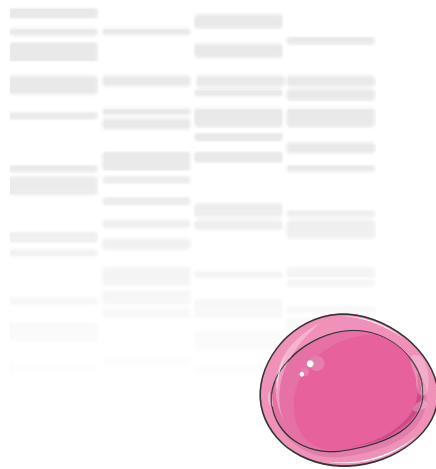
Summary

Transcriptional regulation: a 2-slides introduction

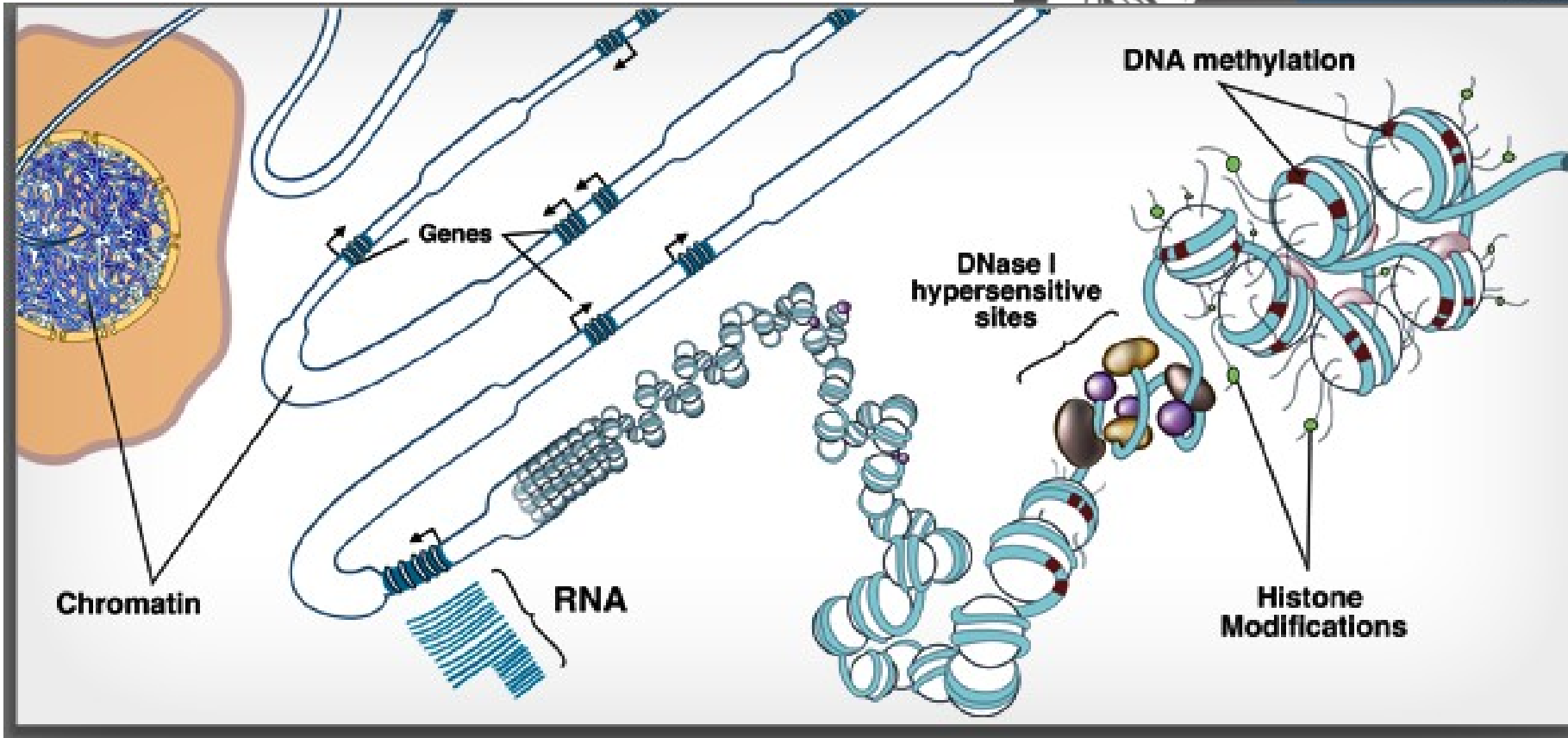
- ❖ MBD2, a major DNA methylation reader
- ❖ Heat*seq web app:

<http://www.heatstarseq.roslin.ed.ac.uk/>

- ❖ Functional annotation transfer using co-expression networks
- ❖ Epigenomics and the human transcript diversity



- ❖ Same genome
- ❖ Distinct cellular phenotypes
- ❖ Differential gene expression
- ❖ Cellular environment + epigenetics





MBD2 binding dynamics during oncogenic transformation

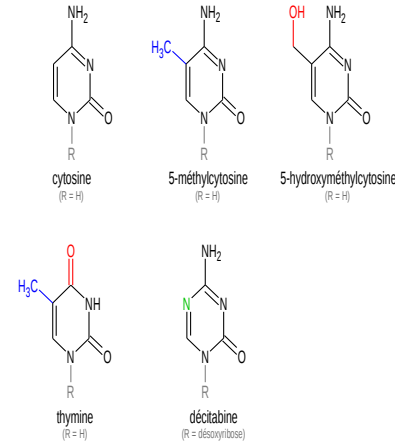
PhD at the Cancer Research Center of Lyon

Supervisor: Robert Dante

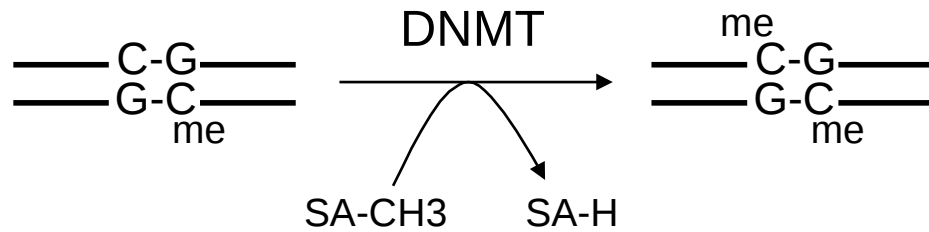
Patrick Mehlen's team

Vertebrate DNA methylation

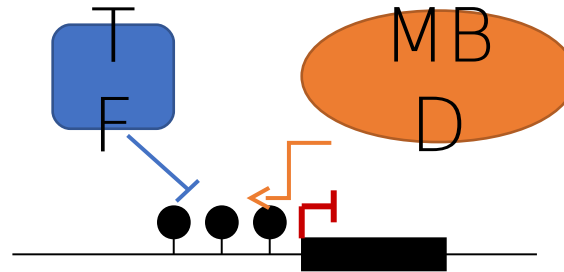
A prototypical epigenetics mark



1- Write



2- Read



3- Erase

- ❖ 5hmC TET + BER
- ❖ BER / NER / MMR
- ❖ passive demethylation

MBD family

MECP2 MBD

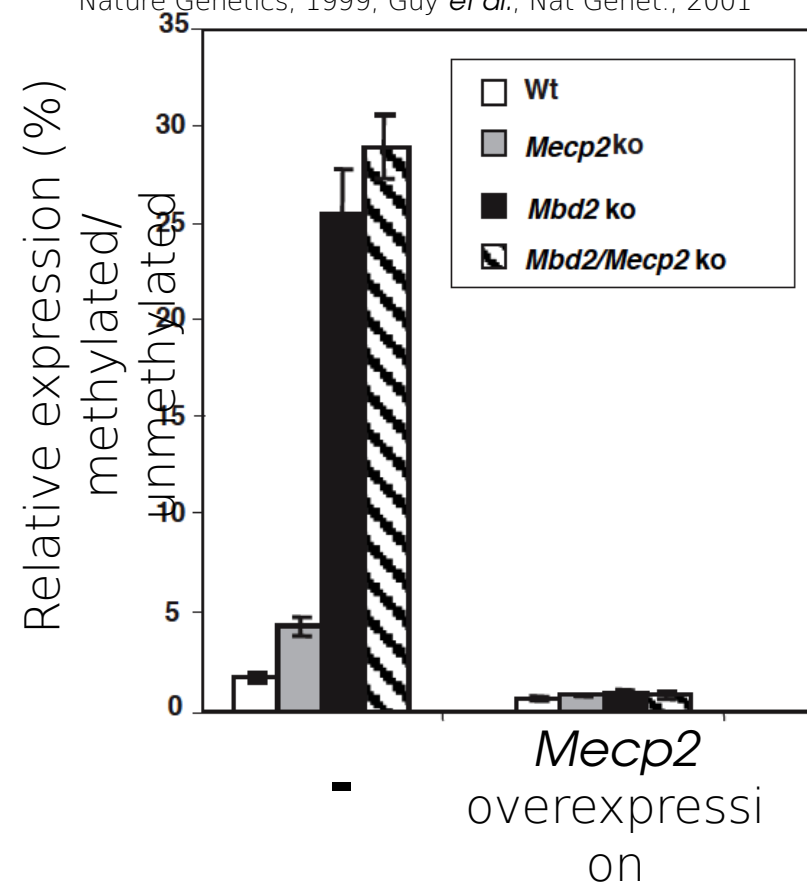
MBD2 1 MBD
4

UHRF family

UHRF UHRF

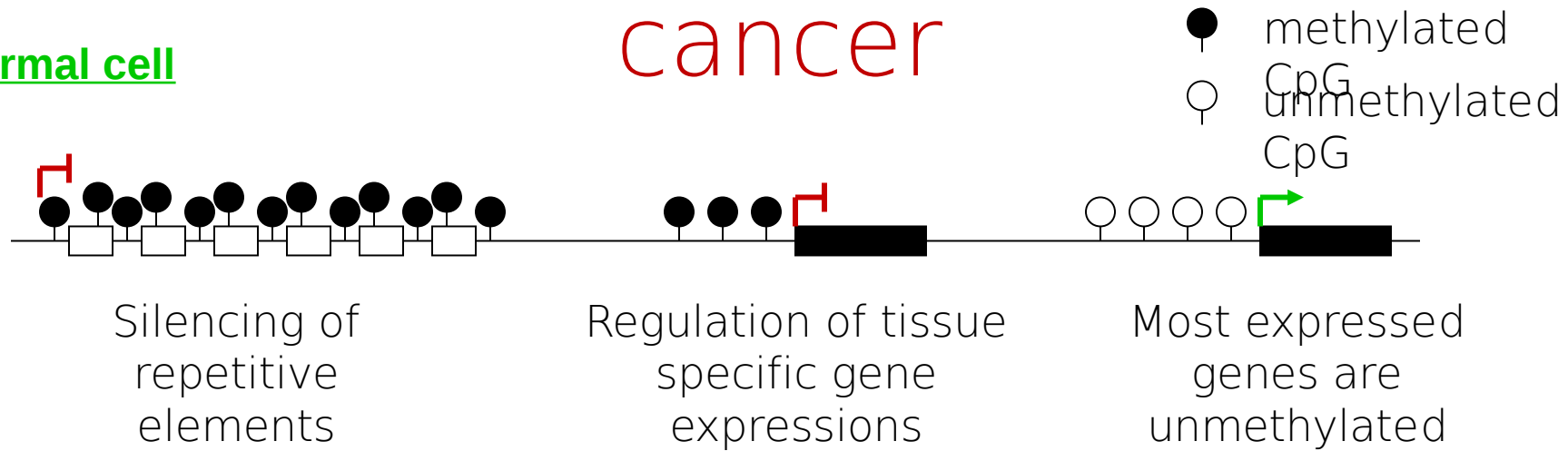
1 2

MBD2 represses transcription of methylated plasmids. Ng *et al.*,
Nature Genetics, 1999, Guy *et al.*, Nat Genet., 2001

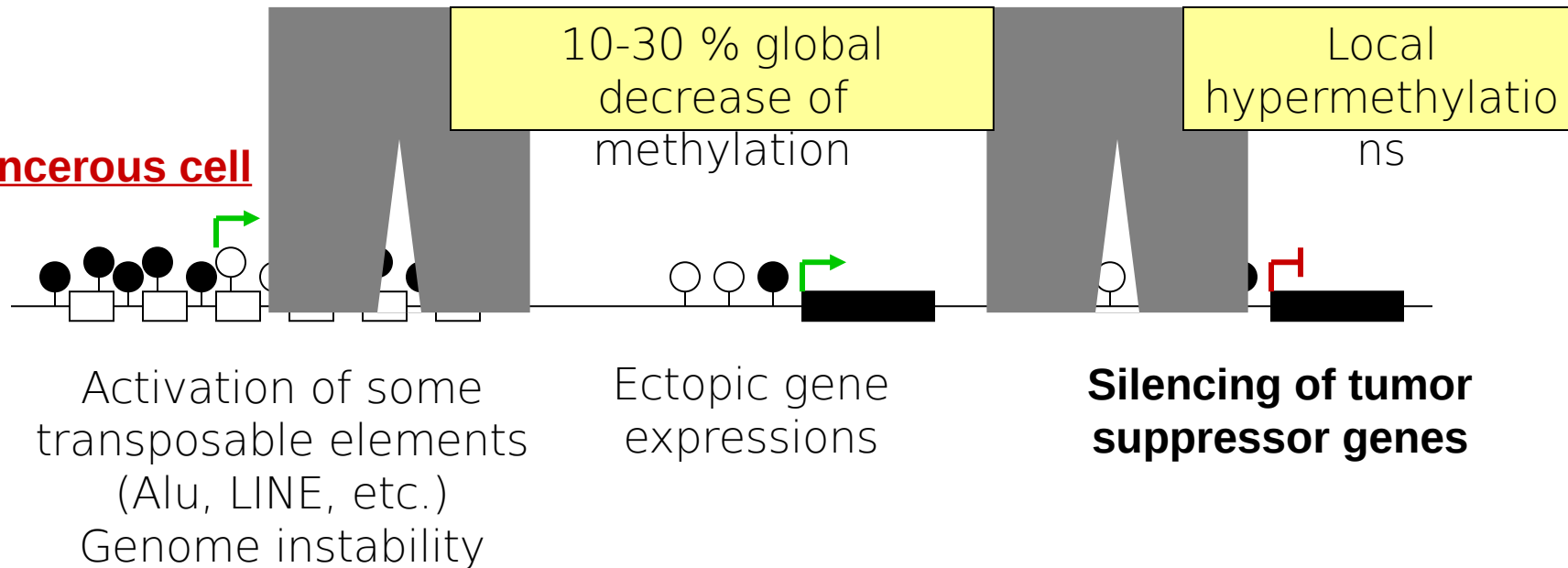


DNA methylation alterations in cancer

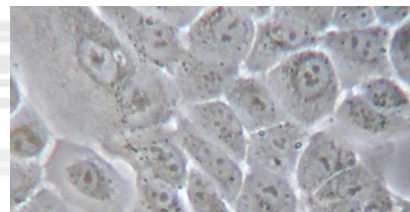
Normal cell



Cancerous cell

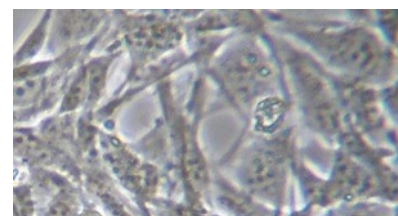


HMEC-hTERT



Immortalized
Non tumorigenic

HMLER



Tumorigenic

→
T/t SV40
HRAS G12V

ChIP-seq:

Endogenous proteins
Pool of 5 experiments

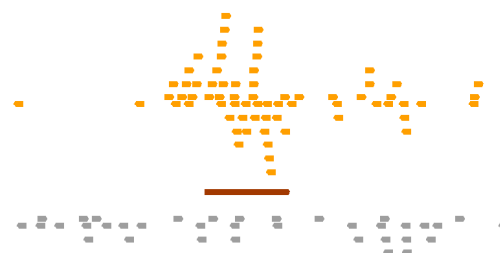
Input
MBD2 ChIP
MeDP (MethylMiner, invitrogen)

RNA-seq:

RNA extract from
untreated cells or cells
treated with **siMBD2** or
DAC

HMEC-hTERT

ChIP
MBD
2
MBD2
peaks
Input



HMLER

ChIP
MBD
2
MBD2
peaks
Input



High throughput

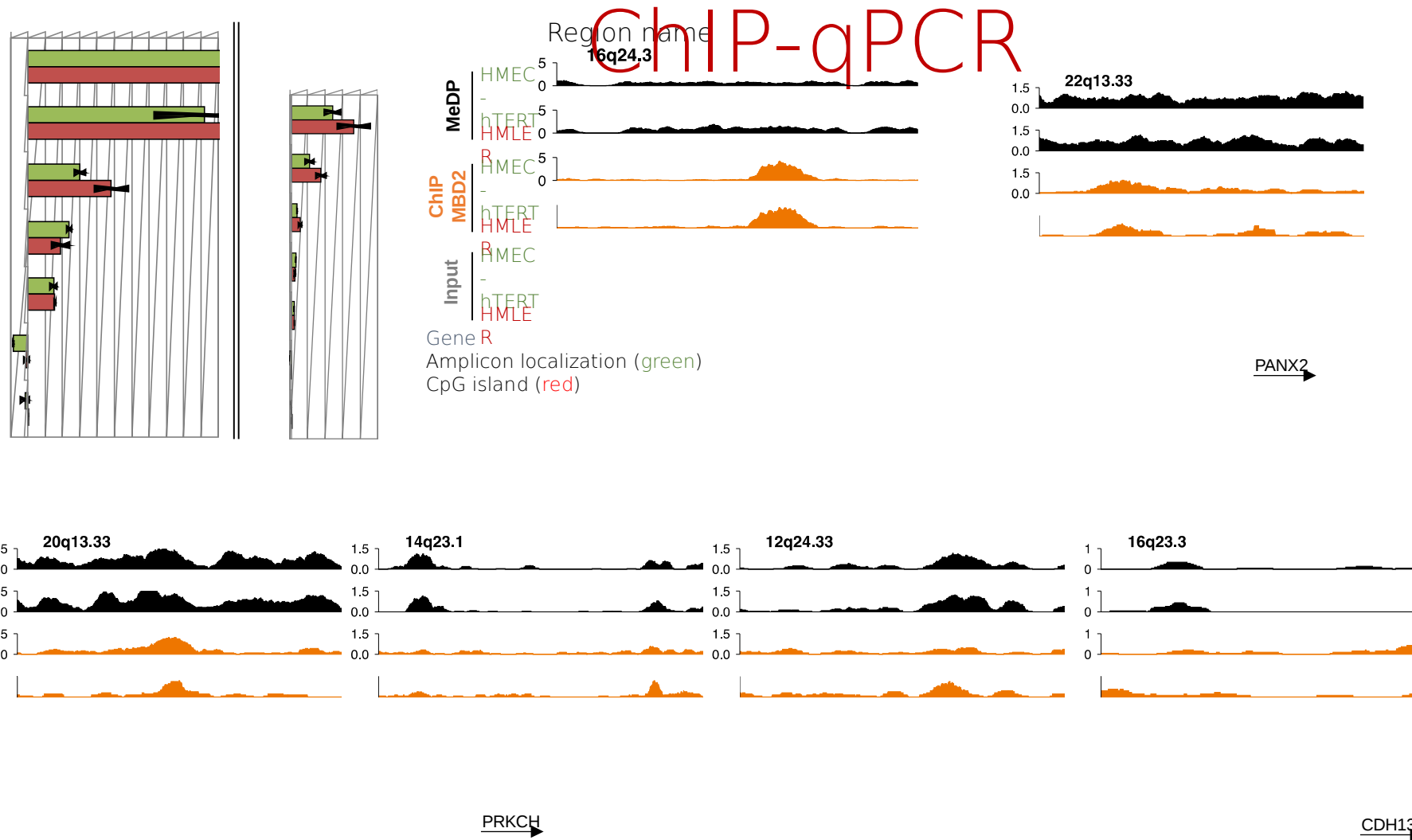
30 - 50 millions of 50 bp reads per
sequencing conditions

Illumina HiSeq 2000



MBD2 ChIPseq validation

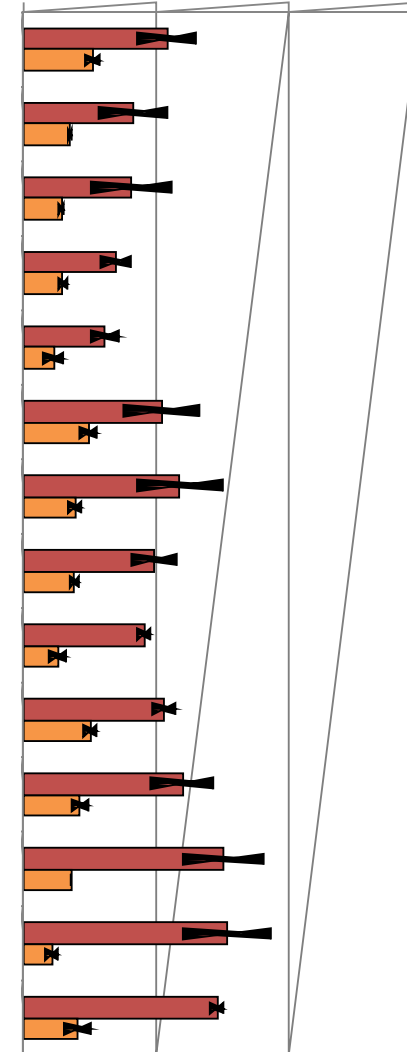
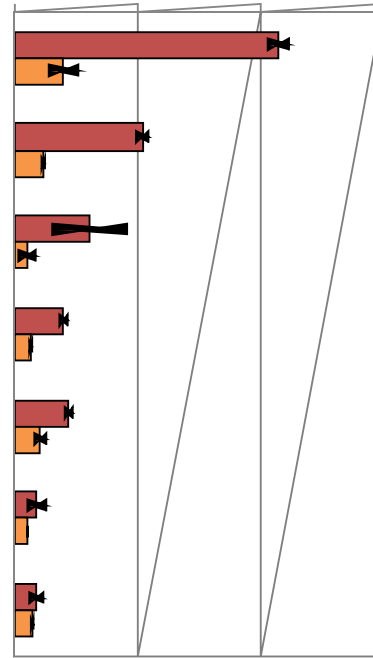
1/3:



MBD2 ChIPseq validation

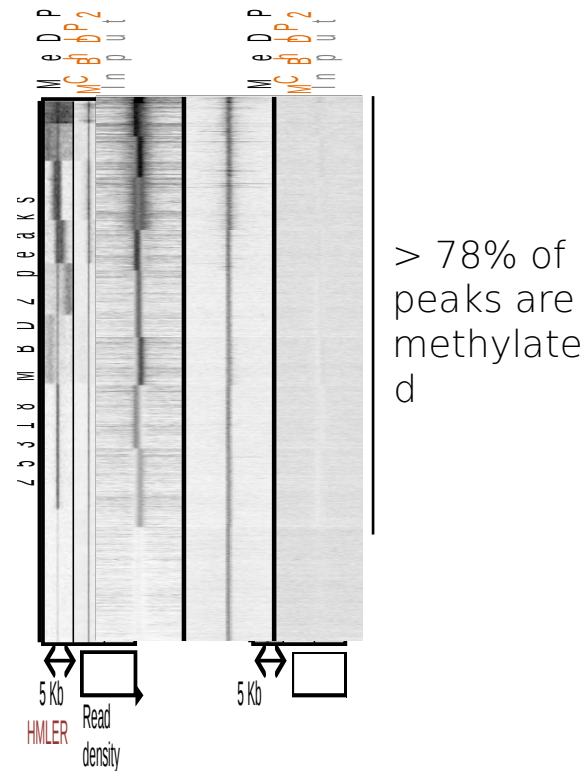
2/3:

MBD2 siRNA



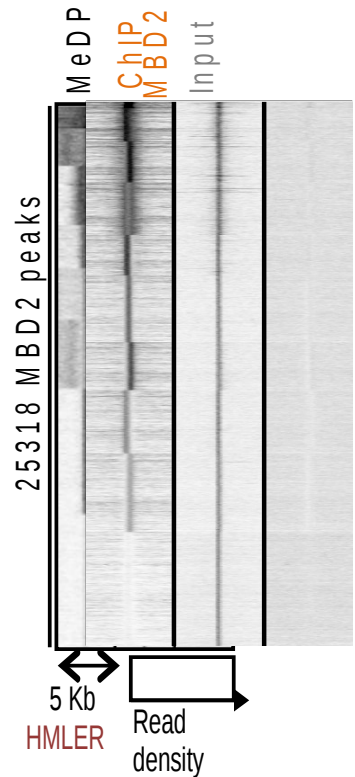
Endogenous MBD2 binds methylated DNA *in vivo*

