



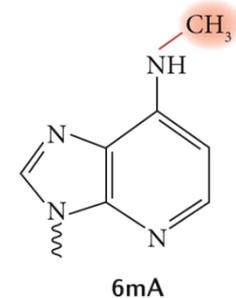
➤ Retour d'expérience
sur l'épigénomique chez les bactéries

➤ Méthylation de l'ADN chez les bactéries

Les différentes méthylations

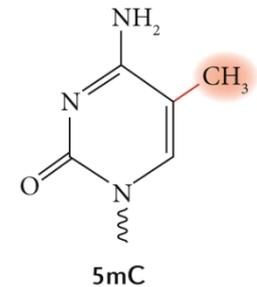
6mA

- Méthylation la plus fréquente chez les bactéries
- Présence chez certains eucaryotes



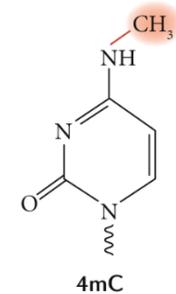
5mC

- Méthylation très étudiée chez les eucaryotes
- Peu étudiée chez les bactéries



4mC

- Méthylation spécifique des bactéries
- Peu étudiée, son rôle est mal connu



➤ Méthylation de l'ADN chez les bactéries

Fonctionnement des méthyltransférases

Absence de déméthylation active connue

- 90 à 99% des motifs sont méthylés
- Absence de méthylation == compétition
- Hémiméthylation lors de la réplication et peut-être lors de compétition

Reconnaissance d'un motif précis

- Grande diversité des motifs
- 3 classes de motifs
 - Type I, motif long en 2 parties (AAGNNNNNNCTC)
 - Type II, motif court palindromique (GATC)
 - Type III, motif court non palindromique (TCAG)

➤ Méthylation de l'ADN chez les bactéries

Rôles de la méthylation

Restriction Modification system (RM)

- Méthyltransférase + Endonucléase sensible à la méthylation
- Protection contre les ADN étrangers

Orpheline

- Régulation du cycle cellulaire
- Régulation de l'expression des gènes
 - compétition avec des activateurs/inhibiteurs
 - Population avec des phénotypes différents

> PacBio

Fonctionnement

Préparation

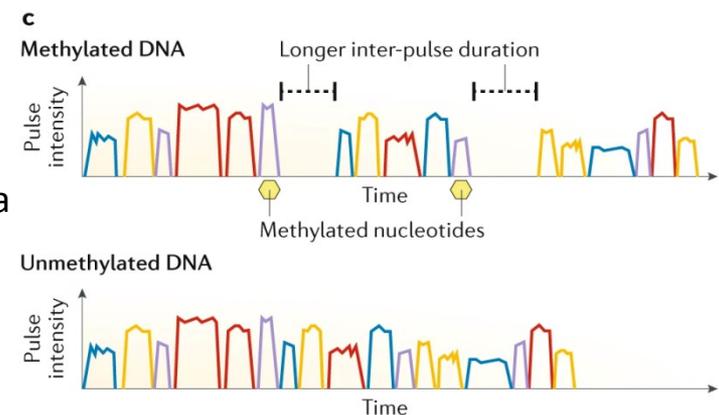
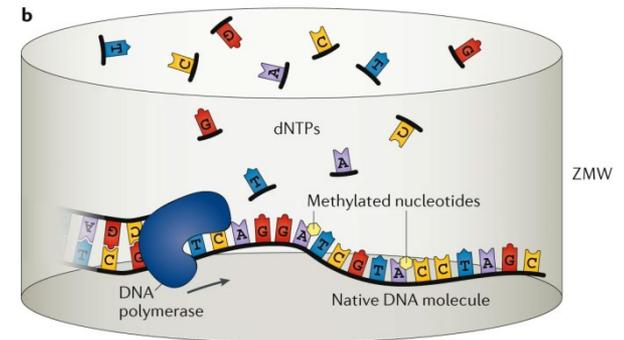
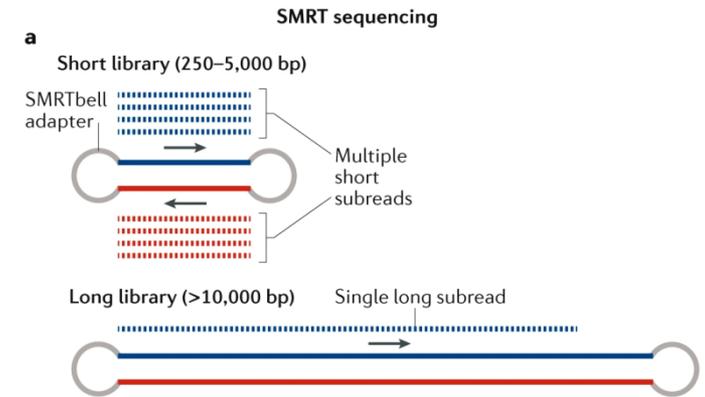
- ADN de haut poids moléculaire
- Taille ADN en fonction des besoins
 - court < 10kb
 - **moyen (HiFi) ~ 20kb**
 - long (CLR) >= 80kb

