



HAL
open science

UF2, module 3 : Bio-informatique pour l'épigénome - Méthylation de l'ADN

Sonia E Eynard, Guillaume Devailly

► To cite this version:

Sonia E Eynard, Guillaume Devailly. UF2, module 3 : Bio-informatique pour l'épigénome - Méthylation de l'ADN. Master. Biologie Computationnelle pour les Biotechnologies, ENSAT Toulouse, France. 2023. hal-04791711

HAL Id: hal-04791711

<https://hal.inrae.fr/hal-04791711v1>

Submitted on 19 Nov 2024

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.



➤ UF2, module 3 : Bio-informatique pour l'épigénome

Guillaume Devailly, Sonia Eynard
sonia.eynard@inrae.fr

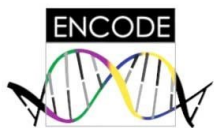
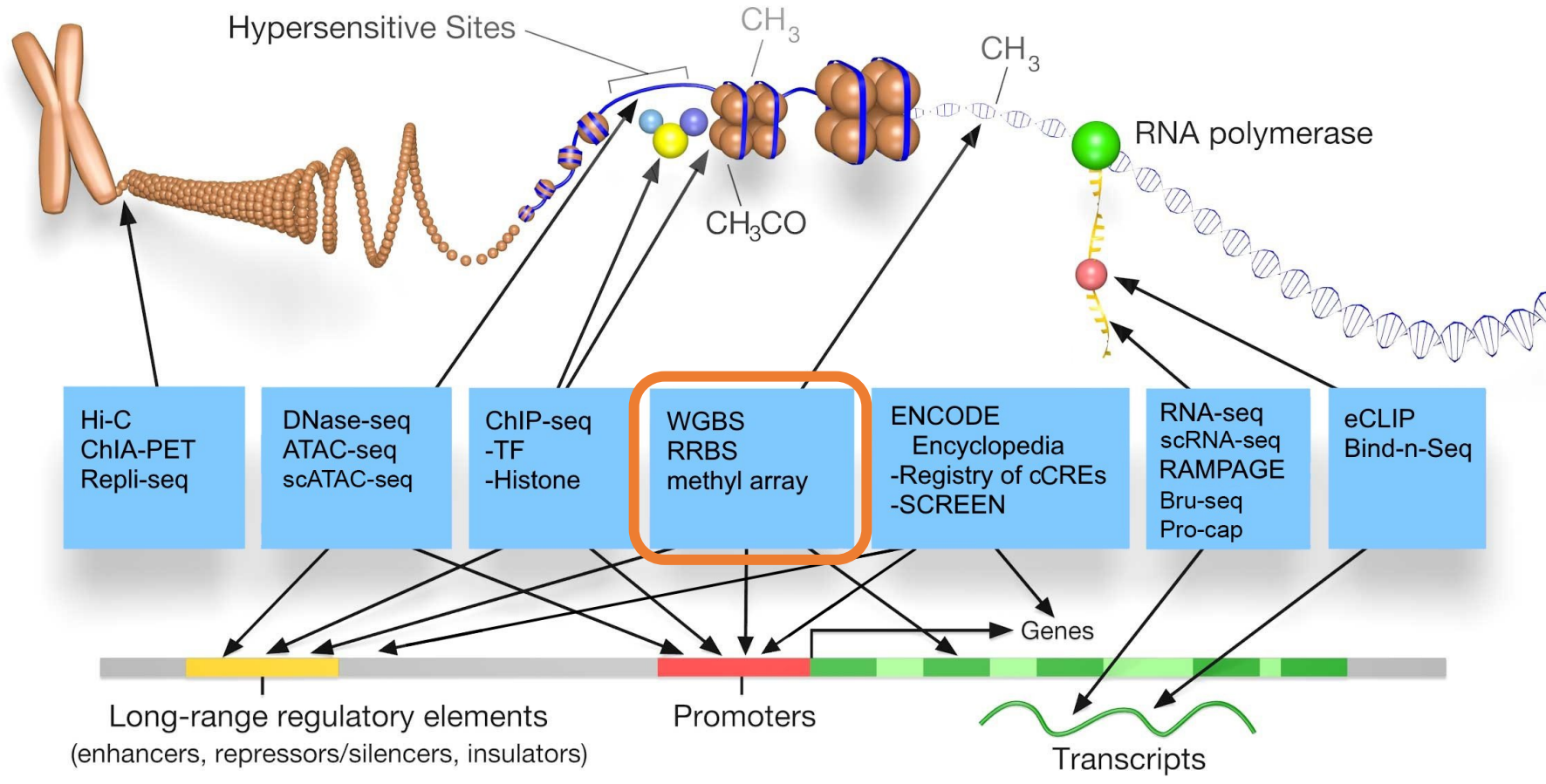
Bio Comp 2023-2024

INRAE

➤ Bio-informatique pour l'épigénome

Méthylation de l'ADN

➤ Cartographie des marques épigénétiques



Based on an image by Darryl Leja (NHGRI), Ian Dunham (EBI), Michael Pazin (NHGRI)



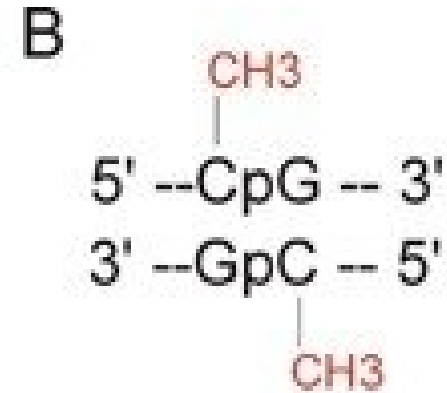
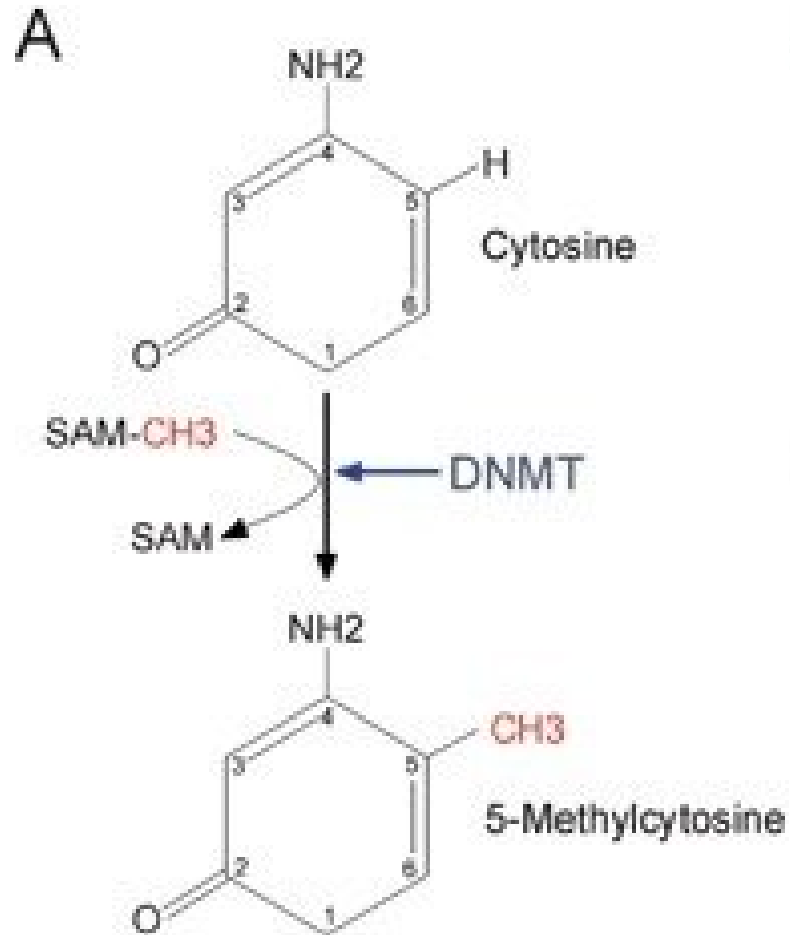
➤ Méthylation de l'ADN

Selon les organismes (plantes, vertébrés, invertébrés, micro-organismes) :

- Différents types de modifications (CG, CHG ou CHH) (dans ce cours focus sur CG)
- Différentes régions du génome affectées selon les organismes (souvent en lien avec la régulation des gènes)
- Transmission ou non à la génération suivante (reprogrammation méiotique)
- Connaissances plus ou moins approfondies



➤ Méthylation de l'ADN (CG)

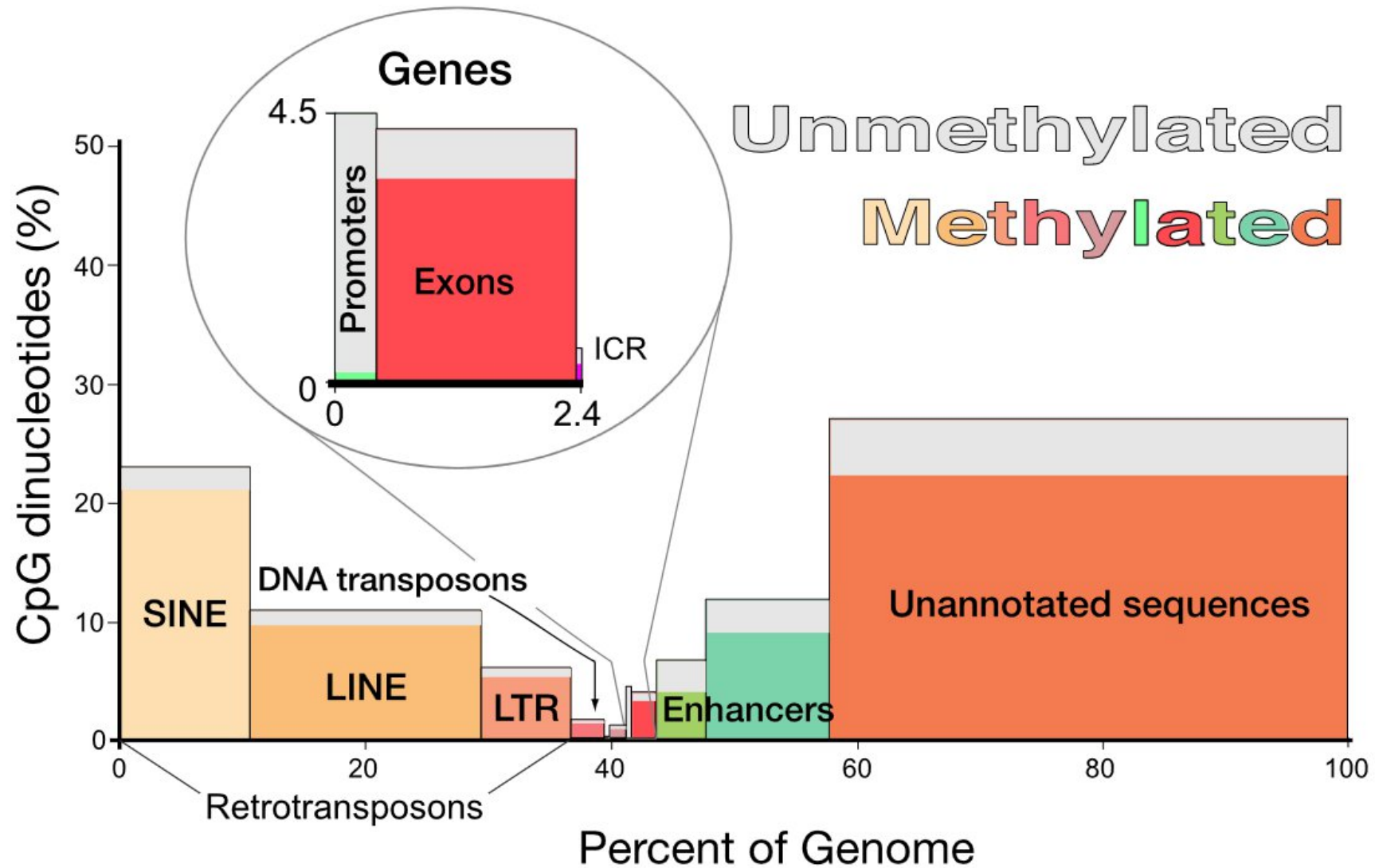


Ilôt CpG : 200 à 500 bp (ou plus)

