Identification de relations enhancer/gène dans les génomes animaux

Mestre Camille

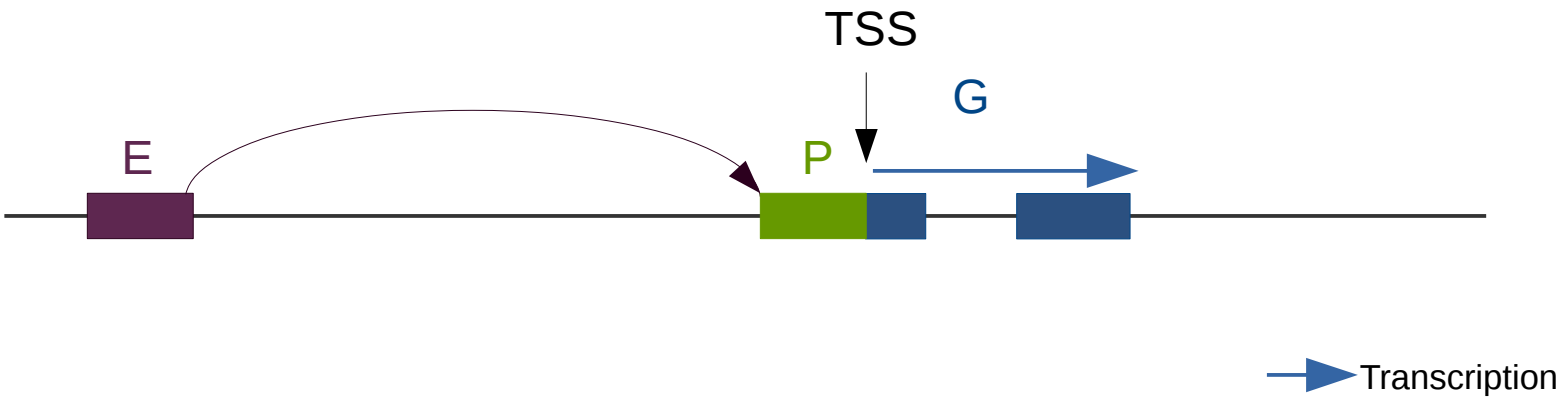
Zouine Mohammed

Djebali Sarah

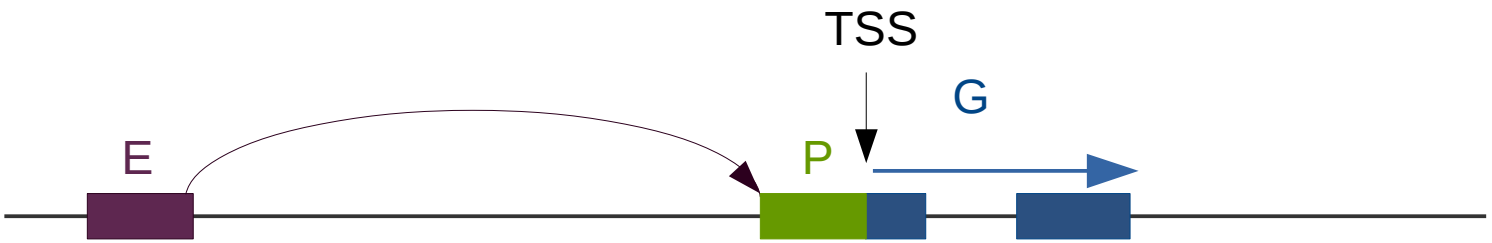
Plan

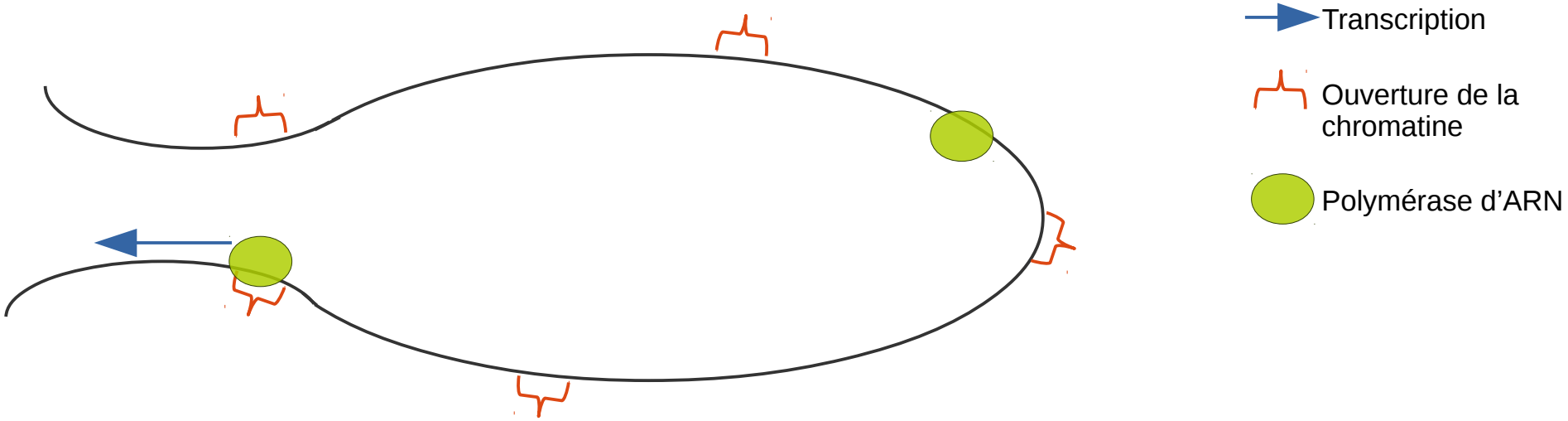
Plan

Contexte et définition

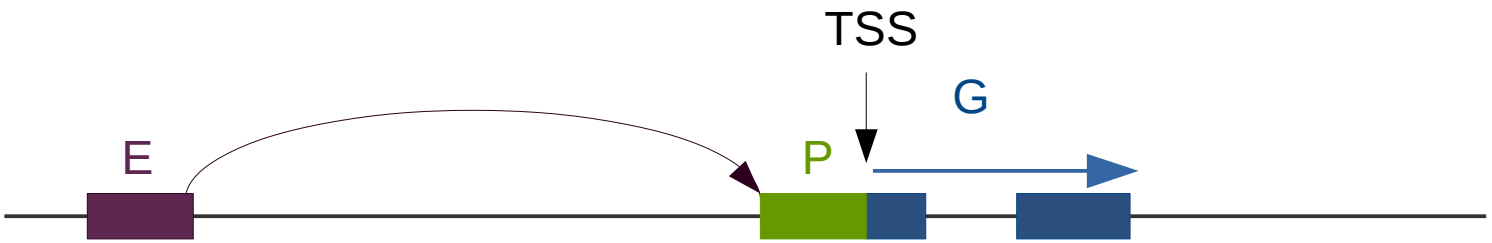


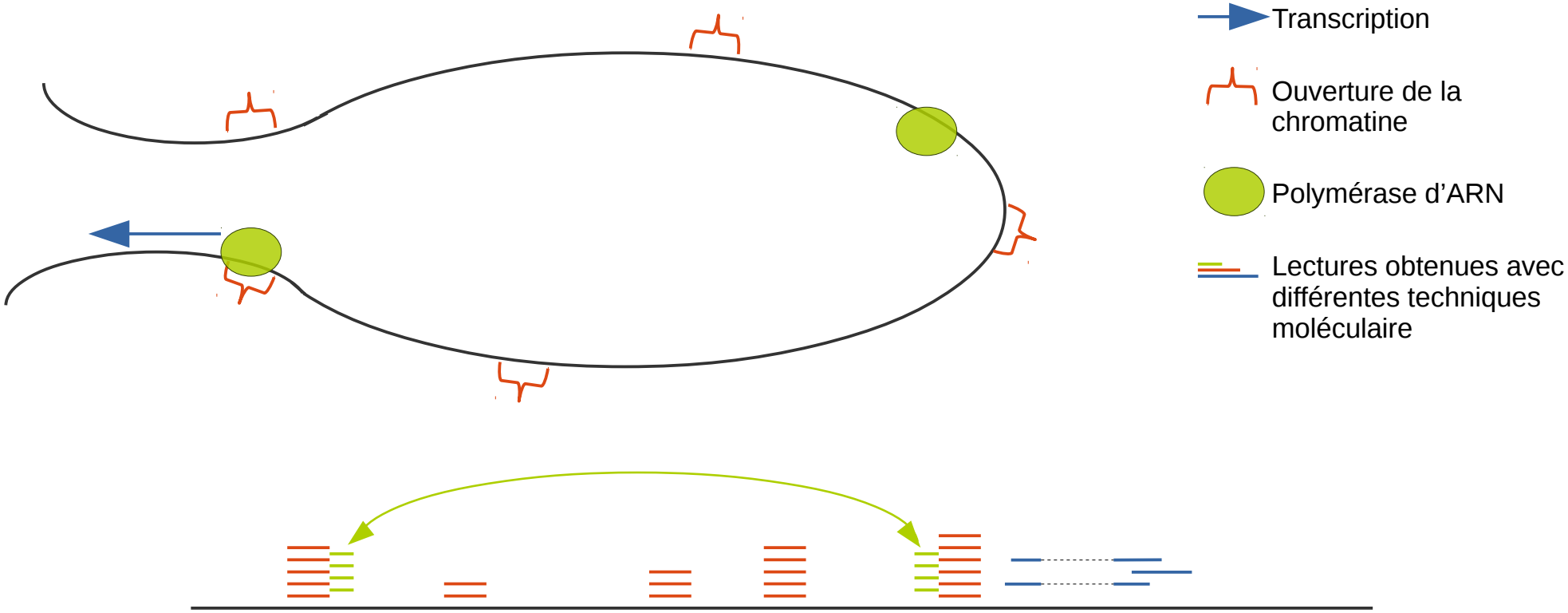
Contexte et définition



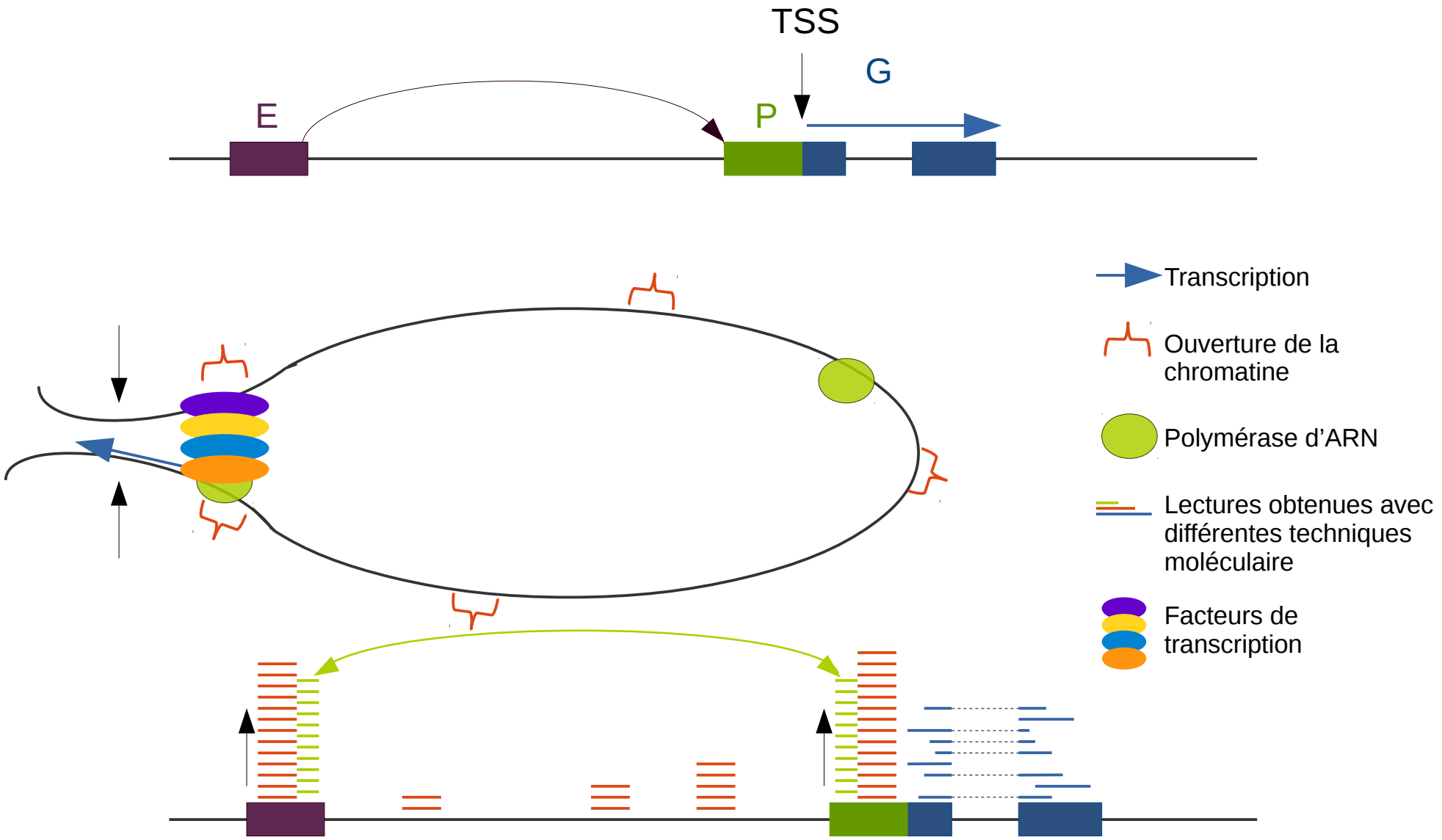


Contexte et définition

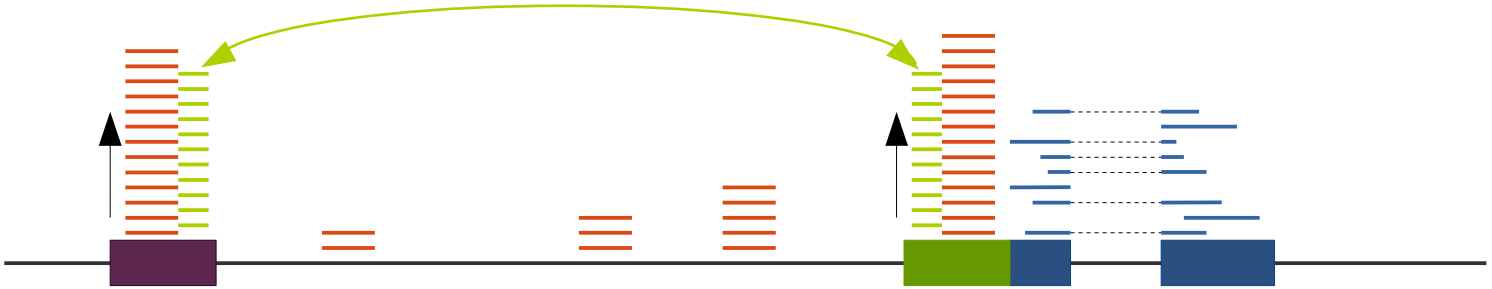




Contexte et définition



Terminologie



- **Enhancer :** un élément régulateur distant
- **TSS :** le départ de la transcription
- **Promoteur :** la région qui se situe à plus ou moins 1kb du TSS
- **Dnase-seq :** la mesure de l'ouverture de la chromatine
- **RNA-seq :** la mesure de la transcription

Contexte et définition

Problématiques

- Est-il possible de constituer un ensemble de relations enhancer/gène de référence ?
- Les données d'ouverture de la chromatine et les données de transcription permettent-elles d'identifier les relations enhancer/gène ?
- Quelles sont les méthodes d'identification simples existantes ?
- Laquelle de ces méthodes est la plus fiable, la plus facile à mettre en œuvre et celle nécessitant le moins de données différentes ?