Séquençage

- 1 molécule d'ADN par puits
- Détection de l'incorporation des dNTP

Détection des modifications

- Comparaison du retard d'incorporation
 - **Natif contre le modèle *in silico* de PacBio**
 - Natif contre WGA/Mutant
- Détection du type de méthylation en fonction de la signature (6mA et 4mC)



> PacBio

Pipeline

Pipeline fourni par PacBio

- Interface graphique et ligne de commande avec SMRT Link
- Besoin du fichier BAM des subreads et d'une référence
- Analyse rapide (~ 2h)

Fichiers en sortie

CSV avec les données pour l'ensemble des bases et par brin

- Difficulté pour fixer un seuil pour considérer une position comme méthylée
 - le score varie avec la couverture
 - ipdRatio est dépendant du type

CSV avec les motifs détectés

- Identification des motifs avec la position de la base modifiée
- Motifs pas toujours propres
 - **CTCGAGC** , HNNNNNNNNN**CTCGAG**, **AGGCCTNNY** (6mA)
 - ACCKGATCH, SGGTANHD
 - **CCTGGAAC** (CCWGG, 5mC)

> PacBio

Résultats

Détection des motifs

- Identifier des motifs sans chercher à connaître l'état de chaque base
- Très bonne sensibilité pour les 6mA et 4mC avec une profondeur assez faible (50x)
- Difficilement utilisable pour le 5mC sauf avec une forte profondeur ($\geq 500x$)
- Validation des motifs un peu artisanale

Profile de méthylation (méthylome)

- Identifier l'état de méthylation de chaque base
- Très bonne sensibilité pour les 6mA et 4mC (80-100x)
- Inutilisable pour le 5mC
 - Conversion TET abandonnée
- Compliqué de trouver un seuil sur les métriques pour fixer l'état de méthylation



➤ Oxford Nanopore

Fonctionnement

Préparation

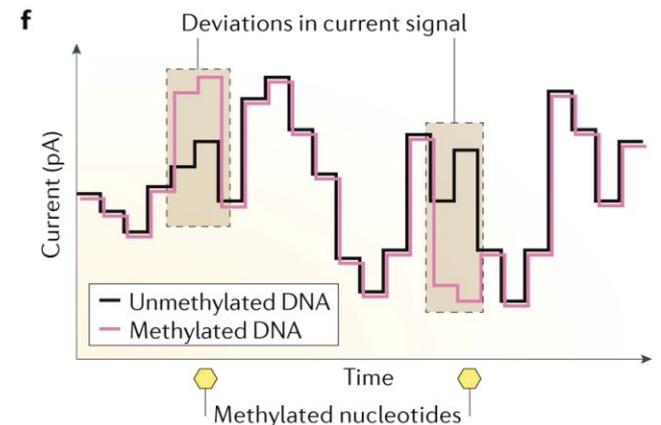
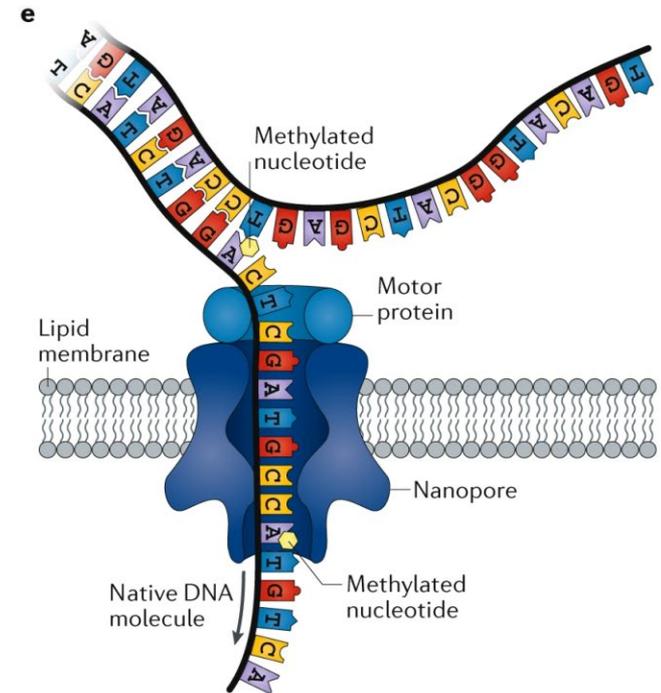
- ADN de haut poids moléculaire

Séquençage

- ADN passe par un pore
- Enregistrement des variations du courant électrique

Détection des modifications

- Comparaison du courant électrique
 - **Natif contre un modèle *in silico***
 - **Natif contre WGA/Mutant**
- Modèle *in silico* à construire à partir d'une comparaison Natif contre WGA/Mutant
- Plusieurs outils fournis par la communauté