Almost all
MBD2 binding
sites are
methylated

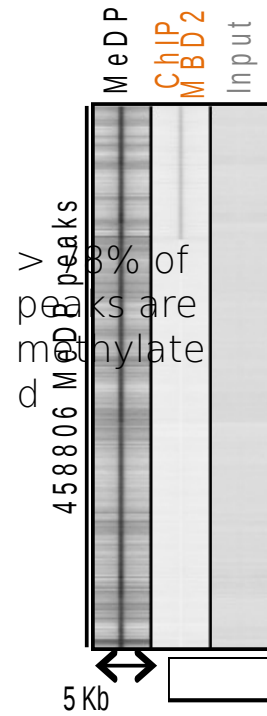


Endogenous MBD2 binds methylated DNA *in vivo*

Almost all
MBD2 binding
sites are
methylated

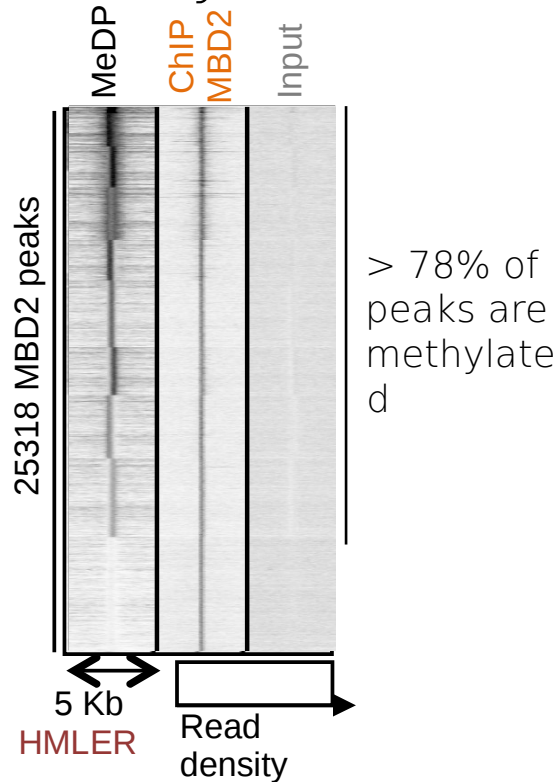


1/4th of
methylated DNA
regions are bound
by endogenous
MBD2

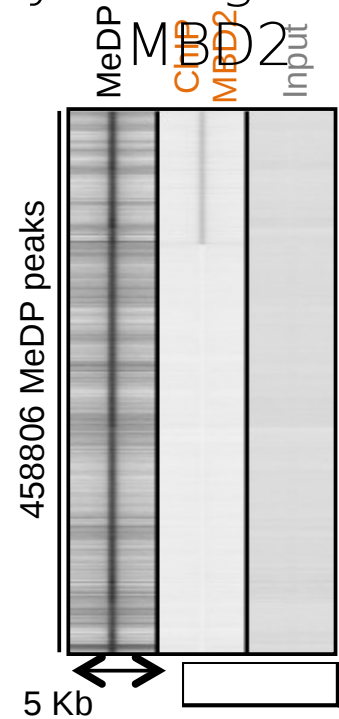


Endogenous MBD2 binds methylated DNA *in vivo*

Almost all
MBD2 binding
sites are
methylated

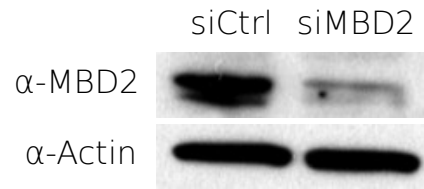


1/4th of
methylated DNA
regions are bound
by endogenous

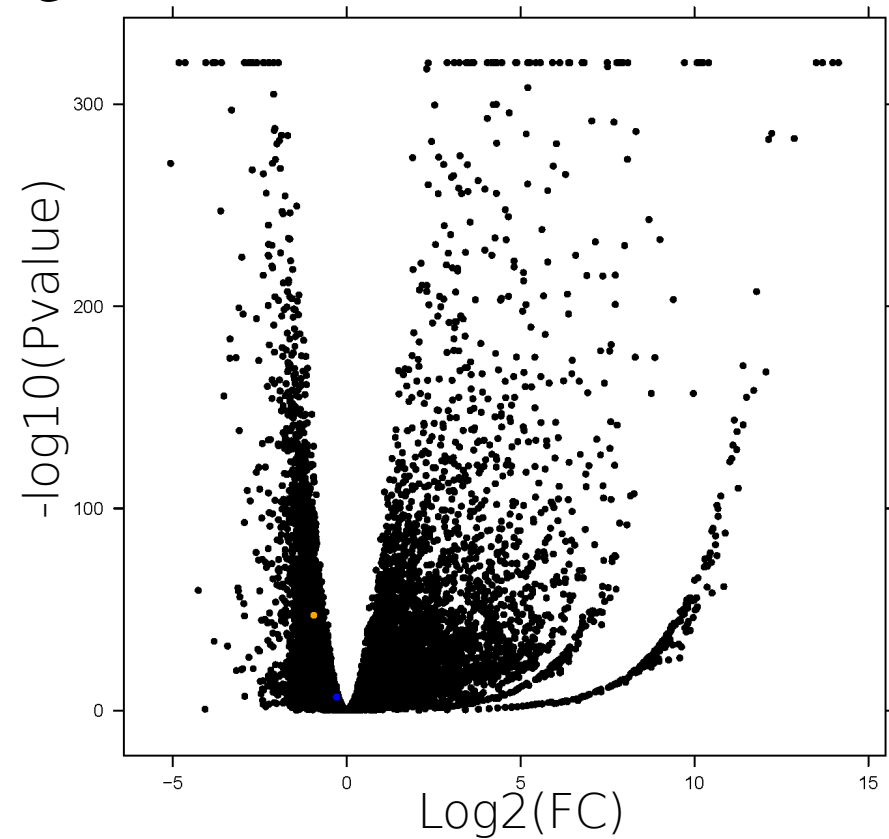
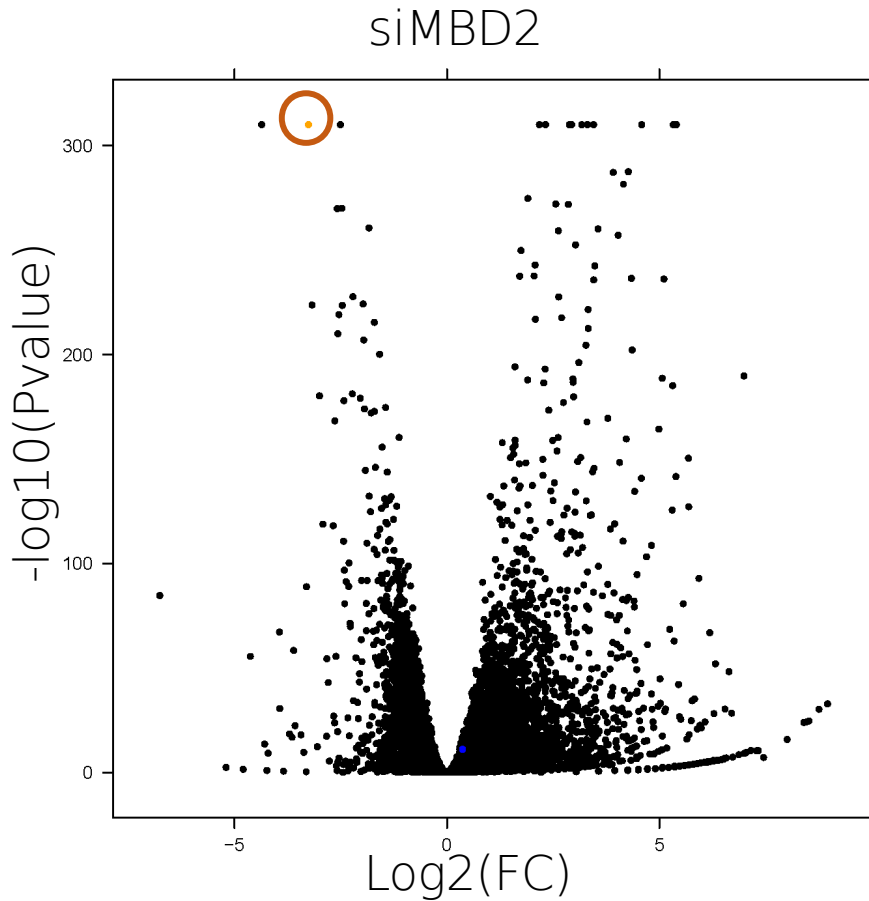


MBD2 binding
on CpG island
depends on
their
methylation
levels

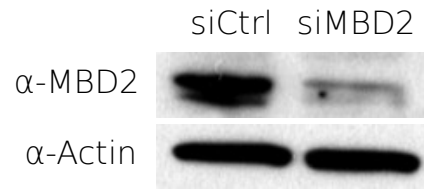
Endogenous MBD2 is a DNA methylation dependent transcriptional repressor



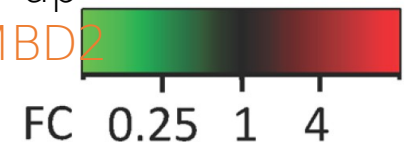
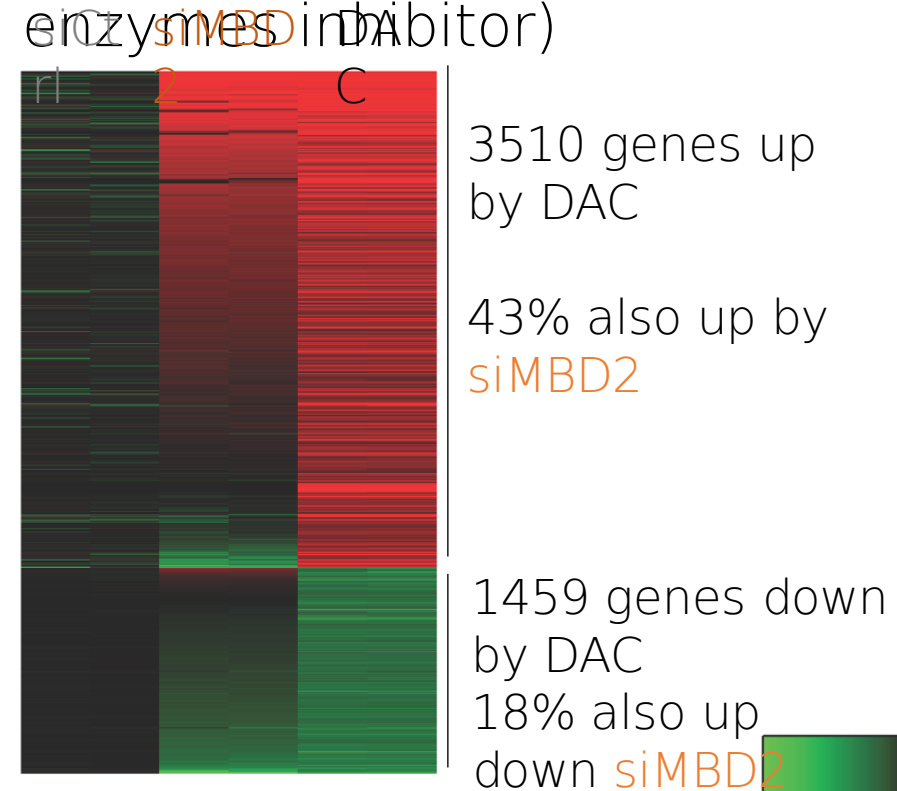
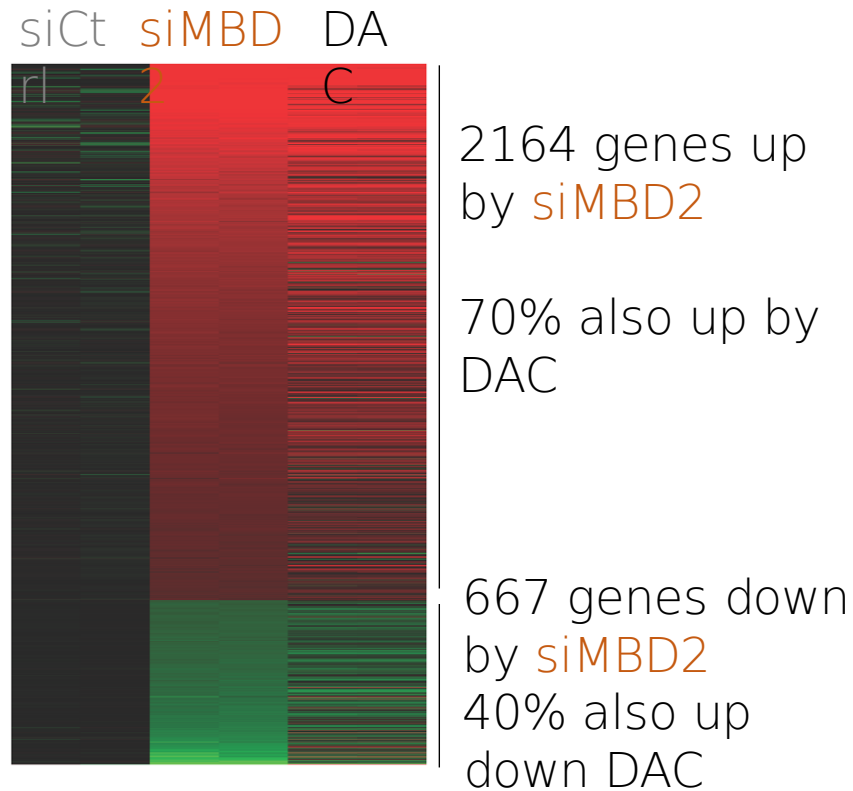
RNAseq after depletion of MBD2 by
siRNA,
or after DAC (5-aza-deoxycytidine)
treatment (DNA methylation
enzymes inhibitor)



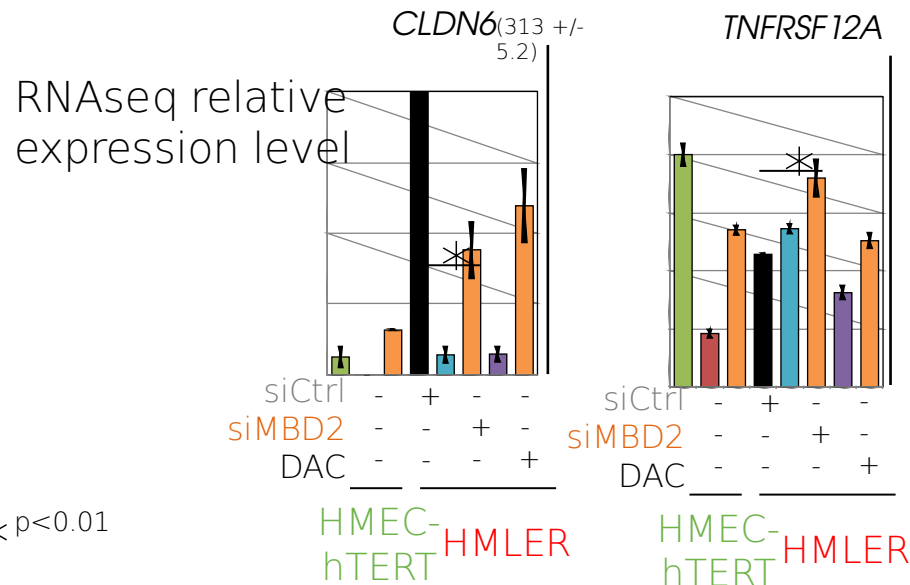
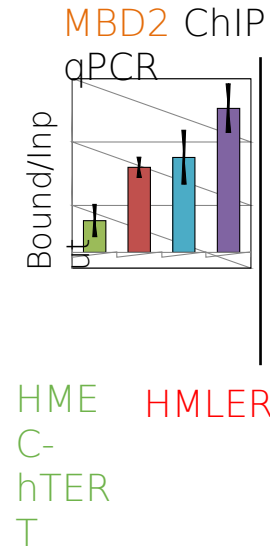
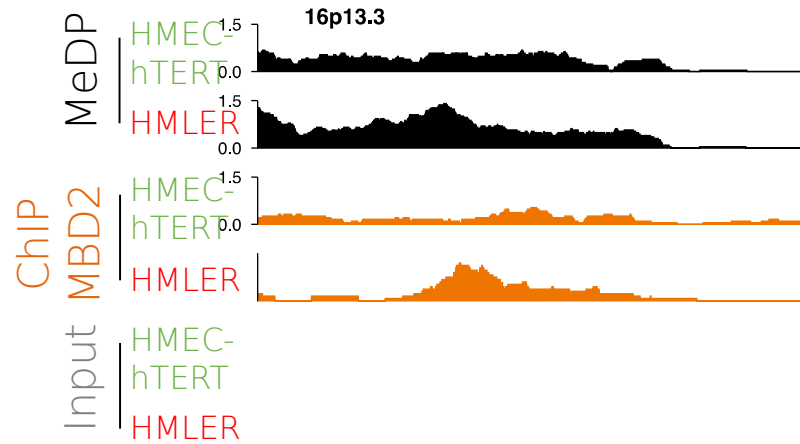
MBD2 is a DNA methylation dependent transcriptional repressor



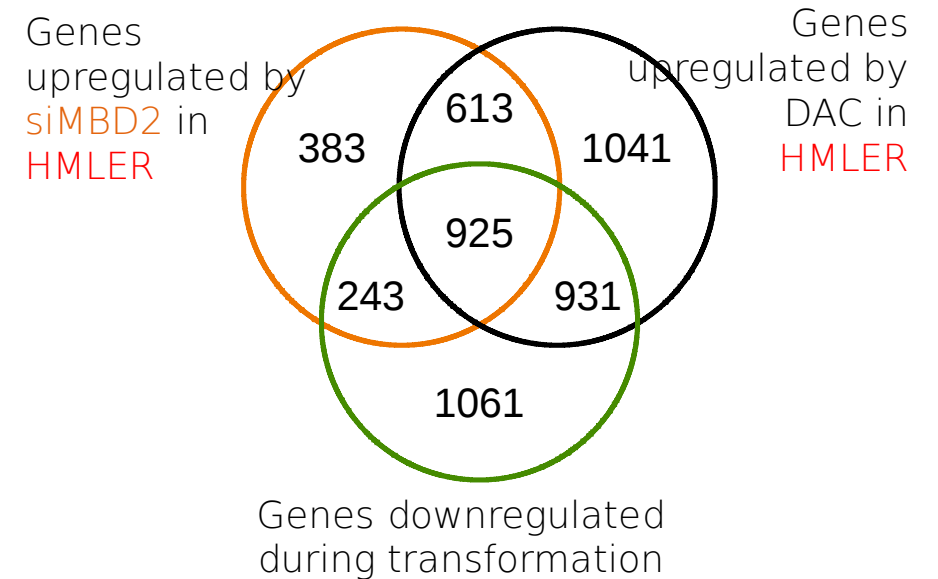
RNAseq after depletion of MBD2 by siRNA,
or after DAC (5-aza-deoxycytidine) treatment (DNA methylation enzymes inhibitor)



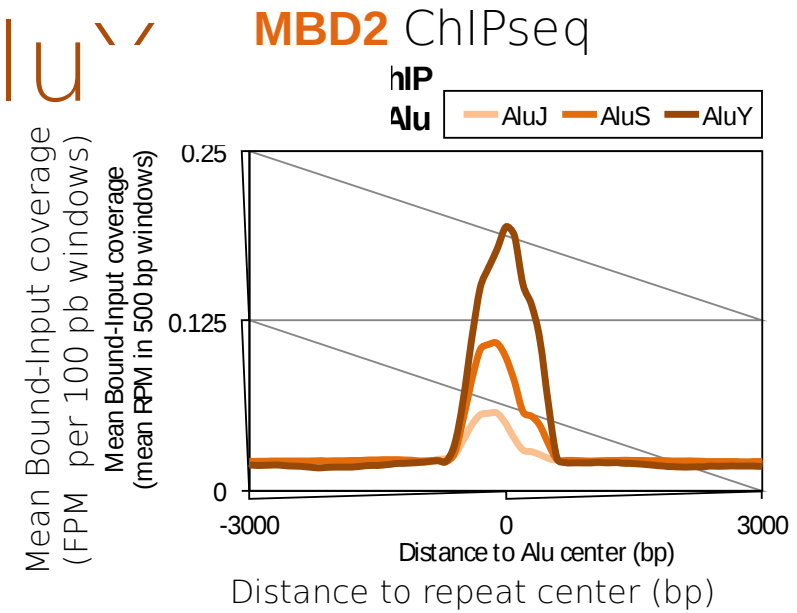
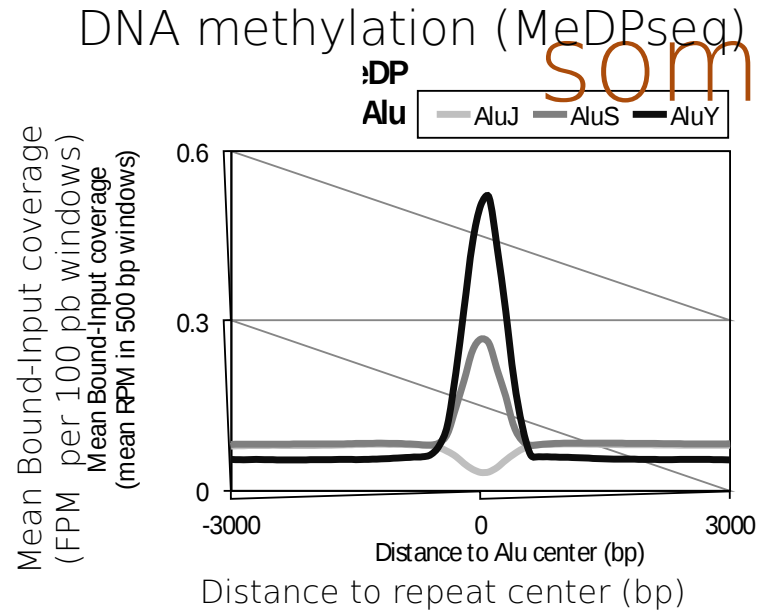
Redistribution of MBD2 during oncogenic transformation: example of *CLDN6*/*TNFRSF12A*



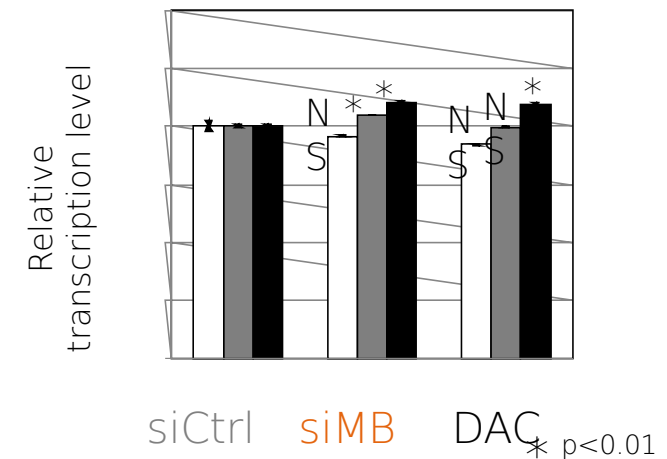
* p<0.01



MBD2 represses transcription of some Alu



Transcription level
(RNAseq)



Conclusions & Perspectives

- ❖ **MBD2** is a major reader of DNA methylation, at least in studied cell lines.
- ❖ **MBD2** is redistributed during oncogenic transformation. This redistribution plays role in the transcriptome modifications of transformed cell lines.
- ❖ What are the mechanism implicated in the redistribution of MBD2?
- ❖ Does targeting MBD2 reduces the transformed phenotype?

