

THESE DE DOCTORAT DE

L'UNIVERSITE DE NANTES

ECOLE DOCTORALE N° 605

Biologie Santé

Spécialité : Bio-informatique

Par

Jennifer DERRIEN

L'hétérogénéité épigénétique intra et inter-patient dans le cas du myélome multiple

Thèse présentée et soutenue à Nantes, le 4 septembre 2020.

Unité de recherche : UMR 1232 Inserm – CRCINA – Université de Nantes

Rapporteurs avant soutenance :

Michele GOODHARDT
Laurent GARDERET

Directrice de recherche, CNRS, Université Paris Diderot
Maître de conférences - praticien hospitalier, Sorbonne, Université APHP

Composition du Jury :

Président : **François PARIS**

Directeur de recherche, INSERM UMR1232, CRCINA

Examineurs : **Michele GOODHARDT**
Laurent GARDERET
Denis PUTHIER

Directrice de recherche, CNRS, Université Paris Diderot
Maître de conférences - praticien hospitalier, Sorbonne, Université APHP
Maître de conférences, TAGC (Inserm U1090)

Dir. de thèse : **Stéphane MINVIELLE**
Co-dir. de thèse : **Florence MAGRANGEAS**

Directeur de recherche, CNRS, INSERM UMR1232, CRCINA
Docteur en sciences, CHU de Nantes, CRCINA

Invité : **Carl HERRMANN**

Assistant professor, Medical Faculty, University Heidelberg

Remerciements

En premier lieu, je tiens avant tout à remercier mes directeurs de thèse, Stéphane Minvielle et Florence Magrangeas, qui m'ont accueilli dans leur équipe il y a maintenant un peu plus de 4 ans, lors de mon stage de master 2. Merci d'avoir toujours été là pour me conseiller, m'encadrer et pour tout ce temps précieux que vous m'avez consacré ces dernières années. Merci pour votre grande disponibilité et pour votre pédagogie qui m'a permis de comprendre certains concepts biologiques qui ne m'étaient pas familiers de par ma formation bio-informatique.

J'adresse tous mes remerciements à l'ensemble des membres du jury, de l'honneur qu'ils m'ont fait en acceptant d'évaluer ce travail de thèse. Merci Michele Goodhardt et Laurent Garderet d'avoir accepté de relire cette thèse et d'en être rapporteurs. Merci Denis Puthier et François Paris d'avoir accepté d'être examinateurs de cette thèse.

Merci Carl Hermann, pour ses nombreux conseils sur le traitement et l'analyse de mes données ainsi que pour ses retours pertinents sur l'article.

Merci à Jérémy Bourdon et Pierre-François Cartron d'avoir accepté de faire partie de mon comité de suivi de thèse, et de m'avoir suivi ces 3 dernières années. Chacune de nos réunions m'ont été très bénéfiques pour la suite de mes travaux.

Je remercie également Catherine qui a énormément participé à tous ces travaux de par ses (magnifiques) figures et ses analyses statistiques sur les données. Mais également, merci pour ta patience, tes conseils, tes propositions et idées, mais également ton soutien tout au long de cette thèse. Ces derniers mois, j'aurais pu m'installer dans ton bureau, tellement j'y ai passé de temps. Merci pour ta relecture scrupuleuse de mon manuscrit et pour la pertinence de tes suggestions. Tes petites bulles m'ont énormément aidé et fait sourire en fin de rédaction.

Merci également à tous les membres de mon équipe : Victor, Magalie, Nathalie, Elise, Emilie, Jean-Baptiste, Jonathan, Anthony, Loïc et Benjamin. Pour la bonne ambiance qui règne chaque jour pendant nos pauses cafés et gourmandises (dont j'attends chaque matin avec impatience le signal de départ sur Slack). Un merci un peu plus particulier à Victor, mon co-bureau, qui a su me supporter pendant toutes ces années. Nos conversations (et débats) m'ont très souvent permis d'avancer dans mes recherches, merci pour tous tes conseils, tes explications sur notre tableau blanc et ton organisation à toutes épreuves pendant nos déplacements. Je n'oublie pas également Wilfried qui m'a aidé à débiter l'analyse des données eRRBS en stage de master 2.

Je remercie toutes les personnes avec qui j'ai partagé mes études et notamment mes camarades bio-informaticiens : Ulysse, Clara, Anne-Sophie, Xavier, Kenzo, Leïla et en particulier Dimitri pour nos moments pauses et échanges très instructifs.

A titre plus personnel, mes remerciements vont aussi à ma famille et mes amis, et particulièrement mes parents qui m'ont toujours soutenu (et poussé) dans mon choix de faire des études, qui se sont avérées un peu plus longues et différentes qu'initialement prévues. Merci pour votre confiance et soutien constant, qui m'ont permis d'en arriver là aujourd'hui. Vous avez su me donner toutes les chances pour que je réussisse.

Et pour finir, quelques mots pour ma petite famille. Félix, merci pour tout ce que tu m'apportes, ta patience, ton calme et ton soutien sans faille. Entre la construction de notre maison, l'arrivée de notre fille, et ma thèse prenant beaucoup de place dans notre vie personnelle, ces dernières années ont été riches en accomplissements et émotions. Je ne pouvais m'imaginer un meilleur partenaire que toi pour vivre tout cela.

Ma fille, Nina, ton arrivée a bien chamboulé ma vie, te voir grandir et t'épanouir, chaque jour qui passe, est un pur bonheur.

Nina, Félix, je vous aime.

Liste des abréviations

3C : Capture de Conformation des Chromosomes.

5fC : 5-formylcytosine.

5hmC : 5-hydroxyméthylcytosine.

5mC : Méthyl-cytosine.

ADN : Acide désoxyribonucléique.

ADNc : ADN double complémentaire.

AID : Désaminase induite par activation.

AM-AR : Modification active - retrait actif de la méthylation.

AML : Leucémie myéloïde aiguë.

AM-PD : Modification active - dilution passive de la méthylation.

ARN : Acide ribonucléique.

ARNm : ARN messenger.

ATAC-seq : Assay for Transposase-Accessible Chromatin with highthroughput sequencing.

BER : Base excision repair.

Bmem : Cellule B mémoire.

BrdU : 5-bromo-2-désoxyuridine.

CEBPD : CCAAT Enhancer Binding Protein Delta.

CGI : Îlot CpG.

ChIP-Seq : Chromatin ImmunoPrecipitation Sequencing.

CLL : Leucémie lymphoïde chronique.

CML : Leucémie myélomonocytaire chronique.

CNV : Variation de nombre de copies.

DAVID : Database for annotation, visualization and integrated discovery.

DLBCL : Lymphome diffus à grandes cellules B.

DMV : Vallée de méthylation de l'ADN.

DNase-seq : Séquençage des sites hypersensibles à la DNase I.

DNMT : ADN méthyltransférase.

EMA : Agence européenne des médicaments.

EPM : Eloci par million, nombre d'eloci détectés par échantillon en fonction du nombre total de loci séquencés.

eRRBS : Enhanced Reduced Representation Bisulfite Sequenced.

ESC : Cellule souche embryonnaire.

EZH2 : Enhancer of zeste homologue 2.

FDA : Food and Drug Administration des Etats-Unis.

FDR : Taux de faux positif.

FGFR3 : Récepteur du facteur de croissance des fibroblastes 3.

GC : Cellule B du centre germinal.

GREAT : Genomic regions enrichment of annotations tool.

HDAC : Histone désacétylase.

Hi-C : High Chromosome Contact map.

HMT : Histone méthyltransférase.

HOMER : Hypergeometric optimization of motif enrichment.

HPC : Cellule précurseur hématopoïétique.

ICE : Iterative Correction and Eigenvector decomposition.

IFM : Intergroupe Francophone du myélome.

IgD : Immunoglobuline de type D.

IgM : Immunoglobuline de type M.

KEGG : Kyoto encyclopedia of genes and genomes.

LADs : Lamina-associated domains.

LncRNA : Long ARN non codant.

MAPK : Mitogen-activated protein kinases.

MBD : methyl-CpG binding domain.

MDS : Syndromes myélodysplasiques.

MeDIP : Methylated DNA immunoprecipitation.

MGUS : Gammapathie monoclonale de signification indéterminée.

MiRNA : Micro ARN.

MM : Myélome multiple.

NBC : Cellule B naïve.

NPC : Cellule normale plasmocytaire.

OR : Odd ratio.

Pb : Paire de bases.

PDR : Pourcentage de reads discordants.

piRNA : ARN interagissant avec Piwi (Piwi-interacting RNA).

PMD : Domaine partiellement méthylé.

PRC2 : Polycomb Repressive Complex 2.

Pré-BC : Cellule B précurseur.

Pro-BC : Cellule B progéniteur.

Rd : Revlimid/lénalidomide, dexaméthason.

RFS : Relapse-free survival.

RLE : Relative Log Expression.

RNA-seq : Séquençage de l'ARN.

RPKM : Reads Per KiloBase Per Million Mapped.

RRBS : Reduced Representation Bisulfite Sequencing.

SAM : S-Adénosyl Méthionine.

siRNA : Petit ARN interférant.

SMM : Myélome multiple latent ou asymptomatique.

SNP : Single-nucleotide polymorphism.

snRNA : Petit ARN nucléaire.

SNV : Single Nucleotide Variant.

TAD : Domaine d'association topologique.

TDG : Thymine DNA glycosylase.

TET : Ten Eleven Translocation proteins.

TSS : Site d'initiation de la transcription.

TTS : Site de terminaison de la transcription.

VMP : Velcade/bortézomib, melphalan, prednisone.

VRD : Bortézomib/Velcade, lénalidomide/Revlimid, dexaméthasone.

VTD : Bortézomib/Velcade, thalidomide, dexaméthasone.

WGBS : Whole Genome Bisulfite Sequencing.

WGS : Whole genome sequencing.

Table des figures

Figure 1 : Structure de la chromatine.....	7
Figure 2 : Modifications post-traductionnelles des histones.....	8
Figure 3 : Distribution des marques d’histone dans les régions régulatrices des gènes .	10
Figure 4 : Conversion de la cytosine en 5’-méthylcytosine par l’action des DNMT	11
Figure 5 : Mécanisme de méthylation de l’ADN	12
Figure 6 : Déméthylation passive de l’ADN.....	14
Figure 7 : Le cycle de la déméthylation active de l’ADN.....	15
Figure 8 : Méthylation de l’ADN et régulation de l’expression	16
Figure 9 : Expansion clonale et progression du myélome multiple	22
Figure 10 : Mécanismes de régulation des gènes oncogéniques et des suppresseurs de tumeur dans les cellules normales et les cellules tumorales.....	25
Figure 11 : Evolution clonale et perturbation de la méthylation de l’ADN proposé par Landau et al. (2014).....	26
Figure 12 : Impact du traitement sur la diversité intraclonale.....	30
Figure 13 : Principe de la technologie HumanMethylation 450k Illumina.....	33
Figure 14 : Les différentes étapes du protocole eRRBS.	35
Figure 15 : Comparaison des techniques de séquençage de la méthylation de l’ADN ..	36
Figure 16 : Définition des 25 états de la chromatine par ChromHMM.....	38
Figure 17 : Protocole du HiC.....	39
Figure 18 : Extrait d’un fichier d’information de la méthylation en sortie de methylKit	42
Figure 19 : Détection des loci de 4 CpGs adjacents et mesure de leur épipolymorphisme avec l’outil methclone	43
Figure 20 : Détection des changements épialléliques.....	45
Figure 21 : Définition des reads concordants et discordants	46
Figure 22 : Exemples de distributions α sur le chromosome 21.	47
Figure 23 : Répartition génomique des compartiments A et B pour les lignées cellulaires de la GM12878, U266 et RPMI-8226.....	50
Figure 24 : Exemple de visualisation des cartes de contact de HiC sur 3 lignées cellulaires	51
Figure 25 : Définition des régions promotrices et autour des CGIs	52
Figure 26 : CpG séquencés par échantillon eRRBS.....	54
Figure 27 : Couverture moyenne par échantillon WGBS	57
Figure 28 : Hypométhylation globale et hyperméthylation localisée	58
Figure 29 : Méthylation moyenne des CpGs aux différents stades du MM	59
Figure 30 : Classification ascendante hiérarchique des profils de méthylation des échantillons eRRBS.....	59
Figure 31 : Distribution génomique des loci sur le génome	60
Figure 32 : Niveaux d’épipolymorphisme en fonction de la méthylation moyenne de l’ADN.....	61
Figure 33 : Niveaux d’épipolymorphisme en fonction de la méthylation moyenne de l’ADN dans les promoteurs-CGIs	62
Figure 34 : Méthylation de l’ADN et épipolymorphisme dans les promoteurs-CGIs...	63

Figure 35 : Organisation des loci en fonction de leur fréquence en épiallèles complètement méthylés ou déméthylés	65
Figure 36 : Organisation des loci en fonction de leur fréquence en épiallèles complètement méthylés ou déméthylés pour les échantillons NPC2 et M#19 au diagnostic	66
Figure 37 : Pourcentage de reads partiellement méthylés par échantillon	67
Figure 38 : Hétérogénéité de la méthylation intra-tumorale au diagnostic et à la rechute par patient	68
Figure 39 : Quantification des changements épialléliques et leur impact sur la survie ..	69
Figure 40 : Distribution génomique des eloci	70
Figure 41 : Analyse d'enrichissement en terme ontologique avec l'outil <i>GREAT</i> , des eloci (NPC versus Diagnostic) compris dans des promoteur-CGIs	71
Figure 42 : Perturbation des vallées de méthylation de l'ADN	72
Figure 43 : Évolution de l'EPM au cours de la progression du MM	73
Figure 44 : Evolution des changements épialléliques au cours de la progression du MM et leur impact sur la survie des patients	74
Figure 45 : Les 4 schémas extrêmes de modifications des patterns de méthylation des eloci	75
Figure 46 : Proportion des eloci présentent une sélection d'un pattern de méthylation ..	76
Figure 47 : Localisation des eloci selon les états de la chromatine	77
Figure 48 : Répartition des 20 échantillons de MM selon leur pourcentage d'eloci hypo- et hyperméthylés	77
Figure 49 : Enrichissement significatif des promoteurs bivalents des cellules souches embryonnaires touchés par un elocus comparé aux promoteurs bivalents des tissus différenciés	79
Figure 50 : Évolution de la méthylation de l'ADN des eloci dans les promoteurs bivalents au cours de l'initiation et de la progression du MM	80
Figure 51 : Évolution de l'expression des gènes ayant un elocus dans leur promoteur bivalent	81
Figure 52 : Pattern de méthylation moyen des eloci dans les promoteurs-CGI bivalents ..	82
Figure 53 : Expression des gènes présentant un elocus hyperméthylé ou ne présentant aucun elocus dans leur promoteur-CGI bivalent	83
Figure 54 : Analyse d'enrichissement d'annotations des gènes dont le promoteur-CGI bivalent est touché par des eloci hyperméthylés	83
Figure 55 : Augmentation variable de la méthylation de l'ADN dans les promoteurs des gènes <i>SFRP2</i> et <i>RASD1</i>	84
Figure 56 : Dispersion des eloci hypométhylés en fonction de leur épipolymorphisme dans les échantillons NPC et au diagnostic pour le patient M#09	85
Figure 57 : Dispersion des eloci hypométhylés en fonction de leur épipolymorphisme dans les échantillons NPC et MM au diagnostic pour tous les patients	86
Figure 58 : Distribution α sur le chromosome 21 des échantillons de WGBS	88
Figure 59 : Comparaison des PMDs des cellules normales et des MM	89
Figure 60 : Caractéristiques clés des régions PMDs	90

Figure 61 : Méthylation moyenne de l'ADN dans les PMDs et en dehors	91
Figure 62 : Intensité de H3K9me3 associé à la longueur des PMDs et à leur déméthylation	91
Figure 63 : Heatmap des PMDs triés par taille croissante, associés au temps de réplication de la GM12878 et des niveaux de méthylation de l'ADN des échantillons de WGBS ..	92
Figure 64 : Intensité de H3K27me3 associé à la déméthylation des PMDs	93
Figure 65 : Variabilité de l'expression des gènes et de la déméthylation dans les PMDs	94
Figure 66 : Analyse fonctionnelle des gènes dans des PMDs proches d'une borne	95
Figure 67 : Méthylation des régions régulatrices dans les PMDs déméthylés	96
Figure 68 : Diminution de la méthylation dans les PMDs et compartimentation du génome	99
Figure 69 : Corrélation entre l'EPM et la méthylation de l'ADN dans les PMDs	101
Figure 70 : Moyenne de la méthylation de l'ADN des PMDs dans les cellules B à différents stades de différenciation	102
Figure 71 : Hétérogénéité transcriptionnelle due au marquage simultané de H3K27ac et H3K27me3 dans le cas de la leucémie lymphoïde chronique	105

Liste des tableaux

Table 1 : Description des données eRRBS 55

Table 2 : Répartition génomique des CpGs séquencés en eRRBS..... 56

Table 3 : Comparaison du nombre de CpGs couverts par les techniques de séquençage
eRRBS et WGBS..... 56

Sommaire

1	Introduction	1
2	Le myélome multiple	2
2.1	Epidémiologie	2
2.2	Physiopathologie	2
2.3	Symptômes et diagnostic	3
2.3.1	Douleurs osseuses	3
2.3.2	Hypercalcémie	3
2.3.3	Insuffisance rénale	4
2.3.4	L'anémie, infections et saignements anormaux	4
2.3.5	Diagnostic du myélome	4
2.4	La prise en charge de la maladie	4
2.5	La rechute	5
3	Les modifications épigénétiques	6
3.1	Organisation de la chromatine	6
3.2	Les modifications post-traductionnelles des histones	8
3.3	La méthylation de l'ADN	11
3.3.1	Les ADN méthyltransférases	12
3.3.2	La déméthylation	13
3.3.3	Les inhibiteurs des DNMTs	15
3.3.4	Régulation de l'expression par la méthylation	16
3.3.5	Évolution des profils de méthylation au cours du temps	17
3.4	Les ARNs non codants	18
4	Hétérogénéité épigénétique et génétique dans le cancer	19
4.1	L'hétérogénéité génétique inter-patient	19
4.2	L'évolution clonale comme modèle de progression du MM	21
4.3	L'hétérogénéité génétique intra-patient	22
4.4	Hétérogénéité de la méthylation	24
4.5	Thérapie épigénétique	29
4.6	L'objectif de la thèse	30
5	Échantillons et techniques utilisées	32
5.1	Description des échantillons	32
5.2	Les différentes méthodes d'étude de la méthylation de l'ADN	32
5.2.1	Les puces à ADN, HumanMethylation 450k d'Illumina	32
5.2.2	Methylated DNA immunoprecipitation	34
5.2.3	Enhanced Reduced Representation Bisulfite Sequencing	34
5.2.4	Whole-genome bisulfite sequencing (WGBS)	35
5.2.5	Comparaison des techniques	36
5.3	Le séquençage de l'ARN (RNA-seq)	37

5.4	Caractérisation des états de la chromatine	38
5.4.1	ChIP-Seq et ChromHMM	38
5.4.2	Hi-C	39
5.4.3	Repli-Seq	40
5.5	Analyse bioinformatique des données	41
5.5.1	Analyse des données eRRBS	41
5.5.1.1	Contrôle qualité et filtration des données	41
5.5.1.2	Alignement contre le génome de référence	41
5.5.1.3	Extraction des CpGs et leurs niveaux de méthylation	42
5.5.1.4	Mesure de l'hétérogénéité de la méthylation	43
5.5.1.5	Identification des changements significatifs en composition épiallélique	44
5.5.2	Analyse des données WGBS	46
5.5.3	Analyse des données de RNAseq	47
5.5.4	Analyse des données d'HiC	49
5.5.4.1	Détermination des compartiments A et B	49
5.5.4.2	Détermination des domaines d'associations topologiques	50
5.5.5	Annotation génomique	52
5.5.6	Annotation fonctionnelle	53
6	Résultats	54
6.1	Description des données	54
6.1.1	Les données eRRBS	54
6.1.2	Les données WGBS	56
6.2	Perturbation globale de la méthylation de l'ADN dans le MM	57
6.3	L'hétérogénéité intra-tumorale de la méthylation dans le MM	60
6.3.1	Méthylation de l'ADN plus désordonnée dans le MM	60
6.4	Les changements significatifs de patterns de méthylation	68
6.4.1	Impact clinique des changements épialléliques	69
6.4.2	Perturbation des gènes impliqués dans le développement	70
6.4.3	Evolution des changements épialléliques au cours de la progression du MM	73
6.5	Impact des changements significatifs de la méthylation sur l'organisation de la chromatine	76
6.5.1	Les promoteurs bivalents	78
6.5.2	Perte de la méthylation de l'ADN dans les domaines partiellement méthylés	85
6.5.3	Perturbation des régions promotrices dans les PMDs hypométhylées	94
6.5.4	Réorganisation de la chromatine due aux perturbations de la méthyla- tion dans les PMDs	97
7	Discussion	100

8 Conclusion et perspectives	106
9 Annexes	127

1 Introduction

Le développement des nouvelles technologies de séquençage à haut débit et l'accès au séquençage du génome entier ont révélé que le niveau de complexité d'un organisme n'est pas lié à son nombre de gènes, ni à la taille du génome. Un exemple parlant, les mammifères et les vers de terre ont un génome comprenant environ 20 000 gènes chacun, alors que leurs comportements et morphologies sont bien distincts. Plusieurs paramètres contribuent à l'augmentation de la diversité, tels que l'étendue des modifications protéiques, les modifications épigénétiques ou encore la diversité des modèles d'épissage. Les données de séquençage complet du génome humain montrent que les régions codantes pour des protéines représentent moins de 5% du génome, et parmi les régions non codantes, 40% seraient des séquences présentant un potentiel régulateur, telles que les enhancers et les promoteurs. Les promoteurs sont des régions d'ADN situées en amont des gènes, sur lesquelles se fixent l'ARN polymérase afin de permettre la transcription de l'ADN en ARN. La fixation et l'activation de l'ARN polymérase sont contrôlées par des facteurs de transcription qui se fixent au niveau de séquences cis-régulatrices spécifiques appelées enhancers, qui se situent généralement à environ 50kb du gène cible. Les enhancers sont tissu spécifique et sont généralement des régions dites pauvres en CpGs contrairement aux promoteurs qui eux sont caractérisés par un enrichissement en CpGs.

Le cancer résulte d'une expression anormale de gènes ou de réseaux de gènes qui entraîne une transformation progressive d'une cellule normale en une cellule maligne. Cette expression aberrante résulte de modifications génétiques qui activent des oncogènes, répriment des gènes suppresseurs de tumeur, altèrent la fonctionnalité de séquences régulatrices, et provoquent des changements au niveau de l'information épigénétique comme les modifications de la chromatine, la méthylation de l'ADN ou l'expression des ARN non codants. Une certaine variabilité inter et intra-patient est couramment décrite dans de nombreux cancers. Tous les patients atteints d'un même type de cancer ne présentent pas tous les mêmes anomalies génétiques ou épigénétiques. Et même au sein d'une même tumeur chez un individu, il existe un ensemble de sous-populations distincts de cellules anormales, ce que l'on appelle des sous-clones. L'hétérogénéité génomique intra-tumorale a été bien décrite dans de nombreux cancers tels que le cancer du pancréas, de la prostate, les cancers colorectaux et pulmonaires. L'hétérogénéité intra-tumorale peut avoir des implications cliniques importantes telles qu'engendrer une résistance thérapeutique ou encore la récurrence de la maladie. Dans plusieurs cas, il a été montré, suite à une pression thérapeutique, une réorganisation et une diversification des sous-populations clonales ou une sélection d'un sous-clone. Bien que la sélection clonale par des aberrations génétiques dans le cancer soit bien documentée, la capacité des altérations épigénétiques à favoriser l'évolution tumorale l'est beaucoup moins. Les projets de notre équipe de recherche visent à mieux comprendre la contribution de ces mécanismes au cours de l'initiation et de la progression du myélome multiple (MM).

2 Le myélome multiple

2.1 Epidémiologie

Le MM est un cancer relativement rare avec environ 5 000 nouveaux cas enregistrés par an en France. Il représente environ 15% des tumeurs malignes hématologiques et moins de 2% des cas de cancers.

Il touche davantage les hommes (52%) que les femmes (48%) (INCa cancers en France 2017), et atteint principalement des personnes âgées de plus de 60 ans. Au diagnostic, l'âge médian des hommes est de 72 ans et l'âge médian des femmes est de 75 ans. Cette maladie est inexistante chez l'enfant et très rare chez les personnes plus jeunes, seulement 3% des cas sont diagnostiqués avant 40 ans.

Malgré l'évolution des traitements au cours de ces dernières années, le MM reste encore aujourd'hui incurable. La survie peut varier de quelques mois à plus de 10 ans, et la trajectoire de la maladie varie pour chaque patient. Les rechutes restent encore inévitables. Le MM est connu pour être très hétérogène aussi bien sur le plan clinique que pathologique, ce qui entraîne une grande variabilité de la réponse au traitement et de la survie des patients.

2.2 Physiopathologie

Le MM est un type de cancer du sang incurable qui survient dans la moelle osseuse, affectant les plasmocytes qui sont des lymphocytes B différenciés capables de produire des anticorps, dont le rôle est de reconnaître tout élément étranger à l'organisme et de le neutraliser.

Dans la moelle osseuse, les cellules B se différencient à partir de cellules précurseurs hématopoïétiques (HPC) impliquant de nombreux réarrangements de gènes codants pour les domaines variables des chaînes lourdes puis des chaînes légères des immunoglobulines à la membrane des cellules, cela en l'absence d'antigène étranger. Au cours de ces réarrangements, les HPC passent par plusieurs stades de maturation. Les réarrangements de la chaîne lourde débutent au stade cellule B progéniteur (pro-BC) et s'achèvent au stade cellule B précurseur (pré-BC). Puis les réarrangements de la chaîne légère débutent au stade pré-BC et s'achèvent avec la différenciation des lymphocytes B immatures, qui présentent à leur surface des immunoglobulines de type M (IgM). La cellule B immature subit une sélection négative permettant d'éliminer par apoptose les lymphocytes B possédant des immunoglobulines spécifiques des antigènes du soi.

Les cellules B immatures quittent la moelle osseuse, passent dans la circulation sanguine et gagnent les organes lymphoïdes secondaires, où les lymphocytes B immatures acquièrent l'immunoglobuline D (IgD) de surface pour devenir matures. Les cellules B matures présentent donc à leur surface des IgM et des IgD possédant la même spécificité antigénique. L'étape suivante de maturation dépend de la rencontre avec un antigène spécifique du récepteur du lymphocyte B. Si la cellule B mature interagit avec un antigène, les lymphocytes B sont activés et se différencient en cellules B mémoires ou en plasmablastes. Ces derniers quittent le centre germinatif et migrent vers la moelle osseuse où ils deviennent des cellules plasmocytaires à longue durée de vie qui produisent des anticorps. Les cellules B mémoires sont capables de se

souvenir d'une rencontre avec un antigène les ayant déjà activées, afin de créer une réponse immunitaire plus rapide dans le cas d'une seconde rencontre avec ce même antigène. Ils sont capables de proliférer et de se différencier en plasmocytes. Les plasmocytes, au contraire des cellules B mémoires, ne présentent pas leurs anticorps au niveau de leur surface, mais sécrètent des anticorps solubles qui vont se fixer sur l'antigène, et le neutraliser. Les cellules plasmocytaires, elles, sont incapables de proliférer dans une situation non pathologique.

Le myélome multiple se manifeste par une prolifération incontrôlée de cellules plasmocytaires anormales qui produisent alors un anticorps monoclonal appelé protéine M, qui contrairement aux anticorps normaux, n'apporte aucun bénéfice pour l'organisme. Les plasmocytes malins prennent peu à peu la place des autres plasmocytes normaux, de ce fait, la production des autres anticorps, nécessaires pour assurer la défense de l'organisme diminue.

Le développement du myélome multiple est un processus à plusieurs étapes, commençant par des états pathologiques précurseurs, tels qu'une gammapathie monoclonale de signification indéterminée (MGUS) et un myélome multiple latent ou asymptomatique (SMM), qui évolue vers un myélome multiple actif cette fois-ci symptomatique. Dans les stades terminaux, le MM devient extramédullaire, c'est à dire indépendant de la moelle osseuse, on parle alors de leucémie à plasmocytes (Kumar et al., 2017).

2.3 Symptômes et diagnostic

Le diagnostic d'un MM peut être réalisé face à plusieurs symptômes notamment osseux et rénaux, décrits ci-dessous, ou alors suite à des examens biologiques ou radiologiques fortuits.

2.3.1 Douleurs osseuses

Les douleurs osseuses, pouvant aller jusqu'aux fractures osseuses, sont les symptômes les plus courants et précoces du myélome, plus de deux tiers des patients en souffrent (?). En temps normal, les os se renouvellent continuellement dans notre corps grâce au travail de deux types de cellules : les ostéoblastes qui forment les os, et les ostéoclastes qui les détruisent. Dans le cas du MM, le développement des douleurs et lésions osseuses sont dûs à un déséquilibre dans l'activité de ces deux types cellulaires. Les cellules myélomateuses sécrètent des cytokines qui stimulent les ostéoclastes et inhibent les ostéoblastes. Ainsi, les ostéoclastes sont plus activés que les ostéoblastes, ce qui conduit à une résorption osseuse plus rapide que la formation osseuse, et donc à une fragilité des os, amenant à la formation de "trous", les os finissent par se briser ou se fracturer.

2.3.2 Hypercalcémie

L'hypercalcémie correspond à une élévation du taux de calcium dans le sang, ceci touche environ 10% des patients. L'hypercalcémie est directement liée à la résorption osseuse, qui provoque une libération accrue de calcium dans le sang. Ceci induit chez le patient une perte de poids, des nausées, des vomissements et des troubles du rythme cardiaque.

2.3.3 Insuffisance rénale

L'insuffisance rénale correspond à un mauvais fonctionnement des reins et est un des symptômes les plus fréquents, cette complication survient chez 20 à 40% des patients (Dimopoulos et al., 2008). Dans 10% des cas, l'insuffisance rénale est suffisamment grave pour nécessiter le remplacement d'un rein avec dialyse au moment du diagnostic.

En temps normal, le rein est chargé de filtrer le sang de toutes ses impuretés, et de les éliminer dans les urines. Dans le cas du myélome, le rein est soumis à de nombreuses agressions, la première cause d'insuffisance rénale est imputable à la prolifération plasmocytaire qui induit une sécrétion en grande quantité de chaînes légères, fragments d'immunoglobuline monoclonale. Leur quantité dépasse les capacités de réabsorption du rein, les chaînes légères des immunoglobulines anormales viennent se précipiter au niveau des tubes distaux et créent des dépôts qui obstruent et réduisent le débit de filtration des reins, et compromettent ainsi ses fonctions. L'hypercalcémie est la deuxième cause d'insuffisance rénale chez les patients.

2.3.4 L'anémie, infections et saignements anormaux

Le myélome se manifeste par une prolifération incontrôlée de plasmocytes anormaux, qui remplacent progressivement les cellules sanguines normalement présentes dans la moelle osseuse. Cela entraîne une diminution du nombre de globules rouges : le patient peut donc être anémique, un faible nombre de plaquette : le patient peut saigner plus facilement, et un faible nombre de globules blancs : le patient peut contracter plus facilement des infections. L'anémie est très souvent le premier symptôme du MM (Mittelman, 2003).

2.3.5 Diagnostic du myélome

Le MM peut être difficile à diagnostiquer, très souvent, le myélome n'est accompagné de symptômes qu'à partir d'un stade avancé. Souvent, on le diagnostique de façon fortuite suite à un résultat anormal lors d'un examen sanguin (présence d'une quantité anormalement élevée de protéines dans le sang). Afin d'affirmer de façon certaine le diagnostic, il est nécessaire de réaliser des analyses sanguines et urinaires, ainsi qu'une biopsie de la moelle osseuse (myélogramme), des tests d'imagerie afin de détecter d'éventuelles atteintes osseuses peuvent venir compléter le diagnostic.

2.4 La prise en charge de la maladie

Lorsque le diagnostic du myélome n'est accompagné d'aucun symptôme, le traitement n'est pas systématique. Toutefois une surveillance régulière du patient sera mise en place afin de suivre l'évolution de la maladie, avec notamment des bilans sanguins, urinaires et radiologiques. Une fois que les symptômes apparaissent, le choix du traitement va être fonction de l'âge du patient, de son état général (maladie pré-existante ou non) et de l'expression du cancer (forme indolente ou agressive), dans le but de contenir son évolution, de prendre en charge les signes de la maladie et d'améliorer la qualité de vie du patient.

Les patients présentant des symptômes vont être traités différemment selon leur âge. En France, l'âge limite est fixé à 65 ans pour bénéficier d'une autogreffe en première ligne de traitement.

Les malades de moins de 65 ans sont éligibles à l'autogreffe qui consiste en un prélèvement des cellules souches de leur moelle osseuse ou dans leur sang, après avoir subi un traitement d'induction de 4 cycles impliquant un inhibiteur du protéasome, un immunomodulateur et de la dexaméthasone (VRD : bortézomib/Velcade, lénalidomide/Revlimid, dexaméthasone, ou VTD : bortézomib/Velcade, thalidomide, dexaméthasone), dans le but de réduire la charge tumorale. A l'issue des 4 cycles d'induction, les cellules souches périphériques sont prélevées sur le patient sont triées et congelées. Ensuite le patient reçoit une chimiothérapie intensifiée par melphalan, au cours de laquelle l'ensemble des cellules de sa moelle osseuse sont détruites, puis ses cellules souches préalablement collectées lui sont réinjectées par transfusion 48h plus tard. Cette ré-injection de cellules souches hématopoïétiques va aider la moelle osseuse à se régénérer. Dans le cadre d'essais thérapeutiques, l'allogreffe peut être envisagée (prélèvement de cellules souches saines chez un donneur compatible). Suite à l'autogreffe, un traitement d'entretien est mis en place en deux phases : un traitement de consolidation, deux mois après la greffe qui consiste en deux cycles d'un traitement identique à celui de l'induction, puis un traitement de maintenance, le lénalidomide par voie orale pendant environ 2 ans.

Pour les malades de plus de 65 ans, si le patient ne présente pas de problème cardiaque, rénal ou pulmonaire, une autogreffe peut être réalisée en routine jusqu'à 70 ans. Pour les patients non éligibles à l'autogreffe, le traitement de référence reste une chimiothérapie classique, la stratégie actuelle repose sur l'association : Velcade/bortézomib - melphalan - prednisone (VMP) ou Revlimid/lénalidomide - dexaméthasone (Rd). Le traitement Rd est un traitement oral, plus simple pour les patients et nécessite moins de visite à l'hôpital.

En parallèle, le traitement des symptômes est également assuré, afin d'améliorer au mieux la qualité de vie du patient. Une radiothérapie peut soulager les douleurs osseuses en traitant les lésions osseuses, ou alors un traitement médicamenteux peut limiter la perte osseuse en renforçant les os. En cas d'insuffisance rénale, le patient peut avoir besoin de dialyse. L'anémie peut être traitée en changeant l'alimentation du patient ou par la prise d'un supplément en fer.

Suite au traitement, le patient est suivi à vie, un bilan sanguin et urinaire est réalisé tous les 3 à 6 mois, et un bilan radiologique tous les 6 mois. Ce suivi permettra une intervention rapide lors de la reprise de l'évolution de la maladie.

2.5 La rechute

Malgré l'évolution des traitements au cours de ces dernières années, dans le but de mieux contrôler l'évolution de la maladie et d'améliorer la vie des patients, le myélome multiple reste encore incurable (Mateos and San Miguel, 2013). Après un temps variable selon les patients, pouvant aller de quelques mois à plus de 10 ans, le MM tend à refaire surface, dû à l'émergence de sous-populations de cellules plasmocytaires myélomateuses plus agressives, résistantes aux agents thérapeutiques. Dans ce cas, la mise en place d'un nouveau traitement, différent du premier, est nécessaire ; on parle alors de traitement de deuxième ligne, dont le

choix sera dicté par plusieurs paramètres, tels que la durée de la première réponse, les maladies associées et le souhait du patient pour préserver sa qualité de vie. Plusieurs molécules sont disponibles : des anticorps anti-CD38 (daratumumab et isatuximab) et des inhibiteurs du protéasome (carfizomib et ixazomib). Au fil des rechutes, le patient a tendance à répondre de moins en moins bien aux traitements, et les périodes de rémission se raccourcissent.

3 Les modifications épigénétiques

Le terme épigénétique dérive d'une préposition grecque "epi" qui signifie "au dessus", "en plus de" la génétique. La notion de paysage épigénétique a été introduit par Conrad Waddington (1905-1975) en 1942 pour décrire des événements impossibles à expliquer par des principes génétiques, "la branche de la biologie qui étudie les interactions causales entre les gènes et leurs produits, ce qui donne naissance au phénotype" (Waddington, 2012). Toutes les cellules de notre organisme possèdent le même patrimoine génétique, mais elles ne l'expriment pas de la même façon, elles vont se différencier de différentes manières pour former des organes différents, aux fonctions différentes. Certains gènes vont être exprimés dans certaines cellules et pas dans d'autres, tous les gènes d'un organisme ne peuvent être actifs dans chaque cellule à tout moment, une cellule doit faire taire l'expression de gènes spécifiques à d'autres types cellulaires pour assurer son destin et ses fonctions.

Dans sa définition moderne, l'épigénétique correspond à des modulations héritées d'expression de gènes qui interviennent sans qu'il n'y ait de modification dans la séquence d'ADN. Ces modifications ne peuvent être imputées à des modifications de séquences d'ADN (mutation, gains et pertes de matériel chromosomique, etc...), et peuvent être réversibles (Berger et al., 2009). Il existe différents mécanismes épigénétiques permettant de réguler l'accès de la machinerie transcriptionnelle et l'état de la chromatine. L'accessibilité des gènes et donc leur expression dépend de différents facteurs et notamment dépend de leur état chromatinien.

3.1 Organisation de la chromatine

Dans une cellule humaine, la chromatine correspond à l'état dans lequel l'ADN est empaqueté dans le noyau. Le nucléosome est l'unité élémentaire de la chromatine, il est composé de deux sous-unités identiques, chacune contenant quatre histones (H2A, H2B, H3 et H4) qui s'associent pour former un octamère autour duquel 147pb d'ADN sont enroulés, ceci correspondant au premier niveau de compaction de l'ADN (Figure 1). Une cinquième histone, l'histone H1, permet une compaction des réseaux de nucléosomes, formant ainsi des fibres de chromatine et correspondant au 2ème degré de condensation (Woodcock and Ghosh, 2010). La chromatine sert à protéger l'ADN et à réguler son accès à des facteurs, protéines de liaisons à l'ADN, jouant un rôle central dans des processus tels que la transcription, la réplication, la recombinaison et la réparation de l'ADN entre autres.

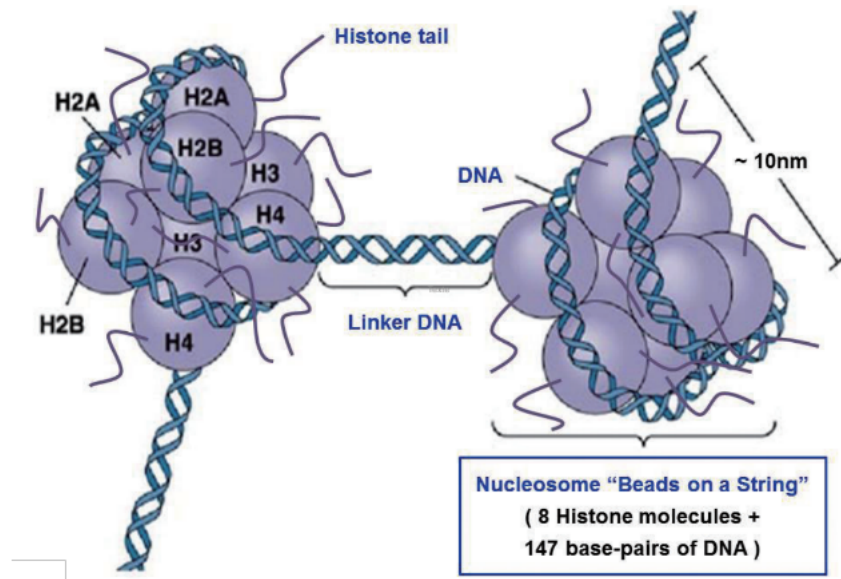


FIGURE 1 – Structure de la chromatine. Figure tirée de Kim (2014).

Les complexes d'histones permettent de réguler l'accessibilité de l'ADN en favorisant la condensation ou relaxation de la chromatine. Dans les génomes eucaryotes, nous distinguons deux types de compartiments :

- l'euchromatine, région ouverte, peu condensée, contenant les gènes actifs (les séquences promotrices sont accessibles à la machinerie transcriptionnelle)
- l'hétérochromatine, région dense, plus compacte, contenant principalement l'ADN inactif.

L'hétérochromatine peut se diviser en deux types :

- l'hétérochromatine constitutive correspondant à des régions génomiques continuellement condensées, ce qui empêche toute transcription, se situant autour des centromères et des télomères, en général ce sont des séquences répétées.
- l'hétérochromatine facultative correspondant à des régions qui peuvent être sous forme hétérochromatine ou euchromatine. La condensation et la décondensation de la chromatine est possible, ce qui permet l'expression temporaire de gènes. Son état est réversible.

Les protéines d'histone, et en particulier leur queue "N-terminale", subissent des modifications post-traductionnelles qui peuvent entraîner une compaction ou une relaxation de la chromatine. Certaines modifications vont perturber les interactions histone-ADN, ce qui va provoquer le déroulement des nucléosomes, une ouverture de la chromatine, permettant la liaison de facteurs de transcription recrutant l'ARN polymérase II et permettant ainsi la transcription de l'ADN (Turner, 2005). À l'inverse d'autres modifications vont renforcer les interactions histone-ADN conduisant à une structure de la chromatine plus compacte, empêchant l'accès des facteurs et de la machinerie transcriptionnelle à l'ADN.

3.2 Les modifications post-traductionnelles des histones

La chromatine est une structure dynamique permettant de réguler l'accès à l'ADN. Les modifications post-traductionnelles des histones jouent un rôle clé dans cette régulation. Ces modifications se produisent principalement sur la queue de l'extrémité N-terminale des histones. Il en existe de nombreux types, les plus courantes sont l'acétylation (ajout d'un groupement acétyle), la phosphorylation (ajout d'un phosphate), la méthylation (ajout d'un groupement méthyle) et l'ubiquitinylation (ajout d'ubiquitine) (Figure 2).

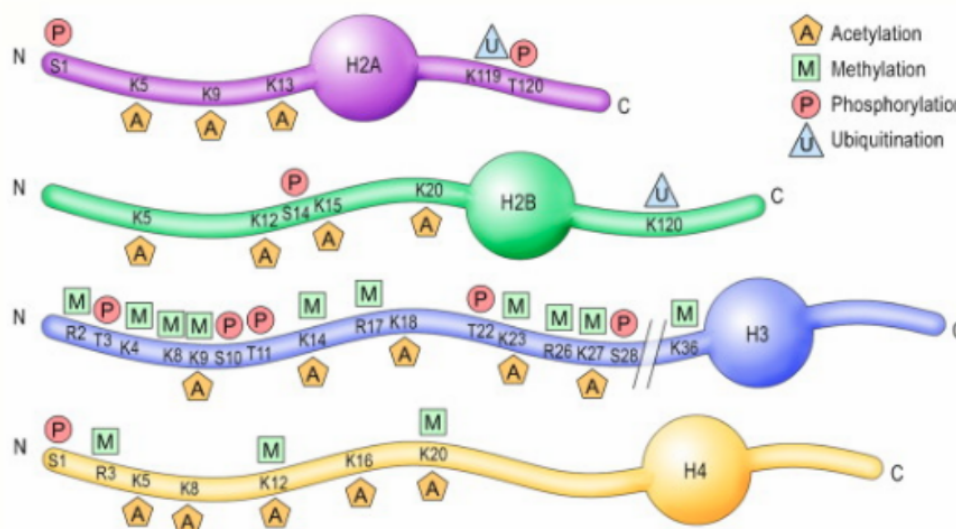


FIGURE 2 – Modifications post-traductionnelles des histones. Figure tirée de Gräff and Mansuy (2008). Représentation des extrémités N et C des histones et leurs modifications épigénétiques spécifiques à chaque résidu.

Ces modifications sont dynamiques, elles peuvent être ajoutées ou retirées des résidus d'acides aminés des histones par un ensemble d'enzymes spécifiques. Ces modifications peuvent affecter la structure de la chromatine, en perturbant l'interaction des histones dans les nucléosomes, ou l'interaction des histones avec l'ADN.

L'acétylation, qui est l'une des premières modifications découvertes, neutralise la charge positive des lysines, ce qui a pour conséquence d'affaiblir les interactions entre les histones et l'ADN et ainsi décondenser l'ADN (Bannister and Kouzarides, 2011). Cette ouverture de la chromatine permet la liaison de facteur de transcription à l'ADN, l'acétylation est associée à une activité transcriptionnelle. Un enrichissement d'acétylation des histones est observé au niveau d'éléments régulateurs, tels que les régions promotrices des gènes et les enhancers. L'acétylation de K27 sur l'histone H3 (H3K27ac) est généralement associé à des enhancers et promoteurs de gènes actifs (Figure 3a). En revanche, la désacétylation a pour conséquence un rétablissement de la charge positive de la lysine, rendant la chromatine plus compacte, l'ADN est moins accessible, ainsi la désacétylation est corrélée à une répression transcriptionnelle.

La phosphorylation se produit principalement sur les sérines, les thréonines et les tyrosines des queues N-terminale des histones. Cette modification ajoute également une charge négative à l'histone conduisant à la neutralisation de sa charge positive, réduisant ainsi son affinité pour l'ADN, et générant une structure de chromatine plus lâche, favorisant une activation transcriptionnelle. Fait intéressant, le même résidu phosphorylé peut avoir des effets significativement différents sur la structure de la chromatine selon le contexte dans lequel il se produit. La phosphorylation de l'histone H3 sur les sérines 10 et 28 est impliquée dans la condensation de la chromatine associée à la mitose et à la méiose, mais aussi à la relaxation de la chromatine liée à l'activation transcriptionnelle (Rossetto et al., 2012).

L'acétylation et la phosphorylation des histones réduisent la charge positive des histones, ce qui peut entraîner une perturbation des interactions électrostatiques entre les histones et l'ADN. En revanche, la méthylation des histones ne modifie pas la charge des histones, et n'impacte donc pas directement les interactions histone-ADN. La méthylation se produit principalement sur les résidus lysines et arginines des histones H3 et H4, et est due à des histones méthyltransférases. Les lysines peuvent être mono, di ou tri-méthylées et les arginines peuvent être mono ou di-méthylées. La méthylation de l'arginine contribue à divers processus cellulaires comme la régulation de la transcription (à la fois dans l'activation et dans la répression de la transcription), l'épissage, et la réparation de l'ADN (Guccione and Richard, 2019).

La méthylation de la lysine est impliquée soit dans une activation soit dans une répression transcriptionnelle, en fonction du site de méthylation ; le degré de méthylation des lysines peut se traduire par différents niveaux de régulation de l'expression des gènes. Les sites de méthylation les plus étudiés sont sur les histones H3 lysine 4 (H3K4), lysine 9 (H3K9), lysine 27 (H3K27), et lysine 36 (H3K36). L'euchromatine, transcrite activement, présente des niveaux élevés d'acétylation et est méthylée sur les sites H3K4, H3K36, H3K79 et H3K20. Les gènes activement transcrits sont caractérisés par des taux élevés de H3K4me3 et H3K27ac dans leur région promotrice et par H3K36me3, H3K79me1 et H4K20me1 le long de la région transcrite (Portela and Esteller, 2010). Les enhancers actifs présentent, eux, des niveaux élevés de H3K4me1 (Figure 3a). En revanche, l'hétérochromatine est associée à des niveaux faibles d'acétylation et à des niveaux élevés de certains sites méthylés : H3K9, H3K27 et H4K20. La tri-méthylation de l'histone H3 au niveau de la lysine 27 ou de la lysine 9 (H3K27me3 ou H3K9me3) induit la formation de l'hétérochromatine et est donc associée à des régions répressives en terme d'expression de gènes (Figure 3b). La triméthylation de H3K27 est induite par EZH2, une protéine du complexe polycomb PRC2 (Polycomb Repressive Complex 2).

Des domaines bivalents ont été décrits présentant à la fois des modifications activatrices et répressives (Rakyan et al., 2010). La présence simultanée de la marque active H3K4me3 et de la marque répressive H3K27me3 au niveau de promoteurs, que l'on nomme promoteurs bivalents, de gènes importants pour le développement, permet aux cellules souches embryonnaires de réguler étroitement et d'activer ou de taire rapidement l'expression des gènes au cours du développement cellulaire (Figure 3c). Lorsque les cellules ES se différencient, elles ont tendance à conserver soit la modification répressive H3K27me3 soit la modification activatrice H3K4me3, mais pas les deux (Kouzarides, 2007).

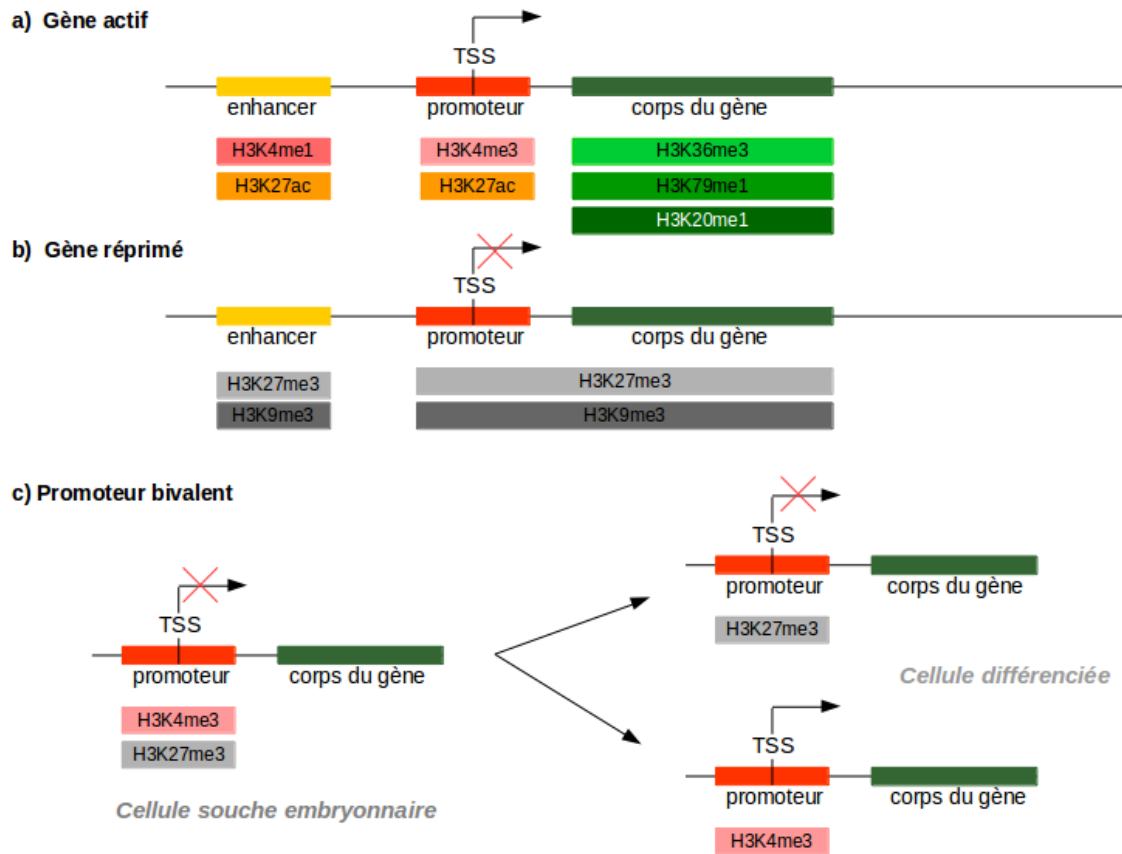


FIGURE 3 – Distribution des marques d’histone dans les régions régulatrices des gènes. (a) Distribution des marques d’histone dans l’enhancer, le promoteur et le corps du gène des gènes activement transcrits. (b) Modifications des histones dans l’enhancer, le promoteur et le corps du gène des gènes réprimés. (c) Promoteur bivalent, présence simultanée des marques H3K4me3 (activatrice) et H3K27me3 (répressive) afin de faciliter un changement rapide d’expression des gènes au cours du développement.

Enfin, l’ubiquitination se produit principalement sur l’histone H2A de la lysine 119 et l’histone H2B de la lysine 120, et consiste en la fixation par une liaison covalente d’une protéine d’ubiquitine (76 acides aminés) à l’histone. L’ubiquitination des histones joue un rôle central dans la réparation de l’ADN (Cao and Yan, 2012; Meas and Mao, 2015).

Ainsi ce grand nombre de modifications possibles des histones permet de contrôler / réguler étroitement la structure de la chromatine. La combinaison de la présence de toutes ces marques d’histone permet de cartographier les états de la chromatine le long du génome (Ernst and Kellis, 2012).

3.3 La méthylation de l'ADN

La méthylation de l'ADN est impliquée dans divers processus biologiques, tels que la régulation génique, la régulation de la structure de la chromatine, la réplication de l'ADN, l'inactivation du chromosome X chez la femme, le développement embryonnaire, et la carcinogenèse. La méthylation de l'ADN est une des marques épigénétiques les plus étudiées chez l'Homme, et c'est une marque épigénétique hautement conservée chez les plantes et les animaux. La méthylation de l'ADN consiste majoritairement en l'ajout d'un groupement méthyle (CH₃) sur le carbone 5' des cytosines (C) précédant un résidu guanine (G) ; c'est ce que l'on appelle un CpG. Cette réaction nécessite une enzyme, l'ADN méthyltransférase (DNMT) et un donneur de groupement méthyle le S-Adénosyl Méthionine (SAM) (Figure 4).

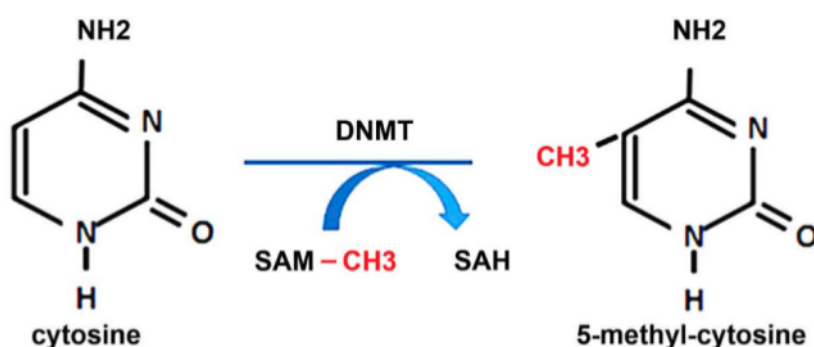


FIGURE 4 – Conversion de la cytosine en 5'-méthylcytosine par l'action des DNMTs. SAM, S-adénosylméthionine ; SAH, S-adénosylhomocystéine ; DNMT, ADN méthyltransférase. Figure tirée de Cui et al. (2016).

