

Impact of cigarette smoke and Modified Risk Tobacco Products (MRTPs) on DNA methylation

Amin Choukrallah, Scientist Genomics, Philip Morris International R&D

Thursday, May 4th
sbv IMPROVER Symposium, Tel Aviv, Israel

Background :

- Environmental impact on the epigenome
- DNA methylation readout
- DNA methylation and cigarette smoke exposure

Results :

Effect of cigarette smoke and MRTPs on DNA methylation in murine lung tissue

- Experimental design
- Results and analytical methods
- Conclusions and perspectives

Environmental factors



Diet



UV radiations



Smoking



Pollution

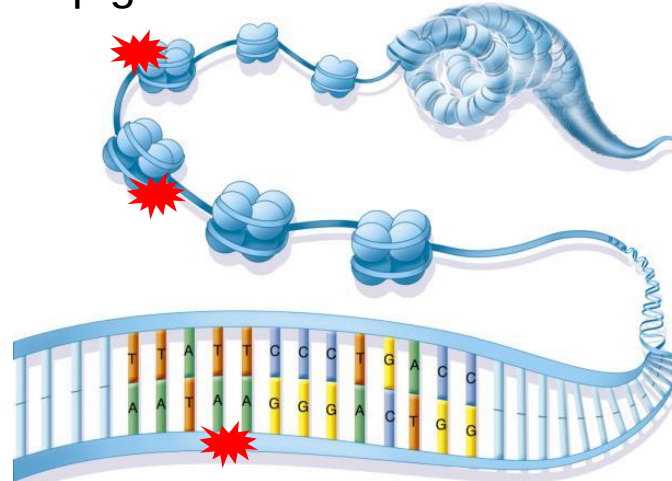


Virus

Toxic chemical



Epigenetic features



Genomic sequence



Phenotype



Gene expression
(RNA)

- Heritable changes via epigenetic mechanisms
- Epigenetic changes may occur before changes in RNA levels (Priming)
- Epigenetic changes may persist after exposure (Epigenetic memory)

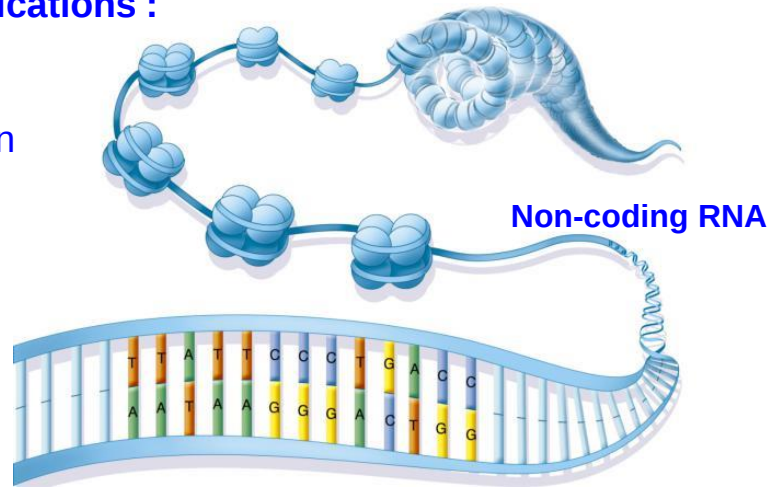
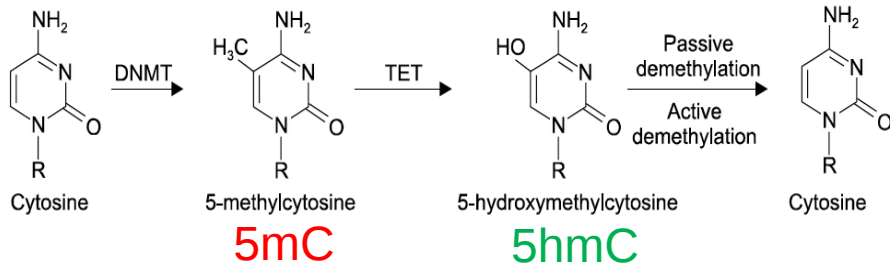
Epigenetic changes as long-term markers of exposure

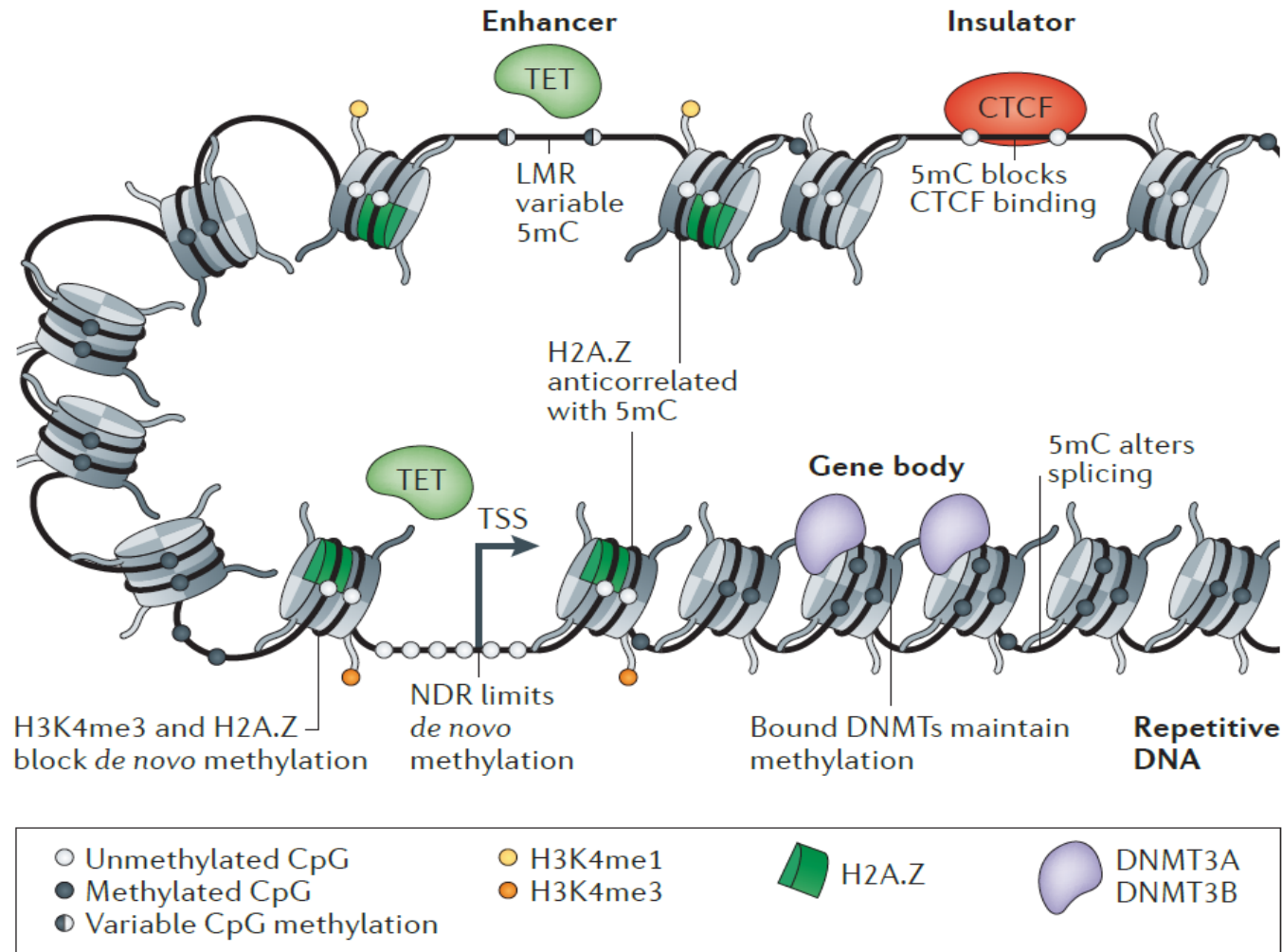
Epigenome : Sequence-independent modifications of DNA, DNA associated proteins and non-coding RNAs

Histone modifications :

Acetylation
Methylation
Phosphorylation
....