<http://ngs-qc.org/navi/index.php>

Accession	Organism	Sample	Experiment	Mapped Reads	QC stamp	
GSM1544114	Homo sapiens	N/A	MBD2 ChIP-seq	57,009,909	BBB	i
GSM1544111	Homo sapiens	N/A	MBD2 ChIP-seq	57,767,880	BBB	i
GSM972973	Mus musculus	ESC, stem cell	MBD2 ChIP-seq	28,211,050	CCB	i
GSM2534656	Homo sapiens	immortalized cell	MBD2 ChIP-seq	23,401,218	CCC	i
GSM1322266	Homo sapiens	N/A	MBD2 ChIP-seq	32,869,475	CCC	i
GSM972994	Mus musculus	ESC, stem cell	MBD2 ChIP-seq	19,671,342	CDD	i
GSM2527610	Homo sapiens	K-562	MBD2 ChIP-seq	23,414,725	CDD	i
GSM1388122	Homo sapiens	N/A	MBD2 ChIP-seq	27,212,212	DCC	i
GSM972978	Mus musculus	N/A	MBD2 ChIP-seq	25,225,988	DCC	i
GSM2527611	Homo sapiens	K-562	MBD2 ChIP-seq	27,605,362	DCC	i
GSM1077269	Mus musculus	ESC, stem cell	MBD2 ChIP-seq	27,360,560	DDC	i
GSM1006707	Homo sapiens	HeLa	MBD2-V5 ChIP-seq	10,943,628	DDD	i

Thanks



Robert Dante
Mélodie Grandin
Pauline Mathot
Véronique Corset
Catherine Guy
Solène Le Guervenel
Duygu Ozmadenci
Laury Perriaud
Anne-Pierre Morel
Patrick Mehlen

Collaborators



CNRS UMR7216, Université Paris 7

Pierre-Antoine Defossez
Olivier Kirsh
Benoit Miotto
Audrey Roussel-Gervais



Inserm
Institut national
de la santé et de la recherche médicale





Heat*seq

Compare your HTS
experiment with

Guillaume Devailly

Anagha Joshi's group, The Roslin Institute, University
of Edinburgh

@G_Devailly

RNA-seq
ChIP-seq datasets
CAGE in



by



THE UNIVERSITY of EDINBURGH



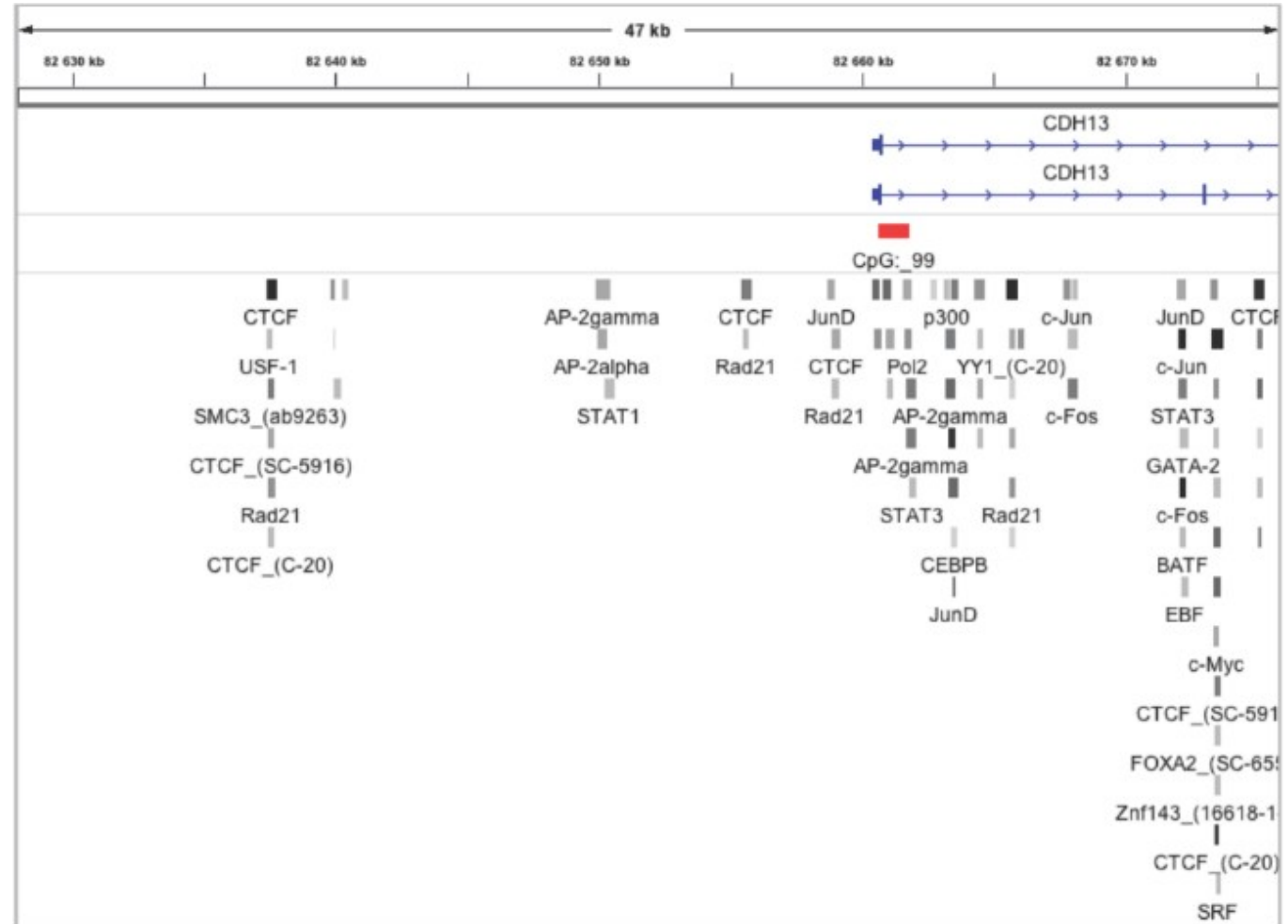
Heat*Seq



HTS is cheap

Big public
datasets

Genome-wide
comparison is
challenging



Workflow



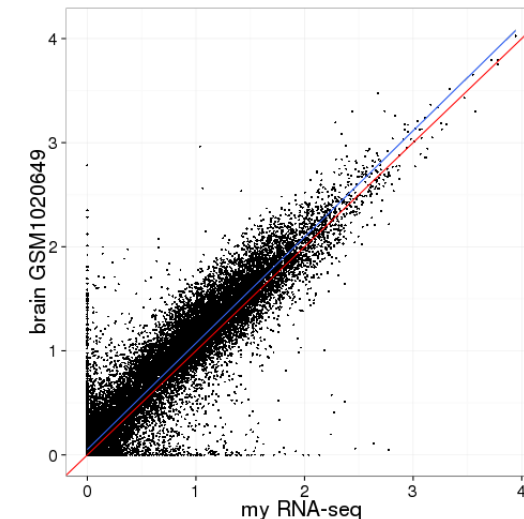
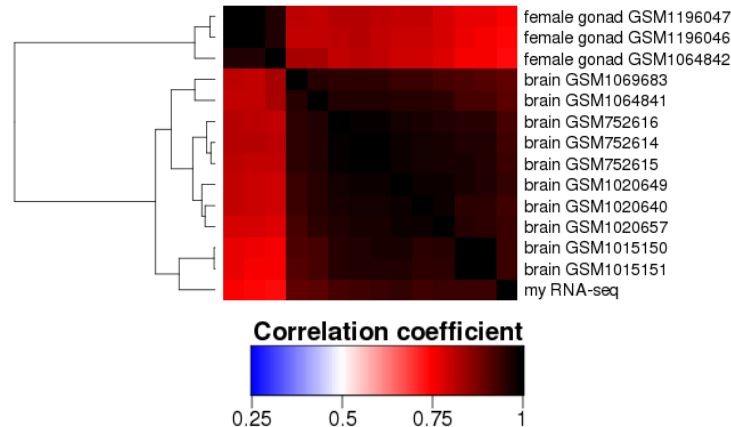
1- Visit website:
www.heatstarseq.roslin.ed.ac.uk

2- Select
dataset

4- Explore
results

3- Upload
processed data

experiment	correlation
brain GSM1020649	0.9496700
Ammon's horn GSM759591	0.9445169
brain GSM1015150	0.9430056
brain GSM1015151	0.9426334
brain GSM1020657	0.9415478
Ammon's horn GSM759593	0.9414239
Ammon's horn GSM759589	0.9400874
Ammon's horn GSM759592	0.9383826
brain GSM752615	0.9375823
brain GSM1020640	0.9374279

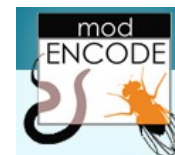


Pearson correlation coefficient: 0.9497
Spearman correlation coefficient: 0.902

Datasets



Assay	Dataset	Organism	Number of experiments
RNA-seq	Bgee	human	77
		mouse	109
	Blueprint epigenome	human	163
	ENCODE	human	302
		mouse	192
	Roadmap Epigenomics	human	57
	Flybase	drosophila	124
TF ChIP-seq	GTEx	human	8555
		human	8555
	ENCODE	human	690
		mouse	156
	CODEX	human	238
		mouse	651
	modENCODE	drosophila	85



HeatRNAseq

Expression matrix

Gene name (~40.000)	Exp 1	Exp 2	Exp 3	...	Exp 77
ENSG00000000000 3	3.983	2.361	10.21 6	...	80.58 3
ENSG00000000000 5	0.071	0.260	0.000	...	3.329
ENSG00000000041 9	10.27 7	2.893	14.15 3	...	42.63 9
...	...	...	...	...	...
ENSG0000027349 3	0.000	0.000	0.000	...	...

Correlation matrix

Pearson's, after log10 scaling

\	Exp 1	Exp 2	Exp 3	...	Exp 77
Exp 1	1	0.94 2	0.93 8	...	0.66 3
Exp 2	0.94 2	1	0.91 7	...	0.68 0
Exp 3	0.93 8	0.91 7	1	...	0.70 6
...	...	...	...	1	...
...	0.66 3	0.68 0	0.70 6	...	1

Clustered heatmap



HeatChIPseq

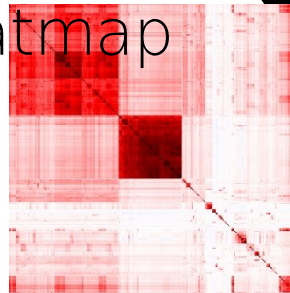
Binary peak

Coordinates (~700.000 non overlapping regions)	Exp 1	Exp 2	Exp 3	...	Exp 690
chr1:10073-10413	F	T	F	...	F
chr1:16110-16390	F	F	T	...	T
chr1:29198-29688	F	F	F	...	F
...	...	...	...	...	...
chrY:28709160- 28709494	T	T	F	...	...

Correlation

	Exp 1	Exp 2	Exp 3	...	Exp 690
Exp 1	1	0.05 9	0.78 6	...	0.03 5
Exp 2	0.05 9	1	0.05 8	...	0.11 8
Exp 3	0.78 6	0.05 8	1	...	0.04 7
...	...	...	...	1	...
	0.03 5	0.11 8	0.04 7	...	1

Clustered heatmap



File format



Tab-delimited text
file

RNA-seq

HeatRNAseq

ChIP-seq

HeatChIPseq

CAGE

HeatCAGEseq

geneID	tpm
ENSG00000134046	120.12
ENSG00000141644	0
ENSG00000169057	85.24
ENSG00000174282	0.54
ENSG00000187098	42
...	...

Ensembl or Flybase gene ID

File format



Tab-delimited text
file

RNA-seq

HeatRNAseq

ChIP-seq

HeatChIPseq

CAGE

HeatCAGEseq

chr	start	end
chr1	125423	125891
chr1	854503 2	854625 4
chr4	452369 8	452478 5
chr12	854120	854870
chrX	245875 0	245987 2
...	...	...

Same genome version
than the dataset

File format



Tab-delimited text
file

RNA-seq

HeatRNAseq

ChIP-seq

HeatChIPseq

CAGE

HeatCAGEseq

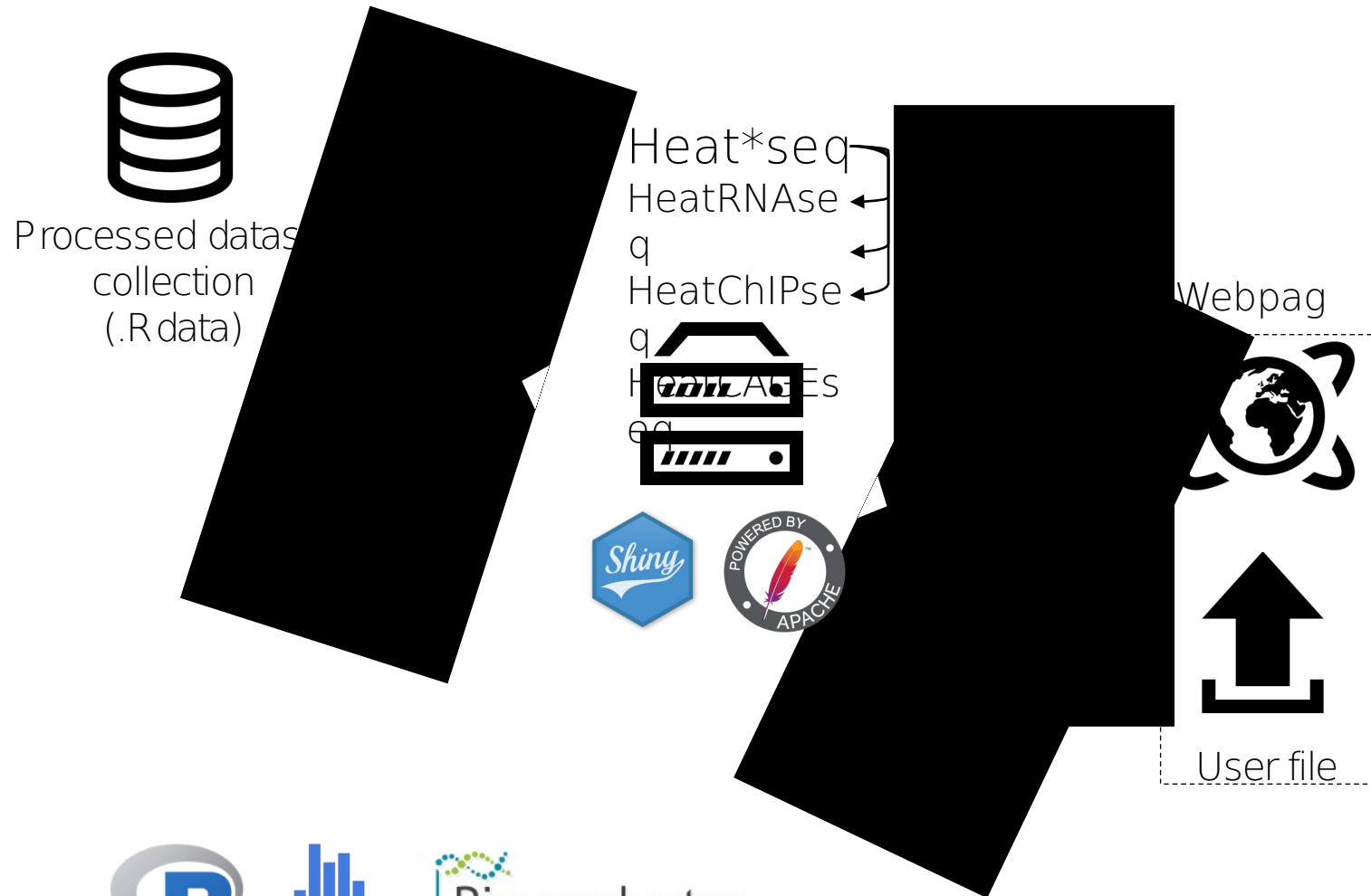
chr	start	end	name	rpm	strand
chr2	25486325	25486487	CAGEpeak _1	458.1 2	+
chr6	5896321	5896380	CAGEpeak _2	25.03	+
chr6	223541	223602	CAGEpeak _3	1.23	-
chr17	5012035	5012100	CAGEpeak _4	45.3	+
chr21	960032	960098	CAGEpeak _5	8.70	-
...	...	...	...	...	...

Implementation



Source code:

https://github.com/gdevailly/HeatStarSeq_gh

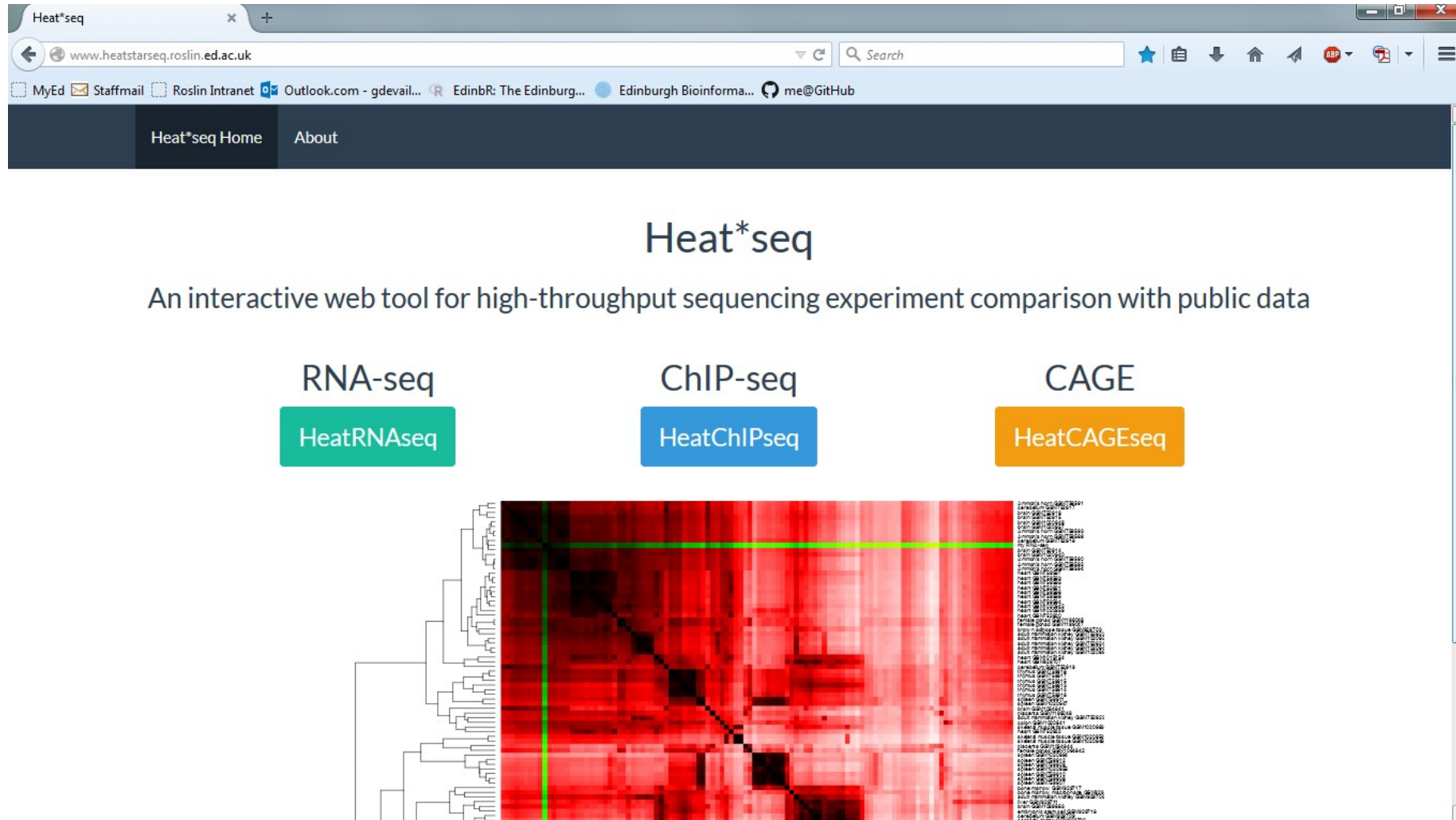


Live demonstration



www.heatstarseq.roslin.ed.ac.uk

If live demo fails...



HeatRNAseq

1 - Select a dataset

Bgee RNA-seq (mouse) ▼

2 - Load your data (optional)

File formatting instructions

☒ Upload your expression file

☐ Use the example file

Choose a file:

Browse...

No file selected.

☒ The expression file contains a header.

Name of your experiment:

my RNA-seq

3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

Uploaded experiment correlation correction:

None ▼

Advanced clustering options

My expression file

Correlation table

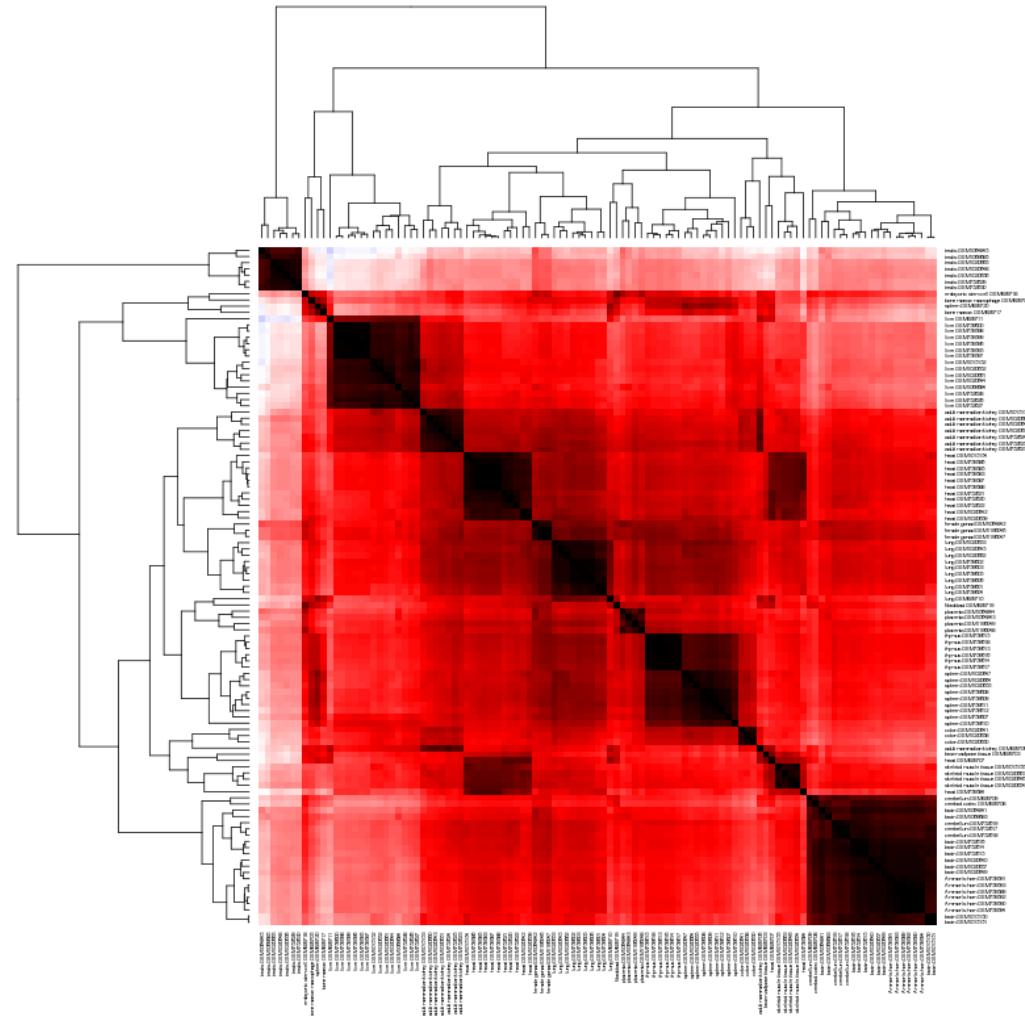
Static heatmap

Responsive heatmap

Tree

Pairwise plot

Samples metadata



HeatRNAseq

1 - Select a dataset

Bgee RNA-seq (mouse)

2 - Load your data (optional)

File formatting instructions

☒ Upload your expression file

☐ Use the example file

Choose a file:

Browse...

No file selected.

☒ The expression file contains a header.

Name of your experiment:

my RNA-seq

3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

Uploaded experiment correlation correction:

None

Advanced clustering options

My expression file

Correla

Samples metadata

HeatRNAseq

1 - Select a dataset

Bgee RNA-seq (mouse)

Bgee RNA-seq (human)

Blueprint RNA-seq (human)

Roadmap Epigenomics RNA-seq (human)

GTEx summary (human)

GTEx - all samples (human)

ENCODE RNA-seq (human)

Bgee RNA-seq (mouse)

ENCODE RNA-seq (mouse)

Browse...

No file selected.



HeatRNAseq

1 - Select a dataset

Bgee RNA-seq (mouse)

2 - Load your data (optional)

File formatting instructions

☒ Upload your expression file

☐ Use the example file

Choose a file:

Browse...

No file selected.

☒ The expression file contains a header.

Name of your experiment:

my RNA-seq

3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

Uploaded experiment correlation correction:

None

Advanced clustering options

My expression file

Correlation table

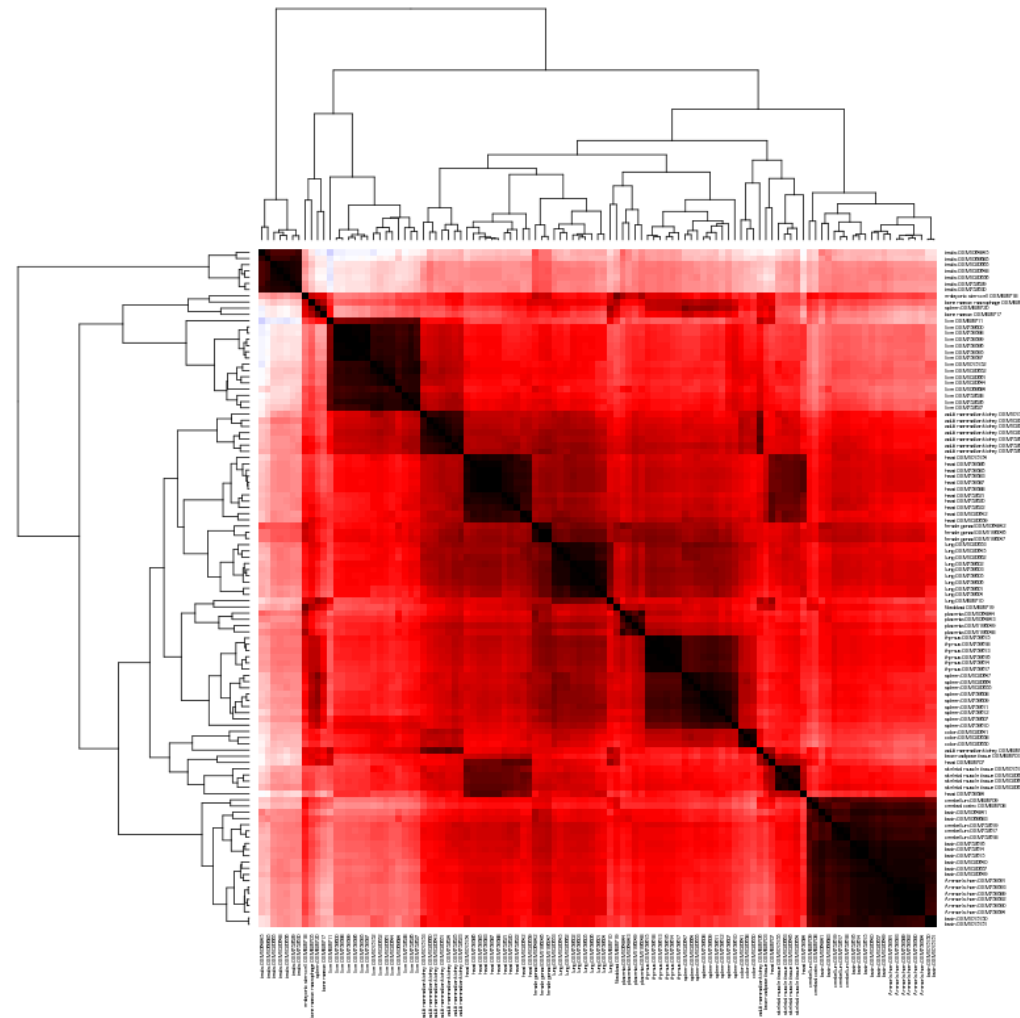
Static heatmap

Responsive heatmap

Tree

Pairwise plot

Samples metadata



HeatRNAseq

1 - Select a dataset

Bgee RNA-seq (mouse) ▼

2 - Load your data (optional)

File formatting instructions

- ☒ Upload your expression file
- ☐ Use the example file

Choose a file:

Browse... mk3_1.txt

Upload complete

☒ The expression file contains a header.