Plan

I. Contexte et définitions

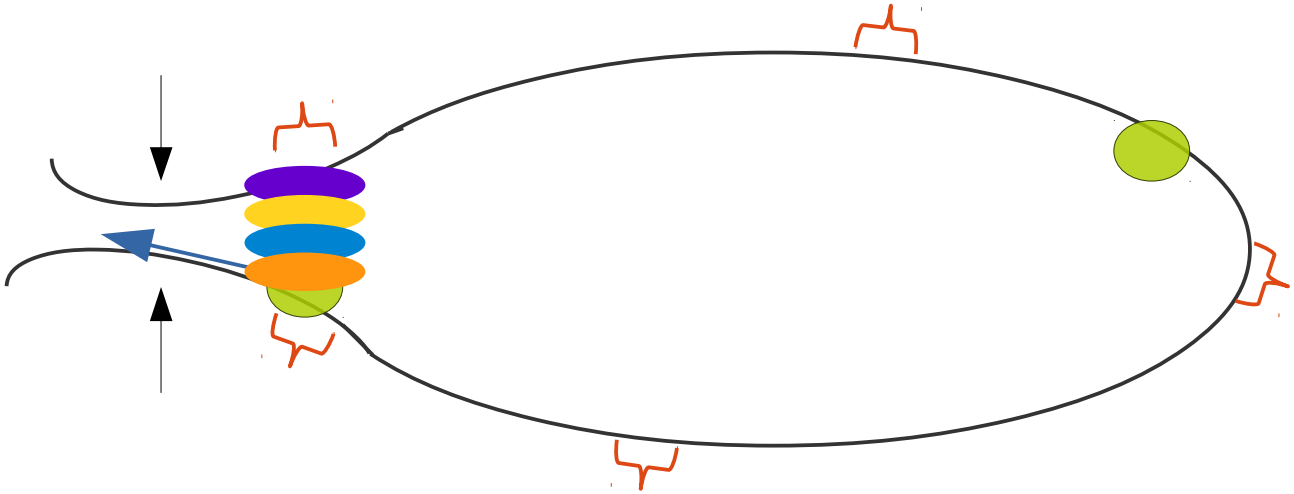
II. Interactions spatiales

III. Ouverture de la chromatine

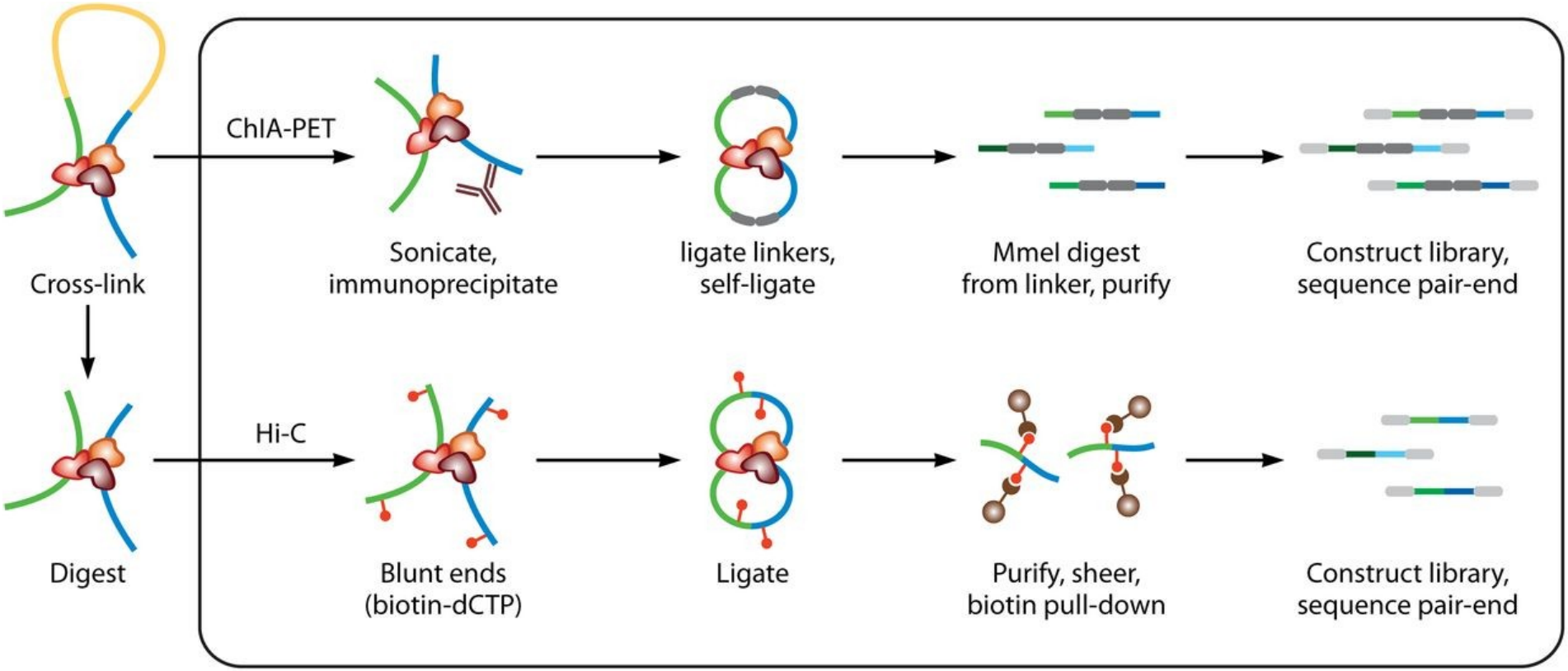
IV. Expression des gènes associée à l'ouverture de la chromatine

V. Conclusions et discussion

Interactions spatiales

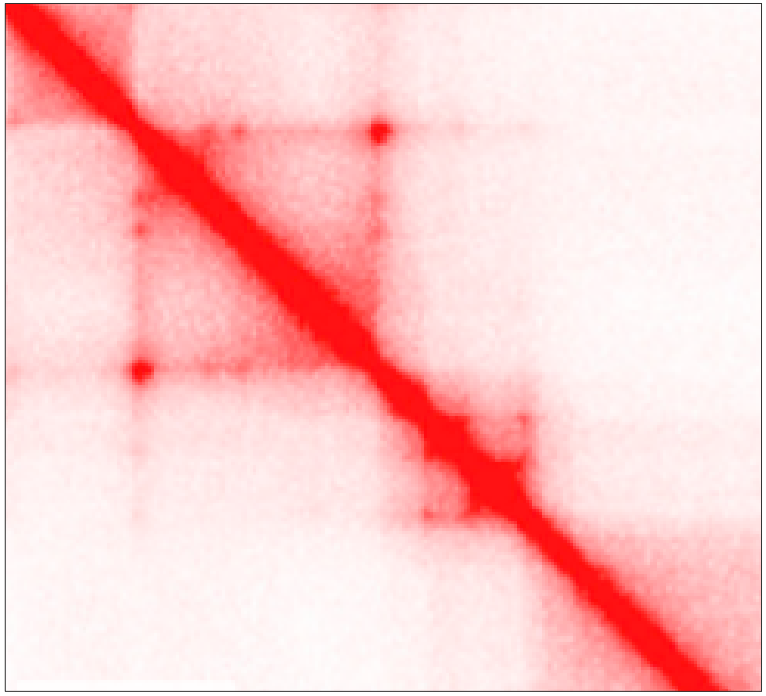


Interactions spatiales

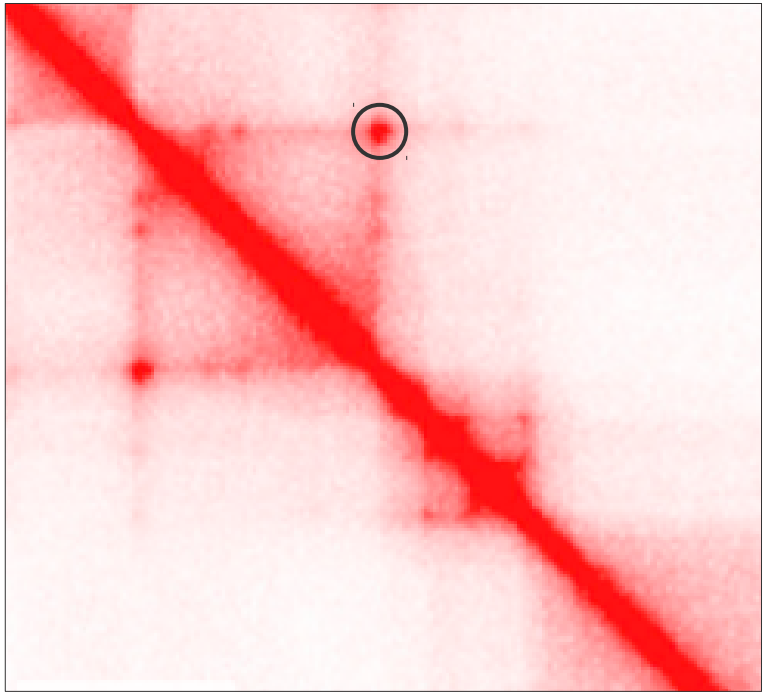


Bouwman and de Laat. Genome biology, 16(1):154, 2015

Interactions spatiales



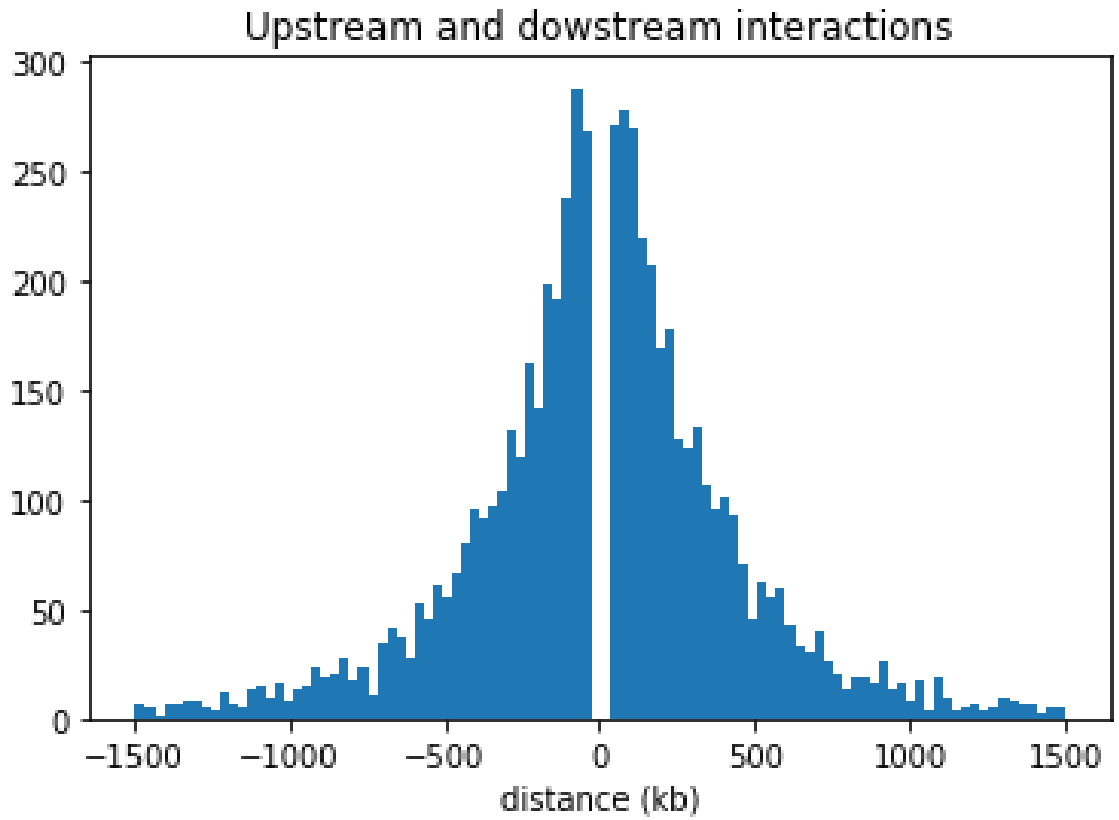
Interactions spatiales



Interactions spatiales

	cHi-C	*in situ Hi-C*	ChIA-PET
Nombre de paires	1 618 000	9 448	113 561
Intrachromosomique	99,9 % 1 617 904	100 %	88 % 100 263
Taille d'intervalles (en bp)	930 (de 5 à 11 000 000)	6 660 (de 5 000 à 10 000)	880 (de 10 à 23 000)
Inter intervalles (en bp)	315 000 (de 10 à 191 000 000)	1 170 000 (de 30 000 à 220 690 000)	850 000 (de 5 à 227 000 000)
Interactions impliquant au moins un promoteur	99 % 1 602 356	51,8 % 4 896	95,3 % 108 192
Score	log(observé/attendu)	Valeur observée, valeur attendue, FDR	Nombre de PET
Espèce/lignée cellulaire	Homme / GM12878		
Référence	Mifsud et al., Mapping long-range promoter contacts in human cells with high-resolution capture hi-c.	Rao et al., A 3d map of the human genome at kilobase resolution reveals principles of chromatin looping	Tang et al., Ctcf-mediated human 3d genome architecture reveals chromatin topology for transcription

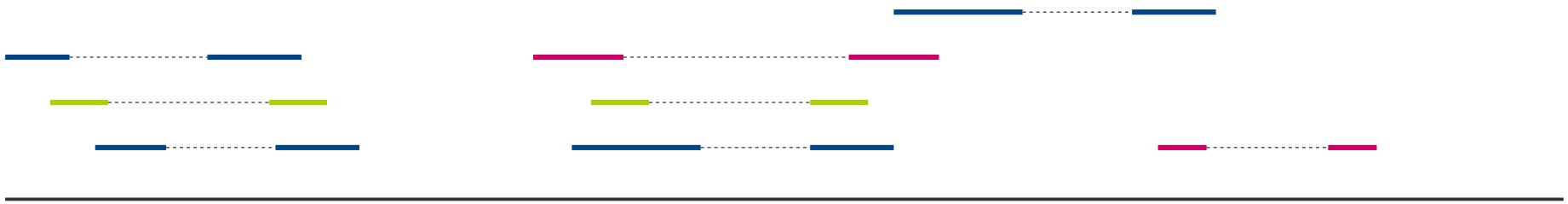
Interactions spatiales



Interactions promoteurs/autres

Effet distance important

Interactions spatiales

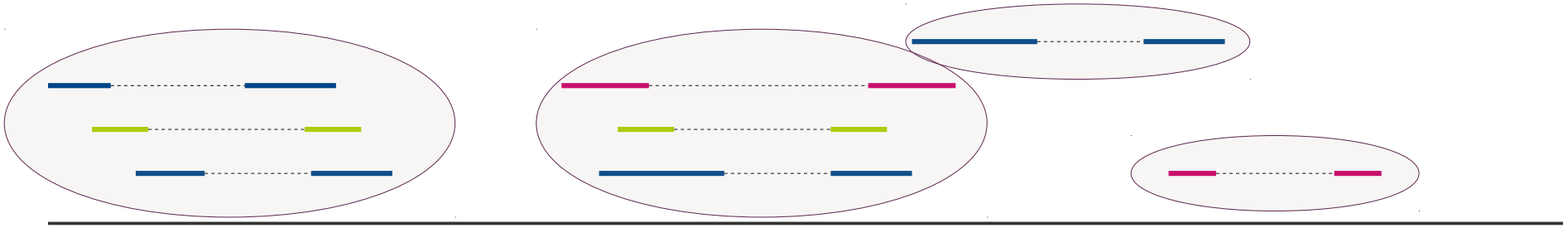


Interaction observée avec *in situ* Hi-C

Interaction observée avec cHi-C

Interaction observée avec ChIA-PET

Interactions spatiales



Interaction observée avec *in situ* Hi-C

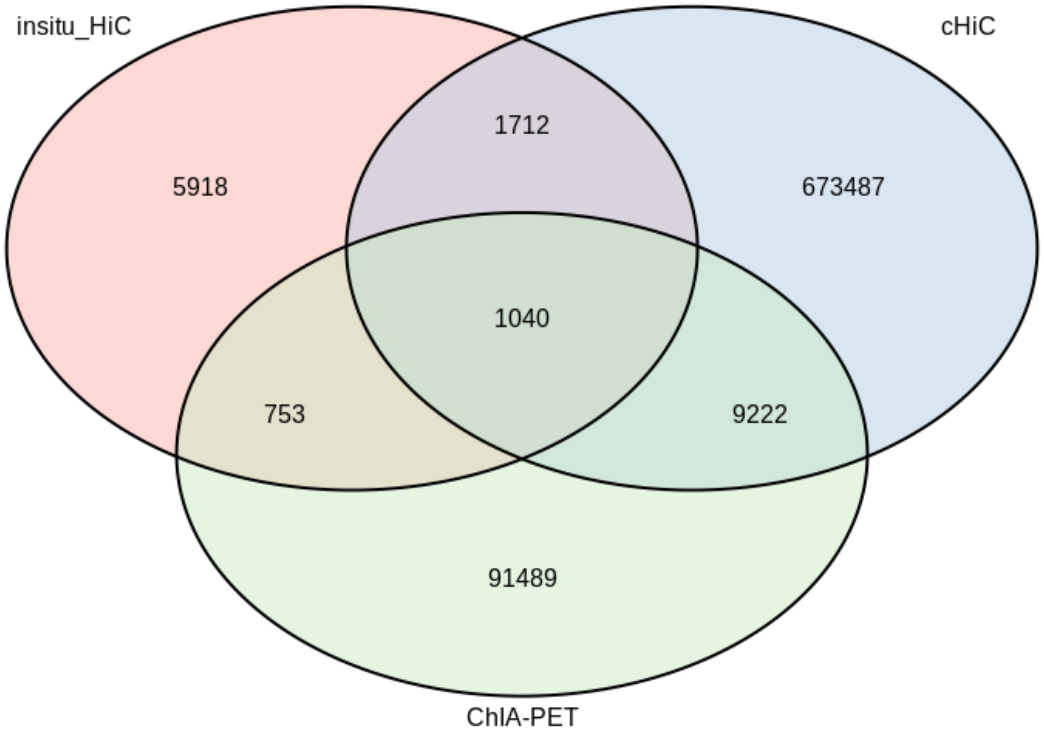
Interaction observée avec cHi-C

Interaction observée avec ChIA-PET

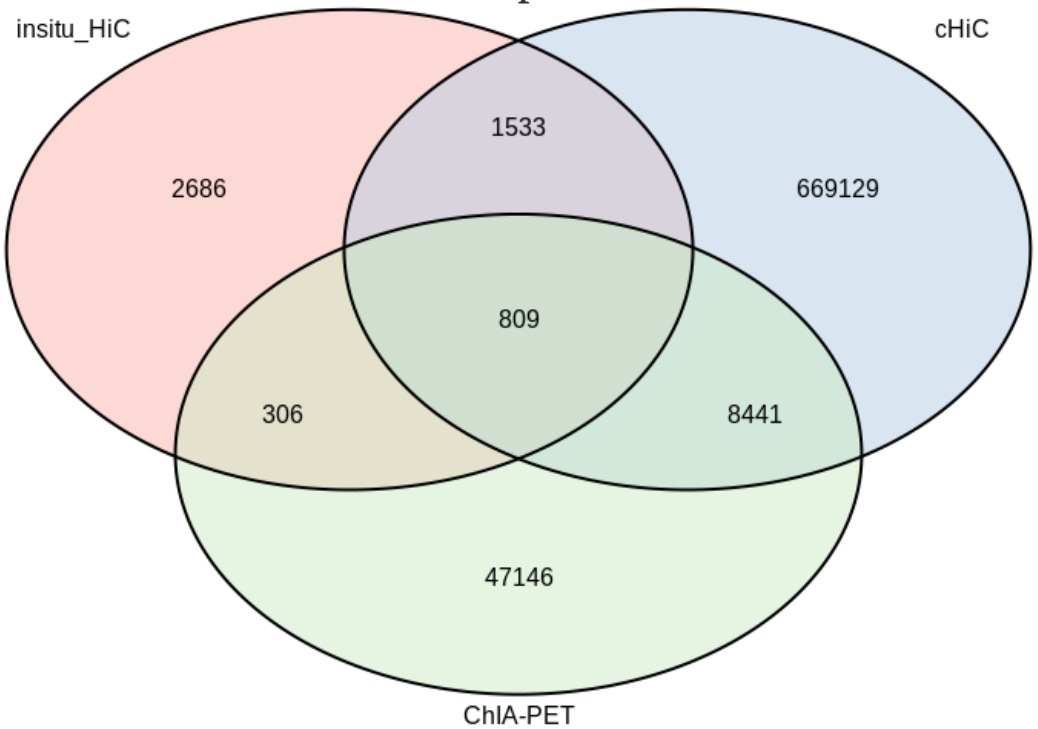
Groupement des interactions en composantes connexes

Interactions spatiales

Toutes les interactions spatiales

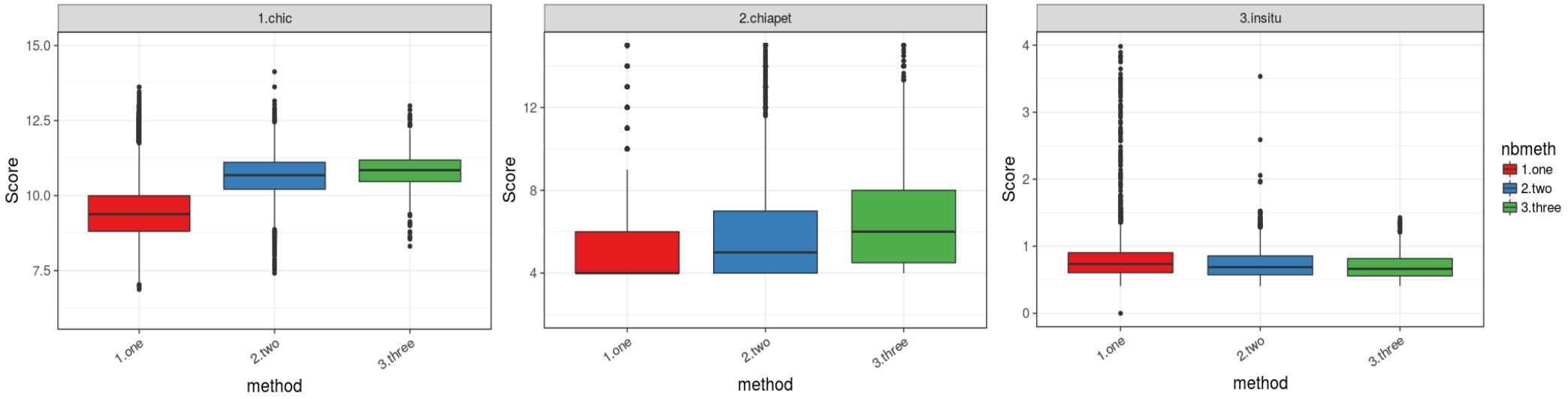


Interactions spatiales incluant au moins un promoteur



Intersection des trois techniques moléculaires : ChIA-PET, *in situ* Hi-C et cHi-C.