La méthylation de l'ADN peut également être présente sur des sites autres que les séquences CpGs, ce que l'on appelle méthylation non-CpG. Les cytosines peuvent être méthylées et suivies d'un résidu adénine (CpA), d'un résidu thymine (CpT) ou d'un autre résidu cytosine (CpC) (Jang et al., 2017). La méthylation non-CpG est facilement détectable dans les cellules souches embryonnaires (ESC), cela représente 15 à 20% de la méthylation totale, mais beaucoup moins présente dans les cellules somatiques différenciées (0,02% de la méthylation totale). Cette méthylation non-CpG se perd au cours de la différenciation cellulaire (Lister et al., 2009).

Dans la majorité des cas, les CpGs sur le génome humain sont méthylés. Plus de 90% des CpG méthylés se situent dans des éléments répétés et les transposons, ce qui représentent plus d'un tiers du génome humain (Nasu et al., 2011). Les sites CpGs et leurs degrés de méthylation sont inégalement répartis dans le génome. Il existe des régions plus riches en CpGs que d'autres, c'est ce que l'on appelle des îlots CpG (CGIs). Ces îlots sont souvent associés à des promoteurs et sont de courts segments d'ADN d'environ 700pb en moyenne, avec une teneur en G + C de 50% et une fréquence CpG (observée / attendue) de 0,6. Les promoteurs sont des régions d'ADN situées en amont des gènes, sur lesquelles se fixent l'ARN polymérase afin de permettre la transcription de l'ADN en ARN. Les CpGs situés dans les CGIs associés

aux promoteurs ont tendance à ne pas être méthylés, mais un sous-ensemble sont méthylés de manière différentielle selon les tissus ou au cours du développement, et entraînent une répression de la transcription des gènes associés. La méthylation du promoteur au niveau des CGIs joue un rôle clé dans l’empreinte génomique, la différenciation des cellules et la régulation de nombreux processus cellulaires. L’impact de la méthylation est tissu spécifique. Si on prend une cellule nerveuse et une cellule musculaire, leur profil de méthylation va être totalement différent, puisque ces deux types de cellules ne vont pas utiliser les mêmes gènes pour leurs fonctions. Les profils de méthylation de l’ADN sont spécifiques du stade du cycle cellulaire, du stade de développement cellulaire, du type de tissu, et de l’âge de l’individu.

3.3.1 Les ADN méthyltransférases

Chez l’homme, la méthylation de l’ADN dans les cellules somatiques est réalisée par des ADN méthyltransférase (DNMT1, DNMT2, DNMT3a, DNMT3b et DNMT3L). Ces DNMTs sont divisées en trois catégories principales : les enzymes impliquées dans le maintien de la méthylation de l’ADN lors de la réplication de l’ADN (DNMT1), les enzymes impliquées dans la réparation de l’ADN (DNMT2) et les enzymes impliquées dans la méthylation de l’ADN de novo (DNMT3A, DNMT3B et leur co-facteur DNMT3L).

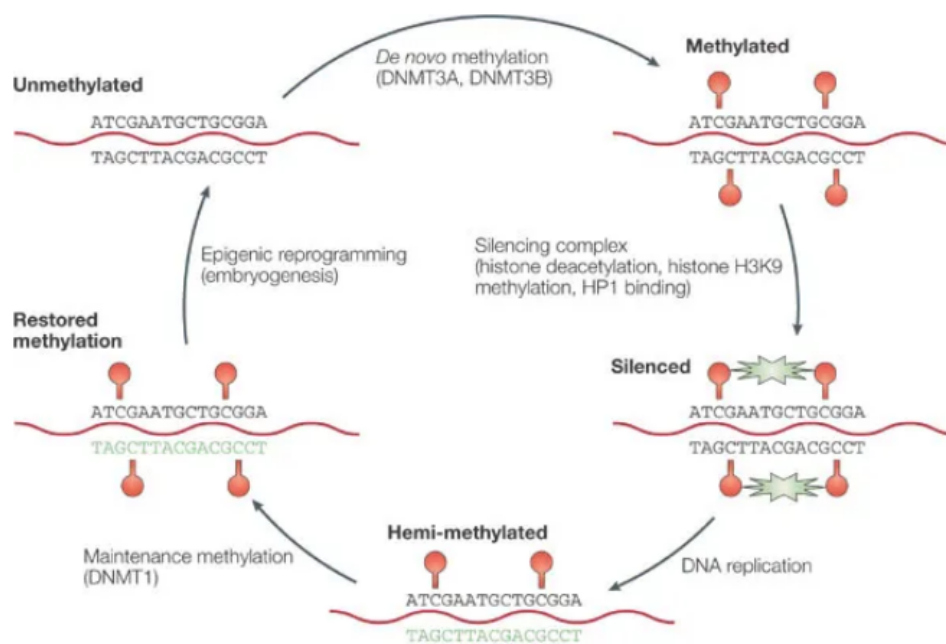


FIGURE 5 – Mécanisme de méthylation de l’ADN. Figure tirée de Issa (2004).

La cytosine méthylée est initialement incorporée à l’ADN au début du développement par les enzymes de novo méthyltransférase DNMT3a et DNMT3b (Okano et al., 1999) (Figure 5). Les sites de méthylation de l’ADN sont occupés par diverses protéines, notamment les protéines MBD (methyl-CpG binding domain), qui interviennent comme répresseurs de la transcription

en empêchant l'accès des facteurs de transcription et par des modifications locales de la chromatine en recrutant des histones désacétylases (HDAC) et des histones méthyltransférases (HMT), résultant en un état de chromatine fermée. Après la réplication de l'ADN, l'ADN nouvellement synthétisé n'est pas méthylé, l'ADN est hémi-méthylé (un seul des 2 brins est méthylé). DNMT1 possédant une affinité supérieure pour l'ADN hémi-méthylé va permettre la maintenance de la méthylation, en copiant les profils de méthylation de l'ADN sur les brins néo-synthétisés au cours de la réplication de l'ADN (Vertino et al., 1996). DNMT1 travaille en complexe, et coordonne la répression des gènes en partenariat avec les histones désacétylases, et est essentiel au maintien de la stabilité de la chromatine. DNMT1 permet le maintien des profils de méthylation au cours des divisions cellulaires, du développement et du vieillissement.

DNMT3L est un co-activateur de DNMT3a et DNMT3b et est impliqué dans l'interaction avec les acteurs de la chromatine, améliorant l'activité de la méthyltransférase. DNMT3L est considéré comme nécessaire pour l'établissement d'empreintes maternelles dans les ovocytes. Ce co-activateur est exprimé dans les ESC, uniquement pendant le développement embryonnaire, mais pas dans les cellules somatiques. Dans les cellules somatiques, DNMT3b peut jouer ce rôle de co-activateur et recruter DNMT3a à la place de DNMT3L, pour initier et maintenir la méthylation de l'ADN dans les régions transcrites. DNMT3a a une faible capacité à restaurer la méthylation de l'ADN par lui-même sans implication du DNMT3b dans des cellules différenciées (Duymich et al., 2016).

La méthylation non-CpG quant à elle est réalisée par les méthyltransférases de novo DNMT3a et DNMT3b (Duymich et al., 2016). La diminution significative de DNMT3 au cours de la différenciation cellulaire coïncide avec la diminution de la proportion de méthylation non-CpG.

3.3.2 La déméthylation

La méthylation des cytosines, bien qu'elle soit transmise au cours des divisions cellulaires, peut être également réversible. Les cytosines peuvent être déméthylées, perte du groupement méthyle CH₃, par plusieurs mécanismes, notamment par un processus actif via les enzymes TET (Ten Eleven Translocation proteins) ou par un processus passif, lors de la réplication de l'ADN en l'absence de méthylation de novo par les DNMTs.

La perte ou la diminution de l'activité de DNMT1 qui permet de maintenir la méthylation lors de la réplication de l'ADN conduit à une perte progressive / une dilution de la méthylation (Figure 6). Au cours des divisions cellulaires, l'ensemble des cellules deviennent hypométhylées.

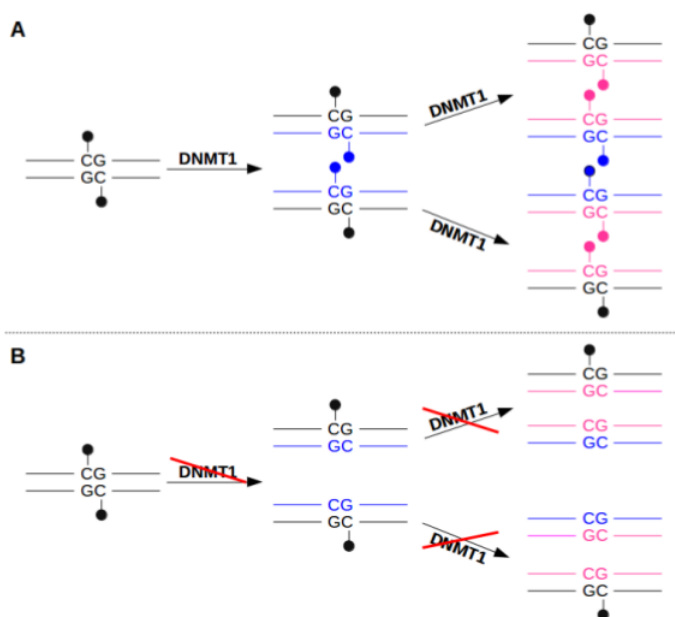


FIGURE 6 – Déméthylation passive de l’ADN. (a) Maintien de la méthylation de l’ADN par DNMT1 lors de la réplication. (b) Déméthylation progressive de l’ADN au cours des réplifications, suite à la perte de DNMT1.

La déméthylation active implique l’oxydation des cytosines méthylées par les enzymes TET (Figure 7). Les méthyl-cytosines (5mC) sont oxydées en 5-hydroxyméthylcytosine (5hmC), qui sont ensuite oxydées en 5-formylcytosine (5fC) puis finalement oxydées en 5-carboxycytosine (5caC). Les 5fC et 5caC peuvent être éliminées de l’ADN par l’enzyme TDG (Thymine DNA glycosylase) et la cytosine non méthylée est restaurée par le biais du système BER (base excision repair) (Wu and Zhang, 2017). Ce processus est indépendant de la réplication de l’ADN, et est nommé AM-AR pour modification active et retrait actif de la méthylation. L’enzyme TDG élimine spécifiquement les molécules 5fC et 5saC, mais pas 5hmC.

Outre le processus AM-AR, la réplication de l’ADN peut conduire à la dilution des 5mC oxydés, ce processus se nomme AM-PD pour modification active – dilution passive. Lors de la réplication, l’ADN est héli-méthylé (un seul des 2 brins est méthylé). Ces sites héli-5mC sont reconnus par une protéine, UHRF1, qui facilite le recrutement de DNMT1. Plusieurs études (Hashimoto et al., 2012; Otani et al., 2013) mettent en avant que UHRF1 et DNMT1 ont moins d’affinités avec les sites 5hmC, 5fC et 5saC qu’avec les sites 5mC. Cela engendre après plusieurs cycles de réplication de l’ADN une dilution de ces sites, et un retour d’une cytosine non méthylée.

La déméthylation active de l’ADN est importante pour différents mécanismes biologiques. Par exemple, peu de temps après la fécondation, on observe une déméthylation globale de l’ADN des génomes paternel et maternel. La déméthylation du génome paternel se fait par un processus AM-PD tandis que la déméthylation du génome maternel se fait lui principalement par le processus passif, la dilution au cours des cycles cellulaires (Sanz et al., 2010).

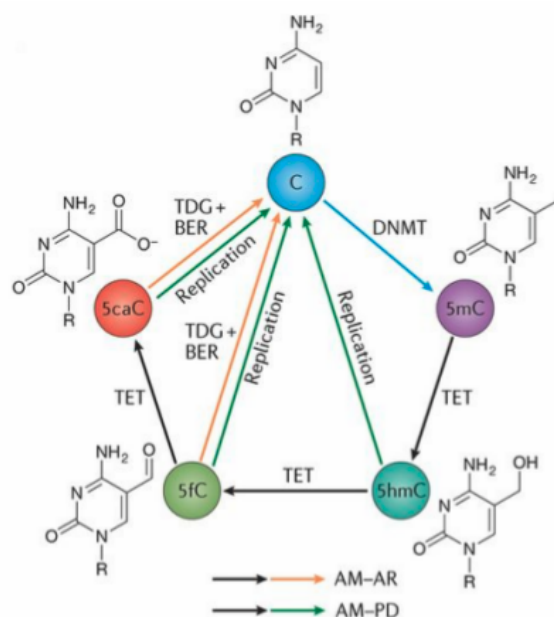


FIGURE 7 – Le cycle de la déméthylation active de l'ADN. Les DNMTs convertissent la cytosine en 5mC, qui lui est converti par oxydation médiée par l'enzyme TET en 5hmC, puis en 5fC et 5caC, qui peuvent être reconverties en cytosine par le complexe TDG+BER ou par une dilution passive lors de la réplication de l'ADN. AM-AR : modification active - retrait actif; AM-PD : modification active - dilution passive. Figure tirée de Wu and Zhang (2017).

3.3.3 Les inhibiteurs des DNMTs

Il existe plusieurs inhibiteurs de la méthylation, et notamment les plus efficaces : la 5-azacytidine (Vidaza) et la 5-aza-2'-désoxycytidine (Décitabine) qui ont été synthétisées pour la première fois par Piskala et Sorm en 1964, dans le cadre du traitement de la leucémie myéloïde aiguë (Šorm et al., 1964). Ce sont des analogues de nucléosides qui, à faibles doses, ont révélé une capacité d'hypométhylation de l'ADN. La 5-azacytidine est un analogue chimique du nucléoside cytidine, qui s'incorpore davantage dans l'ARN que dans l'ADN, alors que la 5-aza-2'-désoxycytidine est un désoxyribonucléoside et ne peut donc être incorporée qu'à l'ADN à la place de la cytosine mais ne peut être méthylée. Après incorporation dans l'ADN, ces analogues de cytidine vont se lier aux DNMTs par une liaison covalente et ainsi inhiber leur fonction enzymatique, produisant ainsi une dilution de la méthylation progressive au cours des divisions cellulaires. Ces inhibiteurs à forte dose présentent une forte toxicité. Ils présentent une faible stabilité chimique, et une faible spécificité, puisqu'elles s'incorporent à l'ADN en remplaçant toutes les cytosines. La zébularine est un autre analogue de la cytidine plus stable et moins toxique, cependant son efficacité est relative, même à de fortes concentrations.

Certains médicaments comme la procainamide (antiarythmique) et la procaine (anesthésiant local), de part leur effet secondaire, ont montré une activité déméthylante de l'ADN au cours de leur utilisation (Castillo-Aguilera et al., 2017). Ces molécules ont montré une affinité pour les régions d'ADN riches en CpG hémi-méthylés, bloquant l'activité des DNMT1.

La 5-azacytidine et la 5-aza-2'-désoxycytidine sont les deux seuls inhibiteurs des DNMTs approuvés par la Food and Drug Administration des Etats-Unis (FDA) et l'Agence Européenne des Médicaments (EMA) pour le traitement de la leucémie myéloïde aiguë (AML), de la leucémie myélomonocytaire chronique (CMML) et des syndromes myélodysplasiques (MDS) (Castillo-Aguilera et al., 2017).

3.3.4 Régulation de l'expression par la méthylation

Le génome humain comprend environ 28 millions de CpGs, qui ne sont pas répartis d'une manière homogène sur le génome, et qui selon leur localisation et leur niveau de méthylation peuvent avoir des impacts différents. Dans les cellules normales, les CpGs méthylés sont principalement localisés dans des régions pauvres en CpG, tandis que les régions riches en CpG (CGI) situées dans les régions régulatrices des gènes sont protégées et non méthylées (Figure 8). La méthylation d'un CpG est une modification thermodynamique très stable ayant une influence sur l'expression des gènes.

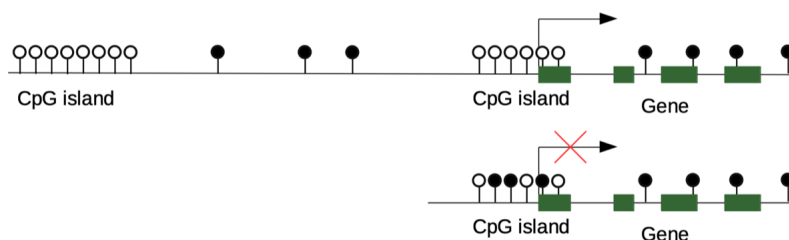


FIGURE 8 – Méthylation de l'ADN et régulation de l'expression.

La plupart des gènes ont dans leur région promotrice un CGI associé. La méthylation des CpGs dans les CGIs associés à un promoteur a été inversement corrélée à l'expression génique (Irizarry et al., 2009b). L'absence de méthylation dans les régions promotrices est corrélée à une configuration active de la chromatine qui favorise l'interaction de l'ADN avec des complexes de transcription menant à l'expression génique. Les gènes exprimés présentent en général une faible méthylation dans leur région promotrice et une méthylation élevée dans leur corps de gène, cette méthylation dans les exons semble être importante dans le contrôle de l'épissage alternatif (Figure 8 partie haute) (Ball et al., 2009; Maunakea et al., 2010).

Au contraire, une méthylation dense dans une région promotrice / CGI est le plus souvent associée à une répression transcriptionnelle (Figure 8 partie basse). Cette répression de la transcription associée à la méthylation peut se produire par deux types de mécanismes non exclusifs mutuellement : la présence de 5mC peut physiquement restreindre la liaison aux facteurs de transcription et à l'ARN polymérase II ou 5mC peut recruter des complexes protéiques qui favorisent la répression de la transcription, comme par exemple des protéines MBD présentant un domaine de liaison à l'ADN méthylé qui ont des fonctions répressives associées au remodelage de la chromatine. La méthylation des cytosines dans les régions promotrices va entraîner une modification de l'architecture de la fibre de chromatine qui

aboutit à la compaction des nucléosomes, empêchant l'accès des facteurs de transcription ou des protéines de liaison à l'ADN (Curradi et al., 2002). Certains facteurs de transcription se liant à l'ADN sur des séquences riches en CpG, tel que *AP-2*, *c-Myc / Myn*, *E2F* et *NFκB*, peuvent être sensibles à une méthylation dans ces régions (Siegfried and Cedar, 1997).

3.3.5 Évolution des profils de méthylation au cours du temps

Suite à la fécondation, le génome des gamètes subit une déméthylation globale, presque totale. Le génome paternel subit un processus de déméthylation actif, tandis que le génome maternel subit un processus de déméthylation passif au début de l'embryogenèse. Dans la blastula, la majeure partie de l'ADN est non méthylée. Après l'implantation de l'embryon, une vague de méthylation *de novo* modifie une grande partie du génome, à l'exception des CGIs qui restent protégés (Siegfried and Cedar, 1997). Une comparaison entre les cellules souches pluripotentes et les cellules somatiques a révélé des zones de méthylation différentielles spécifiques au type cellulaire (Deng et al., 2009).

Au cours de la vie humaine, le contenu en 5mC est plus élevé chez les embryons et diminue progressivement avec l'âge (Goel et al., 2017). Actuellement, la méthylation de l'ADN est le marqueur le plus prometteur pour surveiller le vieillissement et prévoir l'espérance de vie. De nombreux chercheurs ont construit des modèles de prédiction de l'âge avec une grande précision basés sur les changements de méthylation de certains loci de CpGs (Hannum et al., 2013; Horvath, 2013). Des mesures de l'accélération de l'âge ont été mises au point pour évaluer l'état de santé d'une personne. Le vieillissement est largement considéré comme un processus complexe, caractérisé par une accumulation de molécules, de cellules et de tissus endommagés, conduisant à un dysfonctionnement et à une dégénérescence. La vieillesse est associée à une perte de capacité à maintenir l'homéostasie tissulaire initiale et à un risque accru de développer de nombreuses maladies, notamment des cancers.

Les modifications de la méthylation de l'ADN en fonction de l'âge incluent une hypométhylation globale et une hyperméthylation plus localisée (Xiao et al., 2016). L'hypométhylation associée à l'âge se produit plutôt dans l'hétérochromatine, notamment dans les éléments répétés et les transposons, régions qui contiennent la majorité des CpGs méthylés dans l'ADN humain (Bollati et al., 2009; Heyn et al., 2012). En plus des éléments répétés, une perte de méthylation corrélée avec le vieillissement a également été démontrée dans les régions pauvres en CpGs situées à côté des gènes (Day et al., 2013; Heyn et al., 2012). À l'inverse, l'hyperméthylation focale au cours du vieillissement se produit principalement au niveau des CGIs associés aux gènes qui ne sont normalement pas méthylés (Madrigano et al., 2012). Rakyan et al. ont révélé que les régions hyperméthylées au cours du vieillissement sont préférentiellement situées sur des domaines de chromatine bivalents (Rakyan et al., 2010).

Les changements de méthylation de l'ADN associés au vieillissement sont causés en partie par des facteurs environnementaux. Les changements épigénétiques sont dynamiques au cours de la vie et constituent un mécanisme important permettant aux organismes de s'adapter aux changements environnementaux externes et internes. Certains facteurs environnementaux, tels que la pollution, le tabac, l'alcool, les UVs, le stress peuvent faciliter l'hypométhylation globale.

La dynamique du méthylome est présente aussi avant la naissance, les habitudes alimentaires pendant la grossesse déterminent fortement le statut de méthylation du génome de l'enfant (Aagaard-Tillery et al., 2008).

3.4 Les ARNs non codants

Les ARNs non codants jouent également un rôle dans le contrôle de multiples phénomènes épigénétiques, dans la modification de la chromatine et notamment dans la régulation de l'expression des gènes, ils modulent la disponibilité des gènes et leur lecture. Seul 1,2% du génome humain code pour des protéines, le reste des transcrits correspond à des ARNs non codants (ARNs régulateurs, non messagers) (Bernstein and Allis, 2005). Les ARNs non codants sont des molécules d'ARN fonctionnel qui ne codent pas pour une protéine. Il existe différents types d'ARN non codants participant à des phénomènes épigénétiques :

Les micro ARN (miRNA) sont des ARNs simple brin d'une longueur d'environ 22 nucléotides ayant pour rôle principal la régulation post-traductionnelle de l'expression des protéines en dégradant leurs ARNs messagers (ARNm) cibles et / ou en inhibant leur traduction. Les miRNA peuvent inhiber la traduction de nombreuses séquences d'ARNm différentes dû à leur appariement avec leur cible qui n'est pas nécessairement parfait. Outre leur rôle dans le ciblage de gènes pour l'inhibition de la traduction ou la destruction d'ARNm, il semble également que certains miRNA de plantes puissent diriger la méthylation de l'ADN de leurs gènes cibles (Bao et al., 2004).

Les petits ARN interférants (siRNA) sont des ARNs double brin exogènes, entrés par des vecteurs tels que des virus, ils se lient d'une manière parfaite, spécifique, sur leur cible ARNm, cible unique, empêchant ainsi l'expression de ces gènes cibles en clivant l'ARN (Dana et al., 2017)

Les ARN interagissant avec Piwi (Piwi-interacting RNA, piRNA) sont des ARNs simple brin, qui présentent une taille comprise entre 21 et 35 nucléotides, ils inactivent les éléments transposables et régulent l'expression des gènes. Les piRNA guident les protéines PIWI pour cliver l'ARNm cible et favoriser l'assemblage de l'hétérochromatine. Les piARN semblent également affecter certaines méthyltransférases qui effectuent les méthylation nécessaires pour reconnaître et faire taire les transposons (Ozata et al., 2019).

Les petits ARN nucléaire (snRNA) ont une taille comprise entre 60 et 450 nucléotides, ils participent à l'épissage des ARNs messagers et possiblement au maintien des télomères (Franke et al., 2008).

Les longs ARN non codants (lncRNA) font une taille supérieure à 200pb et jouent un rôle dans la régulation post-transcriptionnelle des gènes, comme activateur de la transcription, dans le remodelage de la chromatine et l'inactivation du chromosome X (Lee, 2012). Les LncRNA interviennent dans la méthylation ou la déméthylation par interaction avec diverses méthyltransférases (Frías-Lasserre and Villagra, 2017).

4 Hétérogénéité épigénétique et génétique dans le cancer

Il existe une grande variabilité entre patients atteints d'un même cancer, ce que l'on appelle l'hétérogénéité inter-patient. Deux patients atteints d'un même type de cancer ne présentent pas toujours les mêmes signes cliniques. De plus, dans le cas du MM, il existe une grande variabilité de réponses aux traitements entre les patients. Suite au même traitement, certains patients vont présenter un meilleur pronostic que d'autres.

4.1 L'hétérogénéité génétique inter-patient

Le MM est une maladie génétiquement complexe et hétérogène, résultant de nombreux événements génomiques conduisant à son développement et sa progression. Cette succession d'altérations génétiques commence par des événements précoces prédisposant à la maladie, suivis d'événements primaires impliqués dans le développement de la tumeur, puis d'événements secondaires impliqués dans la progression tumorale et la résistance des patients au traitement.

Au cours de ces 10 dernières années, de nombreuses anomalies génétiques conduisant au MM ont été définies. Les événements primaires sont subdivisés en deux sous-types : les tumeurs hyperdiploïdes et les non hyperdiploïdes. Les tumeurs hyperdiploïdes sont caractérisées par une trisomie des chromosomes 3, 5, 7, 9, 11, 15, 19 et/ou 21. L'hyperdiploïdie est caractérisée par un nombre de chromosomes compris entre 48 et 74, et est détectée chez environ 50% des patients au stade MGUS, SMM et MM. Les tumeurs non hyperdiploïdes sont quant à elles caractérisées par des translocations affectant des gènes codants pour les chaînes lourdes des immunoglobulines, plus de 90% des translocations chromosomiques affectent le chromosome 14, principalement des translocations $t(4; 14)$, $t(6; 14)$, $t(11; 14)$, $t(14; 16)$ et $t(14; 20)$ (Egan et al., 2012).

Lors du développement des cellules B, l'ADN codant pour les régions hypervariables de la chaîne lourde de l'immunoglobuline subit une hypermutation somatique pour produire des anticorps hautement spécifiques. La maturation d'affinité de ces anticorps est augmentée par la commutation de classe qui permet d'exprimer une immunoglobuline de différents isotypes. Ces deux mécanismes, l'hypermutation somatique et la commutation de classe, nécessitent l'expression de la désaminase induite par activation (AID) et sont médiés par des cassures d'ADN double brin dans des loci codant pour l'immunoglobuline. Bien que les cassures d'ADN double brin induites par l'AID soient principalement réparées localement, il peut se produire une jointure à une autre cassure double brin survenant ailleurs dans le génome, ce qui engendre une translocation (Morgan et al., 2012b).

Ces translocations primaires sont observées chez la plupart des MM et impliquent principalement 3 groupes de gènes :

- Les cyclines D qui sont transloquées chez 20 à 25% des MM. La cycline D est impliquée dans la régulation de la progression du cycle cellulaire, sa synthèse entraîne la transition de la phase G1/S. La dérégulation de cette transition est une anomalie moléculaire précoce, importante dans le MM.

- Le récepteur du facteur de croissance des fibroblastes 3 (*FGFR3*) combiné avec la protéine nucléaire *MMSET*, transloqué chez environ 15% des MM,
- Le facteur de transcription *c-MAF* transloqué dans 10% des MM (Kuehl and Bergsagel, 2002).

Les événements secondaires permettent un avantage sélectif particulier nécessaire à la progression tumorale. La moitié des tumeurs avancées du MM présentent des anomalies caryotypiques qui impliquent *c-MYC* ou, occasionnellement, *LMYC* ou *NMYC* (Shou et al., 2000). Des translocations affectant *MYC* sont fréquemment observées dans les tumeurs hyperdiploïdes (environ 60% des cas) et sont de mauvais pronostic. Les mutations affectant la voie ras/MAPKinase sont les mutations les plus fréquemment observées dans le MM et sont détectées chez environ 40% des patients. Des mutations activant la voie NF κ B surviennent chez 20% des MM (Manier et al., 2017) et favorisent la croissance et la survie des cellules de myélome dans le micro-environnement de la moelle osseuse (Demchenko and Kuehl, 2010). Les événements secondaires impliquent également des variations de nombre de copies (CNVs), ainsi qu’une dérégulation de certaines voies de réparation de l’ADN (mutations des gènes *TP53*, *ATM* et *ATR* chez environ 20% des patients, associées à un mauvais pronostic)(Chapman et al., 2011).

Une étude sur une cohorte de 38 patients pour lesquels un séquençage du génome entier (whole genome sequencing, WGS) ou de l’exome entier (whole exome sequencing) a été réalisé, a permis d’identifier 10 gènes présentant significativement des mutations : *NRAS*, *KRAS*, *FAM46C*, *DIS3*, *TP53*, *CCND1*, *PNRC1*, *ALOX12B*, *HLA-A* et *MAGED1* (Chapman et al., 2011). Ces mutations ne sont pas pour autant observées chez tous les patients, seulement une partie. Par exemple, le gène *DIS3* présentait des mutations uniquement chez 4 des 38 patients, ce gène code pour une exonucléase d’ARN hautement conservée, impliquée dans la régulation, le traitement et l’abondance de tous les types d’ARN. Quatre mutations ont été observées se situant dans ces régions hautement conservées, ce qui suggèrent que les mutations de *DIS3* peuvent participer à la dérégulation de la traduction de certaines protéines, et exercer un potentiel effet oncogène dans le cas du MM. Cet éventail de mutations observées avec une certaine occurrence chez plusieurs patients illustre bien l’importance d’étudier dans le cas du MM, non pas un seul, mais de nombreux génomes de patients, de par la diversité des paysages génétiques.

De nombreuses mutations sur des enzymes ayant pour rôle la modification des histones ont été largement décrites dans la littérature pour plusieurs types de cancers. Notamment des translocations et mutations sur SETD2, qui est une histone méthyltransférase spécifique de H3K36, ces perturbations ont été décrites dans plusieurs tumeurs, y compris le MM (Feinberg et al., 2016).

Dans le cas du MM, 15% des patients présentent une translocation t(4;14), qui est l’une des translocations les plus courantes dans le MM et est associée à un mauvais pronostic. Cette translocation conduit à une surexpression systématique du gène *MMSET* et dans 70% des cas à une surexpression également du gène *FGFR3* (dans les 30% des autres cas, le

mauvais pronostic des patients est indépendant de l'expression du gène *FGFR3*) (Xie and Chng, 2014). *MMSET* est un régulateur majeur de la structure de la chromatine et de la transcription dans les cellules de MM, et code pour un répresseur transcriptionnel d'histone méthyltransférase intervenant dans la méthylation de l'histone H3K36 (Mirabella et al., 2013). Une corrélation entre des niveaux élevés de *MMSET* et une augmentation de la méthylation de l'histone H3K36 couplée à une diminution de la méthylation de H3K27 a été mise en évidence, conduisant à une ouverture de la chromatine favorisant l'activation génique (Martinez-Garcia et al., 2011). Il a été montré qu'une perte d'expression de *MMSET* conduit à l'inhibition de la prolifération et à l'apoptose des cellules de MM.

Le développement des nouvelles technologies de séquençage de nouvelle génération, et notamment le séquençage à forte profondeur, a permis de mettre en lumière toute la complexité génétique des tumeurs malignes, et notamment un autre type d'hétérogénéité, l'hétérogénéité intra-tumorale, qui désigne les variations au sein des cellules d'une même tumeur. Les tumeurs ne sont pas constituées d'un ensemble de cellules uniques anormales identiques mais plutôt d'une population de cellules issus de plusieurs cellules anormales.

4.2 L'évolution clonale comme modèle de progression du MM

La première théorie de l'évolution du cancer selon les principes darwiniens d'accumulation progressive de mutations et de sélection naturelle a été évoquée par le Dr Peter Nowell en 1976 (Nowell, 1976). Il propose un modèle évolutif clonal du développement du cancer, une cellule unique accumule des altérations génétiques lui conférant un avantage sélectif permettant une expansion clonale d'une sous-population présentant des avantages de croissance et de survie formant la masse tumorale.

L'hypothèse d'une évolution clonale comme base de la progression tumorale a été étudiée pour plusieurs cancers (Dagogo-Jack and Shaw, 2018) et notamment le myélome multiple. Les cellules propagatrices du myélome représentent un mélange de sous-clones génétiques concurrents, capables de maintenir les fonctions du clone du myélome tout en développant de nouvelles anomalies génétiques lui permettant d'acquérir un avantage sélectif, de s'étendre et de faire évoluer la maladie (Figure 9). Cette évolution clonale se fait par des voies de branchement avec sélection en termes de survie ou de croissance. Chez la plupart des patients, la pression thérapeutique entraîne une réorganisation et une diversification des populations sous-clonales, avec l'émergence et l'expansion de sous-clones résistants au traitement (Magrangeas et al., 2013). Le clone majoritaire suite au traitement peut être éliminé laissant place à l'expansion d'un autre clone présent plus minoritairement mais qui lui peut être encore plus agressif et donc accélérer la progression de la maladie, ou alors un clone plus indolent qui permet de diminuer la progression et les symptômes du MM.

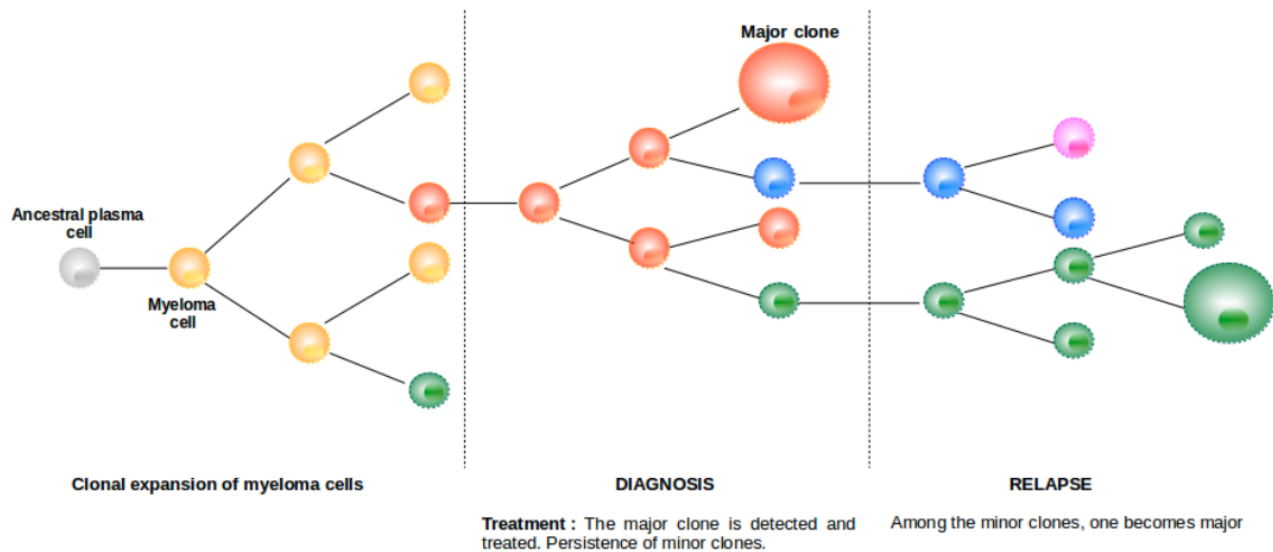


FIGURE 9 – Expansion clonale et progression du myélome multiple.

Des analyses génomiques en série sur des échantillons prélevés à différents temps au cours de la progression du MM ont montré la présence de modèles évolutifs distincts selon les patients : soit une évolution linéaire, soit une évolution hétérogène, ramifiée, avec des clones prédominants en mouvement (Keats et al., 2012).

La contribution de l'hétérogénéité clonale à la progression de la maladie et à la résistance au traitement est de plus en plus reconnue dans le cancer. La connaissance de cette dynamique et hétérogénéité intraclonale est cruciale afin de mieux comprendre la progression du MM, la résistance aux traitements et pouvoir au mieux adopter de nouvelles stratégies thérapeutiques. La dynamique d'évolution des sous-populations clonales peut avoir un impact sur le pronostic et le choix des approches thérapeutiques en améliorant l'évaluation de la réponse au traitement.

4.3 L'hétérogénéité génétique intra-patient

Au cours des dernières années, au vu de la variabilité des translocations et des mutations génétiques observées, des études de séquençage de l'exome entier ont permis d'illustrer l'existence d'une certaine hétérogénéité clonale participant au développement et à l'évolution du MM. Une étude longitudinale a été réalisée sur un patient atteint de MM pour lequel 4 échantillons de tumeur ont été prélevés tout au long de l'évolution de sa maladie (au diagnostic, à sa première rechute, à sa seconde rechute, et au stade terminal) (Egan et al., 2012). Sur ces échantillons, un séquençage du génome entier a été effectué et a permis d'identifier de nouveaux variants de nucléotides (SNVs) et d'illustrer l'existence de l'hétérogénéité tumorale et la dynamique clonale dans un cas de MM. L'analyse du nombre de variants spécifiques aux différents stades de la maladie a permis de montrer qu'il y avait une grande divergence évolutive au cours du temps, ainsi qu'une pression sélective suite aux traitements thérapeutiques. Ils ont mis en évidence que ce patient au diagnostic présentait un clone dominant sensible au lénalidomide, le premier traitement administré, ce qui a engendré une diminution de ce clone

et l'émergence d'un clone résistant à ce premier traitement au cours de la première rechute. Ce patient a été ensuite traité au carfilzomib, le clone de la rechute y étant sensible a donc diminué, mais le clone du diagnostic y étant non sensible a pu ré-émerger. Peu de divergences ont pu être mises en évidence entre les clones du diagnostic et de la seconde rechute, uniquement un SNV les différencie. Au stade terminal de la leucémie à plasmocytes de ce patient, un unique SNV différencie ce stade du précédent, cet unique SNV, non détectable aux autres stades de la maladie, suggère l'acquisition de nouvelles mutations qui permettraient la transformation leucémique. Cette étude illustre parfaitement un modèle d'expansion clonale dans le cas du MM.

Depuis, de nombreuses autres études de l'hétérogénéité et d'évolution clonale ont été réalisées sur des cohortes de patients plus conséquentes. En 2014, Golub et son équipe ont mis en évidence la présence d'au moins 3 sous-clones en plus du clone principal chez la plupart des patients composant sa cohorte (n=203), certains patients pouvaient présenter jusqu'à 7 sous-clones (un sous-clone représentant au moins 10% de l'échantillon tumoral global) (Lohr et al., 2014). Ils ont également mis en évidence que certaines mutations avaient tendance à être plutôt clonales ou sous-clonales. Par exemple, les mutations *DIS3* et *KRAS* ont été identifiées comme étant souvent simultanément clonales, alors que les mutations *BRAF* étaient souvent sous-clonales. Les mutations sont définies comme étant clonales lorsqu'elles sont présentes chez toutes les cellules cancéreuses, et sous-clonales lorsqu'elles se produisent uniquement dans une sous-population de cellules cancéreuses. Ils ont également mis en évidence que les mutations significativement récurrentes étaient plus souvent clonales chez les patients traités par rapport aux patient non traités, suggérant ainsi une influence du traitement sur la sélection d'un certain type de sous-clones. Munshi et son équipe ont également observé que certaines mutations majeures dans le MM (*KRAS*, *NRAS* et *BRAF*) étaient à la fois clonales pour une partie des patients, et sous-clonales pour d'autres (Bolli et al., 2014).

Une étude légèrement plus récente montre que les patients atteints de MM présentent en moyenne 5 sous-clones détectables, certains patients pouvant en présenter jusqu'à 10, et mettent en avant que certains gènes tels que *FAM46C* et *TRAF3* sont plus fréquemment sous-clonaux que d'autres (Walker et al., 2015). Ils montrent aussi que les mutations affectant la voie MAPKinases sont produites plus d'une fois dans la même tumeur mais dans des sous-clones distincts évoluant séparément.

Notre équipe de recherche a réalisé une analyse génomique de comparaison d'échantillons appariés diagnostics et rechutes qui a révélé deux types d'évolution sous-clonale dans le cas du MM (Magrangeas et al., 2013). Deux tiers des patients de l'étude présentaient un modèle linéaire d'évolution, dans lequel un sous-clone principal au diagnostic acquiert des lésions génétiques supplémentaires et évolue vers un nouveau sous-clone dominant à la rechute. Chez un tiers des patients, un modèle non linéaire a été observé dans lequel un sous clone dominant au diagnostic a été éliminé suite au traitement thérapeutique, laissant place à un sous-clone mineur qui a pu émerger au cours de la rechute. Un lien entre ce modèle non linéaire et le traitement par bortézomib a été mis en évidence, suggérant que ce traitement est efficace contre le sous-clone dominant du MM, mais permet à des sous-clones mineurs de persister, et d'avoir une chance d'émerger en rechute.

Une étude de séquençage de l'exome et du génome entier sur une série d'échantillons (MGUS, SMM, MM et leucémie à plasmocytes) a montré que bien que l'hétérogénéité intraclonale soit présente à tous les stades de la maladie, le stade MGUS était moins complexe génétiquement que le stade MM, et que le stade SMM était d'un point de vue génétique très similaire au MM. Le clone prédominant du MM était déjà présent au stade SMM (Walker et al., 2014).

Enfin, une étude de séquençage du génome entier d'échantillons appariés de patients au stade SMM évoluant vers un MM fournit une description des caractéristiques génomiques des SMM ainsi que leur modèle d'évolution (Bolli et al., 2018). Ils ont observé que les aberrations les plus courantes étaient toutes clonales et que les SMM présentent une hétérogénéité sous-clonale importante perturbée au cours de la progression du MM. La complexité génomique des SMM ne semble pas être associée à un phénotype agressif de la maladie. Ils ont constaté deux modèles de progression des SMM. Un premier modèle de progression dit "statique", une même architecture sous-clonale est conservée au cours de l'évolution de la maladie. Le second modèle est dit "spontané", la composition sous-clonale du SMM change sans aucune pression du traitement, en raison d'une accumulation stochastique de mutations conférant un avantage prolifératif à l'un des sous-clones.

Bien qu'un certain nombre de mécanismes génétiques aient été découverts, ils n'expliquent pas à eux seuls les mécanismes de progression tumorale et de résistance du MM. La piste de l'existence de mécanismes épigénétiques participant à l'adaptabilité des cellules cancéreuses est de plus en plus étudiée et caractérisée ces dernières années dans de nombreux cancers. Le non-maintien des marques épigénétiques héréditaires peut entraîner une activation ou une inhibition inappropriée de diverses voies de signalisation et conduire à la transformation des cellules saines en cellules cancéreuses. Le fait que les aberrations épigénétiques, contrairement aux mutations génétiques, soient potentiellement réversibles et puissent retrouver leur état normal rend ce domaine d'étude prometteur pour la mise en place de nouvelles stratégies thérapeutiques.

4.4 Hétérogénéité de la méthylation

La méthylation de l'ADN est le changement épigénétique le plus étudié en cancérologie. Les cellules cancéreuses présentent un épigénome très spécifique : une hypométhylation globale progressive et des sites d'hyperméthylation plus focalisée au niveau de certains CGIs sur le génome (Feinberg and Tycko, 2004). L'hypométhylation globale peut avoir plusieurs impacts, notamment l'activation de l'expression de gènes oncogéniques (qui favorise la prolifération cellulaire) suite à la déméthylation de leur région promotrice (Aithal and Rajeswari, 2013) (Figure 10a et b), ou bien conduire à une instabilité du génome et à des réarrangements chromosomiques. La perte globale de la méthylation semble se faire préférentiellement au niveau des séquences répétées (régions normalement très méthylées dans les cellules saines), ce qui déclenche la réactivation transcriptionnelle de séquences d'ADN parasites pouvant s'intégrer de manière aléatoire dans le génome (Alves et al., 1996), contribuant ainsi à la forte instabilité génétique des cellules cancéreuses. Alors qu'une hyperméthylation locale spécifique,

au niveau de certains promoteurs, permettrait l'inactivation de gènes suppresseurs de tumeurs impliqués dans la prolifération cellulaire, la réparation des dommages de l'ADN et l'apoptose (Figure 10c et d). L'activation de gènes oncogéniques et l'inactivation de gènes suppresseurs de tumeur peuvent provoquer une prolifération cellulaire incontrôlée, caractéristique des cellules cancéreuses. La perte de l'expression de gènes suppresseurs de tumeur confère aux cellules tumorales un puissant avantage prolifératif.

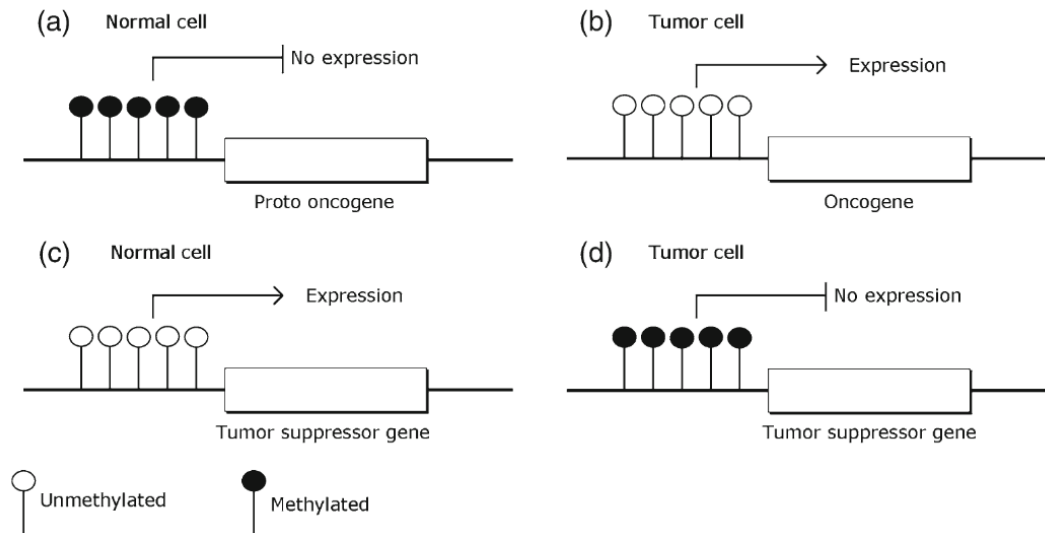


FIGURE 10 – Mécanismes de régulation des gènes oncogéniques et des suppresseurs de tumeur dans les cellules normales et les cellules tumorales. (a) Hyperméthylation de la région promotrice du proto-oncogène dans la cellule normale induisant sa répression transcriptionnelle. (b) Hypométhylation de la région promotrice du proto-oncogène dans la cellule tumorale induisant son expression. (c) Hypométhylation de la région promotrice du gène supresseur de tumeur dans la cellule normale induisant son expression. (d) Hyperméthylation de la région promotrice du gène suppresseur de tumeur dans la cellule tumorale induisant sa répression transcriptionnelle. Figure extraite de Aithal and Rajeswari (2013) .

Dans les cellules cancéreuses, l'hyperméthylation au niveau des promoteurs-CGIs des gènes suppresseurs de tumeur a été associée à une combinaison particulière de marques d'histone : déacétylation des histones H3 et H4, perte de la triméthylation de H3K4, gain de la méthylation d'H3K9 et la triméthylation d'H3K27 (Berdasco and Esteller, 2010). La méthylation de l'ADN au niveau de certains promoteurs et l'hypométhylation globale sont des événements précoces de la formation de la tumeur. Chaque cancer semble avoir un profil de méthylation de l'ADN qui peut être utilisé comme signature pour développer des biomarqueurs diagnostiques. Une étude portant sur le cancer colorectal humain a montré que la dérégulation de l'expression de *DNMT1* ou *DNMT3* ne joue aucun rôle dans l'établissement des profils aberrants de méthylation de l'ADN (Eads et al., 1999).

Hansen et al. ont étudié les régions différentiellement méthylées de plusieurs cancers (du côlon, du sein, du poumon, de la thyroïde et la tumeur de Wilms), et ils ont montré une

variation stochastique de la méthylation dans chacune de ces tumeurs (Hansen et al., 2011). Ils ont également identifié de grands blocs d'hypométhylation représentant pas moins de la moitié du génome et incluant de petites régions hyperméthylées. Ces régions différenciellement méthylées observées dans le cancer ne semblent pas être touchées au hasard mais correspondent à des régions différenciellement méthylées lors de la reprogrammation des cellules souches pluripotentes, et dans la différenciation cellulaire normale. Bien que le niveau global d'expression des gènes dans les blocs hypométhylés reste faible dans le cancer, ils ont observé que ces blocs contiennent les gènes les plus exprimés dans les tumeurs par rapport aux témoins normaux.

Une étude de l'évolution de la méthylation de l'ADN chez des patients atteints de leucémie lymphoïde chronique (CLL) a montré qu'un sous-ensemble de patients présentaient un niveau élevé d'hétérogénéité de la méthylation associé à un mauvais pronostic et à une plus grande diversité génétique sous-clonale (Oakes et al., 2014).

Brocks et al. ont analysé plusieurs sites tumoraux topographiquement distincts les uns des autres dans 5 cas de cancer de la prostate (Brocks et al., 2014). Ils ont montré la présence de plusieurs populations de cellules sous-clonales caractérisées par des profils de méthylation de l'ADN distincts. Ils ont également constaté que l'hypométhylation de l'ADN était un phénomène plutôt sous-clonal alors que les événements d'hyperméthylation étaient plutôt clonaux, c'est-à-dire plus fréquemment partagés entre les sites tumoraux. La méthylation au niveau des sites d'initiation de la transcription (TSS) et des CGIs semble être relativement stable pendant la diversification sous-clonale dans le cas du cancer de la prostate.

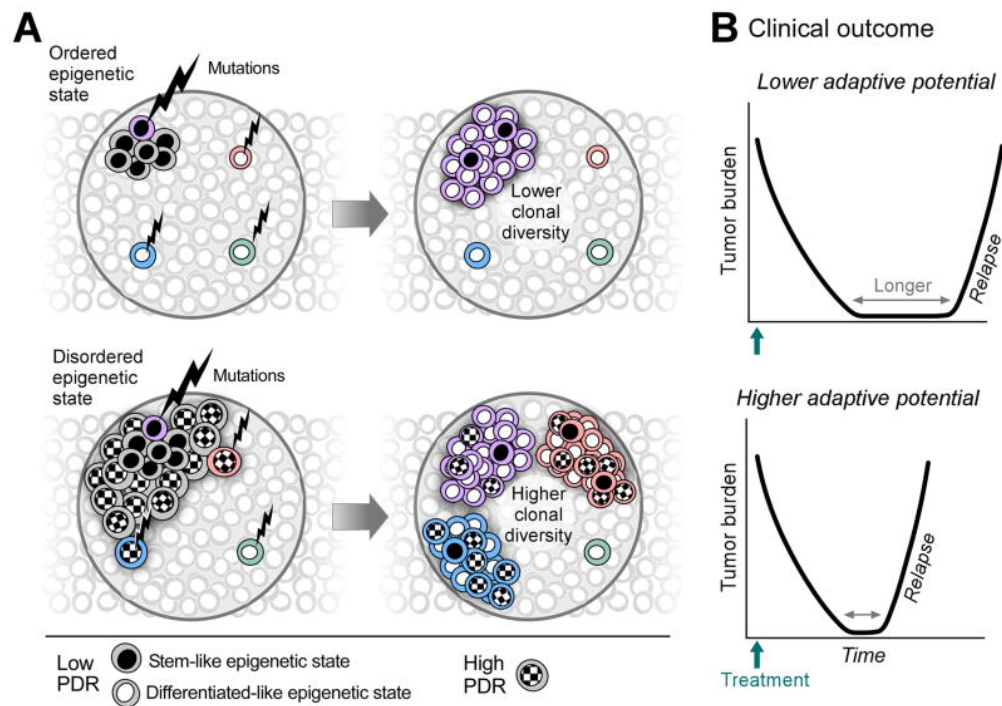


FIGURE 11 – Evolution clonale et perturbation de la méthylation de l'ADN proposé par Landau et al. (2014).

Landau et al. ont réalisé un séquençage au bisulfite du génome entier de cellules B normales, et de 104 leucémies lymphoïdes chroniques secondaires, et ont observé une variabilité intra-échantillon plus élevée des profils de méthylation de l'ADN des CLL par rapport aux cellules normales, provenant d'une méthylation aléatoire, désordonnée, des cellules malignes (Landau et al., 2014). Cette méthylation désordonnée est associée à un mauvais pronostic. Ils ont également étudié la relation entre la méthylation des promoteurs des gènes et leur expression, et ils ont montré qu'une méthylation désordonnée était associée à un paysage transcriptionnel plus "bruyant". Ils ont également proposé un modèle d'évolution clonale présenté sur la figure 11.

Dans les populations cellulaires présentant un faible pourcentage de reads discordants (PDR), donc une méthylation homogène, non désordonnée, lors d'une nouvelle mutation (représentée par des éclairs sur la figure 11) la proportion de cellules capables de participer activement à l'évolution clonale est faible (Figure 11 a et b partie haute). Alors qu'en revanche, dans les populations cellulaires avec un fort PDR, des forts niveaux de méthylation désordonnée, une plus grande proportion de cellules peut donner naissance à de nouveaux sous-clones, ce qui fournit une grande capacité d'adaptation pour la population cancéreuse et entraîne des résultats cliniques plus défavorables (Figure 11 a et b partie basse). Ils suggèrent que la perturbation désordonnée de la méthylation de l'ADN est susceptible d'avoir un rôle similaire à l'instabilité génétique, permettant aux cellules cancéreuses de s'adapter et d'évoluer en réponse à la thérapie.

Une association entre la progression du lymphome diffus à grandes cellules B (DLBCL) et la méthylation de l'ADN a été caractérisée par Pan et al. (2015). A partir de 13 paires d'échantillons diagnostic et rechute, ils ont montré que les patients présentaient une évolution hétérogène de leur profil de méthylation et suggèrent que l'hétérogénéité intra-tumorale des DLBCL au diagnostic est prédictive de la récurrence de la maladie. Ils ont également observé une diminution de l'hétérogénéité intra-tumorale de la méthylation au cours de l'évolution tumorale, ce qui est cohérent avec l'hypothèse d'un processus de sélection clonale après traitement. De plus, ils constatent que l'hétérogénéité de la méthylation n'est pas clairement corrélée à l'hétérogénéité génétique clonale, ces deux phénomènes doivent donc être étudiés en complémentarité.

Quek et al. ont généré pour leur étude des profils de méthylation de 48 régions tumorales spatialement distinctes, à partir de 11 adénocarcinomes pulmonaires et leur tissus pulmonaires normaux associés (Quek et al., 2017), et ils ont observé une hétérogénéité intra-tumorale de la méthylation au sein des mêmes tumeurs (restant tout de même plus faible que l'hétérogénéité inter-patient) associée à une plus grande taille de tumeur et à un risque accru de récurrence. Les différentes régions d'une même tumeur sont plus similaires entre elles que celles d'individus différents.

Dans le cas du glioblastome, Klughammer et al. ont mis en évidence une association entre l'hétérogénéité épigénétique et la survie des patient (Klughammer et al., 2018). Une grande hétérogénéité est associée à un mauvais pronostic. Par exemple, ils ont observé que le promoteur du gène *MGMT* était méthylé pour une partie des patients qui présentaient une survie globale significativement meilleure que les patients pour lesquels le promoteur de ce gène n'était pas méthylé. En revanche, ils n'ont pas trouvé de corrélation entre l'hétérogénéité et la taille de la tumeur.

Dans une étude récente, l'analyse des profils de méthylation de l'ADN de 27 régions tumorales primaires, 7 tissus normaux appariés et 6 métastases ganglionnaires de 7 patients atteints d'adénocarcinome pulmonaire a révélé que les profils de méthylation des tumeurs n'étaient pas plus similaires les uns par rapport aux autres que par rapport à ceux provenant de différents individus (Dietz et al., 2019).