Name of your experiment:

kidney mk3 cell line

3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

My expression file

Correlation table

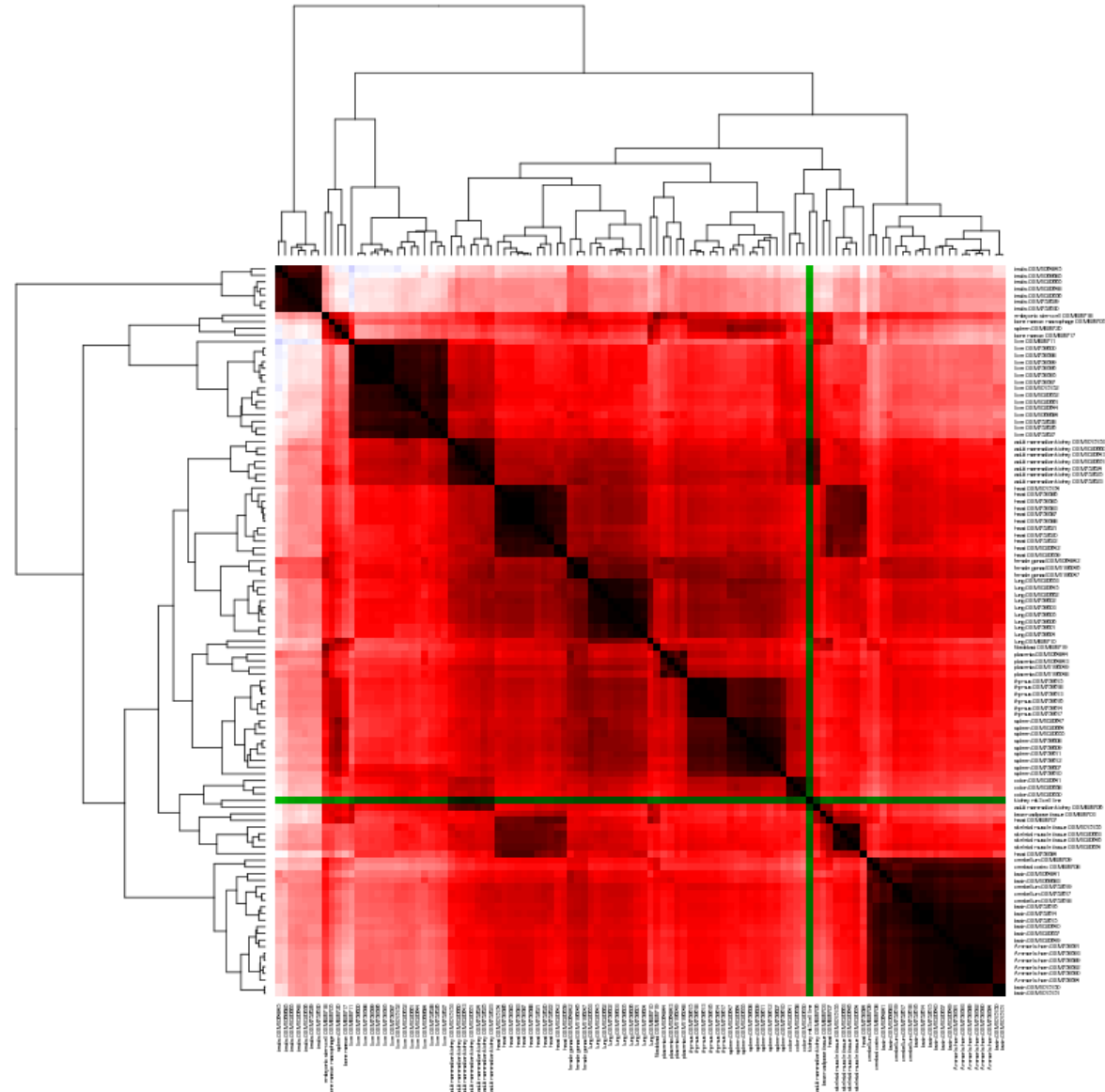
Static heatmap

Responsive heatmap

Tree

Pairwise plot

Samples metadata



3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

Uploaded experiment correlation correction:

None

Advanced clustering options

Label size:

☒ Automatic

☐ Adjust manually

Show dendrogram(s)?

both

Show labels?

both

Customise colours

Tissue (empty to select all):

adult mammalian kidney b

brain

bone marrow

brown adipose tissue

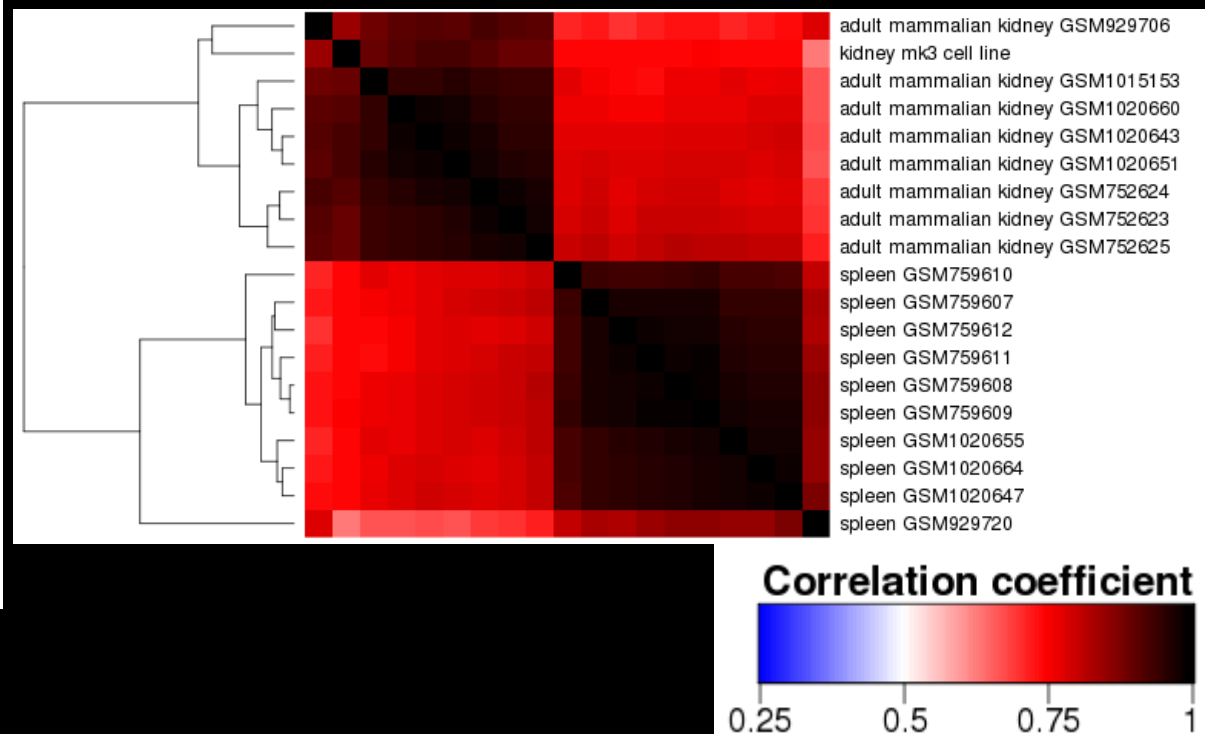
bone marrow macrophage

cerebellum

fibroblast

cerebral cortex

embryonic stem cell



3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

Uploaded experiment correlation correction:

Advanced clustering options

Label size:

- ☒ Automatic
☐ Adjust manually

Show dendrogram(s)?

Show labels?

Customise colours

Distance calculation:

euclidean

Clustering method:

complete

3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

Uploaded experiment correlation correction:

None

Advanced clustering options

Label size:

- ☒ Automatic
☐ Adjust manually

Show dendrogram(s)?

both

Show labels?

both

Customise colours

Customise colours

Colour 1:

#FFFFFF

Value 1:

-1

0.7



Colour 2:

#FFFFFF

Value 2:

0.75

0.8



Colour 3:

#FF8800

Value 3:

0.85

0.9



Colour 4:

#000000

Value 4:

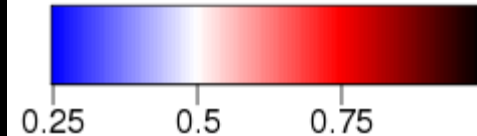
0.95

1

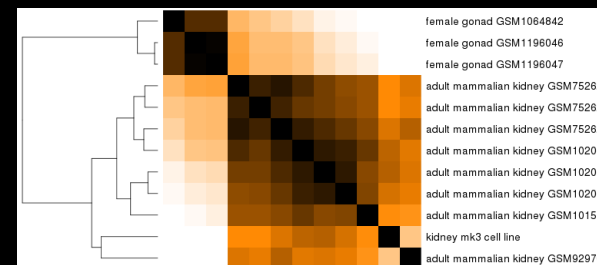
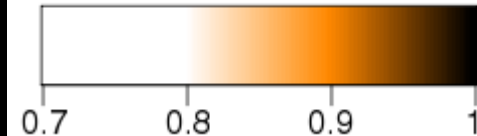


Apply colour changes

Correlation coefficient



Correlation coefficient



File formatting instructions

- ☒ Upload your expression file
- ☐ Use the example file

Choose a file:

No file selected.

☒ The expression file contains a header.

Name of your experiment:

my RNA-seq

3 - Plot customization

☒ Highlight my experiment in the heatmap.

Tissue (empty to select all):

Developmental stage (empty to select all):

Library type (empty to select all):

Uploaded experiment correlation correction:

None

Advanced clustering options

Label size:

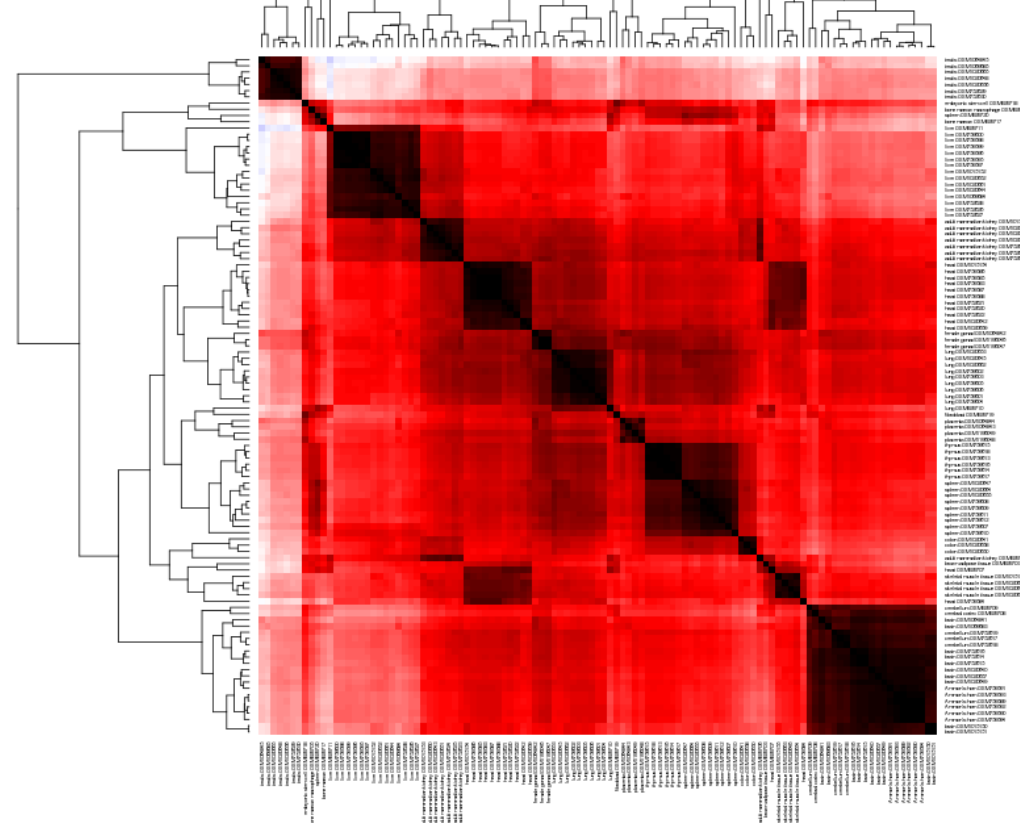
- ☒ Automatic
- ☐ Adjust manually

Show dendrogram(s)?

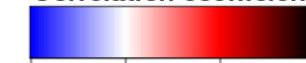
both

Show labels?

both



Correlation coefficient



0.25 0.5 0.75 1

HeatRNAseq

My expression file

Correlation table

Static heatmap

Responsive heatmap

Tree

Pairwise plot

Samples metadata

Show 25 entries

Search:

geneName	value
ENSMUSG000000000001	4.263310e+01
ENSMUSG000000000003	0.000000e+00
ENSMUSG000000000028	7.597450e-01
ENSMUSG000000000031	0.000000e+00
ENSMUSG000000000037	0.000000e+00
ENSMUSG000000000049	1.924590e+01
ENSMUSG000000000056	1.318474e+01
ENSMUSG000000000058	1.964318e+01
ENSMUSG000000000078	7.670450e+00
ENSMUSG000000000085	1.740366e+01
ENSMUSG000000000088	3.714490e+02
ENSMUSG000000000093	2.252150e+01
ENSMUSG000000000094	0.000000e+00
ENSMUSG000000000103	1.650000e-07

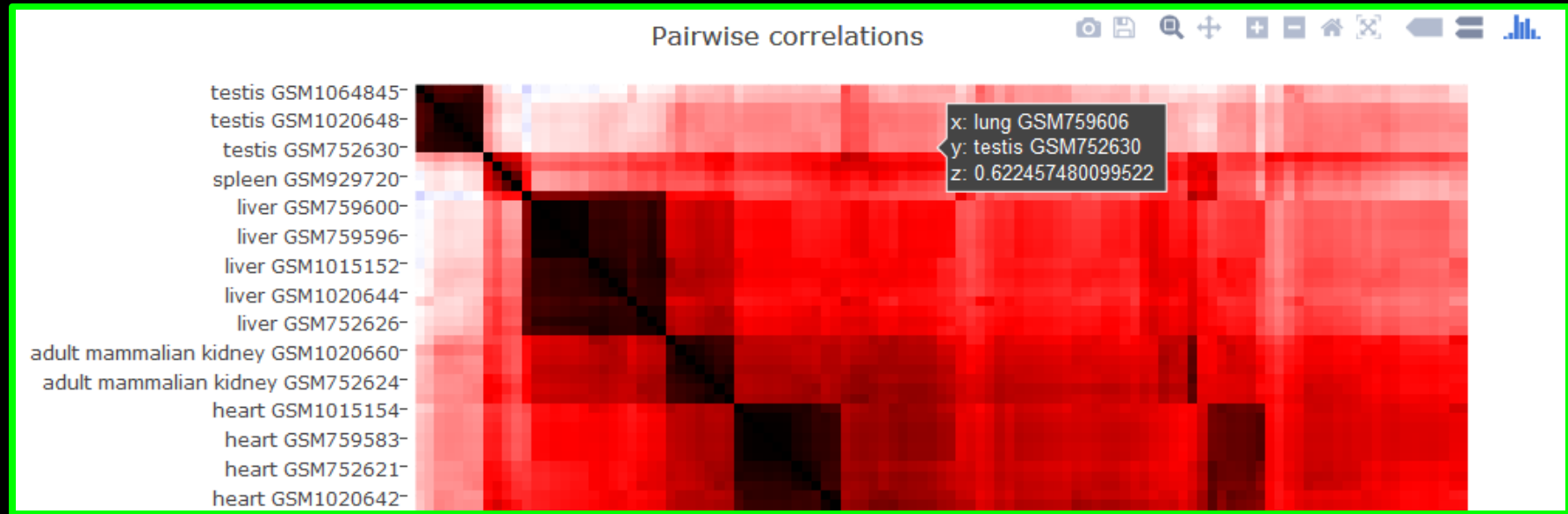
HeatRNAseq

[My expression file](#)[Correlation table](#)[Static heatmap](#)[Responsive heatmap](#)[Tree](#)[Pairwise plot](#)[Samples metadata](#)

Show 25 entries

Search:

experiment	correlation
adult mammalian kidney GSM1020643	0.9280321
adult mammalian kidney GSM1020651	0.9261101
adult mammalian kidney GSM1020660	0.9163906
adult mammalian kidney GSM752624	0.9133589
adult mammalian kidney GSM752625	0.8998140
adult mammalian kidney GSM752623	0.8988885
adult mammalian kidney GSM1015153	0.8955616
adult mammalian kidney GSM929706	0.8497051
female gonad GSM1196047	0.7943039
lung GSM759602	0.7912177
female gonad GSM1196046	0.7909049
colon GSM1020641	0.7901953



Mouse hovering widget

Zoom by drag and drop

Send plot to plotly and share with the world

HeatRNAseq

My expression file

Correlation table

Static heatmap

Responsive heatmap

Tree

Pairwise plot

Samples metadata

Show 25 entries

Search:

geoAccession	tissue	name	stage	libraryType	url
GSM752614	brain	brain GSM752614	post-juvenile adult stage	single	link
GSM752615	brain	brain GSM752615	post-juvenile adult stage	single	link
GSM752616	brain	brain GSM752616	post-juvenile adult stage	single	link
GSM752617	cerebellum	cerebellum GSM752617	post-juvenile adult stage	single	link
GSM752618	cerebellum	cerebellum GSM752618	post-juvenile adult stage	single	link
GSM752619	cerebellum	cerebellum GSM752619	post-juvenile adult stage	single	link
GSM752620	heart	heart GSM752620	post-juvenile adult stage	single	link
GSM752621	heart	heart GSM752621	post-juvenile adult stage	single	link
GSM752622	heart	heart GSM752622	post-juvenile adult stage	single	link

HeatRNAseq

My expression file

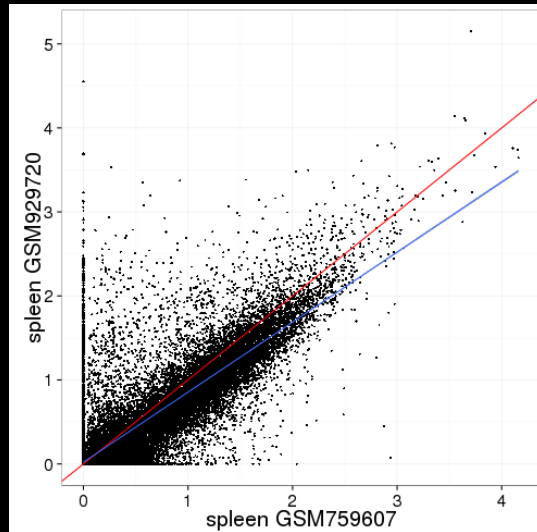
Correlation table

Static heatmap

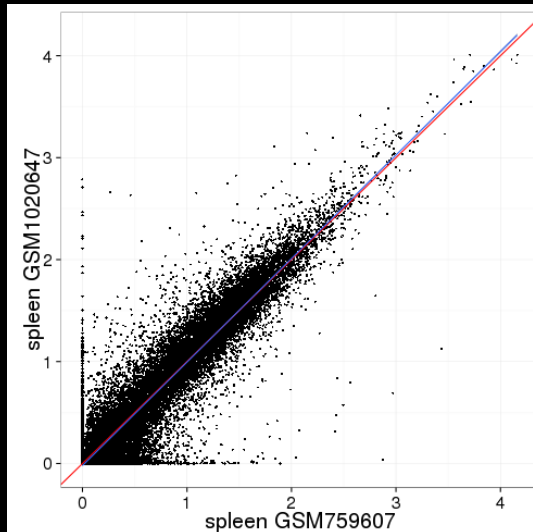
Responsive heatmap

Tree

Pairwise plot



Pearson correlation coefficient: 0.833
Spearman correlation coefficient: 0.7625



Pearson correlation coefficient: 0.944
Spearman correlation coefficient: 0.835

3 - Plot customization

Experiment 1:

adult mammalian kidney GSM1015153

Experiment 2:

adult mammalian kidney GSM1020643

Plot type:

XY

Data scaling:

log10(e + 1)

☒ Add regression line (blue)☒ Add guide line (red)

Conclusions

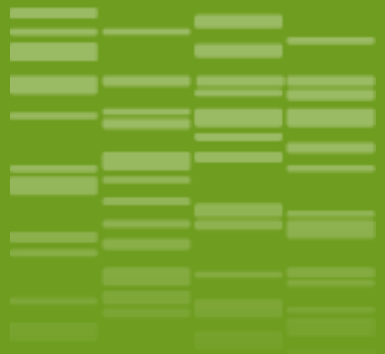


- A lightweight web application
- Quick comparison of RNA-seq, ChIP-seq and CAGE experiments
- First step to guide you before in depth analysis

What's next



- Other datasets (suggestions?)
- Other data types (Histones, DNase1, Gene Lists/Ontologies)
- Multiple user files
- Gene name converter / *liftover* integration



Functionnal annotation transfert using co-expression networks

Pía Francesca Loren Reyes, Tom Michoel, Anagha Joshi,
Guillaume Devailly



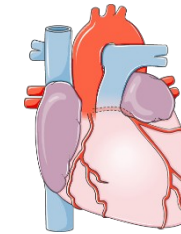
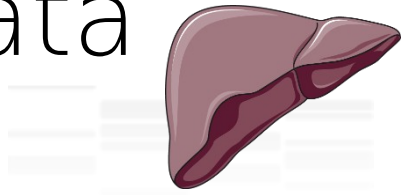
- ❖ Plenty of *omic* data
- ❖ Not enough functional annotations :(
- ❖ Can we use *omic* data to improve functional annotations?

- ❖ Functional annotations are often inferred from other species
- ❖ This lead to over-annotations mistakes, notably in the cases of one-to-many or many-to-many homology groups

Hypothesis: if two orthologs are in the **same co-expression cluster** in different species, then functional annotation are likely to be transferable.

And co-expression cluster are easy to build from public *omic* data! \o/

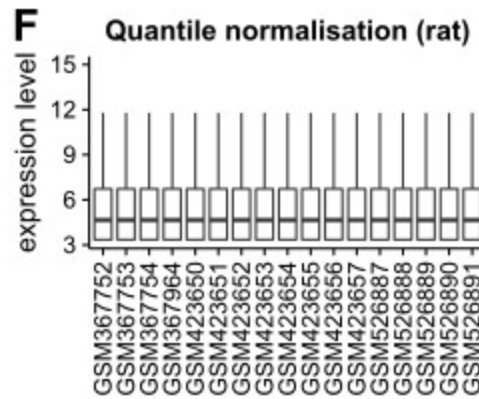
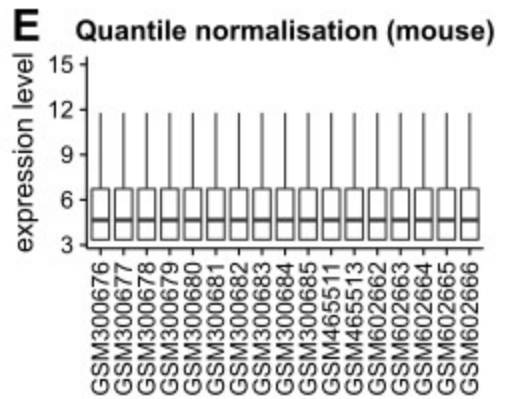
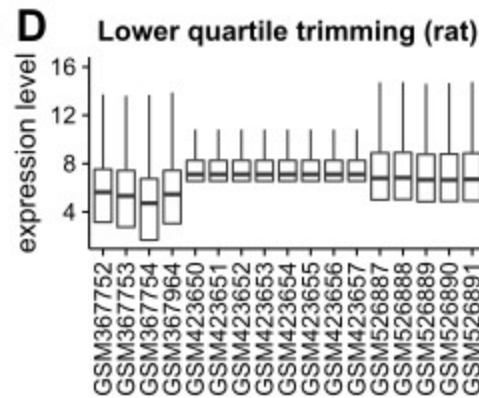
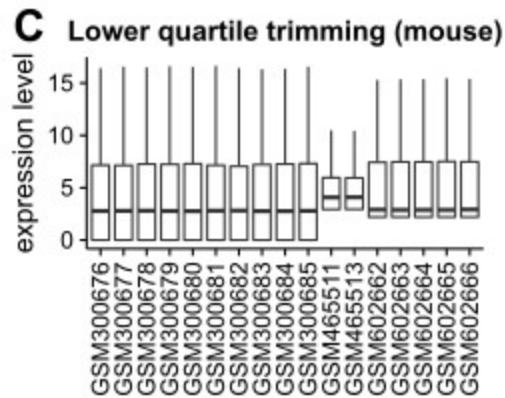
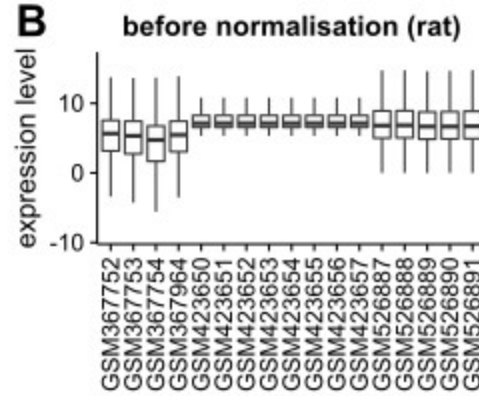
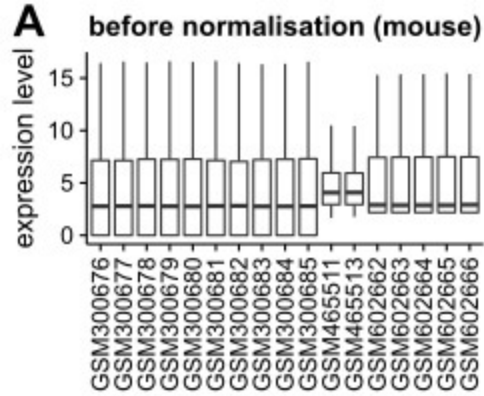
Public microarray expression data



Mous		Rat	
Series ID	n	Series ID	n
<i>total</i>	920	<i>total</i>	620
GSE50789	96	GSE13270	101
GSE9630	59	GSE59495	90
GSE55756	47	GSE24104	47
GSE63027	39	GSE5509	40
GSE51885	27	GSE23748	30
GSE38067	24	GSE27625	30
...	...	...	...