➤ Oxford Nanopore

Pipelines: DeepSignal2

DeepSignal / **DeepSignal2**

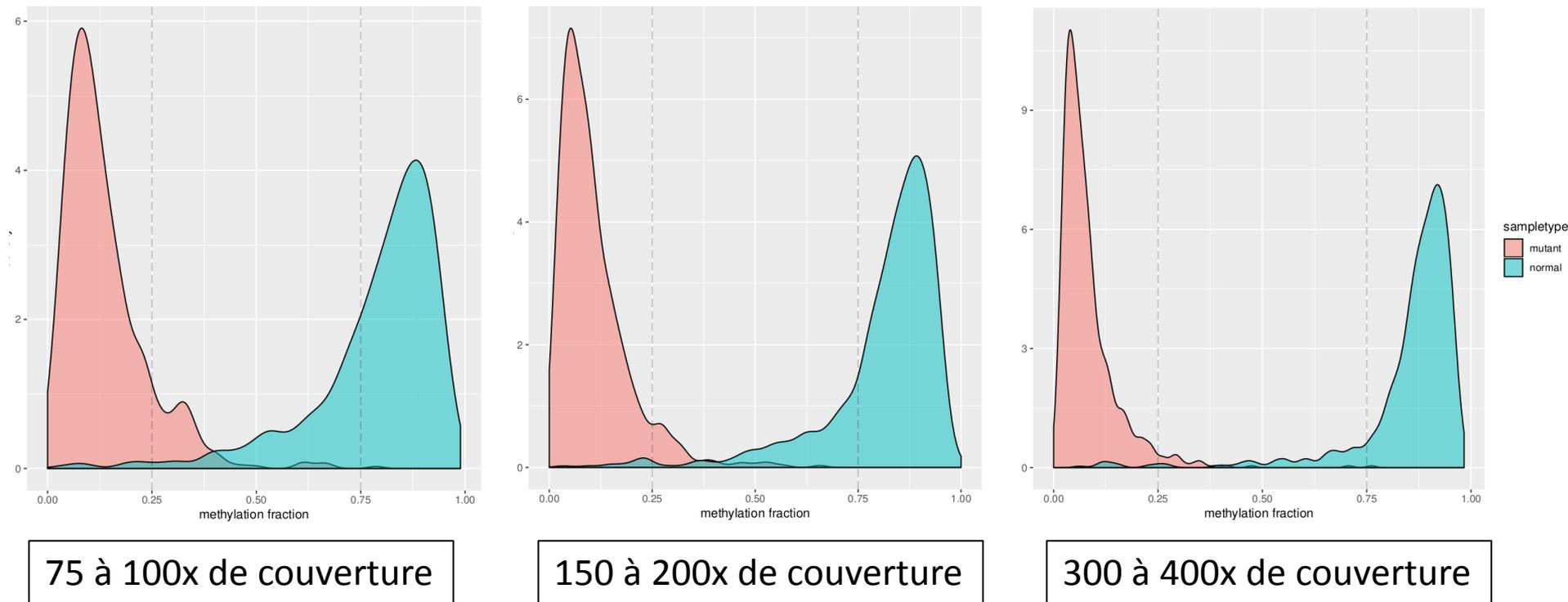
- Séquençage
 - échantillon natif
 - échantillon WGA ou mutant
- Prétraitement des fast5 avec **tombo**
- Extraction des features
 - Signal électrique sur un kmer de 17 bases centré sur le motif
 - échantillon natif est considéré comme méthylé
 - échantillon WGA/Mutant pour un motif considéré comme non méthylé
- Entraînement du modèle
 - training set (60% des données)
 - validation set (40% restant)
- Prédiction des bases méthylées avec le modèle sur l' échantillon natif
- Pipeline assez long et manuel