Dans le cas du MM, de nombreuses études ont rapporté des hyperméthylations fréquentes chez plusieurs patients sur des inhibiteurs du cycle cellulaire p15 et p16 (Ng et al., 1997), des régulateurs de l'apoptose (Death-associated protein kinase (Ng et al., 2001)), et des gènes suppresseurs de tumeurs tels que *RASSF1A* (Ng et al., 2003), *SOC3-1* (Galm et al., 2003), ou *CDH1* (Braggio et al., 2010). L'hyperméthylation dans le promoteur de certains gènes suppresseurs de tumeurs tels que *DAPK* et *RAR β* a également été identifiée comme étant une caractéristique pronostique défavorable chez les patients atteints de MM (Braggio et al., 2010). Enfin, l'hyperméthylation de *RASD1* a été corrélée à la résistance du MM à la dexaméthasone (Nojima et al., 2009).

Walker et al. ont étudié des profils de méthylation de 27 000 sites CpGs de plasmocytes normaux, d'échantillons de MGUS, de MM au diagnostic, et de MM à des stades plus avancés (Walker et al., 2011). Ils ont mis en évidence que les différences majeures dans les profils de méthylation se produisaient lors de la transition MGUS au stade du MM, et du MM au stade de la leucémie à plasmocytes. Entre le stade MGUS et le MM, une hypométhylation globale a été observée, ce qui est souvent associée à une instabilité du génome et une modification de la structure de la chromatine, ainsi qu'une hyperméthylation plus localisée au niveau des régions promotrices de 77 gènes impliqués dans des voies de régulation du développement, du cycle cellulaire et de la transcription. Entre le MM et le stade de leucémie à plasmocytes, aucune vague de déméthylation n'a été observée, mais il a été mis en évidence une augmentation du nombre de gènes impactés par une hyperméthylation (n=1802 gènes). En particulier des gènes impliqués dans la signalisation et l'adhésion cellulaire, altérant ainsi l'adhésion des plasmocytes à leur niche conduisant à une croissance indépendante de la moelle osseuse.

Agirre et al. ont également relevé que l'hypométhylation dans le MM était beaucoup plus étendue et hétérogène que dans les MGUS (Agirre et al., 2015), ce qui suggère que cette déméthylation marquée peut être liée à la progression du MGUS asymptomatique vers le MM symptomatique. Ils ont également montré que les échantillon MGUS étaient moins hétérogènes que les MM mais qu'ils partageaient une signature d'hyperméthylation similaire. Les régions hyperméthylées dans le MM semblent se produire dans des zones qui sont devenues progressivement déméthylées au cours de la différenciation normale des cellules B, ce qui suggère que les cellules de MM acquièrent des caractéristiques de cellules indifférenciées.

4.5 Thérapie épigénétique

Ces dernières années, le progrès des nouvelles techniques génomiques a permis de mieux comprendre la complexité génétique et épigénétique des tumeurs. De plus en plus d'études mettent en avant une relation entre l'hétérogénéité intra-tumorale des tumeurs et certains résultats cliniques. Notamment, la résistance aux médicaments et la récurrence de la maladie sont souvent attribuées à l'émergence de sous-clones tumoraux. En effet, une majorité des thérapies anticancéreuses ne parviennent pas à obtenir des réponses durables, ce qui est souvent attribué à l'hétérogénéité intra-tumorale. Ces constats remettent en cause les approches universelles utilisées pour le traitement des cancers, qui n'évaluent pas toujours l'hétérogénéité intra-tumorale des patients, et met en avant la nécessité de la mise en place d'une médecine individualisée. Les changements épigénétiques sont réversibles, ce qui suggère la possibilité que certains médicaments pourraient agir et rétablir un état normal de ces marques. L'hétérogénéité des tumeurs peut être un obstacle majeur au diagnostic et au traitement efficace du cancer. Notamment, la sous-estimation de l'hétérogénéité des tumeurs peut entraîner une faille importante dans le diagnostic du cancer et la sélection du traitement. L'objectif ultime de la médecine de précision individualisée est de permettre la mise en place d'une intervention préventive ou thérapeutique la plus efficace adaptée au patient. Les changements épigénétiques aberrants pourraient servir de marqueurs de détection précoce et pronostiques dans le cancer, et servir de cibles préventives et thérapeutiques.

Le MM est une maladie caractérisée comme étant complexe et hétérogène, au cours de laquelle une combinaison d'événements génétiques primaires et secondaires se succèdent, accompagnée d'une hétérogénéité clonale. Cela rend complexe l'utilisation de thérapie ciblée qui doit prendre en compte ce contexte d'hétérogénéité. Par exemple, une thérapie ciblant spécifiquement une mutation présente dans seulement une fraction des cellules tumorales ne devrait affecter qu'un seul sous-clone et devrait donc présenter un bénéfice clinique limité, voire potentiellement délétère, si un sous-clone dépourvu de cette mutation ciblée, mais plus agressif, était stimulé suite à ce traitement. Pour une telle approche, il est essentiel de prendre en considération la diversité intra-clonale afin de prévenir l'émergence de sous-clones résistants au traitement. La progression tumorale post-traitement sera déterminée par le type de sous-clone qui va repeupler la niche des plasmocytes dans la moelle osseuse (Figure 12, Morgan et al. (2012b)).

Une thérapie post-traitement, d'entretien, pourrait être ensuite utilisée afin de permettre le repeuplement de la niche des plasmocytes par des clones plus indolents associés à de meilleurs résultats cliniques (Figure 12). En revanche, pour un tel traitement, il est important de stratifier les patients en fonction de leur risque clinique. Cette stratégie sur un patient à haut risque pourrait faciliter la sélection d'un clone plus agressif, ce qui nuirait à la survie du patient post rechute (Morgan et al., 2012b). Des essais sur une thérapie de maintien de la thalidomide ont été réalisés (Morgan et al., 2012a) et ont révélé une amélioration considérable de la survie des patients. L'action de la thalidomide comprend des effets apoptotiques et une modulation du micro-environnement de la moelle osseuse. Actuellement en France, un traitement de maintenance à la lénalidomide pendant deux ans est administré aux patients suite à une autogreffe. La lénalidomide est une molécule dérivée du thalidomide.

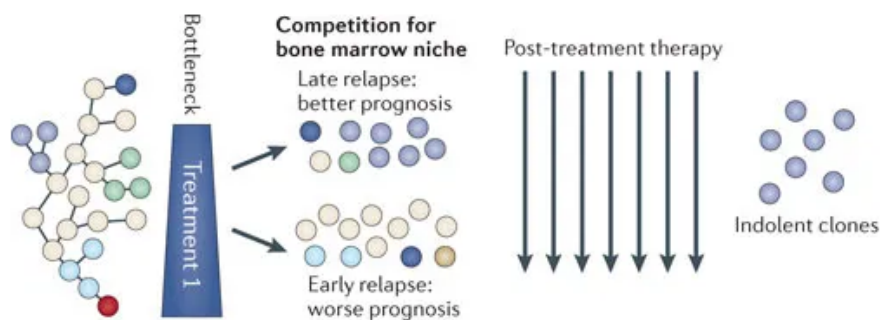


FIGURE 12 – Impact du traitement sur la diversité intraclonale. Figure tirée de Morgan et al. (2012b).

Morgan et al. (2012b) abordent dans leur article l'éventualité d'une intervention thérapeutique précoce dès le stade SMM. Walker et al. (2014) ont montré que le clone dominant du MM était déjà présent au stade SMM. Si ce clone dominant assez agressif pour faire évoluer la maladie était détecté dès ce stade et éliminé par un traitement précoce, il pourrait laisser place dans la niche des plasmocytes du MM à un clone plus indolent, ralentissant ainsi la progression et l'agressivité de la maladie.

4.6 L'objectif de la thèse

Au cours de ces dernières années, le développement des technologies de séquençage de nouvelle génération, et notamment le séquençage à forte profondeur, a permis de mettre en lumière toute la complexité génétique et épigénétique des tumeurs malignes. Les tumeurs ne sont pas constituées d'un ensemble de cellules uniques anormales mais plutôt d'un ensemble de sous-clones issus de plusieurs cellules anormales. La présence d'une certaine hétérogénéité génétique intra-tumorale a déjà été démontrée dans le cas du MM, les cellules propagatrices du myélome représentent un mélange de sous-clones génétiques concurrents, capables de maintenir les fonctions du clone du myélome tout en développant de nouvelles anomalies génétiques lui permettant d'acquérir un avantage sélectif, de s'étendre et de faire progresser la maladie. Chez la plupart des patients, il a été montré que la pression thérapeutique entraîne une réorganisation et une diversification des populations sous-clonales, avec l'émergence et l'expansion de sous-clones résistants au traitement. La connaissance de cette dynamique et hétérogénéité intraclonale est cruciale afin de mieux comprendre la progression du MM, la résistance aux traitements et pouvoir au mieux adapter de nouvelles stratégies thérapeutiques.

Il est bien établi que l'hétérogénéité génétique intra-tumorale contribue à l'évolution clonale et aux mécanismes de résistance dans le cas du MM, mais le faible nombre d'altérations génétiques détectées et pertinentes dans les mécanismes de rechute n'explique pas à lui seul la variabilité phénotypique du MM et les mécanismes moléculaires qui conduisent à l'initiation de la tumeur et sa progression. Dans ce contexte, nous avons émis comme hypothèse l'existence d'une hétérogénéité épigénétique contribuant à l'évolution, à l'agressivité et à la résistance du MM.

Afin d’approfondir cette hypothèse, nous avons réalisé une étude de la méthylation de l’ADN sur une cohorte de 21 patients atteints de myélome et 3 échantillons de cellules plasmocytaires normales, nous servant de contrôle. Le méthylome des échantillons de patients a été caractérisé en utilisant une technique de séquençage au bisulfite de sodium, le eRRBS (enhanced Reduced Representation Bisulfite Sequenced), pour 5 de ces patients nous avons également des données de Whole Genome Bisulfite Sequencing (WGBS, 10 échantillons au total : 5 diagnostics appariés à 5 rechutes). En parallèle, nous avons effectué une étude RNAseq qui a permis d’évaluer la variance épigénétique et transcriptionnelle au cours de la progression de la maladie. Nous avons également utilisé des données de MM disponibles dans le cadre du consortium blueprint : deux échantillons de plasmocytes normaux (NPC4 et NPC5), ainsi que 3 échantillons de MM, présentant des données de WGBS appariées à des données ChIP-seq (H3K9me3 et H3K27me3).

5 Échantillons et techniques utilisées

5.1 Description des échantillons

Pour notre étude, les cellules plasmocytaires de myélome sont issues de moelle osseuse prélevée à différents stades de la maladie chez 21 patients atteints de myélome suivis par l'Intergroupe Francophone du myélome (IFM) et pour lesquels un consentement était disponible. Cette étude inclut 11 femmes et 10 hommes, présentant un âge médian de 60 ans.

Les plasmocytes normaux et anormaux expriment à leur surface, un récepteur de la matrice extra-cellulaire : *CD138* (ou syndecan-1) qui permet de les caractériser. L'expression de ce récepteur chez les cellules hématopoïétiques est limitée aux plasmocytes, ce qui en fait un marqueur de choix. Les cellules de myélome ont donc été purifiées à l'aide de nanobilles et d'un anticorps anti-CD138 (RoboSep, Stem Cell Technologies). Après un tri immunomagnétique, la pureté de la suspension des plasmocytes a été vérifiée et seuls les échantillons contenant au moins 85% de plasmocytes ont été conservés pour les analyses génomiques. La pureté moyenne des cellules de MM était supérieure à 99% (comprise entre 87 et 100).

Des plasmocytes normaux, nous servant tout au long de l'analyse de contrôle, ont été obtenus à partir d'échantillons de moelle osseuse de trois patients suspectés d'être atteints de MGUS, et pour lesquels, un phénotype polyclonal a été observé et l'absence de plasmocytes anormaux a été évaluée par cytométrie en flux. La pureté moyenne des cellules CD138+ était supérieure à 75%.

L'ADN et l'ARN des cellules CD138+ ont été purifiés à l'aide du protocole Qiagen, et la qualité et la quantité des acides nucléiques ont été mesurées sur des profils Nanodrop, Qubit, et Agilent, et stockées dans une biobanque des hôpitaux de Nantes et Toulouse.

5.2 Les différents méthodes d'étude de la méthylation de l'ADN

L'analyse des niveaux de méthylation de l'ADN du génome peut se faire par plusieurs méthodes, qui présentent chacune différents avantages et inconvénients. Dans cette partie, je décrirai quelques exemples de technologies couramment utilisées pour l'analyse de la méthylation, que je finirai par comparer afin de vous expliquer notre choix de technologie pour mon projet de thèse.

5.2.1 Les puces à ADN, HumanMethylation 450k d'Illumina

Comme son nom l'indique, cette technique de séquençage interroge plus de 450 000 sites CpG (485 512 cytosines exactement) sur l'ensemble du génome, et couvre 99% des gènes RefSeq ainsi que 96% des îlot CpGs de la base de données UCSC (Bibikova et al., 2011). Cette technique est basée sur la conversion au bisulfite de sodium, qui permet de convertir les cytosines non méthylées en uraciles. Ainsi, suite à ce traitement, les seules cytosines qui persistent sont celles qui étaient méthylées à l'origine. Cette technologie repose sur deux tests chimiques différents : les tests Infinium I (Figure 13a) et Infinium II (Figure 13b).

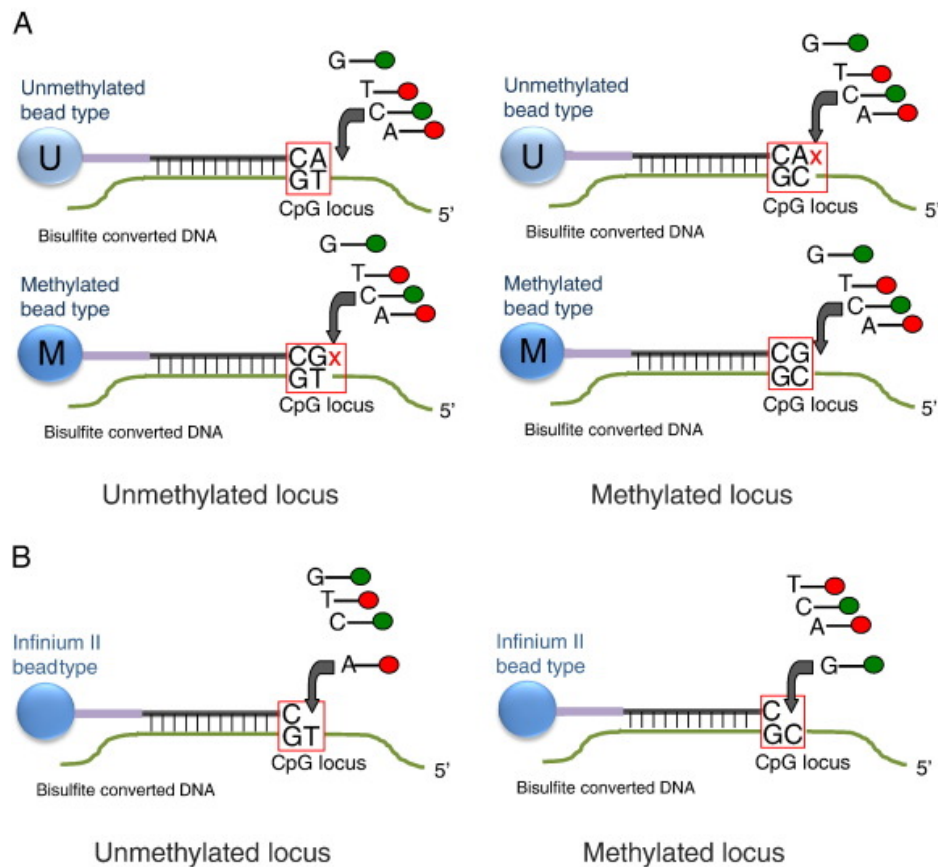


FIGURE 13 – Principe de la technologie HumanMethylation 450k Illumina. Fonctionnement des sondes Infinium I (a) et Infinium II (b). Figure tirée de Bibikova et al. (2011)

Le test Infinium I utilise deux types de sondes différentes de 50pb : une pour l'allèle méthylé ("M") et une pour l'allèle non méthylé ("U") (Figure 13a), situées sur deux types de billes différentes, et les signaux méthylés et non méthylés sont générés dans le même canal de couleur.

Le test Infinium II utilise un seul type de billes avec un type de sonde unique permettant la détection des deux allèles. Les signaux méthylés et non méthylés sont générés respectivement dans les canaux vert et rouge. Dans les deux cas, le pourcentage de méthylation d'une cytosine donnée est indiqué en tant que valeur β correspondant au rapport du signal méthylé sur la somme des signaux méthylés et non méthylés.

Malgré la couverture génomique limitée de cette technique, elle reste, de part son coût avantageux et sa capacité d'analyser jusqu'à 12 échantillons par puce, la plus utilisée à l'heure actuelle.

5.2.2 Methylated DNA immunoprecipitation

La technique de séquençage MeDIP (Methylated DNA immunoprecipitation) ne permet pas d'analyser la méthylation des cytosines individuellement, mais permet d'évaluer l'intensité de la méthylation de groupe de CpGs (Thu et al., 2009; Vucic et al., 2009). Une fois extrait et purifié, l'ADN est fragmenté par sonication, les fragments d'ADN sont ensuite dénaturés afin d'obtenir des ADN simple brin. Puis ces fragments sont immunoprécipités avec des anticorps qui reconnaissent la 5-méthylcytidine. Une fois les anticorps fixés sur les méthylcytosines, les fragments liés sont récupérés et ceux non liés sont éliminés dans le surnageant. Ainsi ces fragments méthylés peuvent être ensuite analysés par une technique de séquençage, qui permet de mesurer une intensité de méthylation pour une résolution de régions de 200 nucléotides en moyenne. L'avantage majeure de cette technique est son très bon rapport couverture du génome / prix, en effet cette technique permet de couvrir en moyenne environ 70% des CpGs du génome. Mais elle présente également un biais majeur pour les régions les plus riches en CpG (CpG islands et régions régulatrices), en effet la précision de cette mesure de la méthylation est faible dans les régions pauvres en CpGs, car les fragments les plus méthylés sont plus facilement précipités.

5.2.3 Enhanced Reduced Representation Bisulfite Sequencing

La technique de séquençage eRRBS (enhanced Reduced Representation Bisulfite Sequencing) correspond à une modification du protocole initial du RRBS (Reduced Representation Bisulfite Sequencing) afin d'élargir la couverture de la méthylation des CpGs (Garrett-Bakelman et al., 2015). L'ADN génomique est digéré en utilisant une enzyme de restriction MspI (Figure 14), elle cible les séquences 5' CCGG 3' et clive les liaisons phosphodiester en amont du CpG. Chaque fragment aura donc un CpG à chaque extrémité. En raison de la façon dont MspI clive l'ADN double brin, les brins se retrouvent avec des extrémités cohésives. Il est nécessaire d'effectuer une étape de « end repair + A-tailing » : les extrémités des fragments sont remplis, il y a ajout d'une adénosine supplémentaire pour chacun des brins. Ensuite, il y a l'ajout des adaptateurs avant le traitement au bisulfite de sodium, suivi d'une amplification des brins par PCR.

La différence majeure dans la préparation des bibliothèques eRRBS et RRBS réside dans la sélection de la taille des fragments. Pour les bibliothèques RRBS, seuls les courts fragments digérés par MspI (40-220bp) sont sélectionnés avant la conversion au bisulfite, ce qui concentre la sélection au niveau des CGIs. Alors qu'au contraire, pour les bibliothèques eRRBS, ce sont des fragments MspI allant de 70 à 320pb qui sont sélectionnés, ce qui augmente la capture des régions au-delà des CGIs. Ainsi, la technique eRRBS par rapport à la technique RRBS donne lieu à un nombre plus élevé de CpGs représentés dans les données générées, ainsi qu'à une augmentation de la couverture de toutes les régions génomiques interrogées. Cette technique permet de mesurer la méthylation à une résolution d'un CpG, mais permet de séquencer qu'environ 4% des CpGs du génome.

5.2.5 Comparaison des techniques

Chacune des techniques de séquençage précédemment décrites présentent leurs lots d'avantages et d'inconvénients (Figure 15).

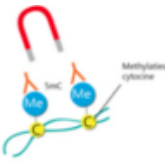
	 Human methylation 450k	 MeDIP-seq	 RRBS	 WGBS
Resolution	Single base	~150 bp	Single base	Single base
Reads per sample	Microarray	~50 million reads	~ 10 million reads	> 500 million reads
Region Covered	~ 1.7 % 485 000 CpGs (96 % CpGs islands, shores and shelves).	~ 70 % ~ 23 million CpGs	~ 3.7 % ~ 2 million CpGs	> 95 % > 28 million CpGs
Cost	Low	Moderate	Moderate	High
Notes	Don't covered regions outside the platform.	Antibody-based selection is biased toward hypermethylated regions.	Restriction enzymes cut at specific sites, providing biased sequence selection.	

FIGURE 15 – Comparaison des techniques de séquençage de la méthylation de l'ADN.

Pour notre étude, nous avons dans un premier temps fait le choix d'étudier la méthylation de l'ADN de nos échantillons de myélome avec comme technique de séquençage l'eRRBS. L'eRRBS étant le parfait compromis qualité/coût. La technique eRRBS a été privilégiée par rapport à la technique RRBS car elle permet d'élargir la couverture des CpGs en dehors des régions riches en CpGs. Elle a également été privilégiée par rapport au MeDIP-seq car elle permet l'analyse des CpGs individuellement, avec, pour un coût abordable, une couverture moyenne de 50X par CpG. Un autre de ces avantages majeurs : cette technique est applicable à partir d'une quantité limitée d'ADN génomique, ce qui rend possible le traitement d'échantillons cliniques humains. Dans le cadre de notre étude, les échantillons ont été obtenus suite à une ponction au niveau de l'os iliaque des patients, les quantités de matériels disponibles sont donc très limitées. Cette technique présente cependant quelques inconvénients : seuls 4% des CpGs du génome sont couverts et la conversion au bisulfite des cytosines non méthylées en uraciles peut être instables, la complexité de la séquence d'ADN est ainsi réduite.

Dans un second temps, nous avons réalisé pour 24 patients (50 échantillons en tout : 24 paires diagnostic / rechute et une paire rechute 1 / rechute 2) une analyse en WGBS, ce qui nous a permis d’avoir une vision globale de la méthylation du génome à une résolution d’un nucléotide, et de valider ce que l’on pouvait observer avec la technique eRRBS. Mais de par le coût de cette technique, nous n’avons pas pu travailler avec une aussi bonne profondeur de séquençage que pour l’eRRBS, nous avons obtenu en moyenne pour nos échantillons une couverture de 15X. Les données eRRBS, malgré leur faible couverture du génome, nous ont permis de réaliser une étude de l’hétérogénéité de la méthylation de nos patients, qu’il n’aurait pas été possible de réaliser en WGBS à cause de la faible profondeur de séquençage. Le WGBS nous a également permis d’avoir une vue d’ensemble de la méthylation sur le génome, au contraire de la technique eRRBS trop focalisée sur les régions riches en CpGs. Ces deux techniques ont été très complémentaires pour notre étude.

Toutefois, une des principales limitations de l’utilisation d’une méthode de séquençage au bisulfite est le risque d’une conversion incomplète. Les cytosines non méthylées sont converties en uraciles, tandis que les cytosines méthylées restent inchangées, si une conversion incomplète des cytosines non méthylées se produit, cela entraîne une surestimation des niveaux de méthylation, et inversement si la protection des cytosines méthylées n’est pas efficace, cela induit une conversion en uracile inattendue et une sous-estimation des niveaux de méthylation (Richards et al., 2018).

5.3 Le séquençage de l’ARN (RNA-seq)

Afin de pouvoir mesurer l’impact sur l’expression génique des modifications de la méthylation de l’ADN, nous avons réalisé le séquençage de l’ARN des échantillons analysés en eRRBS. Nous avons pour cela utilisé comme technique de séquençage à haut débit de l’ARN le RNA-seq, qui est la technologie la plus couramment utilisée pour identifier et quantifier l’expression des transcrits (Mortazavi et al., 2008). Le RNA-seq permet de séquencer et d’identifier les transcrits, et permet également d’avoir accès à leur niveau d’expression avec plus ou moins de précision selon la profondeur de séquençage. Plus la profondeur de séquençage sera importante, plus on augmentera nos chances de séquencer les transcrits les plus faiblement exprimés et plus facile sera la détection des différences d’expression entre deux conditions.

La plupart des ARNs sont en réalité des ARNs ribosomiques (environ 90% des ARNs), il faut donc s’assurer de les éliminer afin de pouvoir étudier uniquement les ARNs messagers, les ARNs codants. Pour cela, la sélection des ARNm matures se fait par leurs queues poly (A), réalisée avec des billes magnétiques ou en cellulose recouvertes de molécules oligonucléotides poly(T). La construction de la librairie comprend ensuite 4 étapes : la fragmentation de l’ARN, la synthèse de l’ADN double complémentaire (ADNc), la ligature d’adaptateurs et la quantification. Nos librairies d’RNA-seq ont été séquencées sur une machine HiSeq2500 d’Illumina, en paired-end pour des lectures de 100pb, par GATC Biotech. Le séquençage en paired-end facilite entre-autre l’alignement des séquences sur le génome.

5.4 Caractérisation des états de la chromatine

L'expression des gènes reposent sur l'accessibilité de la chromatine, qui est médiée par des complexes protéiques et des modifications épigénétiques. L'ensemble des protéines associées à la chromatine et les modifications épigénétiques dans une région constituent l'état chromatinien. Les récentes techniques de séquençage de nouvelles générations nous permettent maintenant de détecter les interactions de la chromatine et leur accessibilité. Je décrirai ci-dessous quelques méthodes actuellement disponibles permettant d'appréhender l'organisation de la chromatine.

5.4.1 ChIP-Seq et ChromHMM

Le ChIP-Seq (Chromatin ImmunoPrecipitation Sequencing) est une méthode utilisée pour analyser les interactions entre les protéines et l'ADN. Dans le cadre de l'étude des états de la chromatine, le but est de cartographier toutes les régions génomiques associées avec des histones portant des modifications post-traductionnelles telles qu'une acétylation ou méthylation, afin de caractériser l'activité transcriptionnelle de ces régions. Cette technique est basée sur l'immunoprécipitation de la chromatine avec des anticorps spécifiques d'histone, suivie par un séquençage à haut débit. On obtient alors, suite au séquençage, les régions d'ADN occupées par ces histones. Des outils ont été développés pour annoter l'état de la chromatine, notamment chromHMM (Ernst and Kellis, 2012), qui est un logiciel d'apprentissage, basé sur un modèle de Markov caché multivarié, qui modélise la présence ou l'absence de chaque marque de la chromatine, en intégrant plusieurs jeux de données de ChIP-seq de diverses modifications d'histone. Ces états de la chromatine sont analysés sur des fenêtres de 200pb, taille qui se rapproche approximativement de la taille des nucléosomes.

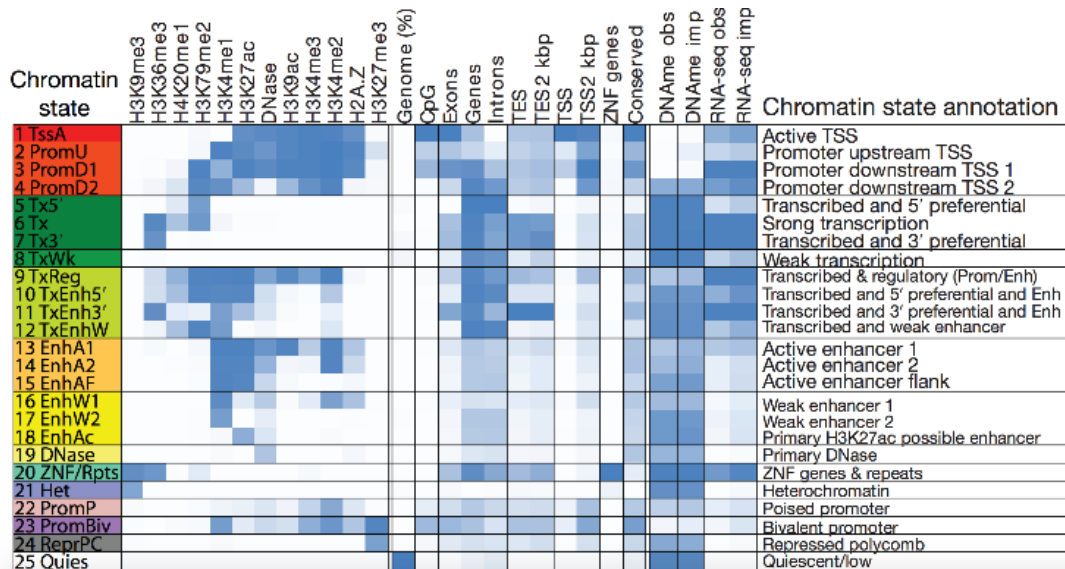


FIGURE 16 – Définition des 25 états de la chromatine par ChromHMM (Ernst et Kellis 2015)

Nous avons téléchargé la définition des états de la chromatine de 127 épigénomes de

référence (111 générés par le projet Roadmap Epigenomics et 16 générés par le projet ENCODE), qui couvrent divers types de cellules et tissus (Ernst and Kellis, 2015). A partir de 12 marques (DNase, H2A.Z, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K9me3, H3K27ac, H3K27me3, H3K36me3, H3K79me2 and H4K20me1), Ernst et son équipe ont pu annoter un ensemble de 25 états de la chromatine, formant ainsi l'information la plus complète sur l'état épigénomique du génome humain disponible à ce jour (Figure 16).

5.4.2 Hi-C

La technique Hi-C (High Chromosome Contact map) permet de détecter toutes les interactions intra- et inter-chromosomiques, et génère des cartes de contact entre toutes les parties du génome. Le HiC est un outil puissant pour révéler la compartimentation du génome. Son protocole repose sur une extension de la technologie de Capture de Conformation des Chromosomes (3C) couplée à un séquençage à haut débit.

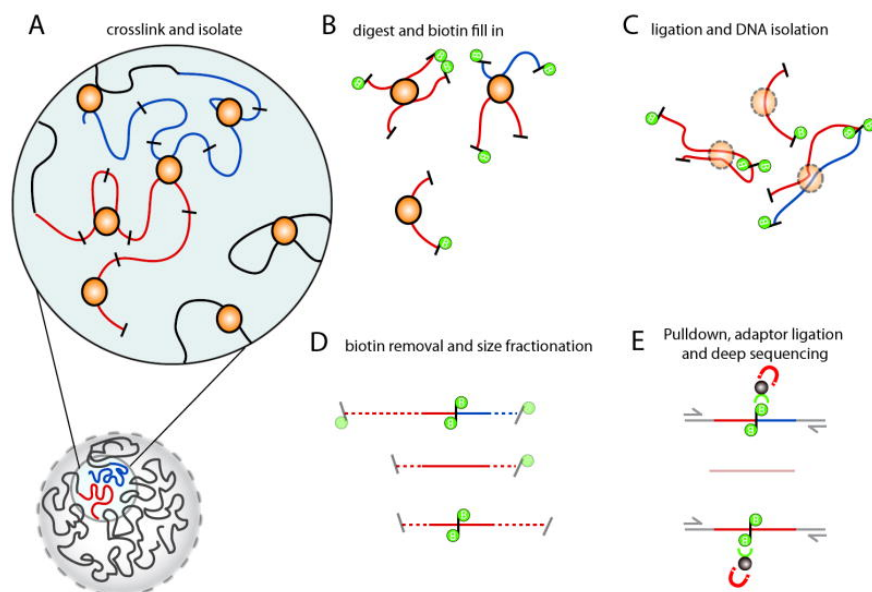


FIGURE 17 – Protocole du HiC. (a) Crosslink et isolation des fragments d'ADN qui interagissent entre eux par réticulation (b) Digestion des fragments d'ADN par une enzyme de restriction, et marquage des extrémités par un nucléotide biotinylé. (c) Ligature des fragments d'ADN proches spatialement dans le noyau et purification. (d) Élimination des extrémités non ligaturées marquées par la biotine et sélection des fragments par leur taille (e) Sélection des fragments avec incorporation interne de biotine par billes magnétiques revêtues de streptavidine, et préparation pour le séquençage. Figure extraite de Belton et al. (2012)

Cette technique repose sur un crosslink qui est réalisé sur l'ADN des cellules, tous les fragments d'ADN proches les uns des autres vont être isolés en réticulant par covalence les complexes protéines/ADN avec du formaldéhyde (Figure 17a). Ensuite, la chromatine est solubilisée et fragmentée. La solubilisation permet d'éliminer les protéines non réticulées et

ouvre la chromatine. La chromatine accessible est ensuite digérée par une enzyme de restriction, en général Hind III, et les extrémités sont marquées avec un nucléotide biotinylé (Figure 17b). Les fragments d'ADN qui interagissaient ensemble sont ensuite ligaturés et purifiés, créant ainsi une librairie de fragments d'ADN spatialement proches dans le noyau (Figure 17c). Les extrémités marquées par la biotine qui n'ont pas été ligaturées ensemble doivent être éliminées afin qu'elles ne soient pas captées par les billes recouvertes de streptavidine avec les véritables produits de ligature (Figure 17d). Les fragments sont ensuite sélectionnés selon leur taille, les produits de ligature HiC ont une taille optimale de 150 à 300pb. Les produits de ligature avec incorporation interne de biotine sont extraits avec des billes magnétiques revêtues de streptavidine, la ligature des adaptateurs est ensuite faite, la librairie est amplifiée par PCR puis séquencée en profondeur afin de pouvoir au mieux quantifier les interactions de la chromatine (Figure 17e) (Belton et al., 2012).

La technique Hi-C permet d'obtenir une matrice de contact du génome qui mesure les interactions entre les régions génomiques, la résolution de cette matrice est déterminée par la profondeur de séquençage. Les données d'HiC permettent de mieux comprendre l'organisation de la chromatine, et notamment de définir la compartimentation du génome en régions de chromatine ouvertes (compartiment A) et fermées (compartiment B).

Ces données permettent également l'identification de domaines d'association topologique (TAD), qui sont des unités fonctionnelles de l'ordre du mégabase, permettant de réguler l'expression des gènes en facilitant les interactions entre les enhancers et les promoteurs des gènes.

5.4.3 Repli-Seq

La technique de séquençage Repli-seq permet d'étudier les temps de réplication à l'échelle du génome (Hansen et al., 2010). A partir de cellules en cultures, les cellules proliférantes sont marquées par du 5-bromo-2-désoxyuridine (BrdU) afin d'identifier l'ADN nouvellement répliqué. Les cellules sont fractionnées en 6 parties correspondant au cycle cellulaire (G1 / G1b, S1, S2, S3, S4, G2), par une cytométrie en flux en fonction de leur teneur en ADN. Puis les brins d'ADN marqués par le BrdU sont immunoprécipités par des anticorps anti-BrdU sur des billes magnétiques. Les fragments sont ensuite séquencés puis alignés sur le génome de référence hg19.

Les signaux de réplication dans chaque fraction du cycle correspondent à la densité des séquences alignées dans une fenêtre génomique coulissante de 50 kb. Pour éviter les variations dues au biais de séquençage, les signaux de réplication sont ensuite normalisés. Nous avons récupéré les temps de réplication de la lignée cellulaire GM12878 sur ENCODE, et plus exactement le fichier : *Wavelet-smoothed Signal* (<http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeUwRepliSeq/>), contenant un signal lissé des six fractions du cycle correspondant à une moyenne pondérée des signaux normalisés de sorte qu'une réplication antérieure coïncide avec les valeurs les plus élevées du fichier.

5.5 Analyse bioinformatique des données

L'expansion du séquençage à haut débit et des analyses épigénétiques ont conduit au développements constants de nouveaux outils d'analyse de données. Ci dessous une présentation des outils existants utilisés ou des pipelines, groupe de logiciels utilisés en série, que nous avons mis au point pour l'analyse de nos données.

5.5.1 Analyse des données eRRBS

Pour analyser les données eRRBS, le pipeline d'analyse que j'ai mis en place est disponible sur gitlab :

<https://gitlab.univ-nantes.fr/JenniferDERRIEN/epipolymorphism-pipeline>.

Ce pipeline regroupe plusieurs logiciels parfaitement adaptés pour des données de type eRRBS et pour l'analyse que nous avons souhaité réaliser.

5.5.1.1 Contrôle qualité et filtration des données

Nous avons utilisé l'outil Trim galore afin de pouvoir contrôler la qualité des données, et de pouvoir détecter et supprimer les contaminations par des adaptateurs Illumina (Krueger, 2012). Pour filtrer les contaminations par les adaptateurs illumina, cet outil utilise par défaut les 13 premières paires de bases de l'adaptateur illumina standard ('AGATCGGAAGAGC'). Pour les librairies eRRBS digérées par l'enzyme de restriction MspI, Trim galore effectue le filtrage des données en 2 étapes, ce qui lui permet d'éliminer les 2 bases supplémentaires, contenant une cytosine, qui ont été artificiellement introduite lors de la préparation de la librairie (Garrett-Bakelman et al., 2015).

Dans le cas de notre analyse, nous avons utilisé comme option pour Trim galore : "-rrbs", "-non directionnal", "-paired" et "-fastqc". L'option "-fastqc" permet suite au filtrage d'effectuer un contrôle qualité des données. Fastqc fournit un aperçu rapide de la qualité des données, et peut aider à situer dans quel domaine les données peuvent présenter des anomalies (Andrews et al., 2010). Il fournit des graphes et des tableaux permettant d'évaluer rapidement les données. Les résultats sont exportés sous forme HTML.

5.5.1.2 Alignement contre le génome de référence

Bismark, qui est un programme spécialement conçu pour aligner des séquences traitées au bisulfite de sodium contre un génome d'intérêt (Krueger and Andrews, 2011), a été utilisé afin d'aligner nos données de séquençage sur le génome de référence hg19 publié par le Genome Reference Consortium en février 2009 (Consortium et al., 2009).

La première étape fut de préparer le génome de référence (hg19) avant l'alignement, grâce à l'outil *bismark genome preparation* qui permet d'indexer le génome pour des données traitées au bisulfite. Puis Bismark utilise l'aligneur bowtie2 (Langmead and Salzberg, 2012) afin d'aligner les reads sur le génome de référence et produit en sortie un fichier BAM d'alignement ainsi qu'un rapport d'exécution contenant :

- le résumé des paramètres d'alignement utilisés
- le nombre de séquences analysées
- l'efficacité de l'alignement
- le nombre de cytosines analysées
- le nombre de cytosines méthylées ou non
- le pourcentage de cytosine méthylée dans les différents contextes (CpG, ou le contexte CHH, ou H peut être A, T ou C).

5.5.1.3 Extraction des CpGs et leurs niveaux de méthylation

Le package R methylKit a été utilisé pour extraire les CpGs des fichiers d'alignements (Akalin et al., 2012). MethylKit permet l'analyse différentielle de la méthylation, le regroupement d'échantillons, l'annotation et la visualisation des niveaux de méthylation. Ce package a été spécialement conçu pour traiter des données de séquençage RRBS et ses variantes. Il prend en entrée un fichier d'alignement SAM trié, il en extrait les informations de chaque CpG contenu dans le fichier d'alignement et calcule leur niveau de méthylation (Figure 18).

chrBase	chr	base	strand	coverage	freqC	freqT
chr1.10563	chr1	10563	F	513	97.66	2.34
chr1.10542	chr1	10542	F	514	97.28	2.72
chr1.10564	chr1	10564	R	632	100.00	0.00
chr1.10577	chr1	10577	F	258	16.67	83.33
chr1.10571	chr1	10571	F	513	98.44	1.56
chr1.10590	chr1	10590	R	314	0.00	100.00
chr1.10526	chr1	10526	R	639	99.37	0.63

FIGURE 18 – Extrait d'un fichier d'information de la méthylation en sortie de methylKit

Nous avons fixé comme seuil de couverture un minimum de 10X pour chaque CpG dont le niveau de méthylation a pu être extrait. La colonne du tableau "freqC" correspond au pourcentage de cytosine alignée pour cette base par rapport à sa couverture, et la colonne "freqT" correspond au pourcentage de thymine. La colonne "freqC" donne donc le niveau de méthylation de chaque CpG, puisque toutes les cytosines non méthylées ont été transformées en thymine lors du traitement par le bisulfite de sodium.

d'épipolymorphisme, différents niveaux d'hétérogénéité. Ces trois exemples de loci avec le même niveau de méthylation d'ADN mais des épipolymorphismes très différents mettent parfaitement en lumière les limites de l'analyse uniquement de la moyenne de la méthylation de l'ADN.

Le choix du nombre de CpGs adjacents composant l'épiallèle, c'est-à-dire quatre, a été optimisé par les créateurs de l'outil methclone. Ils ont mesuré le nombre de CpGs couverts par reads, calculé ensuite le nombre de patterns de méthylation possibles par rapport aux nombres de CpG définissant l'épiallèle, puis ils ont recalculé le nombre de reads qu'ils pouvaient mesurer avec ces différentes tailles d'épiallèles. Ils ont observé une forte baisse du nombre de reads par taille d'épiallèle lors du passage de 4 CpGs à 5 CpGs. Plus l'épiallèle comporte de CpGs et plus le nombre de reads nécessaires pour couvrir l'ensemble des patterns de méthylation possible est élevé. C'est ainsi que le choix s'est porté sur l'analyse de 4 CpGs adjacents formant 16 patterns de méthylation différents. Pour permettre de couvrir l'ensemble des patterns de méthylation, un seuil de minimum 60X de profondeur de séquençage a été fixé.

5.5.1.5 Identification des changements significatifs en composition épiallélique

Afin d'analyser la dynamique épiallélique au cours du développement et de la progression du MM, c'est à dire les changements en composition épiallélique entre deux états/stades de la maladie, par exemple l'état sain (NPC) et le diagnostic du MM, ou entre le diagnostic du MM et sa rechute, nous avons utilisé une autre fonctionnalité de l'outil methclone.

A partir des loci de 4 CpGs adjacents calculés pour chaque échantillon, nous avons mesuré pour chaque locus les changements en composition de patterns de méthylation entre 2 stades de la maladie, notamment le stade sain et le stade diagnostic dont quelques exemples sont présentés sur la figure 20. Pour chaque pattern de méthylation possible, un code couleur est associé.

Methclone mesure les changements de composition en épiallèles de tous les loci de 4 CpGs par une valeur d'entropie combinatoire (ΔS), qui peut varier de 0 à -144. Lorsque les épiallèles ont exactement la même distribution, le delta d'entropie vaut 0 ($\Delta S=0$), et il est maximal lorsque les distributions s'intervertissent complètement : un unique pattern de méthylation pour le premier stade et un unique pattern, différent du premier, pour le second stade (ΔS qui se rapproche de -144). Sur les deux exemples à gauche de la figure 20, les changements en composition épialléliques sont nulles ($\Delta S = 0$) ou assez faible ($\Delta S = -20$). Afin de pouvoir identifier uniquement les changements les plus significatifs, nous avons fixé comme seuil qu'un locus présente un changement significatif de pattern de méthylation entre deux états lorsque le ΔS est inférieur à -70, ce locus est par la suite qualifié d'elocus. A droite de la figure 20, deux exemples d'eloci présentent certaines particularités. Le premier locus est dans un premier temps très hétérogène au stade sain et on observe une sélection d'un pattern de méthylation qui était minoritaire et qui devient majoritaire au diagnostic, l'hétérogénéité épigénétique diminue. Dans le second exemple d'eloci, au contraire, le pattern majoritaire au stade sain disparaît au diagnostic pour laisser place à un mélange hétérogène de patterns de méthylation, il y a une diversification des patterns, l'hétérogénéité épigénétique augmente.

Pour finir, afin de pouvoir comparer les échantillons entre eux, nous avons utilisé une variable nommée EPM, permettant de normaliser le nombre d'eloci sur le nombre total de loci séquencés par échantillon, la formule est présentée en bas de la figure 20 (Li et al., 2014)

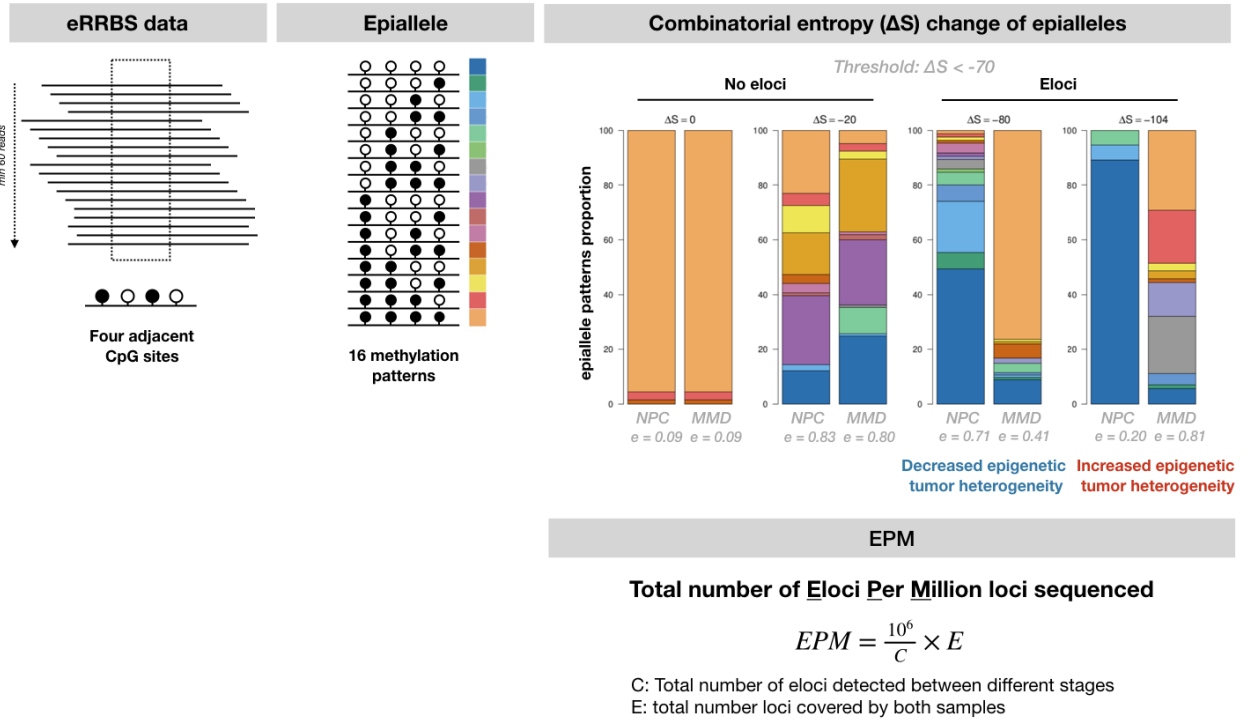


FIGURE 20 – Détection des changements épialléliques. A partir d'une valeur d'entropie combinatoire $\Delta S < -70$, methclone détermine les loci présentant un changement significatif de composition épiallélique (= eloci).

N'ayant pas d'échantillons sains appariés aux échantillons de MM, nous avons défini les loci entre le stade NPC et MM comme étant des eloci uniquement s'ils présentaient un ΔS inférieur à -70 pour nos 3 échantillons de NPCs.

Nous avons défini un eloci hypométhylé ou hyperméthylé comme étant un elocus présentant une différence de méthylation entre les deux états d'au moins 25%.

Enfin, dans le but de définir au mieux l'état de méthylation de l'ADN de ces loci, nous avons utilisé une autre variable nommée PDR pour Proportion of Discordant Reads, la proportion de reads discordants (Landau et al., 2014), comme étant le pourcentage de loci de 4 CpGs adjacents non totalement méthylés ou non totalement déméthylés, c'est à dire le pourcentage de reads présentant une méthylation désordonnée (Figure 21).

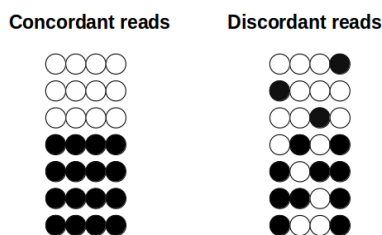


FIGURE 21 – Définition des reads concordants et discordants

Le PDR et l'EPM sont deux mesures essentielles et complémentaires afin de caractériser au mieux l'état de l'hétérogénéité des profils de méthylation pour ce type de données.

5.5.2 Analyse des données WGBS

Les données brutes de WGBS ont été traitées par le DKFZ avec leur pipeline d'analyse disponible sur github à cette adresse :

<https://github.com/DKFZ-ODCF/AlignmentAndQCWorkflows>

Le contrôle qualité des données et la suppression des adaptateurs ont été réalisés avec l'outil trimmomatic (Bolger et al., 2014). L'alignement des reads sur le génome de référence a été réalisée avec l'aligneur methylTools (Hovestadt et al., 2014), qui utilise l'algorithme d'alignement Burrows-Wheeler. MethylTools utilise une approche d'alignement similaire à celle de bismark en améliorant la gestion des grandes quantités de données et la vitesse de l'alignement. L'extraction des CpGs et le calcul de leur niveau de méthylation ont été réalisés avec methylTools et le package R Bseq (Hansen et al., 2012), permettant d'analyser, de gérer et de stocker des données de WGBS. Comme pour les données eRRBS, nous avons fixé comme seuil de couverture un minimum de 10X pour chaque CpG séquencé pour nos analyses.

A partir de ces données, nous avons cherché à définir pour tous les échantillons de WGBS des régions nommées domaines partiellement méthylés (PMDs). La détection des PMDs dans les échantillons de WGBS a été effectuée à l'aide du package R MethylSeekR (Burger et al., 2013). Avant la détection des PMDs, les CpGs chevauchant des SNPs (single-nucleotide polymorphism) communs ont été supprimés des données (dbSNP137commonhg19, version 1.0.0). Un SNP correspond à la variation d'une seule paire de base apparaissant chez plus de 1% des individus d'une population. La distribution des valeurs α a été utilisée pour déterminer la présence ou non de PMDs dans les échantillons. La valeur α caractérise la distribution des niveaux de méthylation dans des fenêtres génomiques coulissantes de 100 CpGs. La fonction `plotAlphaDistributionOneChr()` permet de générer une figure montrant la distribution des valeurs α pour un chromosome (Figure 22), cette distribution est similaire pour tous les chromosomes. Si la distribution des valeurs α est bimodale ou présente une grande fraction de valeur α supérieures ou égales à 1, alors c'est une preuve de présence de PMDs dans l'échantillon (Figure 22a). Une distribution des valeurs α inférieure à 1 est observée dans les méthylomes sans PMD (Figure 22b). Si la distribution α est bien bimodale, alors on peut exécuter la fonction `segmentPMDs()` qui permet d'identifier les PMDs sur le

génomique par un modèle de Markov caché. Cette fonction génère un objet GRanges avec une première colonne indiquant le partitionnement du génome en PMD et régions hors PMD, et une seconde colonne indiquant le nombre de CpG couvert par chaque région. Cette fonction génère une nouvelle distribution α mais cette fois-ci avec deux distributions gaussiennes ajustées du modèle de Markov caché en rouge et vert (Figure 22a). Si le modèle de Markov caché fonctionne correctement, les deux gaussiennes doivent correspondre aux deux pics de la distribution bimodale.

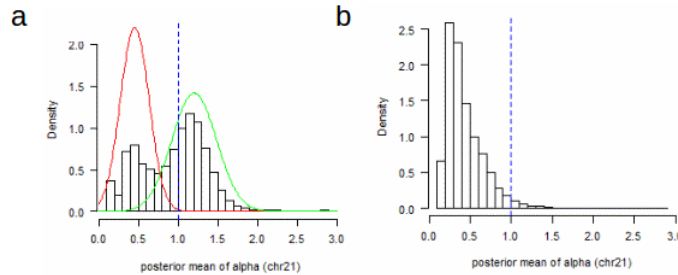


FIGURE 22 – Deux exemples de distributions α sur le chromosome 21. À gauche un exemple de distribution alpha bimodale avec une partie des valeurs supérieures à 1 (ligne bleu en pointillé) ; la bimodalité de la distribution est représentée par les deux courbes rouge et verte ; à droite une distribution α non bimodale, la majorité des valeurs sont inférieures à 1.

5.5.3 Analyse des données de RNAseq

L'application la plus courante de la technique RNAseq est la quantification de l'expression des gènes par l'obtention d'une table de comptage répertoriant par transcrit le nombre de reads séquencés. L'abondance des ARNs dans une cellule est proportionnelle au nombre de reads.

Les bibliothèques de RNAseq ont été séquencées sur une machine Illumina HiSeq2500 en utilisant des reads appariés de 100 pb. Le nombre moyen de reads séquencés par échantillon est d'environ 68 millions ; les reads sont contenus dans des fichiers FASTQ qui ont subi un contrôle qualité grâce à l'outil FASTQC (version 0.11.4) (Andrews et al., 2010), permettant d'éliminer les reads de mauvaise qualité. Les reads ont été ensuite alignés sur le génome de référence hg19 avec l'aligneur STAR (version 2.4.0f1) (Dobin et al., 2013). Les duplicats de PCR ont été identifiés en utilisant l'outil Picard (Pic, 2019) puis éliminés des données. Avec nos échantillons, nous avons obtenu un taux d'alignement moyen des reads de façon unique sur le génome de référence de 66%. Le taux d'alignement n'est jamais parfait pour différentes raisons, comme des erreurs de séquençage, des mutations, et on sait aussi que le génome de référence hg19 lui-même n'est pas parfait puisqu'il correspond à un assemblage de 13 génomes séquencés de personnes volontaires. Suite au séquençage, nous obtenons donc un fichier SAM contenant l'information de l'alignement ou non de chaque read, à partir duquel nous avons compté le nombre de reads alignés sur chaque gène du génome. Ce comptage a été fait à l'aide

de l'outil HTSeq-count (Anders et al., 2015), version 0.6.0, ainsi qu'à partir d'un fichier GTF renseignant les coordonnées de chaque transcrit.

Le nombre de reads bruts par transcrit ne suffit pas à lui seul pour comparer les niveaux d'expression parmi les échantillons, car ces valeurs sont dépendantes de plusieurs facteurs tels que la longueur des gènes, le nombre total de reads séquencés et les biais de séquençage. Plus un gène est long et plus son compte de reads sera important. Afin de pouvoir comparer l'expression de plusieurs gènes dans un même échantillon, il est important de normaliser les comptes de reads par transcrit sur la taille du gène. De même qu'il est important de normaliser les comptes de plusieurs échantillons afin de pouvoir les comparer entre eux, si un échantillon est séquencé plus profondément que l'autre, c'est à dire s'il a un nombre total de reads alignés plus important, les tables de comptage ne seront pas comparables.

La normalisation la plus utilisée en recherche est le RPKM (Reads Per KiloBase Per Million Mapped) :

Formule du RPKM :

$$\text{RPKM} = \frac{\text{numRead}}{\text{sizeGene}/1000 * \text{totRead}/1000000}$$

numRead = nombre de reads alignés pour un gène

sizeGene = longueur du gène en paire de bases (pb).

totRead = Nombre total de reads séquencés de l'échantillon.

Cette approche a été initialement introduite pour faciliter les comparaisons entre les gènes au sein d'un échantillon, et combine la normalisation inter et intra-échantillon, car elle redimensionne les comptes de transcrit en corrigeant les différences de taille de librairie et de longueur de gène (Mortazavi et al., 2008). Cependant, il a été démontré que tenter de corriger les différences de longueur de gène dans une analyse différentielle a en fait pour effet d'introduire un biais dans les variances par gène pour les gènes faiblement exprimés (Dillies et al., 2013).