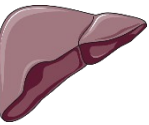
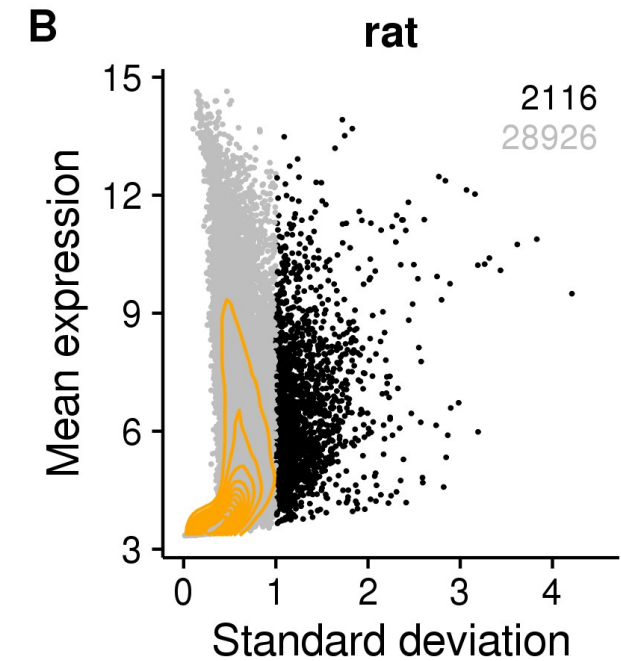
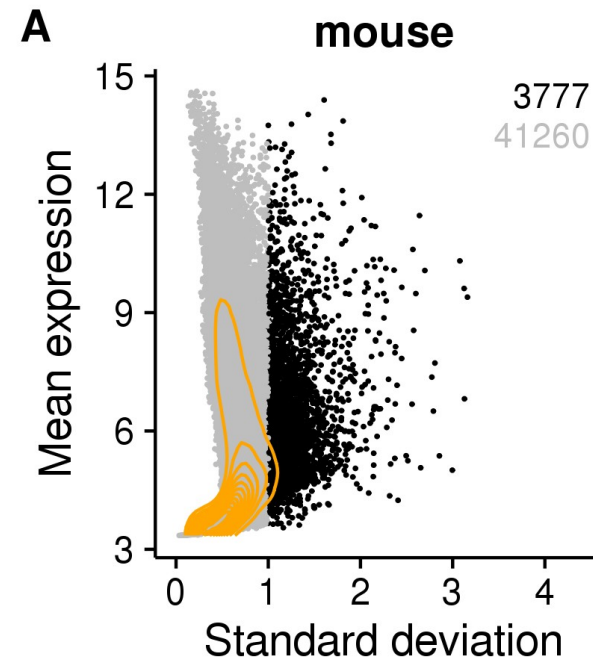
Mous		Rat	
Series ID	n	Series ID	n
<i>total</i>	248	<i>total</i>	1202
GSE1479	36	GSE57822	433
GSE3530	36	GSE57800	429
GSE7487	24	GSE19290	82
GSE5500	21	GSE6104	45
GSE7605	18	GSE7999	30
GSE3440	15	GSE11851	20
...	...	...	...

Affymetrix Mouse Genome 430 2.0
Affymetrix Rat Genome 230 2.0

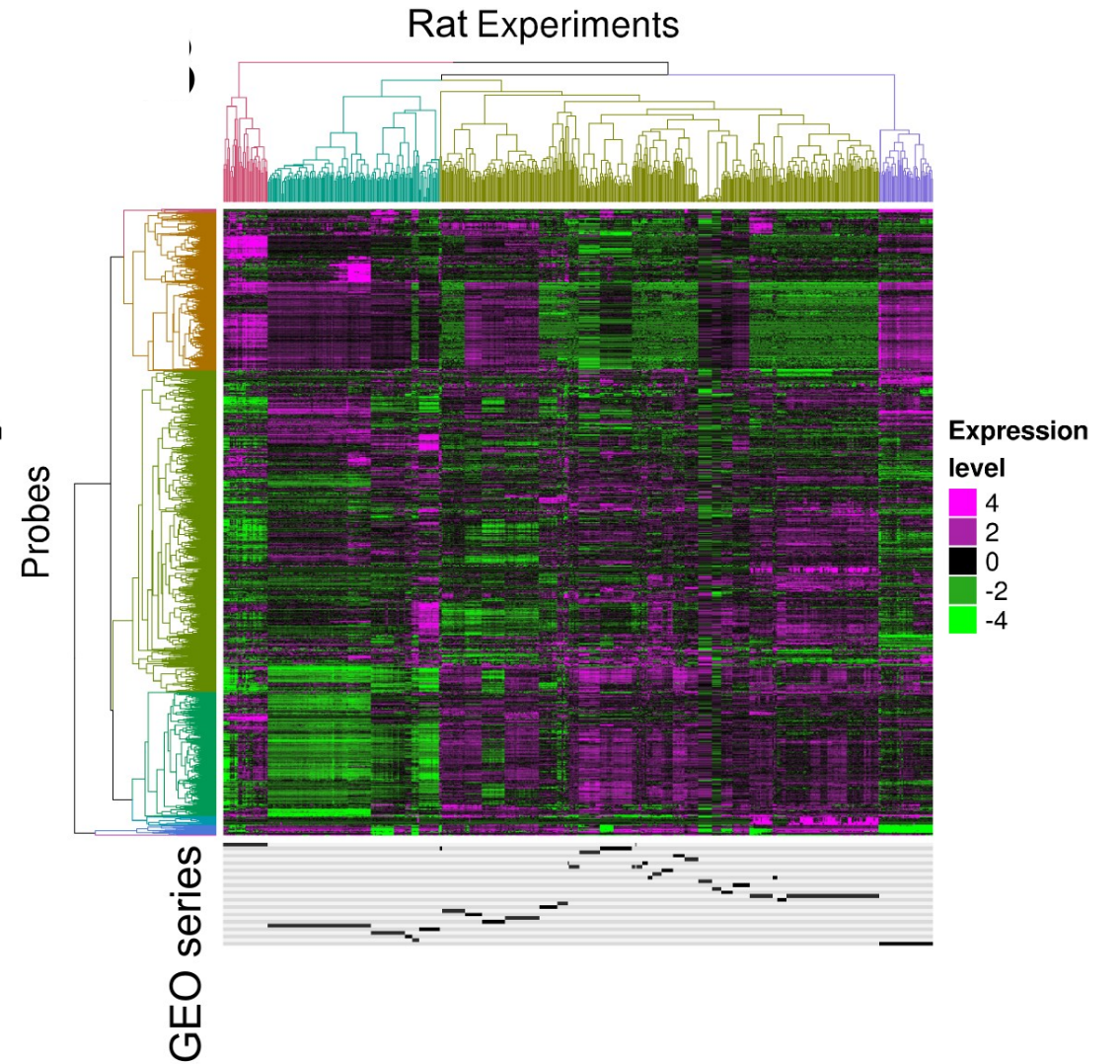
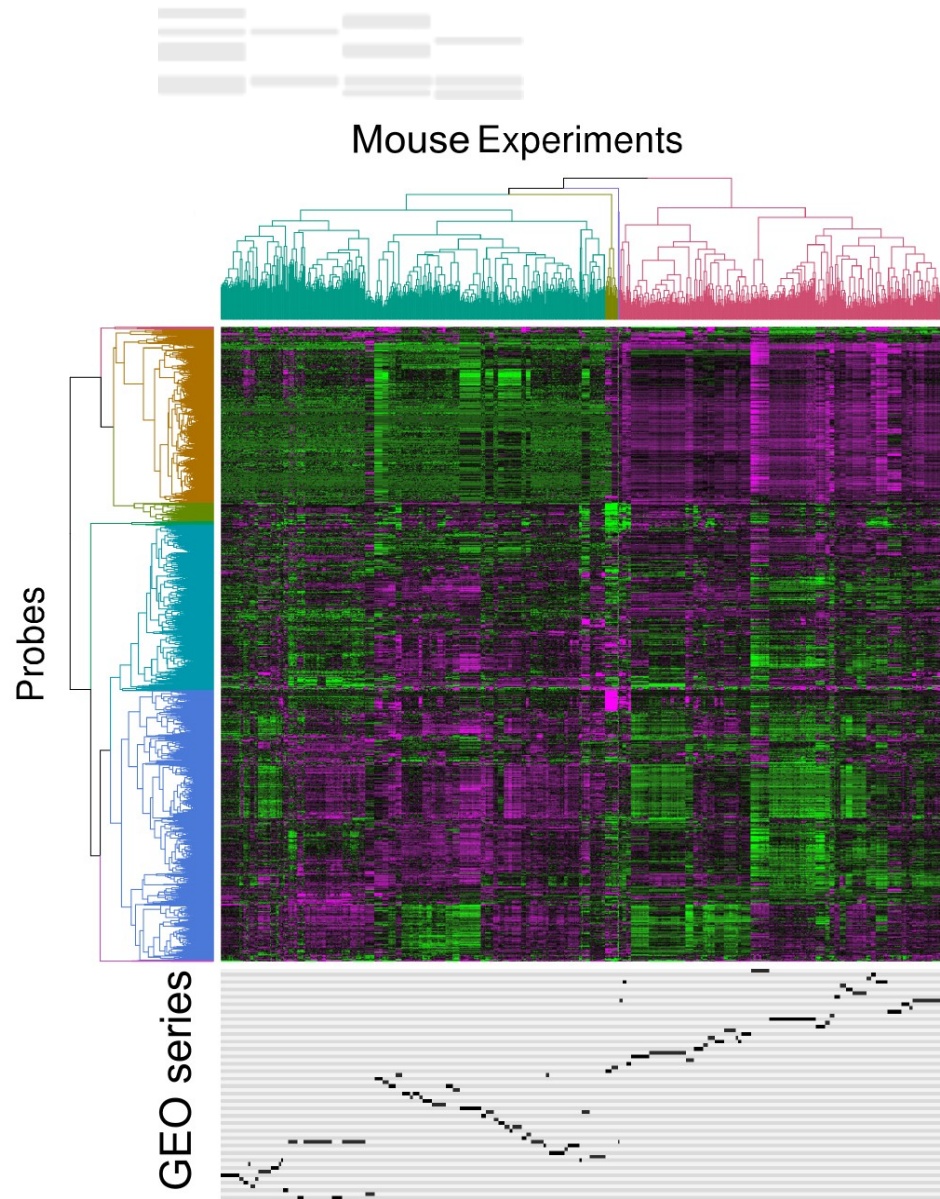


Data normalization

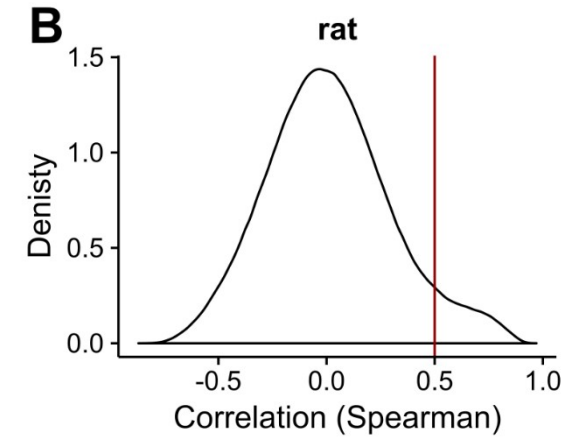
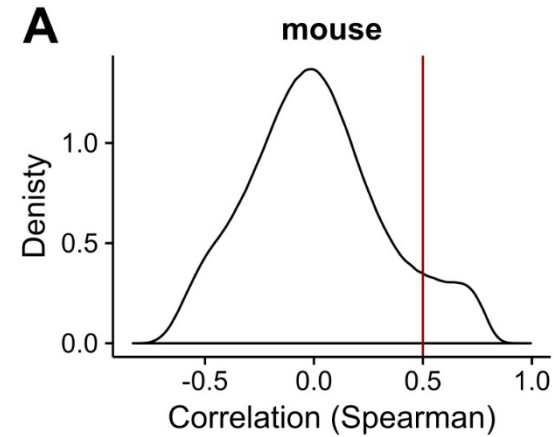
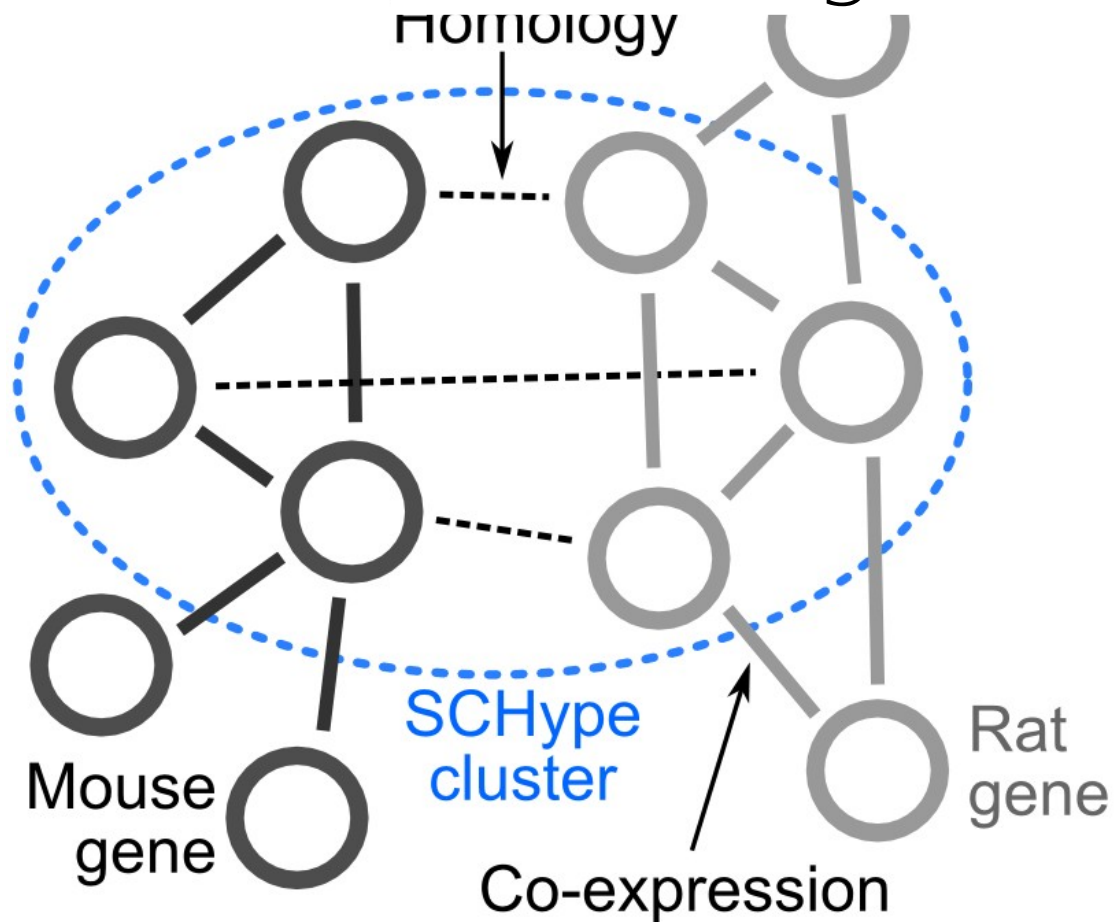
Identification of variable probes



Species	Category	Term	Gene	FE	P-value
Mouse	Reactome	Synthesis of (16-20)-hydroxyeicosatetraenoic acids (HETE)	11	4.78	4.29E-02
		Activation of gene expression by SREBF (SREBP)	15	4.34	5.18E-03
		Regulation of cholesterol biosynthesis by SREBP (SREBF)	17	3.94	4.36E-03
		Cytochrome P450 - arranged by substrate type	27	2.72	7.78E-03
	GO slim BP	Phase 1 - Functionalization of compounds	37	2.55	7.58E-04
		fatty acid metabolic process	52	2.26	2.95E-05
		steroid metabolic process	50	2.18	1.31E-04
Rat	Reactome	Synthesis of bile acids and bile salts via 24-hydroxycholesterol	7	8.63	2.95E-02
		Endosomal/Vacuolar pathway	10	7.93	1.15E-03
		Striated Muscle Contraction	11	6.78	1.48E-03
		ER-Phagosome pathway	10	6.53	6.29E-03
		Activation of gene expression by SREBF (SREBP)	10	6.53	6.29E-03
		Antigen Presentation: Folding, assembly and peptide loading of class I MHC	13	6.27	3.84E-04
		Regulation of cholesterol biosynthesis by SREBP (SREBF)	10	5.55	2.51E-02
		Biological oxidations	25	2.95	3.58E-03
	GO slim BP	Metabolism of lipids and lipoproteins	68	2.13	1.15E-05
		response to biotic stimulus	12	4.16	1.12E-02
		fatty acid metabolic process	22	2.52	2.52E-02

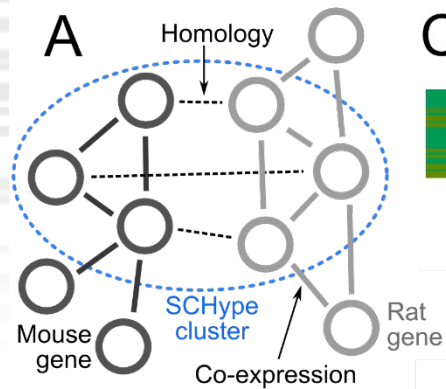


SCHype hyper-graph clustering tool



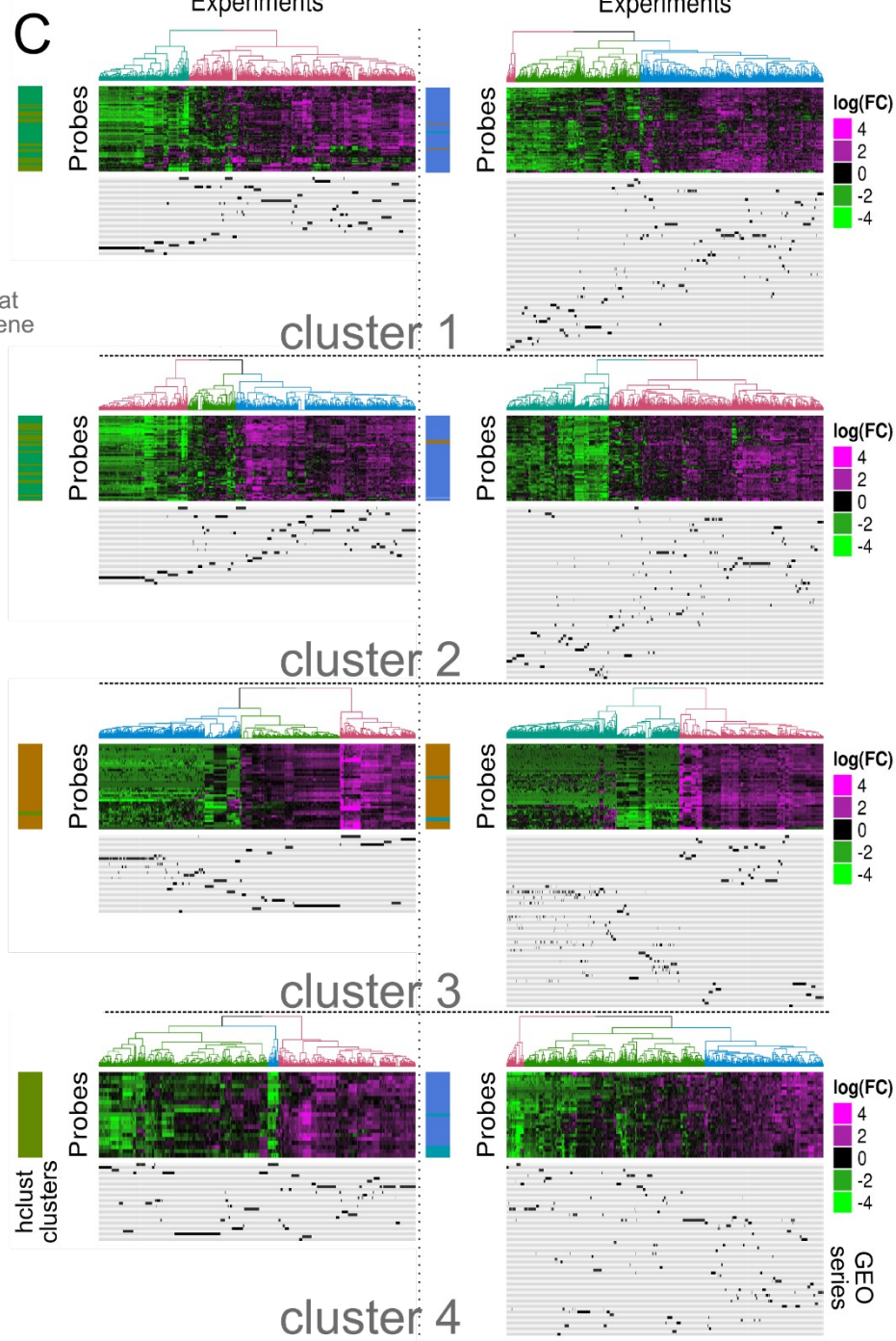
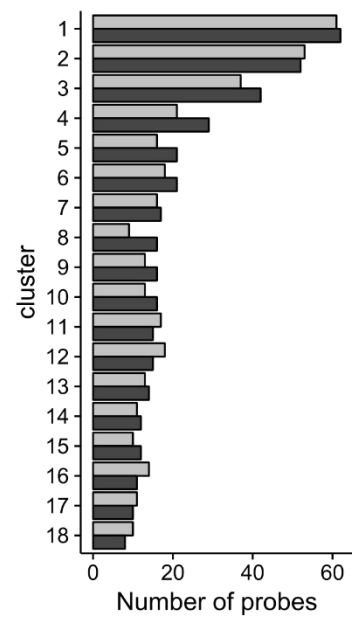
NCBI

HomoloGene



B SCHype clustering

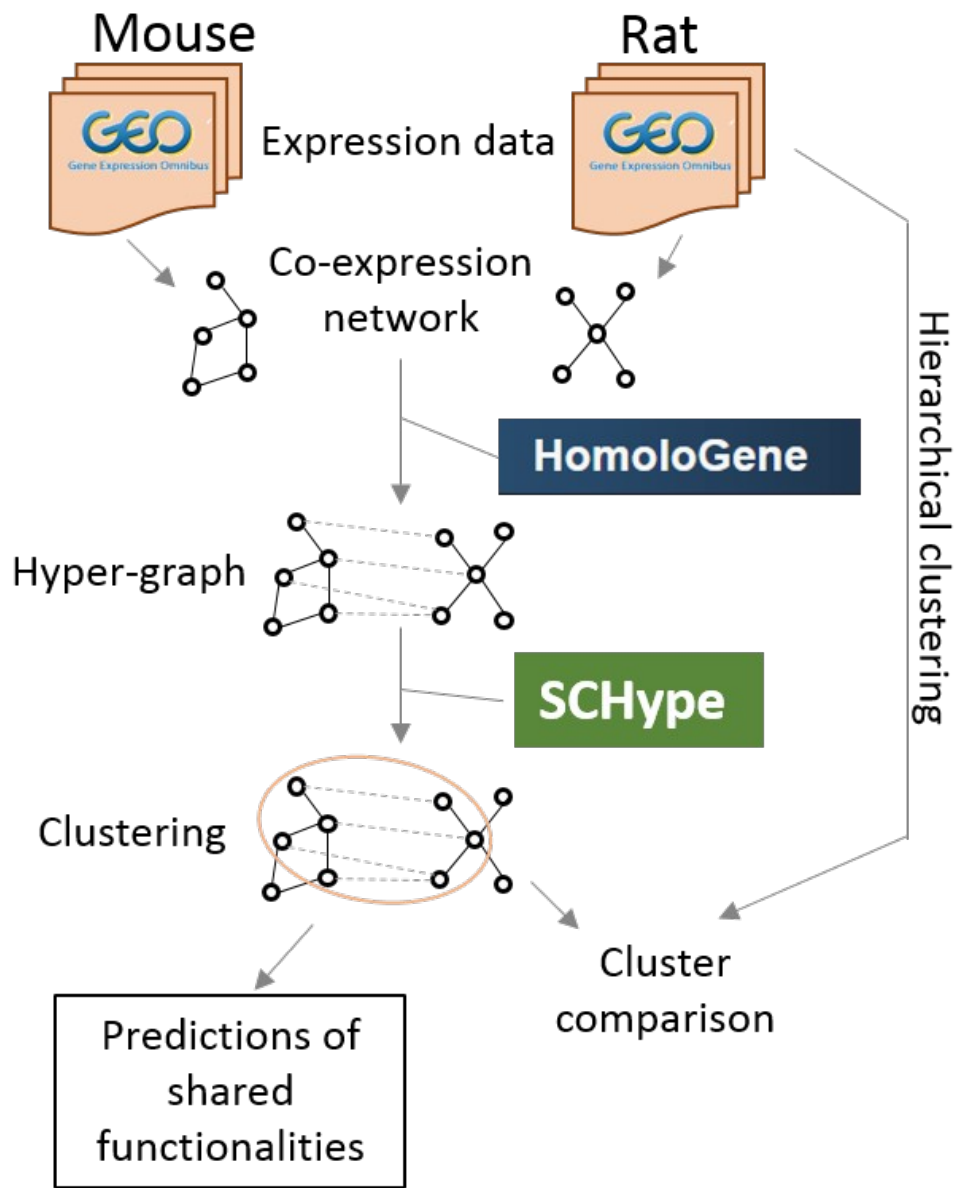
Species ■ Mouse □ Rat



Homology group	Species	Gene name	SCHype cluster	
137229			cluster 69	
	mouse	<i>Anp32a</i>	✓	
	rat	<i>Anp32a</i>	✓	
	rat	<i>LOC100909983</i>		
68982			cluster 7	cluster 30
	mouse	<i>Ccnb1</i>	✓	✓
	mouse	<i>Gm5593</i>		
	rat	<i>Ccnb1</i>	✓	✓
10699			cluster 2	cluster 118
	mouse	<i>Cd248</i>	✓	✓
	rat	<i>Cd248</i>	✓	✓
	rat	<i>LOC100911932</i>		
	rat	<i>LOC100911882</i>		
3938			cluster 1	
	mouse	<i>Ppp1r3c</i>	✓	
	rat	<i>Ppp1r3c</i>	✓	
	rat	<i>LOC100910671</i>		
14108			cluster 2	
	mouse	<i>Rasl10b</i>	✓	
	rat	<i>Rasl10b</i>	✓	
	rat	<i>LOC100912246</i>		

Homology group	Species	Gene name	SCHype cluster		
128630			cluster 9	cluster 12	cluster 45
	mouse	<i>Ceacam1</i>	✓		
	mouse	<i>Ceacam2</i>	✓	✓	✓
	rat	<i>Ceacam1</i>	✓	✓	✓
11456			cluster 5		
	mouse	<i>Elovl6</i>	✓		
	rat	<i>Elovl6</i>	✓		
	rat	<i>LOC102549542</i>	✓		
20277			cluster 35		
	mouse	<i>Rrm2</i>	✓		
	rat	<i>Rrm2</i>	✓		
	rat	<i>LOC100359539</i>	✓		
55991			cluster 1	cluster 119	
	mouse	<i>Tmed2</i>	✓	✓	
	mouse	<i>Gm21540</i>	✓	✓	
	rat	<i>Tmed2</i>	✓	✓	
11890			cluster 10	cluster 43	cluster 81
	mouse	<i>Tnks2</i>	✓	✓	✓
	rat	<i>LOC100910717</i>	✓	✓	✓
	rat	<i>Tnks2</i>	✓	✓	✓

Homology group	Species	Gene name	SCHype cluster
117948			cluster 102
	mouse	<i>Cyp2c38</i>	✓
	mouse	<i>Cyp2c29</i>	
	mouse	<i>Cyp2c39</i>	
	rat	<i>Cyp2c7</i>	✓
104115			cluster 33
	mouse	<i>Hsd3b5</i>	✓
	mouse	<i>Gm10681</i>	
	mouse	<i>Hsd3b4</i>	
	mouse	<i>Gm4450</i>	
	rat	<i>Hsd3b5</i>	✓
	rat	<i>LOC100911116</i>	✓
137425			cluster 2
	mouse	<i>Lce3c</i>	✓
	rat	<i>LOC100361951</i>	✓
	rat	<i>LOC100911982</i>	✓
	rat	<i>Lce3d</i>	
129514			cluster 17
	mouse	<i>Rdh9</i>	✓
	mouse	<i>Rdh1</i>	
	mouse	<i>Rdh16</i>	
	mouse	<i>Rdh19</i>	
	mouse	<i>BC089597</i>	
	rat	<i>Rdh16</i>	✓
	rat	<i>LOC100365958</i>	✓



Conclusions & perspectives

- ❖ Provide evidence for resolving functional annotation transfer in 28 complex homology groups (>100 genes)
- ❖ Using free data!
- ❖ Different array, RNA-seq
- ❖ More tissues!
- ❖ More species!
- ❖ Still need functional annotation in at least one species

Epigenomics and the human transcript diversity

using Roadmap Epigenomics data





- ❖ Epigenetic marks role at TSS and enhancer is (quite) well characterized
- ❖ Putative roles in TES, exon recognition, cryptic TSS repression?
- ❖ Plenty of data generated by Roadmaps Epigenomics (+ENCODE)
- ❖ RNA-seq, WGBS, DNase1, plenty of HisMod in > 30 cell types



What's a gene?