➤ Oxford Nanopore

Pipeline: DeepSignal2

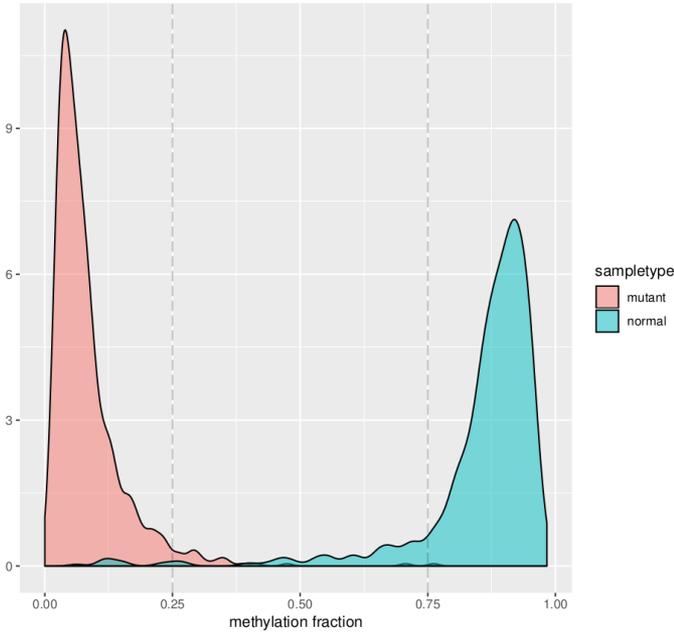
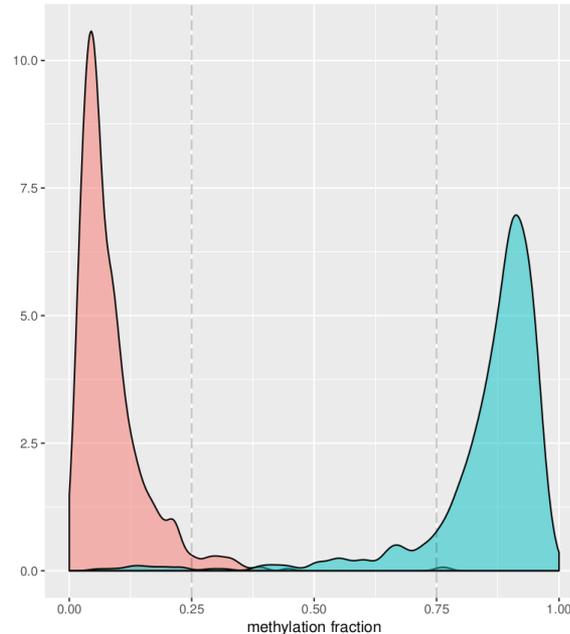
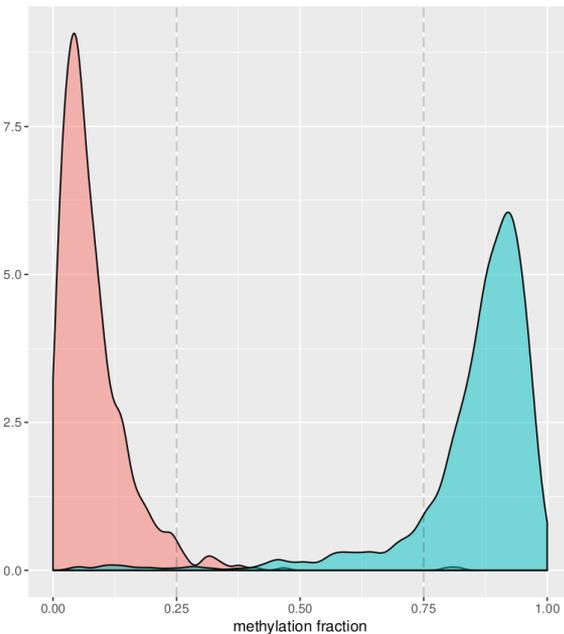
Prédiction des modèles sur l'ensemble des données (training + validation) pour le motif GTWWAC (6mA) chez *Ralstonia solanacearum*



➤ Oxford Nanopore

Pipeline: DeepSignal2

Prédiction du modèle 400x sur les données 100x, 200x et 400x pour le motif GTWWAC (6mA) chez *Ralstonia solanacearum*



75 à 100x de couverture

10	77	697
764	18	2

150 à 200x de couverture

11	74	699
758	24	2

300 à 400x de couverture

10	67	707
760	23	1

➤ Oxford Nanopore

Pipelines

DeepSignal2

- Prédiction de la méthylation 6mA semble bonne avec le modèle 400x
- Le modèle semble réutilisable, mais a valider
- Tester avec un mutant 5mC et un échantillon WGA

Megalodon

- Nécessite de partir depuis le basecalling puis similaire à DeepSignal
- Pas encore testé

Nanodisco

- Pipeline complet sous forme d'image singularity
- Nécessite un échantillon Natif et un WGA/Mutant
- Permet de faire de la prédiction de motif
 - Retrouve bien le motif GTWWAC



➤ WGBS - Bisulfite

Fonctionnement

Préparation

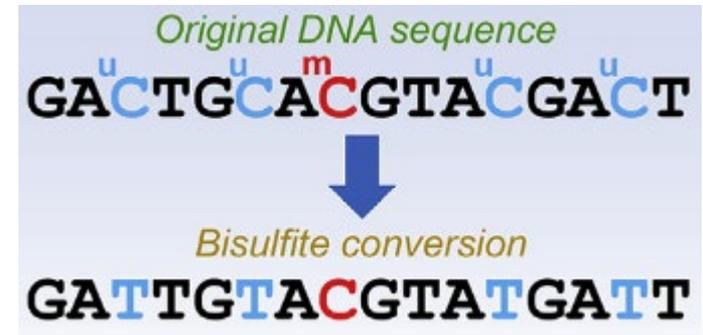
- ADN classique pour du short reads
- Traitement Bisulfite
 - Conversion des C en T
 - Les 5mC restent des C
 - Les 4mC sont partiellement convertis en T
- Abime l'ADN