Pour une analyse différentielle d'expression, comparaison entre échantillons, il est recommandé d'utiliser une autre méthode de normalisation telle que la méthode RLE (Relative Log Expression) que propose le package R DESeq2 (Love et al., 2014). Pour la normalisation de la table de comptage, cette méthode utilise un facteur d'échelle obtenu en calculant la moyenne géométrique des comptages de chaque gène de l'échantillon. Facteur d'échelle par lequel tous les comptes seront multipliés pour corriger le biais de profondeur de séquençage et ainsi rendre les échantillons comparables. Ce package permet également de déterminer les gènes différentiellement exprimés entre deux conditions en se basant sur la distribution négative binomiale. Cette méthode se base sur deux hypothèses : la majorité des gènes ne sont pas différentiellement exprimés et il y a autant de gènes sur-exprimés que sous-exprimés. Pour déterminer les gènes différentiellement exprimés dans ce genre d'étude, la p-value classique n'est pas adaptée. Dans le cas de tests multiples, il est préférable d'utiliser une p-value ajustée par une procédure de Bonferroni (Bland and Altman, 1995) ou de Benjamini-Hochberg (Hochberg and Benjamini, 1990), ces deux méthodes ayant pour but de contrôler le nombre de faux positifs dans les gènes déclarés à tort différentiellement exprimés.

5.5.4 Analyse des données d’HiC

L’analyse des données d’HiC se fait à partir de carte de contact chromosomique qui sont des matrices dans lesquelles sont répertoriées les fréquences de contacts de chaque paire de reads appariés séquencés.

Nous avons récupéré des données brutes d’HiC pour 3 lignées cellulaires : la GM12878 (ENCODE), RPMI-8226 et U266 (Wu et al., 2017). Puis nous avons utilisé le pipeline HiC-Pro (Servant et al., 2015), spécialement conçu pour traiter les données Hi-C, afin de transformer les données brutes (reads appariés, paired-end) en cartes de contact normalisées. HiC-Pro nécessite trois fichiers d’annotation pour son fonctionnement :

- Un fichier BED contenant la position des fragments de restriction après digestion. Ce fichier dépendant à la fois de l’enzyme de restriction et du génome de référence.
- Une table contenant la taille des chromosomes du génome de référence.
- Un index bowtie2 du génome de référence.

Il est également nécessaire de renseigner la résolution de la carte de contact que l’on souhaite définir, c’est à dire la taille des fenêtres utilisées pour découper le génome, dans lesquelles les interactions vont être mesurées. Une résolution comprise entre 10 000 et 50 000pb est idéale pour étudier les TADs et la compartimentation du génome.

Tout d’abord, les deux parties des reads appariés sont alignées indépendamment sur le génome de référence avec l’aligneur bowtie2. Chaque read aligné peut être affecté à un fragment de restriction. Seules les paires de reads impliquant deux fragments de restriction différents sont gardées pour la création de la carte de contact. Les paires de reads dupliquées dues aux artefacts de PCR sont éliminées. Ensuite HiC-Pro effectue un contrôle qualité de l’alignement des reads, les reads ne remplissant pas leurs critères qualités sont éliminés, seules les paires de reads valides sont utilisées pour créer les cartes de contacts.

Afin de générer les cartes de contacts intra et inter-chromosomiques, le génome est divisé en fenêtres de tailles égales et le nombre de contacts observés entre chaque paire de fenêtres génomiques est répertorié dans une carte de contact brute. Pour notre analyse, nous avons choisi de travailler sur des cartes de contacts à une résolution de 20kb.

Les données d’HiC peuvent présenter plusieurs sources de biais, comme toute expérience de séquençage, qui doivent être corrigés. HiC-Pro propose une implémentation rapide de l’algorithme de normalisation ICE (Iterative Correction and Eigenvector decomposition, Imakaev et al. (2012)), aboutissant à une matrice de probabilités de contact relatives, permettant une comparaison des données d’HiC inter et intra-échantillon.

5.5.4.1 Détermination des compartiments A et B

Chaque chromosome peut être décomposé en deux types de compartiments : le compartiment A riche en gènes et présentant des régions de chromatine ouvertes/actives et le compartiment B pauvre en gène et présentant des régions de chromatine plus condensées/fermées. L’appartenance d’une région à l’un ou l’autre des compartiments peut se faire à partir des données d’HiC.

Pour cela nous avons utilisé le package R HiTC (Servant et al., 2012), qui permet d’explorer des données HiC, de les transformer, les normaliser, les annoter et de visualiser des cartes d’interactions. Et notamment nous avons plus spécifiquement utilisé leur fonction *pca.hic*, qui permet d’effectuer une analyse en composante principale sur les cartes de contacts produites par HiC-Pro. La détermination des compartiments A et B est estimée par l’analyse des vecteurs propres de la matrice de contact du génome par la méthode observed-expected, qui normalise les interactions entre deux régions séparées par une distance donnée, par l’interaction moyenne entre toutes les régions séparées par cette même distance. Les changements de signe des vecteurs propres de la matrice de contact correspondent aux changements / limites des compartiments du génome, ainsi sur cette base et en prenant en compte la densité des gènes, nous avons pu définir pour les 3 lignées cellulaires à notre disposition les régions du génome correspondant aux compartiments A et B (Figure 23)

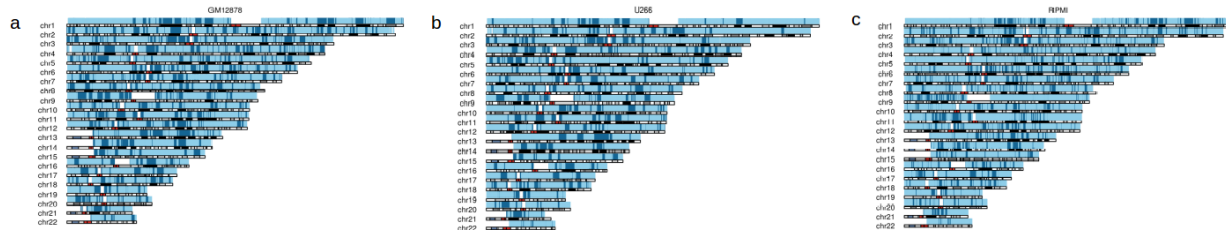


FIGURE 23 – Répartition génomique des compartiments A et B pour les lignées cellulaires de la GM12878 (a), U266 (b) et RPMI-8226 (c). En bleu clair, le compartiment A et en bleu foncé, le compartiment B.

Une étude a montré les possibilités et la fiabilité pour déterminer la compartimentation du génome à partir de différentes données génomiques telles que les puces à ADN (Illumina 450k), la DNase-seq (séquençage des sites hypersensibles à la DNase I), le single-cell ATAC-seq (Assay for Transposase-Accessible Chromatin with highthroughput sequencing) et le single-cell WGBS (Fortin and Hansen, 2015). Fortin et al ont obtenu un accord de 88,8% entre leurs prédictions basées sur des données de puces ADN 450k et leurs prédictions basées sur les données d’HiC. Cependant aucune méthode actuelle, aucun pipeline d’analyse n’a été développé, à ma connaissance afin de prédire la compartimentation du génome à partir de données eRRBS. Jenkinson et al. (2017) ont décrit une méthode de compartimentation du génome à partir de données WGBS, mais aucun outil n’a été développé pour l’instant.

5.5.4.2 Détermination des domaines d’associations topologiques

Pour identifier les TADs à partir des données d’HiC, nous avons utilisé l’outil HiCExplorer, qui est un ensemble de programmes servant à traiter, normaliser, analyser et visualiser les données Hi-C (Ramírez et al., 2018; Wolff et al., 2018).

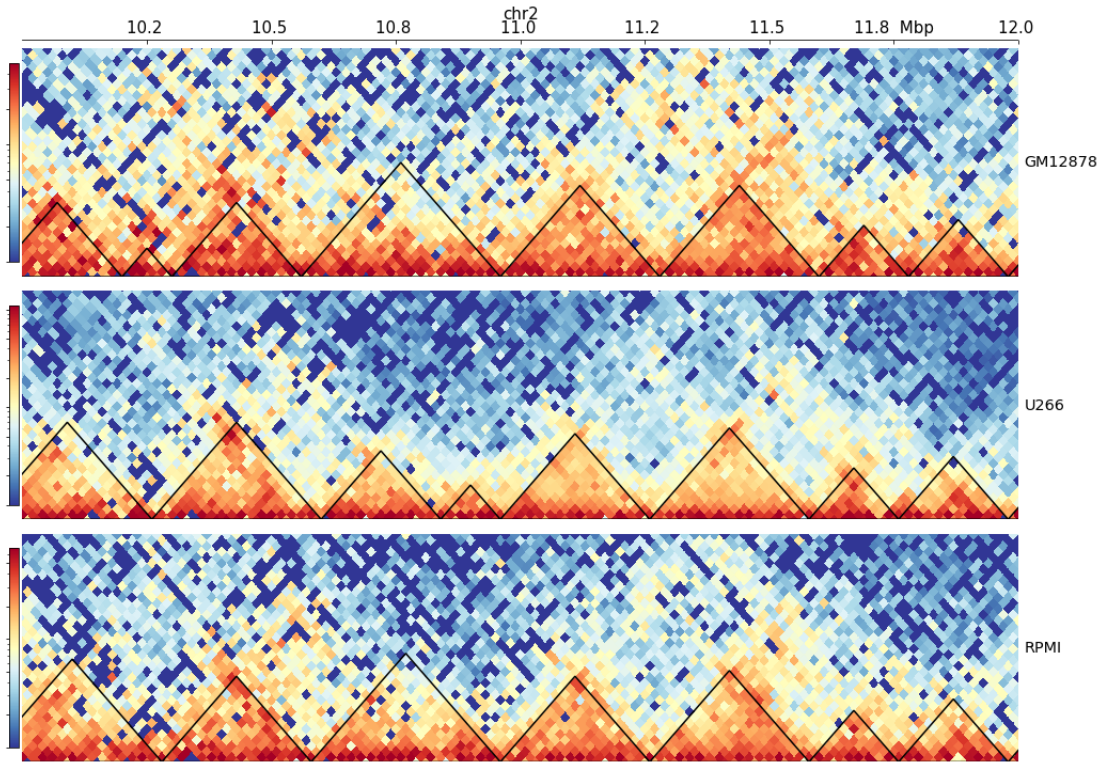


FIGURE 24 – Visualisation des cartes de contact de HiC sur le chr2 :10,000,000 :20,000,000 pour les lignées cellulaires : GM12878, U266 et RPMI.

Cette suite de logiciels prend en charge des données traitées par HiC-Pro à condition de convertir les matrices au format de fichier *.cool* avec l'outil *hicConvertFormat*. HiCExplorer utilise la bibliothèque *hic2cool* pour convertir les fichiers matriciels d'interaction *.hic* au format *.cool*. Le format de fichier *cool* est un format de stockage binaire, compressé et peu dense (Abdennur and Mirny, 2020). Puis les TADs ont été définis avec l'outil *hicFindTADs* qui calcule les régions TADs en deux étapes. Dans un premier temps il calcule un score de séparation (z score) pour toutes les fenêtres génomiques de 20kb, puis dans second temps, les minima locaux du score de séparation des TADs sont évalués par rapport aux autres fenêtres environnantes et une p-value leur est attribuée. L'outil génère plusieurs fichiers :

- Un fichier *domains.bed* contenant les coordonnées génomiques des TADs.
- Un fichier *score* contenant les scores de séparation des TADs.
- Un fichier *tad_score.bm* qui est une matrice contenant les coordonnées génomiques et les scores de séparation des TADs, et qui est utilisé pour visualiser les TADs par l'outil *hicPlotTADs*.

hicPlotTADs est un outil de visualisation pour tracer les TADs dans une région donnée, et peut permettre de comparer plusieurs régions TADs pour différents échantillons. Sur la figure 24, un exemple de régions (chr2 :10,000,000 :20,000,000) avec plusieurs TADs représentés sur

ces heatmaps par des triangles noirs pour les 3 lignées cellulaires étudiées. On remarque une assez grande conservation des TADs entre les différentes lignées. Des pistes supplémentaires peuvent être ajoutées à cette visualisation afin de permettre des comparaisons avec d'autres données, telles que du ChIP-seq, du RNA-seq, ou bien d'autres.

5.5.5 Annotation génomique

Nous avons utilisé les données disponibles sur le site UCSC pour le génome hg19, publié par le Genome Référence Consortium en février 2009 (Consortium et al., 2009), pour la définition des positions des sites d'initiation de la transcription (TSS) des gènes et la définition des îlot CpGs. En ce qui concerne l'étude des îlots CpGs, nous avons défini des régions dites "CpGs shores" bordant jusqu'à 2kb les alentours des îlots CpGs, et des régions dites "CpG shelves" bordant jusqu'à 2kb les alentours des CpG shores (Figure 25a). Nous avons défini les promoteurs comme étant des régions qui s'étendent de 2kb autour du TSS (Figure 25b), et nous appelons promoteur-CGIs, les promoteurs qui intersectent avec un CGI (Figure 25c).

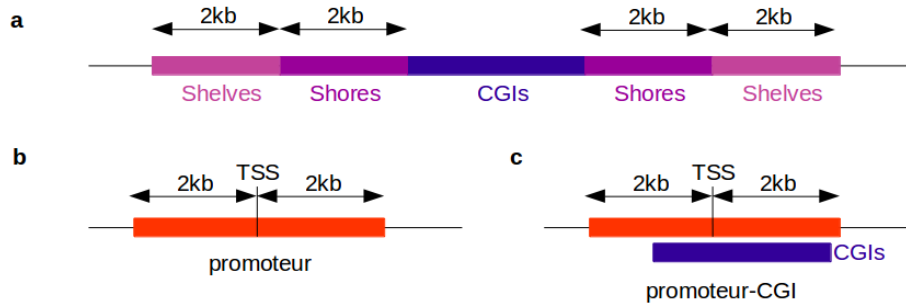


FIGURE 25 – Définition des régions promotrices et autour des CGIs. (a) Définition des régions shores et shelves autour des CGIs. (b) Définition de la région promotrice (c) Définition d'un promoteur-CGI.

Afin de déterminer la localisation génomique des loci et des eloci, nous avons utilisé l'outil *annotatePeaks.pl* de la suite de logiciel HOMER (Heinz et al., 2010) qui détermine le type d'annotation génomique occupée par le centre du locus. Nous avons utilisé leur annotation de base comprenant les TSS, les sites de terminaison de la transcription (TTS), les exons, les introns et les régions intergéniques.

Pour mesurer les intersections entre des régions, par exemple pour savoir si les loci sont dans des promoteurs-CGIs, nous avons utilisé l'outil *bedtools* et plus précisément son option *intersect* (Quinlan and Hall, 2010). Nous avons également utilisé de nombreuses autres options de cet outil, telles que *shuffle* afin de créer des régions aléatoires, ou encore *closest* pour mesurer et déterminer la région la plus proche, et l'option *slop* nous a permis de déterminer les régions shores et shelves des îlots CpGs en allongeant leur taille du nombre de bases souhaitées.

Nous avons également utilisé l'outil *liftover* du UCSC Genome Browser, nous permettant de convertir les coordonnées génomiques de certaines données récupérées dans d'autres études du génome de référence hg38 au génome de référence hg19.

5.5.6 Annotation fonctionnelle

Nos analyses nous ont conduits à la constitution de listes de gènes d'intérêt dont nous avons cherché à caractériser le rôle fonctionnel. Les listes de gènes étant parfois constituées de plusieurs milliers de gènes, une recherche de leur fonction gène par gène n'était pas envisageable, nous avons donc fait une analyse d'enrichissement fonctionnel permettant d'attribuer à des groupes de gènes des fonctions biologiques qui leur sont communes. Pour cela nous avons utilisé l'interface web de *DAVID* (<https://david-d.ncifcrf.gov>, version 6.7) qui permet de travailler avec différentes bases de données, notamment REACTOME, KEGG, PANTHER et de travailler sur les ontologies de gènes (GO terms) (Huang et al., 2009). Sa statistique d'enrichissement est calculée par un test de Fisher basé sur la distribution hypergéométrique pour tester l'hypothèse nulle selon laquelle l'enrichissement constaté d'une annotation est purement aléatoire. Pour le seuil de significativité, nous avons considéré comme fortement enrichies les catégories d'annotation avec une valeur de p value de Fisher inférieure à 0.1 (seuil par défaut de *DAVID*). Nous avons gardé également le seuil par défaut du nombre minimum de gènes appartenant à un terme d'annotation, c'est à dire 2, ce qui exclut les termes ayant un seul gène impliqué (Sherman et al., 2009).

Nous avons également utilisé l'outil *GREAT* (version 3.0.0, <http://great.stanford.edu/>) du laboratoire Bejerano de Stanford qui permet aussi de réaliser une analyse d'enrichissement (McLean et al., 2010) mais cette fois-ci non pas à partir de liste de gènes, mais à partir d'un fichier contenant des positions génomiques de régions d'intérêt (fichier .bed). Cet outil permet d'attribuer une signification biologique à des ensembles de régions génomiques non codantes en analysant les annotations des gènes qui lui sont voisins. Tout d'abord, chaque gène se voit attribuer un domaine régulateur, et ensuite chaque région génomique d'intérêt est associée à tous les gènes dont elle couvre le domaine régulateur. Pour la définition du domaine régulateur des gènes, nous avons utilisé les paramètres par défaut. Le domaine basal de régulation a une distance minimale de 5kb en amont du TSS et 1kb en aval, et est étendu dans les deux directions jusqu'au domaine basal du gène le plus proche, sans excéder l'extension maximale qui est de 1Mb dans une direction. *GREAT* calcule deux mesures d'enrichissement, l'une utilisant un test binomial sur les régions génomiques d'intérêt et une autre utilisant un test hypergéométrique basé sur les gènes. Ces deux tests se complètent, le test binomial corrige le biais du test hypergéométrique en tenant compte de la taille des domaines régulateurs des gènes, et le test hypergéométrique corrige les biais du test binomial en comptant chaque gène une seule fois. Les annotations fonctionnelles sont considérées comme étant significativement enrichies si le taux de faux positifs (valeur du FDR) est inférieur à 0.01.

6 Résultats

6.1 Description des données

6.1.1 Les données eRRBS

Notre cohorte de données eRRBS est composée de 45 échantillons, prélevés à différents stades du MM (SMM, diagnostic et rechute) pour 21 patients inclus dans les essais cliniques de l'intergroupe français du MM (IFM) et pour lesquels un consentement était disponible. Cette étude inclut 11 femmes et 10 hommes, ayant un âge médian de 60 ans.

Le nombre moyen de reads séquencés par patient est de 55216285 ; le taux d'alignement moyen des reads est de 60,81%, avec une moyenne de 2712252 CpGs séquencés par patient à une couverture de 10X et une moyenne de 1538510 CpGs par patient à une couverture de 60X (Figure 26 et Table 1).

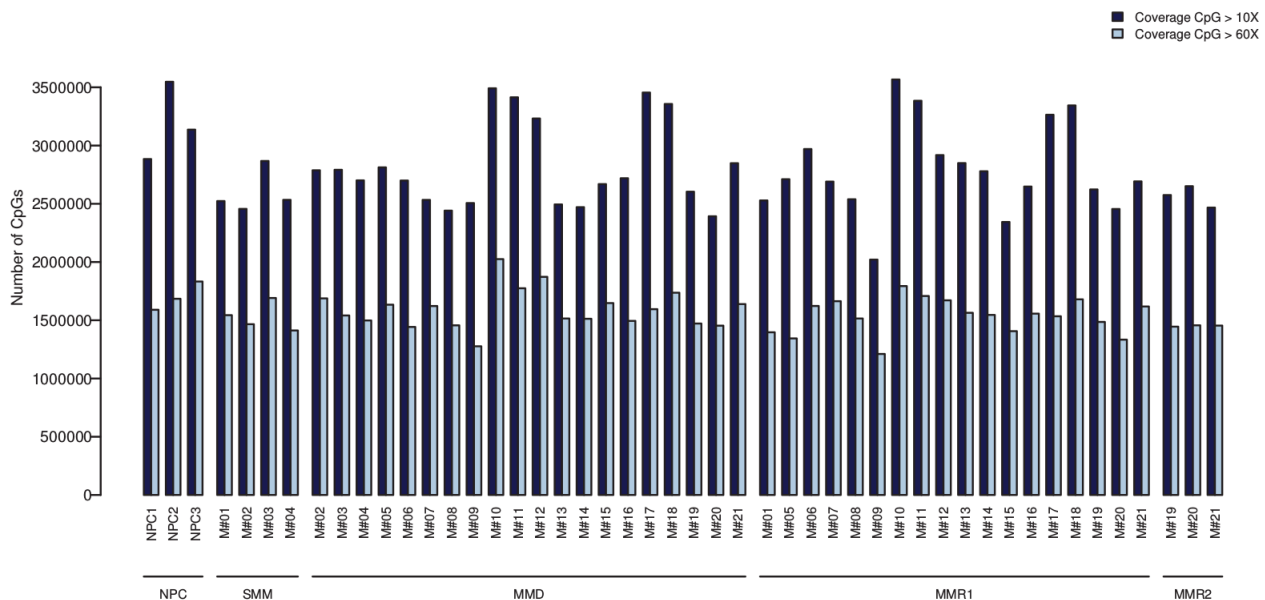


FIGURE 26 – Nombre de CpGs séquencés pour une couverture supérieure à 10x et supérieure à 60x par échantillon eRRBS..

En moyenne, 32% des CpGs séquencés avec une profondeur de 10X en eRRBS se situent dans un CGI contre 36% des CpGs séquencés avec une profondeur de 60X (Table 2).

TABLE 1 – Description des données eRRBS

ID patient	reads before trimming	reads after trimming	Uniquely aligned read count	Mapping efficiency	number of CpG >10X	number of CpG >60X
NPC1	55567561	54027218	34663936	64,2 %	2884414	1590129
NPC2	49150116	48209441	30711256	63,7 %	3547422	1684430
NPC3	51116901	50362568	32886482	65,3 %	3136733	1832548
M#01_SMM	61428280	59207771	35683810	60,3 %	2523053	1543912
M#02_SMM	62504131	57985061	33884808	58,4 %	2456102	1466201
M#03_SMM	79049347	74400738	44635678	60 %	2867303	1691093
M#04_SMM	55666213	51991090	30807429	59,3 %	2533662	1412362
M#02_MMD	60222298	57408485	35502670	61,8 %	2787716	1687971
M#03_MMD	61453332	57349448	34221080	59,7 %	2791587	1541208
M#04_MMD	61804658	57604774	33850985	58,8 %	2700718	1498271
M#05_MMD	54629384	51885205	33076074	63,7 %	2812572	1633624
M#06_MMD	46681613	43397131	27605068	63,6 %	2699457	1442788
M#07_MMD	69469974	65468086	42120566	64,3 %	2533579	1622291
M#08_MMD	54779690	51624695	31338473	60,7 %	2441333	1456238
M#09_MMD	41337053	39151584	23348924	59,6 %	2506353	1276636
M#10_MMD	65235719	64210095	41602689	64,8 %	3491223	2024878
M#11_MMD	57417003	56070614	37123603	66,2 %	3414290	1774615
M#12_MMD	61568981	60638420	37912403	62,5 %	3232145	1872579
M#13_MMD	57745574	54768857	32962736	60,2 %	2494458	1515263
M#14_MMD	58739195	54754077	34582565	63,2 %	2471051	1512614
M#15_MMD	66892182	63816252	39587129	62 %	2668928	1647606
M#16_MMD	58496231	54521562	34270563	62,9 %	2719319	1494593
M#17_MMD	57606174	55598676	35527994	63,9 %	3454934	1595117
M#18_MMD	50666404	52470367	33321457	66,5 %	3356743	1736359
M#19_MMD	57118881	53508937	33090485	61,8 %	2603705	1471932
M#20_MMD	61573436	56817558	34625464	60,9 %	2392815	1453640
M#21_MMD	60212170	57101372	35877159	62,8 %	2848308	1638977
M#01_MMR1	52739274	49708346	29824514	59,9 %	2528667	1396918
M#05_MMR1	50177115	46500488	28649068	61,6 %	2711140	1344075
M#06_MMR1	53912442	51174246	33066619	64,6 %	2969700	1622632
M#07_MMR1	59033111	56016048	35634947	63,6 %	2690526	1663667
M#08_MMR1	58193387	55234338	33738536	61,1 %	2538953	1515398
M#09_MMR1	51550876	48716249	28981223	59,5 %	2021000	1210669
M#10_MMR1	57273549	56007161	36579637	65,3 %	3567294	1793641
M#11_MMR1	52939643	51879778	34858886	67,2 %	3384119	1708211
M#12_MMR1	59998081	58026302	36772293	63,4 %	2918094	1671235
M#13_MMR1	68925567	63526054	37248398	58,6 %	2849316	1563961
M#14_MMR1	61150842	57482751	35359012	61,5 %	2779211	1545898
M#15_MMR1	56237707	52979111	32166568	60,7 %	2343772	1406428
M#16_MMR1	60726394	57950251	36874091	63,6 %	2648115	1557057
M#17_MMR1	53734759	52470367	33250786	63,4 %	3264431	1534358
M#18_MMR1	51615146	50791299	33049147	65,1 %	3344265	1679669
M#19_MMR1	53839425	51066242	31474438	61,6 %	2623120	1485630
M#20_MMR1	51773512	47988803	27841483	58 %	2455181	1333596
M#21_MMR1	67197181	63763682	39928897	62,6 %	2692334	1618269
M#19_MMR2	57089292	53820830	33016254	61,3 %	2575082	1445190
M#20_MMR2	57214046	53248395	32037542	60,2 %	2650854	1456472
M#21_MMR2	58242515	55317315	35332347	63,9 %	2467476	1454306

TABLE 2 – Répartition génomique des CpGs séquencés en eRRBS

Genomic feature	NPC samples				MM diagnosis samples			
	10X		60X		10X		60X	
	Mean	Standard dev	Mean	Standard dev	Mean	Standard dev	Mean	Standard dev
CpG island	1,685434.00	116179.07	1064617.67	70321.87	1554465.12	127211.94	990139.12	81569.69
CpG shores	325652.67	44758.19	146179.00	12743.51	280287.12	50914.35	140120.29	21169.81
CpG shelves	112869.33	15411.05	46133.00	4687.59	95902.00	19193.92	44915.06	7,691.76
3'UTR	32901.33	3223.80	15991.33	1450.87	28837.82	4291.79	15426.59	1964.95
5'UTR	119799.33	7462.21	81664.67	4564.02	111975.53	7771.64	75485.53	5284.99
exon	402853.00	27705.84	228018.00	23546.72	358430.47	38070.83	213453.59	26642.16
intergenic	803420.33	102291.63	363954.00	27996.67	693104.00	127340.43	345887.24	49789.38
intron	1008993.00	120620.80	483153.00	35097.96	884785.12	146022.29	459475.71	57743.30
non-coding	46739.00	4326.29	26285.67	2415.76	41552.47	5010.40	24221.00	3021.85
promoter-TSS	714015.33	63273.34	450953.33	27160.03	658834.71	60672.78	418531.18	32241.87
TTS	60801.67	6363.25	31704.67	2523.74	54315.94	7621.16	30608.65	3678.23

6.1.2 Les données WGBS

Notre cohorte de données WGBS est composée de 50 échantillons, prélevés au diagnostic et à la rechute pour 24 patients également inclus dans les essais cliniques de l'Intergroupe Français du MM (IFM) et pour lesquels un consentement était disponible. Cinq patients et onze échantillons (5 diagnostics, 5 première rechute, une seconde rechute) sont en communs avec notre étude eRRBS.

Nous avons effectué une comparaison des données eRRBS et WGBS pour les 5 échantillons diagnostics en commun en terme de nombre de CpG couverts par ces deux techniques (Table 3). Par le biais de la technique de séquençage eRRBS nous couvrons 23% des CpGs du génome (le génome comporte 28 162 972 CpGs) ; nos échantillons WGBS quant à eux présentent en moyenne une couverture de 96,3% des CpGs du génome. Si on compare également les profondeurs de séquençage pour nos deux techniques, nous avons pu séquencer bien plus profondément en eRRBS qu'en WGBS. Pour nos analyses, afin d'estimer au mieux les niveaux de méthylation des CpGs, nous avons choisi de garder uniquement les CpGs ayant un minimum de couverture de 10X. Ce qui donne pour la technique eRRBS une couverture de 9,9% des CpGs du génome et pour la technique WGBS, une couverture de 63,9% du génome.

TABLE 3 – Comparaison du nombre de CpGs couverts par les techniques de séquençage eRRBS et WGBS.

	eRRBS	WGBS
CpG sequenced	6 478 880	27 114 932
CpG coverage >10X	2 799 845	18 001 525
CpG coverage >60X	1 565 107	254 929
Mean coverage	51.47	17.20

Globalement, nous obtenons une moyenne de couverture des CpGs autour de 15X en WGBS pour nos échantillons (moyenne de 15,43X au diagnostic, 16,16X à la rechute 1, et 15,9X à la rechute 2), les données de moelles normales (NPC4 et NPC5) récupérées d'une autre étude (Kulis et al., 2015) sont quant à elles séquencées bien plus profondément que nos données, avec une couverture par CpG à 38X environ (Figure 27).

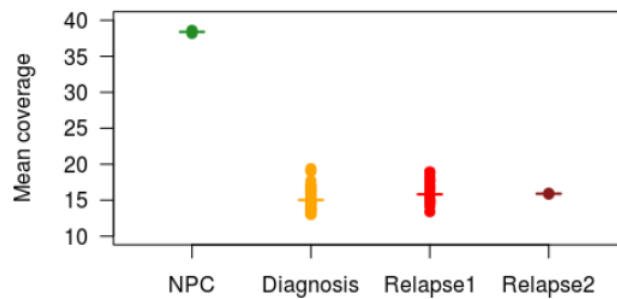


FIGURE 27 – Couverture moyenne par échantillon WGBS

6.2 Perturbation globale de la méthylation de l'ADN dans le MM

Afin d'obtenir une vision globale de la méthylation de l'ADN des patients atteints de MM, nous avons étudié les 5 échantillons WGBS en commun avec notre étude eRRBS et deux échantillons contrôles de cellules plasmocytaires normales (échantillons récupérés de l'étude de Kulis et al. (2015), NPC4 et NPC5). Nous avons observé une déméthylation globale du génome accompagnée d'augmentation localisée de la méthylation dans certaines régions (Figure 28a et 28d), comme précédemment rapporté dans le MM et plusieurs autres cancers (Agirre et al., 2015; Berman et al., 2012; Landau et al., 2014). Cette hypométhylation globale du génome observée en WGBS l'est également en eRRBS, et ce, à tous les stades de la maladie comparés aux cellules normales (Figure 29). Les CpGs présentent un profil de méthylation plutôt bimodal chez les patients atteints de MM (Figure 28b), les CpGs sont soit totalement méthylés soit totalement déméthylés, leur profil est bien distinct du profil des NPCs, qui lui présente un pic moins élevé de CpG à 0% de méthylation, et une augmentation de la densité de CpGs entre 80 et 95% de méthylation.

Pris individuellement, les CpGs sont globalement moins méthylés dans les échantillons de MM que dans les NPCs (Figure 28c). Mais plus la densité de CpG augmente, et plus la tendance s'inverse, les profils de méthylation de l'ADN des MM sont plus méthylés dans les régions avec une forte densité en CpG que les NPCs. Les niveaux de méthylation de l'ADN sont fonction de la densité locale de CpG.

Bien que l'on observe une hypométhylation globale de l'ADN, celle-ci est accompagnée d'une légère augmentation de la méthylation localisée au niveau des CGIs, bornée par une diminution de la méthylation dans les régions adjacentes des CGIs par rapport aux NPCs (Figure 28d). Ce phénomène est illustré sur la figure 28e, sur laquelle on peut observer le promoteur du gène *DOC2B* qui est composé d'un CGI. Les échantillons NPCs ont tendance à être plus méthylés autour du CGI et très peu à l'intérieur, alors que pour les échantillons de MM la tendance est inverse, c'est-à-dire moins méthylés autour et plus méthylés dans le CGI. Irizarry et al. (2009a) ont mis en avant l'importance des modifications de la méthylation de l'ADN dans les régions shores autour des CGIs. Ils ont montré que la plupart des altérations de la méthylation dans le cancer du colon ne se produisaient pas dans les CGIs, mais dans les régions shores qui bordent les CGIs, et que la méthylation dans ces régions shores était fortement liée à l'expression des gènes (Irizarry et al., 2009a).

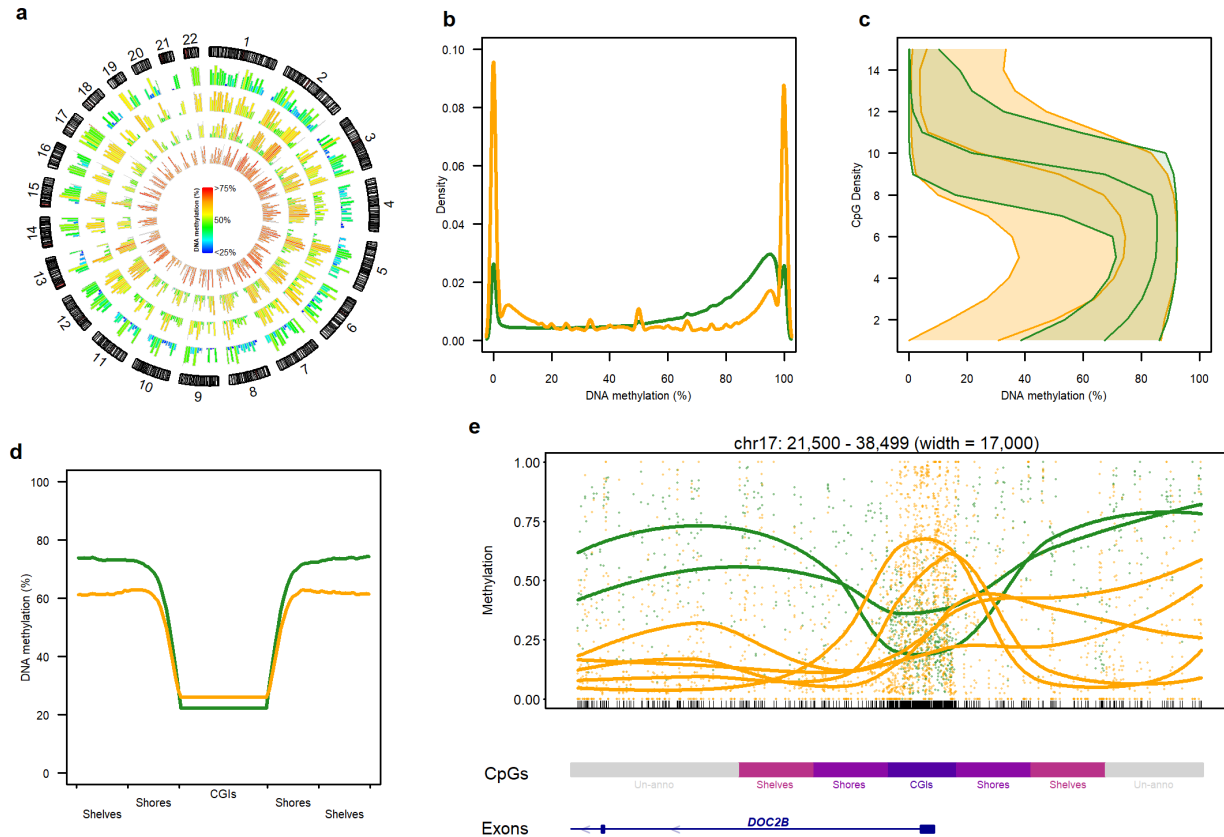


FIGURE 28 – Hypométhylation globale et hyperméthylation localisée. (a) Tracé circulaire des niveaux de méthylation de l'ADN mesurés en WGBS. Du centre vers l'extérieur, les échantillons : NPC4, M#17, M#19, et M#10. Les histogrammes représentent les niveaux de méthylation des CpG moyennés dans des fenêtres génomiques de 10 Mpb. (b) Distribution de la densité des niveaux de méthylation de l'ADN mesurés en WGBS pour les échantillons NPCs (en vert) et de MM au diagnostic (en orange). (c) Distribution des niveaux de méthylation de l'ADN dans des unités génomiques de 100 pb en fonction de la densité de CpG. (d) Méthylation moyenne des CGIs et leurs régions alentours pour les patients atteints de MM (diagnostic en orange) par rapport aux NPCs (en vert). (e) Méthylation de l'ADN entourant le promoteur du gène *DOC2B* dans les échantillons de MM (ligne orange) et dans les NPCs (ligne verte) assignés par dmrseq (Korthauer et al., 2019).

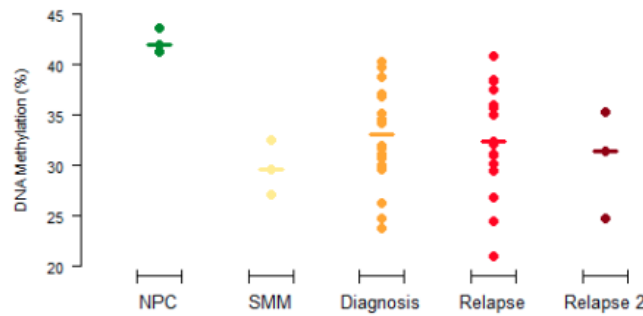


FIGURE 29 – Hypométhylation globale du méthylome des patients atteints de MM dès l’initiation de la maladie. Méthylation moyenne des CpGs aux différents stades du MM à partir des données eRRBS.

Afin de pouvoir regrouper les échantillons présentant un profil de méthylation de l’ADN similaire, nous avons réalisé une classification ascendante hiérarchique non supervisée, basée sur la méthode de Ward, qui consiste à réunir les deux clusters dont le regroupement fera le moins baisser l’inertie interclasse. Cette technique tend à créer des petits groupes, très différents les uns des autres. Cette méthode de classification a permis de montrer que des échantillons provenant de même patient à différents stades de la maladie (SMM, Diagnostic, Rechute 1 et 2) étaient toujours regroupés ensemble (Figure 30). Chez la plupart des patients, les changements de méthylation de l’ADN sont moins marqués entre le stade SMM et diagnostic ou entre la première et seconde rechute comparé aux changements s’effectuant entre le diagnostic et sa rechute, ce qui suggère une reconfiguration plus prononcée du méthylome de l’ADN après un traitement de première ligne. Cette classification hiérarchique des données de méthylation de l’ADN a permis d’établir une distinction claire entre les échantillons de NPCs et de MM. De nombreuses autres études ont également mis en avant que les échantillons d’un même patient étaient généralement plus similaires les uns aux autres que ceux provenant d’individus différents (Brocks et al., 2014; Pan et al., 2015).

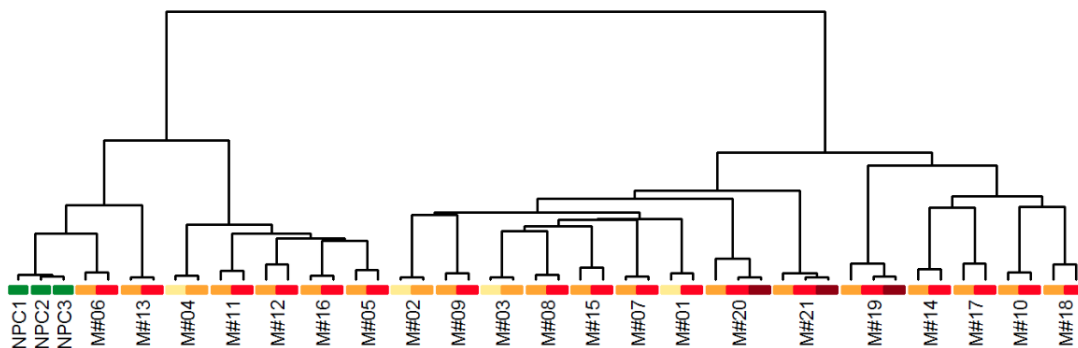


FIGURE 30 – Classification ascendante hiérarchique des profils de méthylation des échantillons eRRBS.

Ces résultats montrent que l'épigénome du MM est caractérisé par une perturbation globale de la méthylation de l'ADN et des états de méthylation intermédiaires variables en fonction de la densité locale en CpG. Nous avons ensuite cherché à savoir si ces changements épigénétiques pouvaient être une source d'hétérogénéité intra-patient.

6.3 L'hétérogénéité intra-tumorale de la méthylation dans le MM

6.3.1 Méthylation de l'ADN plus désordonnée dans le MM

Afin d'évaluer le degré de variabilité épigénétique intra-tumorale, nous avons utilisé l'outil methclone, permettant de détecter au niveau du génome des loci de 4 CpGs adjacents séquencés à une profondeur minimale de 60X (Figure 19). Nous avons détecté en moyenne 1 006 725 loci par échantillon, présentant une taille moyenne de 24 pb, localisés principalement dans les CGIs (Figure 31a), les régions promotrices et dans le corps des gènes (Figure 31b).

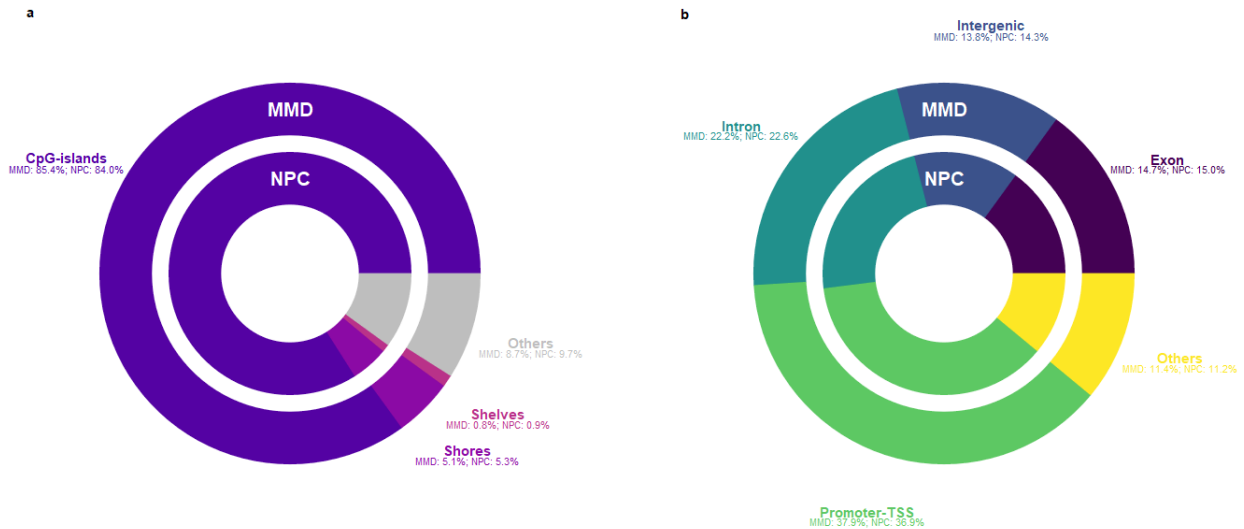


FIGURE 31 – Distribution génomique des loci des 21 MM au diagnostic et des 3 NPCs dans les CGIs et leurs alentours (a) et dans les promoteurs-TSS, les introns, les exons et les régions intergéniques (b).

Pour chacun des loci de 4 CpGs adjacents, nous avons ensuite calculé leur degré d'hétérogénéité de la méthylation de l'ADN (épipolymorphisme), défini comme une distribution de probabilité des 16 patterns de méthylation possibles à partir de 4 CpGs (épiallèle, Figure 19) (Landan et al., 2012). Une valeur élevée d'épipolymorphisme, proche de 1, indique une forte hétérogénéité de la méthylation, une méthylation désordonnée de l'ADN. Alors qu'un niveau faible d'épipolymorphisme, proche de 0, correspond à une méthylation de l'ADN ordonnée, homogène. Nous avons analysé la distribution de cet épipolymorphisme en fonction du niveau de méthylation des loci des échantillons NPCs et diagnostics (Figure 32).

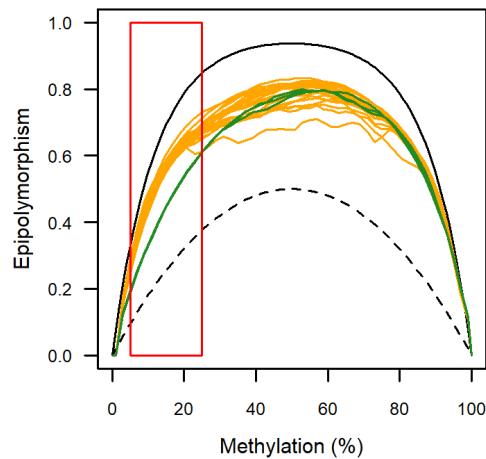


FIGURE 32 – Niveaux d'épipolymorphisme en fonction de la méthylation moyenne de l'ADN pour chaque locus dans les échantillons NPCs (en vert) et de MM au diagnostic (en orange).

Sur la figure 32 est représenté l'épipolymorphisme maximal (ligne noire continue) correspondant au niveau d'hétérogénéité maximum pour chaque valeur de méthylation, et l'épipolymorphisme bimodal (trait noir en pointillé) qui suppose une combinaison d'épiallèles entièrement méthylés et entièrement non méthylés pour chaque niveau de méthylation. Nous avons représenté sur cette figure pour les échantillons NPC et MM, leur valeur médiane de l'épipolymorphisme par niveau de méthylation de l'ADN pour refléter l'hétérogénéité épigénétique du génome des plasmocytes. Nous avons ainsi observé que pour des niveaux de méthylation intermédiaires, la méthylation de l'ADN est hétérogène à la fois dans les échantillons NPCs et de MM au diagnostic. Cependant, une augmentation significative de l'épipolymorphisme est observée pour des niveaux faibles de méthylation, entre 5 et 25% de méthylation dans les échantillons de MM par rapport aux NPCs (encadré rouge sur la figure 32). Les loci étant localisés à environ 80% dans des CGIs, régions caractérisées comme étant peu ou pas méthylées, cette différence de niveau d'épipolymorphisme entre les NPCs et MM, entre 5% et 25% de méthylation, suggère une légère augmentation de la méthylation des CGIs d'une manière désordonnée dans les MM.

Ensuite nous avons étudié plus particulièrement ce phénomène au niveau des promoteurs-CGIs. Dans les promoteurs-CGIs, bien que la majorité des loci suivent également une distribution "en cloche" de cette relation méthylation de l'ADN et épipolymorphisme, un petit nombre de gènes y dérogent (Figure 33a). Certains gènes changent de trajectoire au diagnostic, et vont vers une augmentation de la méthylation de l'ADN dans leur promoteur moins désordonnée qu'attendu pour un tel niveau de méthylation. Parmi eux, nous avons trouvé des gènes suppresseurs de tumeurs tels que *BCL11B*, *SALL4*, *EGFR*, *ETVG* et *BRCA1* comme illustré chez le patient M#05 (Figure 33b) et le patient M#12 (Figure 33c). L'hyperméthylation des promoteurs des gènes suppresseurs est un phénomène récurrent dans les cancers, ce qui conduit à l'inactivation de ces gènes qui sont des régulateurs négatifs de la prolifération cellulaire.

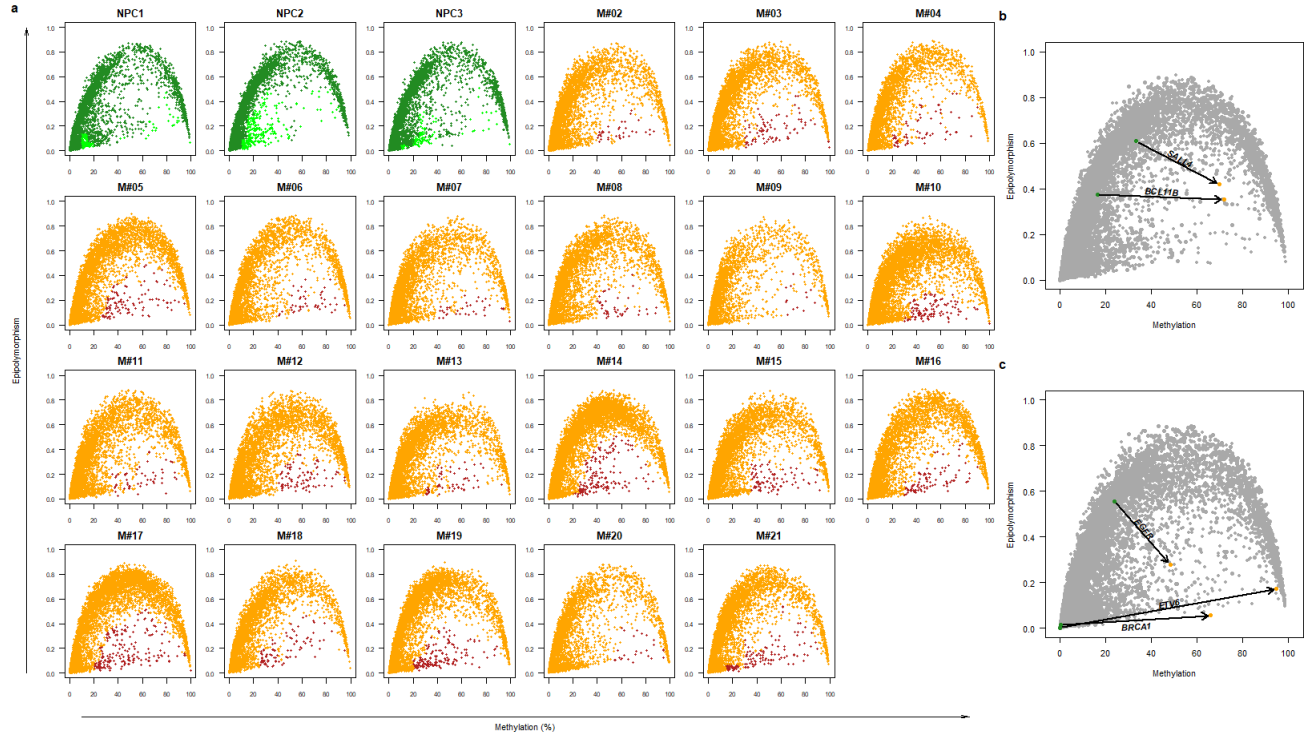


FIGURE 33 – Niveaux d'épipolymorphisme en fonction de la méthylation moyenne de l'ADN au niveau des promoteurs-CGIs, (a) pour l'ensemble des échantillons NPCs et MM au diagnostic, et un focus sur les patients M#05 (b) et M#12 (c). Les flèches sur les figures b et c décrivent le changement de niveau d'épipolymorphisme et de méthylation des gènes suppresseurs de tumeurs, et représentent leur évolution entre l'échantillon NPC2 (point vert) et le MM (point orange).

Chez la majorité des patients atteints de MM, nous avons observé une augmentation de la méthylation de l'ADN dans les CGIs et les régions promotrices couplée à une augmentation de l'épimorphisme (Figure 34a), ainsi qu'une diminution de la méthylation dans les régions bordant les CGIs (shores et shelves) couplée également à une diminution de l'hétérogénéité de la méthylation (Figure 34a partie droite). En revanche, pour tous les patients, l'épimorphisme moyen dans les promoteur-CGIs est fortement corrélé ($r = 0,94$) à une augmentation de la moyenne de la méthylation de l'ADN (Figure 34b).

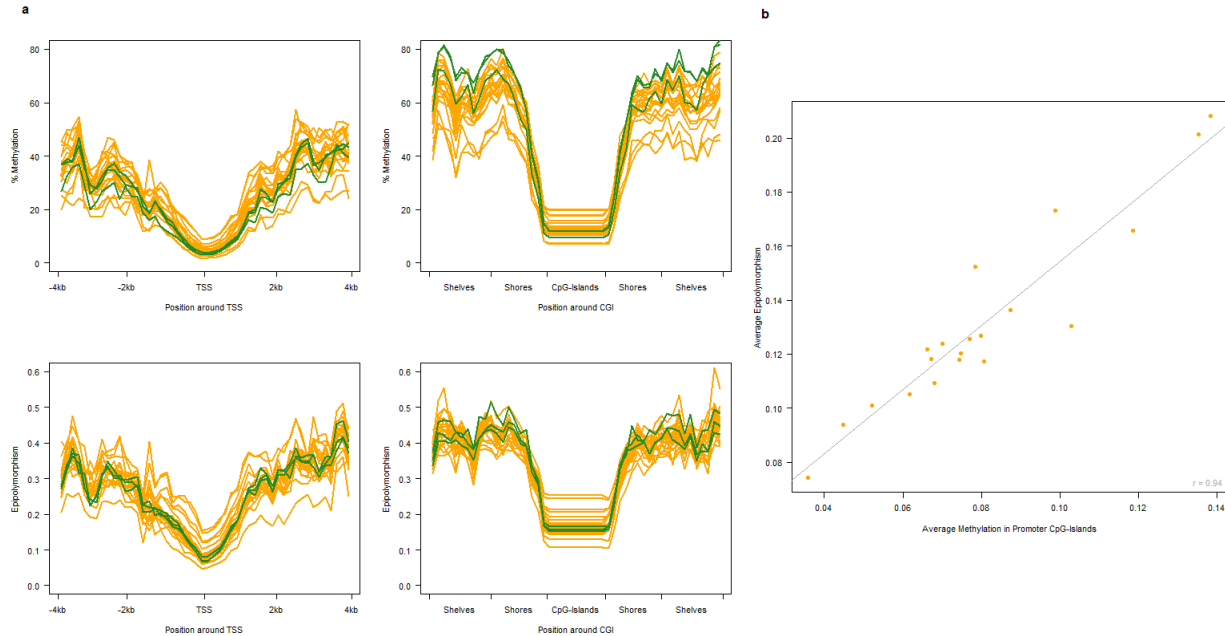
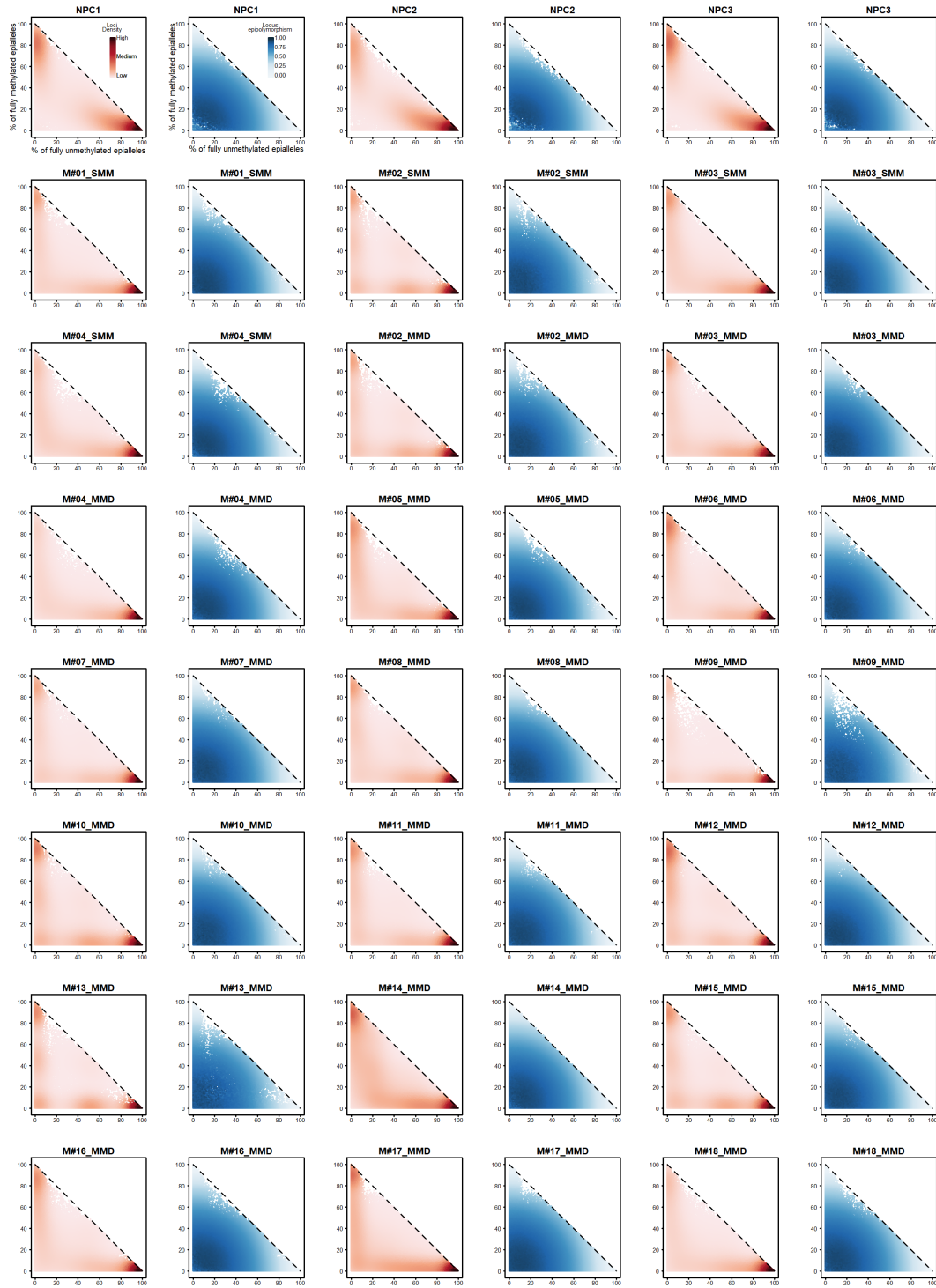


FIGURE 34 – Méthylation de l'ADN et épimorphisme dans les promoteurs-CGIs. (a) Méthylation de l'ADN (à gauche) et épimorphisme (à droite) dans les promoteurs (en haut) et dans les CGIs (en bas) et leurs alentours. (b) Corrélation entre la moyenne de méthylation de l'ADN et l'épimorphisme dans les promoteurs-CGIs par échantillon au diagnostic du MM.