Interactions spatiales



Plan

I. Contexte et définitions

II. Interactions spatiales

III. Ouverture de la chromatine

IV. Expression des gènes associée à l'ouverture de la chromatine

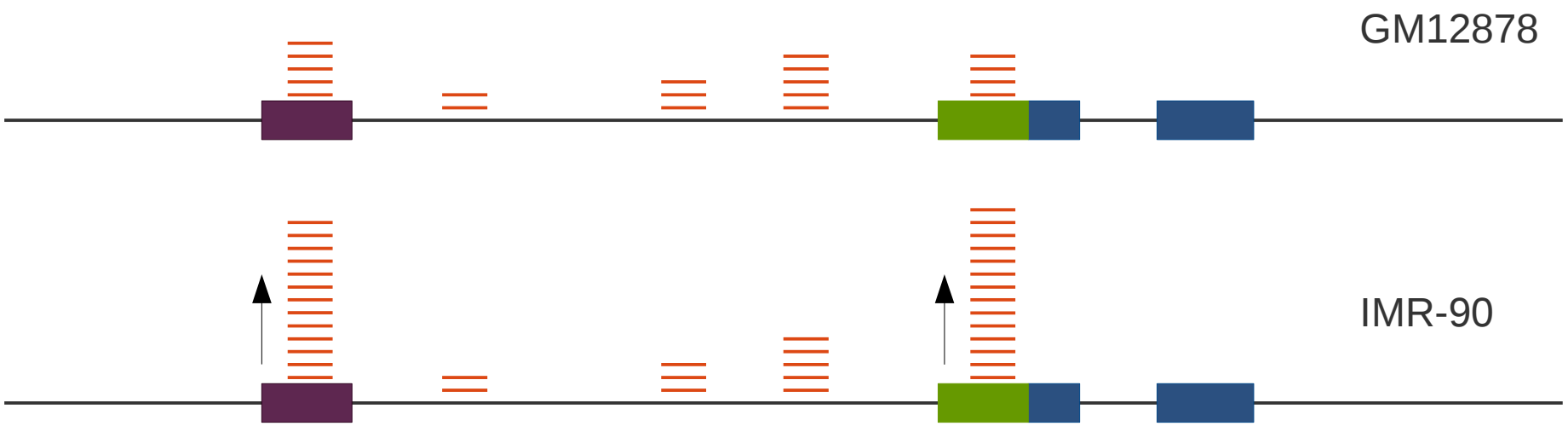
V. Conclusions et discussion

Ouverture de la chromatine

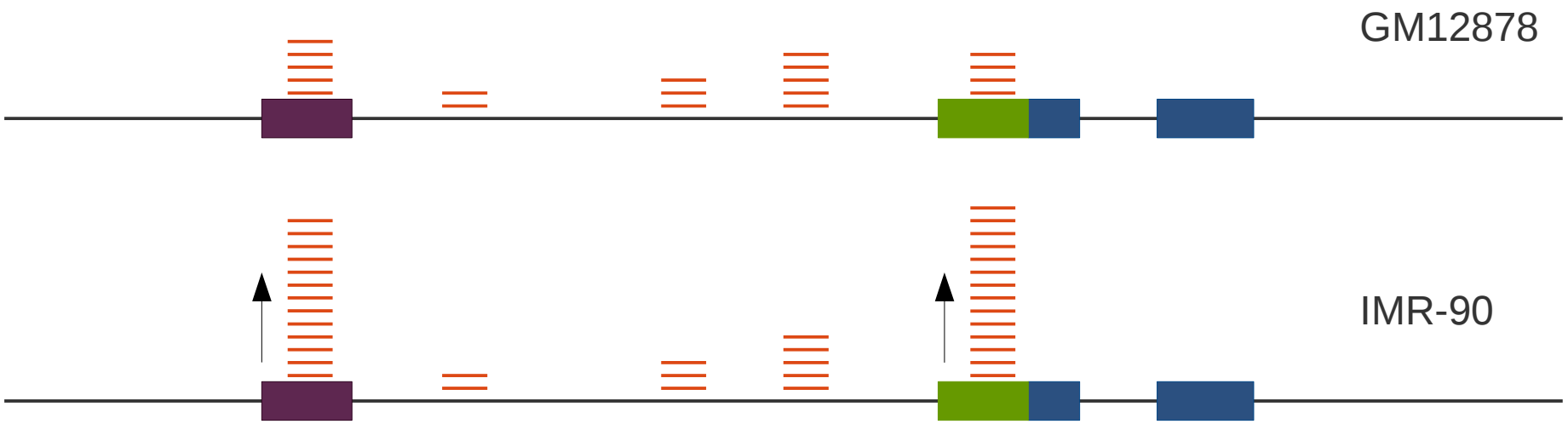
- **10 tissus et lignées cellulaires**
  - GM12878
  - IMR 90
  - small_intestine
  - adrenal_gland
  - HepG2
  - stomach
  - heart
  - thymus
  - K562
  - H1-hESC

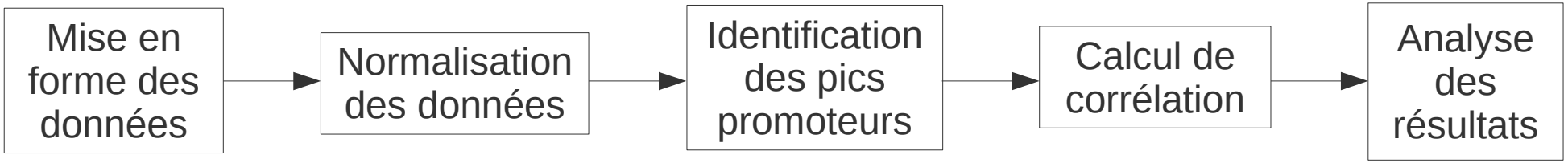
Ouverture de la chromatine

Ouverture de la chromatine



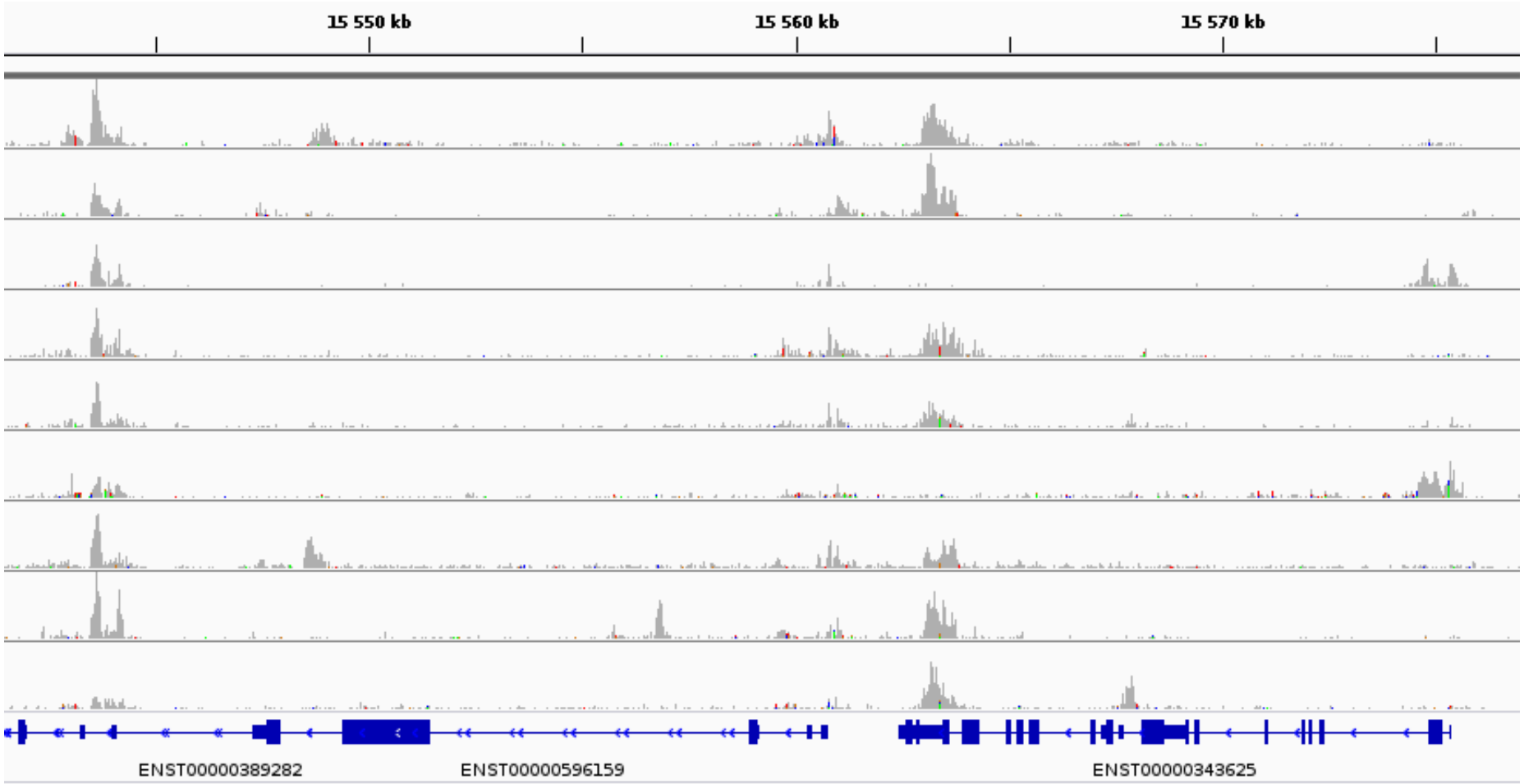
Ouverture de la chromatine



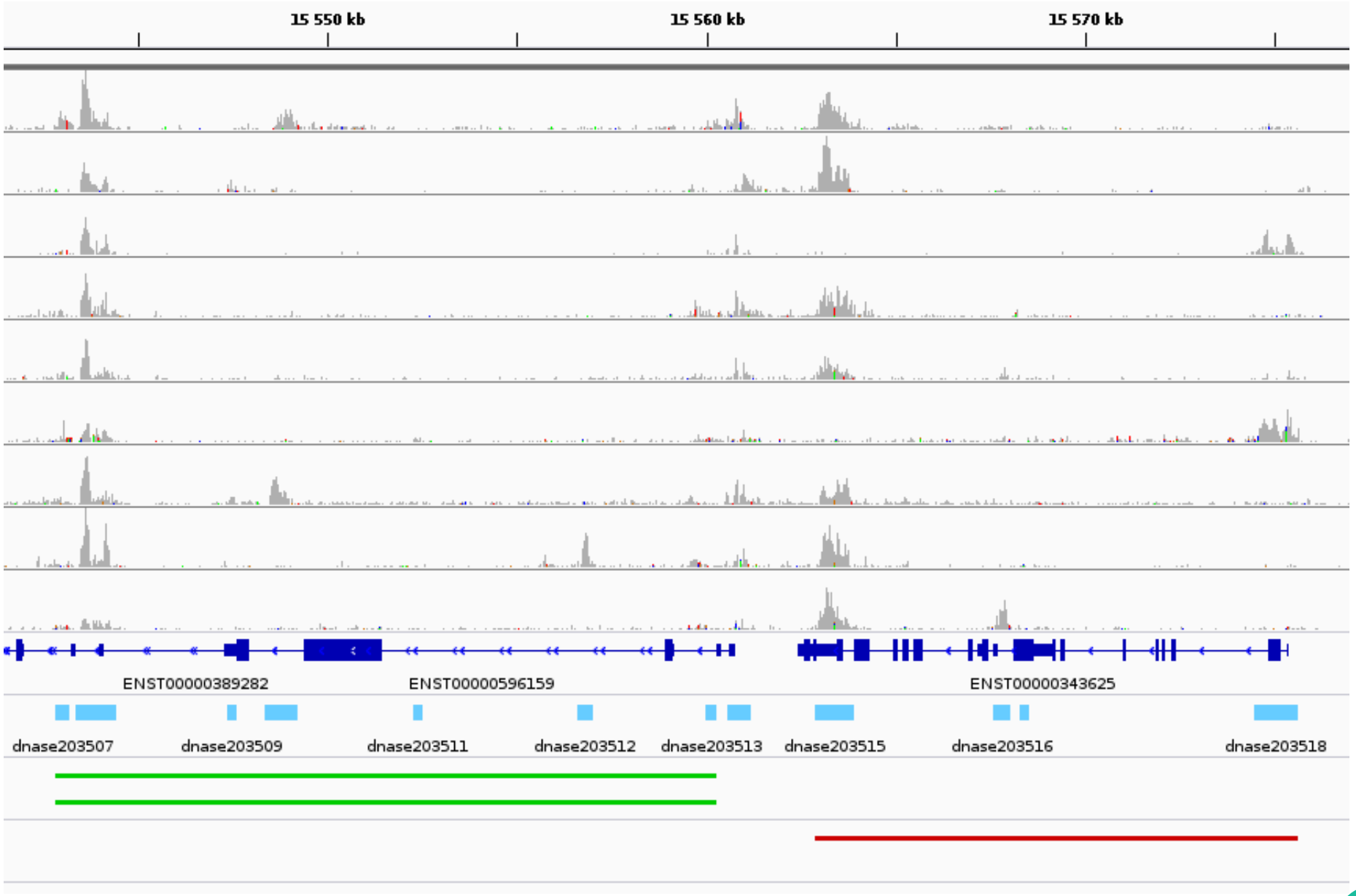


Robert E Thurman et al. Nature, 489(7414):75, 2012.