Version 27 (January 2017 freeze, GRCh38) - Ensembl 90

General stats

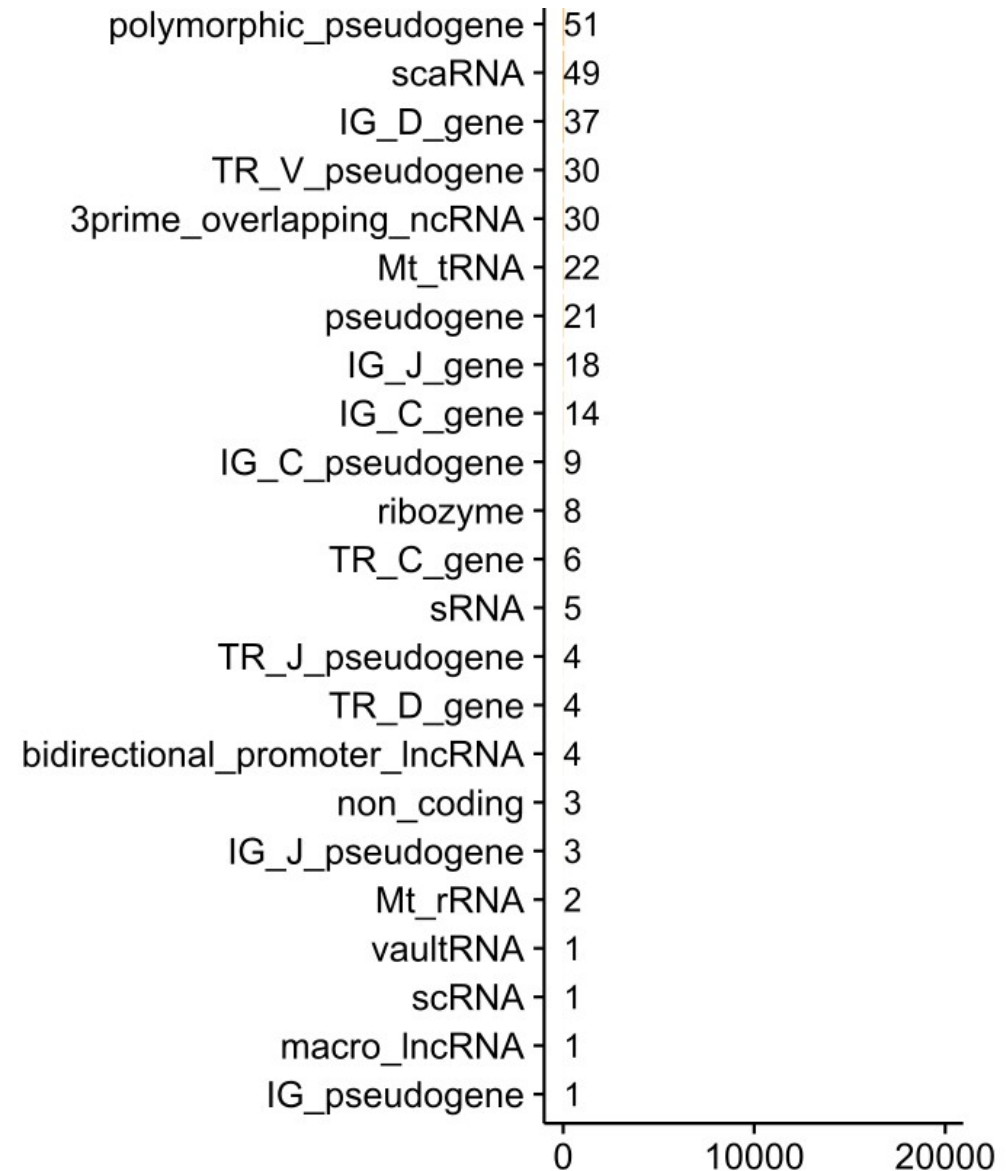
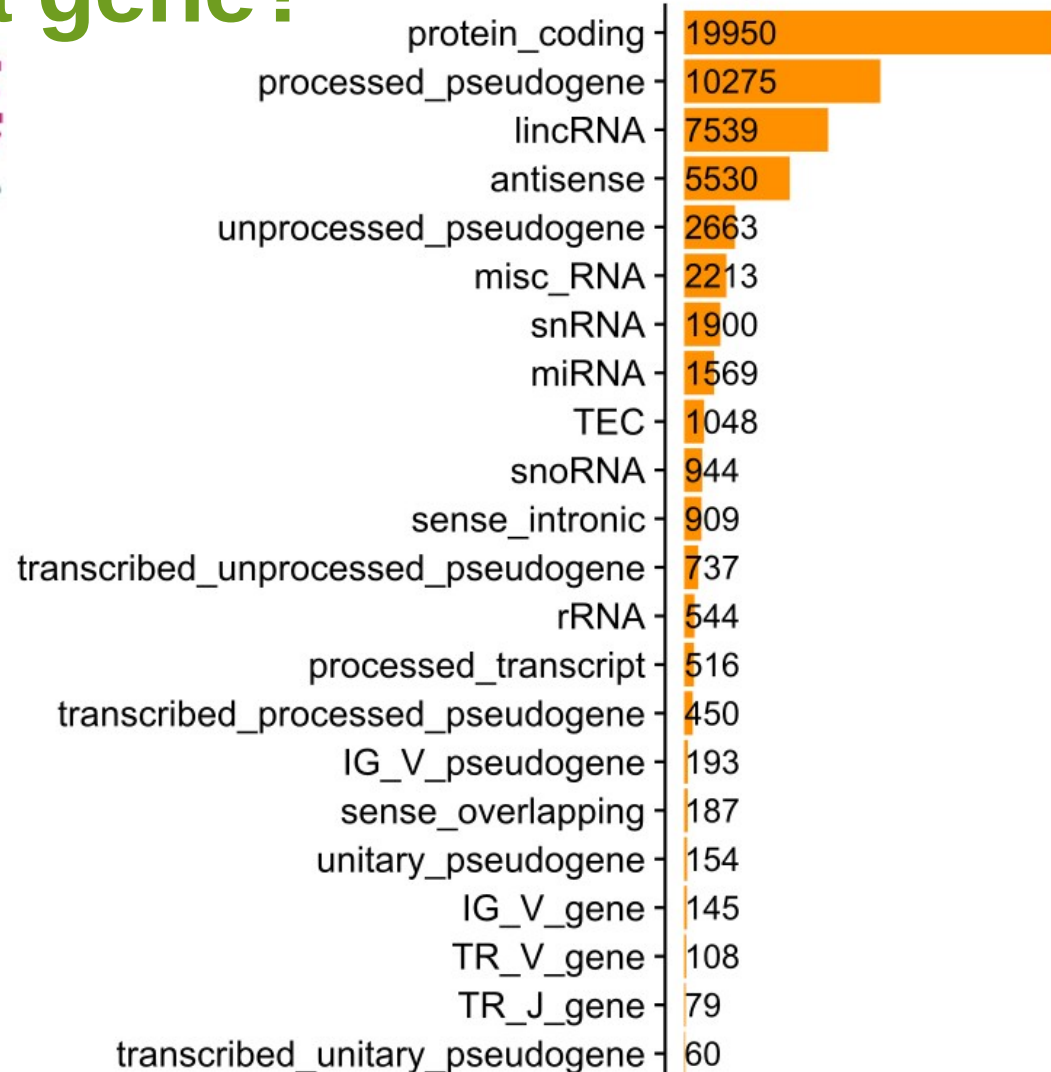
Total No of Genes

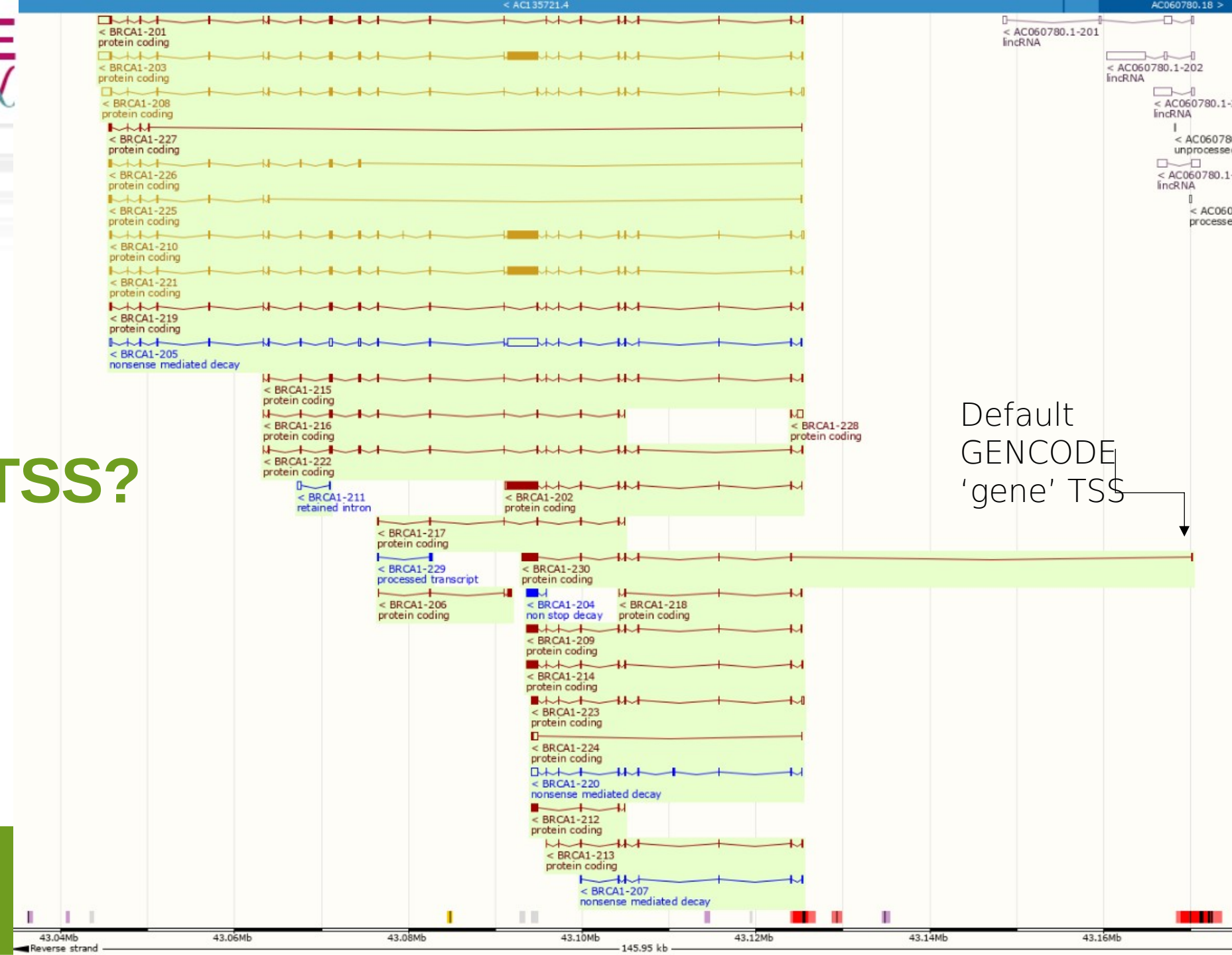
58288

What's a gene?



Le génome humain

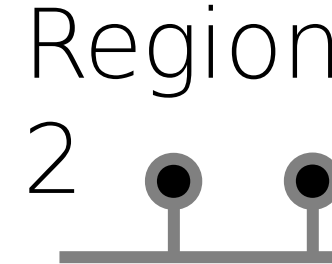
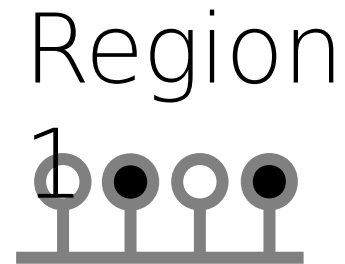




What's a gene TSS?

Default
GENCODE
'gene' TSS

Which metric for DNA methylation?



CpG
density
 $mCpG/CpG$
 $mCpG$
density

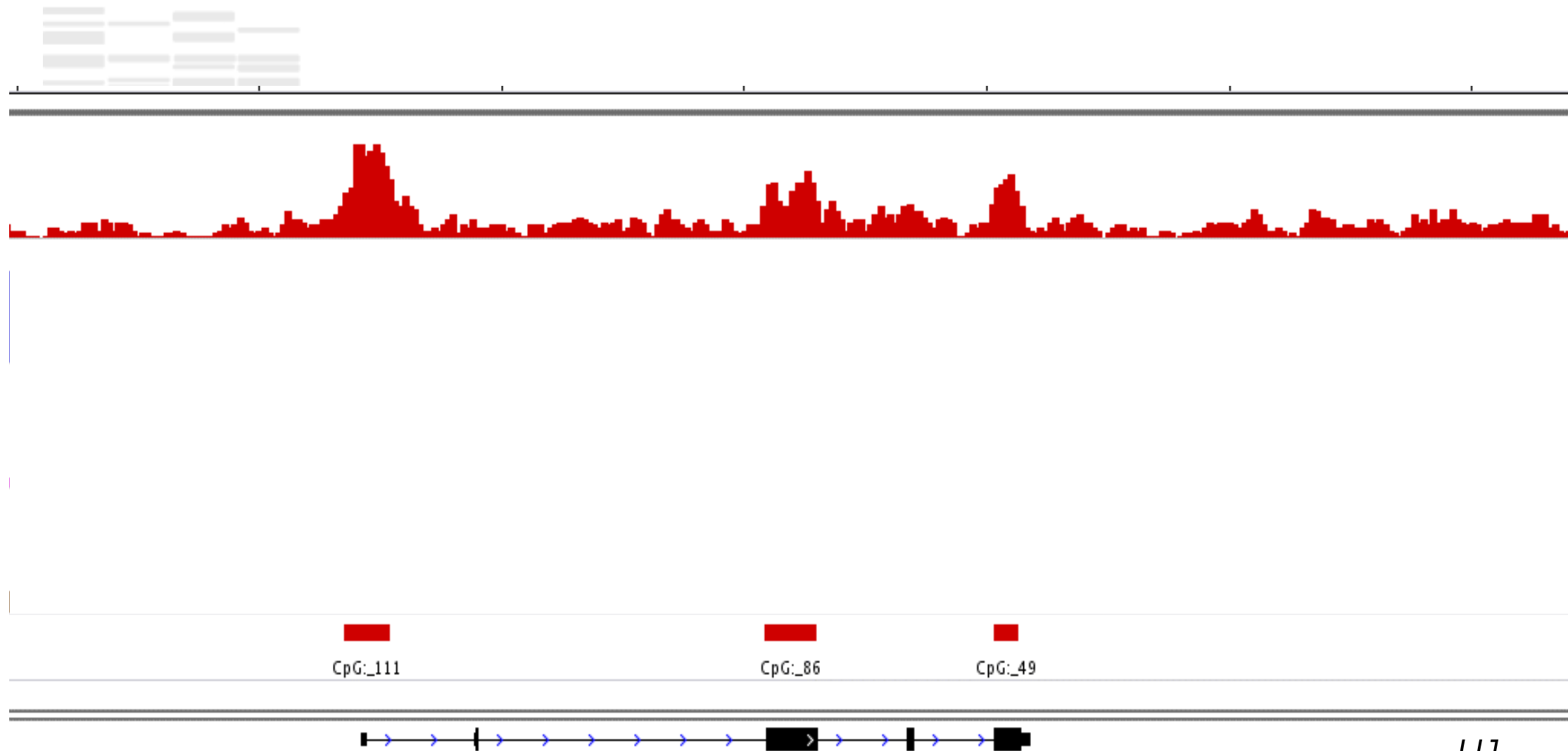
4
0.5
2

2
1
2

mCpG density main driver of methylation

reader?