Séquençage

- Détection de l'incorporation des dNTP par fluorescence

Détection des modifications

- Recherche du meilleur alignement (4 alignements)
- Détermine la méthylation de chaque C en fonction du nombre de T et de C



➤ WGBS - Bisulfite

Pipelines

Pipelines testés

- Bismark
- BWA-meth + MethylDackel
- Extraction des % méthylation pour l'ensemble des C

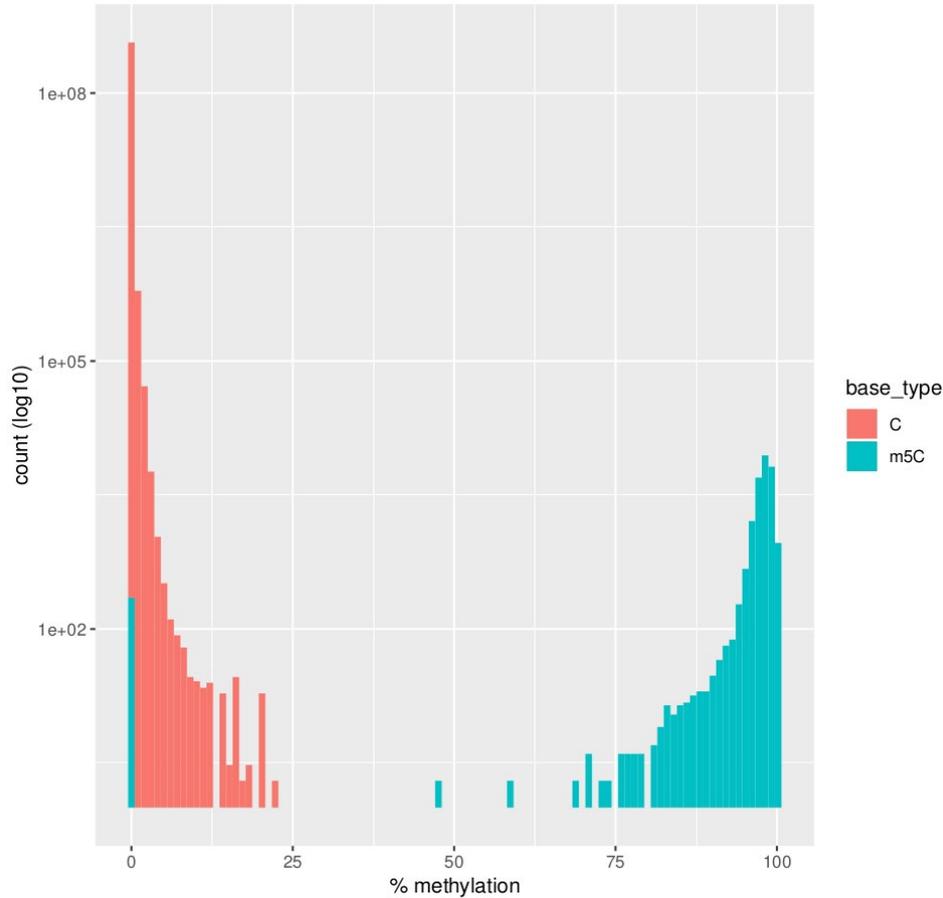
Résultats

- Résultats de méthylation très similaire
- BWA-meth aligne beaucoup plus de reads que Bismark