Afin de caractériser davantage les profils épialléliques de chaque échantillon, nous avons séparé les loci en fonction de leur fréquence en épiallèles complètement méthylés en ordonnées ou complètement non méthylés en abscisses (Figure 35). Les loci des NPCs présentent une distribution quasi bimodale avec un grand nombre de loci entièrement méthylés ou complètement non méthylés (Figure 36 en haut à gauche), la densité forme quasiment une droite près de la ligne en pointillés. Alors que dans les échantillons de myélome, la densité suit plutôt les axes (Figure 36 en bas à gauche), ce qui indique que les loci sont moins entièrement méthylés ou entièrement non méthylés, mais plus hétérogènes. Ils n'ont plus une distribution bimodale et présentent un enrichissement en loci présentant une méthylation plus désordonnée (Figure 36 partie droite) par rapport aux cellules normales.



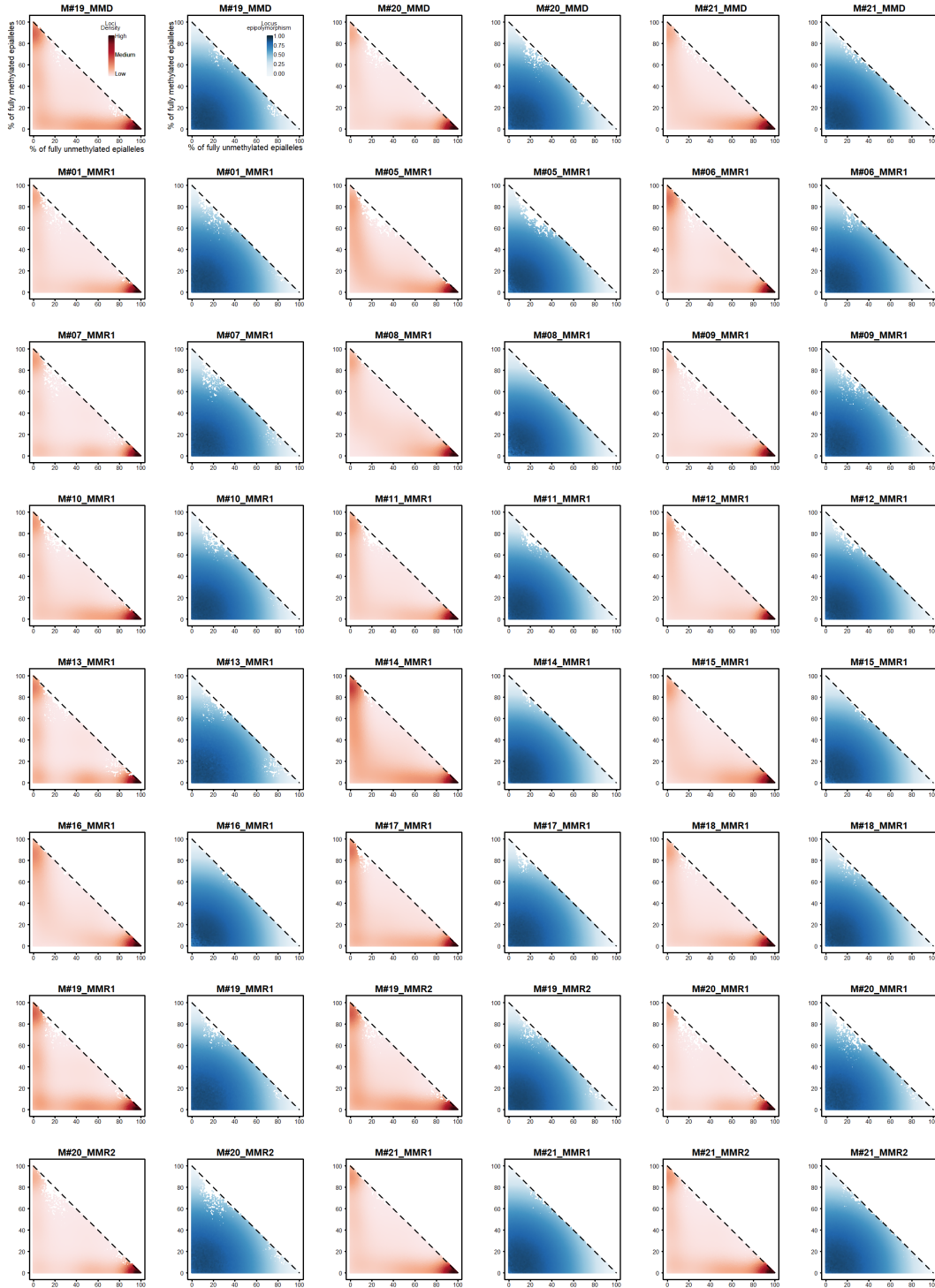


FIGURE 35 – Nuages de points montrant l’organisation des loci en fonction de leur fréquence en épiallèles complètement méthylés ou déméthylés pour tous les échantillons. Chaque point correspond à un locus de 4 CpGs adjacents. Sur chaque paire de figures par échantillon, à gauche, une représentation où chaque point est codé en couleur en fonction de la densité des points environnants, à droite, chaque point est codé en couleur selon son niveau d’épipolymorphisme.

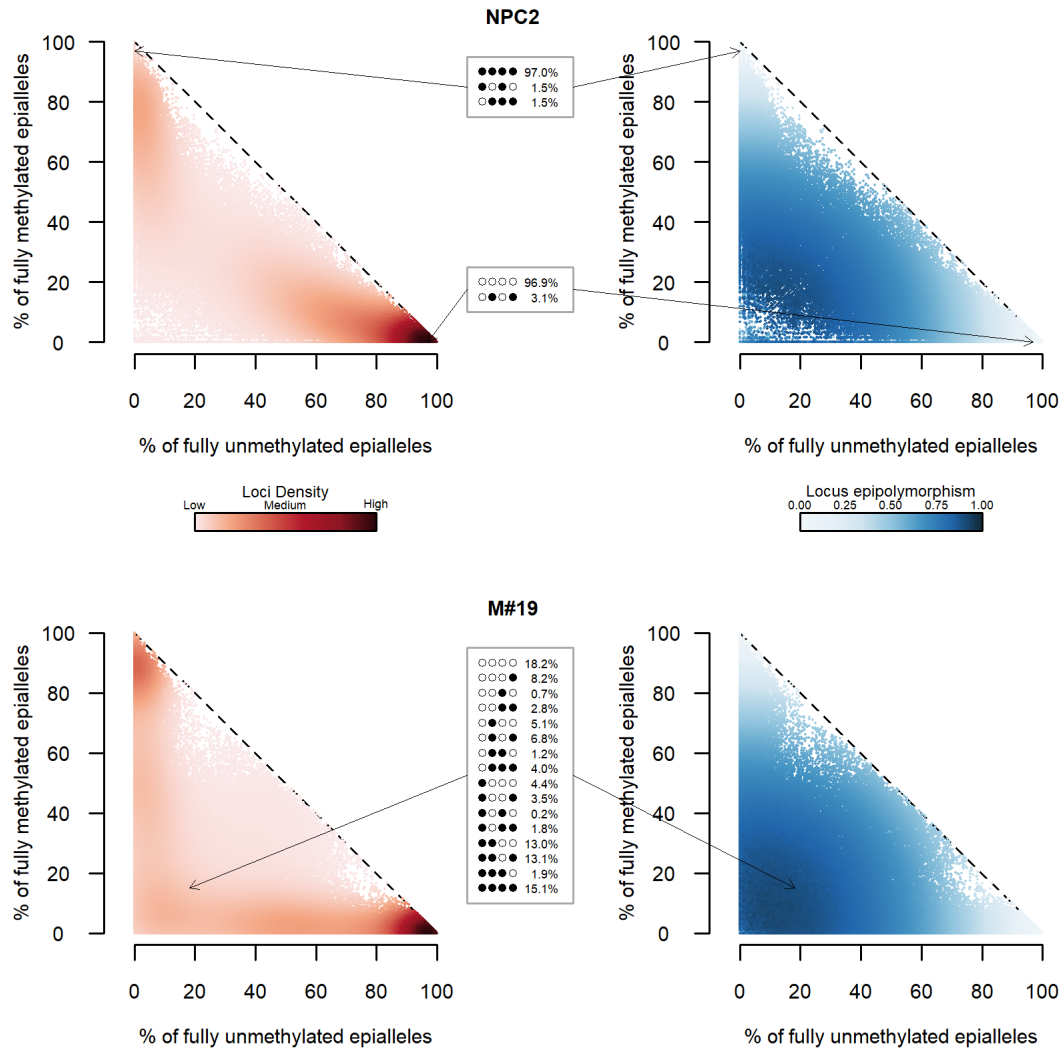


FIGURE 36 – Détails pour un échantillon NPC (NPC2) et un MM au diagnostic (M#19) de l'organisation des loci en fonction de leur fréquence en épiallèles complètement méthylés ou déméthylés. Chaque point correspond à un locus de 4 CpGs adjacents. À gauche, une représentation où chaque point est codé par couleur en fonction de la densité des points environnants, à droite, chaque point est codé en couleur selon son niveau d'épipolymorphisme. Sur chacune des 4 figures, les points en bas à droite (principalement non méthylés) et en haut à gauche (principalement méthylés) représentent les loci avec les épipolymorphismes les plus faibles. Les schémas des épiallèles des loci complètement méthylés, complètement non méthylés et complètement désordonnés sont illustrés au centre de la figure dans les encadrés.

Nous avons utilisé une méthode complémentaire à l'épipolymorphisme pour mesurer l'hétérogénéité de la méthylation : le PDR (le pourcentage de reads discordants) (Landau et al., 2014), qui confirme l'augmentation de l'hétérogénéité dans les cellules de MM par rapport aux NPCs (Figure 37). Un PDR élevé signifie que les loci présentent une grande proportion de reads contenant à la fois des CpGs non méthylés et méthylés (méthylation désordonnée). Nous avons constaté que le PDR moyen des échantillons de MM était plus élevé que le PDR des échantillons NPCs. Cette augmentation de l'hétérogénéité de la méthylation de l'ADN dans le cas du MM apparaît dès le stade précurseur SMM, et persiste également à la rechute. Les cellules de myélomes présentent une méthylation plus hétérogène, plus aléatoire, que les cellules normales au détriment d'états de méthylation plus cohérents et ce, dès le stade précurseur du myélome.

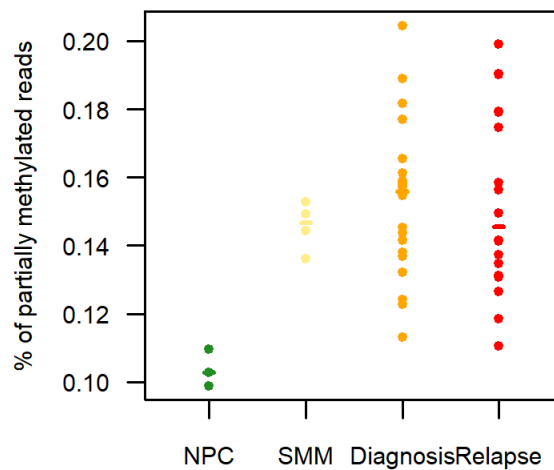


FIGURE 37 – Pourcentage de reads partiellement méthylés par échantillon, NPC ($n = 3$), SMM ($n=4$), Diagnostic ($n=17$), et rechute ($n=17$).

Contrairement à d'autres hémopathies malignes, telles que le lymphome diffus à grandes cellules B (DLBCL, Pan et al. (2015)) dans laquelle l'hétérogénéité épigénétique diminue du diagnostic à la rechute, dans le cas du MM, la variabilité de la méthylation intra-échantillon fluctue d'un patient à l'autre, avec pour une moitié des patients une diminution de l'hétérogénéité à la rechute alors que l'autre moitié présente une augmentation de l'hétérogénéité à la rechute (Figure 38).

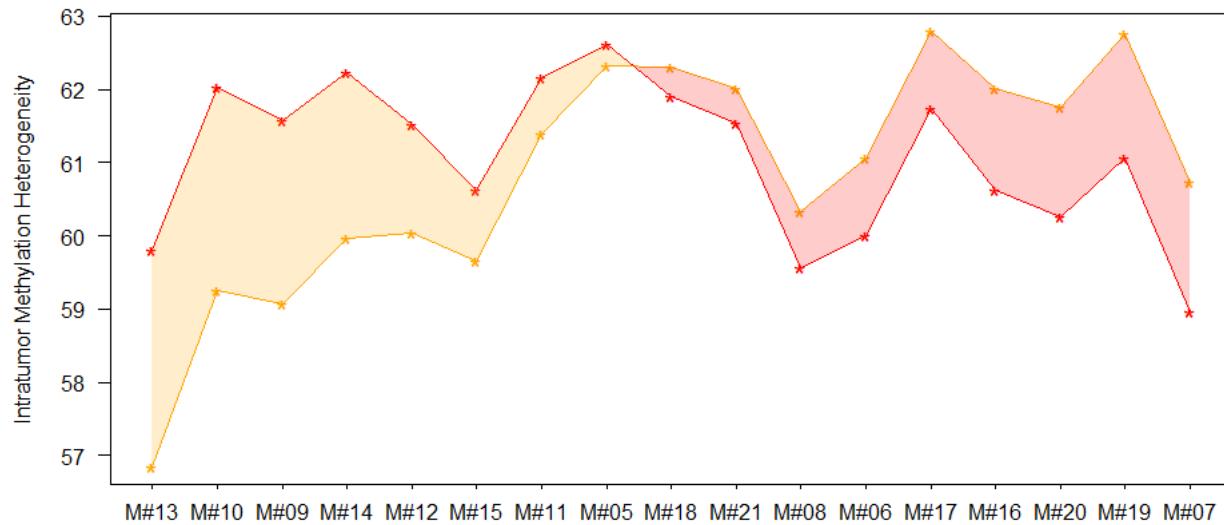


FIGURE 38 – Hétérogénéité de la méthylation intra-tumorale au diagnostic (en orange) et à la rechute (en rouge) par patient. Deux groupes de patients distincts, à gauche, les patients pour lesquels l’hétérogénéité de la méthylation intra-tumorale est plus importante lors de la rechute qu’au diagnostic, et à droite les patients pour lesquels l’hétérogénéité de la méthylation intra-tumorale est plus importante au diagnostic qu’à la rechute.

Les niveaux de variabilité de la méthylation de l’ADN des patients ne sont pas pronostiques, à la différence de ce que l’on observe dans les DLBCL (Pan et al., 2015).

Dans l’ensemble, ces résultats ont révélé qu’au cours de l’initiation et de la progression du MM, les plasmocytes malins semblent accumuler une méthylation de l’ADN plus désordonnée, aléatoire, au détriment d’un état de méthylation plus cohérent présent chez les plasmocytes sains, conduisant à un degré élevé d’hétérogénéité épigénétique intratumorale chez les patients atteints de MM.

6.4 Les changements significatifs de patterns de méthylation

Ayant mis en évidence dans la partie précédente que la perturbation de la méthylation de l’ADN est couplée à une augmentation de l’hétérogénéité épigénétique intra-tumorale dans le cas du MM, nous avons ensuite cherché à analyser cette dynamique épiallélique au cours du développement et de la progression de la maladie. Pour cela, toujours avec l’outil methclone, à partir des loci de 4 CpGs adjacents, nous avons identifié ceux présentant un changement de composition épiallélique significatif entre deux états (normal versus cancer ou diagnostic du MM versus sa rechute), qu’on appelle par la suite eloci (Figure 20).

6.4.1 Impact clinique des changements épialléliques

Afin de pouvoir comparer les changements globaux du méthylome de tous les échantillons de patients, nous avons normalisé le nombre d'eloci détectés par échantillon en fonction du nombre total de loci séquencés, pour obtenir une variable nommée EPM (Li et al., 2014) (Figure 19).

N'ayant pas d'échantillons sains appariés aux échantillons de MM, nous avons mesuré l'EPM entre les échantillons normaux (eloci NPC vs NPC) afin de quantifier l'importance de la variabilité inter-individuelle dans notre méthode d'analyse (Figure 39a, partie gauche, échantillons en vert), qui s'avère être très faible comparée à l'EPM entre les échantillons normaux et cancéreux (eloci NPC vs Diagnostic, Figure 39a, partie droite, échantillons en orange). Pour les eloci NPC versus Diagnostic, nous avons considéré uniquement les loci qui étaient annotés eloci (NPC vs Diagnostic) pour les 3 NPCs, afin de réduire au maximum dans notre analyse la variabilité inter-individuelle. Ces mesures permettent de constater que tous les patients atteints de MM présentent de nombreux changements épialléliques significatifs par rapport aux NPCs avec un degré très variable selon les patients.

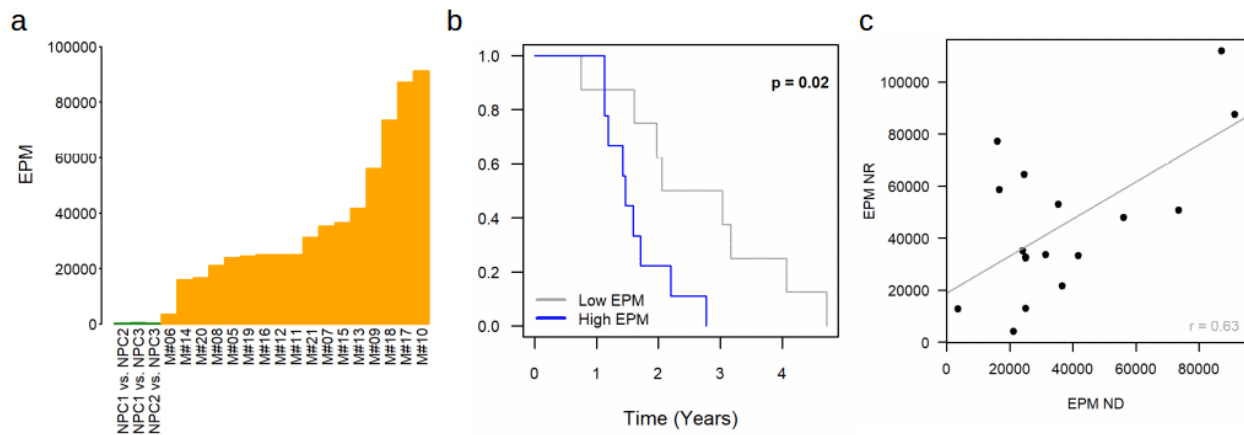


FIGURE 39 – Quantification des changements épialléliques significatifs et leur impact sur la survie des patients. (a) Valeurs d'EPM entre les échantillons NPC versus NPC et NPC versus diagnostic. (b) Association de l'EPM et du temps de rechute suite au traitement des patients. (c) Corrélation entre l'EPM au diagnostic (eloci NPC versus Diagnostic) et à la rechute (eloci NPC versus Rechute) par rapport aux NPCs.

Nous avons ensuite examiné le lien entre ces changements épialléliques et la survie des patients. Pour cela nous avons séparé les patients en deux groupes selon leur valeur médiane d'EPM au diagnostic et nous avons investigué si ces groupes mettaient le même temps ou non à rechuter suite à leur traitement (RFS : relapse-free survival ; Figure 39b). Nous avons constaté qu'un EPM élevé au diagnostic par rapport aux NPCs était associé à un temps de rechute plus court (test de log rank, $p = 0,02$). Ces changements épigénétiques sont indépendants de l'âge, du sexe et de la ploïdie des patients.

De plus, nous avons observé une corrélation entre l'EPM au diagnostic (eloci NPC versus diagnostic) et à la rechute (eloci NPC versus rechute) (coefficient de corrélation $r = 0,63$, Figure 39c). Ces résultats reflètent une perturbation épigénétique globale au diagnostic du MM par rapport aux NPCs, qui persiste lors de la rechute.

6.4.2 Perturbation des gènes impliqués dans le développement

Nous nous sommes ensuite intéressés à la localisation génomique de ces changements significatifs d'épiallèles au cours d'initiation et de la progression du MM.

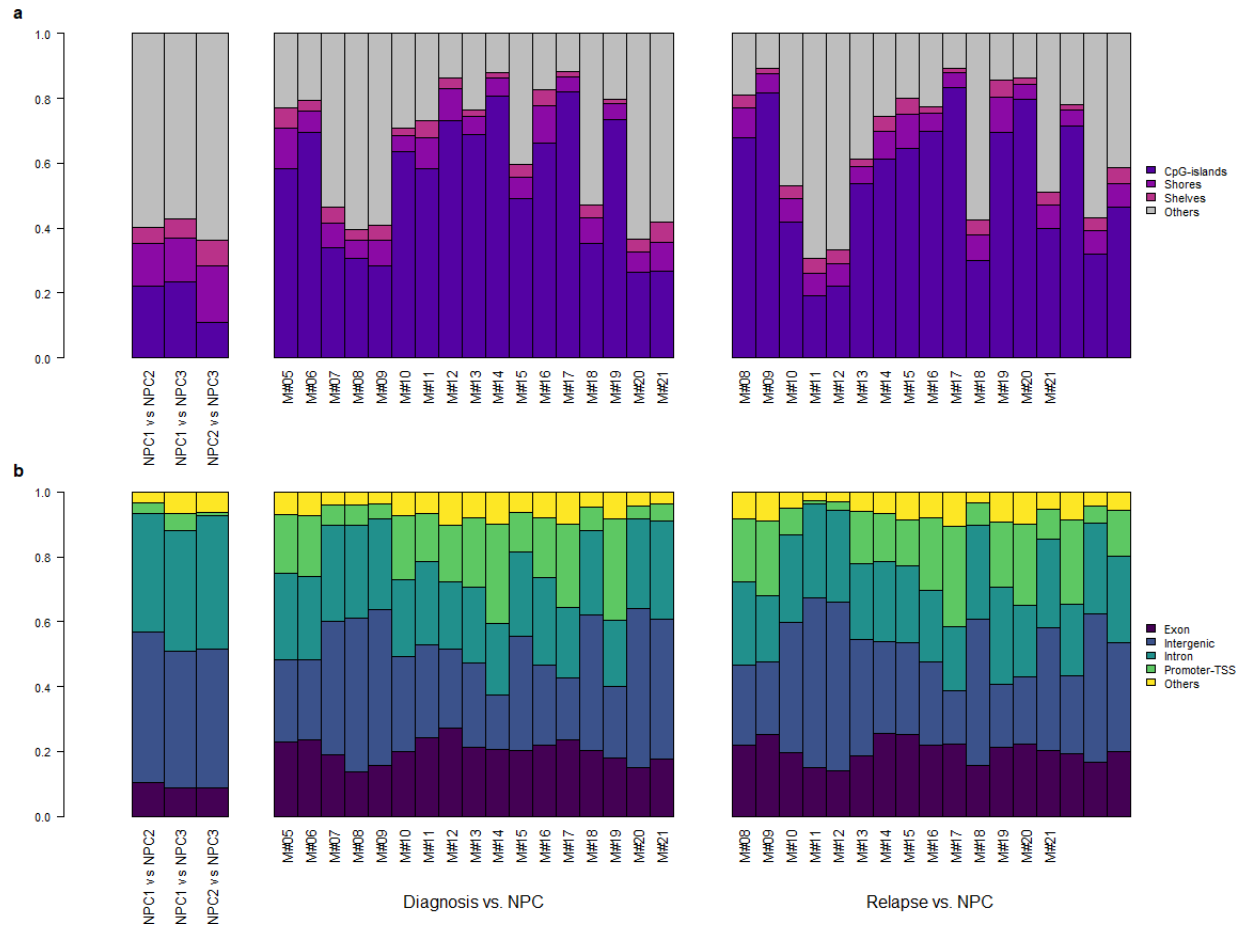


FIGURE 40 – Distribution génomique des eloci NPC versus NPC (à gauche), des eloci NPC versus diagnostic (au milieu) et des eloci NPC versus Rechute (à droite) au niveau des CGIs et leurs alentours (a), et au niveau des principales régions génomiques : promoteurs-TSS, exons, introns et régions intergéniques (b).

Bien que la distribution de ces eloci varie d'un patient à l'autre, la majorité des changements significatifs de méthylation de l'ADN sont localisés dans les CGIs (en moyenne 54% pour les MM (eloci NPC versus diagnostic) et 19% pour les NPCs (eloci NPC versus NPC), test de Wilcoxon $p=0.0018$, figure 40a), dans les promoteurs-TSS (en moyenne 15% pour les MM

et 3% pour les NPCs, test de wilcoxon $p=0.0123$, figure 40b). Une distribution similaire des eloci est également observée à la rechute comparée aux NPCs (Figure 40 partie de droite, test de Wilcoxon, $p=0,82$ pour les promoteurs-TSS et $p=0,75$ pour les CGI).

Afin de pouvoir étudier la fonction biologique de ces perturbations épigénétiques, nous avons effectué une analyse d'enrichissement en terme ontologique avec l'outil *GREAT*, des eloci (NPC versus Diagnostic) compris dans des promoteur-CGIs (Figure 41). Fait intéressant, un enrichissement marqué a été observé en gènes portant la marque répressive H3K27me3 dans leur promoteur, en facteurs de transcription, en régulateurs du développement et en gènes participant à la voie de signalisation Wnt. La voie Wnt joue un rôle dans la prolifération de la cellule tumorale et la formation osseuse, les cellules du MM favorisant le développement de la maladie osseuse qui lui est associée et la maladie osseuse favorisant la croissance tumorale, un cercle vicieux en résulte, qui a pour conséquence l'augmentation de la charge tumorale.

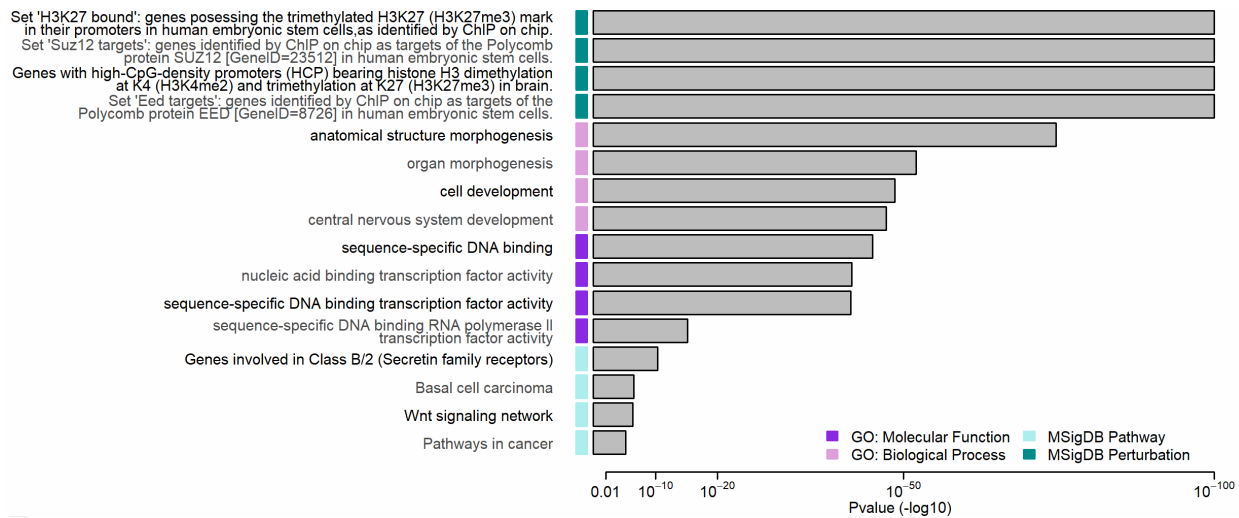


FIGURE 41 – Analyse d'enrichissement en terme ontologique avec l'outil *GREAT*, des eloci (NPC versus Diagnostic) compris dans des promoteur-CGIs.

Il est admis que les régions promotrices des facteurs de transcription et les gènes régulateurs du développement précoce ont tendance à résider dans de grands domaines génomiques qui sont généralement dépourvus de méthylation de l'ADN dans la plupart des lignées cellulaires, appelés vallées de méthylation de l'ADN (DMVs) (Xie et al., 2013), ou encore canyons de méthylation de l'ADN (Jeong et al., 2014). Ces DMVs sont des régions génomiques riches en CpGs qui ne sont pas méthylées dans les cellules normales mais qui acquièrent généralement une certaine méthylation en cas de cancer.

Nous avons donc cherché à déterminer si les changements épigénétiques marqués (eloci NPC vs Diagnostic) dans les promoteurs-CGIs étaient enrichis dans ces DMVs. Nous avons constaté que 60% des gènes associés à un DMV contenaient dans leur promoteur un elocus (Test exact de Fisher, $p=2,2e-16$, Figure 42a). L'examen visuel de la méthylation de l'ADN des DMVs contenant des gènes dont le promoteur présente des eloci (NPC vs Diagnostic) a révélé une hyperméthylation aberrante et variable chez tous les patients atteints de MM,

comparés aux NPCs, du gène *CEBPD* (CCAAT Enhancer Binding Protein Delta), qui code pour un gène candidat supresseur de tumeur, associé à la régulation de l'apoptose et de la prolifération cellulaire (Liu et al., 2019) (Figure 42b).

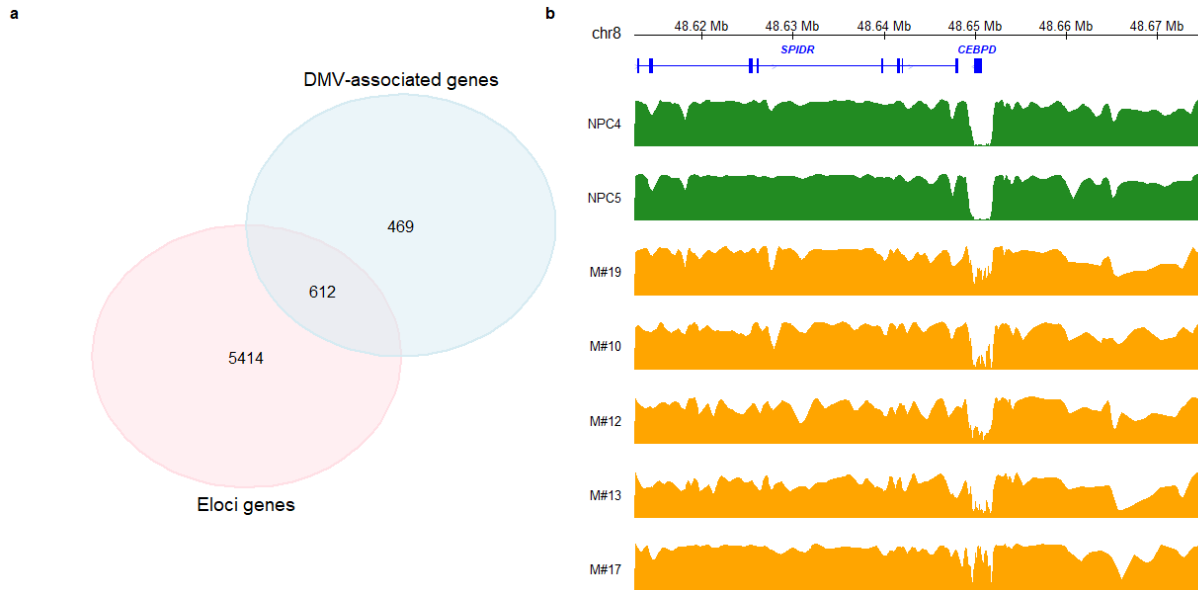


FIGURE 42 – Perturbation des vallées de méthylation de l'ADN. (a) 60% des gènes dans des DMVs contiennent des eloci NPC versus Diagnostic du MM. (DMV-associated genes = gènes présent dans un DMV ; Eloci genes = gènes contenant dans son promoteur un elocus NPC vs Diagnostic). (b) Exemple de perturbation de la méthylation dans un DMV - *CEBPD*.

Bien que la comparaison des NPCs entre eux ait révélé également la présence d'eloci (NPC versus NPC, figure 40 partie gauche) dans des promoteurs de gènes, ils ne se sont révélés enrichis en aucun terme ontologique spécifique.

N'ayant pu déterminer aucun elocus NPC versus diagnostic commun à tous les patients et les modifications de la méthylation de l'ADN semblant être très variables selon les patients (par exemple l'augmentation variable de la méthylation de l'ADN sur le gène *CEBPD*, Figure 42b), on peut penser qu'il est peu probable d'observer la perturbation d'un même gène chez tous les patients, mais qu'il est bien plus probable qu'une même voie de signalisation soit perturbée chez tous les patients. Une hypothèse cohérente suite à ces résultats serait que ce soit la dérégulation des voies de signalisation qui soit pathologiquement importante, telle que la voie de signalisation Wnt, plutôt que la dérégulation d'un gène spécifique (Morgan et al., 2012b).

6.4.3 Evolution des changements épialléliques au cours de la progression du MM

Afin de pouvoir étudier l'évolution des changements épialléliques au cours de l'initiation et de la progression du MM, ainsi que les effets du traitement, nous avons également mesuré les valeurs d'EPM des patients entre tous les stades de la maladie.

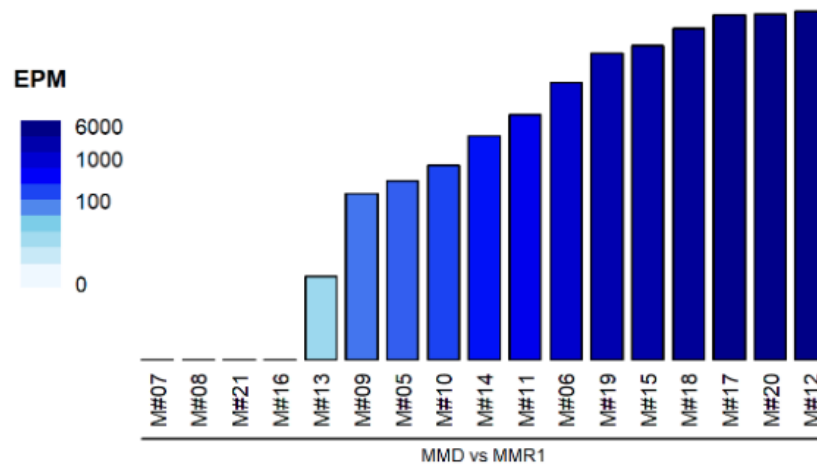


FIGURE 43 – Évolution de l'EPM au cours de la progression du MM.

Aucun changement significatif n'a pu être mis en évidence entre le stade SMM et le stade diagnostique du MM et entre les premières et secondes rechutes. En revanche, entre les échantillons diagnostiques et rechutes, on observe des valeurs d'EPM très variables selon les patients avec une médiane à 277, 24% des patients ne présentent aucun changement significatif ($EPM=0$) entre le diagnostic et sa rechute, tandis que 42% des patients présentent des changements épialléliques très importants ($EPM > 1000$) entre le diagnostic et sa rechute (Figure 43).

L'EPM permet d'évaluer et de mesurer globalement les changements épialléliques à un certain instant mais ne permet pas d'obtenir des informations sur la façon dont les eloci varient au cours de la progression du MM, à plusieurs temps/stades de la maladie. Pour répondre à cela, nous avons catégorisé les eloci de chaque patient (eloci diagnostique versus NPC et rechute vs NPC) en 3 classes :

- les eloci présents spécifiquement au diagnostic,
- les eloci présents spécifiquement à la rechute,
- les eloci présents aux deux stades de la maladie.

Puis nous avons réalisé un regroupement des patients selon leurs proportions de ces 3 classes d'eloci, en réalisant une classification ascendante hiérarchique en utilisant la méthode des kmeans et la gap statistique afin de définir le nombre de groupe de patients le plus juste possible. L'algorithme a été passé 1 000 fois afin de vérifier la stabilité des groupes.

Ainsi par cette méthode d'analyse, nous avons pu séparer les patients en 3 groupes (clusters, Figure 44a) :

- Le cluster 1 : les patients présentant une prédominance significative d'eloci spécifiques au diagnostic,
- Le cluster 2 : les patients ne présentant pas de prédominance d'eloci spécifiques au diagnostic ou à la rechute mais une prédominance d'eloci présent aux 2 stades de la maladie,
- Le cluster 3, les patients présentant une prédominance significative d'eloci spécifiques à la rechute.

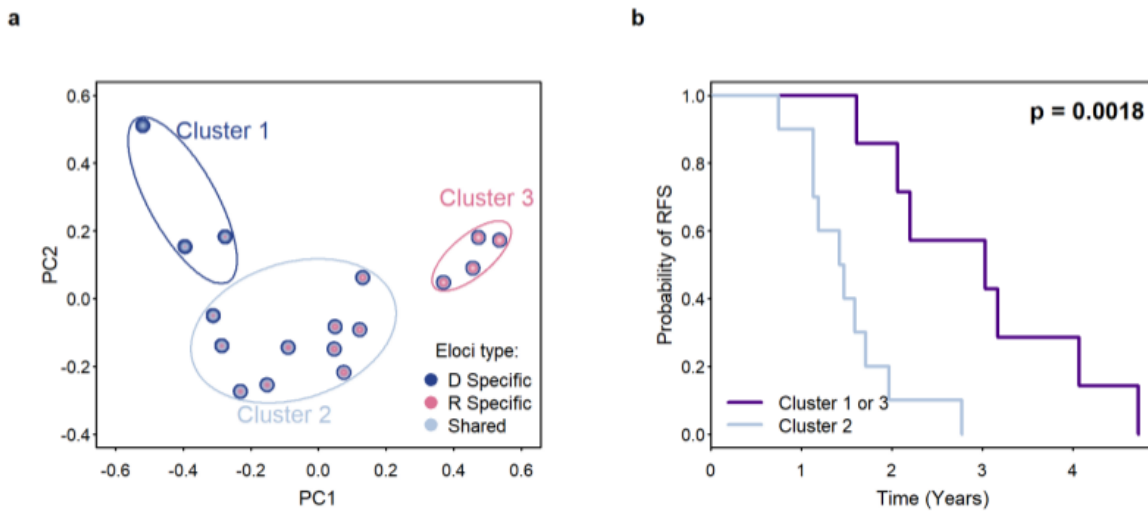


FIGURE 44 – Évolution des changements épialléliques au cours de la progression du MM et leur impact sur la survie des patients. (a) Regroupement des patients selon leurs proportions des 3 classes d'eloci : ceux présents spécifiquement au diagnostic (D Specific), ceux présents spécifiquement à la rechute (R specific) et ceux présents aux deux stades de la maladie (Shared). (b) Temps de rechute suite aux traitements des 3 groupes de patients.

Les patients présentant le plus d'eloci stabilisés durant l'évolution de la maladie (cluster 2) présentent un temps de rechute significativement plus court que les autres (Figure 44b), nous avons également mis en évidence une corrélation entre un fort EPM et ces patients. Les fortes perturbations qui interviennent au diagnostic et qui sont maintenues à la rechute, semblent avoir un impact sur la survie des patients.

Nous avons ensuite analysé plus précisément les changements de méthylation entre le diagnostic et la rechute afin d'étudier l'effet du traitement sur la dynamique des épiallèles (échantillon au diagnostic : avant traitement, et échantillon à la rechute : après traitement). Nous nous sommes concentrés principalement sur 4 schémas extrêmes d'évolution des patterns de méthylation des eloci (Figure 45).

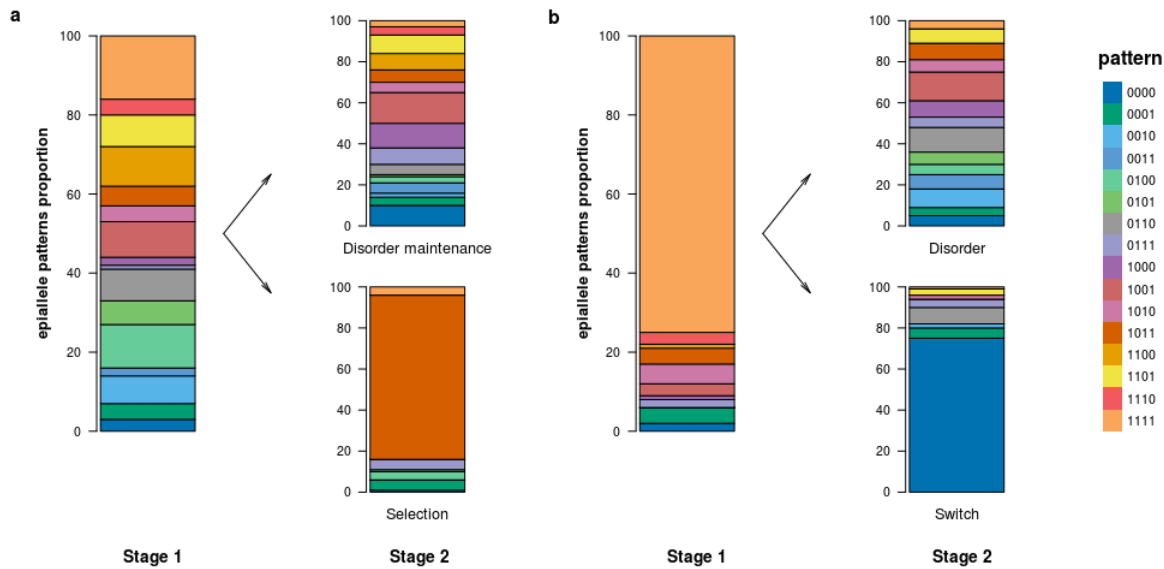


FIGURE 45 – Les 4 schémas extrêmes de modifications des patterns de méthylation des eloci. (a) Au stade 1 de la maladie, le locus ne présente pas de pattern de méthylation dominant, au stade 2, il peut maintenir un état de méthylation désordonnée mais différent ou alors un pattern de méthylation mineur au stade 1 peut être sélectionné et devenir majeur au stade 2. (b) Au stade 1 de la maladie, le locus présente un pattern de méthylation dominant, au stade 2, il peut perdre cette dominance et devenir hétérogène, ou maintenir cette dominance, mais échanger de pattern de méthylation dominant.

A partir d'un locus hétérogène au diagnostic (stade 1, Figure 45a), les évolutions possibles sont :

- La maintenance du désordre : aucun pattern majoritaire au diagnostic du MM et à la rechute (stade 2, Figure 45a, partie haute)
- La sélection : une distribution des épiallèles très hétérogène au diagnostic qui devient ordonnée à la rechute avec un pattern majoritaire présent à plus de 70% (stade 2, Figure 45a, partie basse)

A partir d'un locus présentant un pattern de méthylation majoritaire au diagnostic (stade 1, Figure 45b), les évolutions possibles sont :

- Le désordre : un pattern de méthylation majoritaire au diagnostic qui disparaît à la rechute, et qui laisse place à du désordre (stade 2, Figure 45b, partie haute).
- Le switch : un pattern de méthylation majoritaire au diagnostic qui est remplacé par un autre pattern de méthylation différent mais toujours majoritaire à la rechute (stade 2, Figure 45b, partie basse).

Nous avons ensuite mesuré la proportion de ces 4 modèles d'évolutions possibles des eloci entre le stade NPC et le diagnostic et entre le stade diagnostic et sa rechute. Nous avons observé un enrichissement significatif à la rechute de la sélection d'un pattern de méthylation par rapport au diagnostic (Figure 46, test de Wilcoxon $p = 0.0059$). Entre le stade sain et

le diagnostic du MM (pas de traitement entre ces deux stades), on observe principalement une augmentation de l'hétérogénéité de la méthylation, alors qu'à la rechute (influence du traitement), on observe parmi les eloci une sélection plus marquée d'un pattern de méthylation (phénomène qui reste toutefois très minoritaire sur l'ensemble des eloci), la méthylation est moins désordonnée. Ces résultats montrent une diminution de l'hétérogénéité épigénétique pour certains eloci spécifiques associée à une pression du traitement.

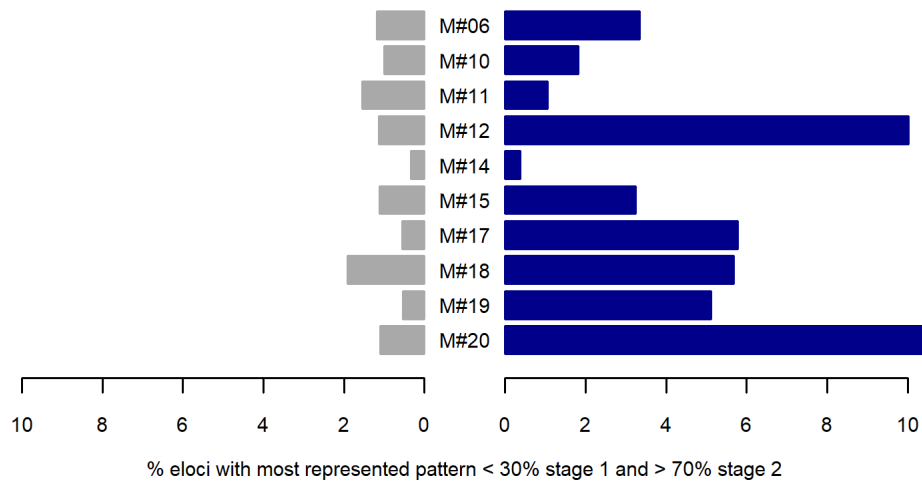


FIGURE 46 – Proportion des eloci (NPC versus diagnostic en gris et diagnostic versus rechute en bleu) présentant une sélection d'un pattern de méthylation.

Ces résultats suggèrent que le traitement peut réduire l'hétérogénéité intratumorale en sélectionnant des sous-clones épigénétiques avec des profils de méthylation plus homogènes.

6.5 Impact des changements significatifs de la méthylation sur l'organisation de la chromatine

Dans cette partie, on cherche à connaître les états de la chromatine associés aux changements d'entropie entre nos échantillons NPCs et MM. Afin d'identifier les caractéristiques de la chromatine liées aux changements significatifs de la méthylation, nous avons analysé la distribution des eloci (NPC versus diagnostic) selon les 25 états de la chromatine définis par l'outil chromHMM sur 127 épigénomes (Figure 47a, données disponibles sur le site UCSC) (Ernst and Kellis, 2015). Nous avons observé une répartition des eloci des patients très homogène sur l'ensemble des 127 épigénomes, et notamment un enrichissement dans les états de la chromatine associés aux régions promotrices actives, aux promoteurs bivalents, aux régions réprimées par les polycomb, et aux régions quiescentes (régions caractérisées par une absence d'enrichissement des 12 marques d'histones utilisées dans le modèle).

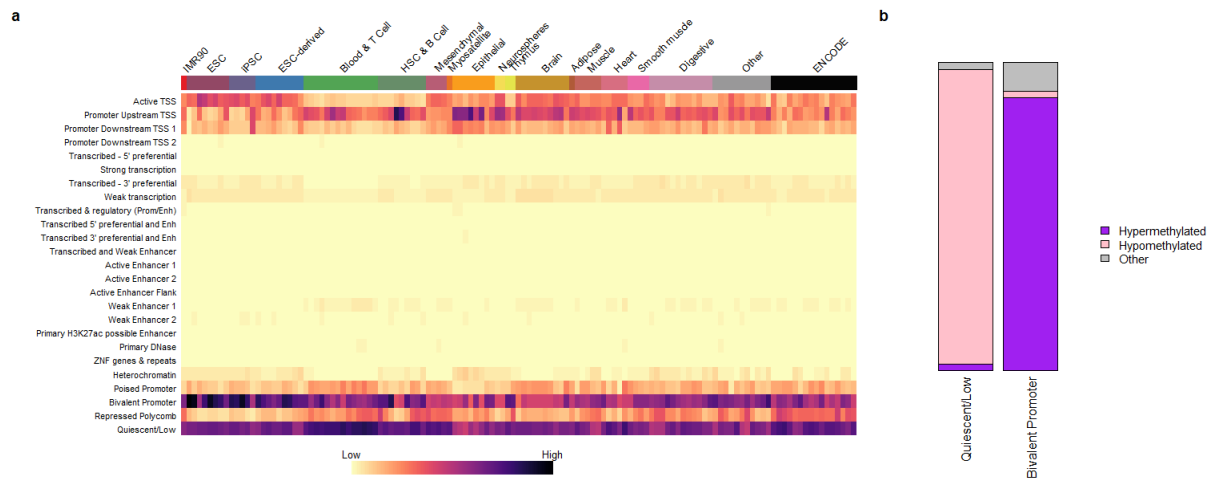


FIGURE 47 – Localisation des eloci selon les états de la chromatine. (a) Annotation des eloci (NPC versus diagnostic) selon les 25 états de la chromatine définis par chromHMM pour 127 épigénomes disponibles sur le site UCSC (Ernst and Kellis, 2015). (b) Proportion d'eloci hypo- et hyperméthylés dans les régions promotrices bivalentes et quiescentes définies par chromHMM.

Les changements épialléliques (eloci NPC versus diagnostic) allant vers une hyperméthylation de l'ADN sont prédominants au niveau des régions promotrices bivalentes, tandis que les eloci allant vers une hypométhylation sont très majoritairement trouvés dans des états de la chromatine dits quiescents (Figure 47b). Les fréquences de ces deux types d'eloci (hyper- et hypométhylés) sont très variables d'un patient à l'autre (Figure 48).

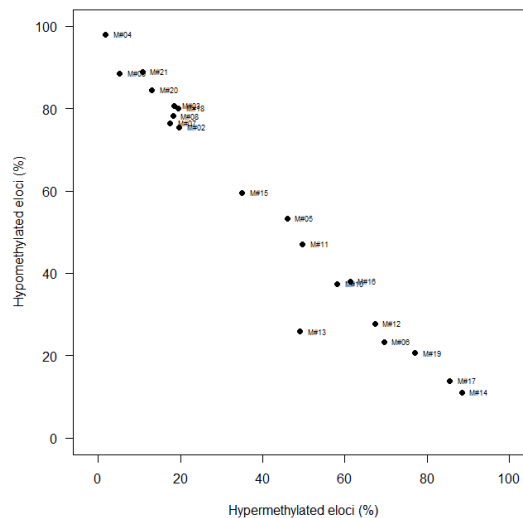


FIGURE 48 – Répartition des 20 échantillons de MM selon leur pourcentage d'eloci hypo- et hyperméthylés.

Pour la suite de notre étude, nous nous sommes intéressés principalement à ces deux types de régions, régions promotrices bivalentes et régions quiescentes, enrichies respectivement en eloci hyperméthylés et hypométhylés entre nos échantillons NPC et MM.

6.5.1 Les promoteurs bivalents

Les promoteurs bivalents sont des états de chromatine abondants dans les cellules souches et les cellules différenciées. La bivalence se produit sur les sites d'initiation de la transcription (TSS) riches en CpG non méthylés, et est caractérisée par la présence de deux modifications d'histones au même endroit, H3K27me3 réprimant la transcription en favorisant une structure de la chromatine compacte et H3K4me3 régulant positivement la transcription et inhibant la méthylation de l'ADN (Kinkley et al., 2016) (Figure 3c). Les gènes présentant un promoteur bivalent jouent un rôle dans la pluripotence des cellules souches embryonnaires (ES). Ce sont des gènes liés au développement embryonnaire précoce et à la spécification de la lignée cellulaire. Les régions bivalentes permettent de faire taire les gènes clés du développement dans les cellules souches embryonnaires tout en les maintenant prêts pour une activation ultérieure (Mas et al., 2018). De tels domaines bivalents sont importants pour maintenir la plasticité épigénomique nécessaire au cours du développement. La bivalence joue un rôle important dans le développement des tissus normaux, mais également permet aux cellules de maintenir leurs états différenciés.

La bivalence de la chromatine est souvent perdue dans les lignées de cellules cancéreuses. Les promoteurs bivalents sont globalement peu méthylés dans les cellules normales, et présentent un niveau de méthylation fréquemment plus élevé dans les cellules cancéreuses qui s'accompagne d'une perte de la bivalence (Bernhart et al., 2016). Cette perte de bivalence au niveau des promoteurs est accompagnée d'un gain de méthylation (lorsque les deux marques H3K4me3 et H3K27me3 sont perdues) ou d'une activation génique (lorsque la marque H3K27me3 est perdue) (Hahn et al., 2014).

Parmi les eloci présents dans ces régions prédites bivalentes par chromHMM (Figure 47a), nous avons constaté que 82,5% de ces eloci se retrouvaient dans des promoteurs bivalents des cellules ES. Nous avons observé un enrichissement significatif des promoteurs bivalents des cellules souches embryonnaires touchés par un elocus par rapport aux promoteurs bivalents des tissus différenciés (Figure 49).

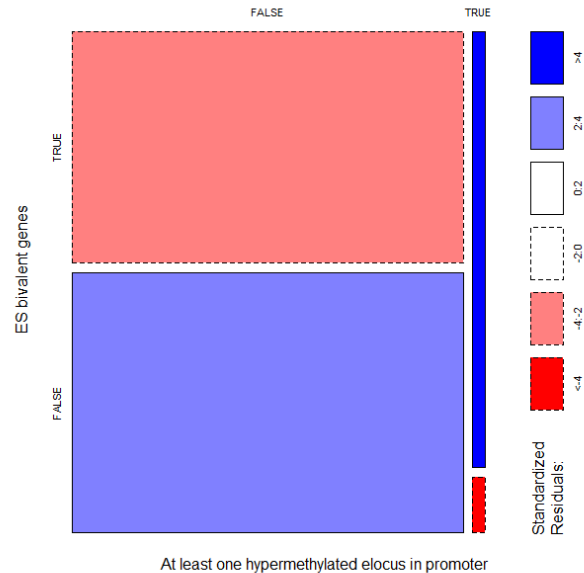


FIGURE 49 – Enrichissement significatif des promoteurs bivalents des cellules souches embryonnaires touchés par un elocus comparé aux promoteurs bivalents des tissus différenciés.