Contenu G+C > 0.5

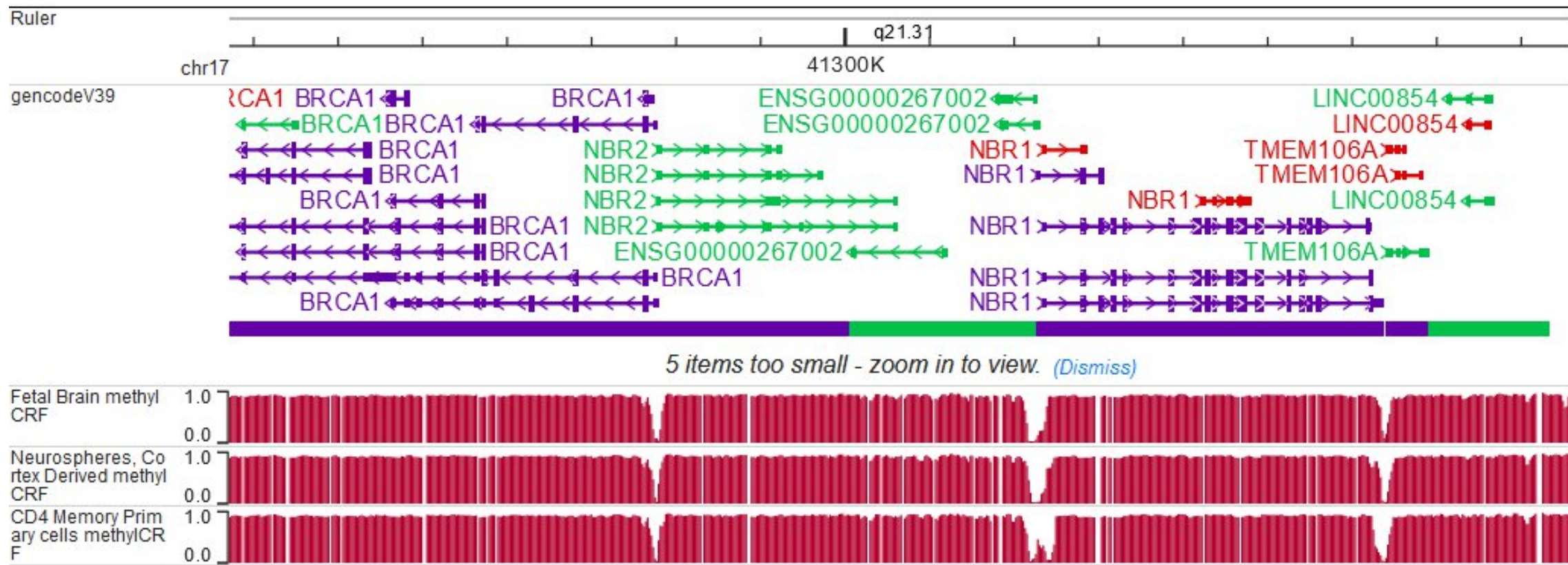
Taux de
CG_theorique/CG_attendu > 0.6

➤ Méthylation de l'ADN (CG)



➤ Méthylation de l'ADN (CG)

Tout le génome est méthylé sauf les éléments régulateurs actifs

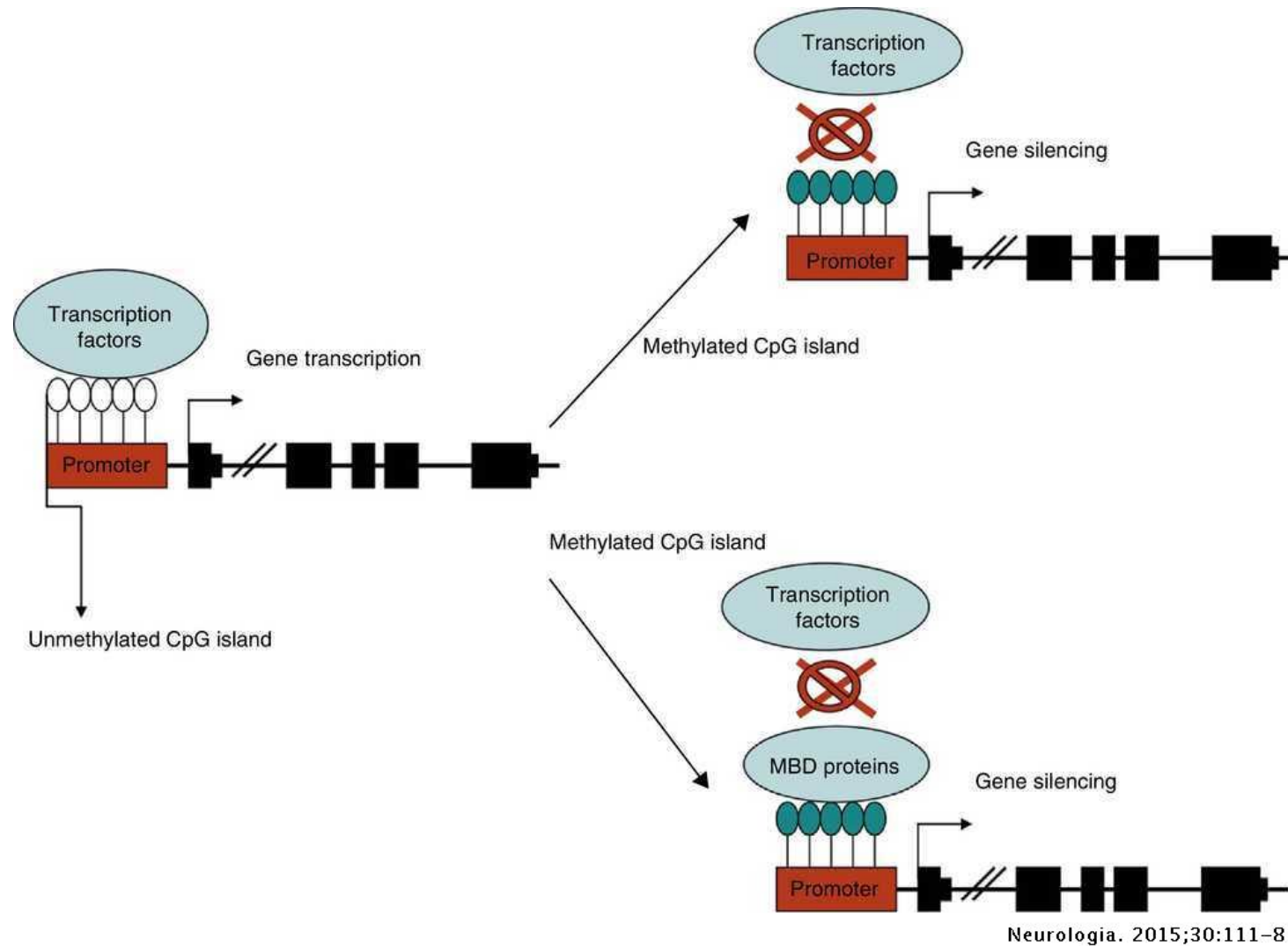


➤ Exemples des effets de la méthylation de l'ADN

- Silencing de certains éléments du génome (gènes, transposons ...)
- Silencing de gènes sur le chromosome X inactif chez les femelles (vu avec Guillaume)
- Expression mono-allélique de gènes sous empreinte

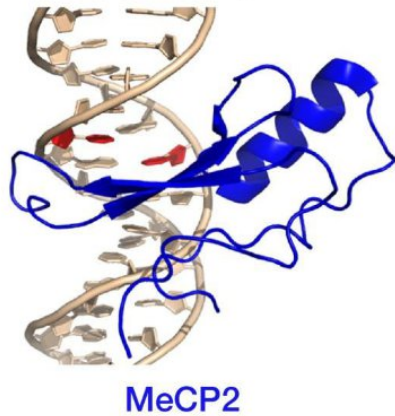


➤ Effet sur l'expression des gènes



➤ Méthylation de l'ADN (CG)

Methyl-Binding Domains (MBDs)



Recognition sequence: 5mCpG, 5mCpA
 Phenotype mutation: Rett syndrome
 Other members: MBD1, MBD2, MBD4

Some C2H2 Zinc Fingers (ZFs)



5'-G5mCGGCA-3'
 demethylation at some ICRs
 >600

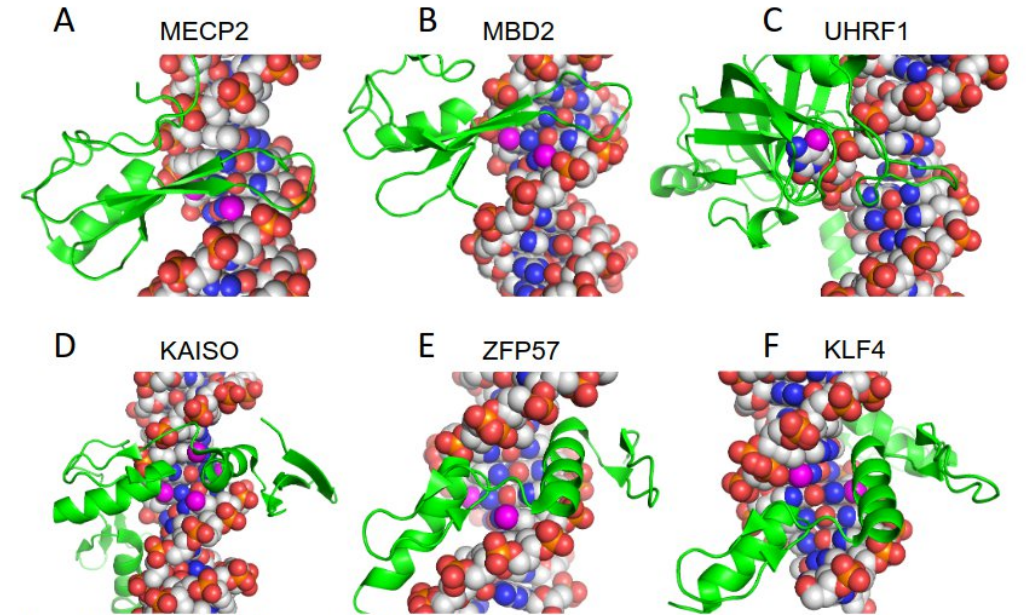
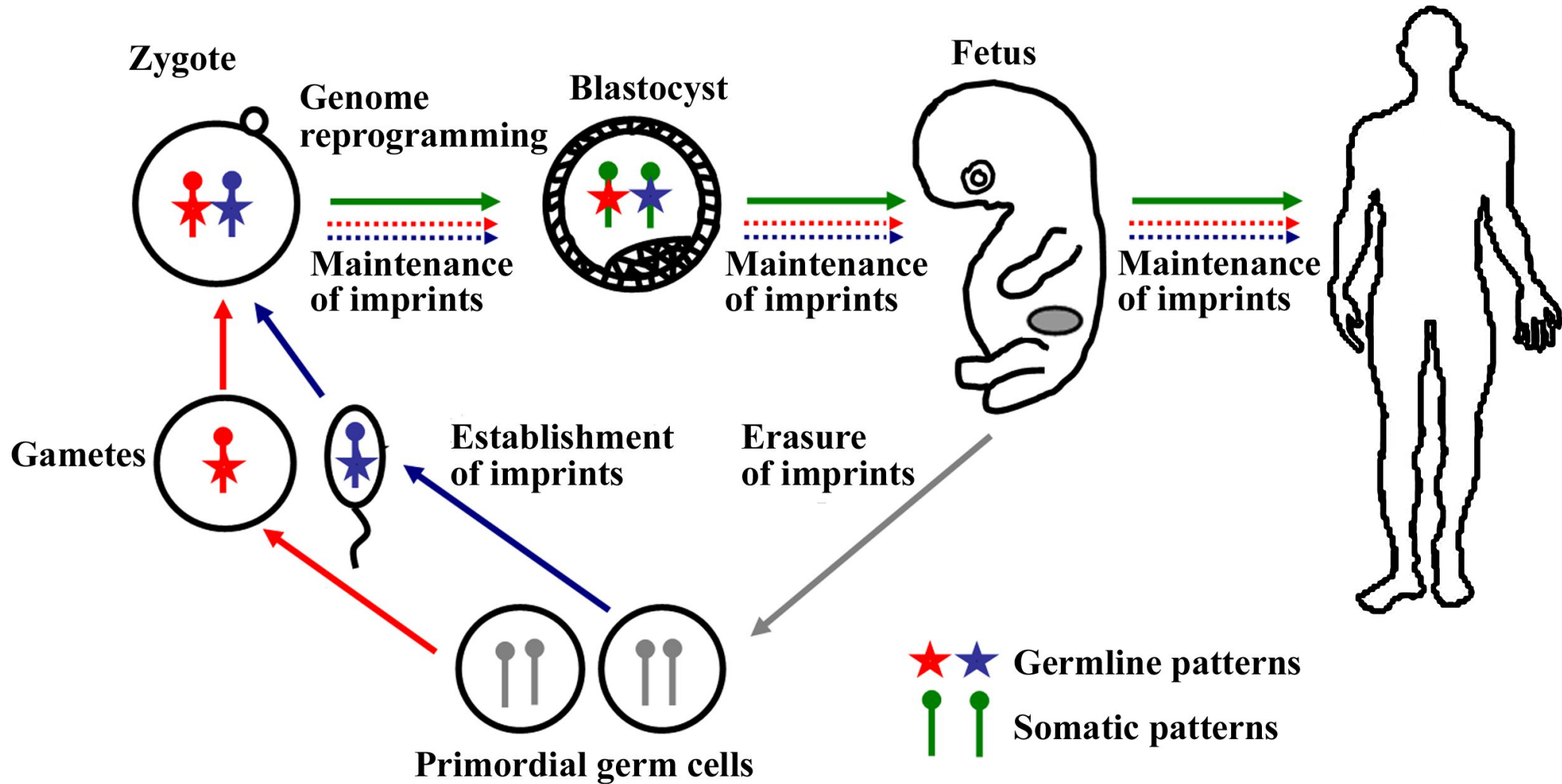


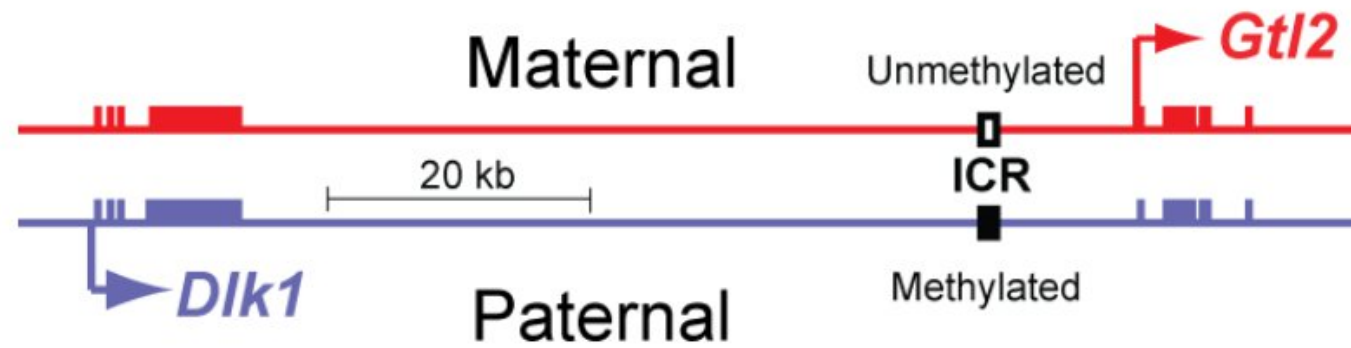
Figure 11 : Domaines de liaisons à l'ADN méthylé.