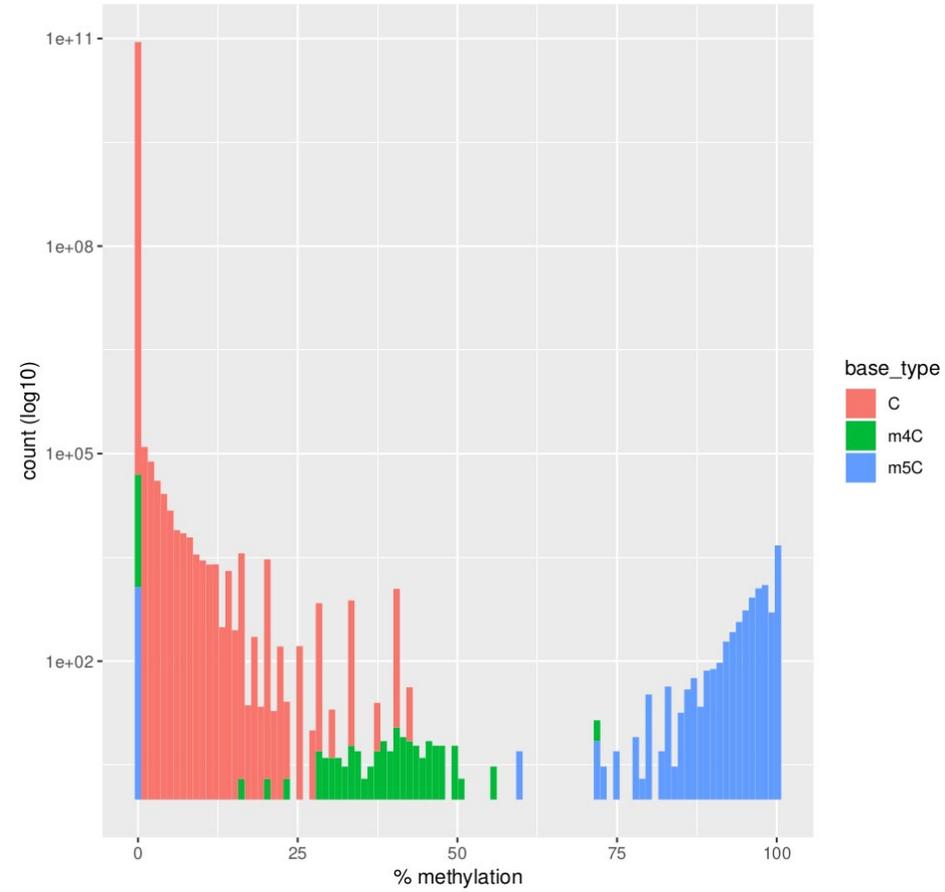
➤ WBGs - Bisulfite

Résultats sur des GC ~50%

E. coli k12



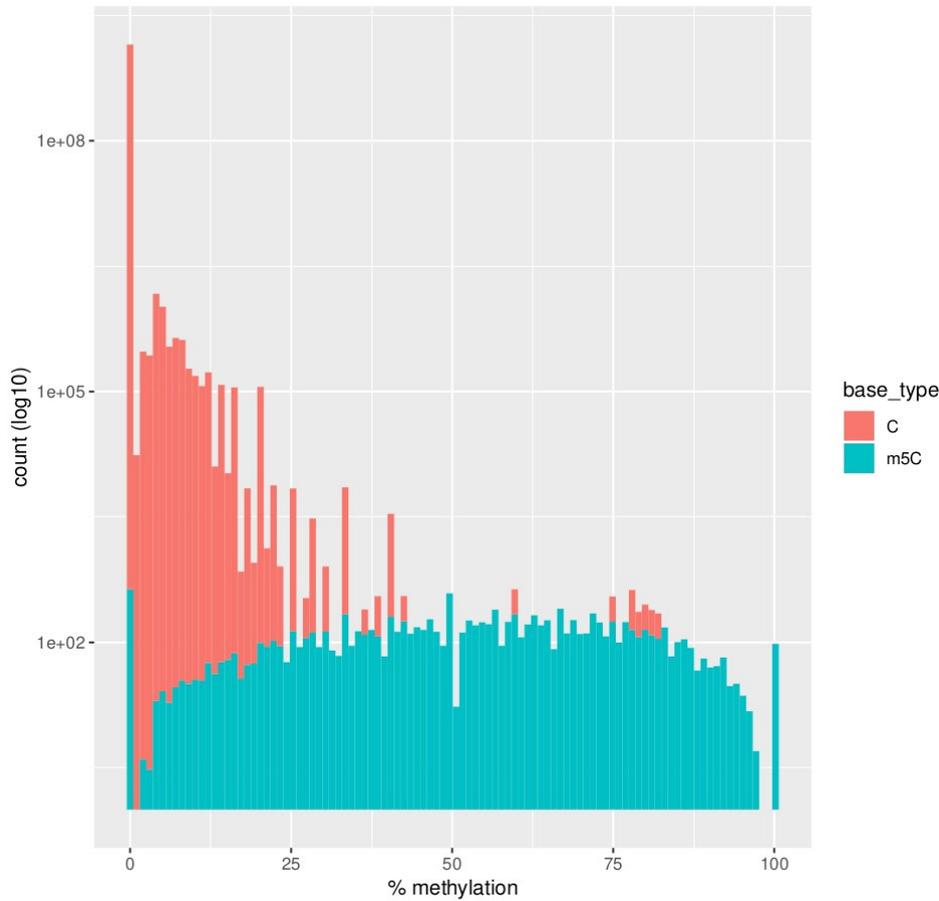
P. luminescens TTO1



➤ WBGs - Bisulfite

Résultats

R. Solanacearum GMI1000



- Semble ne pas être un problème de conversion
- Comportement non attendu pour les 5mC
- Plusieurs manips avec plusieurs kits mais les résultats sont similaires