Ensuite, nous avons étudié l'évolution de la méthylation moyenne de l'ADN des eloci dans les promoteurs bivalents entre les échantillons NPC et MM au diagnostic et entre le diagnostic et la rechute de la maladie (Figure 50). Pour la majorité des patients, on observe une augmentation de la méthylation de l'ADN au cours de la progression du MM. Chez le patient M#14 présentant le plus d'eloci hyperméthylés (Figure 48), nous avons constaté que 93,5% des eloci présentaient un gain important de méthylation de l'ADN, jusqu'à 40% de plus à l'initiation de la maladie comparé aux NPCs. Ce gain de méthylation observé au diagnostic du MM se poursuit dans les échantillons de MM à la rechute (Figure 50).

Puis nous avons analysé les impacts de ces changements significatifs de méthylation de l'ADN dans ces régions promotrices bivalentes sur l'expression des gènes (Figure 51). La très grande majorité des gènes présente une corrélation négative entre l'expression et la méthylation de l'ADN dans leur promoteur bivalent au diagnostic du MM, même si un certain sous-ensemble de gènes présente eux une corrélation positive (augmentation de la méthylation de l'ADN dans le promoteur des gènes couplée à une augmentation de l'expression) au diagnostic du MM (Figure 51). Ce processus de diminution d'expression des gènes dont le promoteur bivalent présente un gain de méthylation augmente à la rechute du MM (Figure 51). Le sous-ensemble de gènes présentant une corrélation positive entre l'expression et la méthylation dans son promoteur au diagnostic tend à disparaître à la rechute du MM. Ce processus de répression génique semble être continu et se stabilise au cours de la progression du MM. Ce phénomène d'augmentation de la méthylation de l'ADN dans des promoteurs bivalents couplée à une répression génique a déjà été décrit dans de nombreuses autres tumeurs (Easwaran et al., 2012).

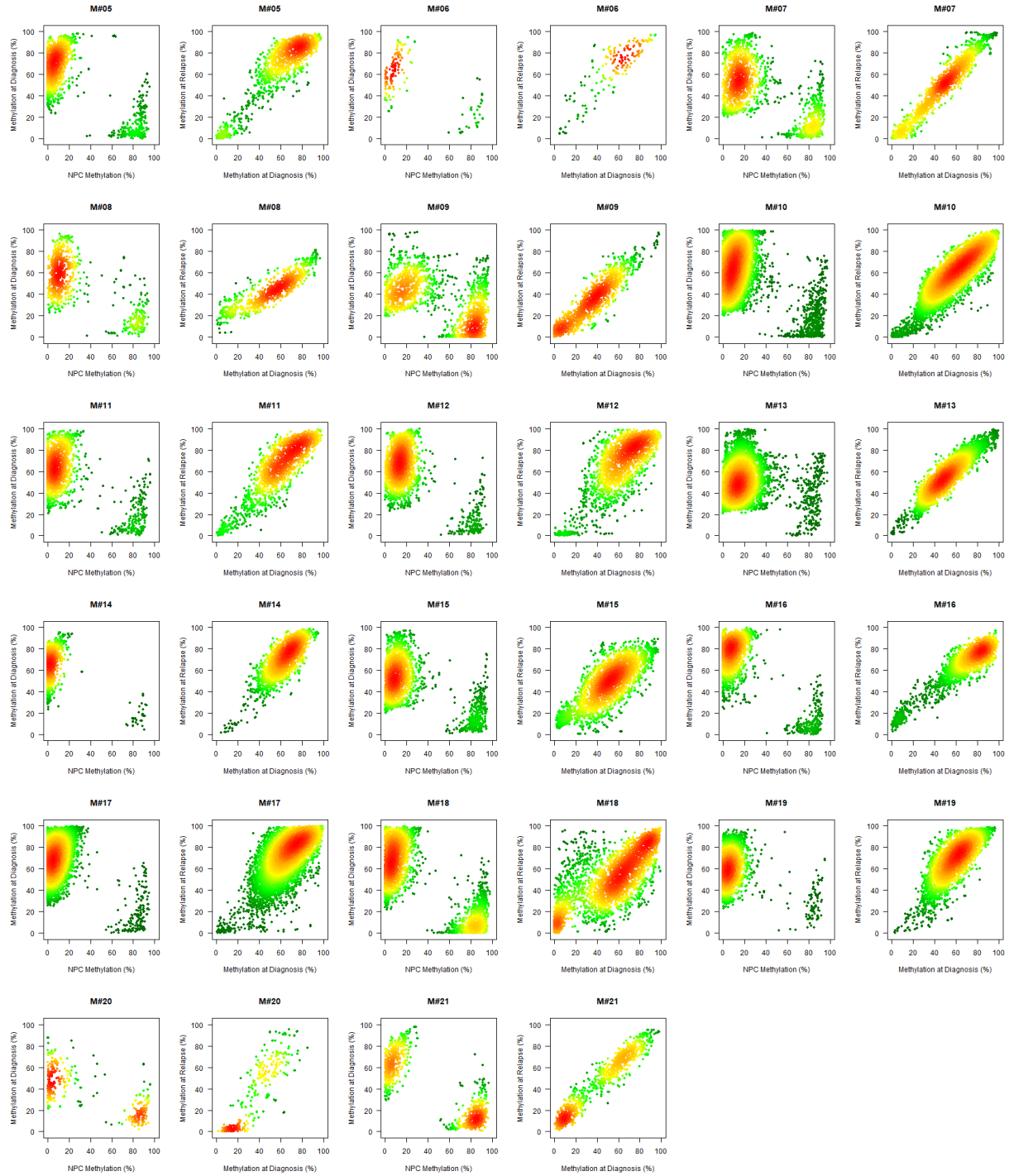


FIGURE 50 – Évolution de la méthylation de l'ADN des eloci dans les promoteurs bivalents au cours de l'initiation et de la progression du MM. Chaque point représente un elocus dans un promoteur bivalent, en fonction de leur méthylation dans les échantillons NPC, MM diagnostic et rechute. Le code couleur correspond à la densité des points, moins il y a de points dans une zone et plus la couleur va vers le vert, plus il y a de points dans une zone du graphique et plus la couleur va vers le rouge.

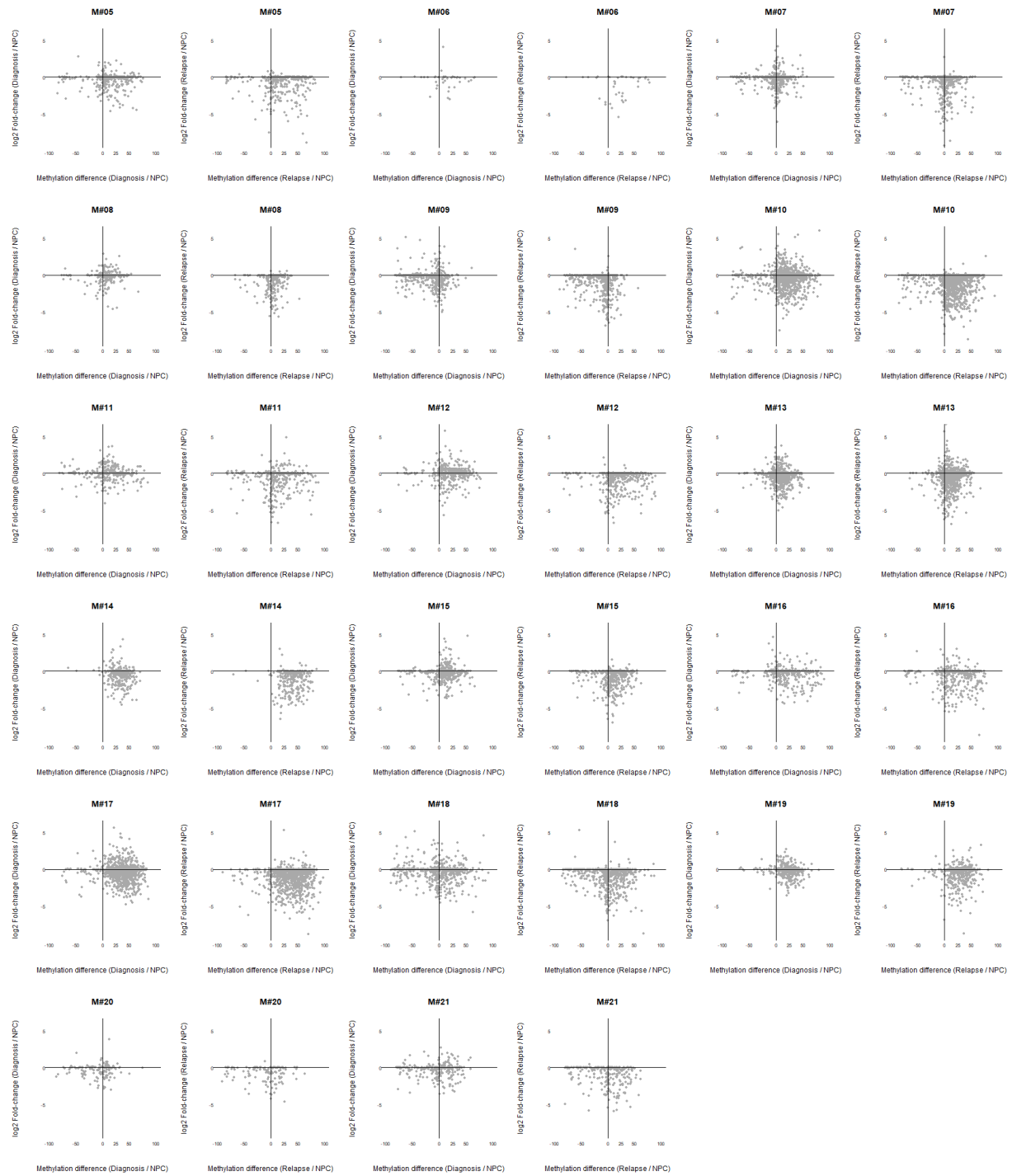


FIGURE 51 – Évolution de l'expression des gènes ayant un elocus dans leur promoteur bivalent en fonction de l'évolution de la méthylation dans ce promoteur au diagnostic et à la rechute par rapport aux NPCs.

Étant donné que plus de 63% des eloci se situent dans des promoteurs-CGI, nous nous sommes intéressés plus particulièrement à ces régions. Et nous avons observé une augmentation significative de l'épipolymorphisme et du PDR des eloci dans les promoteurs-CGIs bivalents au diagnostic par rapport aux NPCs (Figure 52). Ces données suggèrent que l'augmentation de la méthylation dans les promoteurs-CGIs bivalents qui se produit dès l'initiation de la maladie est principalement générée par un processus stochastique qui contribue à une augmentation de l'hétérogénéité intratumorale au détriment d'une méthylation plus bimodale dans les cellules plasmocytaires saines (grande majorité d'épiallèles complètement déméthylés et méthylés dans les échantillons NPCs).

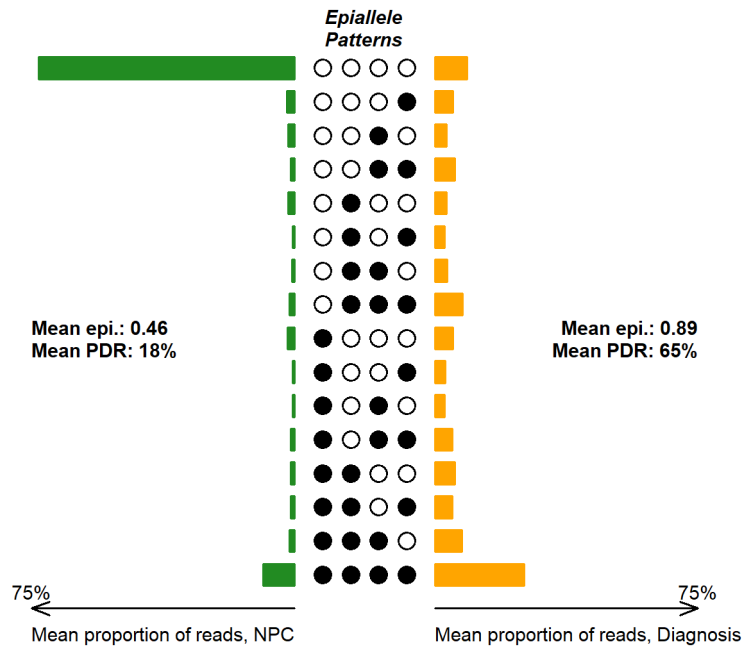


FIGURE 52 – Pattern de méthylation moyen des eloci dans les promoteurs-CGI bivalents pour les échantillons NPCs en vert et pour MM au diagnostic en orange. La variable "mean epi" correspond à l'épipolymorphisme moyen et la variable "mean PDR" correspond à la proportion moyenne de reads discordants des eloci dans les promoteur-CGIs bivalent.

Afin d'approfondir la nature de la relation entre l'expression des gènes et les forts changements de méthylation entre le stade sain et le stade diagnostic du MM, nous avons comparé l'expression des gènes présentant un elocus hyperméthylé dans leur promoteur-CGI bivalent ou ne présentant aucun elocus (no eloci, Figure 53). Nous avons constaté que les gènes présentant un promoteur-CGIs bivalent contenant des eloci hyperméthylés étaient moins exprimés (Figure 53a), mais présentaient une variabilité inter-patient plus élevée que les autres (Figure 53b).

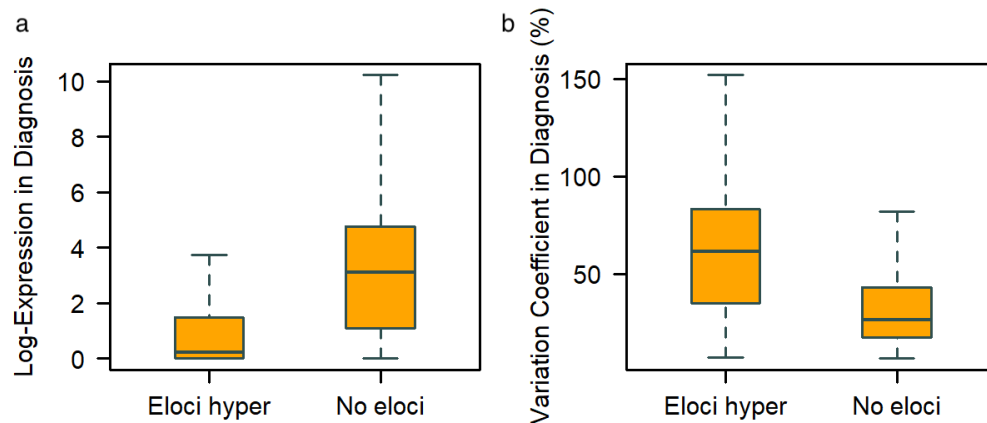


FIGURE 53 – Expression des gènes présentant un elocus hyperméthylé ou ne présentant aucun elocus dans leur promoteur-CGI bivalent. (a) Valeur d’expression des gènes avec un promoteur-CGI bivalent contenant un elocus hyperméthylé ou ne contenant aucun elocus (no eloci). (b) Coefficient de variation de ces gènes entre les échantillons de MM au moment du diagnostic.

Nous avons ensuite réalisé une analyse d’enrichissement d’annotations avec la base de données KEGG de ces gènes contenant des eloci hyperméthylés dans leur promoteur-CGI bivalent, qui a révélé un enrichissement en voies de signalisation importantes dans le cancer, comme les voies de signalisation Wnt et MAPK (Figure 54).

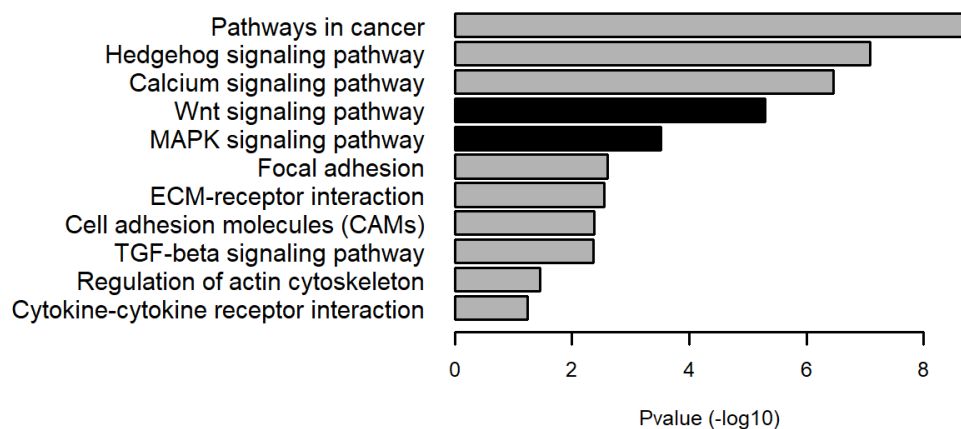


FIGURE 54 – Analyse d’enrichissement d’annotations des gènes dont le promoteur-CGI bivalent est touché par des eloci hyperméthylés.

Des études précédentes ont montré que la plupart des échantillons de MM présentaient une activité aberrante de la voie *Wnt* malgré une faible fréquence de mutation des gènes composants cette voie (3% au total), ce qui suggère que d’autres mécanismes conduisent à cette activation (van Andel et al., 2019). Nous avons pu mettre en évidence dans notre étude qu’un certain nombre de régulateurs négatifs connus de la voie *Wnt*, tels que *SFRP1*, *SFRP2*,

SFRP4, *WIF1*, *MSX1*, *SOX17*, *CDX1* et *CDX2*, étaient touchés par des eloci hyperméthylés dans leur promoteur. Cela suggère que ce mécanisme épigénétique peut être important pour l'activation de la voie *Wnt* dans le MM.

De plus, d'autres études ont montré une méthylation de l'ADN des promoteurs de régulateurs négatifs de la voie de signalisation Ras, tel que *RASSF5* et *RASD1*, une autre voie oncogénique importante dans le MM (De Smedt et al., 2018; Nojima et al., 2009). Nous avons également identifié des eloci hyperméthylés dans les promoteurs de ces deux gènes.

Ci-dessous une illustration montrant une méthylation claire mais variable des promoteur-CGIs des gènes *SFRP2* et *RASD1*, dans tous les échantillons MM par rapport aux NPCs (Figure 55).

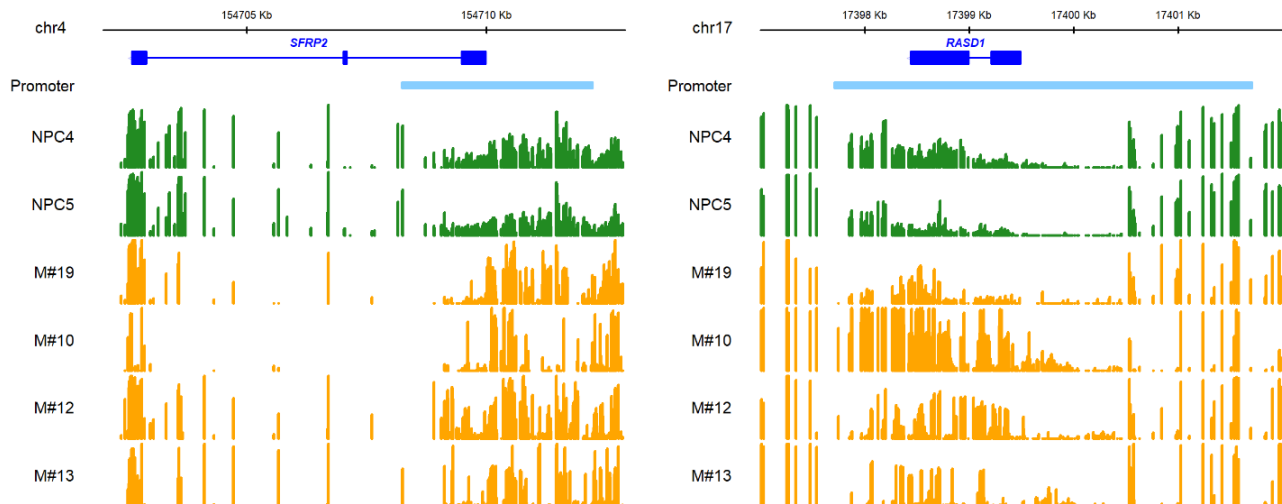


FIGURE 55 – Augmentation variable de la méthylation de l'ADN dans les promoteurs des gènes *SFRP2* et *RASD1* des échantillons de MM comparés aux NPCs.

De plus, nous avons constaté la méthylation du promoteur du gène *CDKN2C*, codant pour un inhibiteur du cycle cellulaire, ainsi que celle des gènes de la famille BMP connus pour induire l'apoptose et l'arrêt de la croissance dans les cellules myélomateuses (Holien and Sundan, 2012).

L'ensemble de ces résultats suggèrent que les changements de méthylation intra et inter-patients au sein de tous ces gènes sont très variables et pourraient jouer un rôle clé dans la transformation tumorale.

6.5.2 Perte de la méthylation de l'ADN dans les domaines partiellement méthylés

Après s'être intéressé aux eloci hyperméthylés, nous avons analysé les eloci hypométhylés localisés principalement dans les régions génomiques dites quiescentes, caractérisées par une absence d'enrichissement en marque d'histone (Ernst and Kellis (2015), Figure 47).

Nous avons étudié les eloci hypométhylés selon l'évolution de leur épipolymorphisme entre le stade sain et le stade diagnostique (Figure 57, partie gauche). Pour une majorité de patients, nous avons constaté deux sous-ensembles d'eloci dont la taille varie selon les patients : un premier sous-ensemble d'eloci complètement déméthylés au diagnostic par rapport aux NPCs et par conséquent présentant un épipolymorphisme nul ou très bas (Figure 56, encadré en bas à droite), et un second sous-ensemble d'eloci dont la perte de la méthylation est très hétérogène (un fort épipolymorphisme) et partielle au diagnostic (Figure 56, encadré en haut à droite).

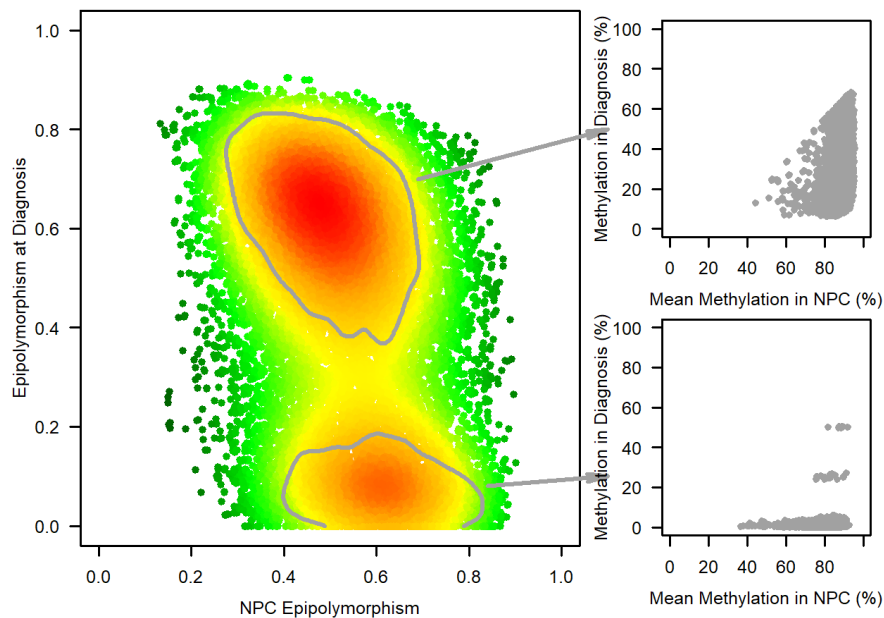


FIGURE 56 – Dispersion des eloci hypométhylés en fonction de leur épipolymorphisme dans les échantillons NPC et au diagnostic pour le patient M#09. Le gradient de couleur correspond à la densité des points, chaque point correspond à un elocus hypométhylé, plus la densité des eloci hypométhylés est élevée et plus la couleur tend vers le rouge (vert si densité peu élevée). Deux populations d'eloci se distinguent sur la figure de gauche, et des informations sur l'évolution de la méthylation de ces deux sous-populations sont données dans les encadrés à droite de la figure.

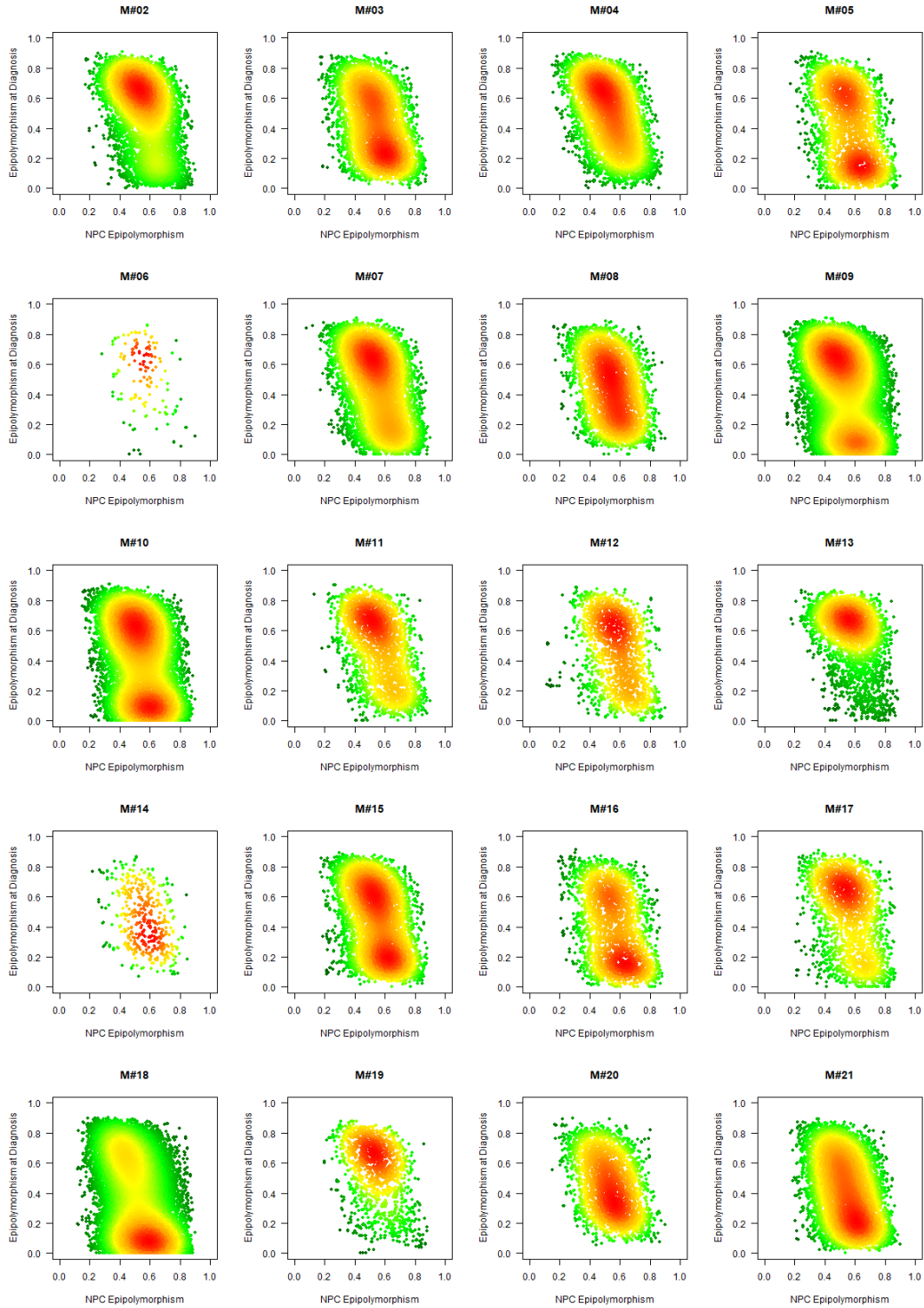


FIGURE 57 – Dispersion des eloci hypométhylés en fonction de leur épipolymorphisme dans les échantillons NPC et MM au diagnostic pour tous les patients. Le gradient de couleur correspond à la densité des points, chaque point correspond à un elocus hypométhylé, plus la densité des eloci hypométhylés est élevée et plus la couleur tend vers le rouge (vert si densité peu élevée).

L'ensemble du génome humain est globalement fortement méthylé (méthylation supérieure à 70%), à l'exception des CGIs et des régions promotrices qui, elles, sont protégées de la méthylation. Cependant, une étude sur des cellules de fibroblastes de poumon foetal humain (IMR90) a montré l'existence de grandes régions génomiques moins méthylées dans les cellules fibroblastiques que dans les cellules souches embryonnaires, qu'ils ont nommé domaines partiellement méthylés (PMDs) (Lister et al., 2009). Les PMDs correspondent à de grandes régions génomiques présentant une méthylation moyenne inférieure à 70% de méthylation, dénotant ainsi avec le reste du génome. Ils sont hypervariables en terme de méthylation, taille et distribution, et affichent des taux de mutation élevés (Brinkman et al., 2019). Les PMDs couvrent environ 50 à 75% du génome des cellules (Salhab et al., 2018).

De nombreuses études ont défini des PMDs dans différents types cellulaires, et différents cancers (Brinkman et al., 2019; Durek et al., 2016; Zhou et al., 2018), et il a été montré qu'un quart des PMDs sont partagés par différents tissus et types cellulaires (Salhab et al., 2018). Les PMDs coïncident avec des domaines répressifs de la chromatine associés à des interactions avec la lamina nucléaire (colocalisation avec les LADs : lamina-associated domains), une réplication tardive et une faible transcription. L'hypométhylation des PMDs est associée à la division cellulaire dès les premiers stades de développement (Salhab et al., 2018; Zhou et al., 2018).

La réplication correspond à la reproduction à l'identique de l'ADN, les profils de méthylation doivent être fidèlement dupliqués. Cette réplication de l'ADN survient pendant la phase S du cycle cellulaire, entre la phase G1 et G2, et dure environ 8h (Weber et al., 2014). Les différentes régions du génome ne se répliquent pas toutes en même temps, mais à des moments différents au cours de la phase S, suivant un programme de temps de réplication lié à l'architecture 3D du génome (Marchal et al., 2017). Les régions répliquées tardivement sont localisées à la périphérie nucléaire; et les régions répliquées précocement sont localisées à l'intérieur du noyau (Weber et al., 2014). En ce qui concerne la méthylation de l'ADN, la réplication donne naissance à un ADN hémi-méthylé dont le brin nouvellement synthétisé doit être méthylé par des méthyltransférases, pour assurer la transmission de la méthylation de l'ADN au cours de la division cellulaire.

Une perte généralisée de méthylation de l'ADN est marquée lors de l'activation des cellules B et conservée lors de la différenciation des cellules B du centre germinal (GC) en cellules B mémoires ou en plasmocytes (Lai et al., 2013). Au cours de leur formation, les plasmocytes de la moelle osseuse subissent plusieurs cycles de division cellulaire, couplés à chaque fois à des pertes de méthylation de l'ADN (Barwick et al., 2016), ce qui conduit à un statut plus hypométhylé que les autres types de cellules (Kulis et al., 2015). Une analyse complète de ces longues régions génomiques présentant une diminution désordonnée des niveaux de méthylation de l'ADN dans les plasmocytes normaux et malins n'a pas encore été réalisée à ma connaissance. Nous avons donc cherché à déterminer si les loci hypométhylés que nous avons observé dans ces régions dites quiescentes coïncidaient avec ces domaines partiellement méthylés.

Afin de déterminer les régions PMDs des plasmocytes et cellules myélomateuses, nous avons utilisé le package R MethylSeekR (Burger et al., 2013) avec ses paramètres par défaut pour analyser les données WGBS de notre cohorte de 24 MM et l'ensemble de données WGBS

disponibles que nous avons pu récupérer (NPC4, NPC5 et 3 patients de MM sur blueprint : M15548, M22965, M23977). MethySeekR permet de définir les régions PMDs, et de les séparer des domaines hautement méthylés et des régions régulatrices faiblement méthylées. Nous avons pu identifier des PMDs dans les deux échantillons de NPCs mais, de manière surprenante, dans seulement cinq de nos 24 échantillons de MM et dans aucun MM de blueprint (Figure 58). Schroeder et al. (2013) ont rapporté que des PMDs pouvaient être détectés dans les ovocytes et les embryons précoces de plusieurs espèces mais qu'ils ne pouvaient pas en détecter dans le placenta, qui est un tissu présentant globalement un faible niveau de méthylation de l'ADN (Schroeder et al., 2013). Ainsi, cela suggère que l'hypométhylation globale marquée observée dans les échantillons de MM empêche la détection de PMDs par MethySeekR (Burger et al., 2013).

Comme expliqué dans le paragraphe Échantillons et techniques utilisées, les échantillons WGBS pour lesquels nous avons pu détecter des PMDs présentent une distribution α bimodale et des valeurs supérieures à 1 (Figure 58). Tandis que les échantillons pour lesquels nous n'avons pas pu en identifier, présentent un seul pic de la distribution des valeurs α qui sont très majoritairement inférieures à 1 (Figure 58).

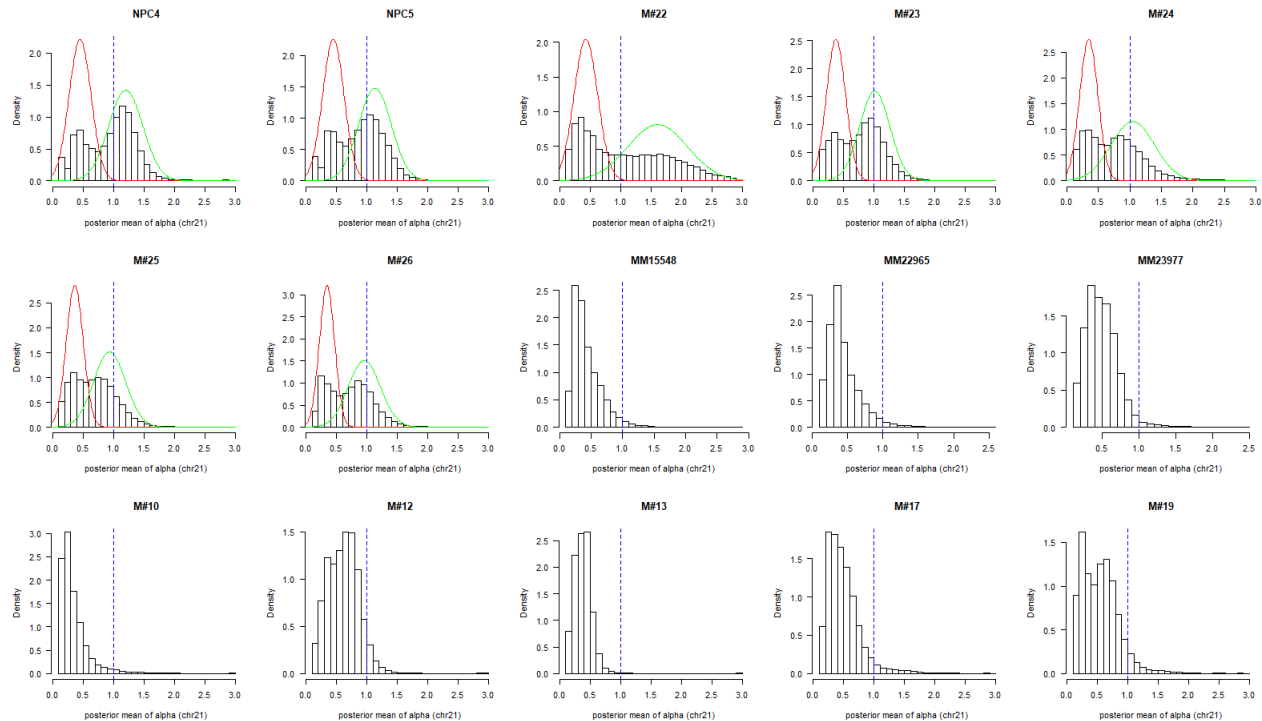


FIGURE 58 – Distribution α sur le chromosome 21 des échantillons de WGBS. Des PMDs ont pu être identifier dans les échantillons présentant une courbe rouge et une courbe verte (distribution bimodale de la distribution α .)

Nous avons observé une structure des PMDs similaire entre les cellules normales et les MM, ceci malgré un niveau très variable de la méthylation de l'ADN (Figure 59a). Les lignes vertes et oranges représentant les niveaux de méthylation de l'ADN des NPC et MM, respectivement. Les lignes grisées correspondent aux MM pour lesquels des PMDs ont pu être identifiés, et présentent des niveaux de méthylation similaires aux NPCs. En revanche, les autres MM pour lesquels nous n'avons pas pu identifier de PMDs présentent une diminution plus ou moins marquée des niveaux de méthylation selon les patients dans ces régions comparées aux NPCs.

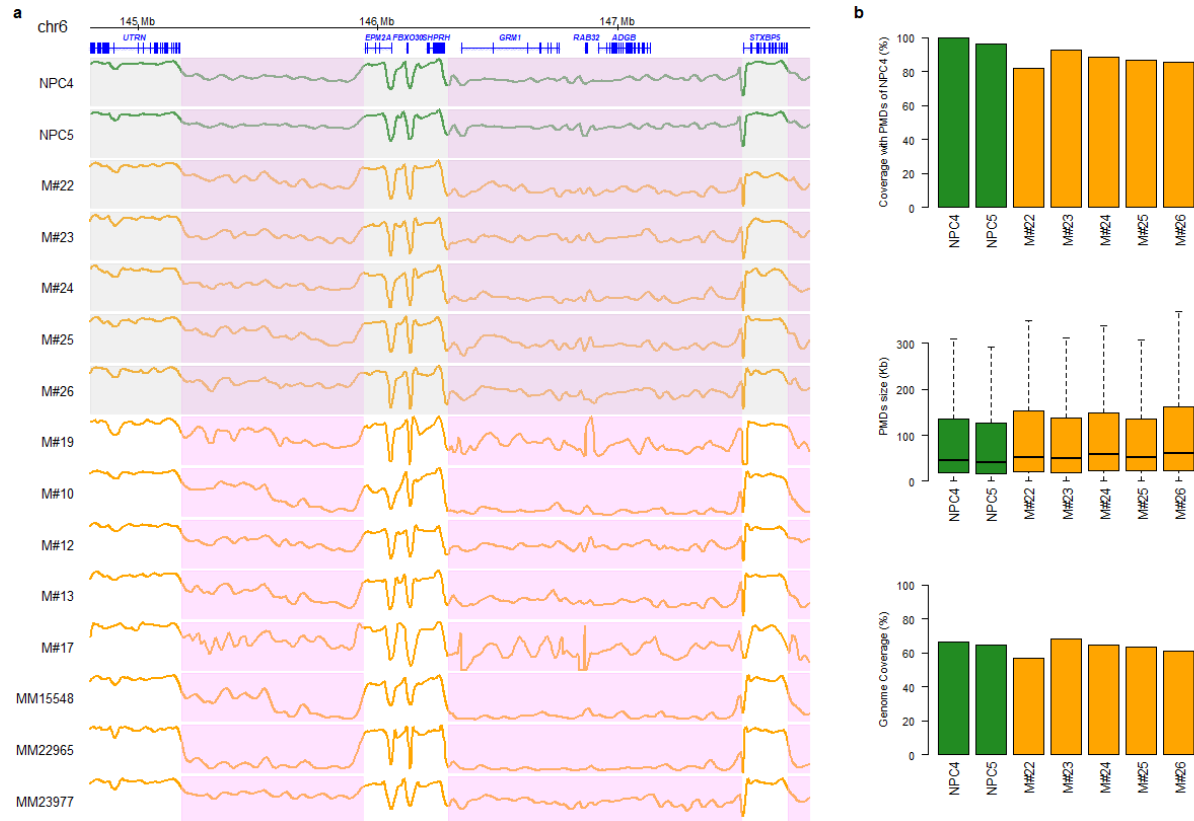


FIGURE 59 – Comparaison des PMDs des cellules normales et des MM. (a) Niveau de méthylation des échantillons de WGBS dans deux régions PMDs (régions rosées) sur le chromosome 6. Les lignes grisées correspondent aux échantillons pour lesquels nous avons pu identifier des PMDs. (b) Le pourcentage de bases communes, la taille et la couverture du génome (en %) des PMDs.

Nous avons observé un très bon chevauchement entre les régions PMDs des deux échantillons de cellules normales (plus de 95% de correspondance), ainsi qu'un bon chevauchement avec les PMDs des MM (plus de 80% de correspondance). Ces régions PMDs par échantillon présentent également une distribution de leur taille et une couverture du génome similaire (Figure 59b). Les PMDs des plasmocytes et des MM couvrent environ 65% du génome et présentent des caractéristiques clés des PMDs de divers types de tumeurs et de tissus normaux (Brinkman et al., 2019; Salhab et al., 2018; Zhou et al., 2018), comme un bon recouvrement avec les

domaines associés à la lamina (LAD : Lamina Associated Domains, Figure 60a), ils sont associés aux domaines de réplication tardive de la GM12878 (Figure 60b), et correspondent à des régions peu denses en gènes (Figure 60c).

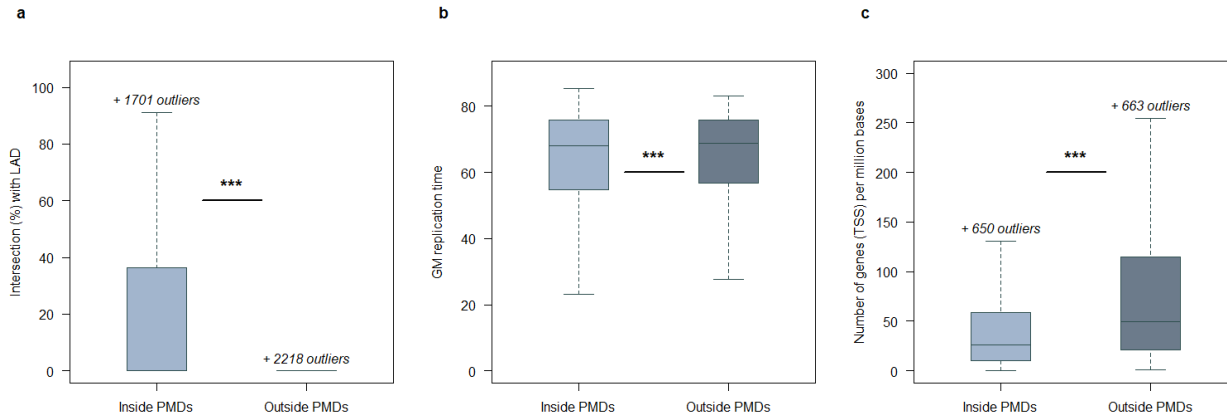


FIGURE 60 – Caractéristiques clés des régions PMDs. (a) Recouvrement des régions PMDs et hors PMDs avec les LADs. (b) Temps de réplication moyen de la lignée GM12878 dans les PMDs et hors des PMDs. Plus un temps de réplication est élevé et plus il est précoce. (c) Densité des gènes dans les PMDs et en dehors.

Les PMDs des cellules plasmocytaires normales présentent 60% de bases communes avec les PMDs identifiés par Zhou et al. (2018) comme étant communs à plusieurs cellules normales et cancéreuses (Zhou et al., 2018), suggérant ainsi qu'un peu moins de la moitié des PMDs que nous avons identifié sont spécifiques des cellules plasmocytaires.

Pour le reste de notre étude, étant donné le très bon recouvrement entre les PMDs des NPCs et ceux des MM pour lesquels des PMDs ont pu être identifiés, nous avons pris les PMDs du NPC4 comme référence. Nous avons observé que plus de 70% des eloci (NPC versus diagnostic) hypométhylés se trouvent dans des PMDs. Nous avons donc repris les 5 échantillons de WGBS en commun avec notre étude eRRBS ainsi que les 3 échantillons de MM récupérés sur blueprint : M15548, M22965, M23977 (Agirre et al., 2015), nous avons regardé leur niveau de méthylation dans les PMDs du NPC4 (Figure 61). Sans surprise, le niveau de méthylation de l'ADN est bien plus faible à l'intérieur qu'à l'extérieur des PMDs, ce qui est une des caractéristiques principales de ces régions.

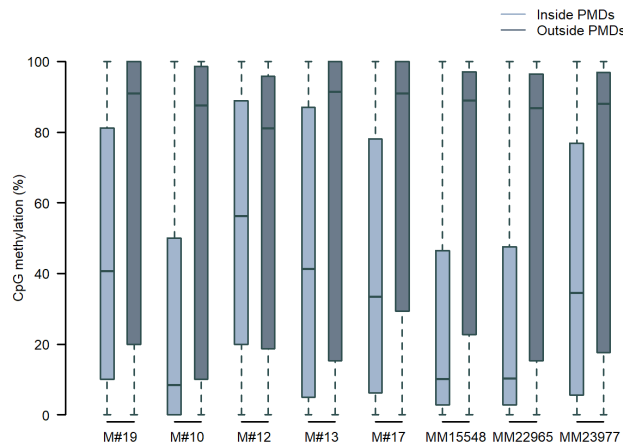


FIGURE 61 – Méthylation moyenne de l’ADN dans les PMDs et en dehors.

L’intérêt premier qui nous a poussé à ajouter à notre étude les 3 échantillons de MM récupérés sur blueprint (Agirre et al., 2015) est la disponibilité pour ces 3 patients de données de WGBS appariées à des données de ChIP-seq, notamment les marques d’histones H3K9me3 et H3K27me3, qui sont deux marques répressives caractéristiques des PMDs (Hawkins et al., 2010).

Nous avons observé pour le patient MM15548 une association entre l’intensité de la marque d’histone H3K9me3 et un faible niveau de méthylation de l’ADN dans des longs PMDs (Figure 62a). Une illustration de ce phénomène est représenté sur la figure 62b, pour ce même patient.

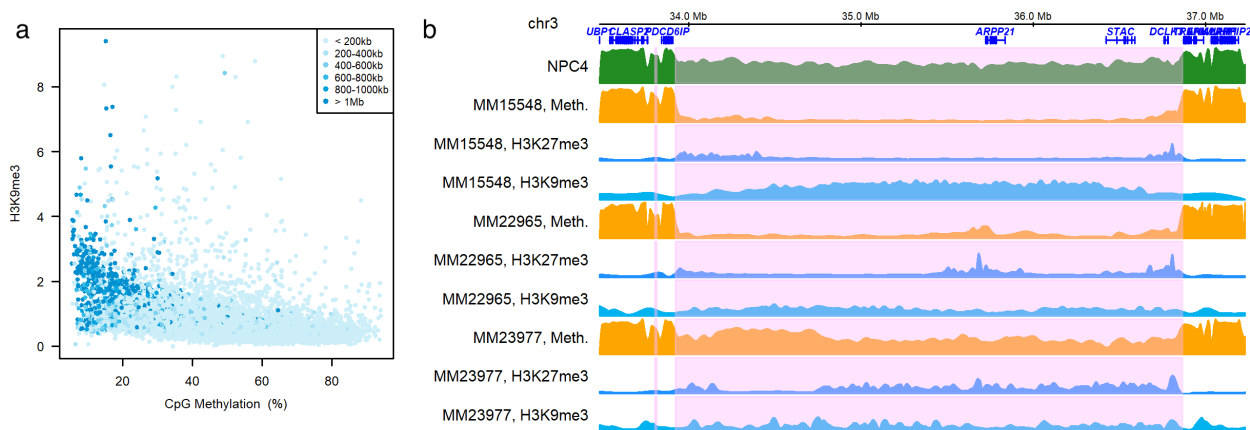


FIGURE 62 – Intensité de H3K9me3 associée à la longueur des PMDs et à leur déméthylation. (a) La méthylation moyenne de l’ADN des PMDs en fonction de l’intensité de la marque H3K9me3, chaque point représente un PMD coloré selon sa taille. (b) Un exemple de long PMD sur le chromosome 3.

Dans un PMD long (3Mb environ, coloré en rose sur la figure 62b), on observe une déméthylation forte pour le patient MM15548 qui est corrélée à une augmentation de la marque H3K9me3. Pour les deux autres patients, la déméthylation est moins complète, l'augmentation de l'intensité de la marque H3K9me3 est moins marquée. Une telle répartition étendue de la marque H3K9me3 a également été observée dans des PMDs de lignée cellulaire de cancer hépatique HepaRG (Salhab et al., 2018) et neuroblastome (Hovestadt et al., 2014). Ils ont constaté que les longs PMDs étaient progressivement déméthylés au cours des divisions cellulaires, et qu'ils présentaient des marques étendues d'H3K9me3 bordées par H3K27me3 (ces deux phénomènes sont constatés dans ce long PMD en exemple pour le patient MM15548), alors que les PMDs les plus courts étaient enrichis en H3K27me3 uniquement.

De plus nous avons observé que les PMDs les plus longs étaient associés à un temps de réplication tardif de la lignée GM12878 (qui a un méthylome similaire aux cellulaires plasmocytaires, Agirre et al. (2015)), ainsi qu'à une hypométhylation variable selon les patients (Figure 63).

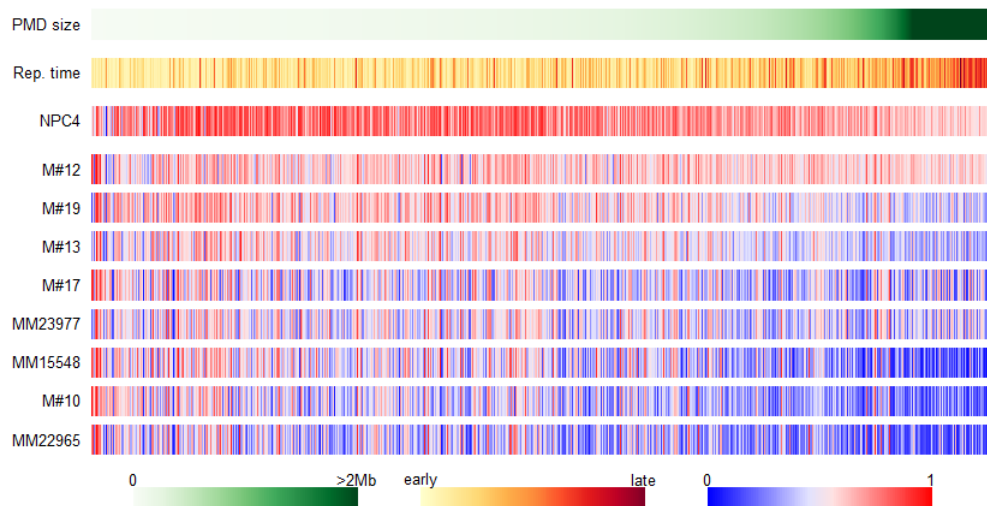


FIGURE 63 – Heatmap des PMDs triés par taille croissante, associés au temps de réplication de la GM12878 et des niveaux de méthylation de l'ADN des échantillons de WGBS.

Baylin and Bestor (2002) ont observé que la perte de la méthylation de l'ADN progressive était plus prononcée dans les domaines à réplication tardive, ce qui est cohérent avec nos résultats. Une hypothèse est que la réplication tardive dans les PMDs donne à la cellule moins de temps pour la méthylation des brins filles nouvellement synthétisés pendant la réplication de l'ADN (Zhou et al., 2018), cela entraînant une perte progressive de la méthylation tout au long de la vie, qui agit comme un biomarqueur du vieillissement cellulaire et qui peut contribuer à l'oncogenèse (Salhab et al., 2018). Zhou et al. (2018) montre que l'hypométhylation du génome dépend du temps de réplication de l'ADN, mais est aussi associée à des territoires nucléaires spécifiques.

La seconde marque caractéristique des PMDs, la marque d'histone H3K27me3, est négativement corrélée à une diminution de la méthylation intermédiaire (coefficient de corrélation $r=0,75$, Figure 64a). Pour une méthylation moyenne des PMDs supérieure à 25% (encadré rouge de la figure 64a), nous observons que plus la méthylation augmente, plus l'intensité de la marque H3K27me3 diminue.

Quelques exemples d'intensité de la marque H3K27me3 dans des PMDs sont illustrés sur la figure 64b. Ici, le patient MM23977 ne présente pas de déméthylation dans le plus grand PMD, à droite, comparé à l'échantillon NPC, et l'intensité de la marque H3K27me3 pour ce même patient est quasi nulle. En revanche, pour les patients MM15548 et MM22965, dans ce même PMD, leur déméthylation est plus marquée comparée au NPC, en parallèle nous observons une augmentation de l'intensité de la marque H3K27me3 sur toute la longueur de ce PMD. Plus la différence de méthylation entre les échantillons NPCs et MM est forte, et plus l'intensité de la marque H3K27me3 est élevée.

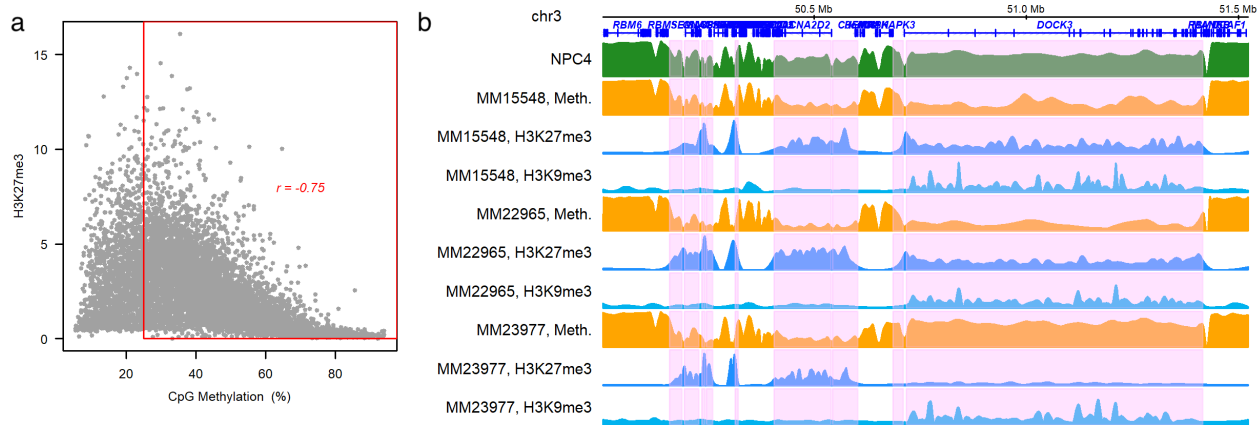


FIGURE 64 – Intensité de H3K27me3 associé à la déméthylation des PMDs. (a) La méthylation moyenne de l'ADN des PMDs en fonction de l'intensité de la marque H3K27me3, chaque point représente un PMD. (b) Exemples de PMDs déméthylés associés à une augmentation de l'intensité de la marque H3K27me3.

Ce phénomène d'augmentation de l'intensité d'H3K27me3 liée à la déméthylation est également visible sur les plus petits PMDs à gauche de la figure 64b. Salhab et al. (2018) ont constaté que les PMDs les plus courts étaient enrichis en H3K27me3, ce qui semble être le cas également dans nos échantillons de MM (Salhab et al., 2018).