Vert : Domaines protéiques de liaison à l'ADN méthylé mettant en évidence les hélices alpha et feuilletés bêta. La double hélice d'ADN est représentée sous forme de boules de couleurs. Rose: groupe méthyl des 5-méthylcytosines. Blanc : autres carbones. Bleu : azotes. Rouge : oxygènes. Orange : phosphores. **A.** domaine MBD de MECP2 (PDB 3C2I). **B.** Domaine MBD de MBD2 (PDB 2KY8). L'interaction du domaine MBD de MBD2 avec l'ADN méthylé est grossièrement similaire à celle de MECP2. **C.** Domaine SRA de UHRF1 (PDB 3F8I). Le domaine SRA sort la cytosine méthylée de la double hélice. **D-F.** Domaines à doigt de zinc de KAISO (**D**, PDB 4F6N), ZFP57 (**E**, PDB 4M9E) et KLF4 (**F**, PDB 4GZN) en contact avec de l'ADN méthylé. La position des deux hélices vis-à-vis des cytosines méthylées diffère d'une protéine à l'autre.

➤ Reprogrammation méiotique



➤ L'empreinte épigénétique

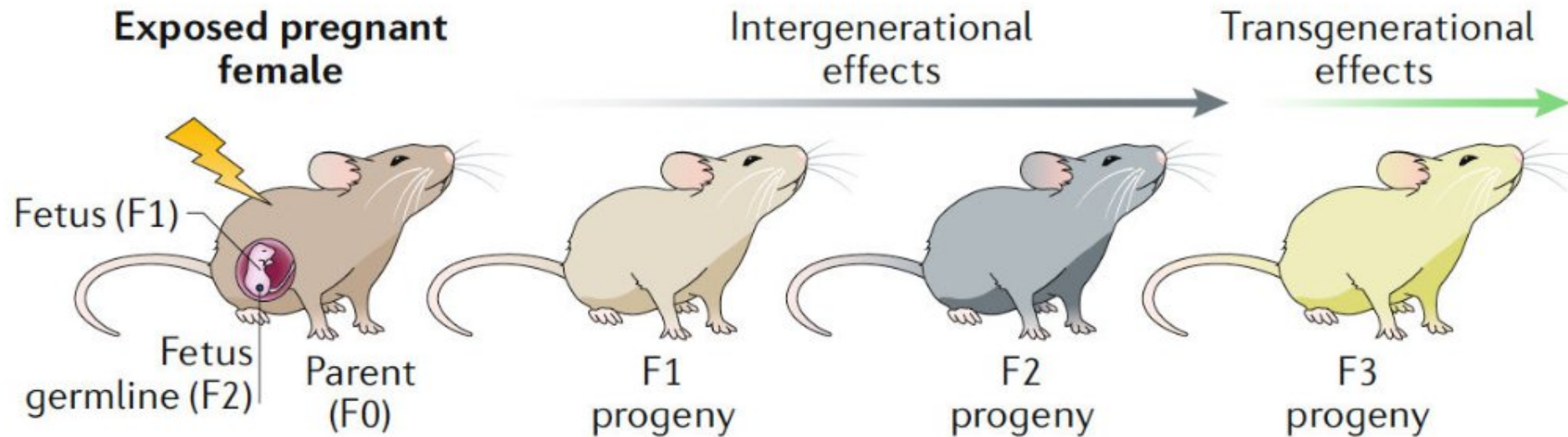


➤ L'épigénétique transgénérationnelle

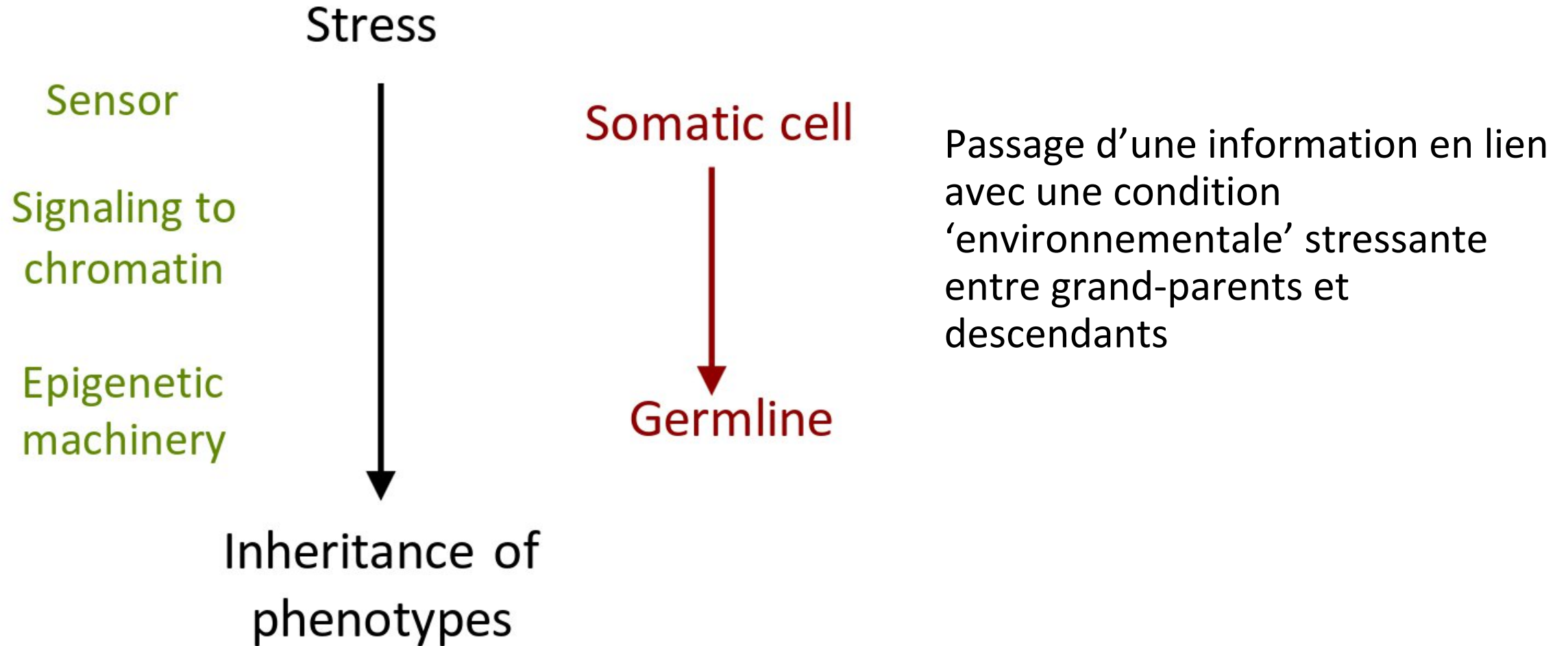
Une définition :

La transmission de marques / modifications épigénétiques d'une génération aux suivantes sans modifications de la structure de l'ADN.

En bref, même ADN mais traits différents.

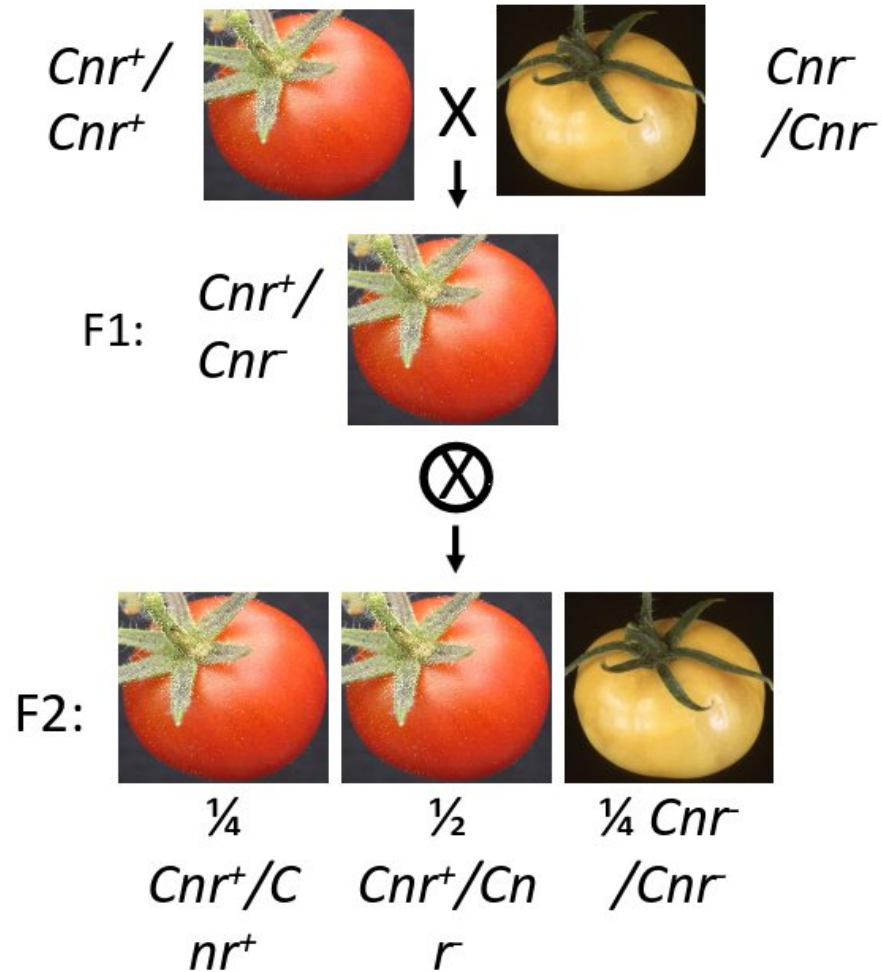


➤ L'épigénétique transgénérationnelle

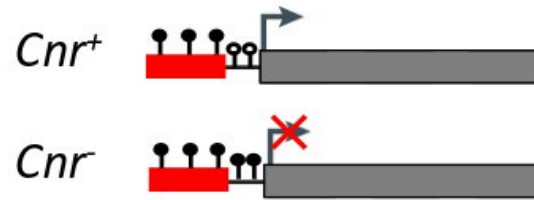


➤ L'épigénétique transgénérationnelle

Un exemple chez les plantes



One allele, two epialleles:



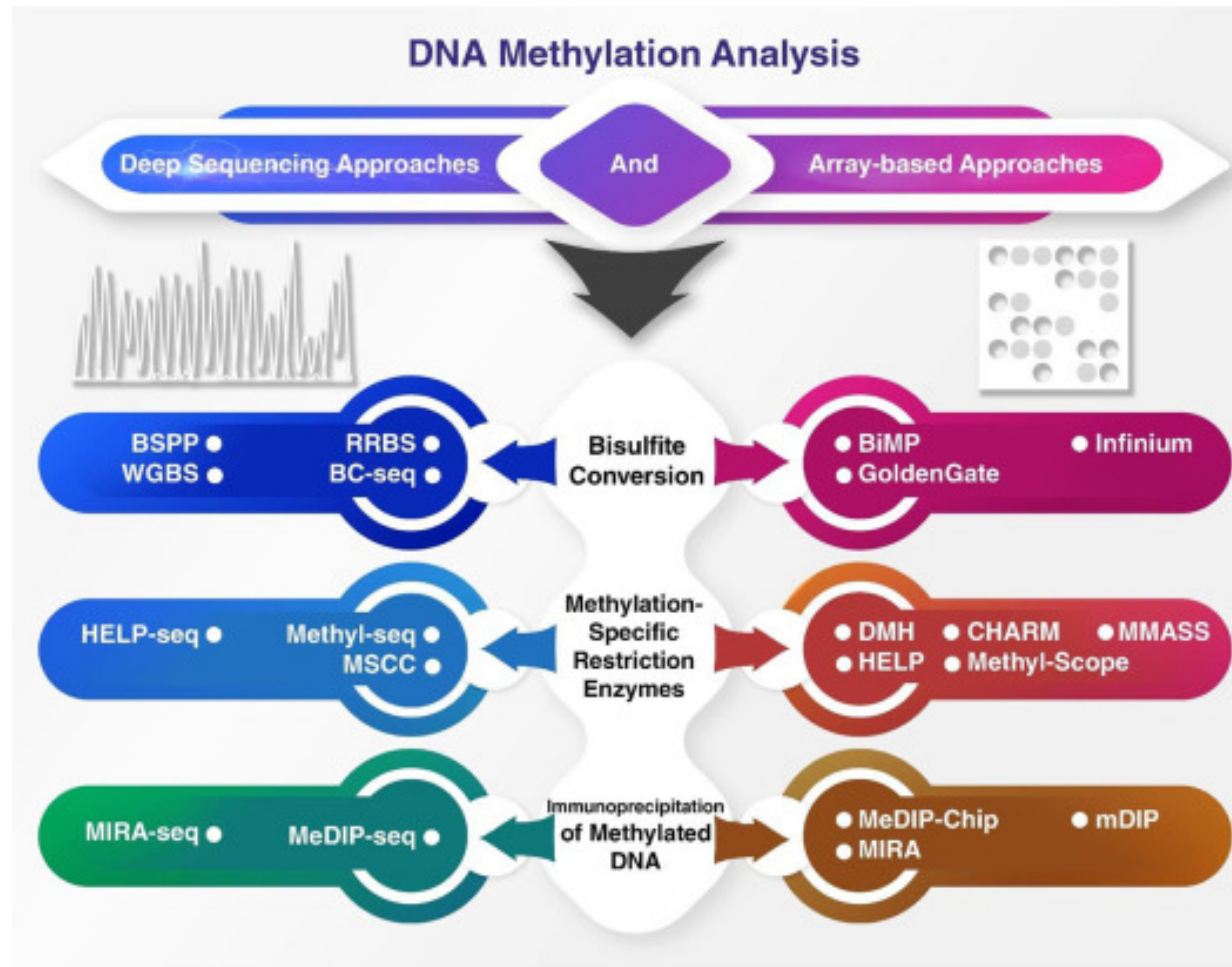
█ : *SBP* gene

█ : TE insertion

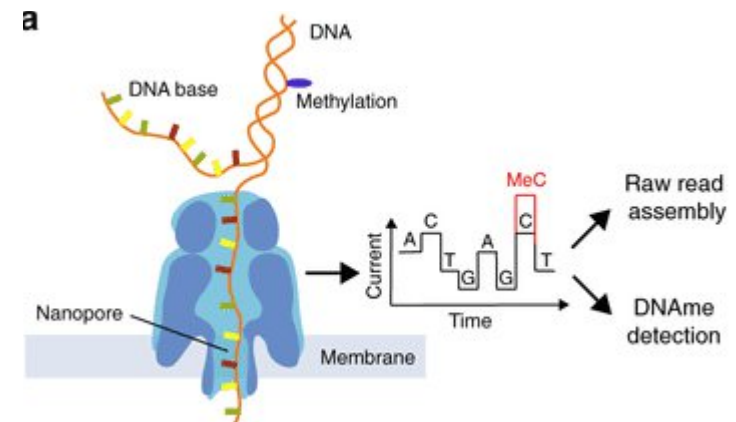
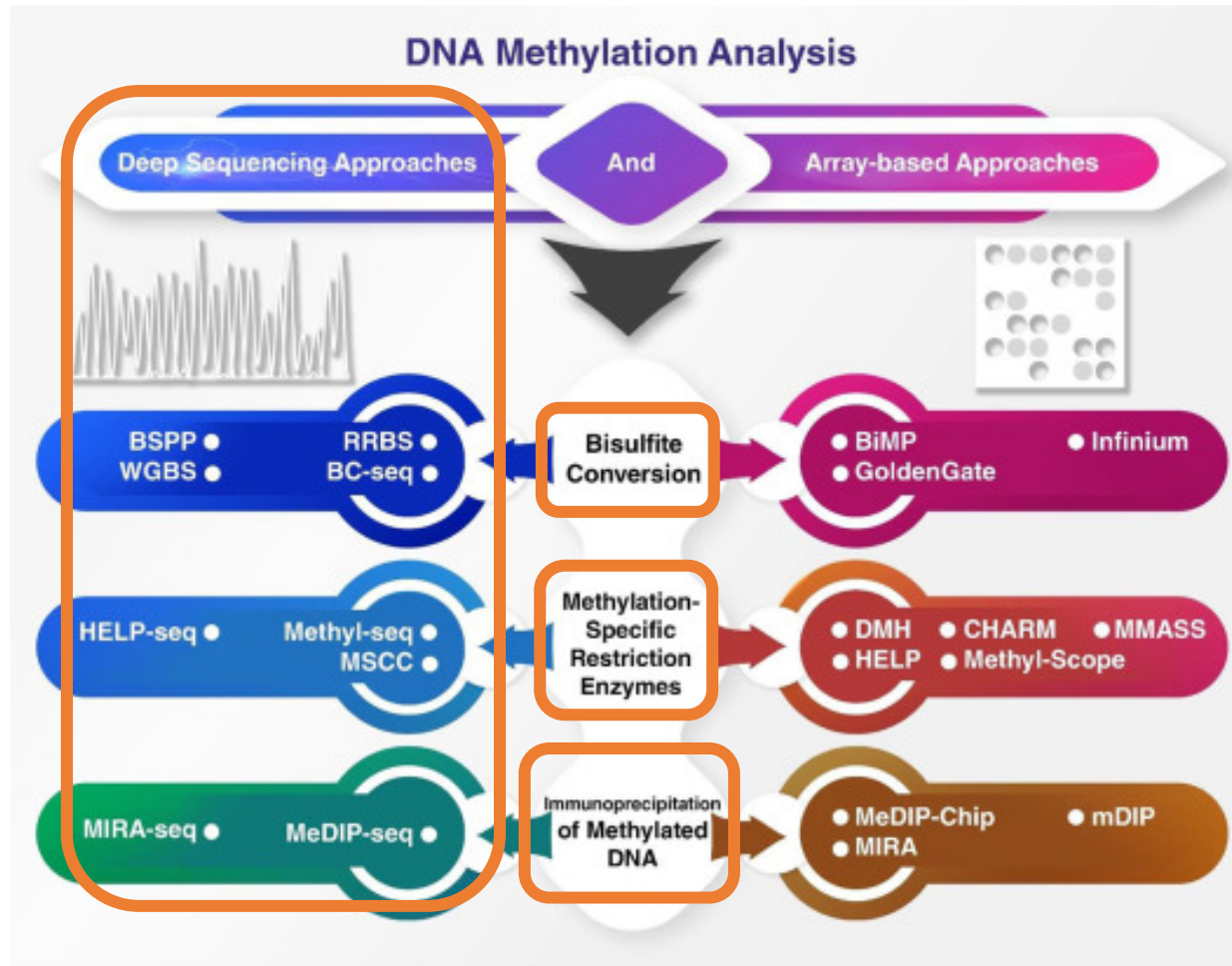
○ : unmethylated C