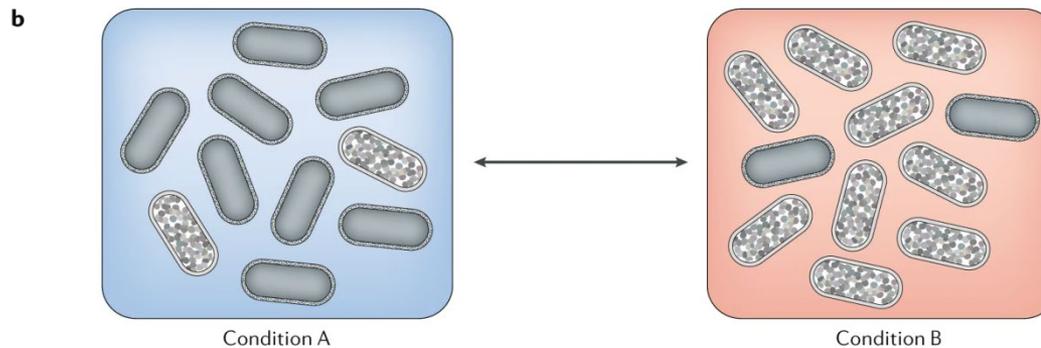
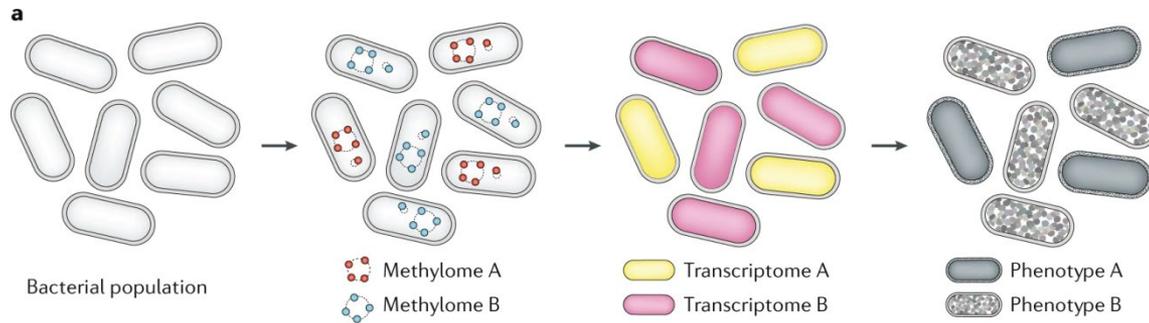
Perspectives

- Tester Oxford nanopore et EM-seq pour les méthylation 5mC



➤ Détection de population

DiscriMet



➤ Détection de populations

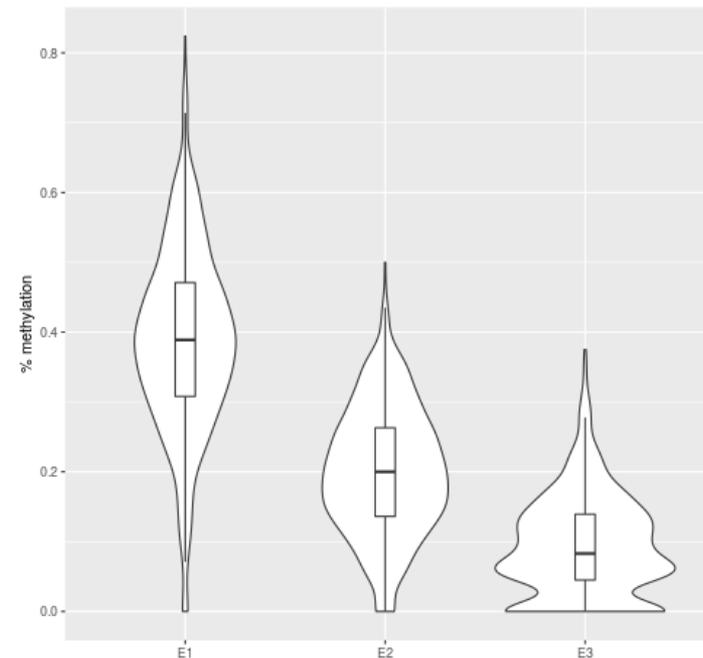
DiscriMet

Pipeline

- Données PacBio avec des fragments court (~2kb sur Sequel I)
 - Objectif : 10 à 20 subreads par fragment
- Analyse de la méthylation puits par puits (single molecule)

Résultats

- E1 : 50% mutant / 50% wt
 - E2 : 75% mutant / 25% wt
 - E3 : 90% mutant / 10% wt
- Fonctionne assez bien globalement mais la dispersion des % méthylation reste élevée



➤ Conclusion & Perspective

Pacbio

- Analyse du méthylome quasiment gratuite
- Bons résultats avec 6mA et 4mC

Oxford Nanopore

- Demande au moins un échantillon contrôle
- Analyse couteuse en temps
- Bons résultats avec 6mA, à évaluer pour 4mC et 5mC.