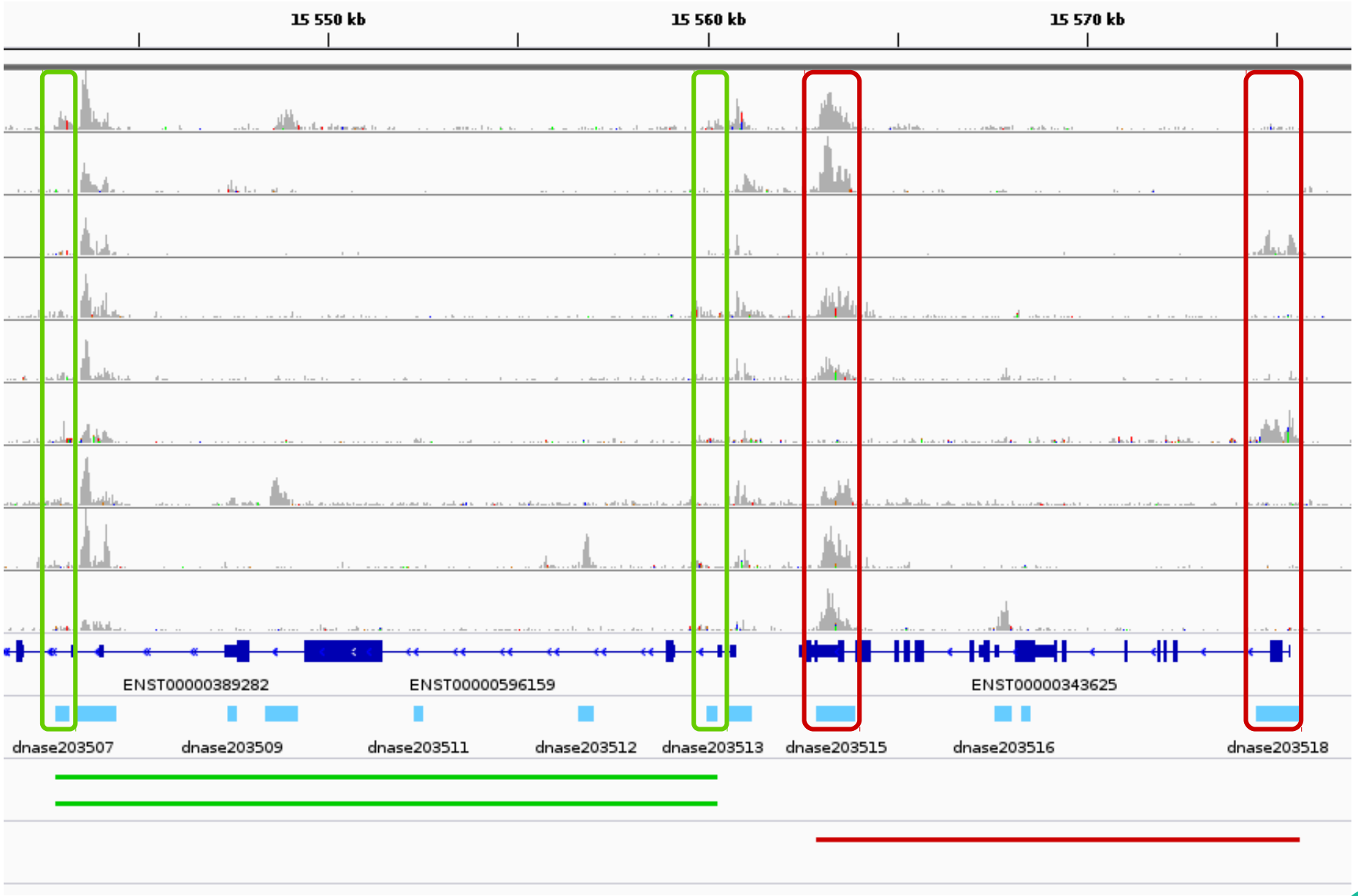
Ouverture de la chromatine



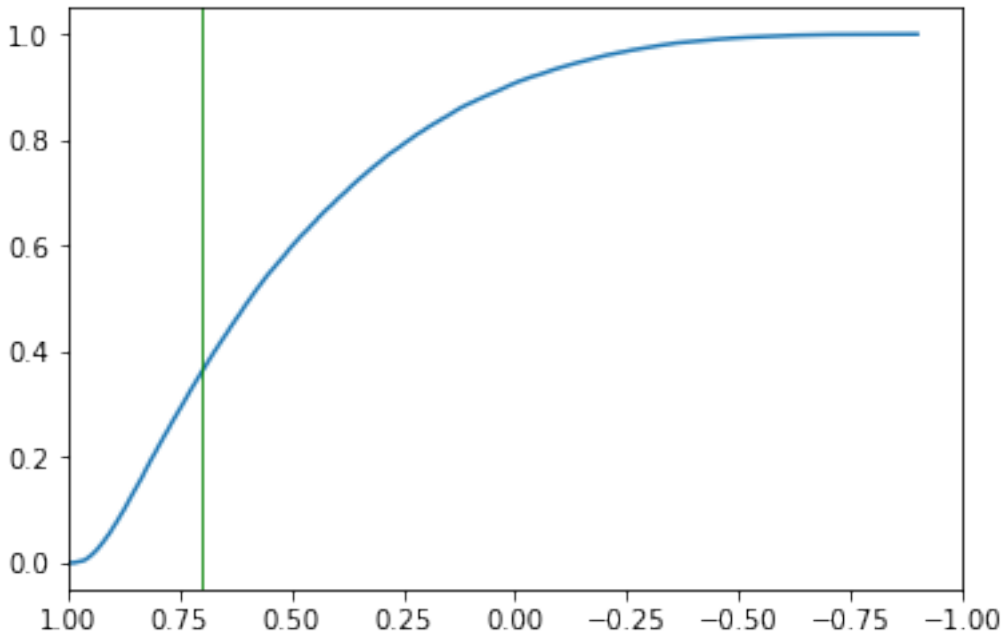
Ouverture de la chromatine

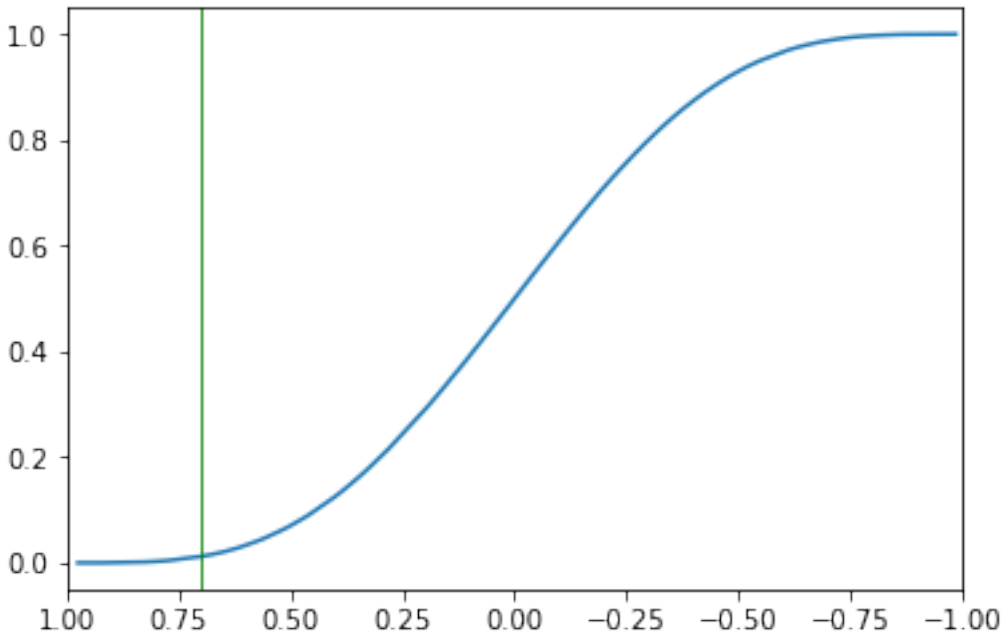


Ouverture de la chromatine

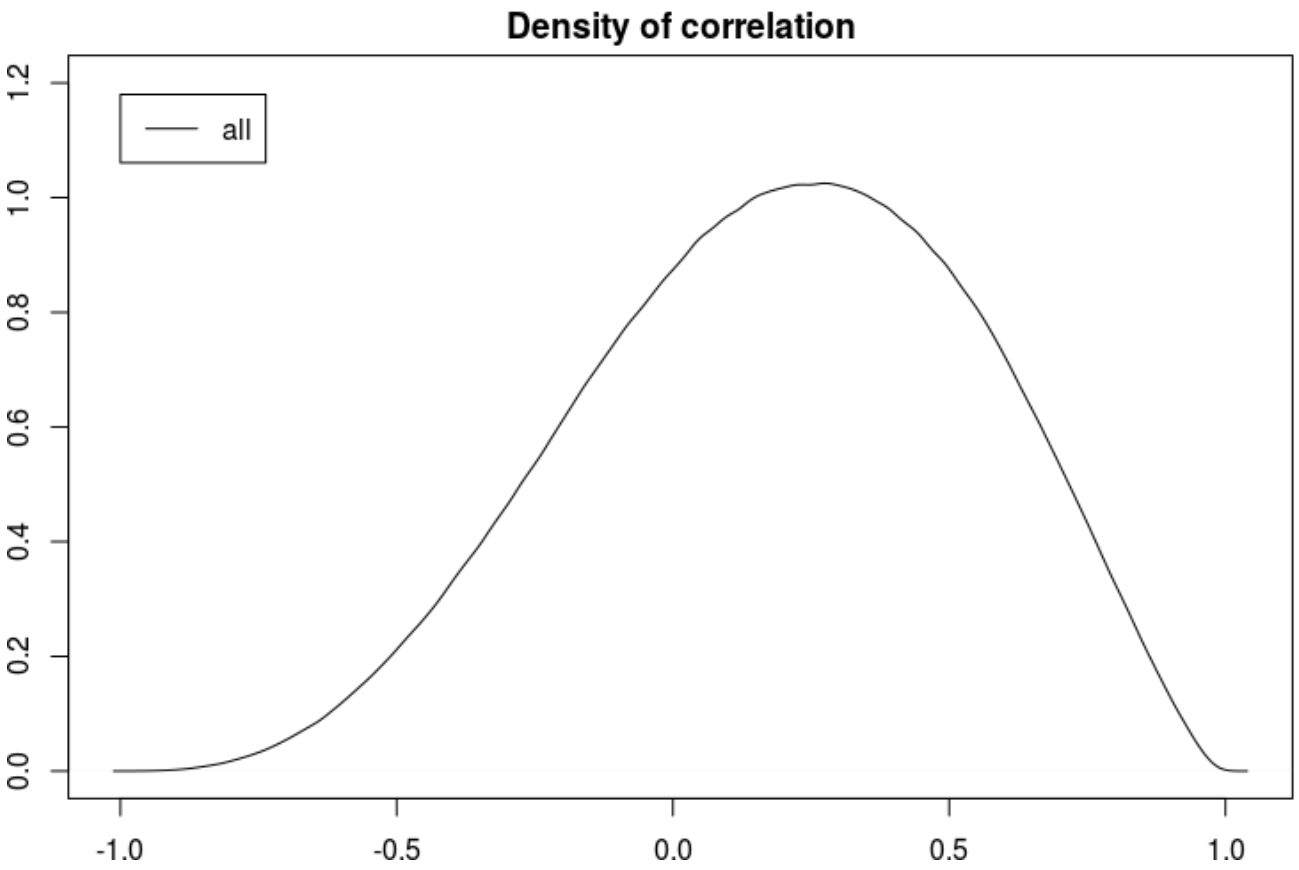


Ouverture de la chromatine

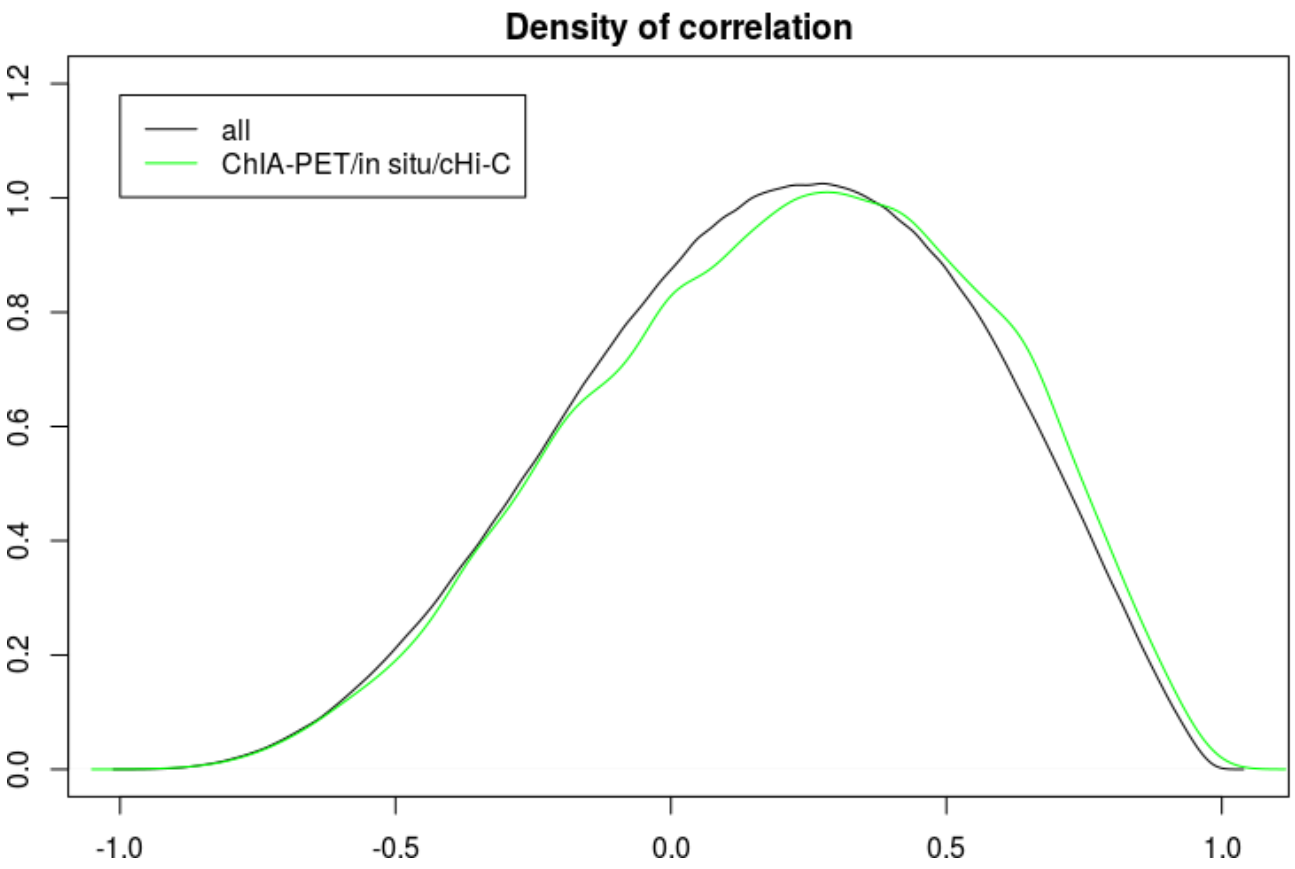




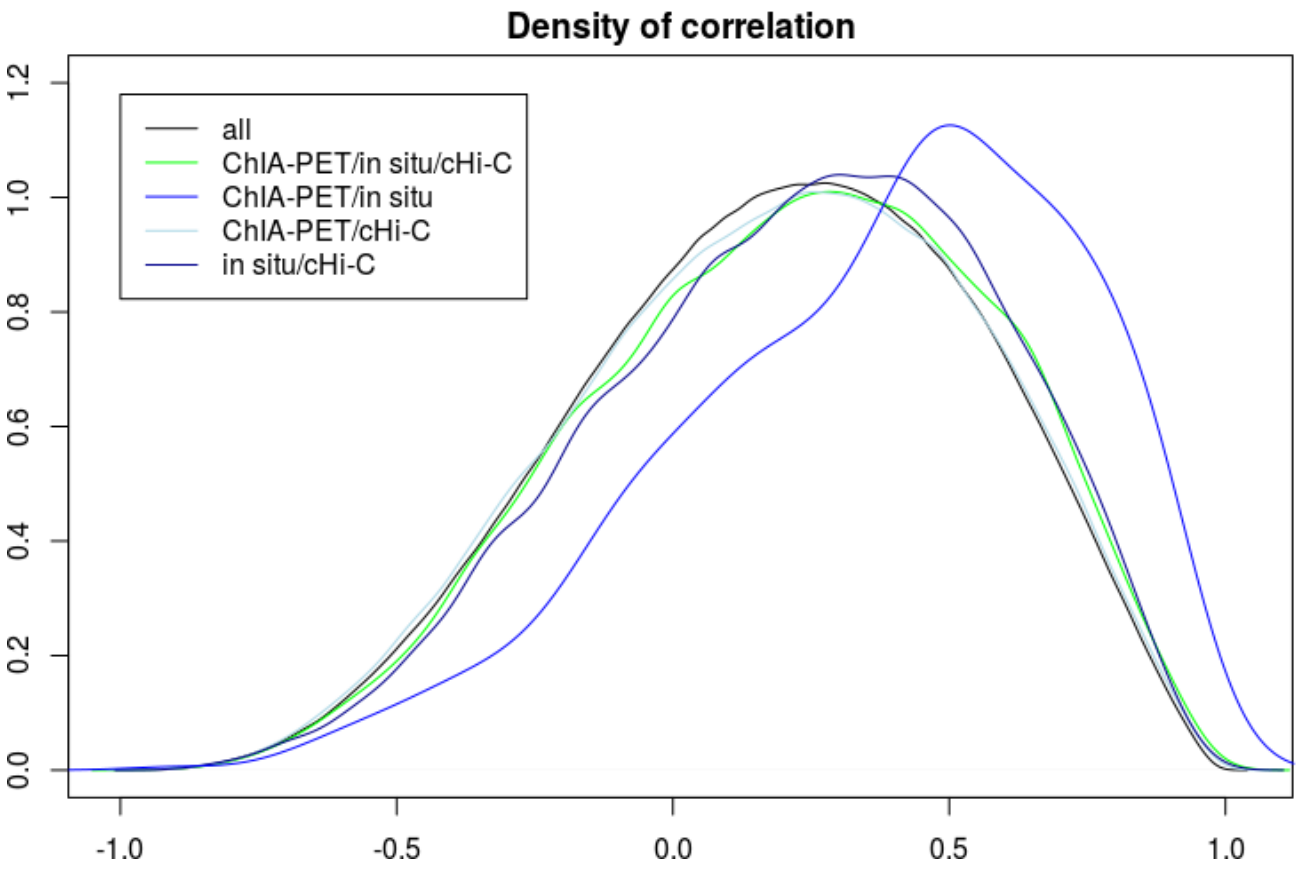
Ouverture de la chromatine



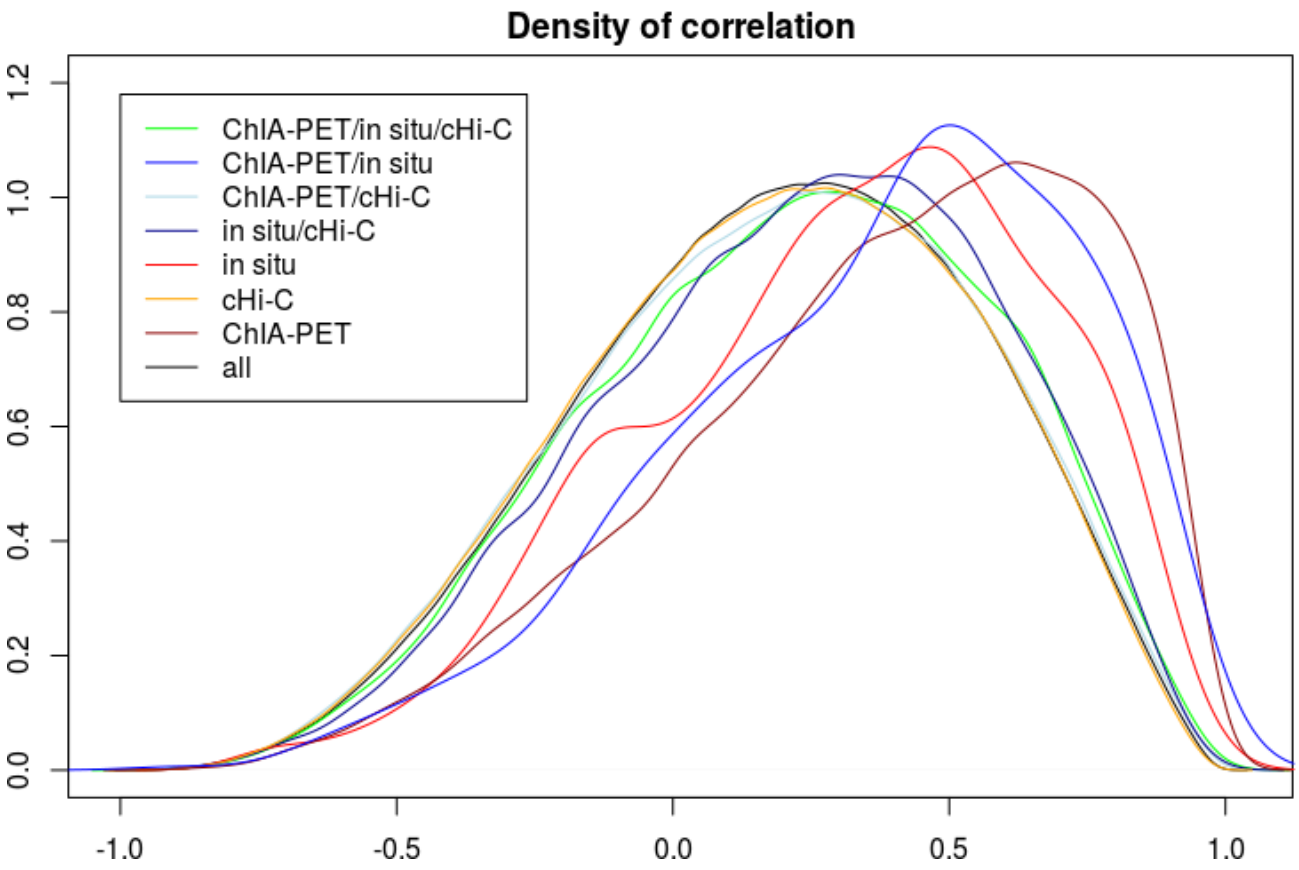
Ouverture de la chromatine



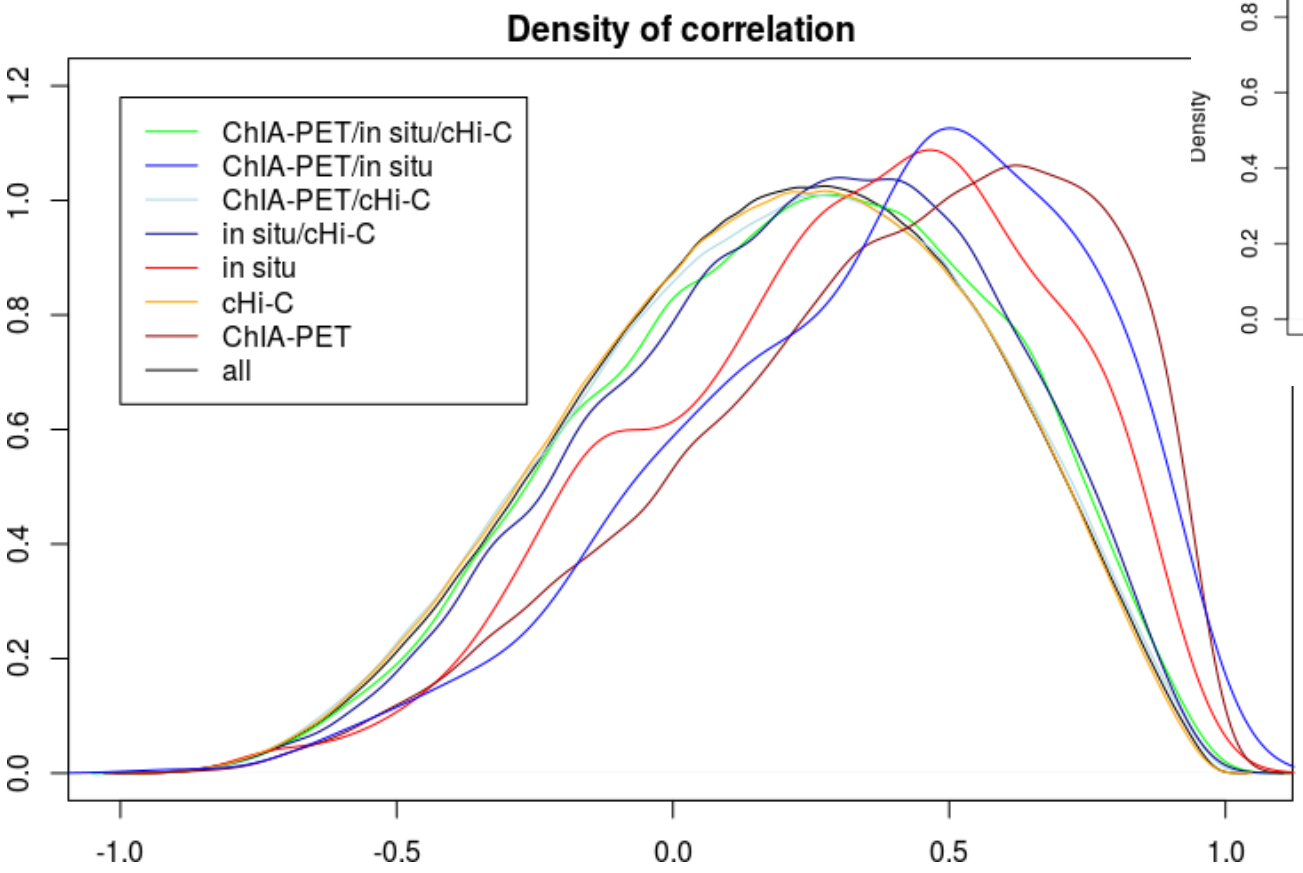