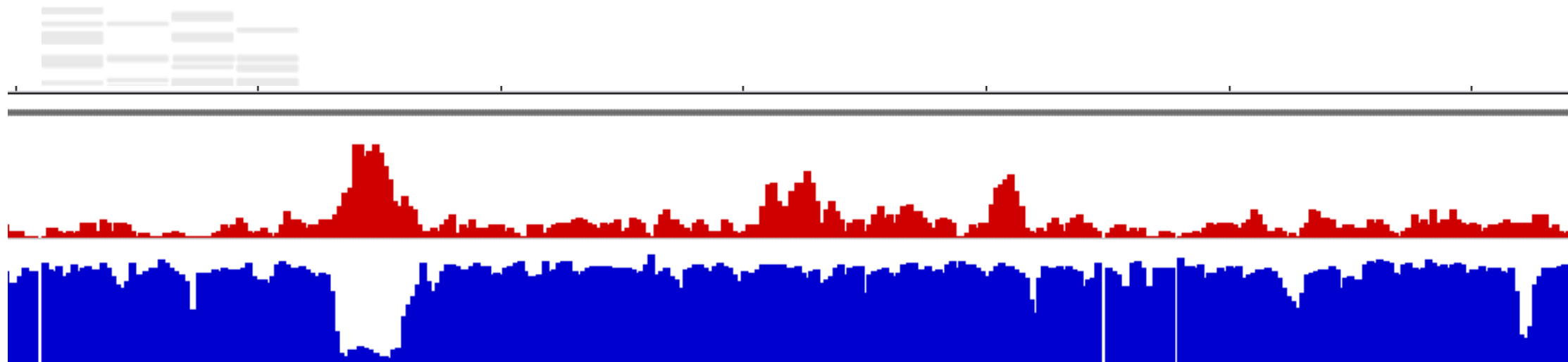
CpG
density



H1-
hESC

CpG
density

mCpG
ratio

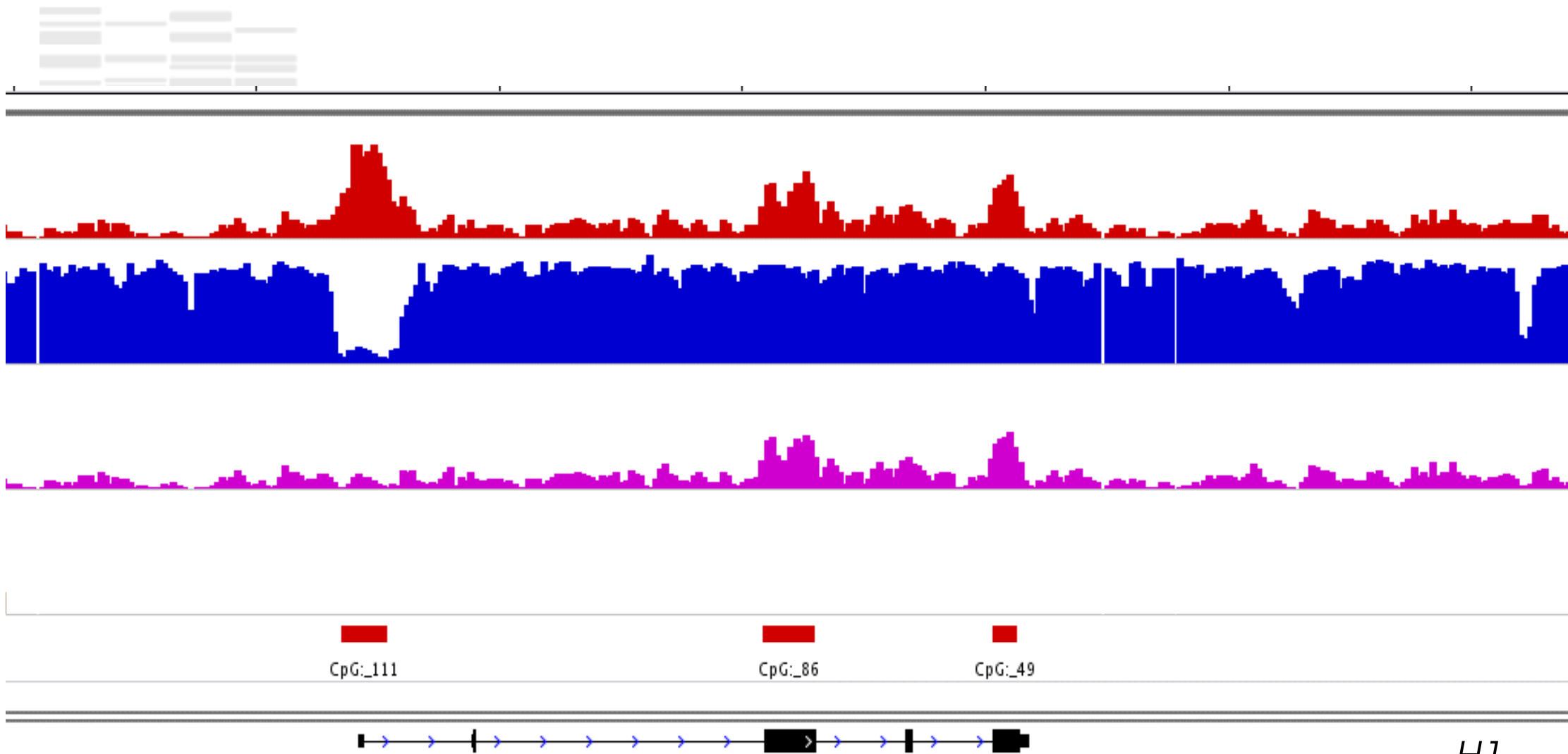


CpG:_111

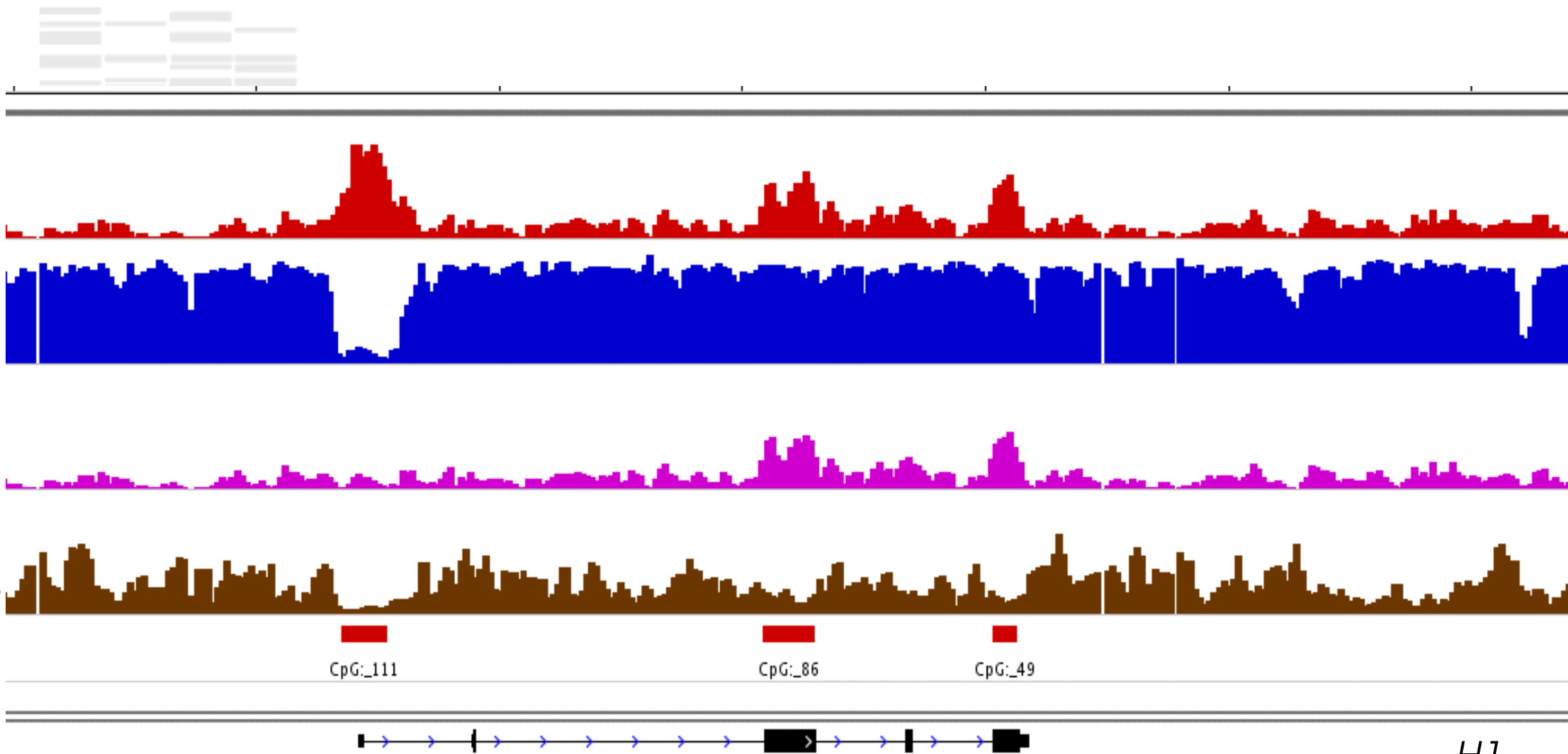
CpG:_86

CpG:_49

*H1-
hESC*

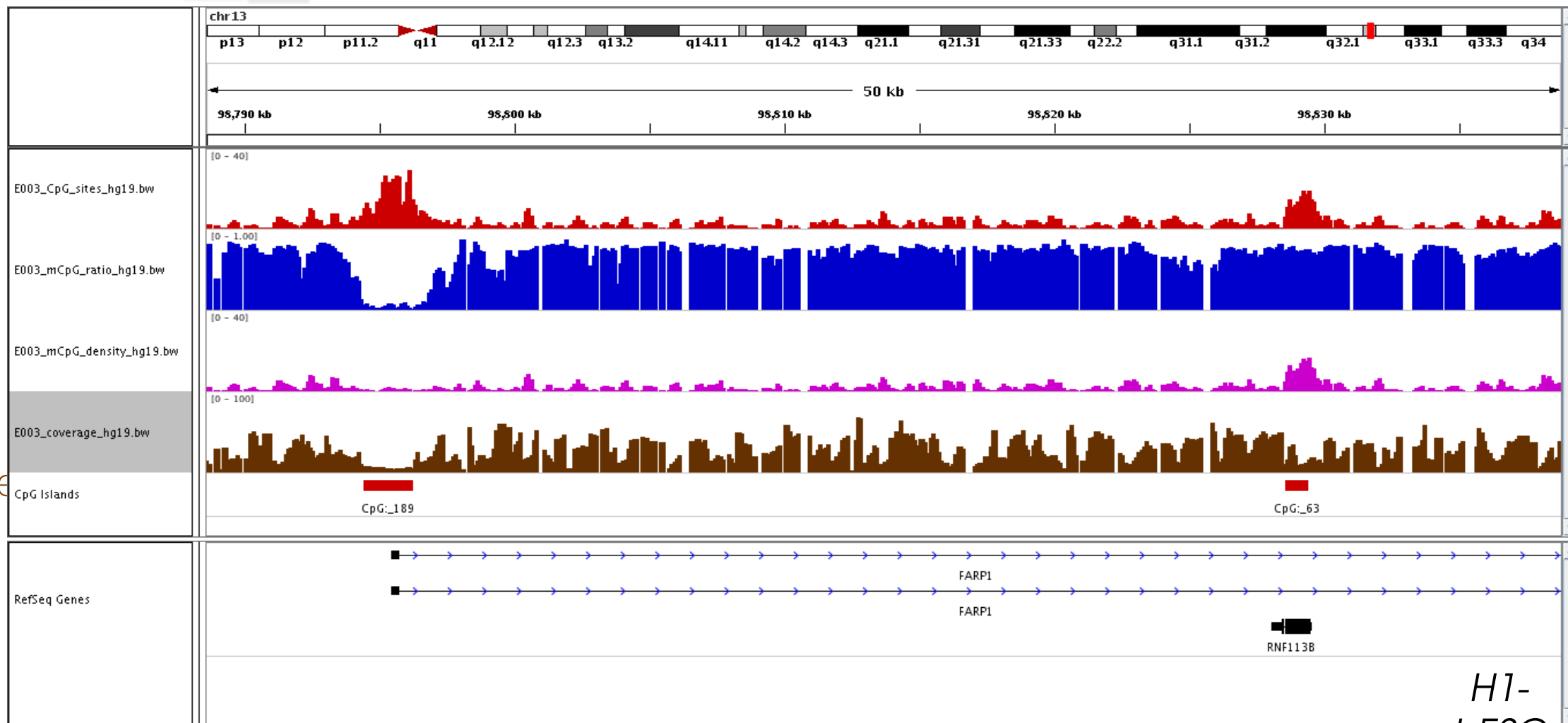


H1-
hESC



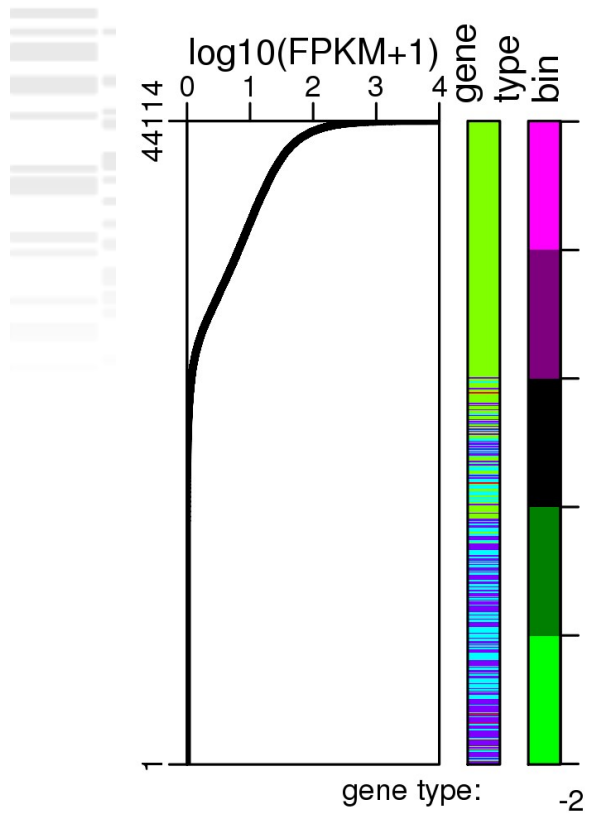
H1-
hESC

CpG
density
mCpG
ratio
mCpG
density
coverage

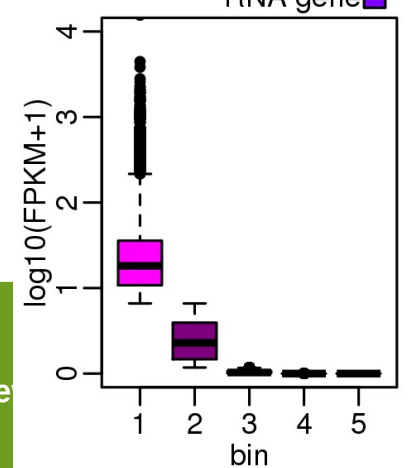


H1-
hESC

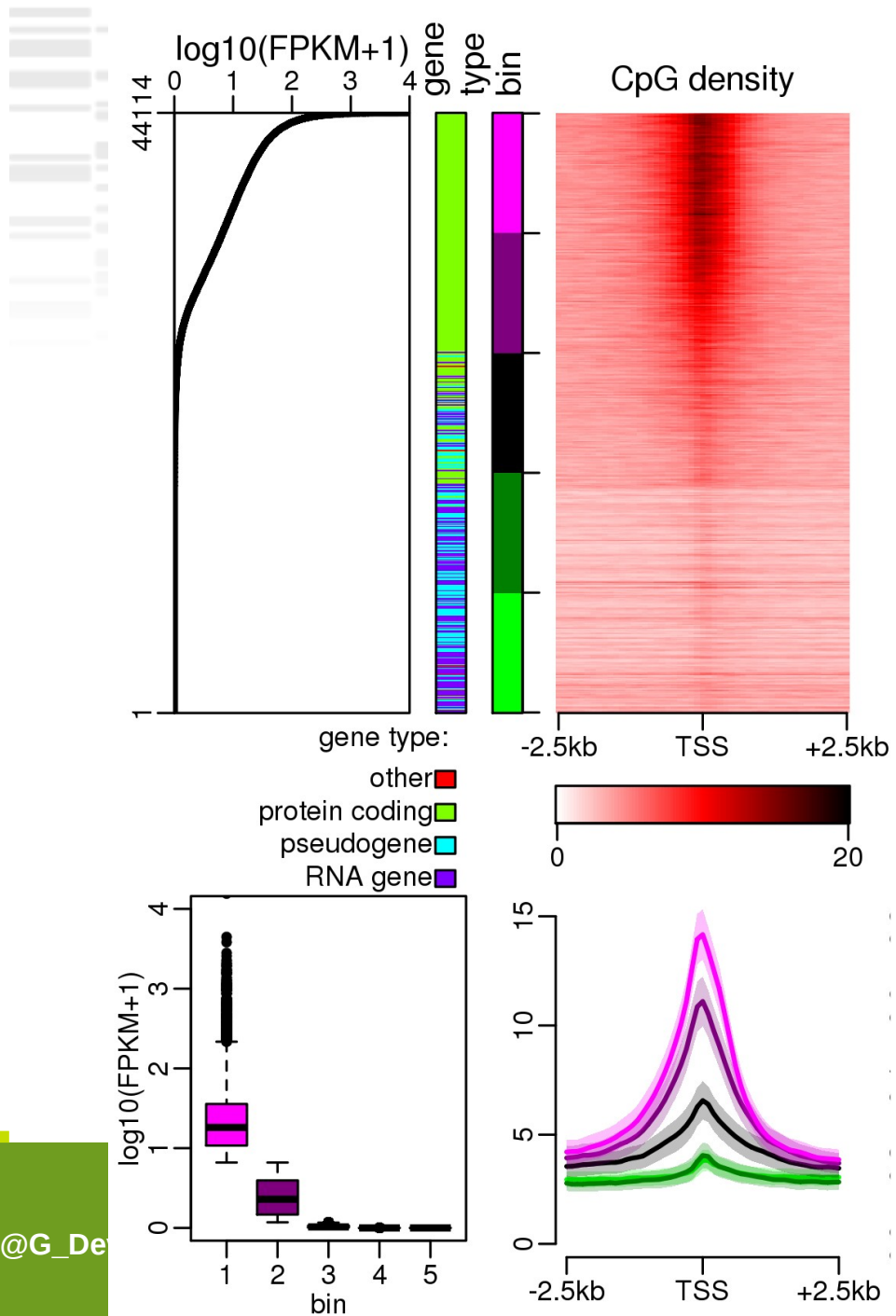
H1-hESC



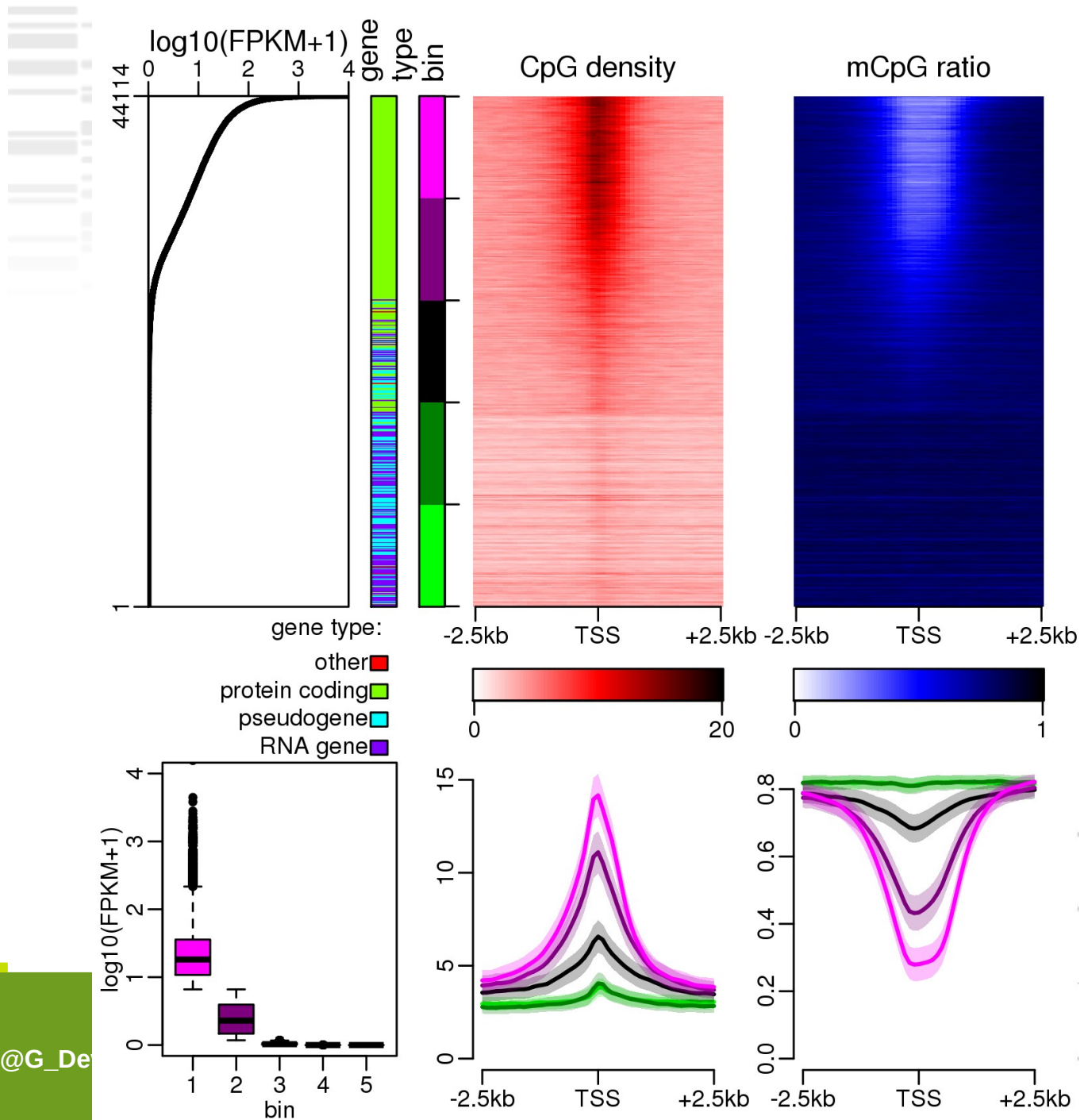
gene type:
other
protein coding
pseudogene
RNA gene



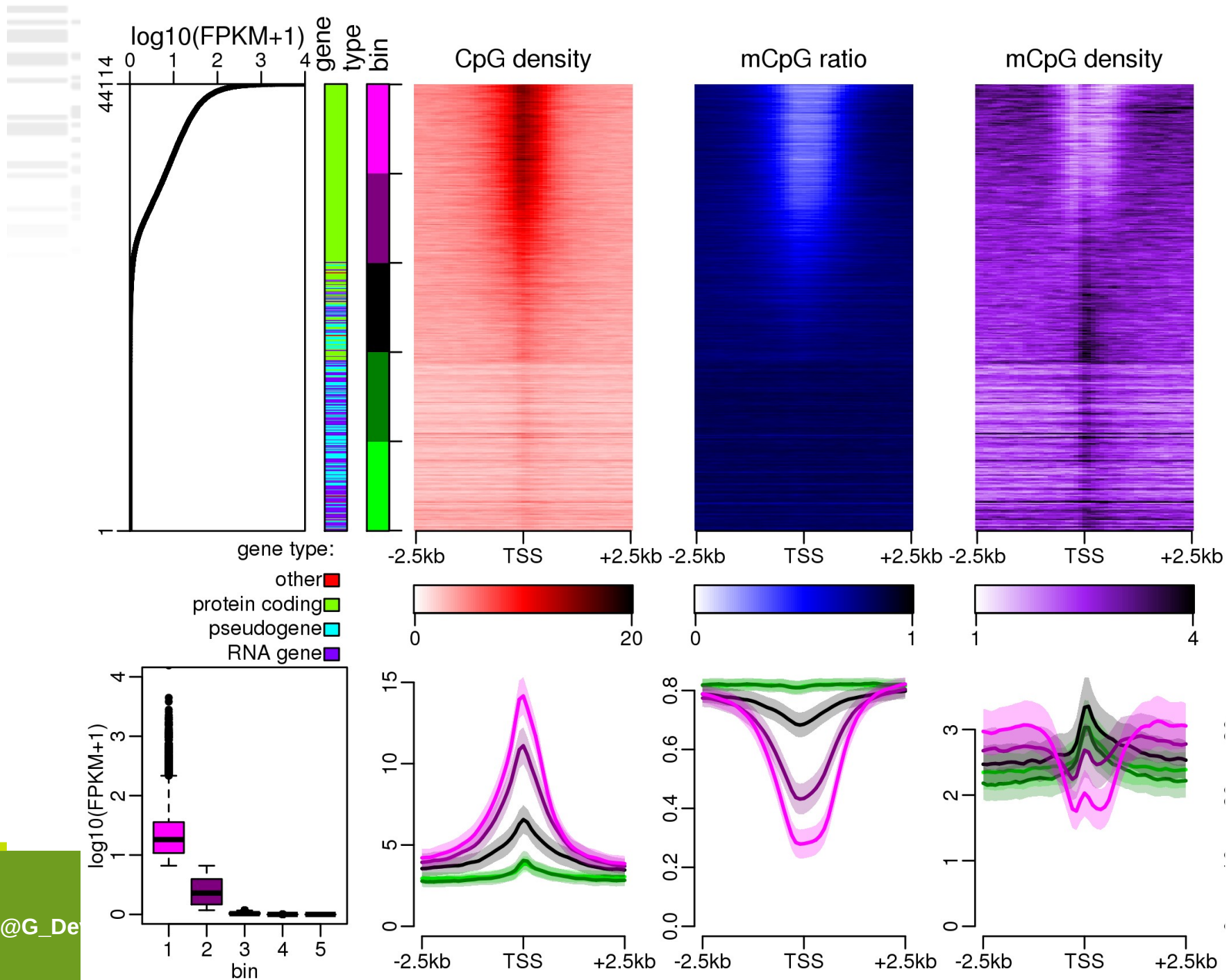
H1-hESC



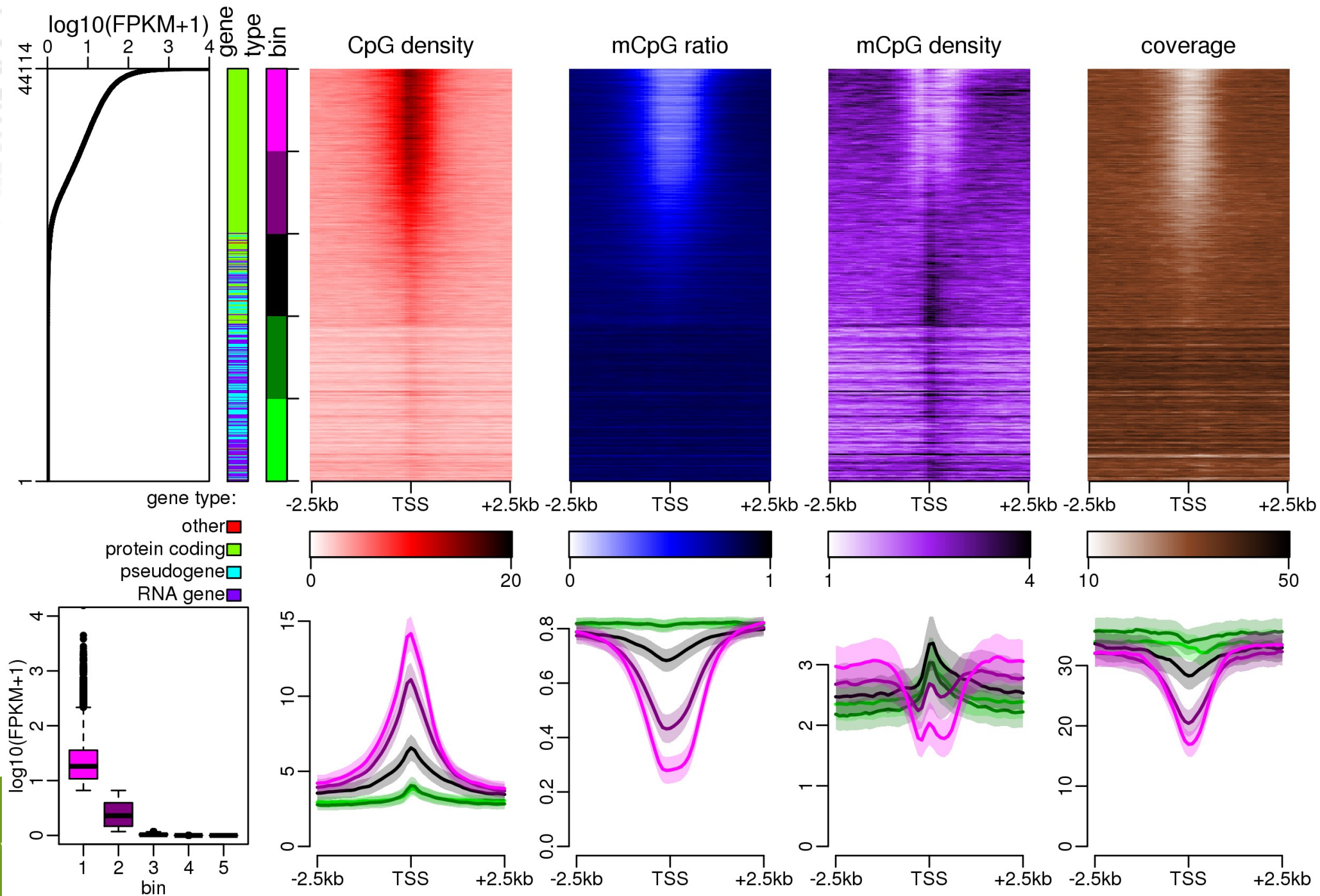
H1-hESC



H1-hESC



H1-hESC





- ❖ Features $\times$ marks $\times$ cell types $\times$ gene types
- ❖ Combinatorial explosion: >5000