6.5.3 Perturbation des régions promotrices dans les PMDs hypométhylées

En plus de la déméthylation majoritaire du génome observée dans les PMDs, nous avons également constaté une grande hétérogénéité de la méthylation entre les patients dans ces domaines (Figure 65a). Suite à ces observations, nous avons ensuite cherché à analyser l'effet de la perturbation de la méthylation de l'ADN dans ces régions sur l'expression des gènes. Nous avons constaté une variabilité d'expression génique plus élevée dans les PMDs qu'en dehors au diagnostic, qui se maintient à la rechute (Figure 65b).

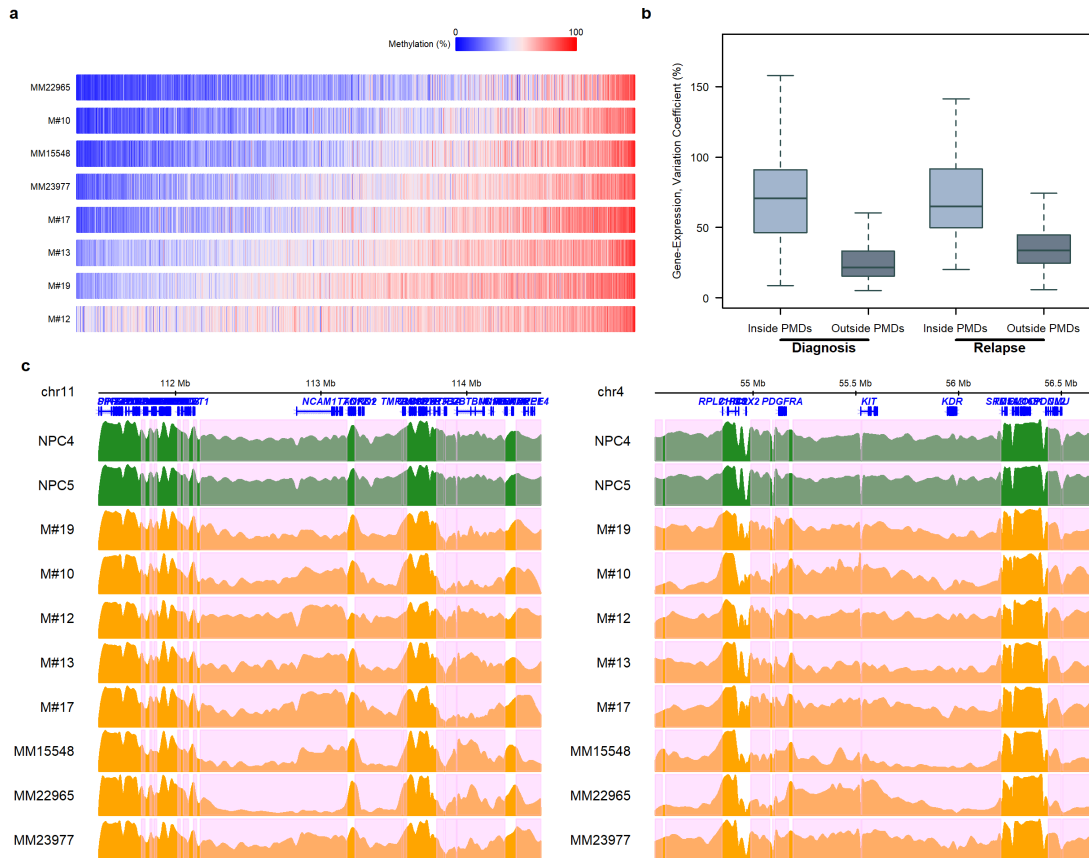


FIGURE 65 – Variabilité de l'expression des gènes et de la déméthylation dans les PMDs. (a) Variabilité inter-patient de la méthylation de l'ADN dans les PMDs. (b) Coefficient de variation de l'expression des gènes dans les PMDs et en dehors au diagnostic et à la rechute. (c) Deux exemples de gènes dans des PMDs présentant une déméthylation variable selon les patients.

Nous avons constaté que les gènes *NCAM1* (CD56) et *KIT* (CD117), deux gènes importants dans le MM, se situent dans des PMDs, et présentent des profils de déméthylation de l'ADN très variables selon les patients (Figure 65c).

CD56 est une glycoprotéine membranaire de la famille des immunoglobulines codée par le gène *NCAM1*, et est une molécule d'adhésion, médiatrice des interactions des cellules de MM entre elles ou avec les cellules ostéoblastiques. Bien que les cellules plasmocytaires normales

n'expriment pas CD56, il est exprimé par les plasmocytes malins chez environ 65 à 80% des patients atteints de MM et peut être associé à une ostéolyse et à un bon pronostic (Bataille et al., 2006). En revanche, l'absence d'expression de CD56 est associée à un mauvais pronostic, à un manque d'ostéolyse et à un état leucémique plasmocytaire (Miyazaki and Suzuki, 2018).

CD117 est un récepteur à activité tyrosine kinase de type III codé par le gène *KIT*. Il est exprimé à la surface d'un sous-ensemble de cellules précurseurs dans la moelle osseuse. L'expression aberrante du CD117 se produit chez environ 30% des patients atteints de MM (Schmidt-Hieber et al., 2011) et est associée à un pronostic favorable. CD117 est exprimé en majorité chez les patients atteints d'un MM indolent, et son expression est significativement moins fréquente à la rechute du MM et dans le cas d'une leucémie plasmocytaire (Bataille et al., 2006).

Une variabilité de la méthylation couplée à une variabilité de l'expression inter-patients de ces gènes observée dans les PMDs pourrait expliquer, en partie, le phénotype aberrant et hétérogène des patients atteints de MM (Capp and Bataille, 2018).

Autre fait intéressant, nous avons remarqué que la majorité des gènes dans les PMDs se trouvaient à moins de 186kb d'une borne d'un PMD (Figure 66a). Ces gènes proches des bornes des PMDs participent, avec un enrichissement significatif, à des voies de signalisation immunitaires (Figure 66b). La présence de ces gènes dans les PMDs pourrait expliquer les mécanismes d'évasions immunitaires qui se produisent au cours de la progression tumorale (Jung et al., 2019; Ryu et al., 2020).

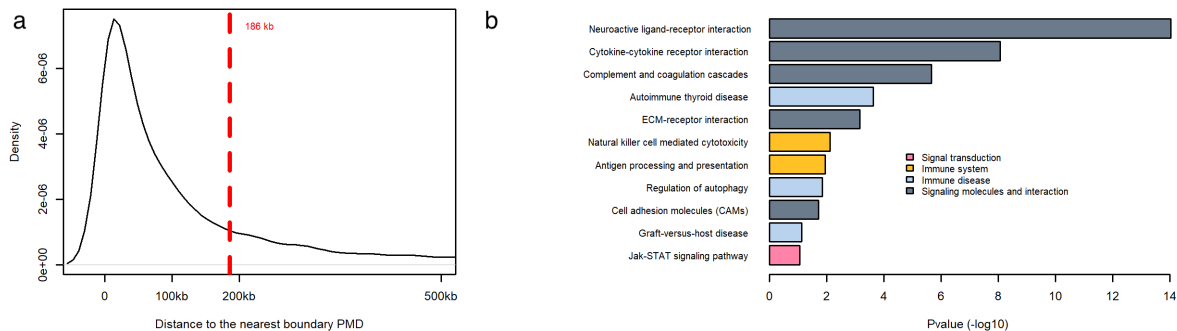


FIGURE 66 – Analyse fonctionnelle des gènes dans des PMD proches d'une borne. (a) Densité de la distance des gènes dans les PMDs à la borne des PMDs la plus proche. La ligne en pointillés rouge correspond en indice au 3ème quartile de la distribution des valeurs. (b) Analyse d'enrichissement d'annotations des gènes dans les PMDs à moins de 186kb d'une borne.

Puis, nous avons caractérisé plus précisément la déméthylation dans les PMDs au niveau des promoteurs et des CGIs. Nous avons constaté qu'à l'intérieur des PMDs, les régions régulatrices (promoteurs et CGIs) perdaient leur état hypométhylé au profit d'un état de méthylation intermédiaire. En revanche, les régions adjacentes (shores et sheaves) hyperméthylées dans les PMDs tendent vers un état intermédiaire de méthylation (Figure 67a).

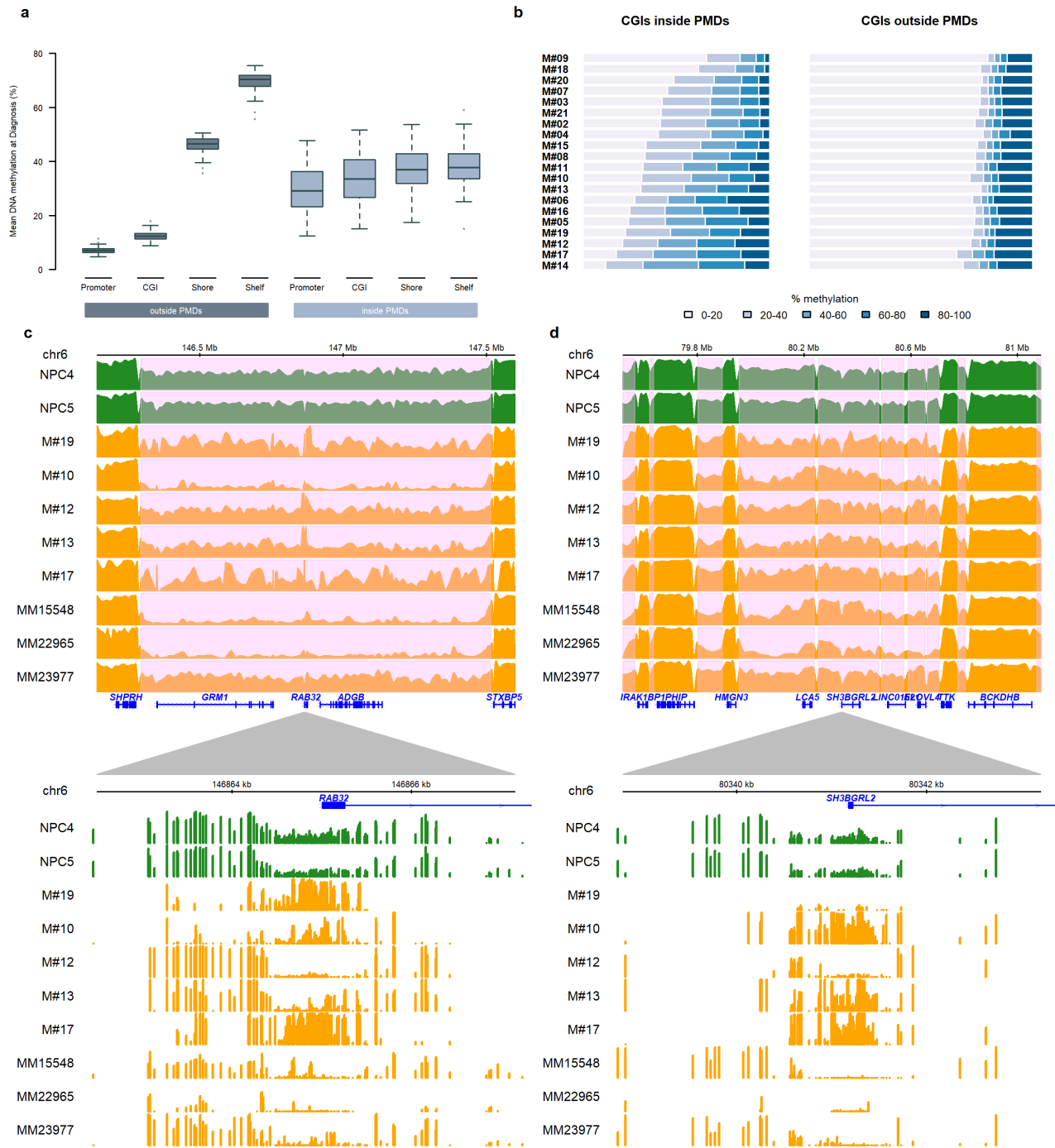


FIGURE 67 – Méthylation des régions régulatrices dans les PMDs déméthylés. (a) Niveau moyen de la méthylation de l'ADN des échantillons au diagnostic dans les régions promotrices, les CGIs, shores et sheaves selon leurs positions dans un PMD ou en dehors. (b) Détails des niveaux de méthylation des CGIs dans un PMD ou en dehors. (c et d) Exemples de promoteurs hyperméthylés situés dans des PMDs hypométhylés au diagnostic du MM comparé aux NPCs.

Globalement, dans les PMDs, toutes les catégories génomiques présentent un niveau de méthylation intermédiaire, alors qu'en dehors des PMDs, les promoteurs-CGIs sont bien déméthylés, et les régions autour sont bien méthylées. Ainsi, il semble que les états de méthylation bimodaux classiquement observés sont perdus dans les PMDs au profit d'états de méthylation de l'ADN plus intermédiaires. On observe dans les PMDs une augmentation significative d'états de la méthylation entre 20% et 80%, dans les CGI au détriment d'état de la méthylation strictement méthylés (plus de 80%) ou non méthylés (moins de 20%), cela, dans tous les échantillons de MM (Figure 67b). Ces résultats sont en accord avec d'autres études dans le cancer du sein, qui ont également révélé que la distribution de la méthylation presque bimodale normalement observée était perdue à l'intérieur des PMDs (Brinkman et al., 2019).

Nous avons remarqué que près de 70% des promoteurs-CGI bivalents présentant une méthylation de l'ADN perturbée (contenant des eloci NPC versus diagnostic), étaient situés dans de grandes régions hypométhylées comme illustré sur les figures 67c et d. Sur ces exemples, les gènes *RAB32* et *SH3BGRL2* présentent dans leur promoteur-CGI bivalent des eloci hyperméthylés et se trouvent dans de grands PMDs déméthylés d'une manière très variable pour tous les patients. Nous avons également constaté, dans les promoteurs de ces gènes, une variabilité de l'hyperméthylation des promoteurs selon les patients de MM.

Dans l'ensemble, ces résultats suggèrent que les cellules myélomateuses présentent une instabilité épigénomique élevée due à une hypométhylation variable et une grande variabilité transcriptomique des PMDs, qui pourrait être bénéfique aux cellules malignes.

6.5.4 Réorganisation de la chromatine due aux perturbations de la méthylation dans les PMDs

Des études récentes montrent une diminution globale de la méthylation de l'ADN entraîne une réorganisation 3D de la chromatine dans les cellules souches embryonnaires (McLaughlin et al., 2019). Nous nous sommes donc demandés si l'hypométhylation globale du génome que l'on observe à l'initiation de la maladie pouvait remodeler l'organisation 3D de la chromatine des cellules de MM.

Lieberman-Aiden et al. ont établi que le génome était divisé en deux compartiments appelés A et B, que les interactions entre régions génomiques étaient plus susceptibles de se produire entre des régions appartenant au même compartiment et que le découplage de ces compartiments étaient spécifiques à un type cellulaire (Lieberman-Aiden et al., 2009). Le compartiment A est associé à la chromatine ouverte, est dense en gènes et actif sur le plan transcriptionnel, il est organisé près du noyau. Le compartiment B est associé lui à la chromatine fermée, peu enrichi en gènes, inactif sur le plan transcriptionnel et organisé près de l'enveloppe nucléaire. Un grand degré de plasticité spatiale dans la disposition des compartiments A et B entre les types de cellules a été mis en évidence, 36% du génome change de compartiment au cours du développement et ces changements de compartiment sont associés à l'expression génique (Dixon et al., 2015).

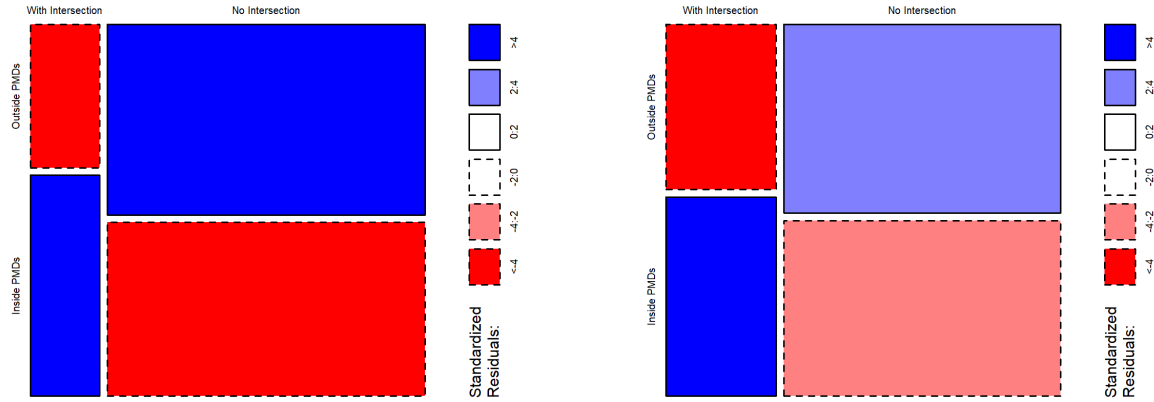
Dans le cas du développement du cancer, il a été observé une réorganisation de l'architecture de la chromatine, notamment des mouvements de régions génomiques passant du compartiment A au compartiment B ou vice versa. La proportion des commutations des compartiments A et B varie selon le type des tumeurs (Barutcu et al., 2015; Wu et al., 2017). L'analyse de la conformation de la chromatine a permis de révéler que des changements de compartiments (A/B) entre cellules normales et cancéreuses sont associés à des changements d'expression de gènes. Dans le cas du MM, Wu et al. (2017) ont mis en évidence qu'environ 20% des régions du génome changeaient de compartiment, et que ces changements de compartiments étaient également associés à des modifications d'expressions géniques (Wu et al., 2017).

Jenkinson et al. ont mis en évidence que le compartiment B présentait en terme de méthylation une plus grande hétérogénéité que le compartiment A, et donc que le compartiment B serait le plus susceptible d'être modifié en terme de méthylation au cours de la cancérogenèse (Jenkinson et al., 2017).

Afin de pouvoir étudier les relations entre les PMDs et les changements de compartiments, nous avons récupéré des données d'HiC pour 3 lignées cellulaires : la GM12878 (couramment utilisée pour représenter les cellules B normales, présentant un méthylome similaire aux cellules plasmocytaires), RPMI-8226 et U266, deux lignées de myélomes générées par Wu et al. (2017). Puis nous avons annoté les régions génomiques en compartiment A et B, à une résolution de 20 kb grâce au package R HiTC (Servant et al., 2012).

Ainsi, nous avons pu observer que les PMDs étaient significativement plus enclins à changer de compartiments que les régions génomiques en dehors des PMDs (Figure 68a). En illustration de ce phénomène, sur la figure 68b, est représenté un exemple de région, passant du compartiment B de la GM12878 au compartiment A dans les lignées de MM, dans un long PMD déméthylé d'une manière très variable selon les patients par rapport aux NPCs. Le gène *IGF1R* code pour le récepteur membranaire du facteur de croissance IGF qui joue un rôle clé dans la croissance, la prolifération et la survie des cellules cancéreuses dans le cas du MM, mais également contribue à l'angiogenèse et à l'ostéolyse associées au MM (Bieghs et al., 2016). *IGF1R* est capable d'activer, en plus de ces voies principales, la voie JAK / STAT et la signalisation Wnt et NF κ B. De plus l'augmentation de l'expression du récepteur IGF1R à la surface des cellules de MM est corrélée à un mauvais pronostic chez certains patients (Bataille et al., 2006). Sur cet exemple, nous observons que le gène *IGF1R* se situe à cheval entre le compartiment A et B dans la GM12878, et se retrouve entièrement dans le compartiment A des deux lignées cellulaires de MM, suggérant une possible activation de ce gène.

a



b

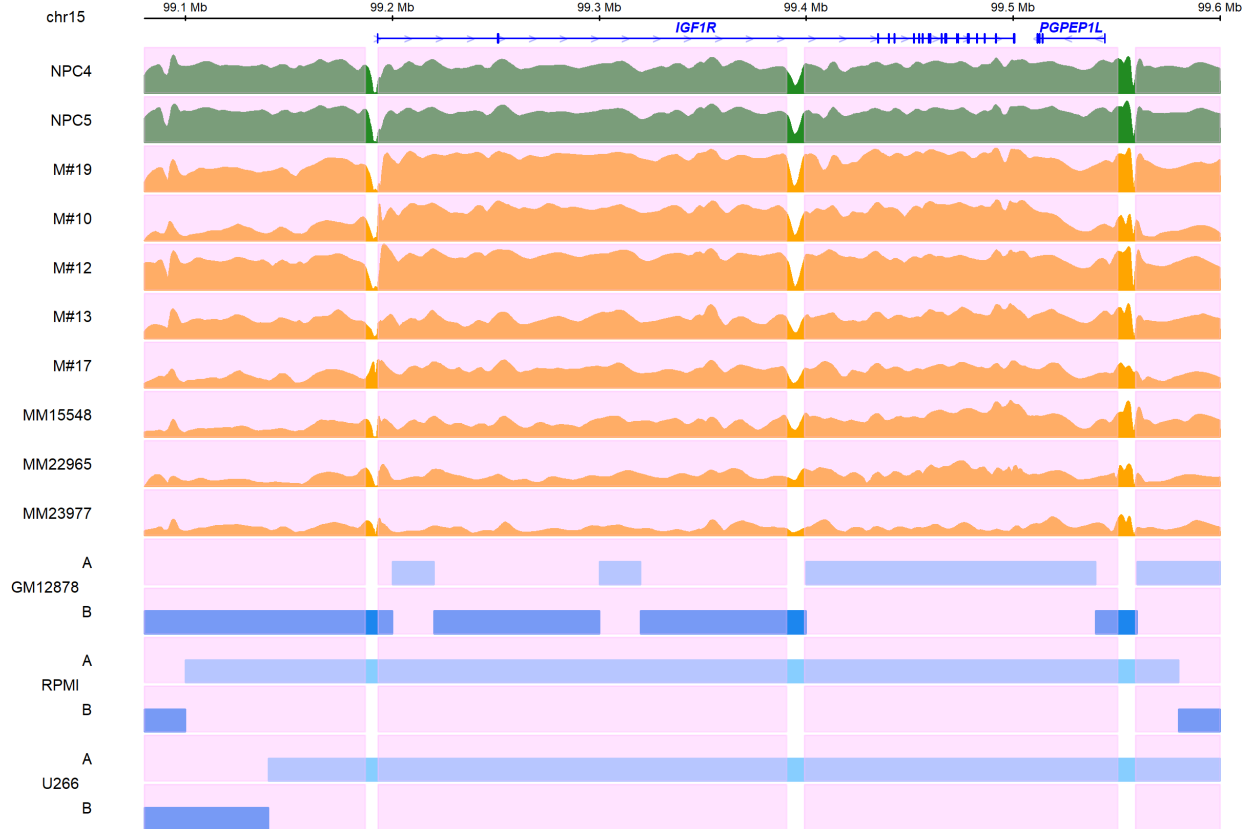


FIGURE 68 – Diminution de la méthylation dans les PMDs et compartimentation du génome. (a) Mosaïque plot révélant à gauche la proportion des régions qui commutent du compartiment B (GM12878) au compartiment A (U266 et RPMI, et à droite la proportion de régions qui commutent du compartiment A (GM12878) au compartiment B (U266 et RPMI) dans ou en dehors des PMDs. (b) Exemple du gène *IGFR1* situé dans un PMD dont une partie du gène commute du compartiment B au compartiment A (GM12878 versus U266 et RPMI).

Ces résultats suggèrent que l'hypométhylation globale du génome observée principalement dans les PMDs pourrait engendrer une réorganisation 3D de la chromatine des cellules plasmocytaires malignes, et leur permettre de moduler l'expression de gènes clés favorisant leur prolifération et survie.

L'ensemble des résultats font l'objet d'un article soumis à une revue scientifique, que vous pouvez retrouver en annexe.

7 Discussion

Le myélome multiple est connu pour être une hémopathie maligne hétérogène. Cependant, jusqu'à présent, seules des études sur des anomalies génétiques ont été réalisées pour démontrer l'existence de populations sous-clonales présentant une évolution complexe au cours de la maladie et leur permettant d'acquérir certains avantages de survie. Une récente étude met en évidence l'implication de l'hétérogénéité épigénétique intra-tumorale dans l'évolution de la leucémie lymphoïde chronique (CLL) suite au traitement (Gaiti et al., 2019). Par mon travail de thèse, nous avons pu mettre également en évidence l'existence et l'implication d'une hétérogénéité épigénétique intra- et interpatient dans le cas du MM. Nous avons réalisé une étude intégrative de la méthylation de l'ADN, des modifications post-traductionnelles des histones et du transcriptome dans des échantillons de MM.

Afin d'analyser les profils de méthylation de l'ADN des patients atteints de MM, nous avons utilisé l'eRRBS comme technique de séquençage. Cette technique de séquençage au bisulfite permet de capturer le statut de méthylation des CpGs individuellement, à une profondeur de séquençage suffisante pour permettre l'étude de l'hétérogénéité épigénétique des tumeurs. Ainsi à partir de ces données, nous avons pu mesurer l'hétérogénéité intra-tumorale de la méthylation de l'ADN à l'initiation et au cours de la progression du MM, à l'aide de l'outil methclone (Li et al., 2014).

Notre étude suggère que l'épigénome du MM est caractérisé par une forte hétérogénéité intra-tumorale qui se produit dans le contexte d'une hypométhylation globale du génome associée à une hyperméthylation plus localisée au niveau de certaines régions régulatrices. Nous avons constaté qu'un fort EPM à l'initiation de la maladie est associé à un risque accru de rechute chez les patients de notre étude. Cela suggère que ces forts changements épigénétiques peuvent avoir un impact sur les mécanismes de rechutes des patients, et pourraient être un potentiel marqueur pronostic. La grande majorité des changements épigénétiques entre les plasmocytes normaux et malins sont associés à une perte de la méthylation de l'ADN localisée dans les PMDs (Figure 69), suggérant que les PMDs des NPCs délimitent des régions génomiques dans lesquelles les informations de méthylation de l'ADN ne sont pas maintenues avec précision au cours des cycles cellulaires. Ce phénomène a également été récemment observé dans une étude sur le compartiment B, qui, en raison d'une dépense énergétique réduite dans ces régions, échoue à maintenir avec précision les informations de la méthylation, ce qui explique le recouvrement de ce compartiment avec les grands blocs d'hypométhylation retrouvés dans les cellules cancéreuses (Jenkinson et al., 2017).

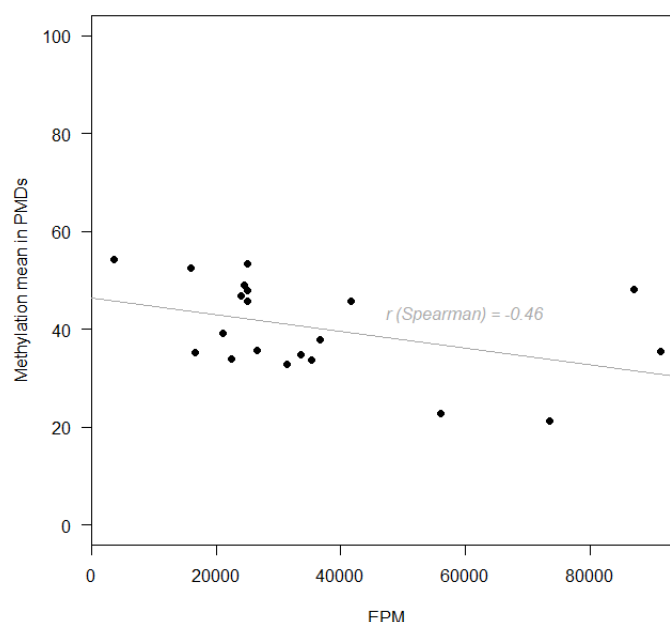


FIGURE 69 – Corrélation entre l'EPM (eloci NPC versus diagnostic) et la méthylation moyenne de l'ADN dans les PMDs au diagnostic du MM.

Les PMDs couvrent environ 70% du génome des cellules plasmocytaires saines et semblent être la principale source de variabilité chez les patients atteints de MM. Les bornes des PMDs ne sont pas disposées aléatoirement sur le génome, elles sont coordonnées à l'organisation 3D de la chromatine.

Les TADs correspondent à des portions de fibres chromatiniennes à l'intérieur desquelles de très nombreuses interactions locales se produisent. Les séquences d'ADN au sein d'un TAD interagissent les unes avec les autres plus fréquemment qu'avec des séquences en dehors du TAD. Wu et al. ont constaté, comme dans d'autres tumeurs, que les génomes de MM contiennent plus de TADs que les cellules B normales et que leur taille moyenne est plus petite (Wu et al., 2017). Nous avons identifié les TADs dans la lignée GM12878 et deux lignées de MM (RPMI et U266), et nous sommes parvenus aux mêmes conclusions. De plus, nous avons observé que les limites des TADs étaient plus proches des bornes réelles des PMDs que des bornes prises au hasard sur le génome (Test de wilcoxon, $p < 0,001$). Jenkinson et al. (2017) ont suggéré que les limites des TADs pouvaient prédire efficacement l'organisation de la chromatine notamment en terme de compartiment A et B (Jenkinson et al., 2017). Il semblerait que les limites des TADs peuvent également prédire l'organisation des PMDs.

Les cellules plasmocytaires malignes sont fortement déméthylées dans les PMDs des cellules plasmocytaires normales, eux-mêmes déjà fortement déméthylés, ceci faisant du MM l'une des tumeurs les plus hypométhylées (Zhou et al., 2018). La méthylation des PMDs dépend de la capacité des cellules à maintenir la méthylation sur le brin d'ADN nouvellement synthétisé et est étroitement liée au temps de réplication et à sa localisation dans le noyau (Zhou et al., 2018). Les PMDs des cellules B naïves présentent une forte perte progressive de la

méthylation de l'ADN au cours de la différenciation en plasmocytes (Figure 70). Au stade de la différenciation terminale des cellules B, les plasmocytes sortent du cycle cellulaire, et ne sont plus capables de proliférer. Les cellules myélomateuses, elles, le sont et re-rentrent dans le cycle cellulaire, ce qui renforce leur hypométhylation globale au cours des cycles cellulaires successifs. Nous avons observé que les PMDs les plus longs, les plus déméthylés, présentent des temps de réplifications plus tardifs, qui ne semblent pas maintenir les profils de méthylation. Ryba et al. (2012) ont observé un degré d'instabilité dans le programme de répllication des échantillons leucémiques comparés à des lignées cellulaires témoins présentant des profils de répllication remarquablement stables et spécifiques des types cellulaires. La principale cause de variabilité de la méthylation pourrait donc être due à un état de prolifération différent entre les cellules cancéreuses ou à des temps de répllication globalement modifiés (Ryba et al., 2012).

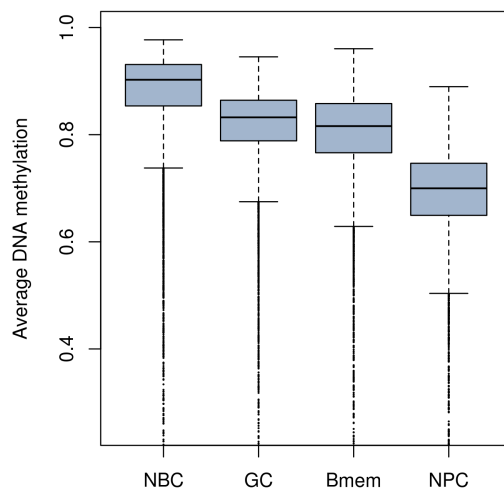


FIGURE 70 – Moyenne de la méthylation de l'ADN des PMDs dans les cellules B à différents stades de différenciation. Cellules B naïves (NBC), cellules B du centre germinatif (GC), cellules B mémoires (Bmem), plasmocytes de la moelle osseuse (NPC).

Les PMDs sont liées à d'autres aberrations épigénétiques clés telles que l'hyperméthylation des régions régulatrices (promoteurs et CGIs), nous avons notamment observé que cette méthylation se produisait d'une manière désordonnée dans le cas du MM (valeur de PDR élevée). Plusieurs études ont montré que la méthylation du promoteur est associée à du bruit transcriptionnel (Huh et al., 2013), qui faciliterait la résistance au traitement (Farquhar et al., 2019). Dans le cas de la leucémie lymphoïde chronique, Landau et al. (2014) proposent que la méthylation désordonnée joue un rôle similaire à l'instabilité génétique, permettant aux cellules cancéreuses une plus grande diversité de population, facilitant leur capacité à rechercher de nouvelles trajectoires évolutives leur étant plus favorables. Ils ont montré qu'un fort PDR était associé à un état transcriptionnel intermédiaire qui interfère avec les états transcriptionnels habituels, c'est-à-dire une répression complète ou un fort niveau d'expression

(Landau et al., 2014). Au vue des résultats de cette étude, et le fort PDR également observé dans le MM, nous pouvons supposer que la méthylation désordonnée dans les PMDs pourrait être associée à du bruit transcriptionnel, permettraient aux cellules tumorales de mieux s'adapter et de trouver de nouvelles trajectoires facilitant la résistance au traitement.

Les profils de méthylation de l'ADN des patients atteints de MM présentent des similitudes avec les profils de méthylation des cellules placentaires. Notamment une hyperméthylation des CGIs localisés dans de grandes régions hypométhylées, ce qui suggère que, comme dans d'autres cancers, la programmation nucléaire placentaire pourrait être cooptée dans les cellules myélomateuses (Lorincz and Schübeler, 2017). Le fait que les tissus placentaires et cancéreux adoptent un méthylome similaire est particulièrement intéressant, car ils partagent des caractéristiques clés telles que la modulation immunitaire, l'angiogenèse et l'invasion cellulaire (Novakovic and Saffery, 2013).

Nous avons constaté que les CGIs hyperméthylés impactent des promoteurs de gènes impliqués dans le développement. Certains de ces gènes sont des inhibiteurs de voies oncogéniques importantes dans le MM comme les voies Wnt et Ras / MAPK. L'hyperméthylation des promoteurs des gènes du développement, en réponse aux changements environnementaux ou au vieillissement, ont été décrits comme favorisant le processus oncogène dans les cancers du poumon et du côlon respectivement (Tao et al., 2019; Vaz et al., 2017). Nos résultats supposent qu'un mécanisme identique pourrait exister dans MM.

Les promoteur-CGIs contrôlant les gènes du développement sont réprimés par le complexe PRC2 (Polycomb Repressive Complex 2), qui est responsable de la triméthylation de la lysine située en position 27 de l'histone 3 (H3K27me3) via l'activité catalytique d'un de ces composants : EZH2 (enhancer of zeste homologue 2). La méthylation de l'ADN et le complexe PRC2 sont des mécanismes qui jouent un rôle important dans le maintien de la répression transcriptionnelle. Nos résultats suggèrent que l'hypométhylation des PMDs est associée à une perturbation de la marque H3K27me3 qui est variable selon les régions génomiques, les patients et également les niveaux de méthylation de l'ADN. Selon le modèle de Reddington et al. et à la vue de nos résultats, nous pouvons émettre l'hypothèse que l'hypométhylation globale observée à l'initiation de la maladie pourrait entraîner une redistribution de H3K27me3 responsable de la dé-répression de gènes cibles du complexe polycomb et de la répression de nouveaux gènes (Reddington et al., 2013). Cette redistribution pourrait expliquer en partie la dé-répression du gène *HOXA9*, normalement réprimé par H3K27me3 dans les cellules plasmocytaires normales. Dans le cas du MM, cette marque répressive se trouve absente ou très faiblement présente dans la plupart des lignées cellulaires de MM (Chapman et al., 2011). Cette redistribution pourrait également expliquer, en partie, les promoteurs bivalents *de novo* observés par Agarwal et al. (2016) dans le MM (les promoteurs bivalents étant caractérisés par la présence simultanée des marques H3K27me3 et H3K4me3) (Agarwal et al., 2016).

Dans un sous-ensemble de MM présentant une translocation t(4; 14), la redistribution d'EZH2 est également provoquée par la sur-expression du gène *MMSET*, qui est un régulateur majeur de la structure de la chromatine et de la transcription dans les cellules de MM (Popovic et al., 2014).

Fait intéressant, dans un modèle murin de leucémie myéloïde aiguë (AML), il a été observé des fonctions contradictoires et opposées d'EZH2, fonction d'oncogène et de suppresseur de tumeur, dans différentes phases de la même tumeur maligne (Basheer et al., 2019). Extrapolant ce concept du rôle spécifique d'EZH2 évoluant selon le stade de la tumeur dans le cas du MM, nous pouvons faire l'hypothèse que l'hypométhylation de l'ADN globale et progressive qui se produit entre le stade MGUS et le diagnostic du MM (Agirre et al., 2015; Walker et al., 2011) pourrait perturber la liaison d'EZH2, conduisant les cellules MGUS à perdre leur identité, et à acquérir progressivement l'identité des cellules myélomateuses en raison de la redistribution d'EZH2. Suite à ces travaux, l'étude du rôle d'EZH2 au cours des différents stades d'évolution du MM pourrait présenter un réel intérêt, en prenant toutefois en compte le double rôle d'EZH2 qui maintient à la fois l'identité des cellules normales et tumorales. Des études récentes suggèrent que la sur-expression d'EZH2 chez les patients atteints de MM peut-être liée à la prolifération et associée à un mauvais pronostic (Goldsmith et al., 2019; Pawlyn et al., 2017). L'inhibition d'EZH2 a été identifiée comme étant une stratégie thérapeutique prometteuse dans le cas du MM : des essais cliniques sont en cours (Tremblay-LeMay et al., 2018).

Nous avons constaté que l'hypométhylation de l'ADN est variable entre les cellules de myélome d'un individu. De plus, cette diminution de la méthylation modifie la distribution de H3K27me3 d'une manière variable selon les individus. De par ces observations, nous pouvons émettre l'hypothèse que la redistribution de H3K27me3 peut être également variable entre les cellules de myélome (à l'image de la méthylation variable observée), conduisant à une variabilité épigénétique importante entre les cellules myélomateuses. Par conséquent, la masse tumorale d'un patient atteint de MM serait composée d'un mélange de cellules myélomateuses aux identités épigénétiques divergentes comme cela a été démontré récemment dans la leucémie lymphoïde secondaire (Pastore et al., 2019).

Dans le cas de la leucémie lymphoïde chronique (CLL), qui est une tumeur maligne des cellules B, il a été observé des états de chromatine marqués simultanément par deux modifications d'histone : H3K27ac (marque activatrice) et H3K27me3 (marque répressive), qui s'excluent généralement mutuellement. Ces états particuliers de la chromatine ont été observés avec un enrichissement significatif dans les cellules malignes par rapport aux cellules B saines, suggérant ainsi une diversité épigénétique intra-tumorale dans les cellules tumorales. Ce marquage simultané de H3K27ac et H3K27me3 dans une même région est plus vraisemblablement dû à des modifications des histones divergeant d'une cellule à l'autre plutôt qu'à une co-occurrence de ces 2 marques mutuellement exclusives dans les mêmes cellules (Pastore et al., 2019). Dans les cellules B normales, l'organisation bien coordonnée des marques d'histones stabilise l'expression des gènes et l'identité cellulaire, alors qu'au contraire dans les cellules malignes, la diversité épigénétique entraîne une variabilité de l'expression des gènes entre les cellules (Figure 71, tirée de Pastore et al. (2019)).

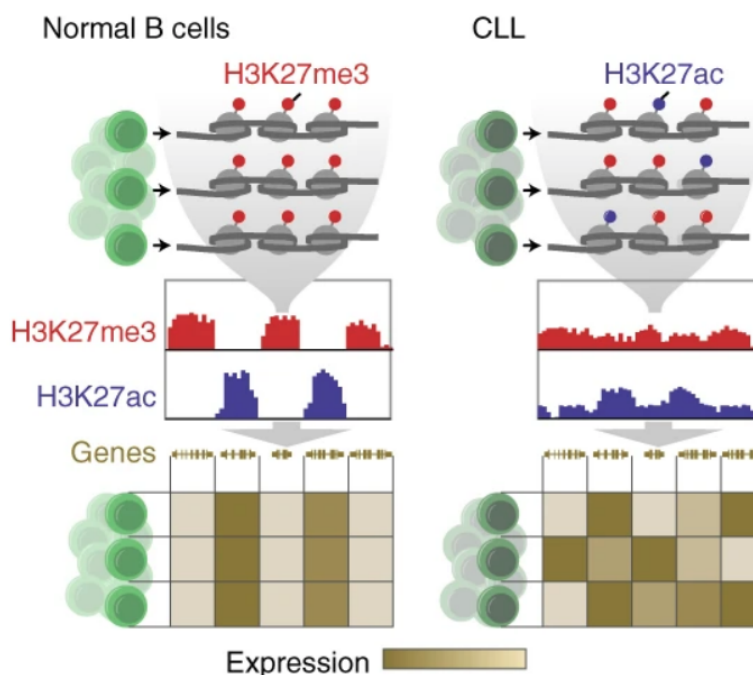


FIGURE 71 – Hétérogénéité transcriptionnelle due au marquage simultané de H3K27ac et H3K27me3 dans le cas de la leucémie lymphoïde chronique. Figure tirée de Pastore et al. (2019).

Dans le cas du cancer du sein, il a été mis en évidence une certaine hétérogénéité des modifications post-traductionnelles des histones parmi les populations de cellules tumorales (Grosselin et al., 2019). Il a été montré que les profils des marques d'histones H3K4me3 et H3K27me3 des cellules sensibles au traitement étaient plutôt homogènes alors que la population de cellules résistantes présentait des états de chromatine plus hétérogènes, et notamment un appauvrissement en H3K27me3 par rapport aux cellules sensibles. Cette étude suggère l'émergence de populations hétérogènes de cellules qui sont résistantes au traitement avec des caractéristiques de chromatine bien distinctes.

De plus, la redistribution de H3K27me3 due à l'hypométhylation de l'ADN pourrait également jouer un rôle important sur la décompaction de la chromatine comme cela a été montré dans les cellules souches embryonnaires (McLaughlin et al., 2019) et pourrait expliquer l'augmentation des régions de chromatine ouvertes dans des états d'hétérochromatine des échantillons de MM par rapport aux échantillons normaux (Jin et al., 2018). Jin et al. (2018) suggèrent qu'une conversion généralisée de régions normalement hétérochromatiques en chromatine active accessible serait une étape clé dans le développement du MM.

8 Conclusion et perspectives

Nous avons généré un ensemble complet de données de méthylation à haute résolution, d'une cohorte de patients atteints de MM comprenant des échantillons appariés à différents stades de la maladie (SMM, diagnostic et rechute). Ce travail a révélé une perte globale de la méthylation de l'ADN dans de grandes régions du génome intégrant des régions régulatrices méthylées d'une manière désordonnée. Une grande variabilité inter-patient a été observée d'un point de vue épigénétique et transcriptionnel. Ces résultats suggèrent que lors de la prolifération des cellules malignes, le manque de maintien précis de la méthylation de l'ADN dans les PMDs conduit à de fortes instabilités épigénétiques et transcriptomiques permettant aux cellules tumorales de s'adapter aux changements environnementaux et d'emprunter de nouvelles trajectoires d'évolution en réponse aux traitements thérapeutiques.

Ce travail a permis la génération d'un pipeline d'analyse complet de données eRRBS permettant, à partir de données brutes de deux échantillons, d'y extraire et de mesurer les niveaux de méthylation des CpGs, d'identifier des loci de 4 CpGs adjacents, de mesurer leur niveau d'hétérogénéité (épipolymorphisme) et d'identifier les principaux changements épialléliques entre les deux échantillons (eloci).

Le faible nombre d'échantillons ($n=4$), à notre disposition, de patients au stade SMM nous a empêché de tirer des conclusions sur cette étape de la maladie. Toutefois, nous avons pu remarquer que l'augmentation de l'hétérogénéité de la méthylation semblait se produire dès le stade SMM. Il aurait été intéressant d'intégrer à notre étude, un nombre plus conséquent de patients au stade SMM pour vérifier si cette tendance se confirmait dans une cohorte de patients plus importantes. De plus, ayant remarqué la présence d'une hétérogénéité de la méthylation déjà très présente au stade SMM, il aurait été intéressant d'intégrer également à notre étude des patients au stade MGUS, afin de pouvoir déterminer à quel stade de la maladie l'apparition de cette hétérogénéité de la méthylation a lieu.

Nous disposons de données de WGBS de 24 échantillons appariées diagnostic et rechute de patients. Une étude approfondie de l'ensemble des données de WGBS sera importante à réaliser prochainement. Nous avons pour le moment étudié qu'une partie des données nous permettant d'appuyer et de valider nos observations faites à partir des données eRRBS.

Suite à ces résultats, il sera également intéressant d'aller étudier la redistribution de H3K27me3 à l'initiation du MM, corrélée aux niveaux de méthylation de l'ADN. Pour cela, il faudrait générer des données de ChIP-seq avec la marque H3K27me3 dans des échantillons NPC et MM pour lesquels nous avons des données de méthylation de l'ADN appariées.

Et pour finir, nous avons mis en évidence une variabilité de la méthylation intra et inter-patient grâce aux données eRRBS séquencées profondément et à l'outil methclone. En revanche, dans les PMDs, nous n'avons pu mettre en évidence qu'une variabilité transcriptomique inter-patient due aux données de séquençage RNAseq en bulk de cellules. Le développement des méthodes de séquençage en cellule unique, notamment le single-cell RNA-seq (Tang et al., 2009) et les nombreux outils bio-informatique associés permettraient de coupler les techniques eRRBS et single-cell RNA-seq afin d'examiner les relations entre les promoteurs ayant une méthylation désordonnée ou ordonnée et l'expression des gènes correspondants dans chaque cellule de la population totale de cellules tumorales (Landau et al., 2014).

Suite au développement des nouvelles méthodes de séquençage à haut débit, notre étude aurait pu être réalisée en single-cell bisulfite sequencing (Smallwood et al., 2014). De plus, il existe de nombreuses autres techniques d'étude du génome entier en cellule unique, telles que le single-cell ATAC (Buenrostro et al., 2015), le single-cell ChIP-seq (Rotem et al., 2015), le single-cell HiC (Nagano et al., 2013) et le single-cell Repli-seq (Miura et al., 2019).

Références

- Picard toolkit. <http://broadinstitute.github.io/picard/>, 2019.
- Aagaard-Tillery Kjersti M ; Grove Kevin ; Bishop Jacalyn ; Ke Xingrao ; Fu Qi ; McKnight Robert, and Lane Robert H. Developmental origins of disease and determinants of chromatin structure : maternal diet modifies the primate fetal epigenome. *Journal of molecular endocrinology*, 41(2) :91, 2008.
- Abdennur Nezar and Mirny Leonid A. Cooler : scalable storage for hi-c data and other genomically labeled arrays. *Bioinformatics*, 36(1) :311–316, 2020.
- Agarwal Prasoon ; Alzrigat Mohammad ; Párraga Alba Atienza ; Enroth Stefan ; Singh Umashankar ; Ungerstedt Johanna ; Österborg Anders ; Brown Peter J ; Ma Anqi ; Jin Jian, and others . Genome-wide profiling of histone h3 lysine 27 and lysine 4 trimethylation in multiple myeloma reveals the importance of polycomb gene targeting and highlights ezh2 as a potential therapeutic target. *Oncotarget*, 7(6) :6809, 2016.
- Agirre Xabier ; Castellano Giancarlo ; Pascual Marien ; Heath Simon ; Kulis Marta ; Segura Victor ; Bergmann Anke ; Esteve Anna ; Merkel Angelika ; Raineri Emanuele, and others . Whole-epigenome analysis in multiple myeloma reveals dna hypermethylation of b cell-specific enhancers. *Genome research*, 25(4) :478–487, 2015.
- Aithal Madhuri GS and Rajeswari Narayanappa. Role of notch signalling pathway in cancer and its association with dna methylation. *Journal of genetics*, 92(3) :667–675, 2013.
- Akalin Altuna ; Kormaksson Matthias ; Li Sheng ; Garrett-Bakelman Francine E ; Figueroa Maria E ; Melnick Ari, and Mason Christopher E. methylkit : a comprehensive r package for the analysis of genome-wide dna methylation profiles. *Genome biology*, 13(10) :R87, 2012.
- Alves Gilda ; Tatro Alicia, and Fanning Thomas. Differential methylation of human line-1 retrotransposons in malignant cells. *Gene*, 176(1-2) :39–44, 1996.
- Anders Simon ; Pyl Paul Theodor, and Huber Wolfgang. Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2) :166–169, 2015.
- Andrews Simon and others . Fastqc : a quality control tool for high throughput sequence data, 2010.
- Ball Madeleine P ; Li Jin Billy ; Gao Yuan ; Lee Je-Hyuk ; LeProust Emily M ; Park In-Hyun ; Xie Bin ; Daley George Q, and Church George M. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nature biotechnology*, 27(4) :361, 2009.
- Bannister Andrew J and Kouzarides Tony. Regulation of chromatin by histone modifications. *Cell research*, 21(3) :381–395, 2011.

- Bao Ning ; Lye Khar-Wai, and Barton M Kathryn. Microrna binding sites in arabidopsis class iii hd-zip mrnas are required for methylation of the template chromosome. *Developmental cell*, 7(5) :653–662, 2004.
- Barutcu A Rasim ; Lajoie Bryan R ; McCord Rachel P ; Tye Coralee E ; Hong Deli ; Messier Terri L ; Browne Gillian ; van Wijnen Andre J ; Lian Jane B ; Stein Janet L, and others . Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome biology*, 16(1) :214, 2015.
- Barwick Benjamin G ; Scharer Christopher D ; Bally Alexander PR, and Boss Jeremy M. Plasma cell differentiation is coupled to division-dependent dna hypomethylation and gene regulation. *Nature immunology*, 17(10) :1216, 2016.
- Basheer Faisal ; Giotopoulos George ; Meduri Eshwar ; Yun Haiyang ; Mazan Milena ; Sasca Daniel ; Gallipoli Paolo ; Marando Ludovica ; Gozdecka Malgorzata ; Asby Ryan, and others . Contrasting requirements during disease evolution identify ezh2 as a therapeutic target in aml. *Journal of Experimental Medicine*, 216(4) :966–981, 2019.
- Bataille Régis ; Jégo Gaëtan ; Robillard Nelly ; Barillé-Nion Sophie ; Harousseau Jean-Luc ; Moreau Philippe ; Amiot Martine, and Pellat-Deceunynck Catherine. The phenotype of normal, reactive and malignant plasma cells. identification of" many and multiple myelomas" and of new targets for myeloma therapy. *Haematologica*, 91(9) :1234–1240, 2006.
- Baylin Stephen and Bestor Timothy H. Altered methylation patterns in cancer cell genomes : cause or consequence? *Cancer cell*, 1(4) :299–305, 2002.
- Belton Jon-Matthew ; McCord Rachel Patton ; Gibcus Johan Harmen ; Naumova Natalia ; Zhan Ye, and Dekker Job. Hi-c : a comprehensive technique to capture the conformation of genomes. *Methods*, 58(3) :268–276, 2012.
- Berdasco María and Esteller Manel. Aberrant epigenetic landscape in cancer : how cellular identity goes awry. *Developmental cell*, 19(5) :698–711, 2010.
- Berger Shelley L ; Kouzarides Tony ; Shiekhattar Ramin, and Shilatifard Ali. An operational definition of epigenetics. *Genes & development*, 23(7) :781–783, 2009.
- Berman Benjamin P ; Weisenberger Daniel J ; Aman Joseph F ; Hinoue Toshinori ; Ramjan Zachary ; Liu Yaping ; Noushmehr Houtan ; Lange Christopher PE ; van Dijk Cornelis M ; Tollenaar Rob AEM, and others . Regions of focal dna hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature genetics*, 44(1) :40, 2012.
- Bernhart Stephan H ; Kretzmer Helene ; Holdt Lesca M ; Jühling Frank ; Ammerpohl Ole ; Bergmann Anke K ; Northoff Bernd H ; Doose Gero ; Siebert Reiner ; Stadler Peter F, and others . Changes of bivalent chromatin coincide with increased expression of developmental genes in cancer. *Scientific reports*, 6 :37393, 2016.

- Bernstein Emily and Allis C David. Rna meets chromatin. *Genes & development*, 19(14) : 1635–1655, 2005.
- Bibikova Marina ; Barnes Bret ; Tsan Chan ; Ho Vincent ; Klotzle Brandy ; Le Jennie M ; Delano David ; Zhang Lu ; Schroth Gary P ; Gunderson Kevin L, and others . High density dna methylation array with single cpg site resolution. *Genomics*, 98(4) :288–295, 2011.
- Bieghs Liesbeth ; Johnsen Hans E ; Maes Ken ; Menu Eline ; Van Valckenborgh Els ; Overgaard Michael T ; Nyegaard Mette ; Conover Cheryl A ; Vanderkerken Karin, and De Bruyne Elke. The insulin-like growth factor system in multiple myeloma : diagnostic and therapeutic potential. *Oncotarget*, 7(30) :48732, 2016.
- Bland J Martin and Altman Douglas G. Multiple significance tests : the bonferroni method. *Bmj*, 310(6973) :170, 1995.
- Bolger Anthony M ; Lohse Marc, and Usadel Bjoern. Trimmomatic : a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15) :2114–2120, 2014.
- Bollati Valentina ; Schwartz Joel ; Wright Robert ; Litonjua Augusto ; Tarantini Letizia ; Suh Helen ; Sparrow David ; Vokonas Pantel, and Baccarelli Andrea. Decline in genomic dna methylation through aging in a cohort of elderly subjects. *Mechanisms of ageing and development*, 130(4) :234–239, 2009.
- Bolli Niccolo ; Avet-Loiseau Hervé ; Wedge David C ; Van Loo Peter ; Alexandrov Ludmil B ; Martincorena Inigo ; Dawson Kevin J ; Iorio Francesco ; Nik-Zainal Serena ; Bignell Graham R, and others . Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature communications*, 5 :2997, 2014.
- Bolli Niccolò ; Maura Francesco ; Minvielle Stephane ; Gloznik Dominik ; Szalat Raphael ; Fullam Anthony ; Martincorena Inigo ; Dawson Kevin J ; Samur Mehmet Kemal ; Zamora Jorge, and others . Genomic patterns of progression in smoldering multiple myeloma. *Nature communications*, 9(1) :1–10, 2018.
- Braggio Esteban ; Maiolino Angelo ; Gouveia Maria E ; Magalhães Roberto ; Souto Filho João T ; Garnica Márcia ; Nucci Marcio, and Renault Ilana Zalcborg. Methylation status of nine tumor suppressor genes in multiple myeloma. *International journal of hematology*, 91(1) :87–96, 2010.
- Brinkman Arie B ; Nik-Zainal Serena ; Simmer Femke ; Rodriguez-Gonzalez F German ; Smid Marcel ; Alexandrov Ludmil B ; Butler Adam ; Martin Sancha ; Davies Helen ; Glodzik Dominik, and others . Partially methylated domains are hypervariable in breast cancer and fuel widespread cpg island hypermethylation. *Nature communications*, 10(1) :1–10, 2019.
- Brocks David ; Assenov Yassen ; Minner Sarah ; Bogatyrova Olga ; Simon Ronald ; Koop Christina ; Oakes Christopher ; Zucknick Manuela ; Lipka Daniel Bernhard ; Weischenfeldt Joachim, and others . Intratumor dna methylation heterogeneity reflects clonal evolution in aggressive prostate cancer. *Cell reports*, 8(3) :798–806, 2014.