Ouverture de la chromatine

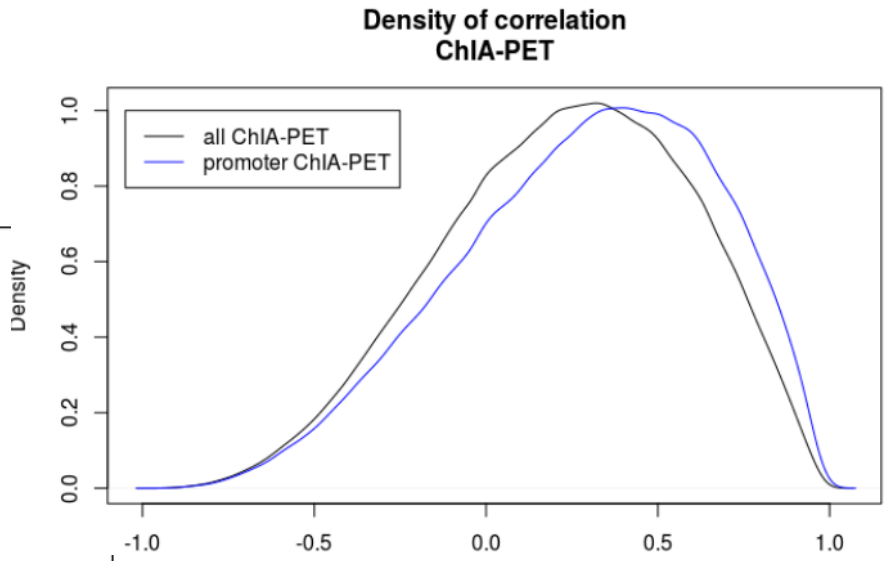


Ouverture de la chromatine

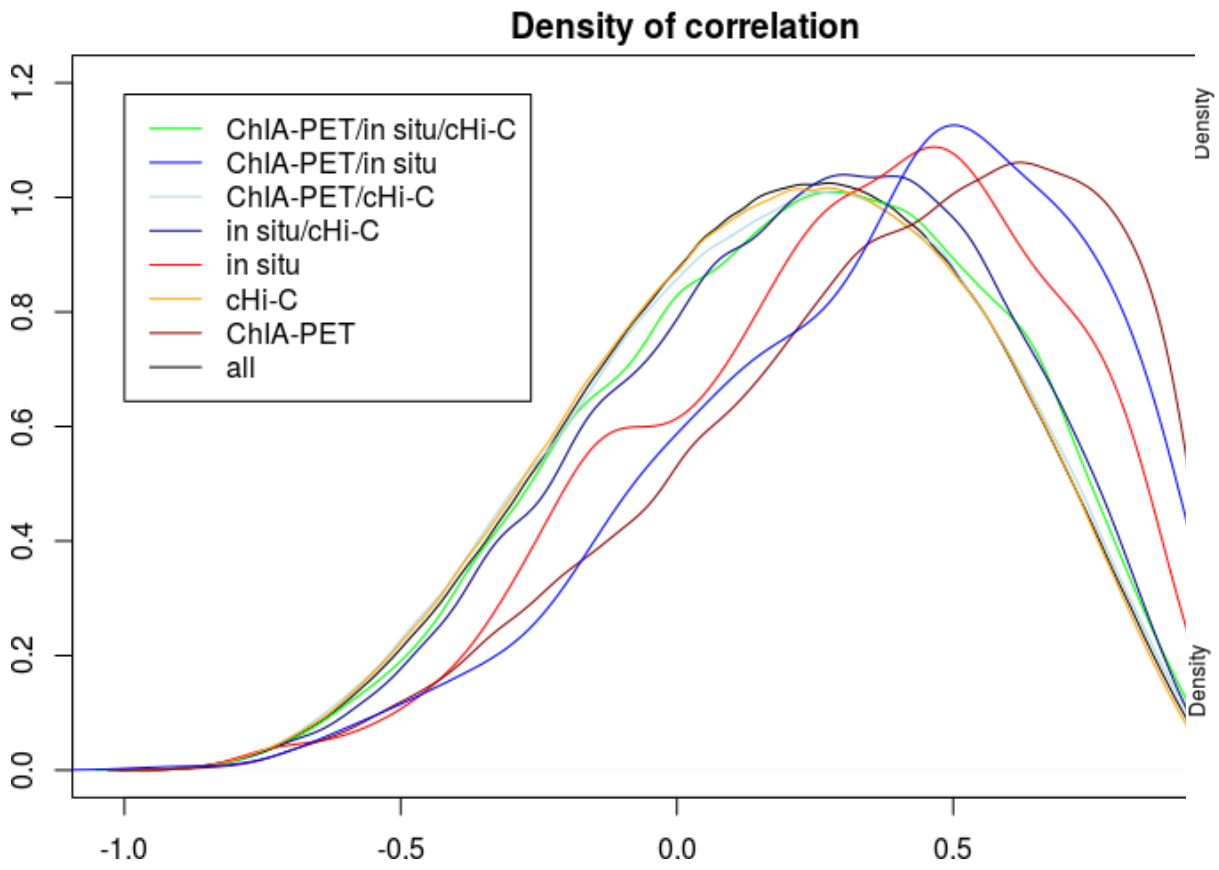


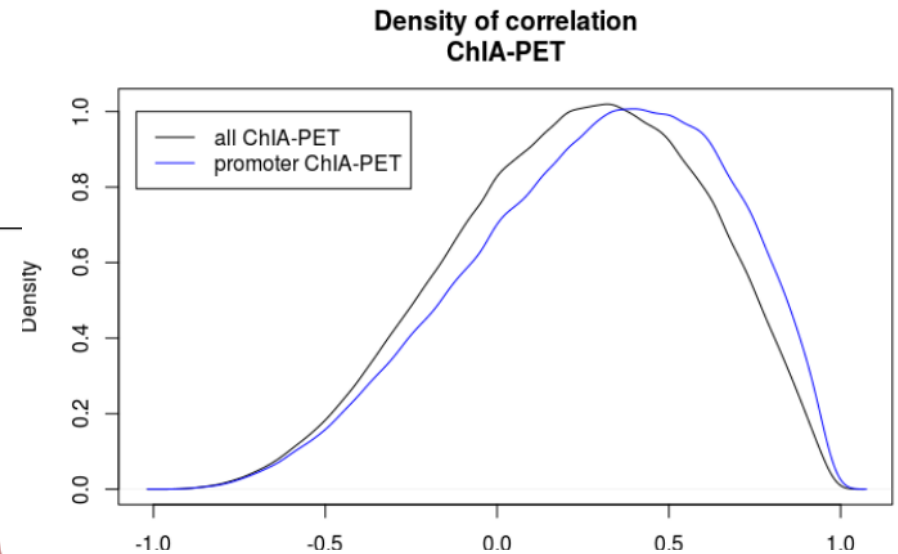
Ouverture de la chromatine

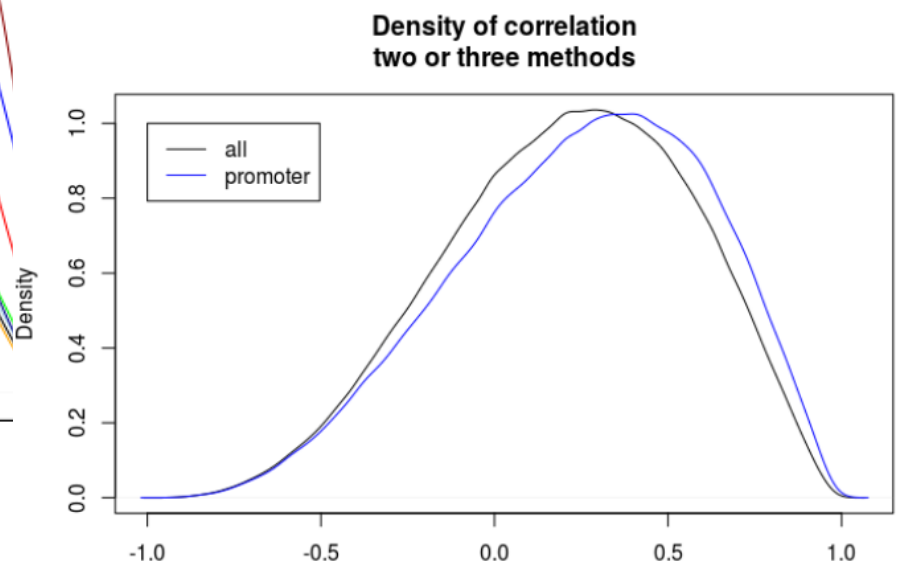




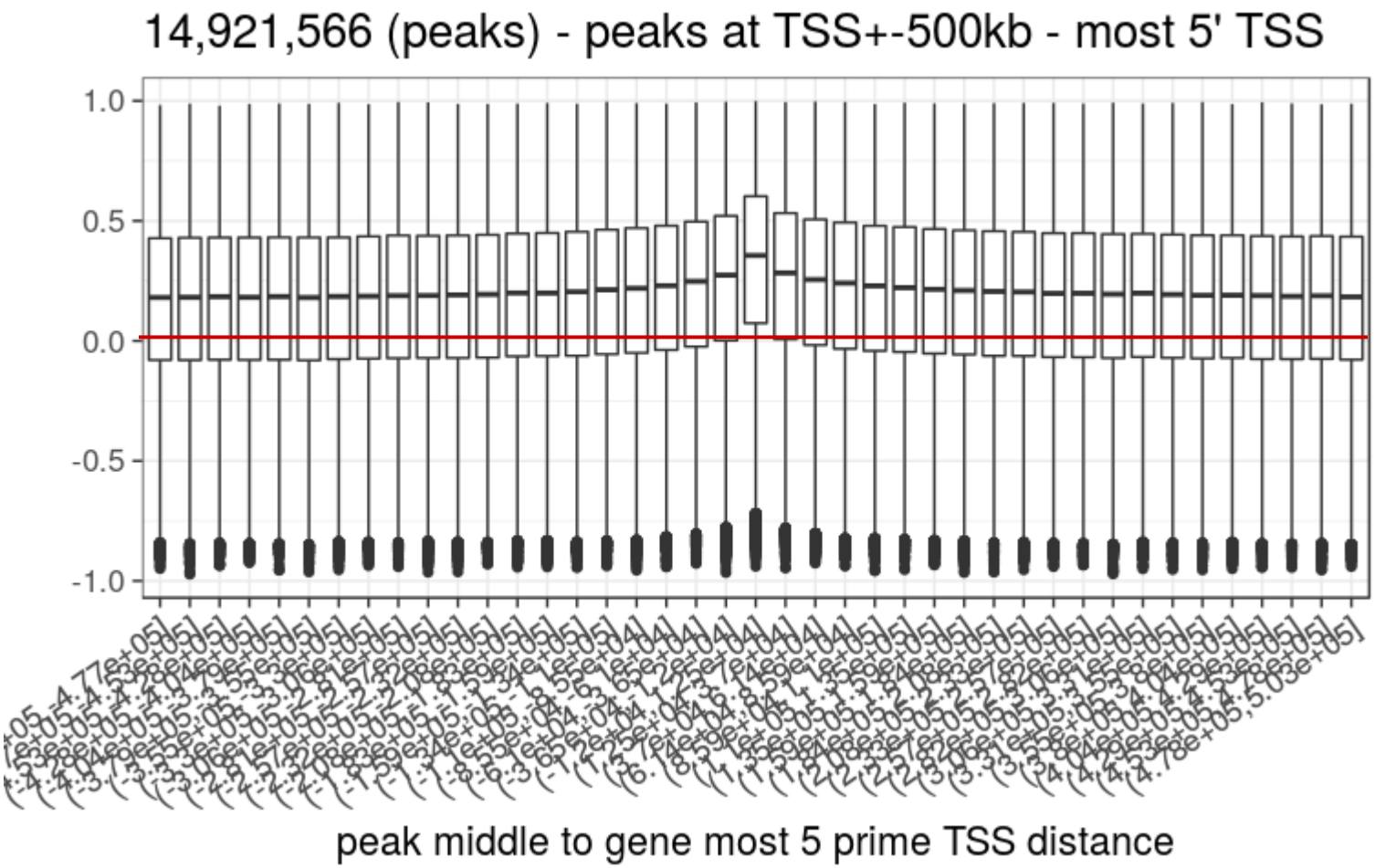
Ouverture de la chromatine







Ouverture de la chromatine



Plan

I. Contexte et définitions

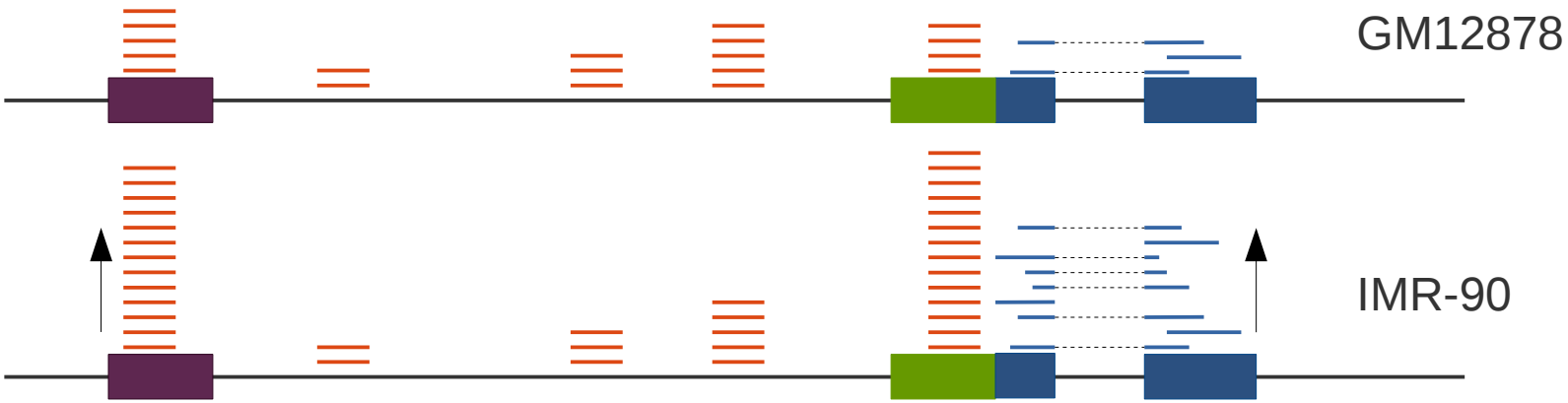
II. Interactions spatiales

III. Ouverture de la chromatine

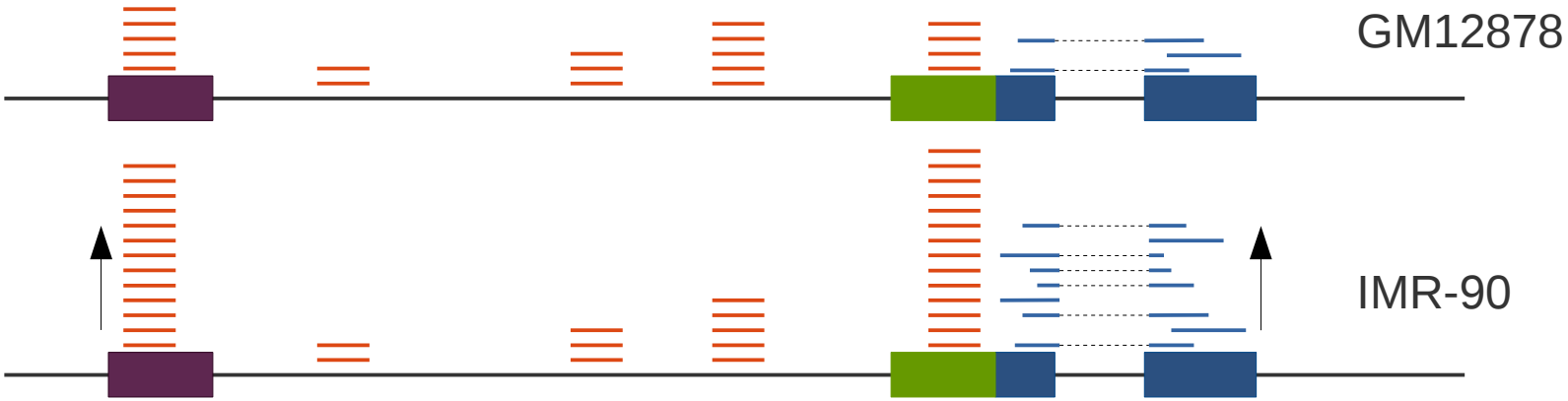