● : methylated C

➤ Méthodes moléculaires pour analyser la méthylation

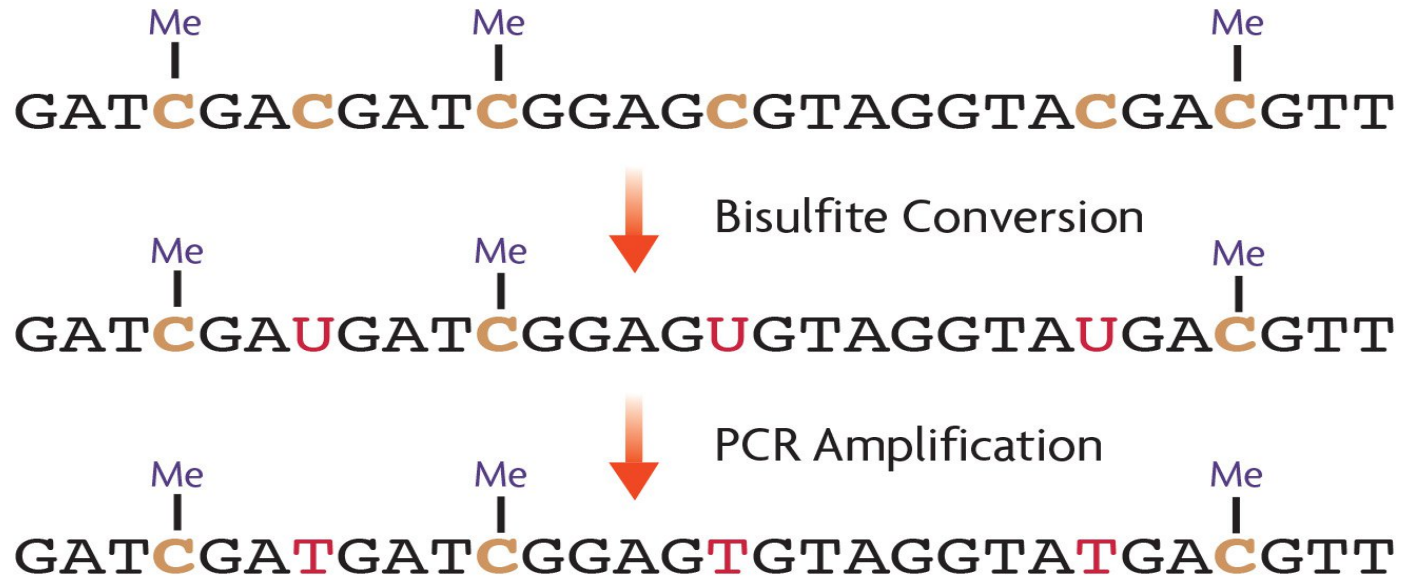


➤ Méthodes moléculaires pour analyser la méthylation

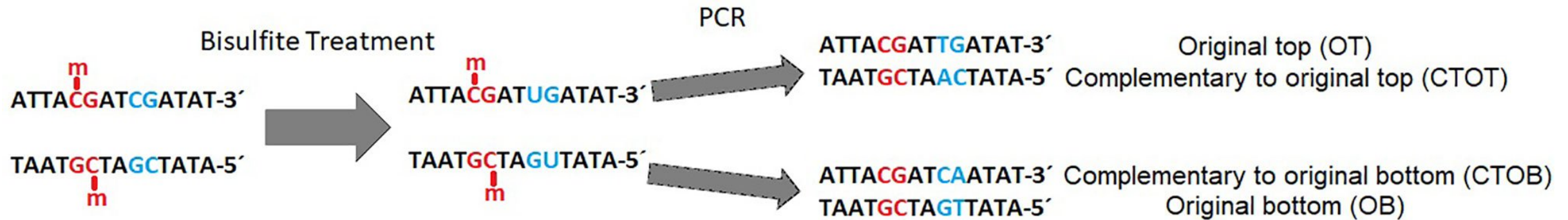


Séquençage long lectures
avec Oxford Nanopore

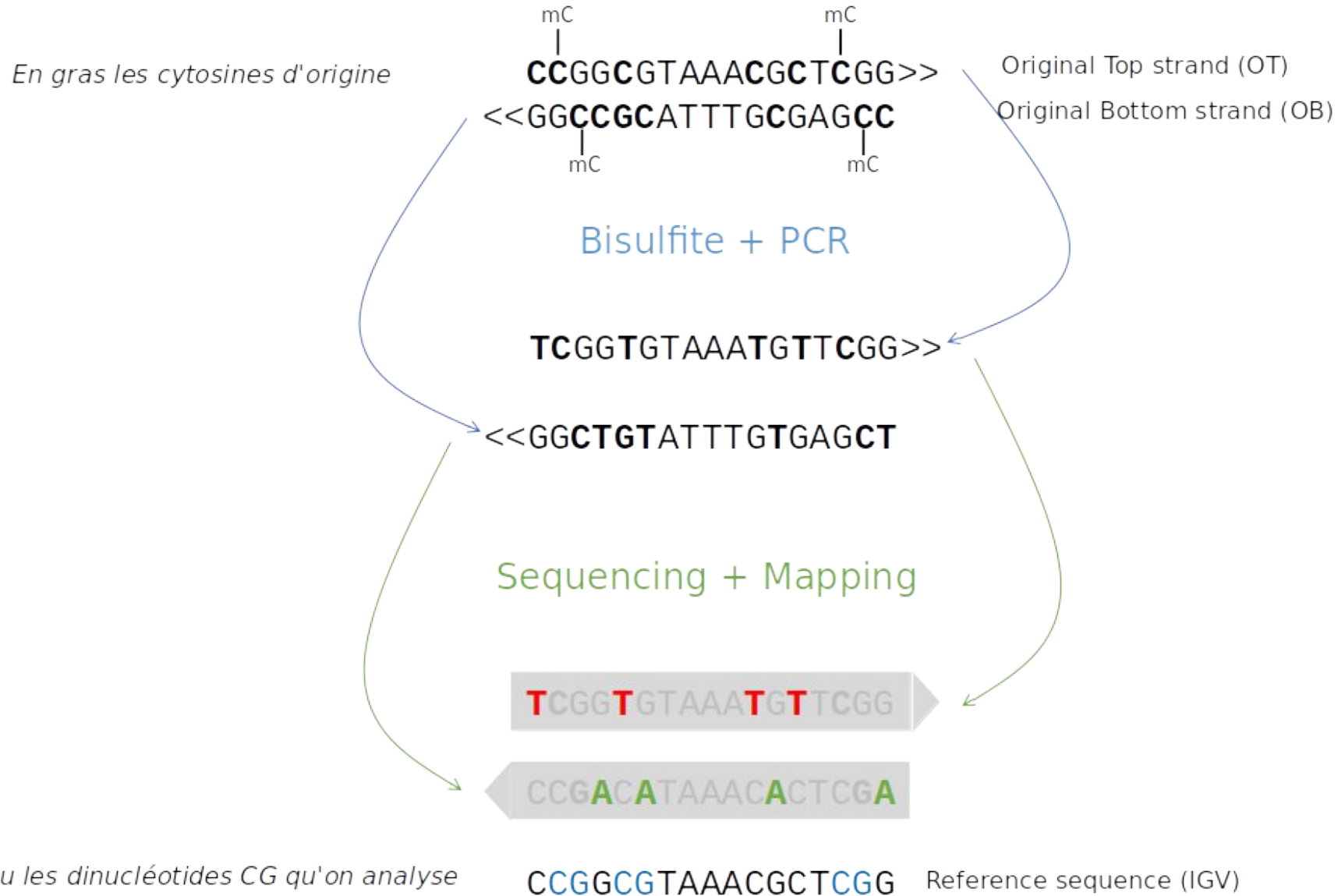
➤ Conversion bisulfite



➤ Conversion bisulfite



➤ Conversion bisulfite



➤ Conversion bisulfite - exercice1

Calculez le taux de méthylation, pour chaque position (en couleur), pour une séquence après traitement au bisulfite, après amplification PCR

```
GATTGATGATTGGAGTGTAGGTACGACGTT
GATTGATGATTGGAGCGTAGGTACGACGTT
GATTGATGATCGGAGTGTAGGTATGACGTT
GATTGATGATCGGAGCGTAGGTATGACGTT
GATCGATGATTGGAGTGTAGGTACGACGTT
GATTGATGATTGGAGTGTAGGTATGACGTT
GATTGATGATTGGAGCGTAGGTACGACGTT
GATTGATGATTGGAGTGTAGGTACGACGTT
GATTGATGATTGGAGTGTAGGTATGATGTT
```



➤ Conversion bisulfite - exercice1 - Réponse

GATTGATGATTGGAGTGTAGGTACGACGTT
GATTGATGATTGGAGCGTAGGTACGACGTT
GATTGATGATCGGAGTGTAGGTATGACGTT
GATTGATGATCGGAGCGTAGGTATGACGTT
GATCGATGATTGGAGTGTAGGTACGACGTT
GATTGATGATTGGAGTGTAGGTATGACGTT
GATTGATGATTGGAGCGTAGGTACGACGTT
GATTGATGATTGGAGTGTAGGTACGACGTT
GATTGATGATTGGAGTGTAGGTATGATGTT

$$\% \text{ méthylation} = C / (C+T)$$



➤ Conversion bisulfite - exercice1 - Réponse

GATT**T**GAT**T**GATT**T**GGAG**T**GTAGG**T**AC**G**AC**G**GTT
GATT**T**GAT**T**GATT**T**GGAG**T**GTAGG**T**AC**G**AC**G**GTT
GATT**T**GAT**T**GAT**C**GGAG**T**GTAGG**T**AT**T**GAC**G**GTT
GATT**T**GAC**G**AT**C**GGAG**C**GTAGG**T**AT**T**GAC**G**GTT
GAT**C**GAT**T**GAT**C**GGAG**T**GTAGG**T**AC**G**AC**G**GTT
GATT**T**GAT**T**GATT**T**GGAG**T**GTAGG**T**AT**T**GAC**G**GTT
GATT**T**GAT**T**GATT**T**GGAG**C**GTAGG**T**AC**G**AC**G**GTT
GATT**T**GAT**T**GATT**T**GGAG**T**GTAGG**T**AC**G**AC**G**GTT
GATT**T**GAT**T**GATT**T**GGAG**T**GTAGG**T**AT**T**GAT**T**GTT
10 10 30 20 60 90



➤ Conversion bisulfite - exercice2

GATCGACGATCGGAGCGTAGGTACGACGTT

- Écrire cette séquence non méthylée / partiellement méthylée / complètement méthylée
- Écrire les séquence forward / reverse
- Écrire les séquences après traitement bisulfite
- Écrire les séquences complémentaires après traitement bisulfite
- Calculer pour chacune de ces séquences les fréquences nucléotidiques
- Pour plus d'info sur la fréquence des dinucléotides : <https://bioinfo-fr.net/frequences-des-dinucleotides-dans-le-genome-dorganismes-modeles>



➤ Conversion bisulfite - exercice2

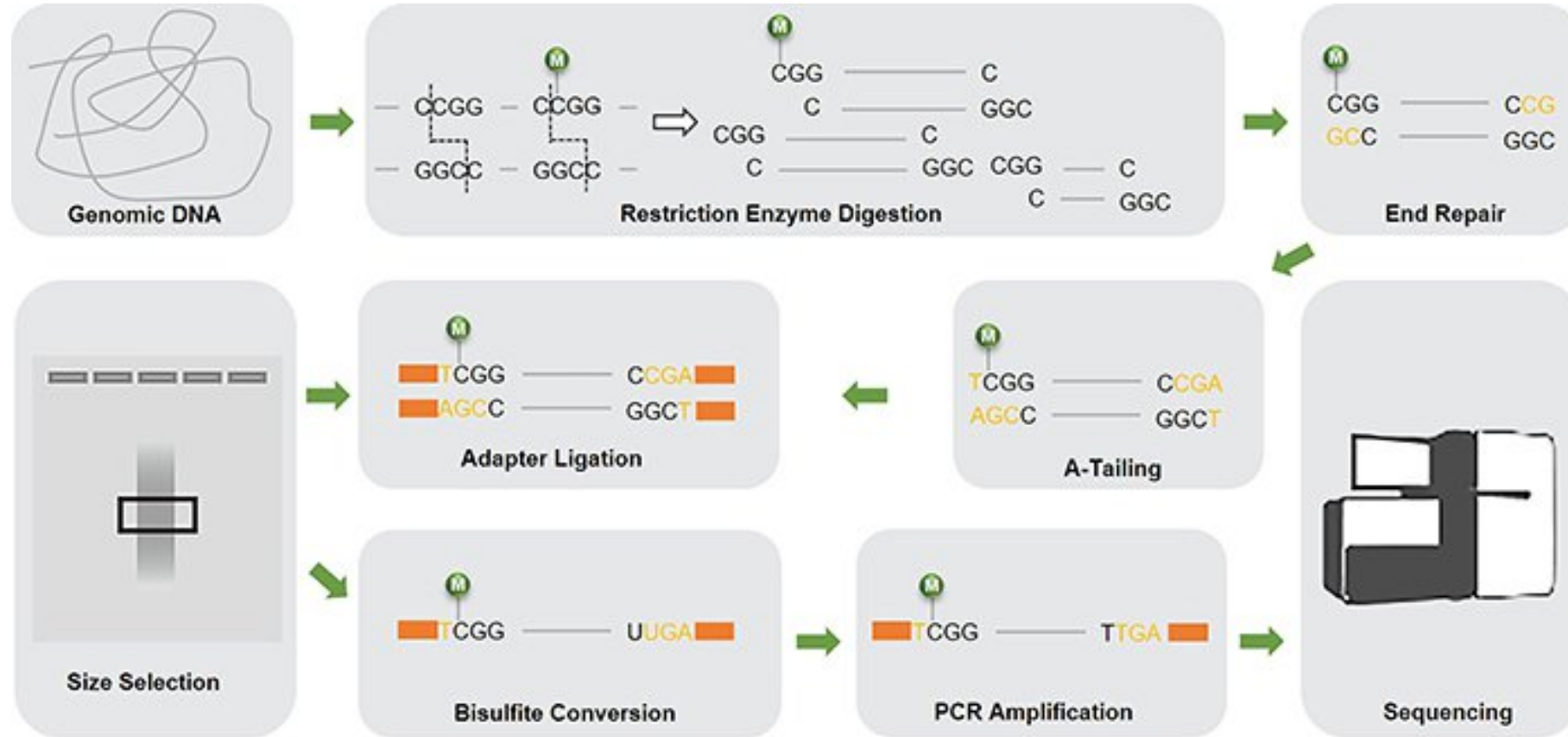
																	tot	A	T	C	G	fA	fT	fC	fG																																
Seq_F	>>>	G	A	T	C	G	A	C	G	A	T	C	G	G	A	G	C	G	T	A	G	G	T	A	C	G	A	C	G	T	T	>>>	30	7	6	6	11	0.23	0.2	0.2	0.37																
Seq_R	<<<	C	T	A	G	C	T	G	C	T	A	G	C	C	T	C	G	C	A	T	C	C	A	T	G	C	T	G	C	A	A	<<<	30	6	7	11	6	0.2	0.23	0.37	0.2																
Meth0	>>>	G	A	T	T	G	A	T	G	A	T	T	G	G	A	G	T	G	T	A	G	G	T	A	T	G	A	T	G	T	T	>>>	30	7	12	0	11	0.23	0.4	0	0.37																
		C	T	A	A	C	T	A	C	T	A	A	C	C	T	C	A	C	A	T	C	C	A	T	A	C	T	A	C	A	A		30	12	7	11	0	0.4	0.23	0.37	0																
	<<<	C	T	A	G	T	T	G	T	A	G	T	C	T	C	G	T	A	T	C	C	A	T	G	T	T	G	T	A	A	<<<	30	6	13	5	6	0.2	0.43	0.17	0.2																	
		G	A	T	C	A	A	C	A	A	T	C	A	G	A	G	C	A	T	A	G	G	T	A	C	A	A	C	A	T	T		30	13	6	6	5	0.43	0.2	0.2	0.17																
						me										me																																									
Meth_partie	>>>	G	A	T	C	G	A	T	G	A	T	T	G	G	A	G	C	G	T	A	G	G	T	A	C	G	A	T	G	T	T	>>>	30	7	9	3	11	0.23	0.3	0.1	0.37																
		C	T	A	G	C	T	A	C	T	A	A	C	C	T	C	G	C	A	T	C	C	A	T	G	C	T	A	C	A	A		30	9	7	11	3	0.3	0.23	0.37	0.1																
						me										me																																									
	<<<	C	T	A	G	C	T	G	T	A	G	T	C	T	C	G	C	A	T	C	C	A	T	G	C	T	G	T	A	A	<<<	30	6	10	8	6	0.2	0.33	0.27	0.2																	
		G	A	T	C	G	A	C	A	A	T	C	A	G	A	G	C	G	T	A	G	G	T	A	C	G	A	C	A	T	T		30	10	6	6	8	0.33	0.2	0.2	0.27																
						me	me			me						me								me	me																																
Meth_full	>>>	G	A	T	C	G	A	C	G	A	T	C	G	G	A	G	C	G	T	A	G	G	T	A	C	G	A	C	G	T	T	>>>	30	7	6	6	11	0.23	0.2	0.2	0.37																
		C	T	A	G	C	T	G	C	T	A	G	C	C	T	C	G	C	A	T	C	C	A	T	G	C	T	G	C	A	A		30	6	7	11	6	0.2	0.23	0.37	0.2																
						me	me			me						me								me	me																																
	<<<	C	T	A	G	C	T	G	C	T	A	G	C	C	T	C	G	C	A	T	C	C	A	T	G	C	T	G	C	A	A	<<<	30	6	7	11	6	0.2	0.23	0.37	0.2																
		G	A	T	C	G	A	C	G	A	T	C	G	G	A	G	C	G	T	A	G	G	T	A	C	G	A	C	G	T	T		30	7	6	6	11	0.23	0.2	0.2	0.37																



> RRBS

Reduced Representation Bisulfite Sequencing :

Enzyme = MspI



C/CGG

GGC/C




➤ Hands on: visualisation de données de séquençage bisulfite

- Récupérer les données RRBS poule mises à disposition sur Genobioinfo
`/work/user/gdevailly/cours/ensat/toydata/rrbs/RRBS_igv.tar.gz`

```
cd Downloads
scp
yourname@genobioinfo.toulouse.inrae.fr:/work/user/gdevailly/cours/ensat/toydata/rrbs/RRBS_igv.tar.gz .
tar -xf RRBS_igv.tar.gz
```

- Visualiser les données RRBS dans IGV
(<https://igv.org/app/>)