DNA methylation

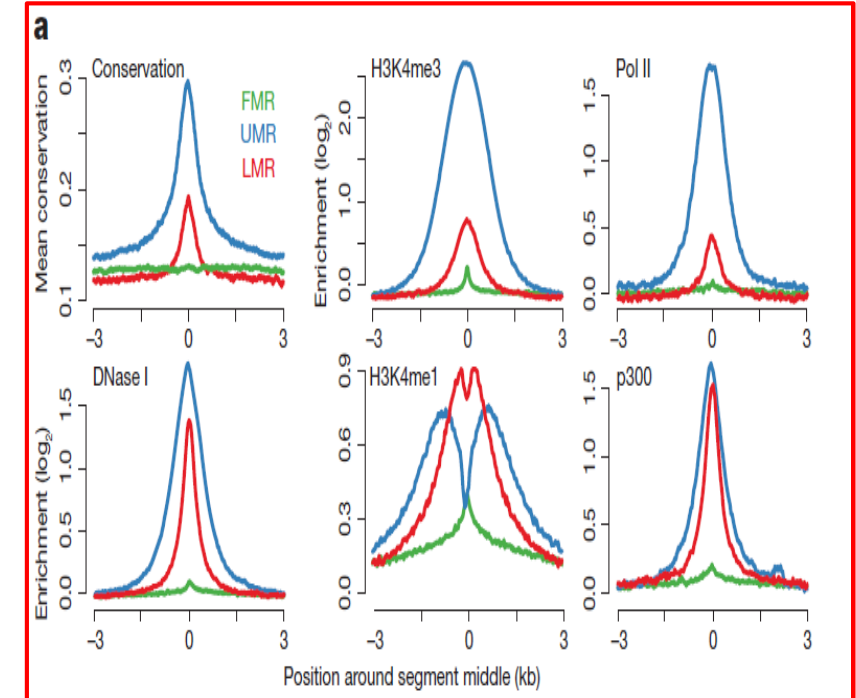
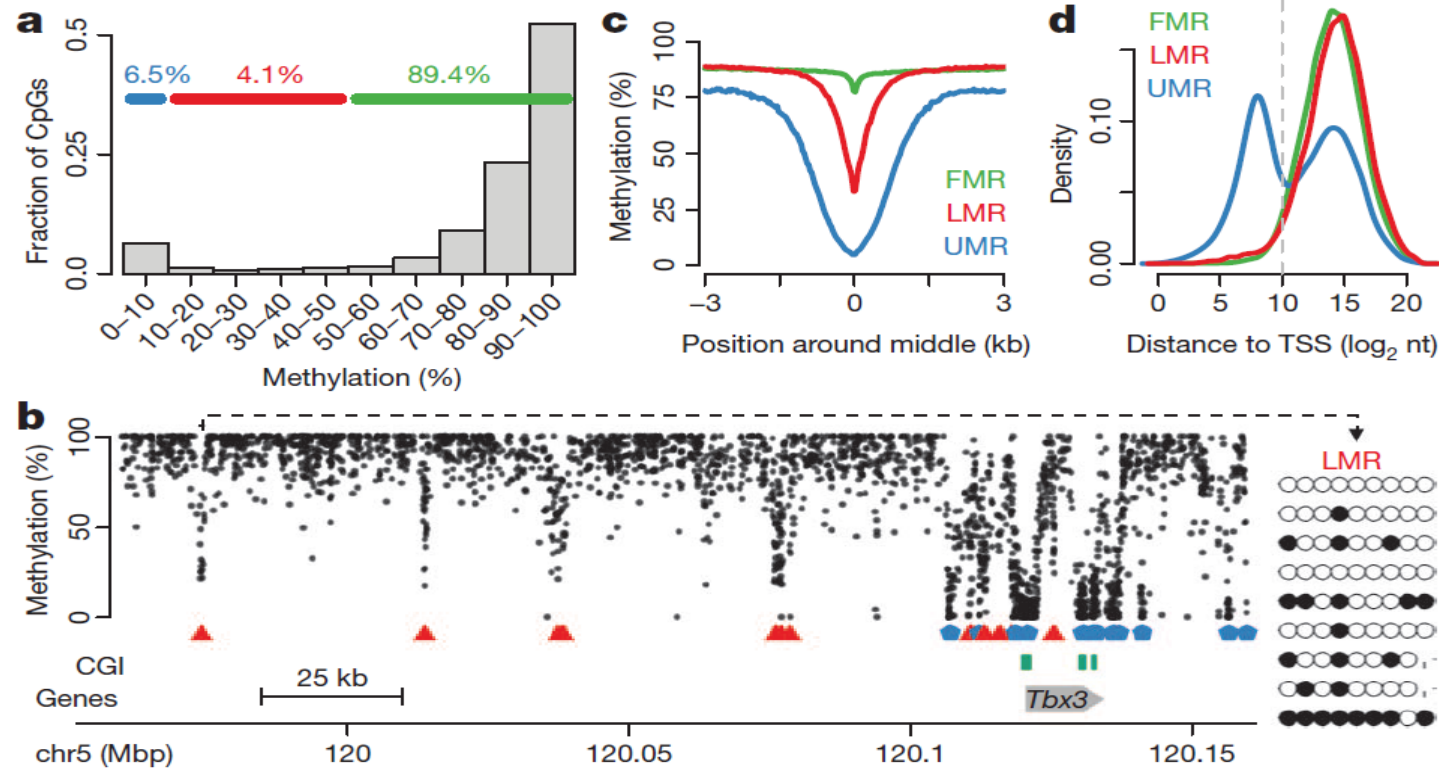




DNA Methylation != Repression

Peter A. Jones 2012

WGBS : Whole-Genome Bisulfite Sequencing



FMR : Fully Methylated Regions

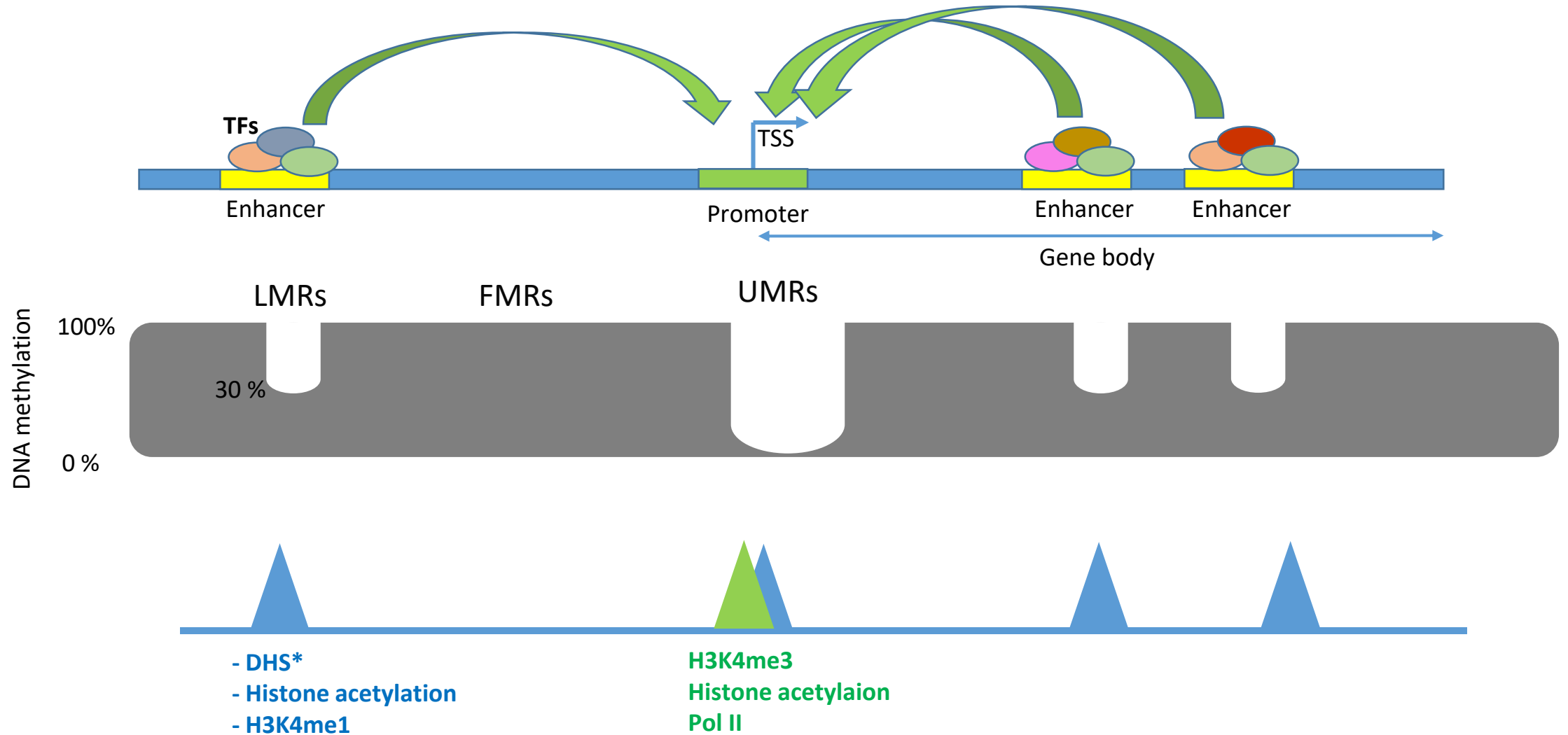
UMRs: UnMethylated Regions

LMRs: Low Methylated Regions

-Active promoters
-CpG islands

LMRs = Enhancers

DNA Methylation provides information about the genomic context



Breitling et al, 2011

Human peripheral **blood** cells

177 samples

27K BeadChips Illumina

Hypomethylation of **1** single CpG

Shenker et al, 2012

Blood

374 samples

450K BeadsChIP Illumina

Hypomethylation **9 CpGs**

Joubert et al; 2012

Cord **blood** of newborns

1062 samples

450K BeadsChIP Illumina

26 CpGs mapped to 10 genes

Joeheanes et al, 2016

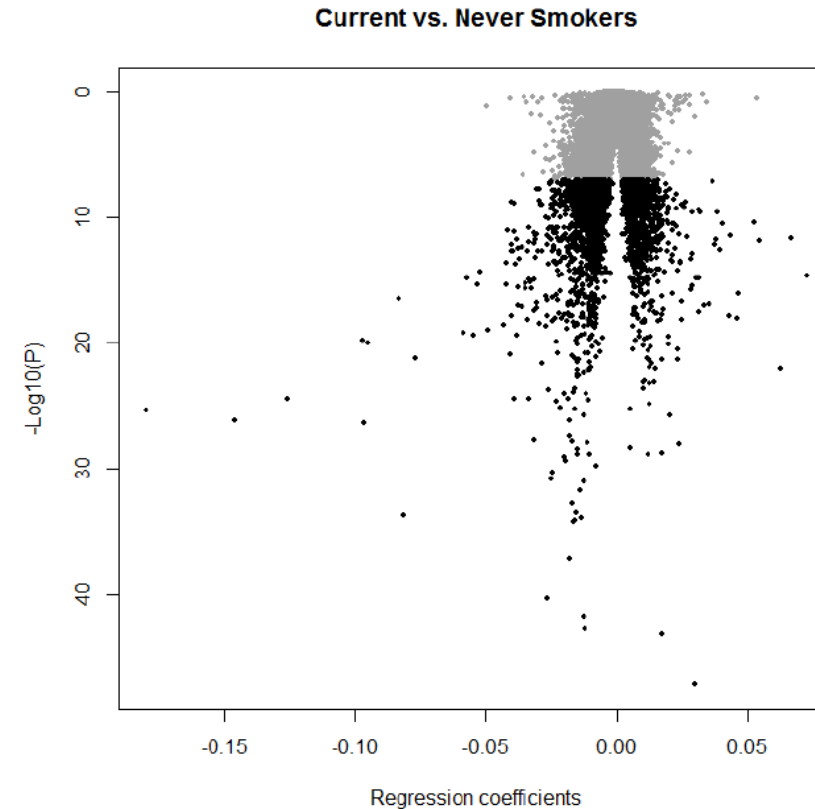
Blood / CD4 T cells

15 907 samples

450K BeadsChIP Illumina

2623 CpG / 1405 genes

Supplemental Figure 3. Volcano plot for CpG site association with respect to current versus never smoker



HumanMethylation450 array content.