IV. Expression des gènes associée à l'ouverture de la chromatine

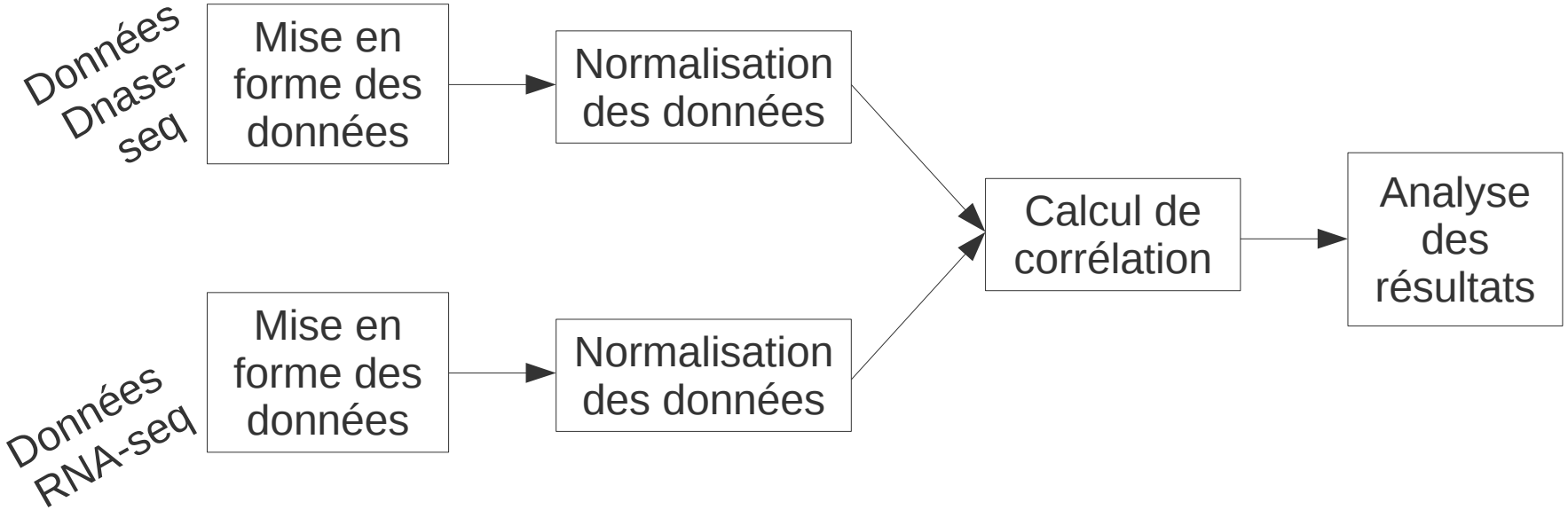
V. Conclusions et discussion

Expression des gènes associée à l'ouverture de la chromatine



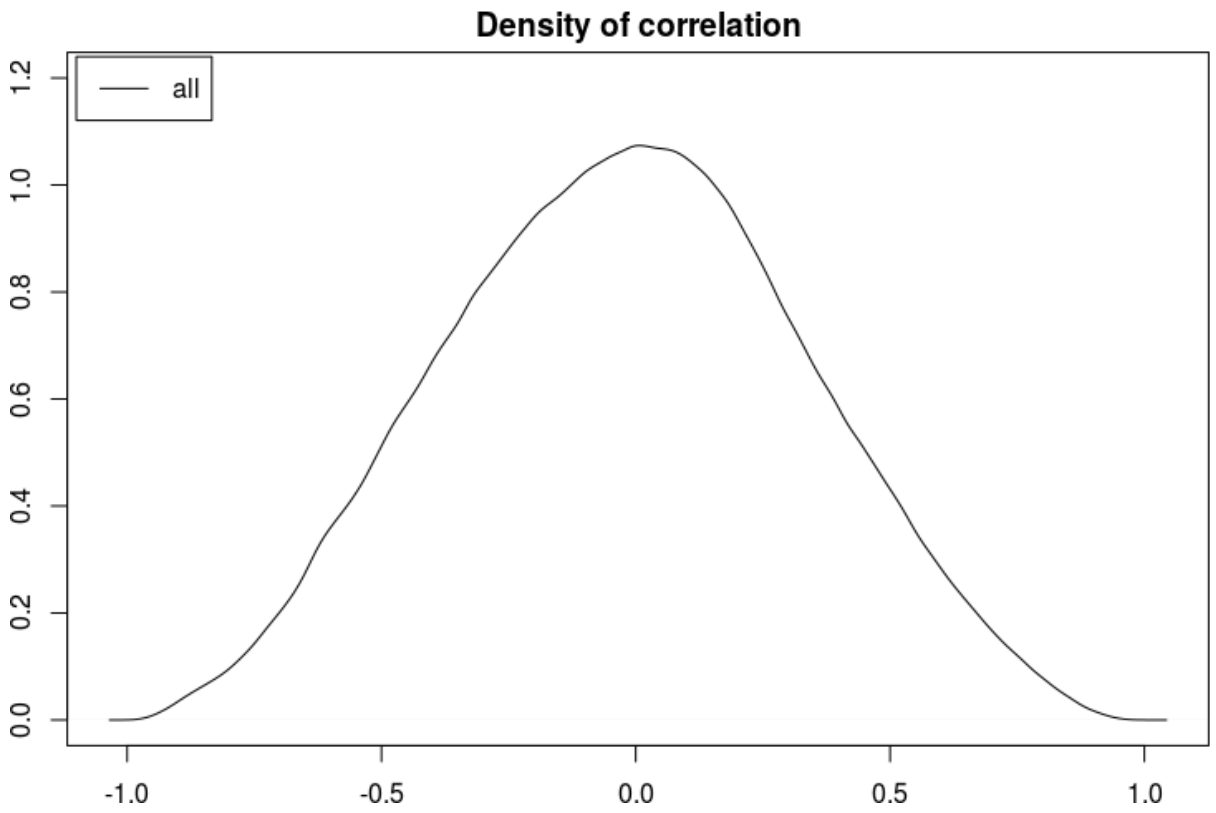
Expression des gènes associée à l'ouverture de la chromatine



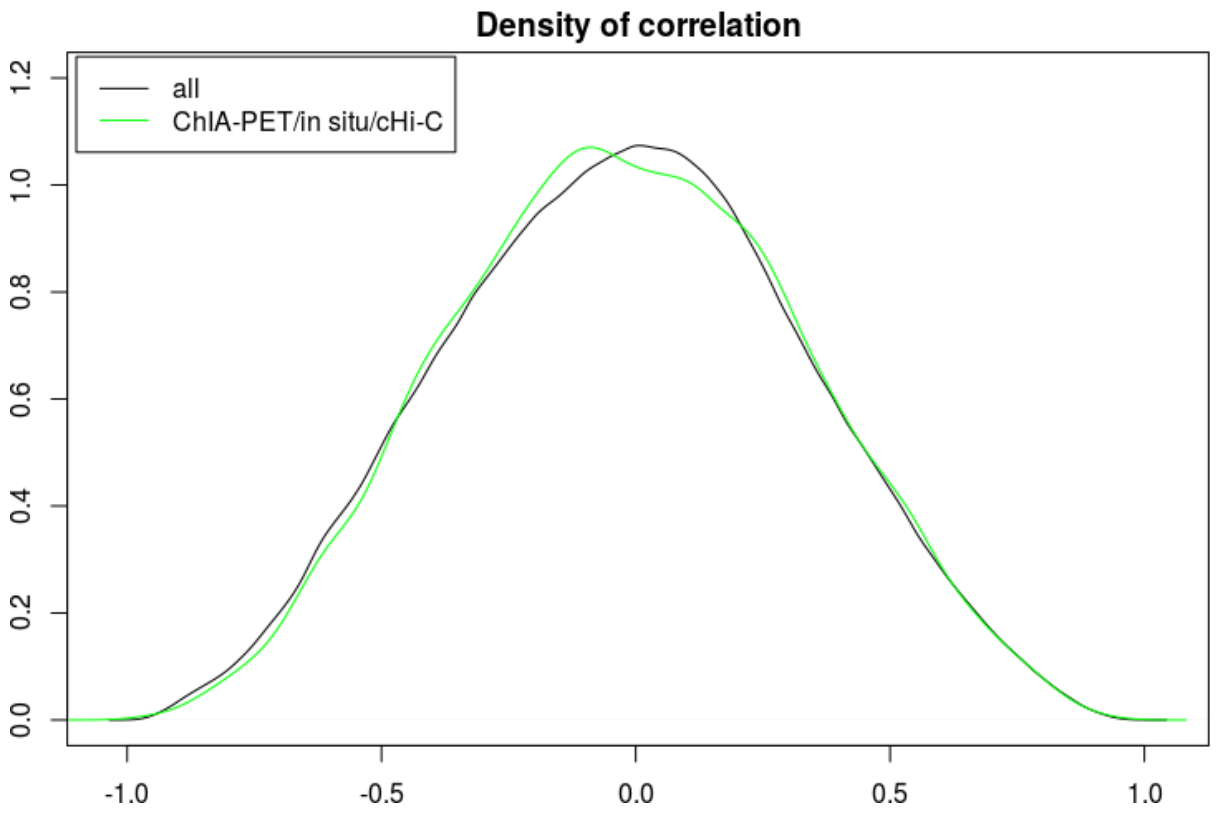


Nathan C. Sheffield et al., Genome research 23.5 (2013)

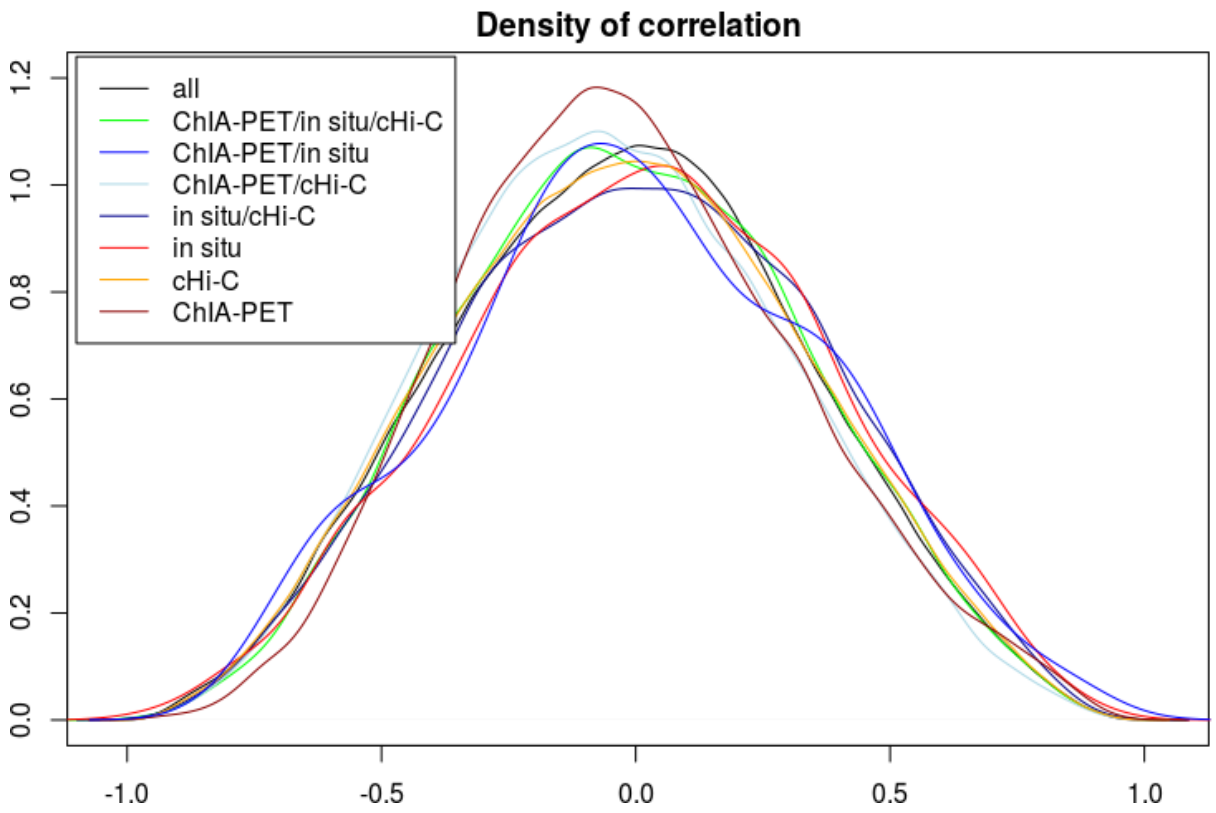
Expression des gènes associée à l'ouverture de la chromatine