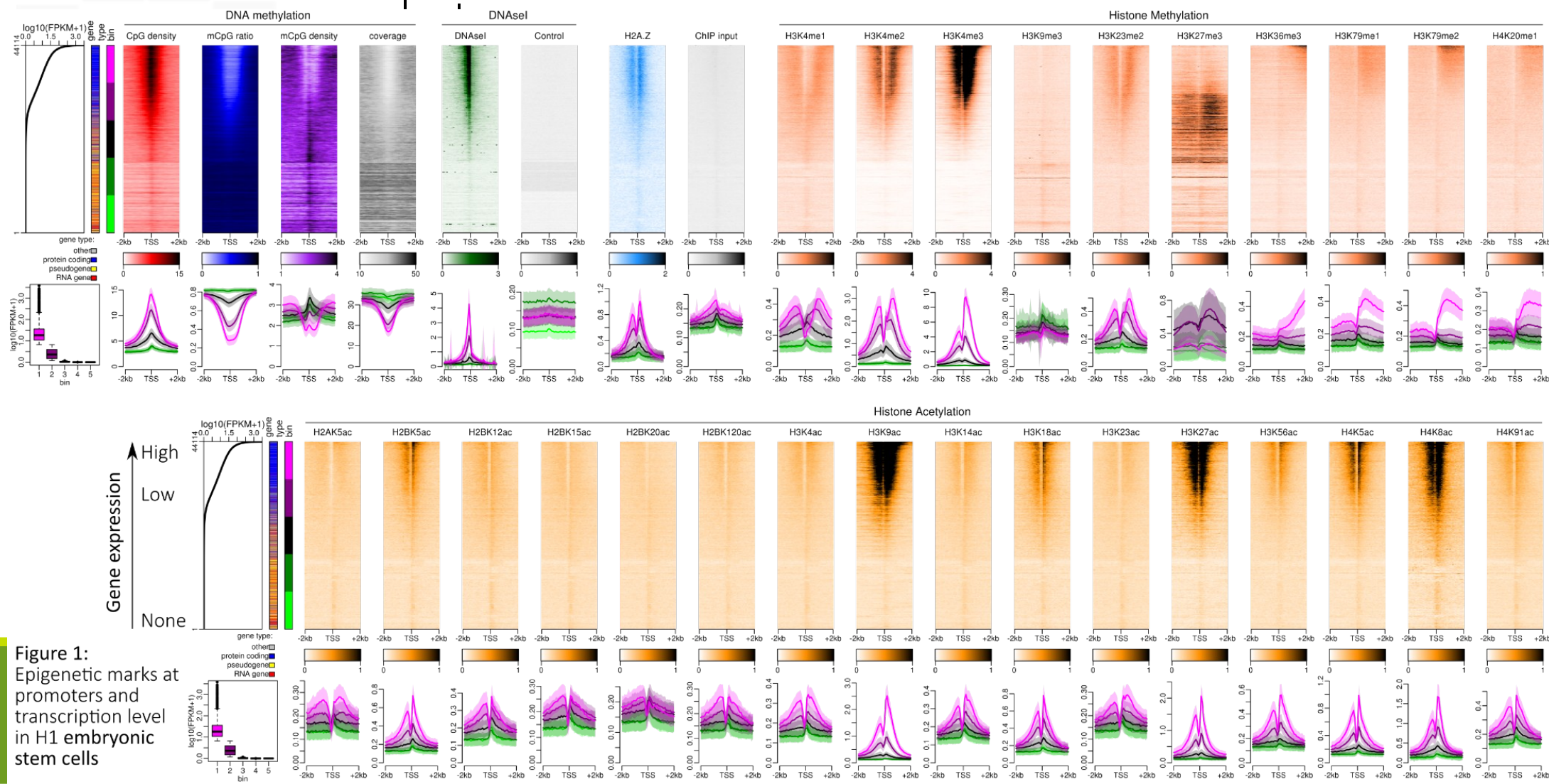
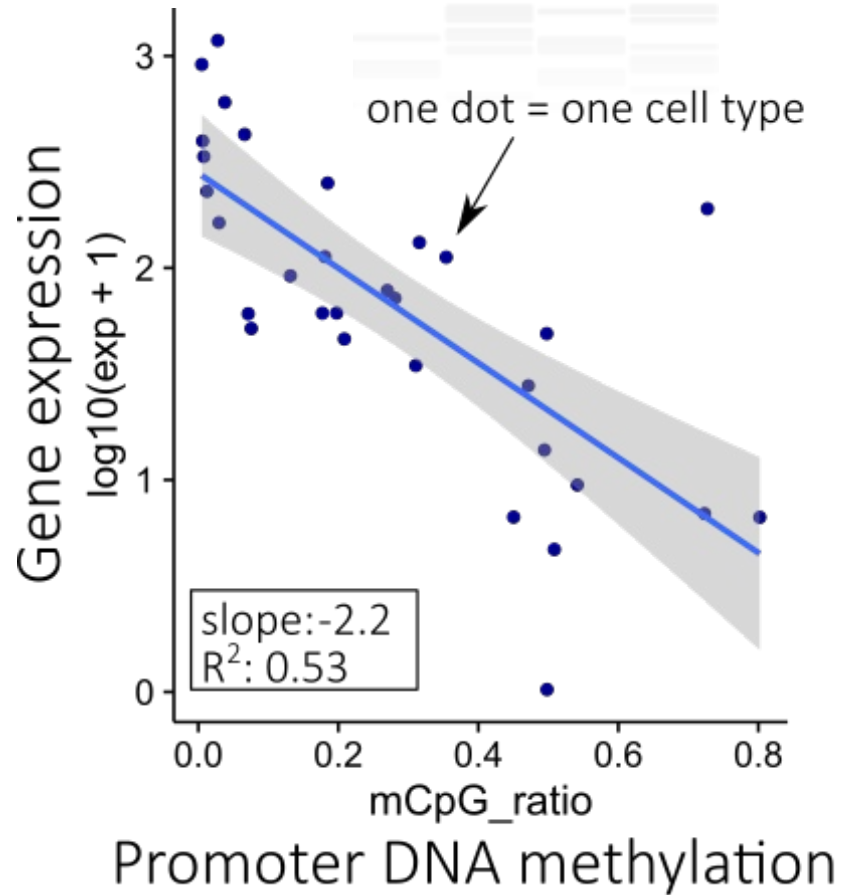
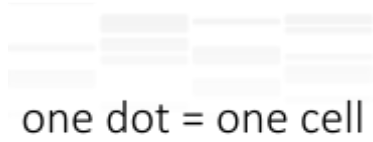
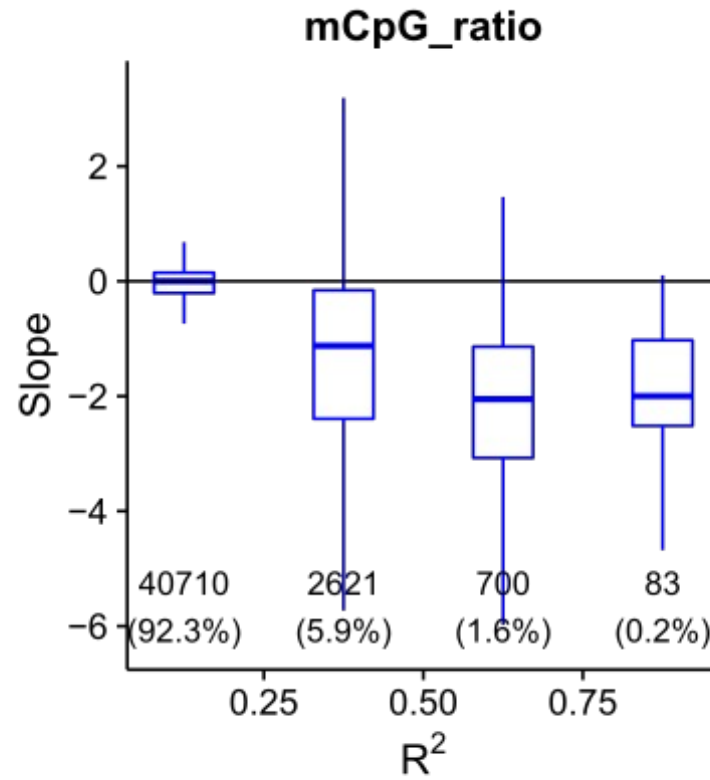
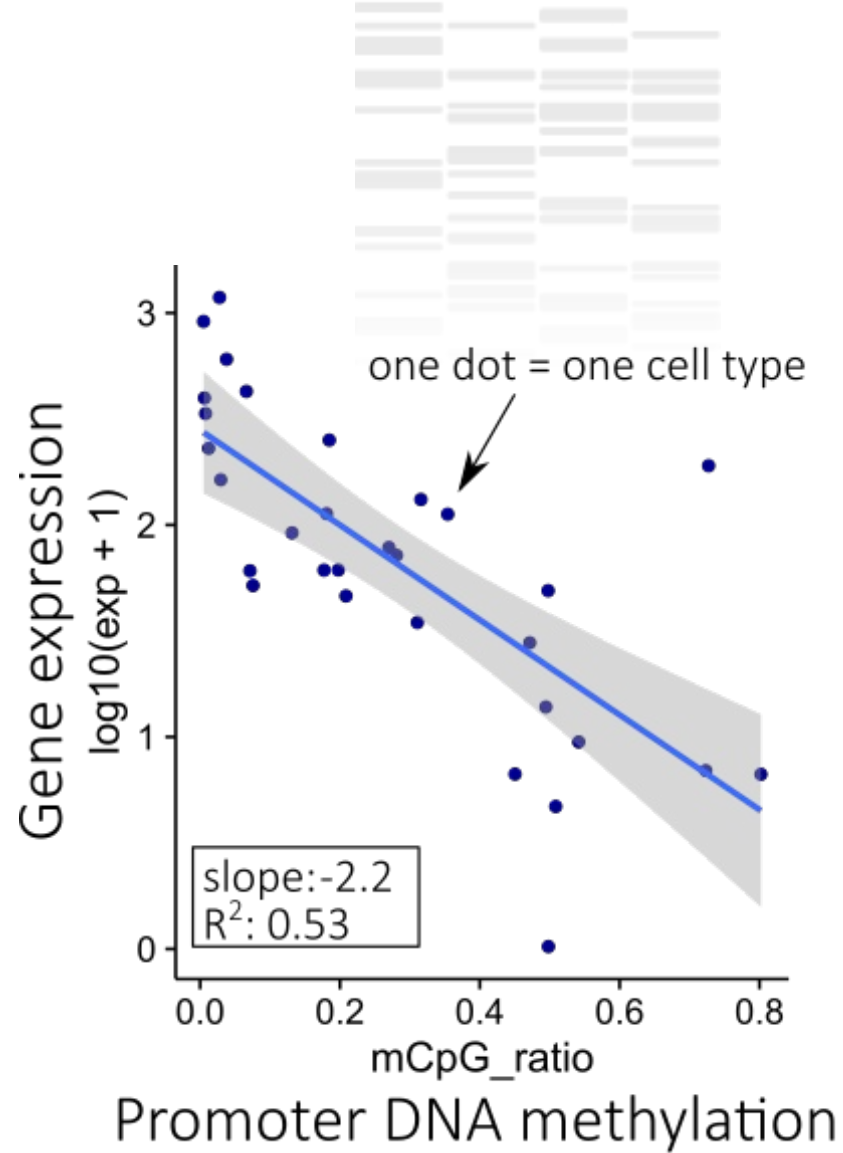


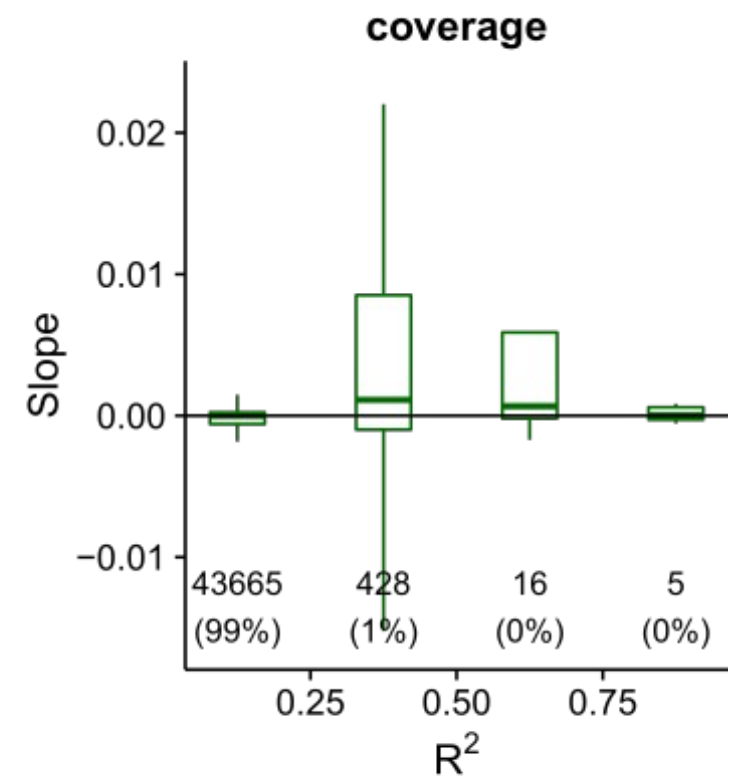
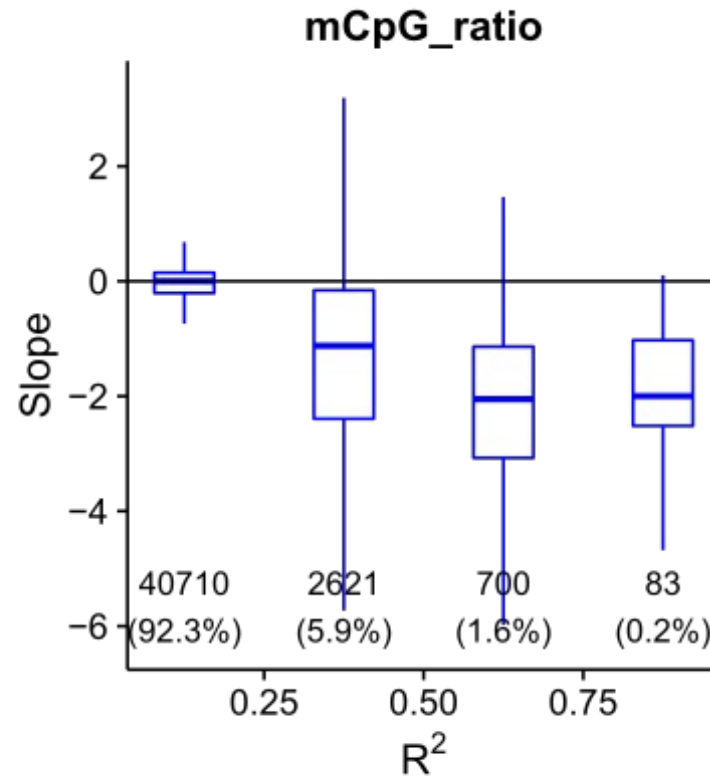
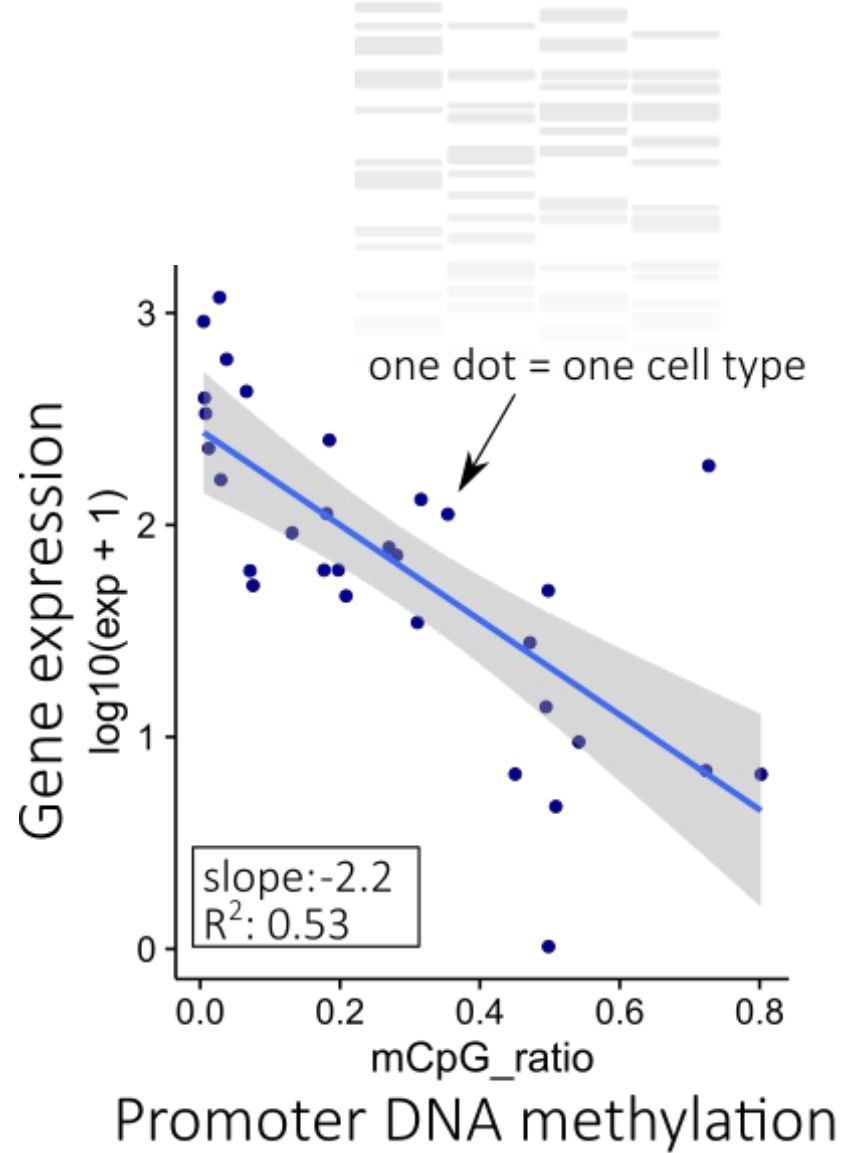
Figure 1:
Epigenetic marks at
promoters and
transcription level
in H1 embryonic
stem cells



GJA1 (ENSG00000152661)

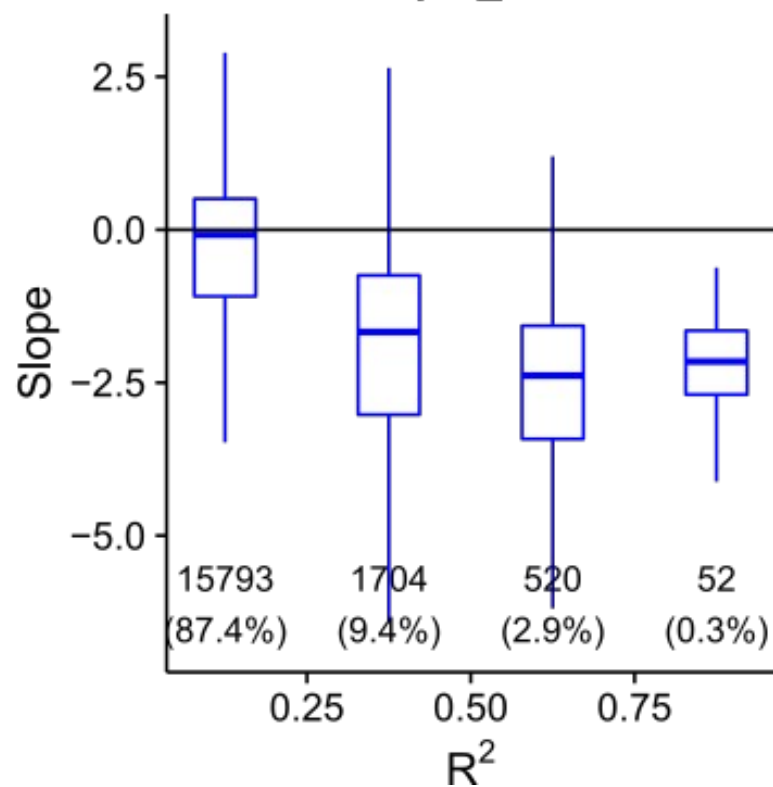




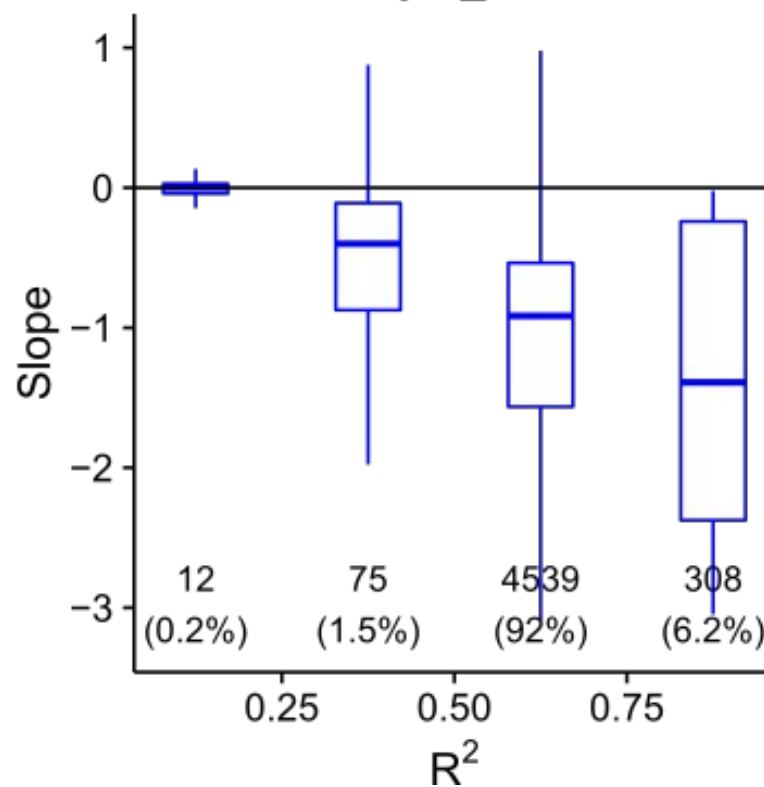




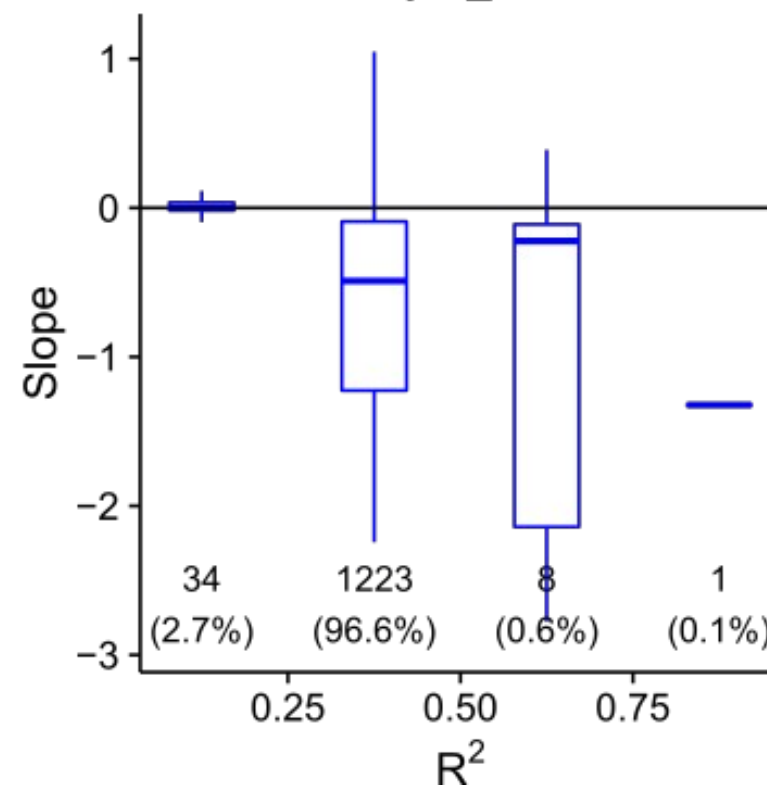
protein coding genes
mCpG_ratio

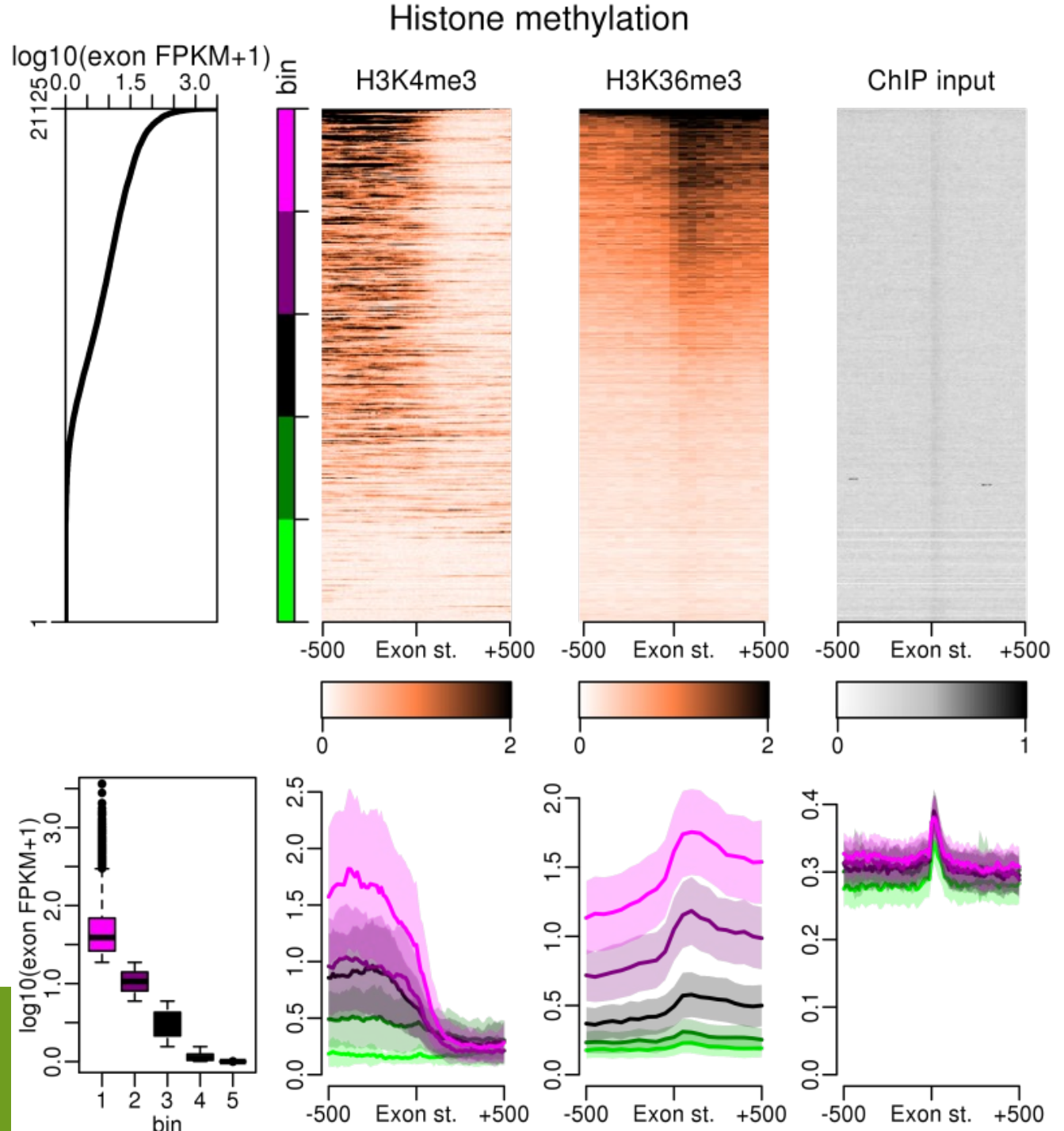


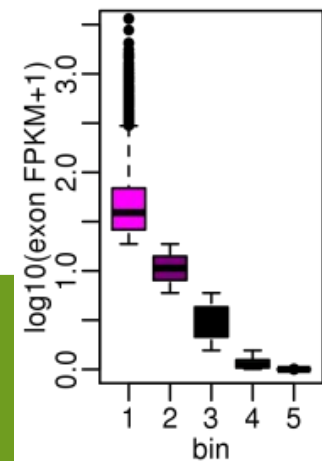
lincRNA genes
mCpG_ratio



miRNA genes
mCpG_ratio









Perspectives

- ❖ Epigenetic marks and **alternative** splicing?
- ❖ Gene by gene tool in webapp
- ❖ Polish, write and submit

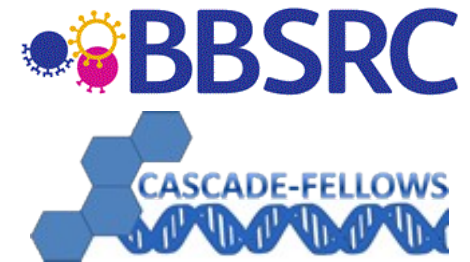
Thanks



Anagha Joshi Arina Mantsok Barry Horne Tom Michoel



Angeles Arzalluz-Luque Deepti Vipin Pia Francesca Loren Reyes





Epigenetics and heritability

- ❖ missing heritability
- ❖ Improving livestock epigenome

Non-heritable epigenetics

- ❖ Early prediction?



- ❖ Variant annotation

(physiology and nutrition)