Feature type	Included on array
Total number of sites	485,577
RefSeq genes	21,231 (99%)
CpG islands	26,658 (96%)
CpG island shores (0–2 kb from CGI)	26,249 (92%)
CpG island shelves (2–4 kb from CGI)	24,018 (86%)
HMM islands ^a	62,600
FANTOM 4 promoters (High CpG content) ^a	9426
FANTOM 4 promoters (Low CpG content) ^a	2328
Differentially methylated regions (DMRs) ^a	16,232
Informatically-predicted enhancers ^a	80,538
DNase hypersensitive sites	59,916
Ensemble regulatory features ^a	47,257
Loci in MHC region	12,334
HumanMethylation27 loci	25,978
Non-CpG loci	3091

Bibikova et al; 2011

- The arrays contain mainly annotated loci
- Limited coverage of distal regulatory elements (Enhancers)
- Context dependent read out of DNA methylation

Mouse models :

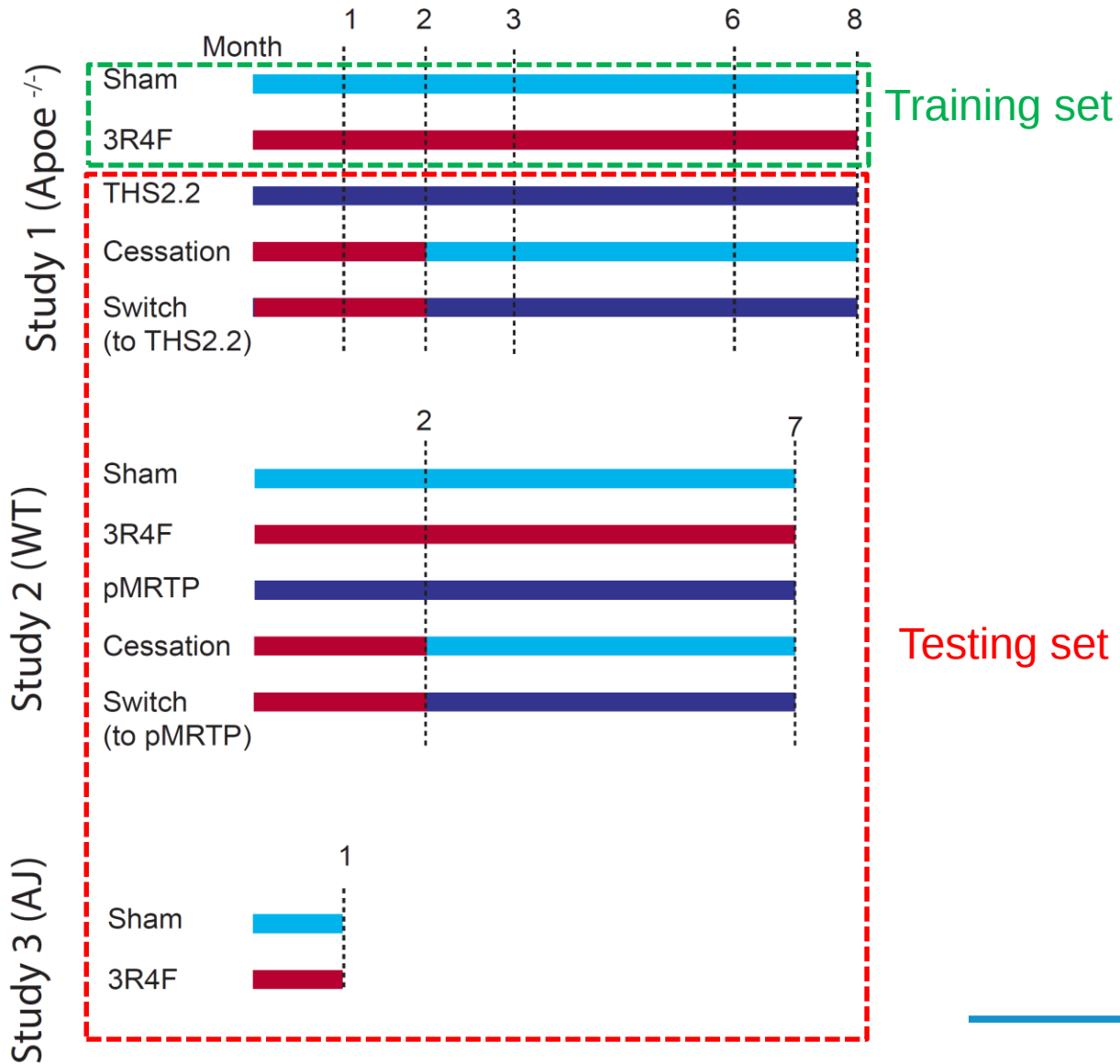
- WT BL6 mice
- *Apoe* -/- (Atherosclerosis-prone apolipoprotein E-deficient mice)
- A/J strain

Tissues :

- Lung

Sequencing technique:

- Whole Genome Bisulfite Sequencing (WGBS)



- **Sham** : Fresh air (control)
- **3R4F** : Conventional cigarette smoke extract
- **THS2.2** : Aerosol from Tobacco Heating System 2.2
- **pMRTP** : Aerosol from prototype MRTP

THS2.2 and pMRTP : Heat-no-burn products

1-Can a smoke exposure signature be extracted from DNA methylation levels of cis-regulatory elements (CREs)?

2-Can a smoke exposure signature be extracted from expression data of genes controlled by differentially methylated DNA cis-regulatory elements?

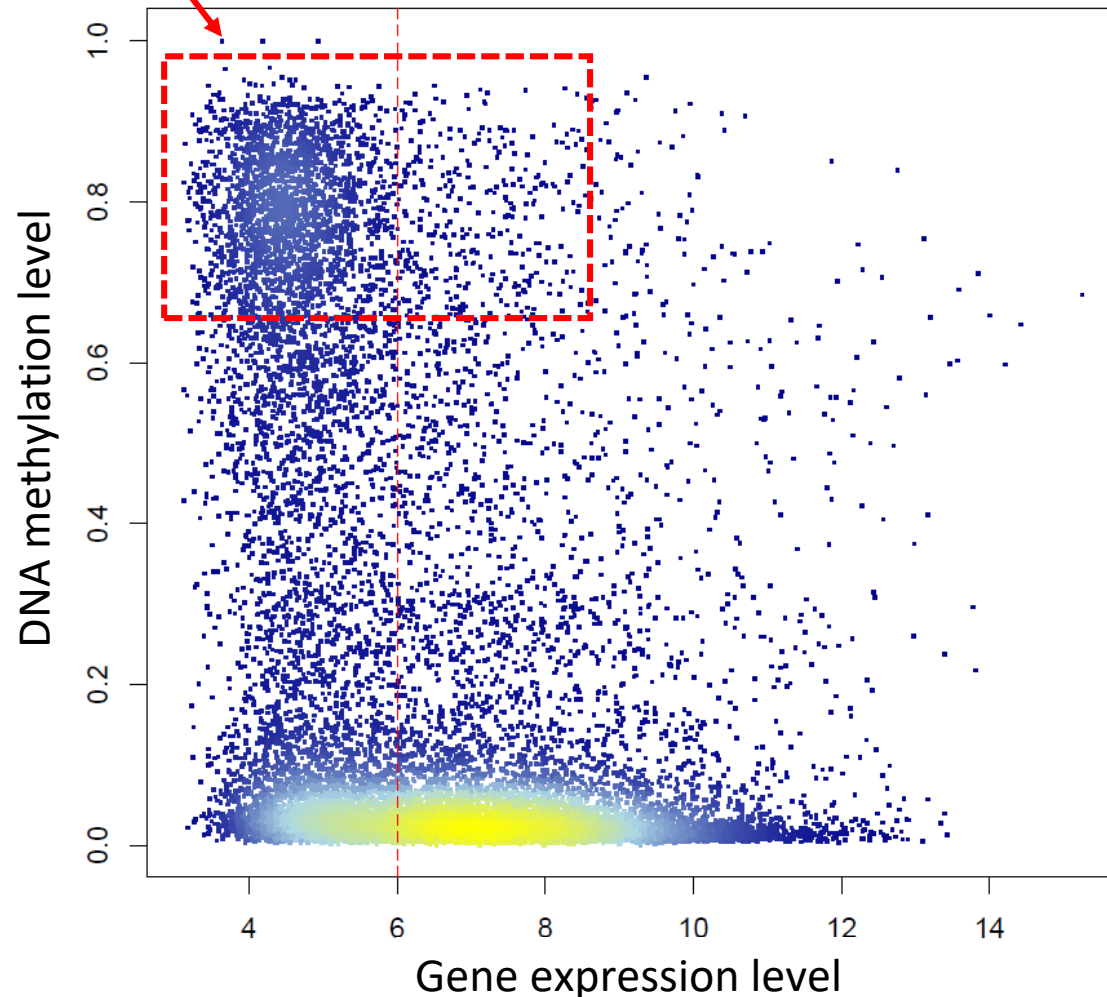
1-Can a smoke exposure signature be extracted from DNA methylation levels of cis-regulatory elements (CREs)?

- Identify differentially methylated CREs, including annotated (e.g. promoters) and unannotated (e.g. enhancers, insulators...) elements between smoke and fresh air exposed samples

Correlation between gene expression and DNA methylation at promoters

Highly methylated promoters

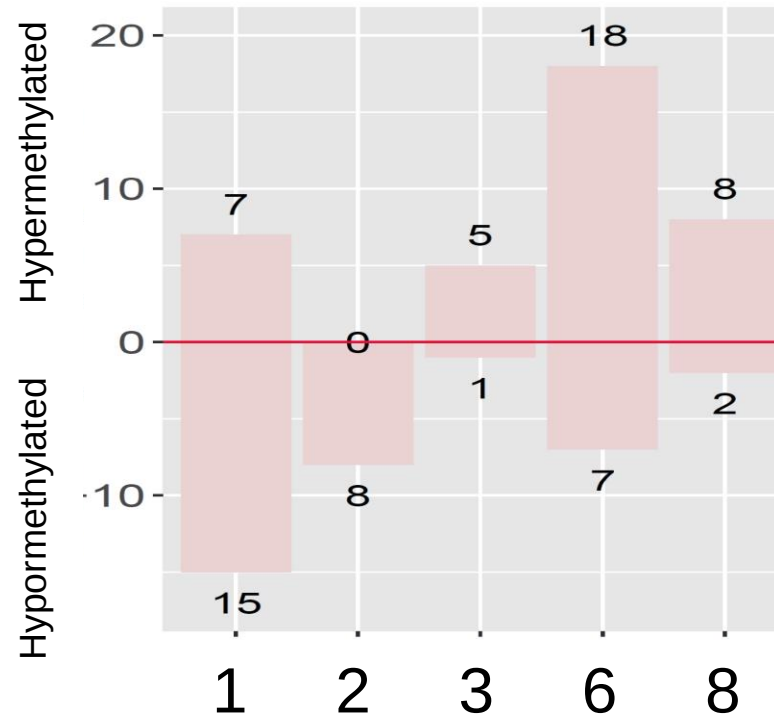
Promoters : Refseq Data base



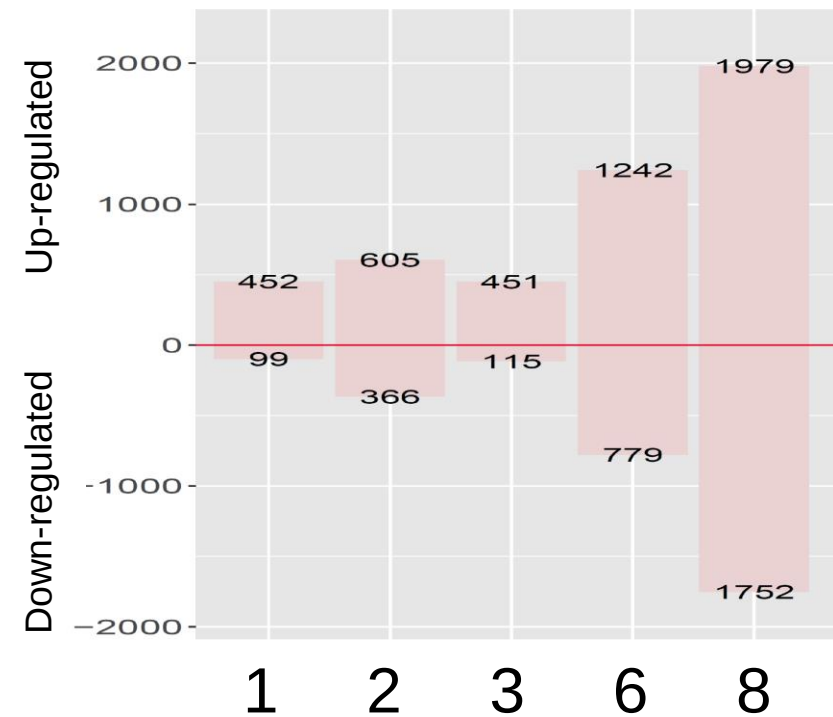
- DNA methylation at promoters shows a bimodal distribution highlighting an excellent signal-to-noise ratio that allows accurate detection of methylated promoters.
- The vast majority of methylated promoters are not expressed. This observation further validates the biological meaning of DNA methylation signal in this study.