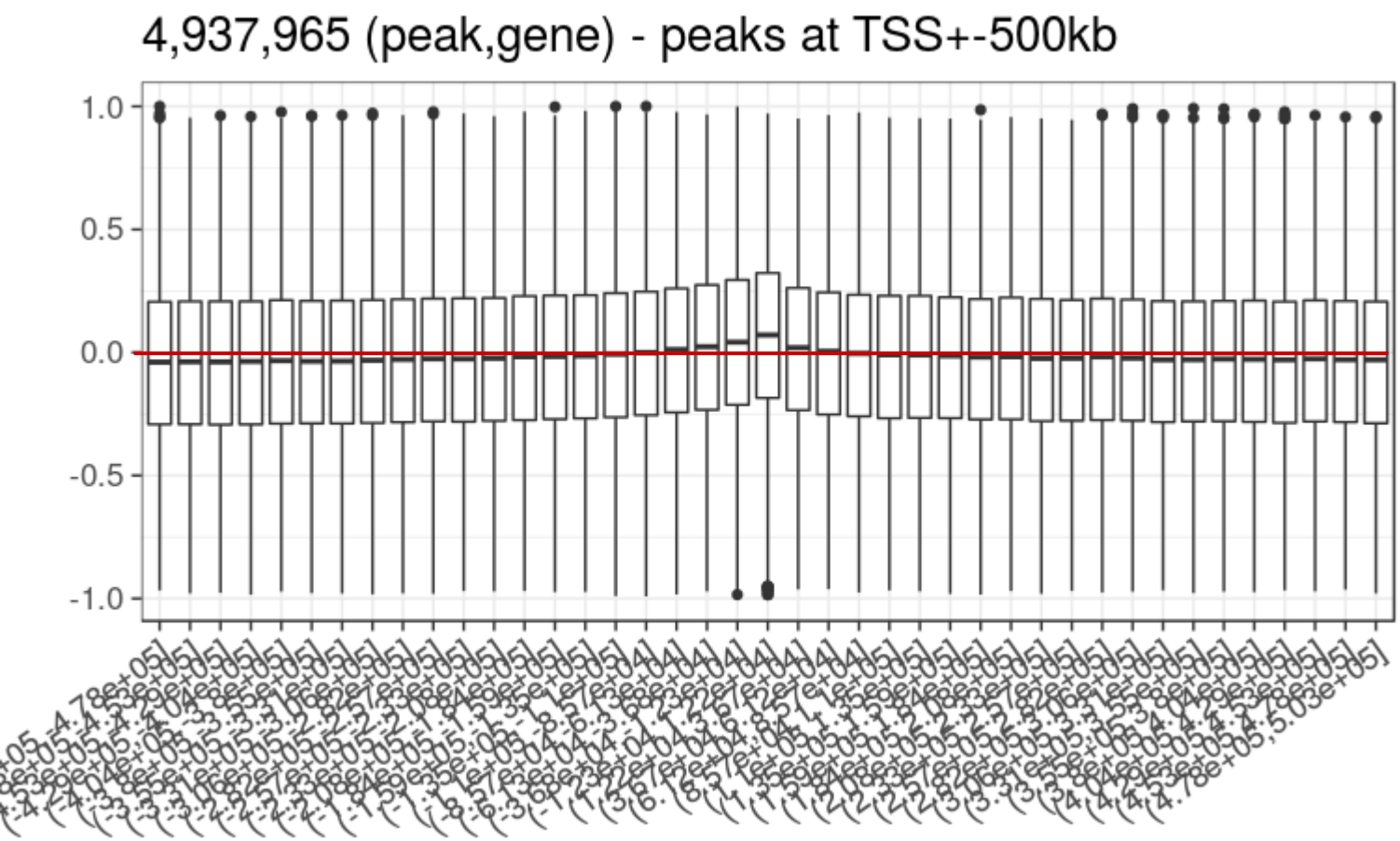
Expression des gènes associée à l'ouverture de la chromatine



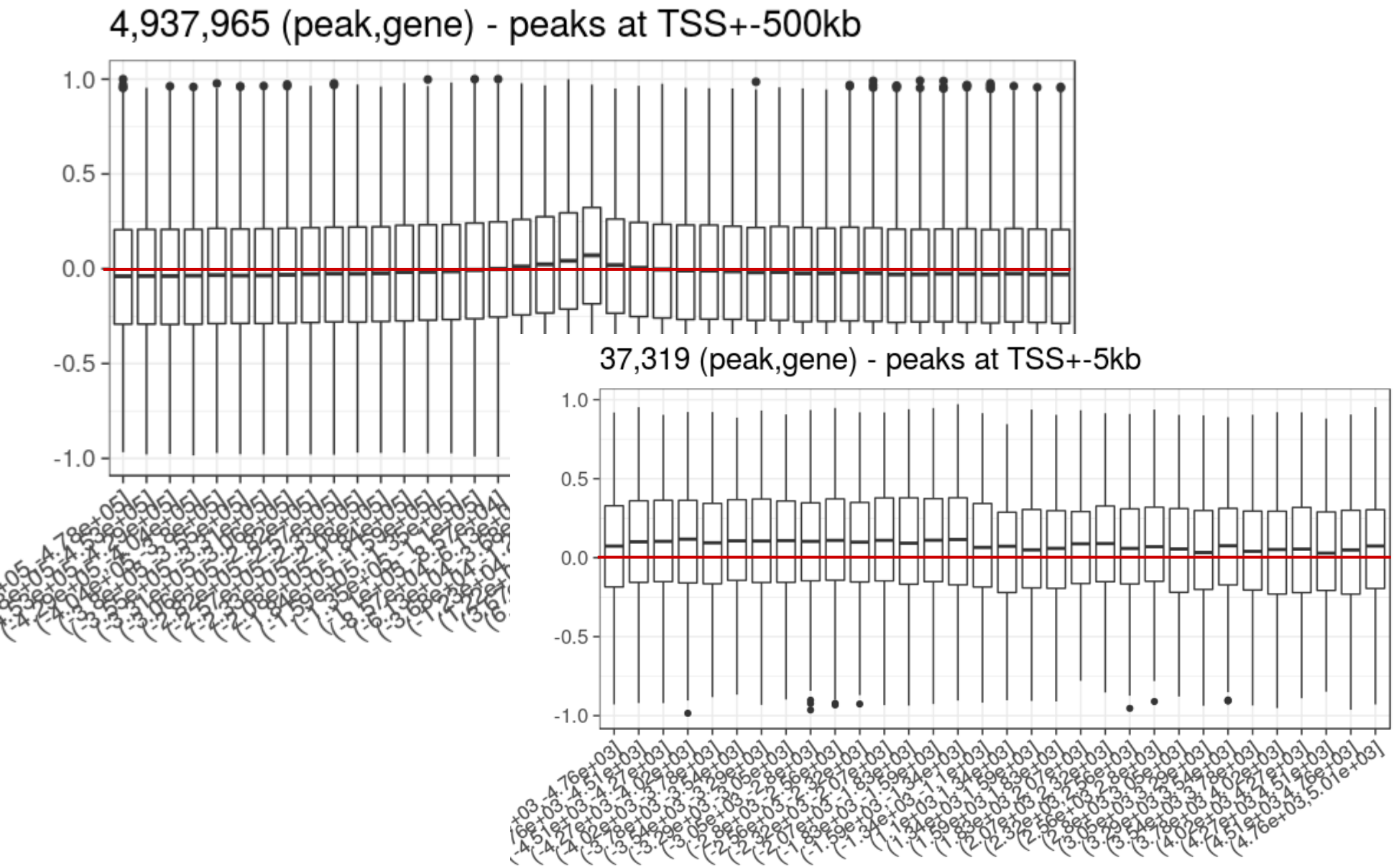
Expression des gènes associée à l'ouverture de la chromatine



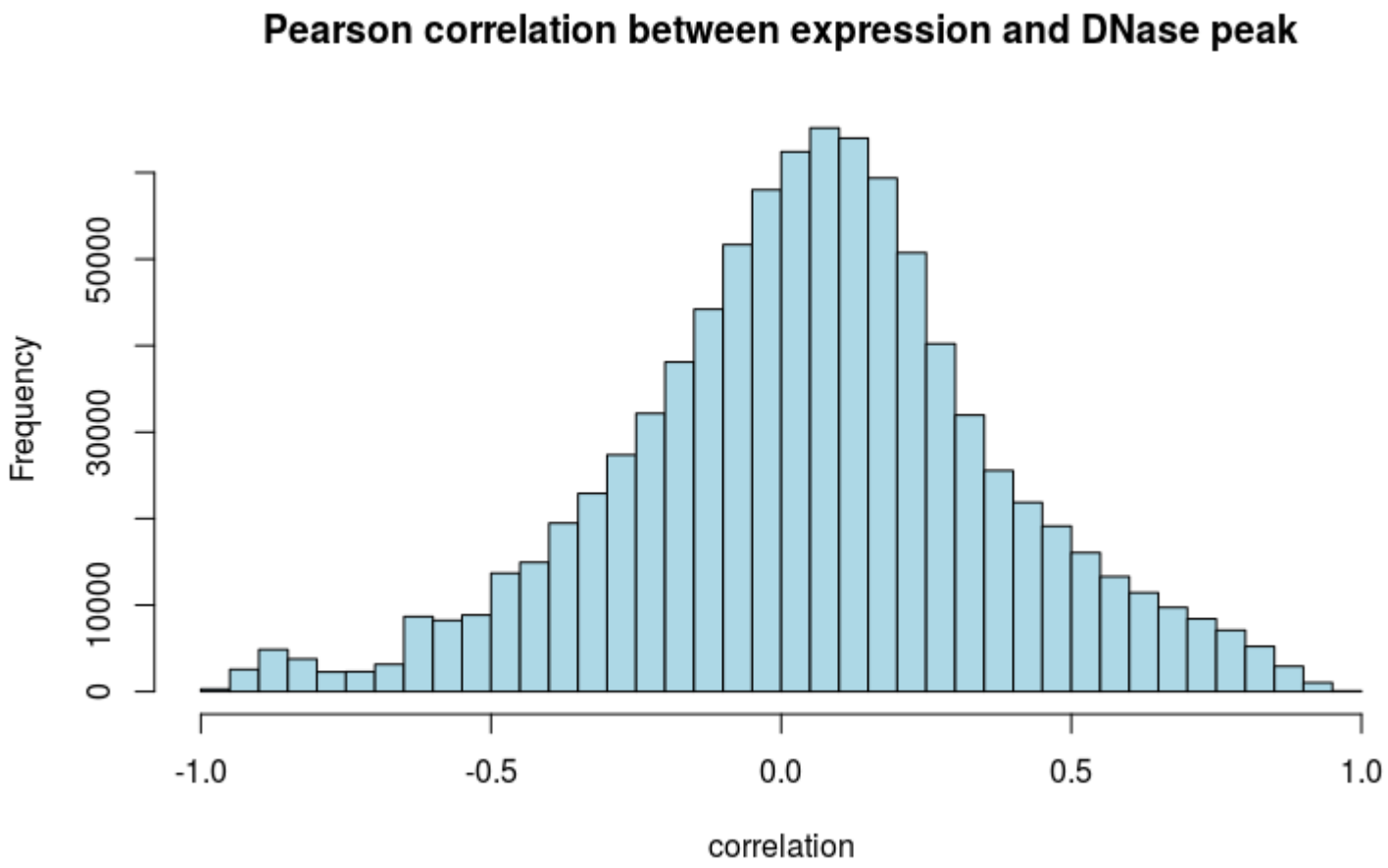
Expression des gènes associée à l'ouverture de la chromatine



Expression des gènes associée à l'ouverture de la chromatine

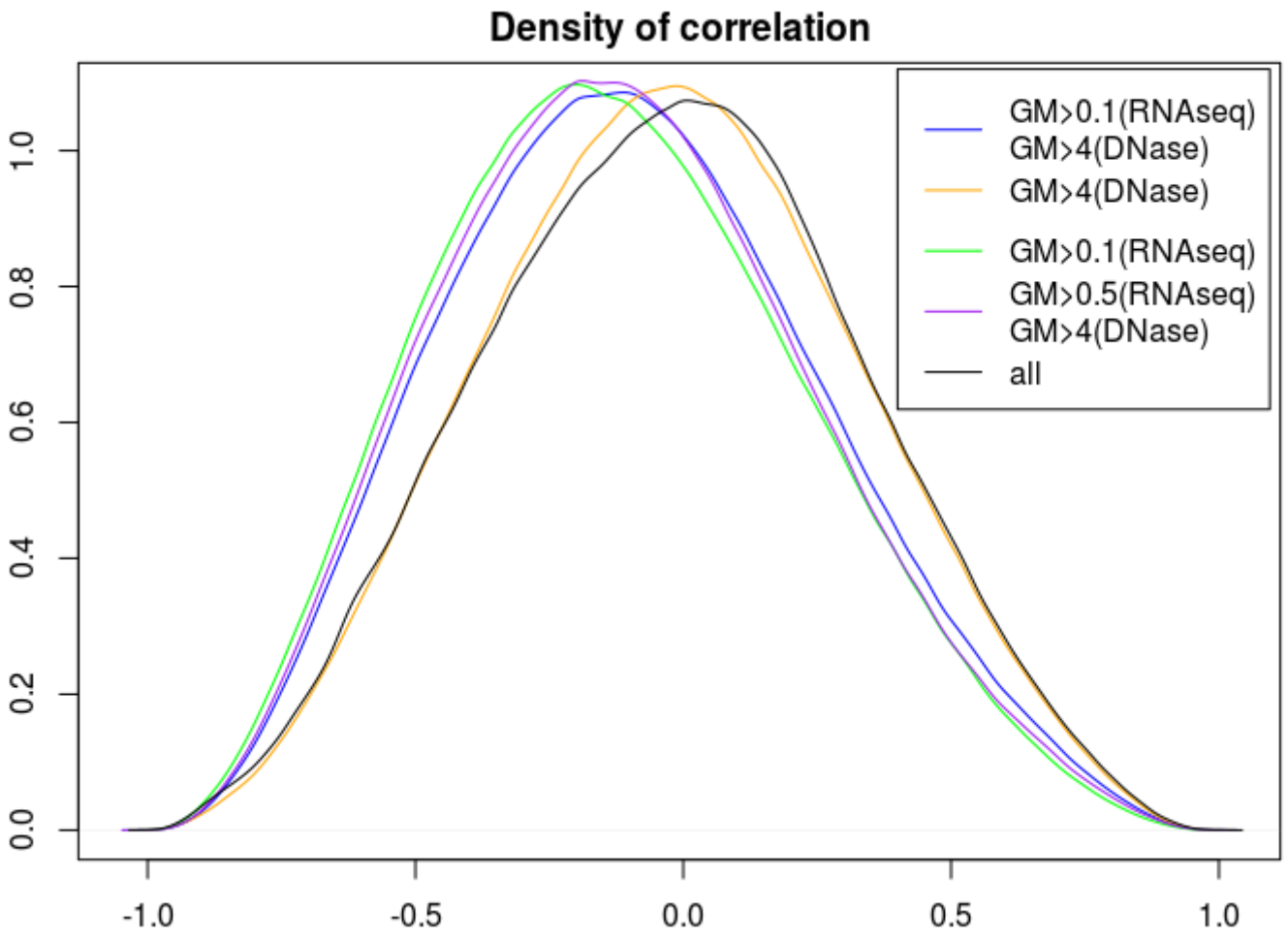


Expression des gènes associée à l'ouverture de la chromatine



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
-0.99130	-0.14120	0.05920	0.05066	0.24290	1.00000