Bisulfite – WGS

- Séquençage spécifique
- Analyse peu couteuse
- Bons résultats pour 5mC pour certaines bactéries
- Résultats inexploitable pour Ralstonia
- Evaluer EM-seq



LIPME

Alice Guidot
Rekha Gopalan-Nair
Léa Casagrande
Boris Taillefer

DGMI

Julien Brillard
Alain Givaudan
Anne Lanois Nouri
Amaury Payelleville

GeT-Plage

Olivier Bouchez
Céline Lopez-Roques
Céline Vandecasteele
Paul Terzian
Alain Roulet

REacTION

Nadia Ponts
Fabrice Legeai
Stéphanie Robin
Gautier Richard

➤ Annexes

Signatures

5mC

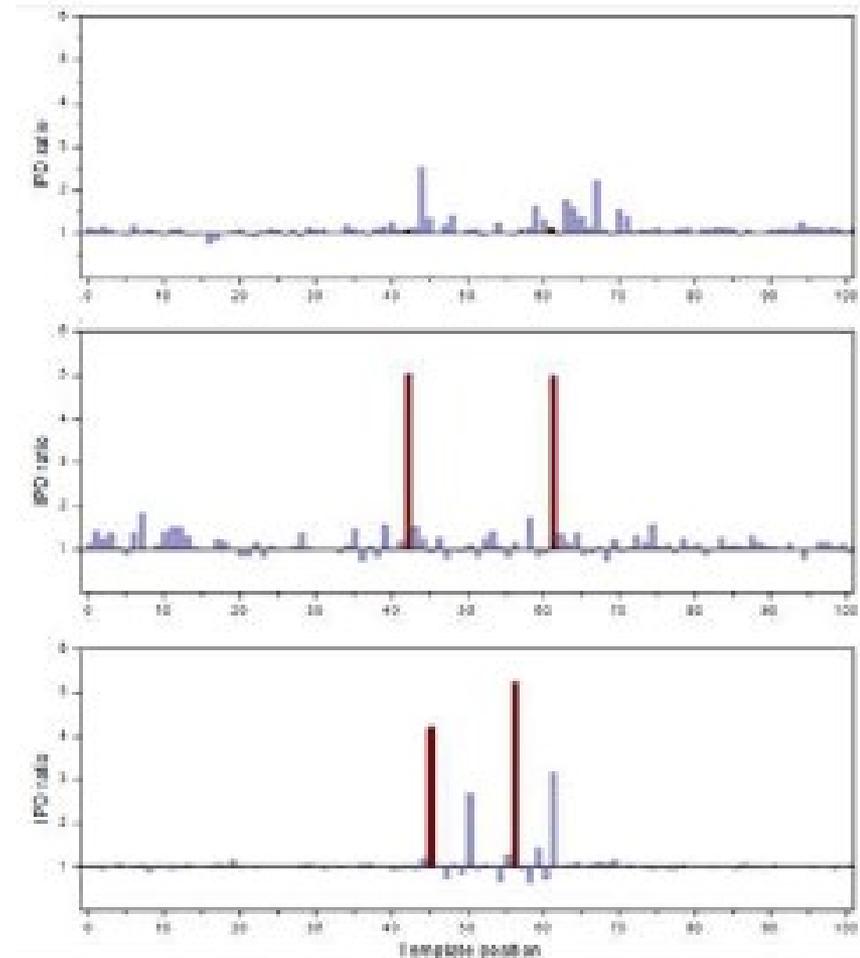
- Retard +2 et +6

4mC

- Retard uniquement sur la position

6mA

- Retard sur la position et +5



➤ Annexes

Modèles et données DeepSignal2

	Training set	Validation set	Total
100x	79 547	53 030	132 577
200x	158 844	105 898	264 742
400x	317 604	211 737	529 341

	Epoch	Accuracy	Precision	Recall
100x	6	0,827	0,8407	0,7899
200x	6	0,8557	0,8685	0,8197
400x	7	0,8873	0,9085	0,8528

modèle	données	AG1 (wt)			mAG4 (mutant GTWWAC)		
		unmeth	unkown	meth	unmeth	unknown	meth
100x	100x features	1,4% (11)	26,15% (205)	72,45% (568)	88,52% (694)	11,35% (89)	0,13% (1)
200x	200x features	1,53% (12)	18,11% (142)	80,36% (630)	93,24% (731)	6,76% (53)	0% (0)
400x	400x features	1,28% (10)	8,55% (67)	90,18% (707)	96,94% (760)	2,93% (23)	0,13% (1)
400x	400x fast5	1,28% (10)	9,18% (72)	89,54% (702)	96,17% (754)	3,83% (30)	0% (0)
400x	200x features	1,4% (11)	9,44% (74)	89,16% (699)	96,68% (758)	3,06% (24)	0,26% (2)
400x	100x features	1,28% (10)	9,82% (77)	88,9% (697)	97,45% (764)	2,3% (18)	0,26% (2)