Cigarette smoke has a limited effect on DNA methylation at promoters

Differentially methylated promoters
(FDR* < 0.05)



Differentially expressed genes
(FDR < 0.05)

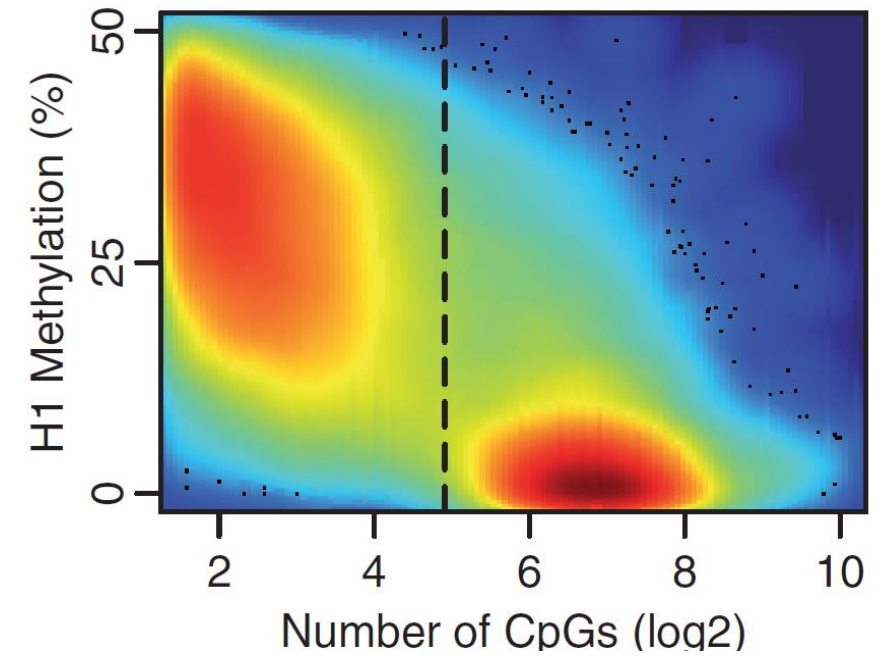
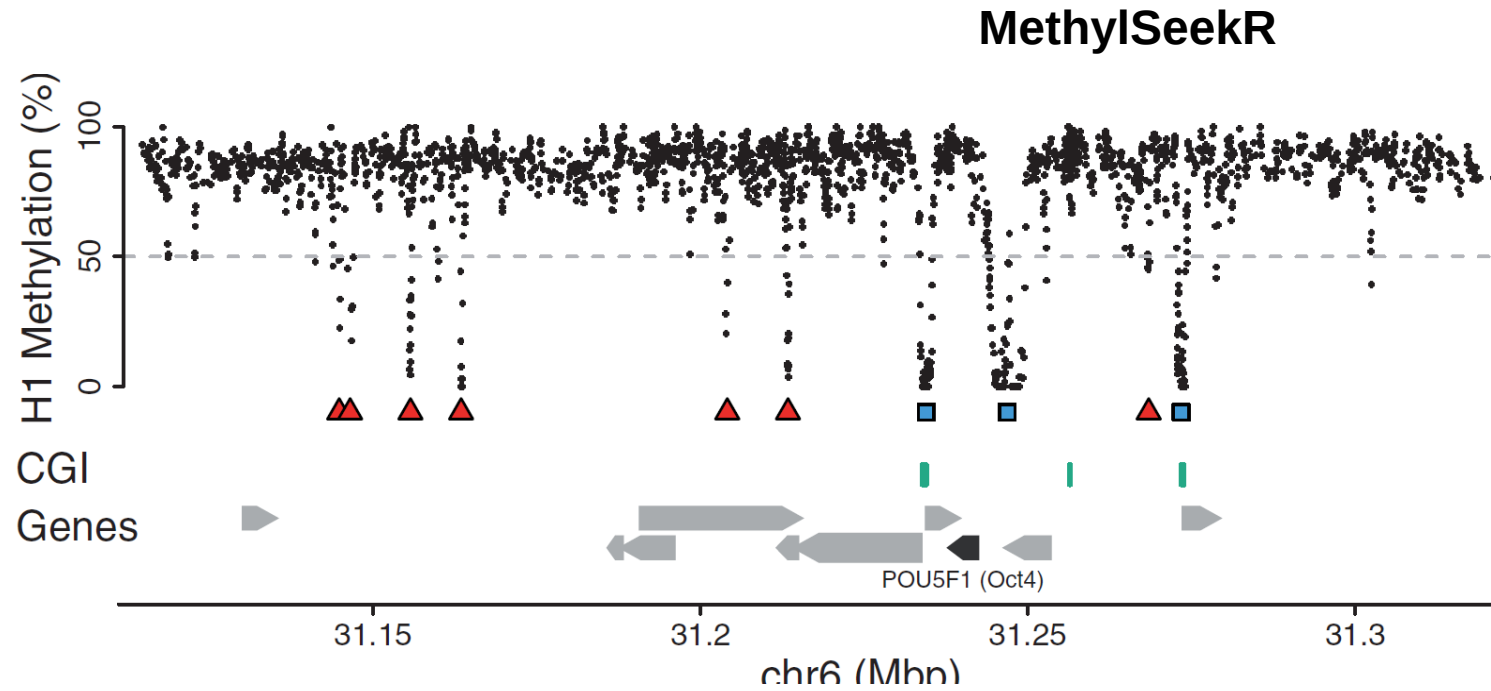


Study time (months)