- Génome : Poule (*Gallus gallus*)
- Chromosome : 4
- Positions : 2 Mb à 3 Mb début du chromosome
- A voir : DMCs et DMRs
A voir en plus : les SNPs

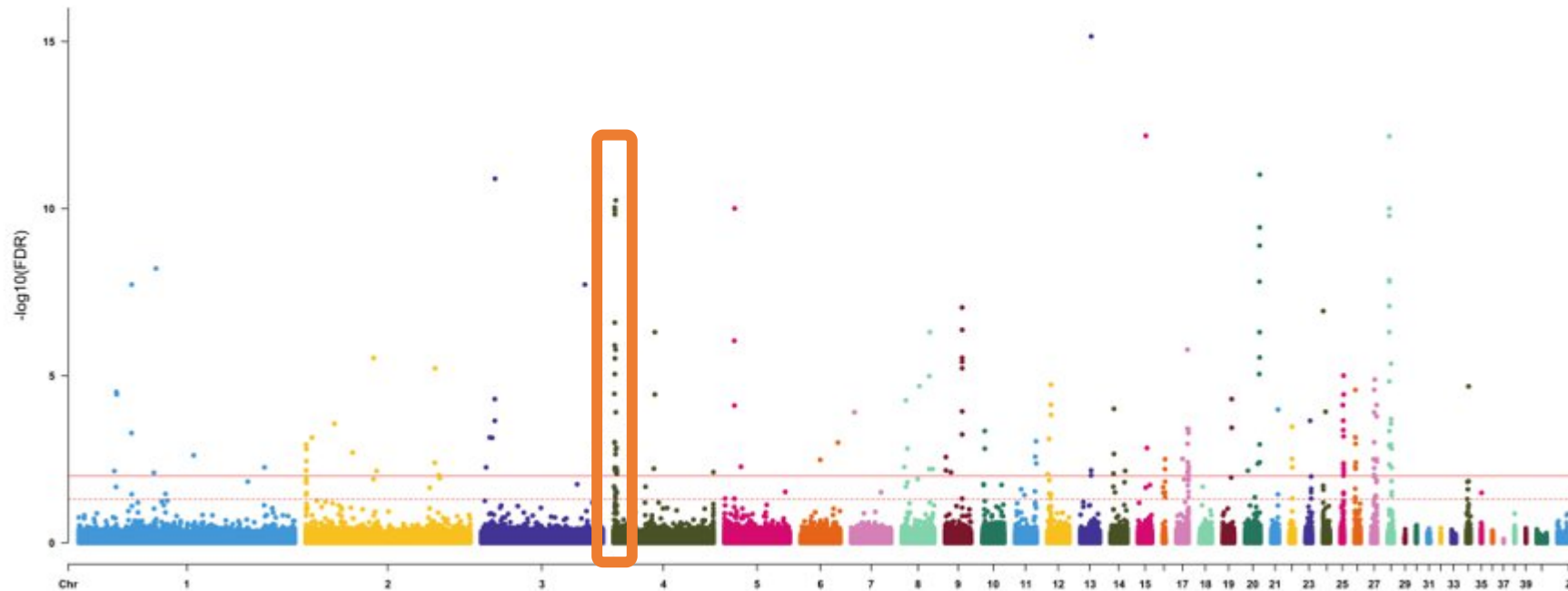


Chromosome	GenBank	RefSeq	Size (bp)	GC content (%)	Unlocalized count	Action
1	CM028482.1	NC_052532.1	196,449,156	40.5	4	⋮
2	CM028483.1	NC_052533.1	149,539,284	40	0	⋮
3	CM028484.1	NC_052534.1	110,642,502	40	0	⋮
4	CM028485.1	NC_052535.1	90,801,225	40.5	2	⋮
5	CM028486.1	NC_052536.1	59,506,338	41.5	0	⋮
6	CM028487.1	NC_052537.1	36,220,557	42	0	⋮
7	CM028488.1	NC_052538.1	35,382,834	41.5	0	⋮
8	CM028489.1	NC_052539.1	29,578,256	42	0	⋮
9	CM028490.1	NC_052540.1	23,733,309	43	0	⋮
10	CM028491.1	NC_052541.1	20,453,248	43.5	0	⋮
11	CM028492.1	NC_052542.1	19,638,187	42.5	0	⋮
12	CM028493.1	NC_052543.1	20,119,077	44	0	⋮
13	CM028494.1	NC_052544.1	17,905,061	45	0	⋮
14	CM028495.1	NC_052545.1	15,331,188	45.5	0	⋮
15	CM028496.1	NC_052546.1	12,703,657	45.5	0	⋮
16	CM028497.1	NC_052547.1	2,706,039	57.5	5	⋮
17	CM028498.1	NC_052548.1	11,092,391	48.5	0	⋮
18	CM028499.1	NC_052549.1	11,623,896	47.5	0	⋮
19	CM028500.1	NC_052550.1	10,455,293	47.5	0	⋮
20	CM028501.1	NC_052551.1	14,205,659	46	0	⋮
21	CM028502.1	NC_052552.1	6,970,754	48	0	⋮
22	CM028503.1	NC_052553.1	4,686,657	48.5	0	⋮
23	CM028504.1	NC_052554.1	6,253,421	51	0	⋮
24	CM028505.1	NC_052555.1	6,478,339	49.5	0	⋮
25	CM028506.1	NC_052556.1	3,067,737	57	0	⋮
26	CM028507.1	NC_052557.1	5,349,051	52.5	0	⋮
27	CM028508.1	NC_052558.1	5,228,753	52.5	0	⋮
28	CM028509.1	NC_052559.1	5,437,364	53	0	⋮
29	CM028510.1	NC_052560.1	726,478	54	2	⋮
30	CM028511.1	NC_052561.1	755,066	59.5	0	⋮
31	CM028512.1	NC_052562.1	2,457,334	52	3	⋮
32	CM028513.1	NC_052563.1	125,424	55	6	⋮
33	CM028514.1	NC_052564.1	3,839,931	49	2	⋮
34	CM028515.1	NC_052565.1	3,469,343	57.5	0	⋮
35	CM028516.1	NC_052566.1	554,126	56.5	5	⋮
36	CM028517.1	NC_052567.1	358,375	58.5	6	⋮
37	CM028518.1	NC_052568.1	157,853	55.5	6	⋮
38	CM028519.1	NC_052569.1	667,312	62	0	⋮
39	CM028520.1	NC_052570.1	177,356	68.5	2	⋮
W	CM028521.1	NC_052571.1	9,109,940	46.5	19	⋮
Z	CM028522.1	NC_052572.1	86,044,486	41.5	0	⋮
MT	CM028585.1	NC_053523.1	16,784	46	0	⋮

➤ Hands on: visualisation de données de séquençage bisulfite

DMR : 2,858,109-2,858,165

DMC : 2,008,788 / 2,435,902



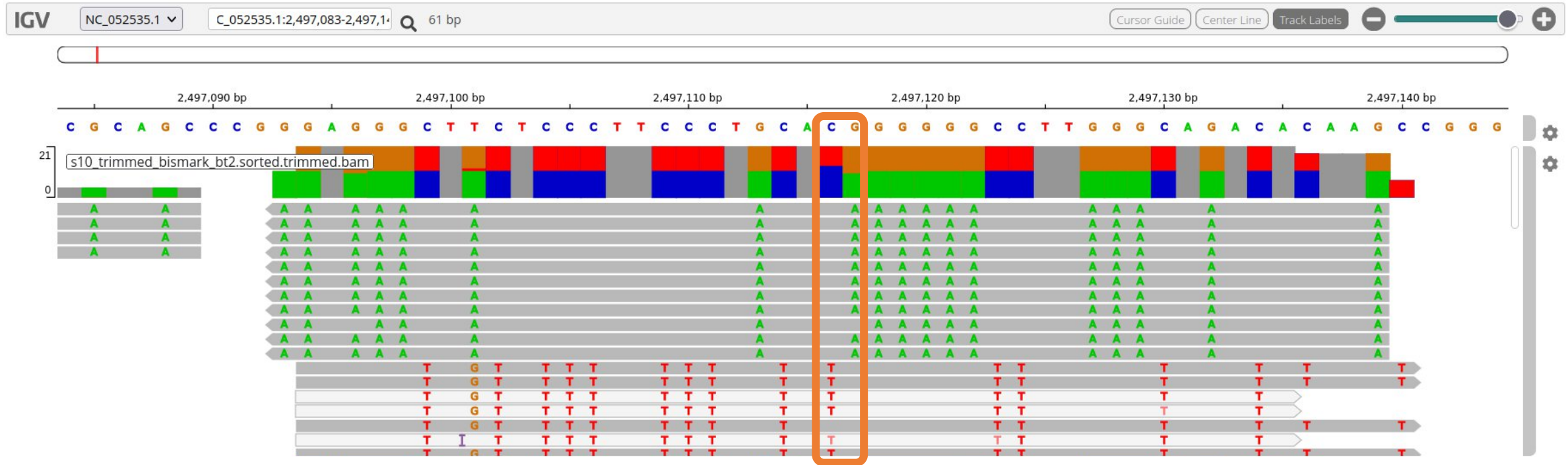
➤ Hands on: visualisation de données de séquençage bisulfite