Expression des gènes associée à l'ouverture de la chromatine



Plan

I. Contexte et définitions

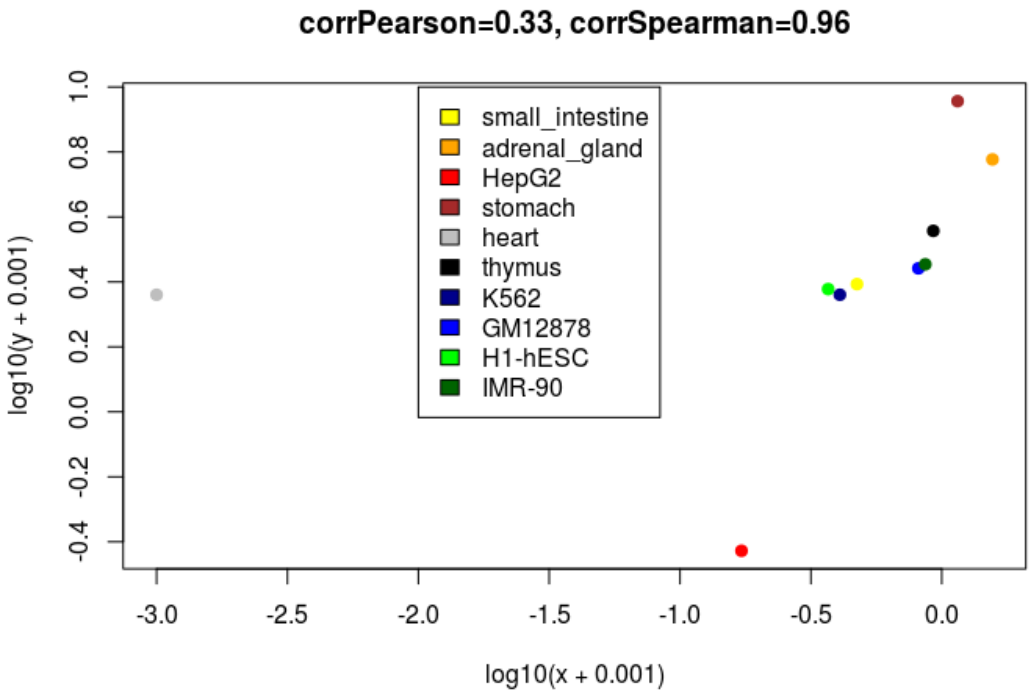
II. Interactions spatiales

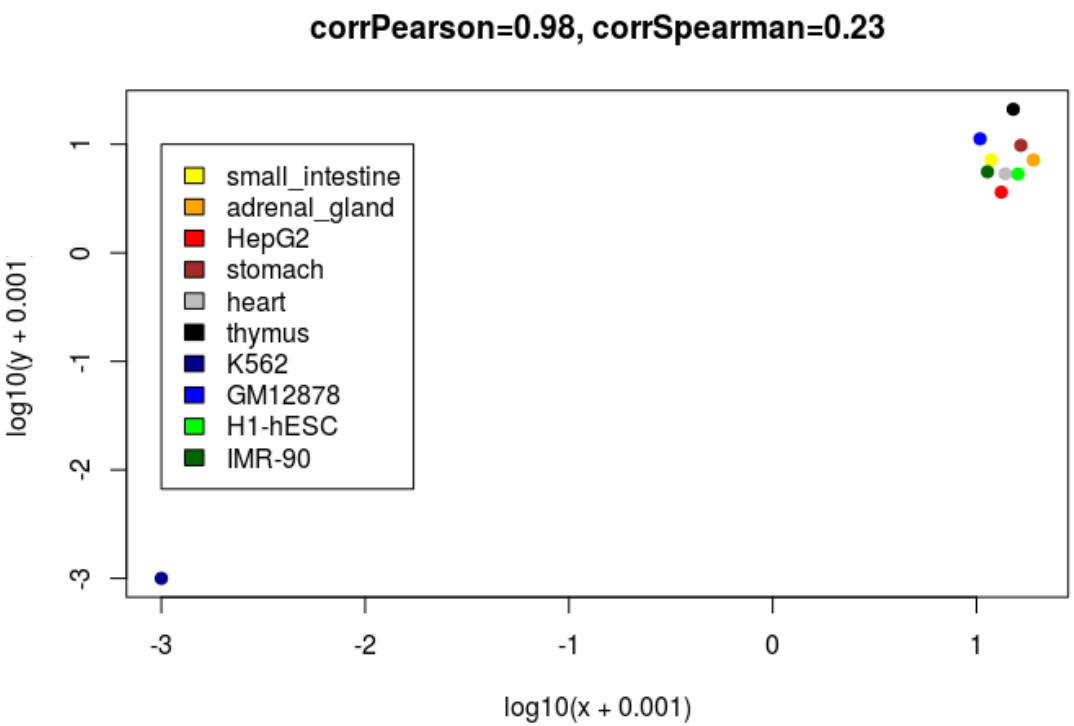
III. Ouverture de la chromatine

IV. Expression des gènes associée à l'ouverture de la chromatine

V. Conclusions et discussion

Conclusions et discussion

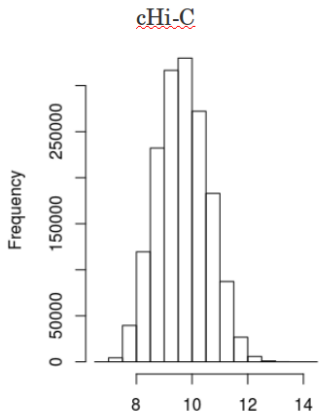


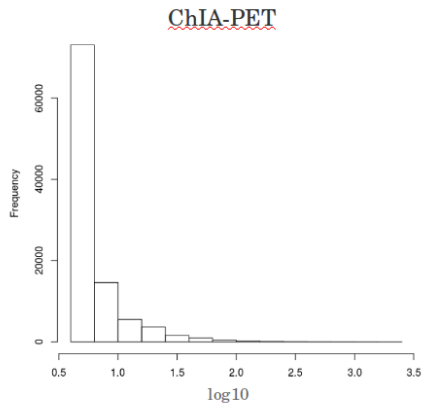


Merci de votre attention

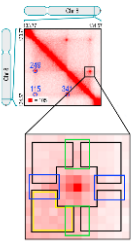
Mestre Camille

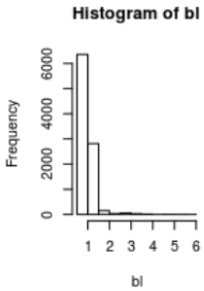
Interactions spatiales

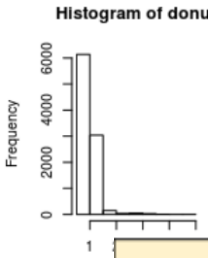


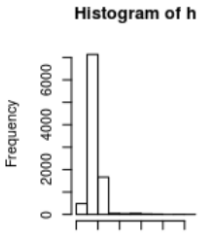


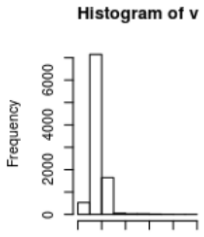
In situ Hi-C



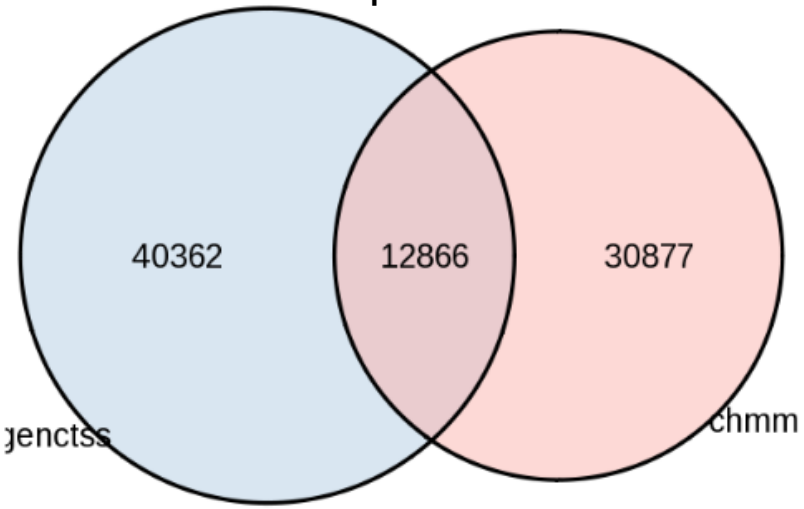






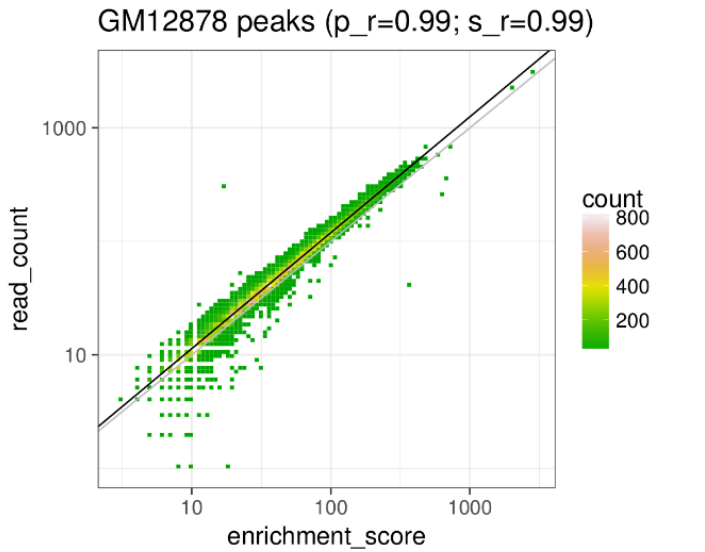


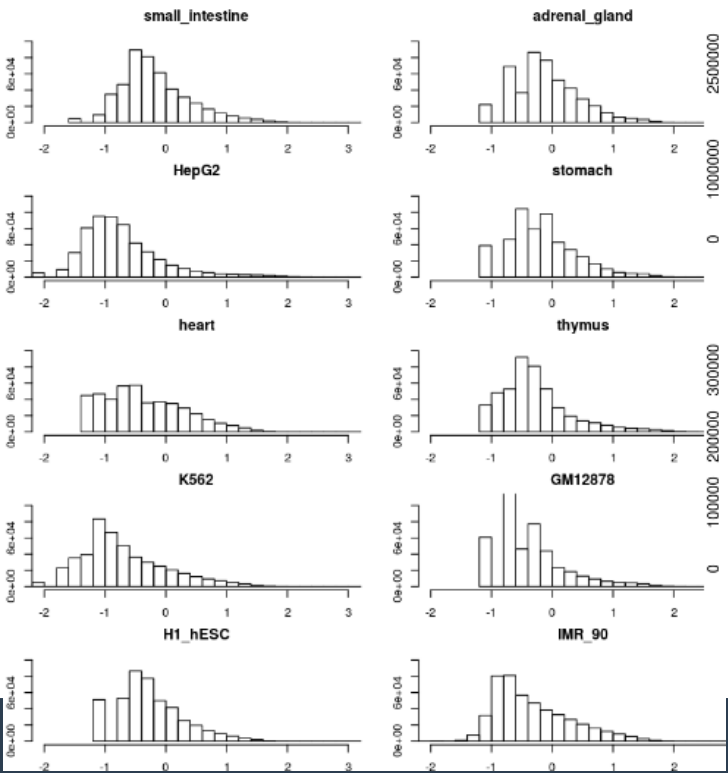
Comparing gencode tss+-1kb with chromHMM promoter states



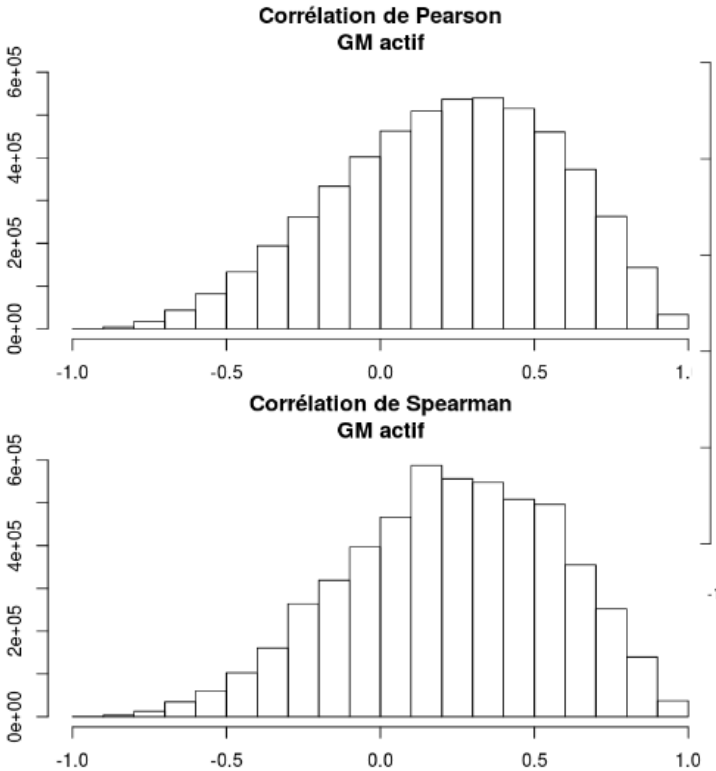
Connection set	# connections	with 1 part overlapping a promoter state from the segmentation		with 1 part overlapping an enhancer state from the segmentation		with one part overlapping a promoter state of the segmentation and the other part overlapping an enhancer state of the segmentation	
		#	% of total	#	% of total	#	% of total
chiapet	113,572	61,674	54.3	70,212	61.8	35,411	31.2
chic	739,520	410,323	55.5	458,087	61.9	153,130	20.7
insitu	9,448	3,844	40.7	6,479	68.6	2,371	25.1
chiapet.only	91,489	44,104	48.2	53,826	58.8	23,783	26.0
chic.only	673,487	368,068	54.7	412,418	61.2	134,313	19.9
insitu.only	180	38	21.1	107	59.4	0	0.0
chiapet.chic	9,222	8,882	96.3	9,048	98.1	8,039	**87.2**
chiapet.insitu	753	396	52.6	646	85.8	320	**42.5**
chic.insitu	1,712	1,353	79.0	1,493	87.2	982	**57.4**
chiapet.chic.insitu	1,040	988	95.0	1,034	99.4	929	**89.3**

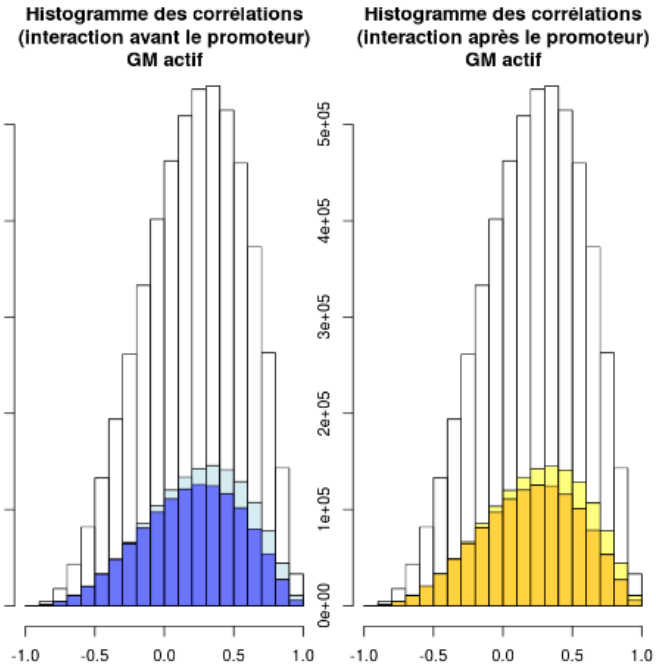
Ouverture de la chromatine





Différence P et S





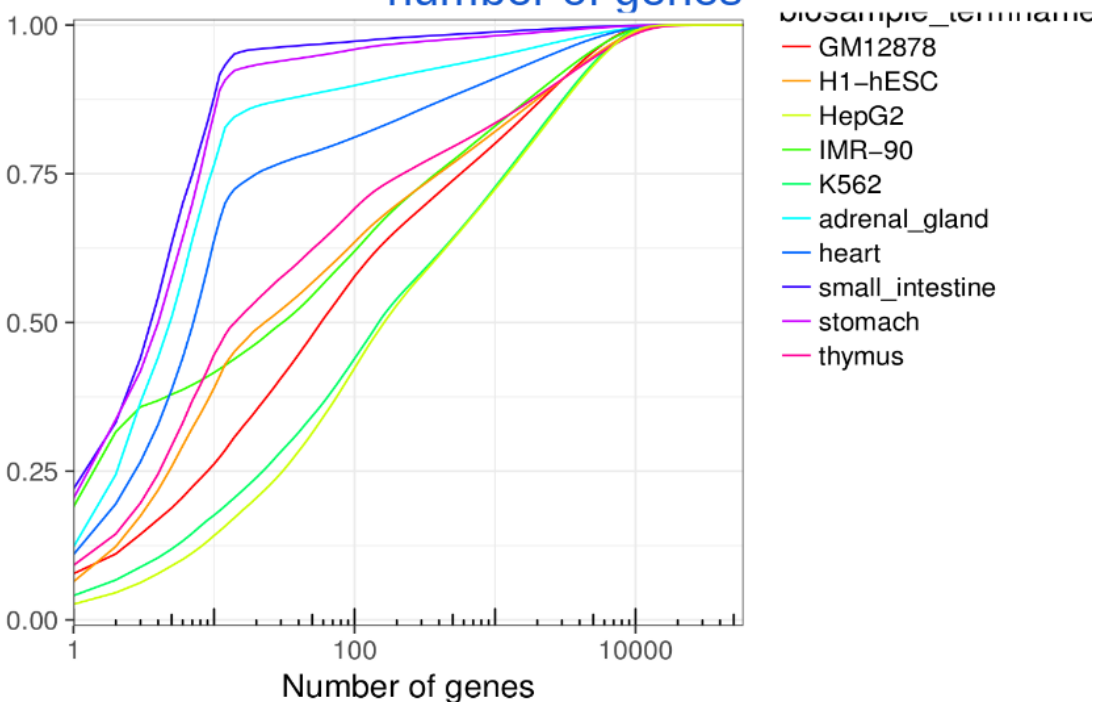
[-500kb ; -250kb] = blue
[-250kb ; promoter] = light blue
[promoter ; +250kb] = yellow
[+250kb ; +500kb] = orange

Histogramme avecla corrélation de Pearson au-dessus

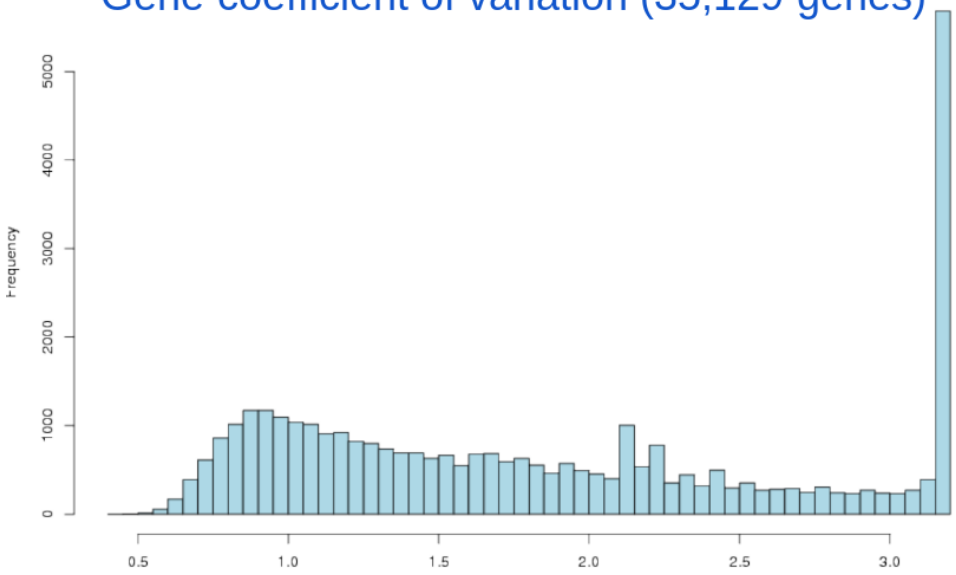
Effet distance

Expression des gènes associée à l'ouverture de la chromatine

Fraction of the total gene TPM captured by each number of genes

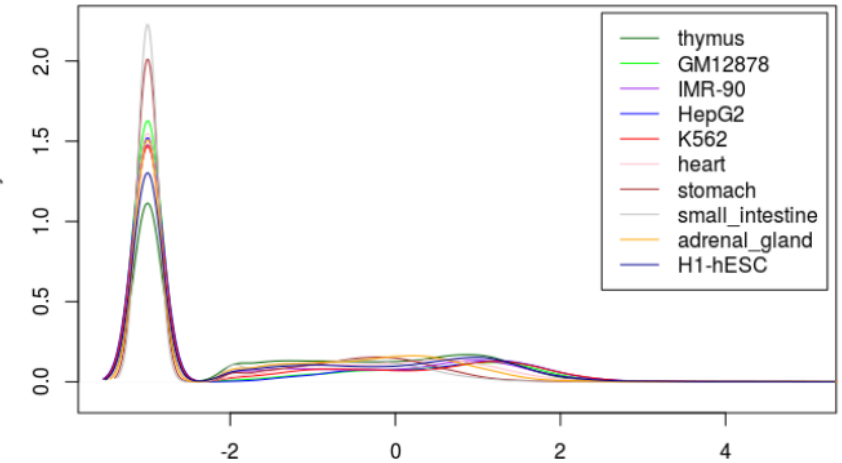


Gene coefficient of variation (35,129 genes)



Écart-type/moyenne

Density of expression per tissu (log10)



pseudocount = log10(count+0.001)

La suite du stage

DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes, Aran et al., 2013

Model for the relationship promoter methylation/gene expression -> machine learning algorithm (SVM-MAP code)

→

Apply the derived model to methylation sites away from gene promoters (1Mb upstream of the TSS and 1Mb downstream of the transcription end site) Characterizes chromatin structure and binding profile of discovered sites

→

Comparison with a condition (here cancer)

Use of CpG methylation sites for model developpement

"True" pairs with randomized pairs (interchromosomique) = inputs

Score for each pair: if score >0.85 => model successfully paired 87.2% of "true" pairs, null expectation: 50%.

- [-500;+2000] region from TSS considered for training
- Only used VMS (variable methylation sites) for the analysis
- A CpG site had to be covered 10 times in at least 60% (80%?) of the cell types

Validation by 5C
Code: http://projects.yisongyue.com/svmmap/