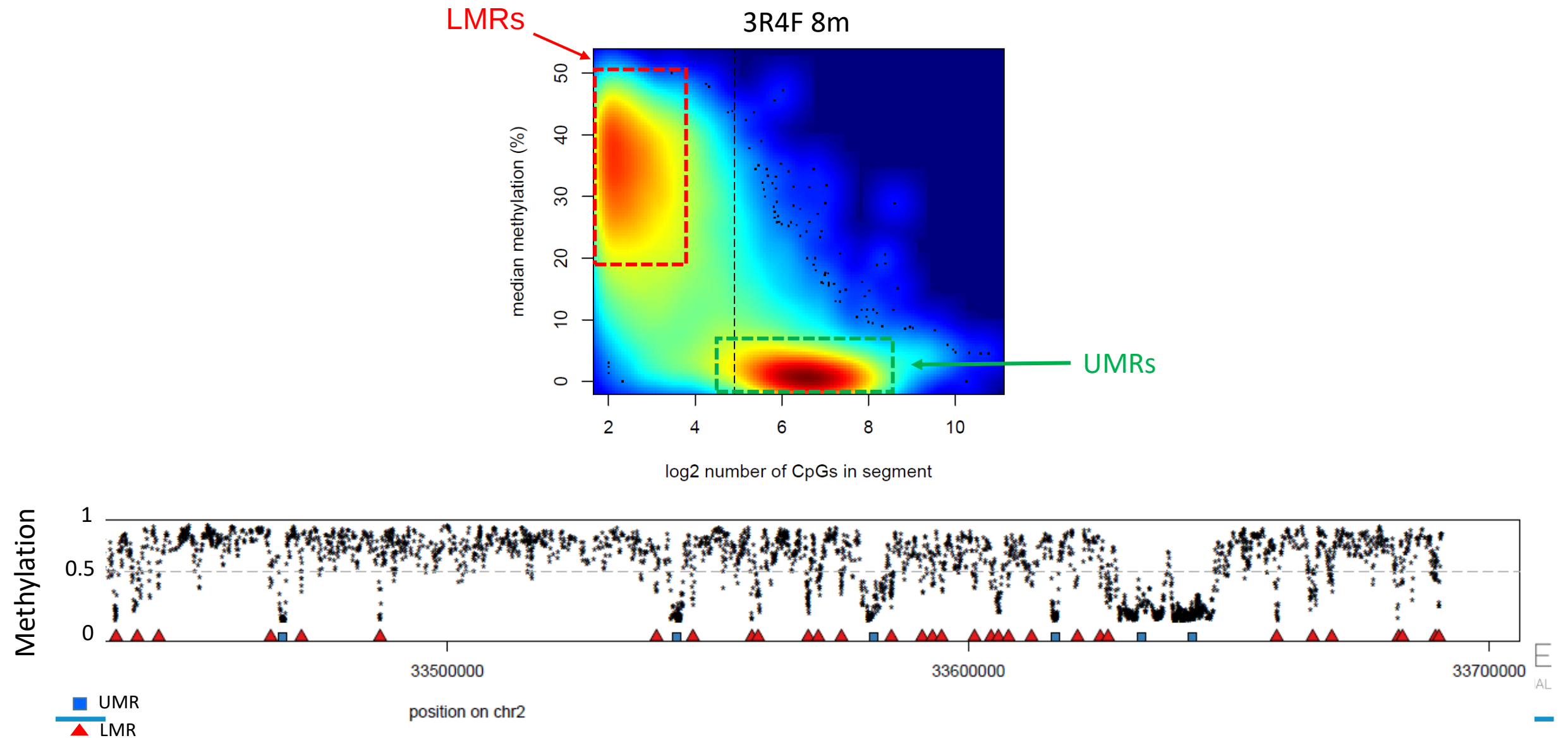
CS (3R4F)-exposed vs. sham

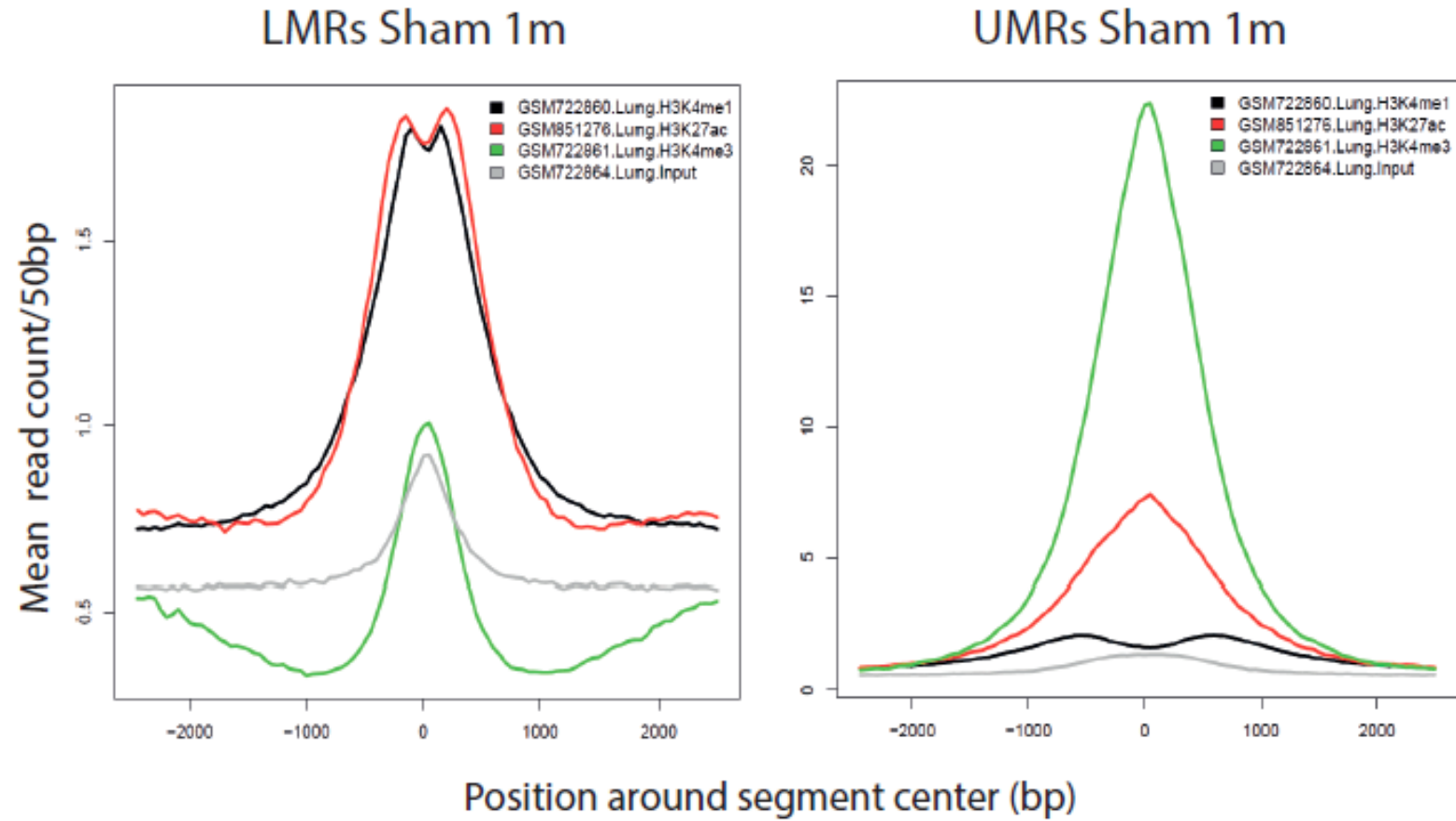
*Betabinomial test

LMR (candidate enhancers) identification from methylome data



Identification of LMRs (candidate enhancers) in our methylomes





Workflow of LMR identification and differential methylation assessment

LMRs in every condition and time point



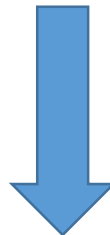
Merge overlapping elements
(60% overlap)

Combined catalog of LMRs (130k)



DNA methylation calculation

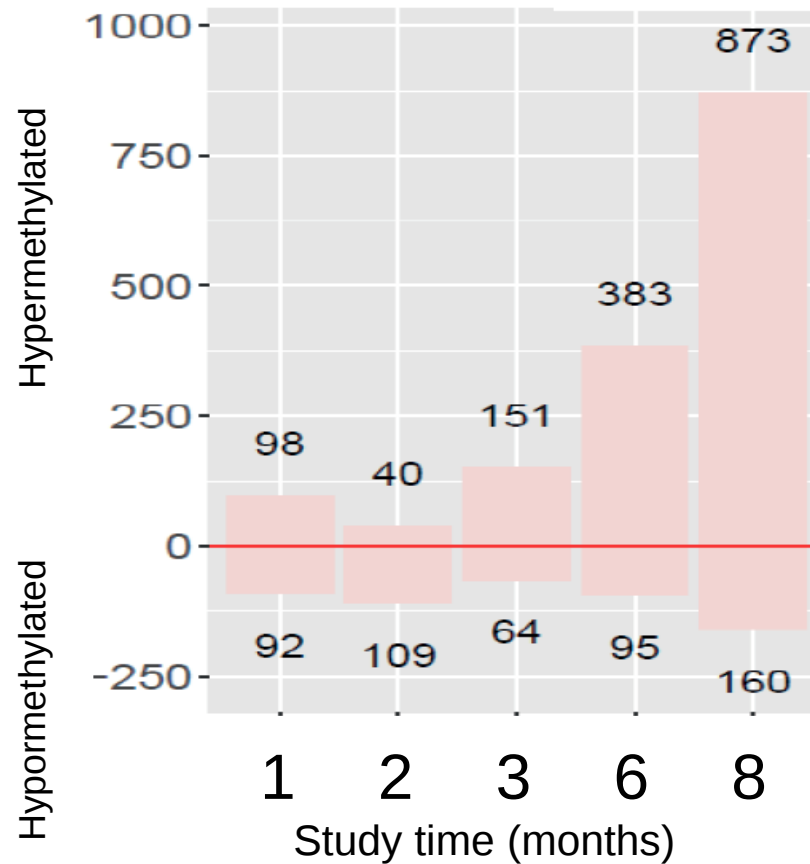
DNA methylation level for every condition



Betabinomial test

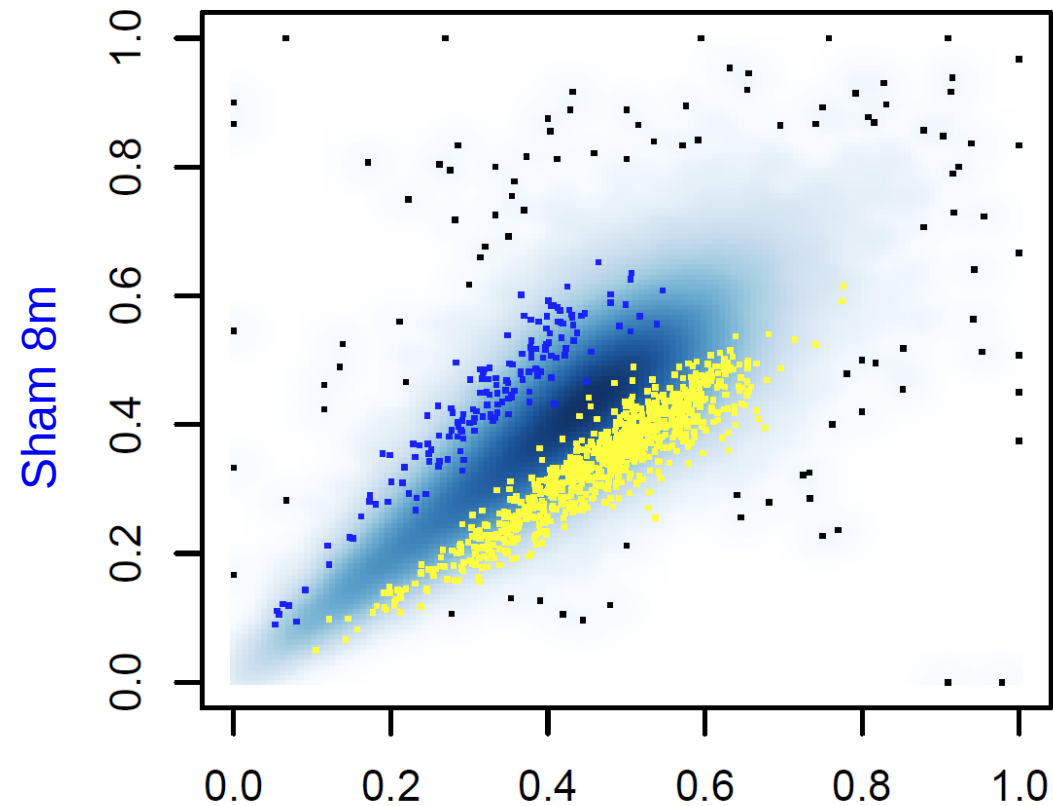
Differentially methylated LMRs
FDR < 0.05

Differentially methylated LMRs (FDR < 0.05)



CS (3R4F)-exposed vs. sham

Methylation at LMRs



3R4F 8m

1-Can a smoke exposure signature be extracted from DNA methylation levels of cis-regulatory elements (CREs)?