Hands on: visualisation de données de séquençage bisulfite

TCGGTGTAAATGTTCGG

CCGACATAAACACTCGA



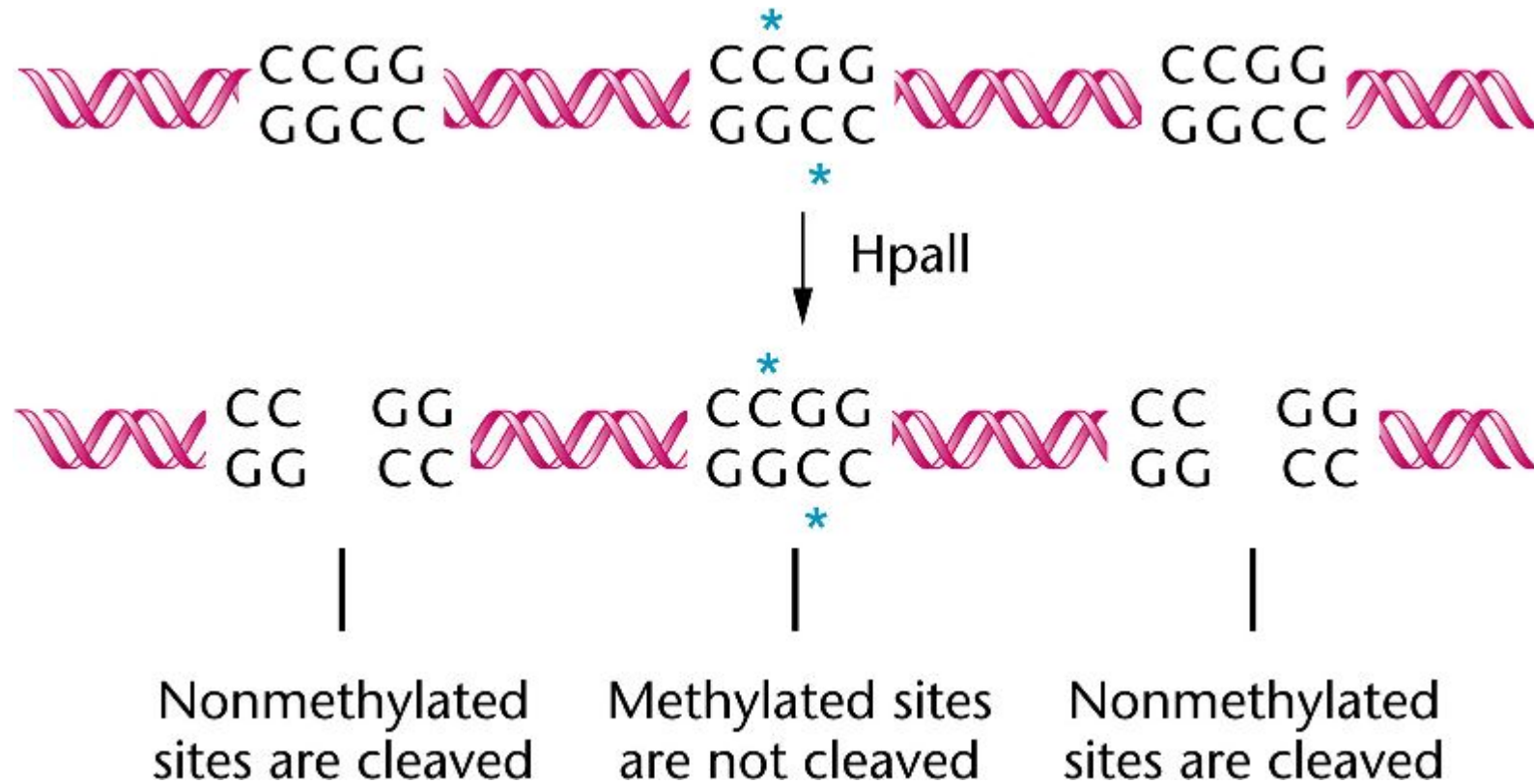
➤ WGBS



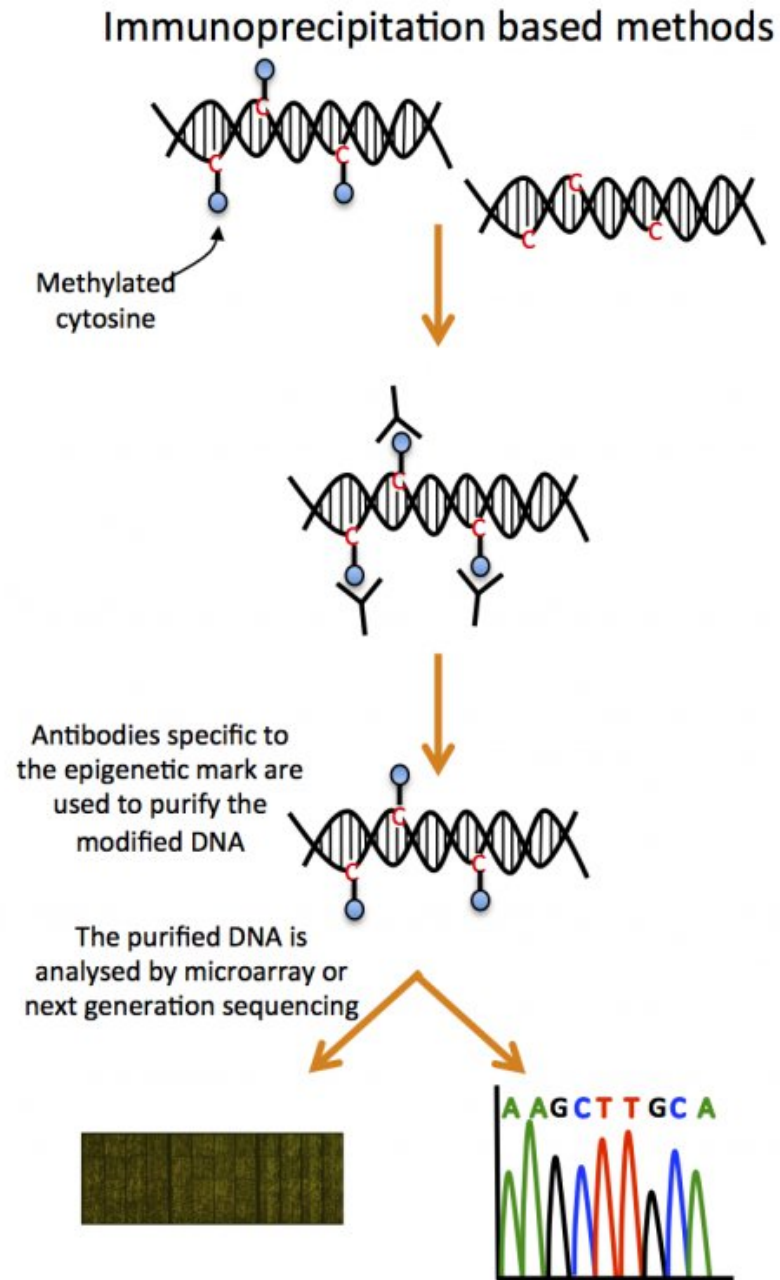
4 C / 5 ou 4 G / 5 →

8 C méthylés sur une position couverte par 5 lectures = taux de methylation 80%

➤ Utilisation enzymes de restriction spécifiques de la méthylation



➤ Immunoprecipitation



➤ Avantages / Limitations

– Avantage:

- Adaptable
- Peu couteuse si on accepte une sous representation ou une couverture faible
- Analyses standard, beaucoup de développement

– Inconvénients:

- Repose sur un traitement de l'ADN (Bisulfite, enzymes ...)
- Difficulté d'identifier un SNP versus une marque épigénétique au niveau des sites CpG



➤ Avantages / Limitations

– Avantage:

- Adaptable
- Peu couteuse si on accepte une sous representation ou une couverture faible
- Analyses standard, beaucoup de développement

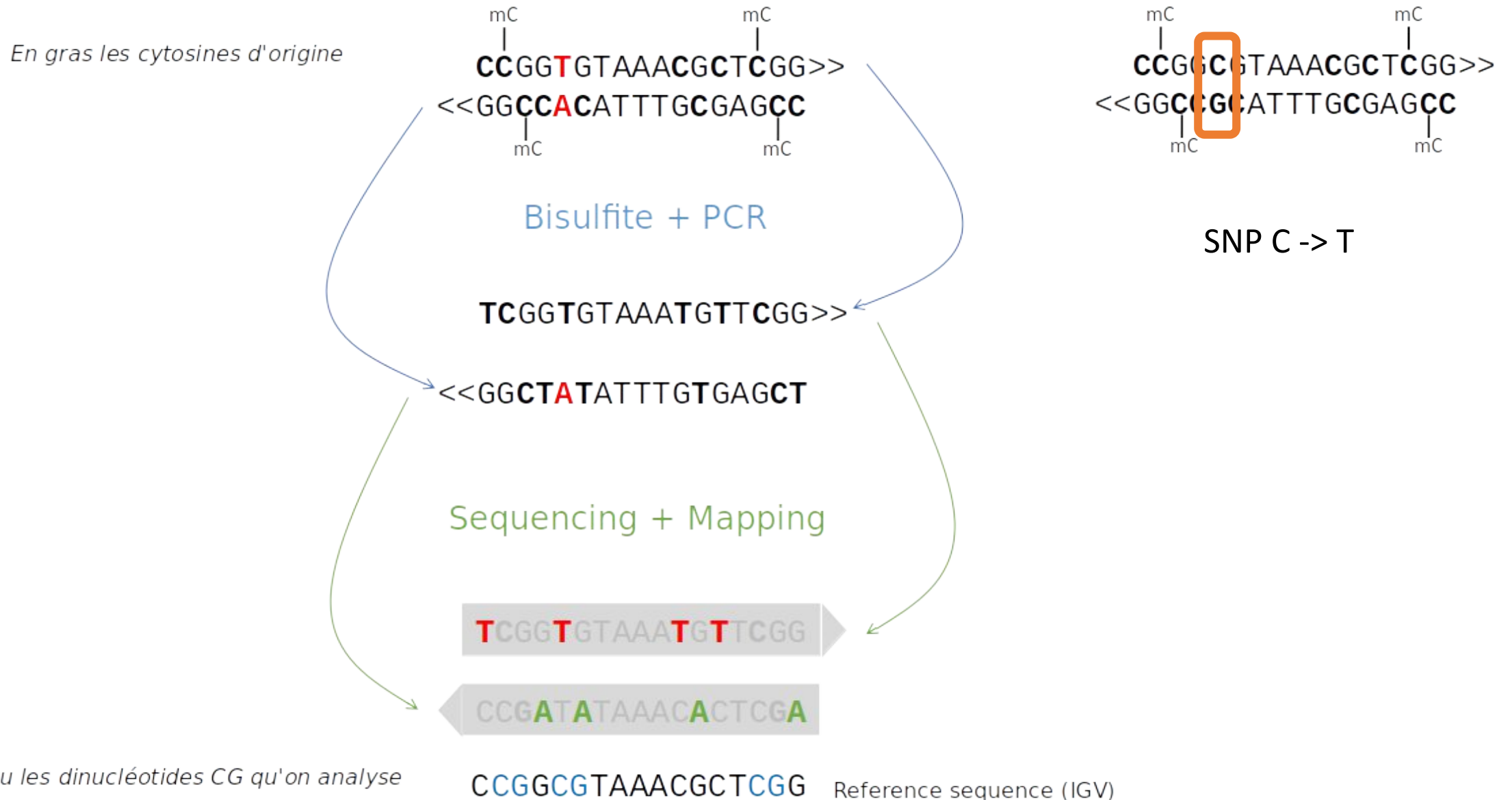
– Inconvénients:

- Repose sur un traitement de l'ADN (Bisulfite, enzymes ...)
- Difficulté d'identifier un SNP versus une marque épigénétique au niveau des sites CpG

Pourquoi ?



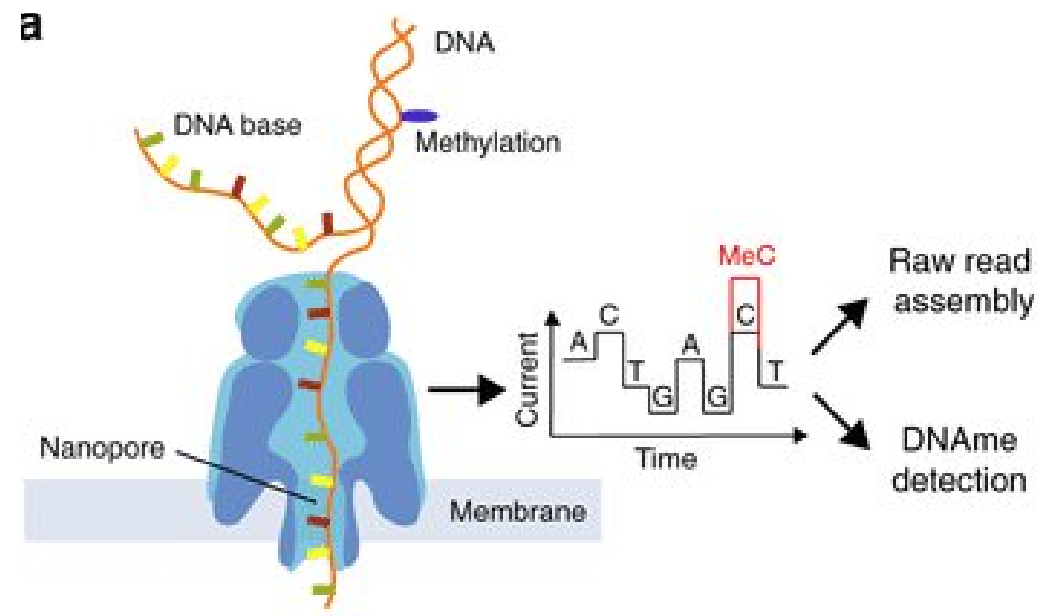
➤ Méthylation versus SNP pour CpG



➤ Séquençage

Oxford Nanopore Sequencing

- Lectures longues (plusieurs kbp)
- Utilisation du signal électrique pour identifier la méthylation
- Distinction entre SNPs et C-méthylé possible et direct
- Beaucoup d'outils en développement (méthodologie d'analyse pas encore stable)



> Séquençage

Oxford Nanopore Sequencing - Outils

- Alignement : Megalodon, guppy, Dorado (directement au moment du séquençage, sur la machine ou a posteriori)
- Methylation : Nanopolish, modbam2bed, Dorado



➤ Méthodes bioinformatique pour analyser la méthylation

Pipeline standard pour l'analyse de la méthylation

????

➤ Méthodes bioinformatique pour analyser la méthylation

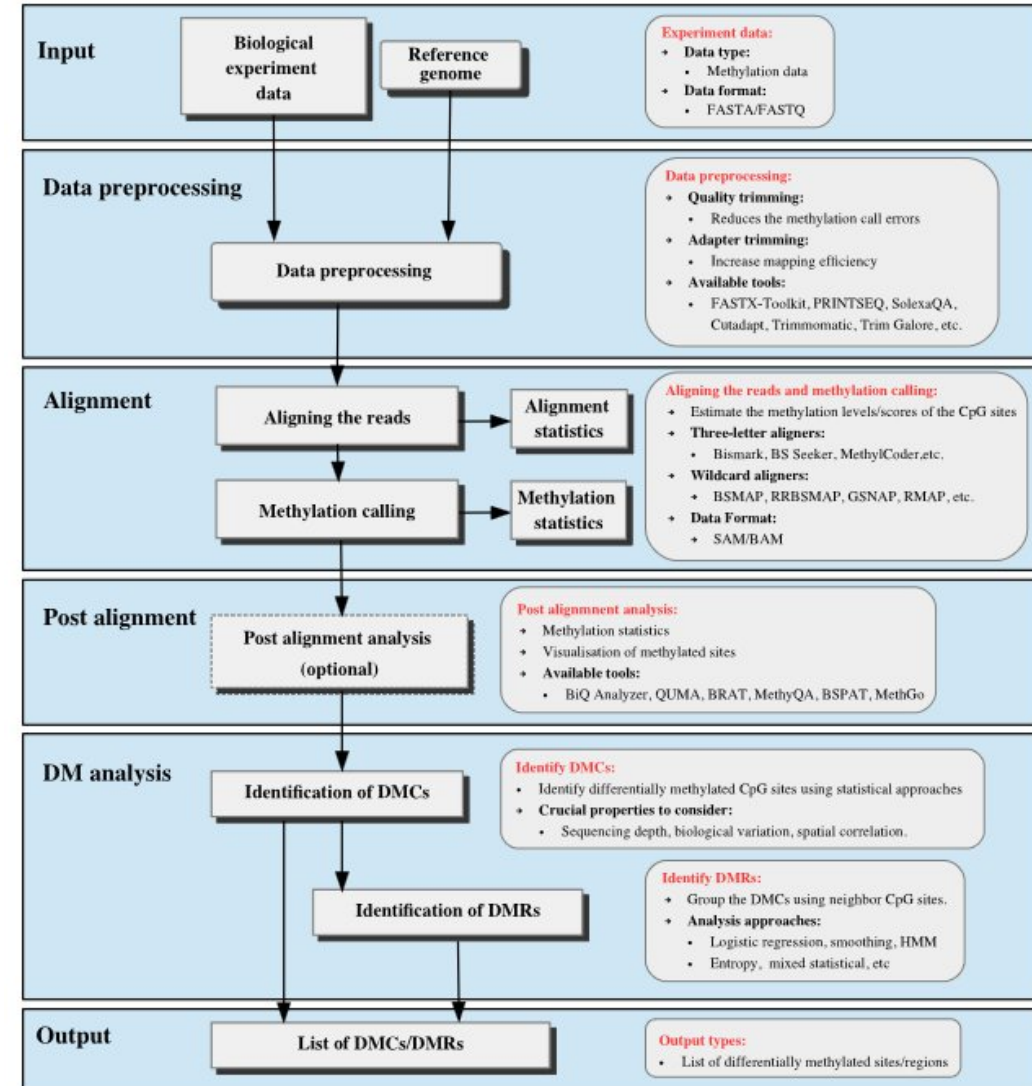
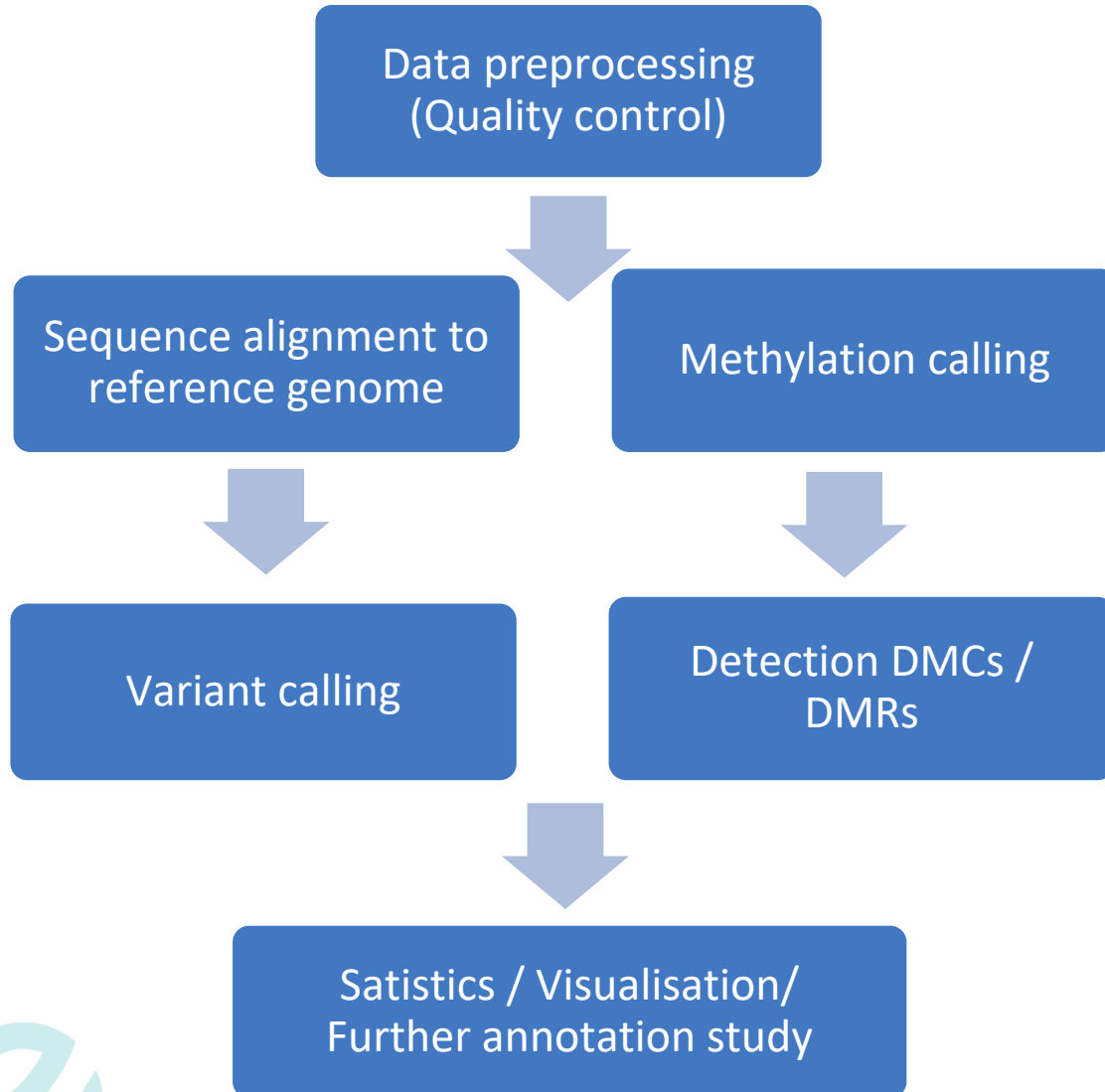


Figure 2: The workflow of analyzing DNA methylation using bisulfite sequencing data.

➤ Méthodes bioinformatique pour analyser la méthylation

Calcul taux de méthylation

- à partir du tableau ci dessous calculez, pour chaque position, le taux de méthylation

Chromosome	Position	X (couverture)	N (# C méthylé)
1	10	16	16
1	15	12	10
1	17	20	2
1	22	18	2
1	35	7	2



➤ Méthodes bioinformatique pour analyser la méthylation

Calcul taux de méthylation

Chromosome	Position	X (couverture)	N (# C méthylé)	Taux de méthylation
1	10	16	16	1 = 100 %
1	15	12	10	0.83 = 83 %
1	17	20	2	0.1 = 10 %
1	22	18	2	0.11 = 11 %
1	35	7	2	0.28 = 28 %



➤ Méthodes statistiques pour analyser la méthylation

Différentiel de méthylation

- Changement du taux de méthylation entre 2 groupes/traitements ...
- Identification
 - DMCs (Differentially Methylated Cytosines)
 - DMRs (Differentially Methylated Regions) = regions comportant plusieurs DMCs significativement différentes entre groupes/traitements/individus.
- Test statistique
 - Test exact de Fisher
 - Test Man-Whitney
 - T-tests



➤ Méthodes statistiques pour analyser la méthylation

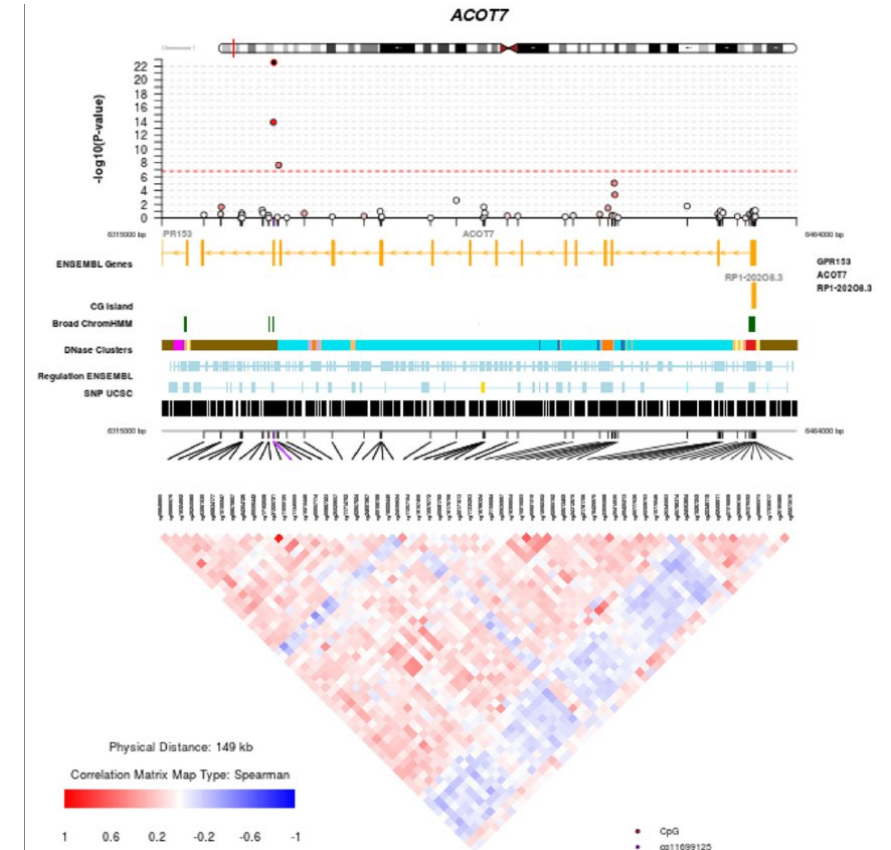
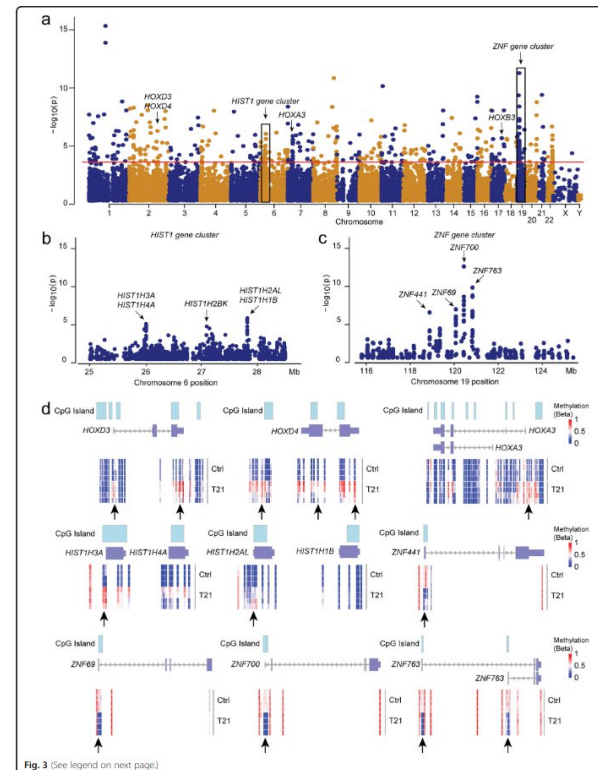
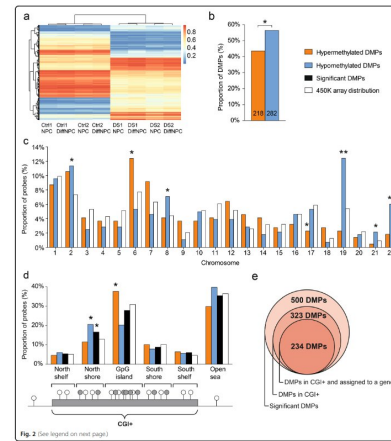
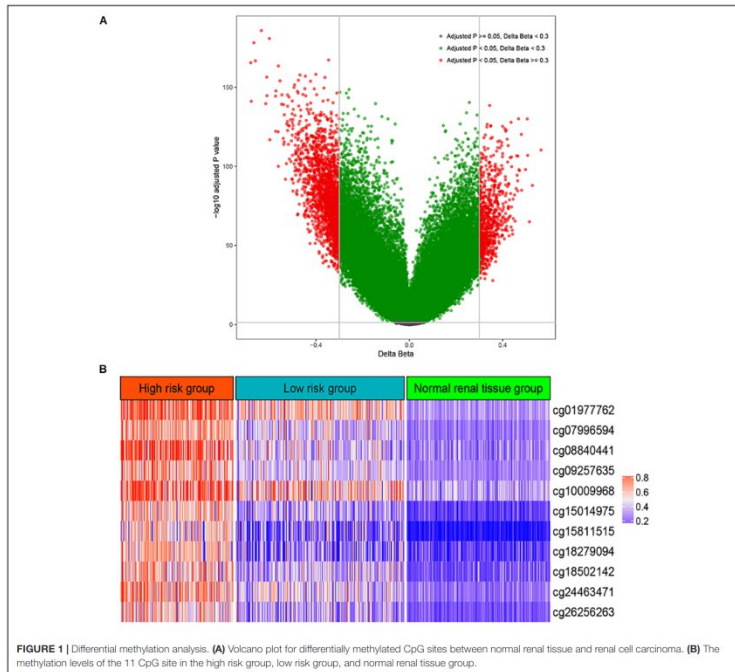
Différentiel de méthylation

– Test statistique

- Beaucoup de marqueurs mesurés
- Correction pour tests multiples nécessaire
- Bonferonni / Benjamin-Hochberg (BH) / FDR (False Discovery Rate)
 - Ajuste la p-value (valeur de significativité) pour le nombre de tests que l'on effectue
 - Seuil de significativité diminue
 - Test plus strict



➤ Représentations graphiques



➤ Développements futur

L'analyse de la méthylation pour :

- Voir l'impact de traitements, environnements ... Sur la méthylation
- Comprendre la régulation des gènes dans différents tissus, à différents temps, sous différentes conditions
- Comprendre les marques parentales non génétiques transmises
- Ajouter ces informations à la sélection, la conservation ... Comme processus d'adaptation, de plasticité