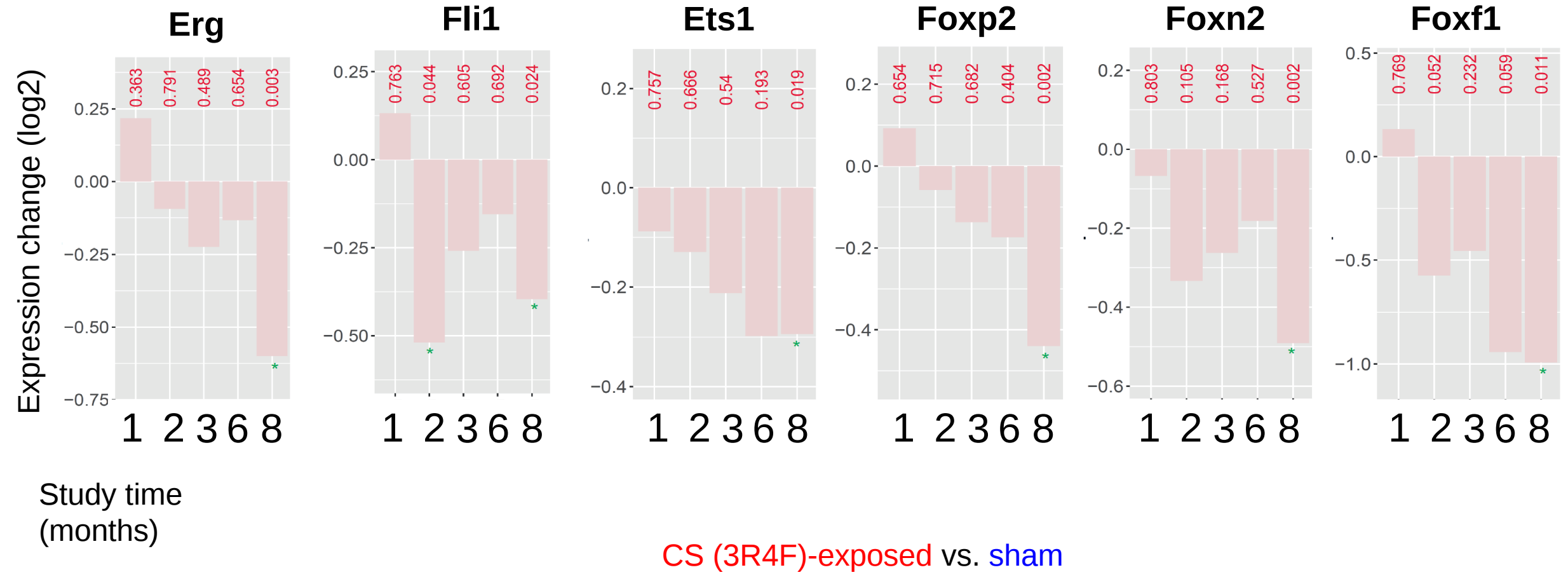
- Identify differentially methylated CREs, including annotated (e.g. promoters) and unannotated (e.g. enhancers, insulators...) elements between smoke and fresh air exposed samples
- Identify transcription factors potentially regulating the activity of the differentially methylated CREs

Transcription factor motifs enriched in LMRs hypermethylated in 3R4F group at 8 months

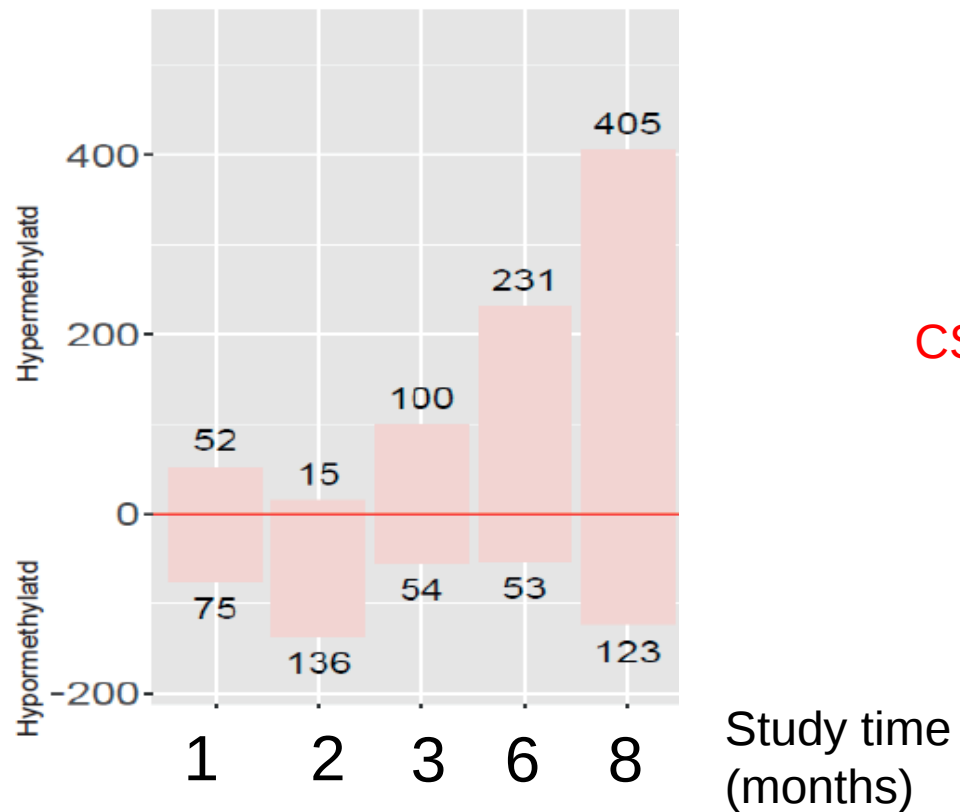
Motif name	Motif logo	P-value
ERG		1e-37
ETS1		1e-26
EWS:ERG		1e-25
FLI1		1e-24
ETV1		1e-24
ETV2		1e-23
GABPA		1e-20
FOXL2		1e-20
Foxo1		1e-17
EWS:FLI1		1e-15

<http://homer.ucsd.edu/homer/index.html>

Heinz et al; 2010



Differentially methylated enhancers
(FDR < 0.05)



CS (3R4F)-exposed vs. sham

1-Can a smoke exposure signature be extracted from DNA methylation levels of cis-regulatory elements (CREs)?

- Identify differentially methylated CREs, including annotated (e.g. promoters) and unannotated (e.g. enhancers, insulators...) elements between smoke and fresh air exposed samples
- Identify transcription factors potentially regulating the activity of the differentially methylated CREs
- Extract a smoke exposure signature from DNA methylation levels of the identified CREs
- Classify each sample in the test set using the CRE smoke exposure signature extracted from DNA methylation data, providing the probability that a sample belongs to the 3R4F exposed group

Classifier: Linear Discrimination Analysis (LDA)

Training set

Signal ranking :

- Moderate t-test between groups

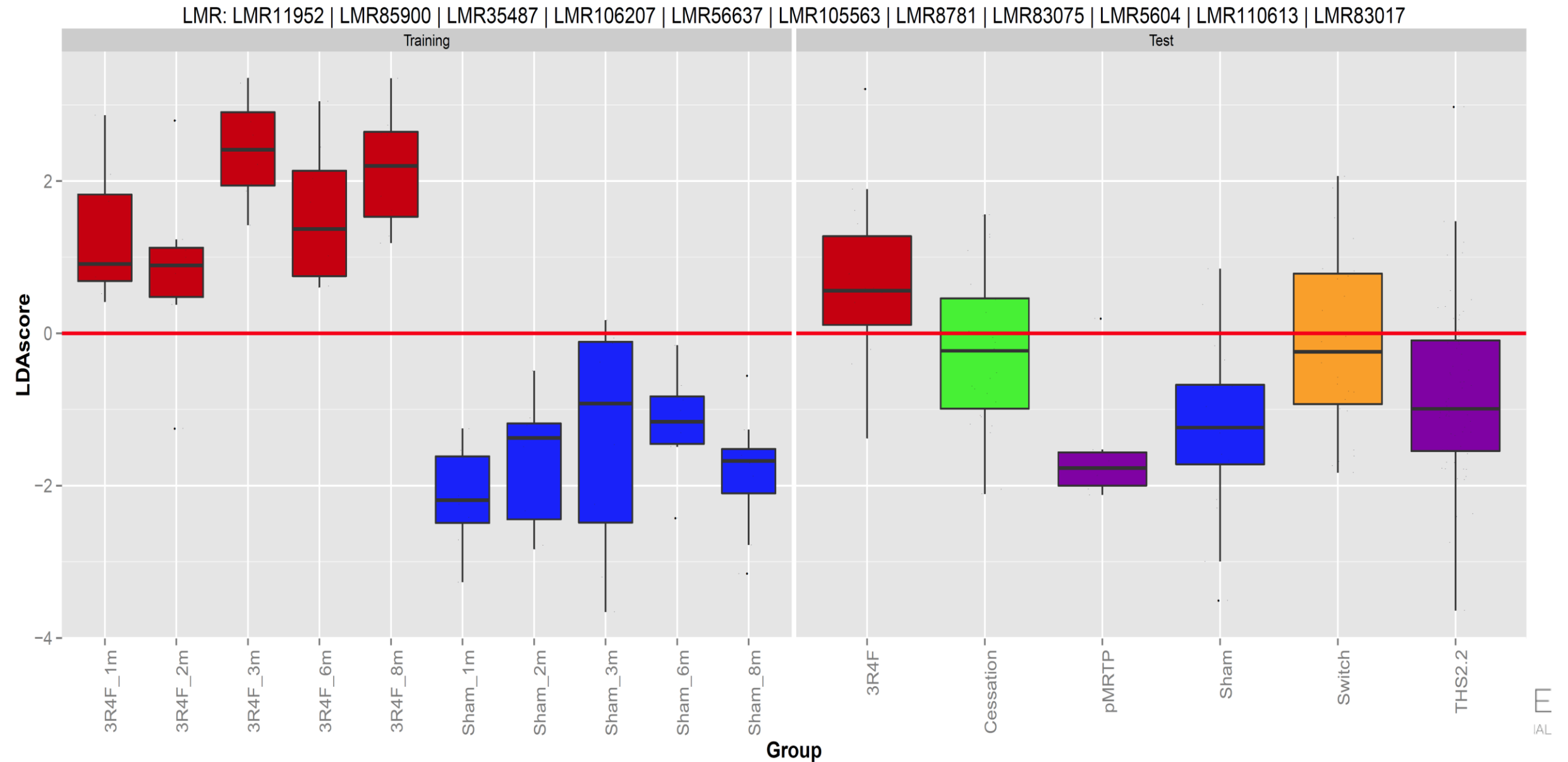
Feature selection:

- Begin with the top **d=2** features and move sequentially through the list, one element at a time
- The best value of **d** is chosen by Maximizing a 5-fold cross-validated performance (MCC) with 5 iterations

Testing set

Classification:

Apply the LDA model and make predictions assuming equal priors of each class in the testing set



1-Can a smoke exposure signature be extracted from DNA methylation levels of cis-regulatory elements (CREs)?

- Identify differentially methylated CREs, including annotated (e.g. promoters) and unannotated (e.g. enhancers, insulators...) elements between smoke and fresh air exposed samples
- Identify transcription factors potentially regulating the activity of the differentially methylated CREs
- Extract a smoke exposure signature from DNA methylation levels of the identified CREs
- Classify each sample in the test set using the CRE smoke exposure signature extracted from DNA methylation data, providing the probability that a sample belongs to the 3R4F exposed group

2-Can a smoke exposure signature be extracted from expression data of genes controlled by differentially methylated DNA cis-regulatory elements?

- Identify the target genes controlled by the CREs from Question 1 Step 1. above
- Extract a smoke exposure signature from the expression data of genes controlled by the 1000 most differentially methylated CREs between smoke and fresh air exposed samples
- Classify each sample in the test set using the smoke exposure signature extracted from expression data, providing the probability that a sample belongs to the 3R4F exposed group



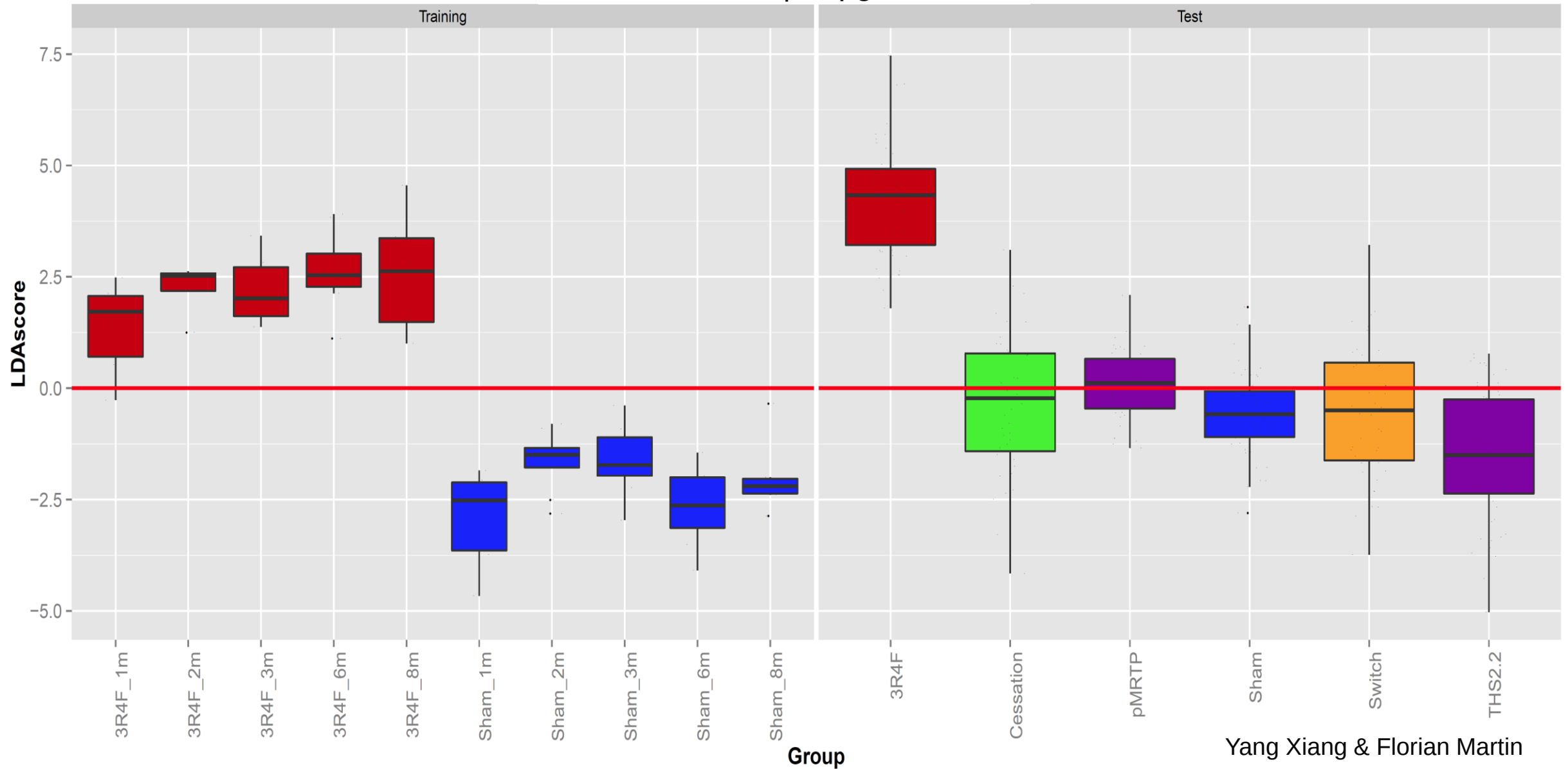
LMR selection :

- Differentially methylated LMRs (FDR : 0.05)
- Consistent direction (Hyper / Hypo) through the 5 time points

Gene selection :

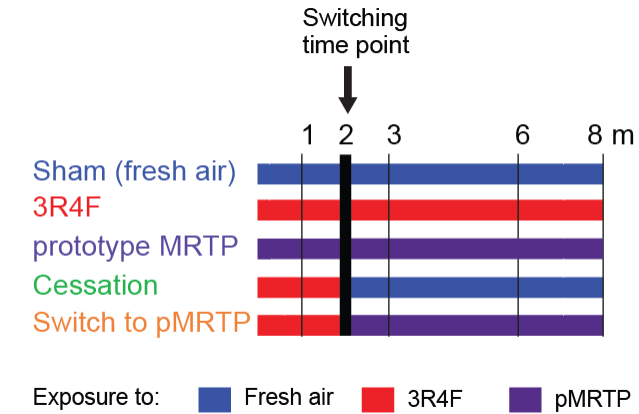
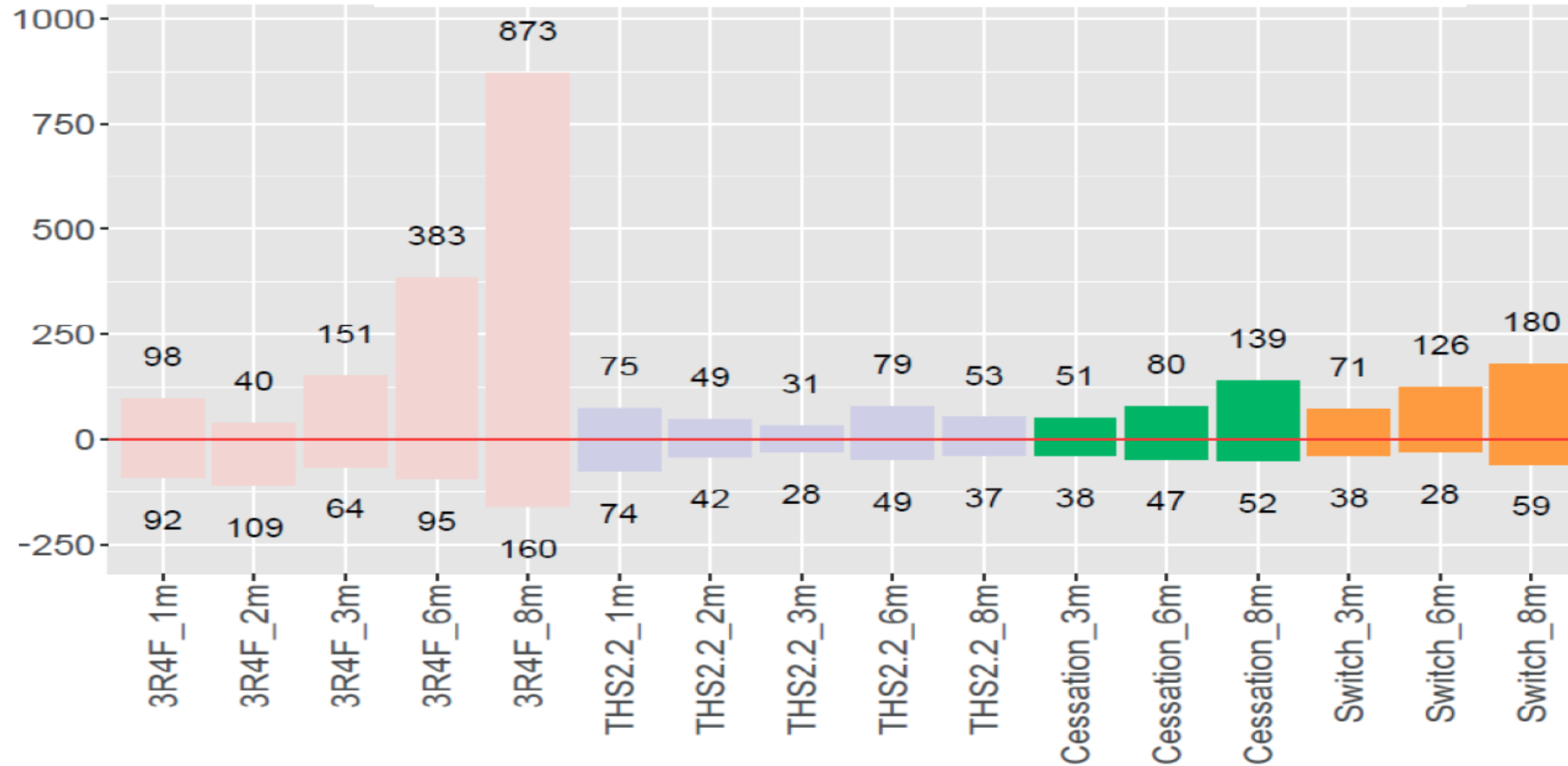
- Select the closest promoter / gene
- Select genes with expression change anticorrelating with methylation change of the corresponding LMRs

F7 | Capg



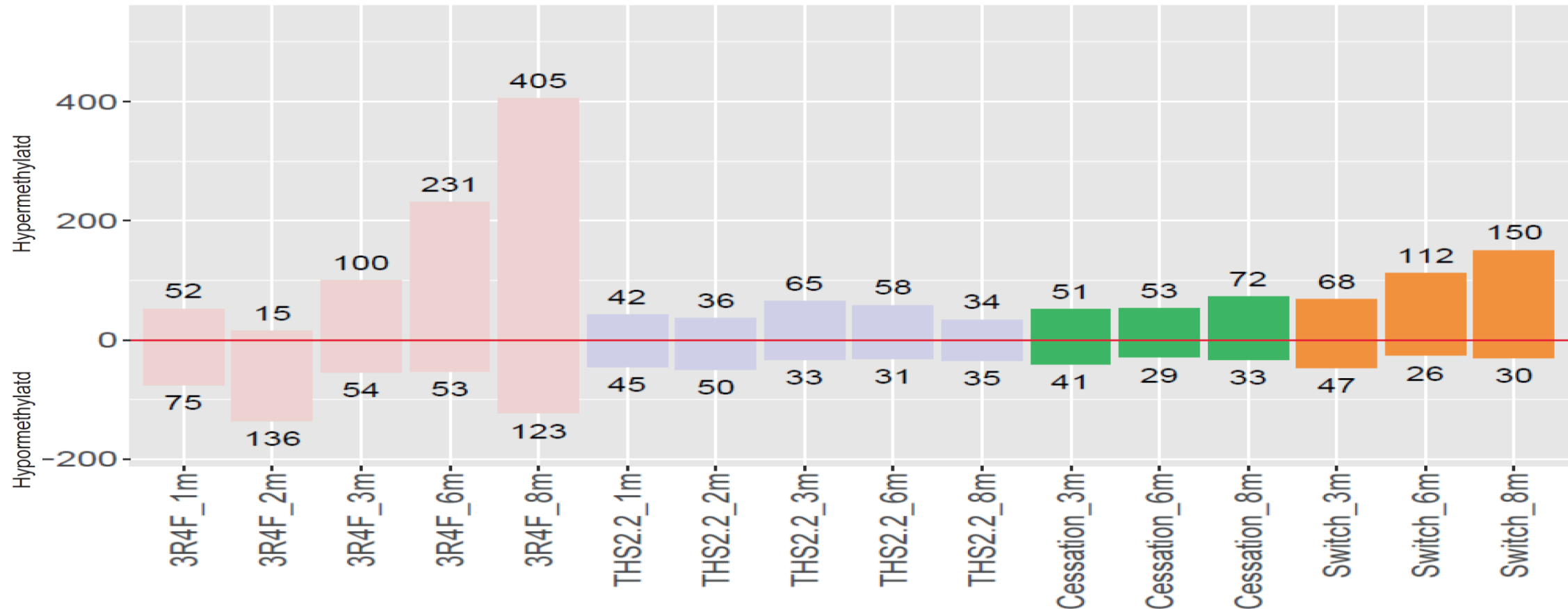
Cigarette smoke but not THS induces hypermethylation at enhancers

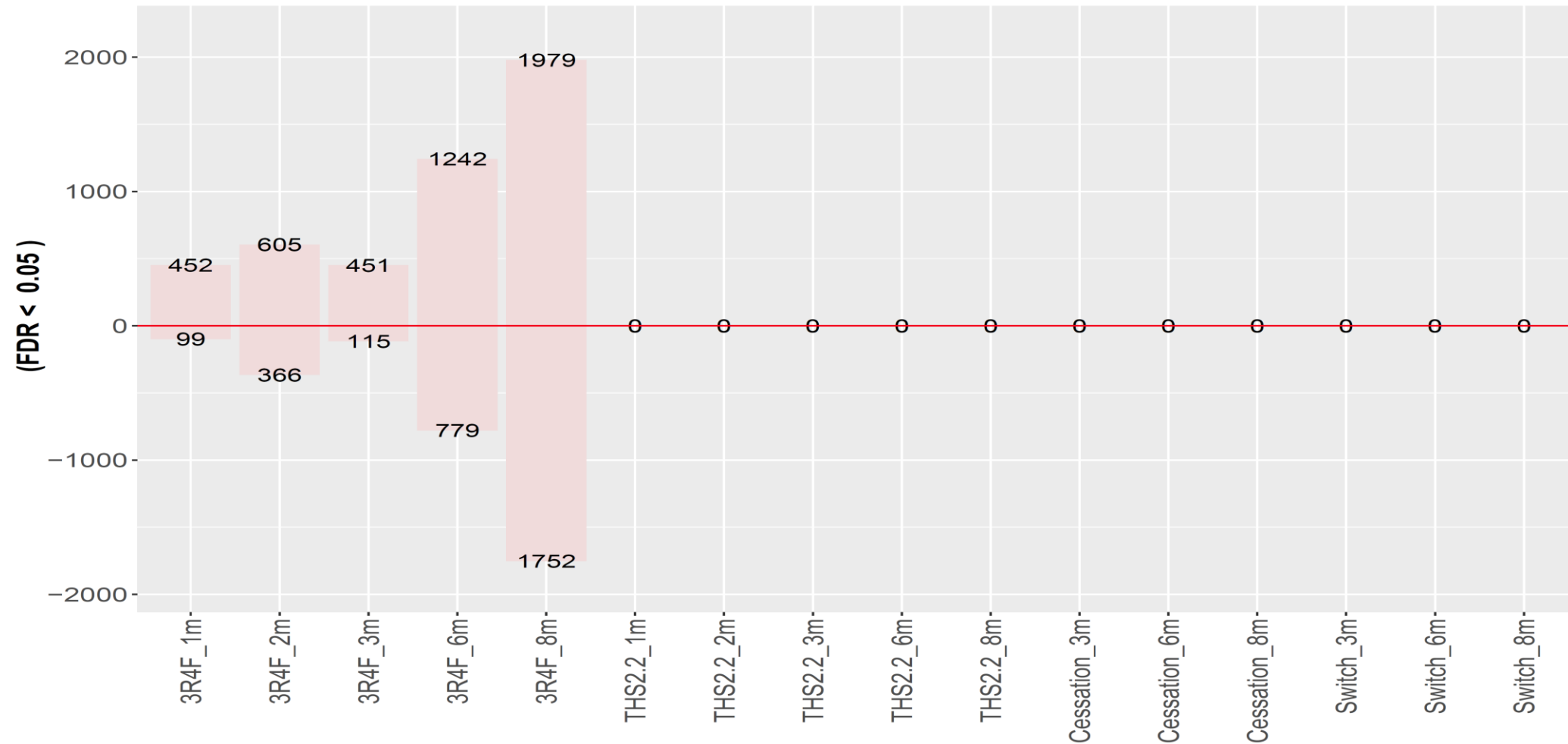
Differentially methylated LMRs (FDR < 0.05)



Cigarette smoke but not THS induces hypermethylation at enhancers

Number of differentially methylated enhancers as defined by chromatin signature





Apoe study

Phillips et al ; 2016

- Cigarette smoke affects methylation level of very few promoters
- Cigarette smoke exposure affects the methylation of hundreds of LMRs (Enhancers)
- This effect is not observed for THS2.2 exposure and is strongly reduced upon cessation or switching to THS2.2
- Hypermethylated LMRs in 3R4F exposed samples are mainly enriched for ETS and Fox motifs in agreement with their role in lung function
- DNA methylation can be used as a marker for cigarette smoke exposure
- Further epigenetic investigations are required to better understand the underlying mechanisms (Mapping of transcription factors, histone marks, chromatin organization, etc ...)

Team members

Nicolas Sierro

Sonia Ouadi

Jerome Thomas

Statistics and Machine Learning

Florian Martin

Yang Xiang

Nikolai Ivanov

Julia Hoeng

Manuel Peitsch

SBV Improver Team

Thank you for your attention !



Support Slides

Choosing the optimal statistical model:

➤ **Hidden Markov models (HMMs)** [ComMet](#), [Bisulfighter...](#)

➤ **Fisher's Exact test** [methykit...](#)

Simple comparison (sample vs control)

Does not take into account biological variability

Regression methods:

Take into account the overdispersion

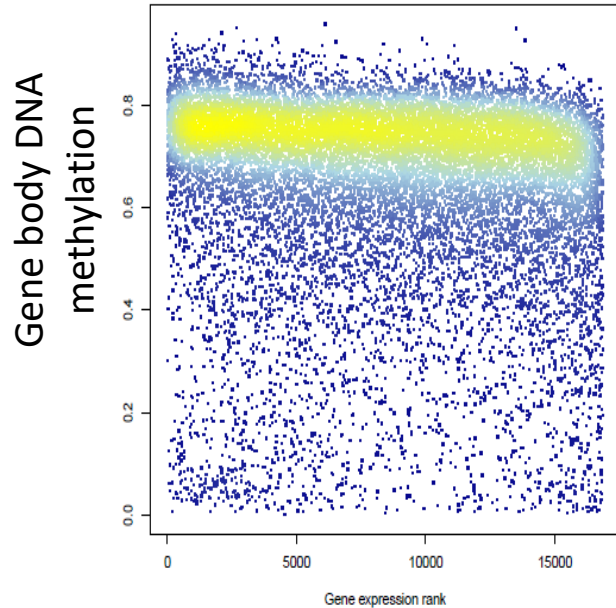
➤ **Linear regression** [Limma](#), [RnBeads](#), [BSmooth](#)

➤ **Logistic** [Methyl kit](#)

➤ **Betabinomial** [DSS](#), [RADMeth](#), [BiSeq...](#)

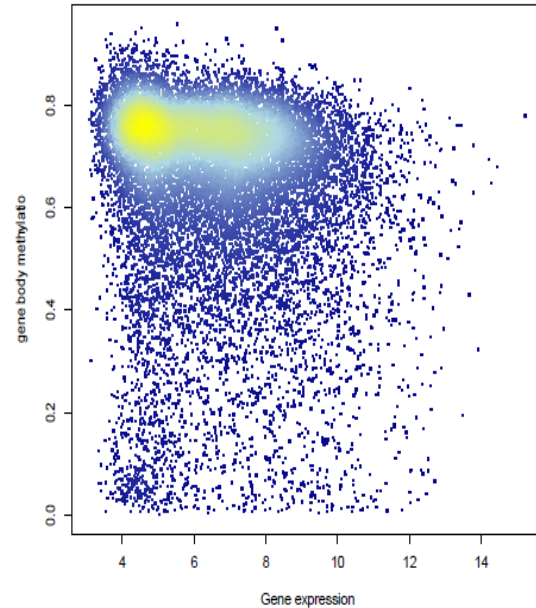
Performs better when there is more variance than expected

Sham – 8m



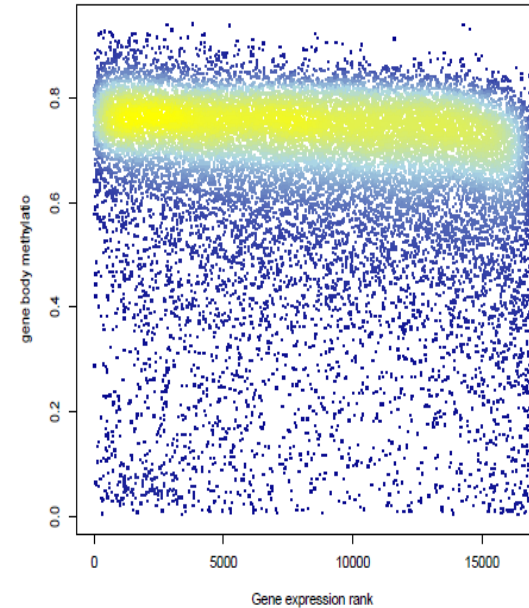
Ranked Gene expression

Sham – 8m



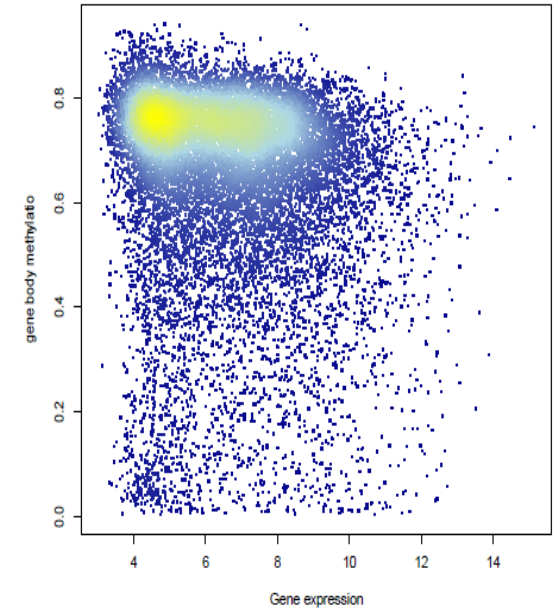
Raw Gene expression

CS (3R4F) – 8m



Ranked Gene expression

CS (3R4F) – 8m



Raw Gene expression

No correlation between gene expression and gene body methylation was observed in our samples