- Buenrostro Jason D ; Wu Beijing ; Litzenburger Ulrike M ; Ruff Dave ; Gonzales Michael L ; Snyder Michael P ; Chang Howard Y, and Greenleaf William J. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, 523(7561) :486–490, 2015.
- Burger Lukas ; Gaidatzis Dimos ; Schübeler Dirk, and Stadler Michael B. Identification of active regulatory regions from dna methylation data. *Nucleic acids research*, 41(16) : e155–e155, 2013.
- Cao Jian and Yan Qin. Histone ubiquitination and deubiquitination in transcription, dna damage response, and cancer. *Frontiers in oncology*, 2 :26, 2012.
- Capp Jean-Pascal and Bataille Régis. Multiple myeloma exemplifies a model of cancer based on tissue disruption as the initiator event. *Frontiers in oncology*, 8 :355, 2018.
- Castillo-Aguilera Omar ; Depreux Patrick ; Halby Ludovic ; Arimondo Paola B, and Goossens Laurence. Dna methylation targeting : the dnmt/hmt crosstalk challenge. *Biomolecules*, 7(1) :3, 2017.
- Chapman Michael A ; Lawrence Michael S ; Keats Jonathan J ; Cibulskis Kristian ; Sougnez Carrie ; Schinzel Anna C ; Harview Christina L ; Brunet Jean-Philippe ; Ahmann Gregory J ; Adli Mazhar, and others . Initial genome sequencing and analysis of multiple myeloma. *Nature*, 471(7339) :467–472, 2011.
- Consortium GR and others . Grch37 genome reference consortium human reference 37 (gca000001405.1), 2009.
- Cui Xiangrong ; Jing Xuan ; Wu Xueqing ; Yan Meiqin ; Li Qiang ; Shen Yan, and Wang Zhenqiang. Dna methylation in spermatogenesis and male infertility. *Experimental and therapeutic medicine*, 12(4) :1973–1979, 2016.
- Curradi Michela ; Izzo Annalisa ; Badaracco Gianfranco, and Landsberger Nicoletta. Molecular mechanisms of gene silencing mediated by dna methylation. *Molecular and cellular biology*, 22(9) :3157–3173, 2002.
- Dagogo-Jack Ibiayi and Shaw Alice T. Tumour heterogeneity and resistance to cancer therapies. *Nature reviews Clinical oncology*, 15(2) :81, 2018.
- Dana Hassan ; Chalbatani Ghanbar Mahmoodi ; Mahmoodzadeh Habibollah ; Karimloo Rezvan ; Rezaiean Omid ; Moradzadeh Amirreza ; Mehmandoost Narges ; Moazzen Fateme ; Mazraeh Ali ; Marmari Vahid, and others . Molecular mechanisms and biological functions of sirna. *International journal of biomedical science : IJBS*, 13(2) :48, 2017.
- Day Kenneth ; Waite Lindsay L ; Thalacker-Mercer Anna ; West Andrew ; Bamman Marcos M ; Brooks James D ; Myers Richard M, and Absher Devin. Differential dna methylation with age displays both common and dynamic features across human tissues that are influenced by cpg landscape. *Genome biology*, 14(9) :R102, 2013.

- De Smedt Eva; Maes Ken; Verhulst Stefaan; Lui Hui; Kassambara Alboukadel; Maes Anke; Robert Nicolas; Heirman Carlo; Cakana Andrew; Hose Dirk, and others . Loss of rassf4 expression in multiple myeloma promotes ras-driven malignant progression. *Cancer research*, 78(5) :1155–1168, 2018.
- Demchenko Yulia N and Kuehl W Michael. A critical role for the nfkb pathway in multiple myeloma. *Oncotarget*, 1(1) :59, 2010.
- Deng Jie; Shoemaker Robert; Xie Bin; Gore Athurva; LeProust Emily M; Antosiewicz-Bourget Jessica; Egli Dieter; Maherali Nimet; Park In-Hyun; Yu Junying, and others . Targeted bisulfite sequencing reveals changes in dna methylation associated with nuclear reprogramming. *Nature biotechnology*, 27(4) :353, 2009.
- Dietz Steffen; Lifshitz Aviezer; Kazdal Daniel; Harms Alexander; Endris Volker; Winter Hauke; Stenzinger Albrecht; Warth Arne; Sill Martin; Tanay Amos, and others . Global dna methylation reflects spatial heterogeneity and molecular evolution of lung adenocarcinomas. *International journal of cancer*, 144(5) :1061–1072, 2019.
- Dillies Marie-Agnès; Rau Andrea; Aubert Julie; Hennequet-Antier Christelle; Jeanmougin Marine; Servant Nicolas; Keime Céline; Marot Guillemette; Castel David; Estelle Jordi, and others . A comprehensive evaluation of normalization methods for illumina high-throughput rna sequencing data analysis. *Briefings in bioinformatics*, 14(6) :671–683, 2013.
- Dimopoulos MA; Kastritis E; Rosinol L; Blade J, and Ludwig H_. Pathogenesis and treatment of renal failure in multiple myeloma. *Leukemia*, 22(8) :1485–1493, 2008.
- Dixon Jesse R; Jung Inkyung; Selvaraj Siddarth; Shen Yin; Antosiewicz-Bourget Jessica E; Lee Ah Young; Ye Zhen; Kim Audrey; Rajagopal Nisha; Xie Wei, and others . Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539) :331–336, 2015.
- Dobin Alexander; Davis Carrie A; Schlesinger Felix; Drenkow Jorg; Zaleski Chris; Jha Sonali; Batut Philippe; Chaisson Mark, and Gingeras Thomas R. Star : ultrafast universal rna-seq aligner. *Bioinformatics*, 29(1) :15–21, 2013.
- Durek Pawel; Nordström Karl; Gasparoni Gilles; Salhab Abdulrahman; Kressler Christopher; De Almeida Melanie; Bassler Kevin; Ulas Thomas; Schmidt Florian; Xiong Jieyi, and others . Epigenomic profiling of human cd4+ t cells supports a linear differentiation model and highlights molecular regulators of memory development. *Immunity*, 45(5) :1148–1161, 2016.
- Duymich Christopher E; Charlet Jessica; Yang Xiaojing; Jones Peter A, and Liang Gangning. Dnmt3b isoforms without catalytic activity stimulate gene body methylation as accessory proteins in somatic cells. *Nature communications*, 7 :11453, 2016.

- Eads Cindy A ; Danenberg Kathleen D ; Kawakami Kazuyuki ; Saltz Leonard B ; Danenberg Peter V, and Laird Peter W. CpG island hypermethylation in human colorectal tumors is not associated with dna methyltransferase overexpression. *Cancer research*, 59(10) : 2302–2306, 1999.
- Easwaran Hariharan ; Johnstone Sarah E ; Van Neste Leander ; Ohm Joyce ; Mosbrugger Tim ; Wang Qiuju ; Aryee Martin J ; Joyce Patrick ; Ahuja Nita ; Weisenberger Dan, and others . A dna hypermethylation module for the stem/progenitor cell signature of cancer. *Genome research*, 22(5) :837–849, 2012.
- Egan Jan B ; Shi Chang-Xin ; Tembe Waibhav ; Christoforides Alexis ; Kurdoglu Ahmet ; Sinari Shripad ; Middha Sumit ; Asmann Yan ; Schmidt Jessica ; Braggio Esteban, and others . Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood, The Journal of the American Society of Hematology*, 120(5) :1060–1066, 2012.
- Ernst Jason and Kellis Manolis. Chromhmm : automating chromatin-state discovery and characterization. *Nature methods*, 9(3) :215–216, 2012.
- Ernst Jason and Kellis Manolis. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology*, 33(4) :364, 2015.
- Farquhar Kevin S ; Charlebois Daniel A ; Szenk Mariola ; Cohen Joseph ; Nevozhay Dmitry, and Balázsi Gábor. Role of network-mediated stochasticity in mammalian drug resistance. *Nature communications*, 10(1) :1–14, 2019.
- Feinberg Andrew P and Tycko Benjamin. The history of cancer epigenetics. *Nature Reviews Cancer*, 4(2) :143, 2004.
- Feinberg Andrew P ; Koldobskiy Michael A, and Göndör Anita. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nature Reviews Genetics*, 17(5) :284, 2016.
- Fortin Jean-Philippe and Hansen Kasper D. Reconstructing a/b compartments as revealed by hi-c using long-range correlations in epigenetic data. *Genome biology*, 16(1) :180, 2015.
- Franke Jacqueline ; Gehlen Jessica, and Ehrenhofer-Murray Ann E. Hypermethylation of yeast telomerase rna by the snrna and snorna methyltransferase tgs1. *J Cell Sci*, 121(21) : 3553–3560, 2008.
- Frías-Lasserre Daniel and Villagra Cristian A. The importance of ncRNAs as epigenetic mechanisms in phenotypic variation and organic evolution. *Frontiers in microbiology*, 8 : 2483, 2017.
- Gaiti Federico ; Chaligne Ronan ; Gu Hongcang ; Brand Ryan M ; Kothén-Hill Steven ; Schulman Rafael C ; Grigorev Kirill ; Risso Davide ; Kim Kyu-Tae ; Pastore Alessandro, and

- others . Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature*, 569(7757) :576–580, 2019.
- Galm Oliver ; Yoshikawa Hirohide ; Esteller Manel ; Osieka Rainhardt, and Herman James G. Socs-1, a negative regulator of cytokine signaling, is frequently silenced by methylation in multiple myeloma. *Blood, The Journal of the American Society of Hematology*, 101(7) : 2784–2788, 2003.
- Garrett-Bakelman Francine E ; Sheridan Caroline K ; Kacmarczyk Thadeous J ; Ishii Jennifer ; Betel Doron ; Alonso Alicia ; Mason Christopher E ; Figueroa Maria E, and Melnick Ari M. Enhanced reduced representation bisulfite sequencing for assessment of dna methylation at base pair resolution. *JoVE (Journal of Visualized Experiments)*, (96) :e52246, 2015.
- Goel Neelam ; Karir Priya, and Garg Vivek Kumar. Role of dna methylation in human age prediction. *Mechanisms of ageing and development*, 166 :33–41, 2017.
- Goldsmith Scott R ; Fiala Mark A ; O’Neal Julie ; Souroullas George P ; Toama Wael ; Vij Ravi, and Schroeder Mark A. Ezh2 overexpression in multiple myeloma : Prognostic value, correlation with clinical characteristics, and possible mechanisms. *Clinical Lymphoma Myeloma and Leukemia*, 19(11) :744–750, 2019.
- Gräff Johannes and Mansuy Isabelle M. Epigenetic codes in cognition and behaviour. *Behavioural brain research*, 192(1) :70–87, 2008.
- Grosselin Kevin ; Durand Adeline ; Marsolier Justine ; Poitou Adeline ; Marangoni Elisabetta ; Nemati Fariba ; Dahmani Ahmed ; Lameiras Sonia ; Reyat Fabien ; Frenoy Olivia, and others . High-throughput single-cell chip-seq identifies heterogeneity of chromatin states in breast cancer. *Nature genetics*, 51(6) :1060–1066, 2019.
- Guccione Ernesto and Richard Stephane. The regulation, functions and clinical relevance of arginine methylation. *Nat. Rev. Mol. Cell Biol*, 20 :642–657, 2019.
- Hahn Maria A ; Li Arthur X ; Wu Xiwei ; Yang Richard ; Drew David A ; Rosenberg Daniel W, and Pfeifer Gerd P. Loss of the polycomb mark from bivalent promoters leads to activation of cancer-promoting genes in colorectal tumors. *Cancer research*, 74(13) :3617–3629, 2014.
- Hannum Gregory ; Guinney Justin ; Zhao Ling ; Zhang Li ; Hughes Guy ; Sadda Srinivas ; Klotzle Brandy ; Bibikova Marina ; Fan Jian-Bing ; Gao Yuan, and others . Genome-wide methylation profiles reveal quantitative views of human aging rates. *Molecular cell*, 49(2) : 359–367, 2013.
- Hansen Kasper D ; Langmead Benjamin, and Irizarry Rafael A. Bsmooth : from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome biology*, 13(10) : R83, 2012.

- Hansen Kasper Daniel ; Timp Winston ; Bravo Héctor Corrada ; Sabunciyan Sarven ; Langmead Benjamin ; McDonald Oliver G ; Wen Bo ; Wu Hao ; Liu Yun ; Diep Dinh, and others . Increased methylation variation in epigenetic domains across cancer types. *Nature genetics*, 43(8) :768, 2011.
- Hansen R Scott ; Thomas Sean ; Sandstrom Richard ; Canfield Theresa K ; Thurman Robert E ; Weaver Molly ; Dorschner Michael O ; Gartler Stanley M, and Stamatoyannopoulos John A. Sequencing newly replicated dna reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences*, 107(1) :139–144, 2010.
- Hashimoto Hideharu ; Liu Yiwei ; Upadhyay Anup K ; Chang Yanqi ; Howerton Shelley B ; Vertino Paula M ; Zhang Xing, and Cheng Xiaodong. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic acids research*, 40(11) : 4841–4849, 2012.
- Hawkins R David ; Hon Gary C ; Lee Leonard K ; Ngo QueMinh ; Lister Ryan ; Pelizzola Mattia ; Edsall Lee E ; Kuan Samantha ; Luu Ying ; Klugman Sarit, and others . Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell stem cell*, 6 (5) :479–491, 2010.
- Heinz Sven ; Benner Christopher ; Spann Nathanael ; Bertolino Eric ; Lin Yin C ; Laslo Peter ; Cheng Jason X ; Murre Cornelis ; Singh Harinder, and Glass Christopher K. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Molecular cell*, 38(4) :576–589, 2010.
- Heyn Holger ; Li Ning ; Ferreira Humberto J ; Moran Sebastian ; Pisano David G ; Gomez Antonio ; Diez Javier ; Sanchez-Mut Jose V ; Setien Fernando ; Carmona F Javier, and others . Distinct dna methylomes of newborns and centenarians. *Proceedings of the National Academy of Sciences*, 109(26) :10522–10527, 2012.
- Hochberg Yosef and Benjamini Yoav. More powerful procedures for multiple significance testing. *Statistics in medicine*, 9(7) :811–818, 1990.
- Holien Toril and Sundan Anders. Oncogene addiction to c-myc in myeloma cells. *Oncotarget*, 3(8) :739, 2012.
- Horvath Steve. Dna methylation age of human tissues and cell types. *Genome biology*, 14 (10) :3156, 2013.
- Hovestadt Volker ; Jones David TW ; Picelli Simone ; Wang Wei ; Kool Marcel ; Northcott Paul A ; Sultan Marc ; Stachurski Katharina ; Ryzhova Marina ; Warnatz Hans-Jörg, and others . Decoding the regulatory landscape of medulloblastoma using dna methylation sequencing. *Nature*, 510(7506) :537–541, 2014.
- Huang Da Wei ; Sherman Brad T, and Lempicki Richard A. Bioinformatics enrichment tools : paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research*, 37(1) :1–13, 2009.

- Huh Iksoo ; Zeng Jia ; Park Taesung, and Soojin V Yi. Dna methylation and transcriptional noise. *Epigenetics & chromatin*, 6(1) :9, 2013.
- Imakaev Maxim ; Fudenberg Geoffrey ; McCord Rachel Patton ; Naumova Natalia ; Goloborodko Anton ; Lajoie Bryan R ; Dekker Job, and Mirny Leonid A. Iterative correction of hi-c data reveals hallmarks of chromosome organization. *Nature methods*, 9(10) :999, 2012.
- Irizarry Rafael A ; Ladd-Acosta Christine ; Wen Bo ; Wu Zhijin ; Montano Carolina ; Onyango Patrick ; Cui Hengmi ; Gabo Kevin ; Rongione Michael ; Webster Maree, and others . The human colon cancer methylome shows similar hypo-and hypermethylation at conserved tissue-specific cpg island shores. *Nature genetics*, 41(2) :178, 2009a.
- Irizarry Rafael A ; Wang Chi ; Zhou Yun, and Speed Terence P. Gene set enrichment analysis made simple. *Statistical methods in medical research*, 18(6) :565–575, 2009b.
- Issa Jean-Pierre. Cpg island methylator phenotype in cancer. *Nature Reviews Cancer*, 4(12) : 988–993, 2004.
- Jang Hyun Sik ; Shin Woo Jung ; Lee Jeong Eon, and Do Jeong Tae. Cpg and non-cpg methylation in epigenetic gene regulation and brain function. *Genes*, 8(6) :148, 2017.
- Jenkinson Garrett ; Pujadas Elisabet ; Goutsias John, and Feinberg Andrew P. Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nature genetics*, 49(5) :719, 2017.
- Jeong Mira ; Sun Deqiang ; Luo Min ; Huang Yun ; Challen Grant A ; Rodriguez Benjamin ; Zhang Xiaotian ; Chavez Lukas ; Wang Hui ; Hannah Rebecca, and others . Large conserved domains of low dna methylation maintained by dnmt3a. *Nature genetics*, 46(1) :17, 2014.
- Jin Yi ; Chen Kenian ; De Paepe Ayla ; Hellqvist Eva ; Krstic Aleksandra D ; Metang Lauren ; Gustafsson Charlotte ; Davis Richard E ; Levy Yair M ; Surapaneni Rakesh, and others . Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood, The Journal of the American Society of Hematology*, 131(19) :2138–2150, 2018.
- Jung Hyunchul ; Kim Hong Sook ; Kim Jeong Yeon ; Sun Jong-Mu ; Ahn Jin Seok ; Ahn Myung-Ju ; Park Keunchil ; Esteller Manel ; Lee Se-Hoon, and Choi Jung Kyoon. Dna methylation loss promotes immune evasion of tumours with high mutation and copy number load. *Nature communications*, 10(1) :1–12, 2019.
- Keats Jonathan J ; Chesi Marta ; Egan Jan B ; Garbitt Victoria M ; Palmer Stephen E ; Braggio Esteban ; Van Wier Scott ; Blackburn Patrick R ; Baker Angela S ; Dispenzieri Angela, and others . Clonal competition with alternating dominance in multiple myeloma. *Blood, The Journal of the American Society of Hematology*, 120(5) :1067–1076, 2012.

- Kernaleguen Magali ; Daviaud Christian ; Shen Yimin ; Bonnet Eric ; Renault Victor ; Deleuze Jean-François ; Mauger Florence, and Tost Jörg. Whole-genome bisulfite sequencing for the analysis of genome-wide dna methylation and hydroxymethylation patterns at single-nucleotide resolution. In *Epigenome Editing*, pages 311–349. Springer, 2018.
- Kim Young Zoon. Altered histone modifications in gliomas. *Brain tumor research and treatment*, 2(1) :7–21, 2014.
- Kinkley Sarah ; Helmuth Johannes ; Polansky Julia K ; Dunkel Ilona ; Gasparoni Gilles ; Fröhler Sebastian ; Chen Wei ; Walter Jörn ; Hamann Alf, and Chung Ho-Ryun. rechip-seq reveals widespread bivalency of h3k4me3 and h3k27me3 in cd4+ memory t cells. *Nature communications*, 7(1) :1–13, 2016.
- Klughammer Johanna ; Kiesel Barbara ; Roetzer Thomas ; Fortelny Nikolaus ; Nemc Amelie ; Nenning Karl-Heinz ; Furtner Julia ; Sheffield Nathan C ; Datlinger Paul ; Peter Nadine, and others . The dna methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. *Nature medicine*, 24(10) :1611–1624, 2018.
- Korthauer Keegan ; Chakraborty Sutirtha ; Benjamini Yuval, and Irizarry Rafael A. Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics*, 20(3) :367–383, 2019.
- Kouzarides Tony. Chromatin modifications and their function. *Cell*, 128(4) :693–705, 2007.
- Krueger Felix. Trim galore : a wrapper tool around cutadapt and fastqc to consistently apply quality and adapter trimming to fastq files, with some extra functionality for mspi-digested rrbs-type (reduced representation bisulfite-seq) libraries. URL http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/. (Date of access : 28/04/2016), 2012.
- Krueger Felix and Andrews Simon R. Bismark : a flexible aligner and methylation caller for bisulfite-seq applications. *bioinformatics*, 27(11) :1571–1572, 2011.
- Kuehl W Michael and Bergsagel P Leif. Multiple myeloma : evolving genetic events and host interactions. *Nature Reviews Cancer*, 2(3) :175–187, 2002.
- Kulis Marta ; Merkel Angelika ; Heath Simon ; Queirós Ana C ; Schuyler Ronald P ; Castellano Giancarlo ; Beekman Renée ; Raineri Emanuele ; Esteve Anna ; Clot Guillem, and others . Whole-genome fingerprint of the dna methylome during human b cell differentiation. *Nature genetics*, 47(7) :746, 2015.
- Kumar Shaji K. ; Rajkumar Vincent ; Kyle Robert A. ; Duin Mark van ; Sonneveld Pieter ; Mateos María-Victoria ; Gay Francesca, and Anderson Kenneth C. Multiple myeloma. *Nature Reviews Disease Primers*, 3(1) :1–20, 2017.
- Lai Anne Y ; Mav Deepak ; Shah Ruchir ; Grimm Sara A ; Phadke Dhiral ; Hatzi Katerina ; Melnick Ari ; Geigerman Cissy ; Sobol Steve E ; Jaye David L, and others . Dna methylation

- profiling in human b cells reveals immune regulatory elements and epigenetic plasticity at alu elements during b-cell activation. *Genome research*, 23(12) :2030–2041, 2013.
- Landan Gilad ; Cohen Netta Mendelson ; Mukamel Zohar ; Bar Amir ; Molchadsky Alina ; Brosh Ran ; Horn-Saban Shirley ; Zalcenstein Daniela Amann ; Goldfinger Naomi ; Zundeleovich Adi, and others . Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nature genetics*, 44(11) :1207, 2012.
- Landau Dan A ; Clement Kendell ; Ziller Michael J ; Boyle Patrick ; Fan Jean ; Gu Hongcang ; Stevenson Kristen ; Sougnez Carrie ; Wang Lili ; Li Shuqiang, and others . Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer cell*, 26(6) :813–825, 2014.
- Langmead Ben and Salzberg Steven L. Fast gapped-read alignment with bowtie 2. *Nature methods*, 9(4) :357, 2012.
- Lee Jeannie T. Epigenetic regulation by long noncoding rnas. *Science*, 338(6113) :1435–1439, 2012.
- Li Sheng ; Garrett-Bakelman Francine ; Perl Alexander E ; Luger Selina M ; Zhang Chao ; To Bik L ; Lewis Ian D ; Brown Anna L ; D’Andrea Richard J ; Ross M Elizabeth, and others . Dynamic evolution of clonal epialleles revealed by methclone. *Genome biology*, 15(9) :472, 2014.
- Lieberman-Aiden Erez ; Van Berkum Nynke L ; Williams Louise ; Imakaev Maxim ; Ragoczy Tobias ; Telling Agnes ; Amit Ido ; Lajoie Bryan R ; Sabo Peter J ; Dorschner Michael O, and others . Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950) :289–293, 2009.
- Lister Ryan ; Pelizzola Mattia ; Downen Robert H ; Hawkins R David ; Hon Gary ; Tonti-Filippini Julian ; Nery Joseph R ; Lee Leonard ; Ye Zhen ; Ngo Que-Minh, and others . Human dna methylomes at base resolution show widespread epigenomic differences. *nature*, 462(7271) : 315–322, 2009.
- Liu Pengyu ; Cao Wanlu ; Ma Buyun ; Li Meng ; Chen Kan ; Sideras Kostandinos ; Duitman Jan-Willem ; Sprengers Dave ; Khe Tran TC ; Ijzermans Jan NM, and others . Action and clinical significance of ccaat/enhancer-binding protein delta in hepatocellular carcinoma. *Carcinogenesis*, 40(1) :155–163, 2019.
- Lohr Jens G ; Stojanov Petar ; Carter Scott L ; Cruz-Gordillo Peter ; Lawrence Michael S ; Auclair Daniel ; Sougnez Carrie ; Knoechel Birgit ; Gould Joshua ; Saksena Gordon, and others . Widespread genetic heterogeneity in multiple myeloma : implications for targeted therapy. *Cancer cell*, 25(1) :91–101, 2014.
- Lorincz Matthew C and Schübeler Dirk. Evidence for converging dna methylation pathways in placenta and cancer. *Developmental cell*, 43(3) :257–258, 2017.

- Love Michael I ; Huber Wolfgang, and Anders Simon. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology*, 15(12) :550, 2014.
- Madrigano Jaime ; Baccarelli Andrea A ; Mittleman Murray A ; Sparrow David ; Vokonas Pantel S ; Tarantini Letizia, and Schwartz Joel. Aging and epigenetics : longitudinal changes in gene-specific dna methylation. *Epigenetics*, 7(1) :63–70, 2012.
- Magrangeas Florence ; Avet-Loiseau Hervé ; Gouraud Wilfried ; Lodé Laurence ; Decaux Olivier ; Godmer Pascal ; Garderet Laurent ; Voillat Laurent ; Facon Thierry ; Stoppa Anne Marie, and others . Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia*, 27(2) :473–481, 2013.
- Manier Salomon ; Salem Karma Z ; Park Jihye ; Landau Dan A ; Getz Gad, and Ghobrial Irene M. Genomic complexity of multiple myeloma and its clinical implications. *Nature reviews Clinical oncology*, 14(2) :100, 2017.
- Marchal Claire ; Sasaki Takayo ; Vera Daniel ; Wilson Korey ; Sima Jiao ; Rivera-Mulia Juan-Carlos ; Garcia Claudia Trevilla ; Nogues Coraline ; Nafie Ebtesam, and Gilbert David M. Repli-seq : genome-wide analysis of replication timing by next-generation sequencing. *bioRxiv*, page 104653, 2017.
- Martinez-Garcia Eva ; Popovic Relja ; Min Dong-Joon ; Sweet Steve MM ; Thomas Paul M ; Zamdborg Leonid ; Heffner Aaron ; Will Christine ; Lamy Laurence ; Staudt Louis M, and others . The mmset histone methyl transferase switches global histone methylation and alters gene expression in t (4 ; 14) multiple myeloma cells. *Blood, The Journal of the American Society of Hematology*, 117(1) :211–220, 2011.
- Mas Glòria ; Blanco Enrique ; Ballaré Cecilia ; Sansó Miriam ; Spill Yannick G ; Hu Deqing ; Aoi Yuki ; Le Dily François ; Shilatifard Ali ; Marti-Renom Marc A, and others . Promoter bivalency favors an open chromatin architecture in embryonic stem cells. *Nature genetics*, 50(10) :1452–1462, 2018.
- Mateos María-Victoria and San Miguel Jesús F. How should we treat newly diagnosed multiple myeloma patients? *Hematology 2013, the American Society of Hematology Education Program Book*, 2013(1) :488–495, 2013.
- Maunakea Alike K ; Nagarajan Raman P ; Bilenky Mikhail ; Ballinger Tracy J ; D’Souza Cletus ; Fouse Shaun D ; Johnson Brett E ; Hong Chibo ; Nielsen Cydney ; Zhao Yongjun, and others . Conserved role of intragenic dna methylation in regulating alternative promoters. *Nature*, 466(7303) :253–257, 2010.
- McLaughlin Katy A ; Flyamer Ilya M ; Thomson John P ; Mjoseng Heidi K ; Shukla Ruchi ; Williamson Iain ; Grimes Graeme R ; Illingworth Robert S ; Adams Ian R ; Pennings Sari, and others . Dna methylation directs polycomb-dependent 3d genome re-organisation in naive pluripotency. *bioRxiv*, page 527309, 2019.

- McLean Cory Y ; Bristor Dave ; Hiller Michael ; Clarke Shoa L ; Schaar Bruce T ; Lowe Craig B ; Wenger Aaron M, and Bejerano Gill. Great improves functional interpretation of cis-regulatory regions. *Nature biotechnology*, 28(5) :495, 2010.
- Meas Rithy and Mao Peng. Histone ubiquitylation and its roles in transcription and dna damage response. *DNA repair*, 36 :36–42, 2015.
- Mirabella F ; Wu P ; Wardell CP ; Kaiser MF ; Walker BA ; Johnson DC, and Morgan GJ. Mmset is the key molecular target in t (4; 14) myeloma. *Blood cancer journal*, 3(5) : e114–e114, 2013.
- Mittelman Moshe. The implications of anemia in multiple myeloma. *Clinical lymphoma*, 4 : S23–S29, 2003.
- Miura Hisashi ; Takahashi Saori ; Poonperm Rawin ; Tanigawa Akie ; Takebayashi Shin-ichiro, and Hiratani Ichiro. Single-cell dna replication profiling identifies spatiotemporal developmental dynamics of chromosome organization. *Nature genetics*, 51(9) :1356–1368, 2019.
- Miyazaki Kanji and Suzuki Kenshi. Cd56 for multiple myeloma : Lack of cd56 may be associated with worse prognosis. *Acta haematologica*, 140(1) :40–41, 2018.
- Morgan Gareth J ; Gregory Walter M ; Davies Faith E ; Bell Sue E ; Szubert Alexander J ; Brown Julia M ; Coy Nuria N ; Cook Gordon ; Russell Nigel H ; Rudin Claudius, and others . The role of maintenance thalidomide therapy in multiple myeloma : Mrc myeloma ix results and meta-analysis. *Blood, The Journal of the American Society of Hematology*, 119 (1) :7–15, 2012a.
- Morgan Gareth J ; Walker Brian A, and Davies Faith E. The genetic architecture of multiple myeloma. *Nature Reviews Cancer*, 12(5) :335–348, 2012b.
- Mortazavi Ali ; Williams Brian A ; McCue Kenneth ; Schaeffer Lorian, and Wold Barbara. Mapping and quantifying mammalian transcriptomes by rna-seq. *Nature methods*, 5(7) : 621, 2008.
- Nagano Takashi ; Lubling Yaniv ; Stevens Tim J ; Schoenfelder Stefan ; Yaffe Eitan ; Dean Wendy ; Laue Ernest D ; Tanay Amos, and Fraser Peter. Single-cell hi-c reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469) :59–64, 2013.
- Nasu Kaei ; Kawano Yukie ; Tsukamoto Yoshiyuki ; Takano Masayuki ; Takai Noriyuki ; Li Haili ; Furukawa Yuichi ; Abe Wakana ; Moriyama Masatsugu, and Narahara Hisashi. Aberrant dna methylation status of endometriosis : epigenetics as the pathogenesis, biomarker and therapeutic target. *Journal of Obstetrics and Gynaecology Research*, 37(7) :683–695, 2011.
- Ng Margaret HL ; To KW ; Lo KW ; Chan S ; Tsang KS ; Cheng SH, and Ng HK. Frequent death-associated protein kinase promoter hypermethylation in multiple myeloma. *Clinical cancer research*, 7(6) :1724–1729, 2001.

- Ng Margaret HL ; Lau KM ; Wong WS ; To KW ; Cheng SH ; Tsang KS ; Chan Natalie PH ; Kho Bonnie CS ; Lo KW ; Tong Joanna HM, and others . Alterations of ras signalling in chinese multiple myeloma patients : absent braf and rare ras mutations, but frequent inactivation of rassf1a by transcriptional silencing or expression of a non-functional variant transcript. *British journal of haematology*, 123(4) :637–645, 2003.
- Ng MHL ; Chung YF ; Lo KW ; Wickham NWR ; Lee JCK, and Huang DP. Frequent hypermethylation of p16 and p15 genes in multiple myeloma. *Blood, The Journal of the American Society of Hematology*, 89(7) :2500–2506, 1997.
- Nojima Masanori ; Maruyama Reo ; Yasui Hiroshi ; Suzuki Hiromu ; Maruyama Yumiko ; Tarasawa Isao ; Sasaki Yasushi ; Asaoku Hideki ; Sakai Hajime ; Hayashi Toshiaki, and others . Genomic screening for genes silenced by dna methylation revealed an association between rasd1 inactivation and dexamethasone resistance in multiple myeloma. *Clinical Cancer Research*, 15(13) :4356–4364, 2009.
- Novakovic Boris and Saffery Richard. Placental pseudo-malignancy from a dna methylation perspective : unanswered questions and future directions. *Frontiers in genetics*, 4 :285, 2013.
- Nowell Peter C. The clonal evolution of tumor cell populations. *Science*, 194(4260) :23–28, 1976.
- Oakes Christopher C ; Claus Rainer ; Gu Lei ; Assenov Yassen ; Hüllein Jennifer ; Zucknick Manuela ; Bieg Matthias ; Brocks David ; Bogatyrova Olga ; Schmidt Christopher R, and others . Evolution of dna methylation is linked to genetic aberrations in chronic lymphocytic leukemia. *Cancer discovery*, 4(3) :348–361, 2014.
- Okano Masaki ; Bell Daphne W ; Haber Daniel A, and Li En. Dna methyltransferases dnmt3a and dnmt3b are essential for de novo methylation and mammalian development. *Cell*, 99(3) :247–257, 1999.
- Otani Junji ; Kimura Hironobu ; Sharif Jafar ; Endo Takaho A ; Mishima Yuichi ; Kawakami Toru ; Koseki Haruhiko ; Shirakawa Masahiro ; Suetake Isao, and Tajima Shoji. Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PloS one*, 8(12), 2013.
- Ozata Deniz M ; Gainetdinov Ildar ; Zoch Ansgar ; O’Carroll Dónal, and Zamore Phillip D. Piwi-interacting rnas : small rnas with big functions. *Nature Reviews Genetics*, 20(2) : 89–108, 2019.
- Pan Heng ; Jiang Yanwen ; Boi Michela ; Tabbò Fabrizio ; Redmond David ; Nie Kui ; Ladetto Marco ; Chiappella Annalisa ; Cerchietti Leandro ; Shaknovich Rita, and others . Epigenomic evolution in diffuse large b-cell lymphomas. *Nature communications*, 6 :6921, 2015.

- Pastore Alessandro ; Gaiti Federico ; Lu Sydney X ; Brand Ryan M ; Kulm Scott ; Chaligne Ronan ; Gu Hongcang ; Huang Kevin Y ; Stamenova Elena K ; Béguelin Wendy, and others . Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in cll. *Nature communications*, 10(1) :1–11, 2019.
- Pawlyn C ; Bright MD ; Buros AF ; Stein CK ; Walters Z ; Aronson LI ; Mirabella F ; Jones JR ; Kaiser MF ; Walker BA, and others . Overexpression of ezh2 in multiple myeloma is associated with poor prognosis and dysregulation of cell cycle control. *Blood cancer journal*, 7(3) :e549–e549, 2017.
- Popovic Relja ; Martinez-Garcia Eva ; Giannopoulou Eugenia G ; Zhang Quanwei ; Zhang Qingyang ; Ezponda Teresa ; Shah Mrinal Y ; Zheng Yupeng ; Will Christine M ; Small Eliza C, and others . Histone methyltransferase mmset/nsd2 alters ezh2 binding and reprograms the myeloma epigenome through global and focal changes in h3k36 and h3k27 methylation. *PLoS genetics*, 10(9), 2014.
- Portela Anna and Esteller Manel. Epigenetic modifications and human disease. *Nature biotechnology*, 28(10) :1057, 2010.
- Quek Kelly ; Li Jun ; Estecio Marcos ; Zhang Jiexin ; Fujimoto Junya ; Roarty Emily ; Little Latasha ; Chow Chi-Wan ; Song Xingzhi ; Behrens Carmen, and others . Dna methylation intratumor heterogeneity in localized lung adenocarcinomas. *Oncotarget*, 8(13) :21994, 2017.
- Quinlan Aaron R and Hall Ira M. Bedtools : a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6) :841–842, 2010.
- Rakyan Vardhman K ; Down Thomas A ; Maslau Siarhei ; Andrew Toby ; Yang Tsun-Po ; Beyan Huriya ; Whittaker Pamela ; McCann Owen T ; Finer Sarah ; Valdes Ana M, and others . Human aging-associated dna hypermethylation occurs preferentially at bivalent chromatin domains. *Genome research*, 20(4) :434–439, 2010.
- Ramírez Fidel ; Bhardwaj Vivek ; Arrigoni Laura ; Lam Kin Chung ; Grüning Björn A ; Villaveces José ; Habermann Bianca ; Akhtar Asifa, and Manke Thomas. High-resolution tads reveal dna sequences underlying genome organization in flies. *Nature communications*, 9(1) :1–15, 2018.
- Reddington James P ; Perricone Sara M ; Nestor Colm E ; Reichmann Judith ; Youngson Neil A ; Suzuki Masako ; Reinhardt Diana ; Dunican Donncha S ; Prendergast James G ; Mjoseng Heidi, and others . Redistribution of h3k27me3 upon dna hypomethylation results in de-repression of polycomb target genes. *Genome biology*, 14(3) :R25, 2013.
- Richards Rebecca ; Patel Jayshree ; Stevenson Kate, and Harbison SallyAnn. Evaluation of massively parallel sequencing for forensic dna methylation profiling. *Electrophoresis*, 39(21) :2798–2805, 2018.

- Rossetto Dorine ; Avvakumov Nikita, and Côté Jacques. Histone phosphorylation : a chromatin modification involved in diverse nuclear events. *Epigenetics*, 7(10) :1098–1108, 2012.
- Rotem Assaf ; Ram Oren ; Shores Noam ; Sperling Ralph A ; Goren Alon ; Weitz David A, and Bernstein Bradley E. Single-cell chip-seq reveals cell subpopulations defined by chromatin state. *Nature biotechnology*, 33(11) :1165–1172, 2015.
- Ryba Tyrone ; Battaglia Dana ; Chang Bill H ; Shirley James W ; Buckley Quinton ; Pope Benjamin D ; Devidas Meenakshi ; Druker Brian J, and Gilbert David M. Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. *Genome research*, 22(10) :1833–1844, 2012.
- Ryu Daeun ; Kim Seok Jin ; Hong Yourae ; Jo Areum ; Kim Nayoung ; Kim Hee-Jin ; Lee Hae-Ock ; Kim Kihyun, and Park Woong-Yang. Alterations in the transcriptional programs of myeloma cells and the microenvironment during extramedullary progression affect proliferation and immune evasion. *Clinical Cancer Research*, 26(4) :935–944, 2020.
- Salhab Abdulrahman ; Nordström Karl ; Gasparoni Gilles ; Kattler Kathrin ; Ebert Peter ; Ramirez Fidel ; Arrigoni Laura ; Müller Fabian ; Polansky Julia K ; Cadenas Cristina, and others . A comprehensive analysis of 195 dna methylomes reveals shared and cell-specific features of partially methylated domains. *Genome biology*, 19(1) :150, 2018.
- Sanz Lionel A ; Kota Satya K, and Feil Robert. Genome-wide dna demethylation in mammals. *Genome biology*, 11(3) :110, 2010.
- Schmidt-Hieber Martin ; Perez-Andres Martin ; Paiva Bruno ; Flores-Montero Juan ; Perez Jose J ; Gutierrez Norma C ; Vidriales Maria-Belen ; Matarraz Sergio ; San Miguel Jesus F, and Orfao Alberto. Cd117 expression in gammopathies is associated with an altered maturation of the myeloid and lymphoid hematopoietic cell compartments and favorable disease features. *Haematologica*, 96(2) :328–332, 2011.
- Schroeder Diane I ; Blair John D ; Lott Paul ; Yu Hung On Ken ; Hong Danna ; Crary Florence ; Ashwood Paul ; Walker Cheryl ; Korf Ian ; Robinson Wendy P, and others . The human placenta methylome. *Proceedings of the national academy of sciences*, 110(15) :6037–6042, 2013.
- Servant Nicolas ; Lajoie Bryan R ; Nora Elphège P ; Giorgetti Luca ; Chen Chong-Jian ; Heard Edith ; Dekker Job, and Barillot Emmanuel. Hic : exploration of high-throughput ‘c’ experiments. *Bioinformatics*, 28(21) :2843–2844, 2012.
- Servant Nicolas ; Varoquaux Nelle ; Lajoie Bryan R ; Viara Eric ; Chen Chong-Jian ; Vert Jean-Philippe ; Heard Edith ; Dekker Job, and Barillot Emmanuel. Hic-pro : an optimized and flexible pipeline for hi-c data processing. *Genome biology*, 16(1) :259, 2015.
- Sherman Brad T ; Lempicki Richard A, and others . Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature protocols*, 4(1) :44, 2009.

- Shou Yaping ; Martelli Maria L ; Gabrea Ana ; Qi Ying ; Brents Leslie A ; Roschke Anna ; Dewald Gordon ; Kirsch Ilan R ; Bergsagel P Leif, and Kuehl W Michael. Diverse karyotypic abnormalities of the c-myc locus associated with c-myc dysregulation and tumor progression in multiple myeloma. *Proceedings of the National Academy of Sciences*, 97(1) :228–233, 2000.
- Siegfried Zahava and Cedar Howard. Dna methylation : a molecular lock. *Current Biology*, 7 (5) :R305–R307, 1997.
- Smallwood Sébastien A ; Lee Heather J ; Angermueller Christof ; Krueger Felix ; Saadeh Heba ; Peat Julian ; Andrews Simon R ; Stegle Oliver ; Reik Wolf, and Kelsey Gavin. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature methods*, 11 (8) :817–820, 2014.
- Šorm F ; Piskala A ; Čihák A, and Veselý J. 5-azacytidine, a new, highly effective cancerostatic. *Experientia*, 20(4) :202–203, 1964.
- Tang Fuchou ; Barbacioru Catalin ; Wang Yangzhou ; Nordman Ellen ; Lee Clarence ; Xu Nanlan ; Wang Xiaohui ; Bodeau John ; Tuch Brian B ; Siddiqui Asim, and others . mrna-seq whole-transcriptome analysis of a single cell. *Nature methods*, 6(5) :377–382, 2009.
- Tao Yong ; Kang Byunghak ; Petkovich Daniel A ; Bhandari Yuba R ; In Julie ; Stein-O’Brien Genevieve ; Kong Xiangqian ; Xie Wenbing ; Zachos Nicholas ; Maegawa Shinji, and others . Aging-like spontaneous epigenetic silencing facilitates wnt activation, stemness, and brafv600e-induced tumorigenesis. *Cancer cell*, 35(2) :315–328, 2019.
- Thu Kelsie L ; Vucic Emily A ; Kennett Jennifer Y ; Heryet Cameron ; Brown Carolyn J ; Lam Wan L, and Wilson Ian M. Methylated dna immunoprecipitation. *JoVE (Journal of Visualized Experiments)*, (23) :e935, 2009.
- Tremblay-LeMay Rosemarie ; Rastgoo Nasrin ; Pourabdollah Maryam, and Chang Hong. Ezh2 as a therapeutic target for multiple myeloma and other haematological malignancies. *Biomarker research*, 6(1) :34, 2018.
- Turner Bryan M. Reading signals on the nucleosome with a new nomenclature for modified histones. *Nature structural & molecular biology*, 12(2) :110–112, 2005.
- van Andel Harmen ; Kocemba Kinga A ; Spaargaren Marcel, and Pals Steven T. Aberrant wnt signaling in multiple myeloma : molecular mechanisms and targeting options. *Leukemia*, 33 (5) :1063–1075, 2019.
- Vaz Michelle ; Hwang Stephen Y ; Kagiampakis Ioannis ; Phallen Jillian ; Patil Ashwini ; O’Hagan Heather M ; Murphy Lauren ; Zahnow Cynthia A ; Gabrielson Edward ; Velculescu Victor E, and others . Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by kras mutations. *Cancer cell*, 32(3) :360–376, 2017.

- Vertino Paula M ; Yen Ray-Whay ; Gao Jin, and Baylin Stephen B. De novo methylation of cpg island sequences in human fibroblasts overexpressing dna (cytosine-5-)-methyltransferase. *Molecular and cellular biology*, 16(8) :4555–4565, 1996.
- Vucic Emily A ; Wilson Ian M ; Campbell Jennifer M, and Lam Wan L. Methylation analysis by dna immunoprecipitation (medip). In *Microarray Analysis of the Physical Genome*, pages 141–153. Springer, 2009.
- Waddington Conrad H. The epigenotype. *International journal of epidemiology*, 41(1) :10–13, 2012.
- Walker Brian A ; Wardell Christopher P ; Chiecchio Laura ; Smith Emma M ; Boyd Kevin D ; Neri Antonino ; Davies Faith E ; Ross Fiona M, and Morgan Gareth J. Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood, The Journal of the American Society of Hematology*, 117(2) :553–562, 2011.
- Walker Brian A ; Wardell Christopher P ; Melchor Lorenzo ; Brioli Annamaria ; Johnson David C ; Kaiser Martin F ; Mirabella Fabio ; Lopez-Corral Lucia ; Humphray Sean ; Murray Lisa, and others . Intracloal heterogeneity is a critical early event in the development of myeloma and precedes the development of clinical symptoms. *Leukemia*, 28(2) :384–390, 2014.
- Walker Brian A ; Boyle Eileen M ; Wardell Christopher P ; Murison Alex ; Begum Dil B ; Dahir Nasrin B ; Proszek Paula Z ; Johnson David C ; Kaiser Martin F ; Melchor Lorenzo, and others . Mutational spectrum, copy number changes, and outcome : results of a sequencing study of patients with newly diagnosed myeloma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*, 33(33) :3911, 2015.
- Weber Tom Serge ; Jaehnert Irene ; Schichor Christian ; Or-Guil Michal, and Carneiro Jorge. Quantifying the length and variance of the eukaryotic cell cycle phases by a stochastic model and dual nucleoside pulse labelling. *PLoS computational biology*, 10(7), 2014.
- Wolff Joachim ; Bhardwaj Vivek ; Nothjunge Stephan ; Richard Gautier ; Renschler Gina ; Gilsbach Ralf ; Manke Thomas ; Backofen Rolf ; Ramírez Fidel, and Grüning Björn A. Galaxy hicexplorer : a web server for reproducible hi-c data analysis, quality control and visualization. *Nucleic acids research*, 46(W1) :W11–W16, 2018.
- Woodcock Christopher L and Ghosh Rajarshi P. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology*, 2(5) :a000596, 2010.
- Wu Pengze ; Li Tingting ; Li Ruifeng ; Jia Lumeng ; Zhu Ping ; Liu Yifang ; Chen Qing ; Tang Daiwei ; Yu Yuezhou, and Li Cheng. 3d genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nature communications*, 8(1) : 1–11, 2017.
- Wu Xiaoji and Zhang Yi. Tet-mediated active dna demethylation : mechanism, function and beyond. *Nature Reviews Genetics*, 18(9) :517, 2017.

- Xiao Fu-Hui ; Kong Qing-Peng ; Perry Benjamin, and He Yong-Han. Progress on the role of dna methylation in aging and longevity. *Briefings in functional genomics*, page elw009, 2016.
- Xie Wei ; Schultz Matthew D ; Lister Ryan ; Hou Zhonggang ; Rajagopal Nisha ; Ray Pradipta ; Whitaker John W ; Tian Shulan ; Hawkins R David ; Leung Danny, and others . Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*, 153(5) : 1134–1148, 2013.
- Xie Zhigang and Chng Wee Joo. Mmset : role and therapeutic opportunities in multiple myeloma. *BioMed research international*, 2014, 2014.
- Zhou Wanding ; Dinh Huy Q ; Ramjan Zachary ; Weisenberger Daniel J ; Nicolet Charles M ; Shen Hui ; Laird Peter W, and Berman Benjamin P. Dna methylation loss in late-replicating domains is linked to mitotic cell division. *Nature genetics*, 50(4) :591, 2018.

9 Annexes

Jennifer Derrien, Catherine Guérin-Charbonnel, Victor Gaborit, Loïc Campion, Magali Devic, Elise Douillard, Nathalie Roi, Hervé Avet-Loiseau, Olivier Decaux, Thierry Facon, Jan-Philipp Mallm, Roland Eils, Nikhil C. Munshi, Philippe Moreau, Carl Herrmann, Florence Magrangeas and Stéphane Minvielle. « The DNA methylation landscape of multiple myeloma shows extensive inter- and inpatient heterogeneity that fuels transcriptomic variability ».

RESEARCH

The DNA methylation landscape of multiple myeloma shows extensive inter- and inpatient heterogeneity that fuels transcriptomic variability

Jennifer Derrien¹, Catherine Guérin-Charbonnel^{1,2}, Victor Gaborit^{1,3}, Loïc Champion^{1,2}, Magali Devic^{1,4}, Elise Douillard^{1,4}, Nathalie Roi^{1,4}, Hervé Avet-Loiseau⁵, Olivier Decaux⁶, Thierry Facon⁷, Jan-Philipp Mallm⁸, Roland Eils^{9,10}, Nikhil C. Munshi¹¹, Philippe Moreau^{1,4}, Carl Herrmann⁹, Florence Magrangeas^{1,4†} and Stéphane Minvielle^{1,4*†}

* Correspondence:

stephane.minvielle@univ-nantes.fr

¹Université de Nantes, CNRS, INSERM, CRCINA, F-44000 Nantes, France

Full list of author information is available at the end of the article

† Co-senior

Abstract

Background: Cancer evolution depends on epigenetic and genetic diversity. Historically, in multiple myeloma (MM), subclonal diversity and tumor evolution have been investigated mostly from a genetic perspective.

Results: Here, we combined the notions of epipolymorphism and epiallele switching to analyze DNA methylation heterogeneity in MM patients. We show that MM is characterized by the continuous accumulation of stochastic methylation at the promoters of development-related genes. High entropy change is associated with poor outcomes and depends predominantly on partially methylated domains (PMDs). These PMDs, which represent the major source of inter- and inpatient DNA methylation heterogeneity in MM, are linked to other key epigenetic aberrations, such as CpG island (CGI)/transcription start site (TSS) hypermethylation and H3K27me3 redistribution as well as 3D organization alterations. In addition, transcriptome analysis revealed that intratumor methylation heterogeneity was associated with low-level expression and high variability.

Conclusion: We propose that disordered methylation in MM is responsible for high epigenetic and transcriptomic instability allowing tumor cells to adapt to environmental changes by tapping into a pool of evolutionary trajectories.

Keywords: multiple myeloma; disordered DNA methylation; epipolymorphism; epiallele switching; inter- and inpatient heterogeneity; transcriptomic variability

Background

Multiple myeloma (MM) is a neoplasm of plasma cells (PCs) with an incidence rate of approximately 5/100,000 in Europe. The median survival time of patients has improved substantially over the past decade. This is due to the establishment of high-dose therapy followed by autologous stem cell transplantation as a routine procedure, significant improvements in supportive care strategies, and the introduction and widespread use of drugs including immunomodulatory drugs, proteasome inhibitors, histone deacetylase inhibitors and monoclonal antibodies. Nevertheless, almost all patients ultimately relapse due to the emergence of more aggressive subpopulations of myeloma PCs resistant to therapeutic agents.

Several mechanisms have been suggested to explain the capacity of the subpopulations of myeloma cells within an individual to survive the pressure of frontline therapy and proliferate. These mechanisms include: the emergence of myeloma cells that achieve bortezomib resistance by decommitment from immunoglobulin synthesis [1] or by the derepression of growth factor receptors typically not associated with the plasma cell lineage [2], somatic mutations that emerge during disease progression involving key driver genes in MM such as the mono- or biallelic loss of TP53 [3, 4] or the biallelic loss of TRAF3 [5]. These mechanisms facilitate the expansion of proliferative subclonal populations. Genetic intratumor heterogeneity increases the evolutionary fitness potential of a rare subset of myeloma cells harboring a particular combination of molecular aberrations to survive when challenged by multiagent chemotherapy [4]. The disease evolves predominantly through a Darwinian process of clonal expansion, and the population of tumor PCs represents an admixture of competing genetic subclones [6, 5, 7, 8]. In most patients, the treatment pressure causes the profound reorganization and diversification of subclonal populations with complex dynamics of tumor evolution raising the possibility of biologically and clinically important cross-talk between subclones [9]. However, the small number of genetic alterations detected in MM and relevant in relapse mechanisms do not alone explain the profound phenotypic variability across patients [10, 4]. Beyond genetic diversity, other processes generate the intratumoral functional heterogeneity of cancer cells, including global epigenetic changes [11].

To address this question in MM, we conducted an analysis of the DNA methylation landscape of 47 myeloma samples from 26 patients. This study was motivated by the facts that (1) MM is characterized by extreme heterogeneity in median DNA methylation levels compared with normal plasma cells (NPCs) [12], (2) intratumor methylome variability is higher in various cancer types including other B-cell malignancies [13, 14, 15, 16], than in normal cells, and (3) genome-wide analysis of DNA methylation at single-base-pair resolution combined with analysis of the methylation states of neighboring CpGs (in the local sequence context) is now available [17].

Results

Intrapatient epigenetic heterogeneity arises from stochastic DNA methylation

To explore the global DNA methylation spectrum of MM patients, we compared whole-genome bisulfite sequencing (WGBS) datasets of myeloma cells from patients with those of controls (Additional file 1: Table S1). We found a marked global hypomethylation (Fig. 1a), as previously reported in MM and other cancers [12, 13, 15]. Individual CpG sites showed a predominantly bimodal pattern in MM patients (Fig. 1b) with methylation levels depending on the local density of CpG (Fig. 1c and Additional file 2: Table S2). Although MM displayed no evidence of aberrant DNA methylation at many genes (Additional file 3: Figure S1), we observed slight methylation gain at CpG islands (CGIs) and methylation loss at adjacent shores and shelves regions compared to NPCs (Fig. 1d), as illustrated in the genomic region of *DOC2B* (Fig. 1e).

These results show that the MM epigenome is characterized by global DNA hypomethylation and variable intermediate methylation states depending on the local

CpG content. We investigated whether these epigenetic changes could be a source of inpatient heterogeneity. To assess the degree of epigenetic variability within patients, we employed enhanced reduced representation bisulfite sequencing (eRRBS) technology on 21 MM patients because eRRBS provides an average depth of 50X per covered CpG and approximately 2.7 million CpGs per sample, with more than 10X sequencing coverage (Additional file 4: Table S3 and Additional file 5: Table S4), and is cost effective for a large number of samples (48 in our study, Additional file 1: Table S1 and Additional file 3: Figure S2). We then applied a computational method that analyzes DNA methylation modifications at a genomic locus defined as a group of four contiguous CpGs covered by single sequence reads [17]. We computed for each genomic locus the epipolymorphism, defined as a probability distribution of the 16 possible methylation patterns (epialleles): a high value of epipolymorphism indicates a stochastic process of DNA methylation [18] (Additional file 3: Figure S3). Almost all genomic loci were located in CGIs (85% in MM samples) and, to a lesser extent, in promoter transcription start sites (TSSs) and gene bodies (74% in MM samples, Additional file 3: Figure S4). We plotted epipolymorphism distribution as a function of methylation level for all loci in NPCs and MM samples (Fig. 2a). Remarkably, loci with modest methylation levels (5% - 25%) showed a higher degree of epipolymorphism in cancer cells than in control cells whereas at higher methylation levels, epipolymorphism distribution was similar in MM and NPCs, indicating enrichment of a fraction of loci-CGIs with a stochastic methylation state at the expense of a bimodal methylation profile in MM samples. In all cases, promoter CGI methylation gain was associated with an increase in epipolymorphism (Additional file 3: Figure S5).

To further characterize epiallele patterns of each sample, we stratified loci according to the frequency of fully methylated or fully unmethylated epialleles (Fig. 2b and Additional file 3: Figure S6). NPC loci displayed near-bimodal distribution, with a large number of homogeneous loci either fully unmethylated or fully methylated (Fig. 2b, top). Conversely, in MM samples, a large number of loci showing high epipolymorphism levels were along the axes, indicating marked enrichment for stochastic (i.e., neither fully methylated nor fully unmethylated) methylation patterns (Fig. 2b, bottom). Notably, this intrasample methylation heterogeneity was already present at the early stage of the disease, recognized as smoldering MM (SMM), and persisted at relapse (Additional file 3: Figure S6). A complementary approach that measures the proportion of discordant reads (PDR) [15] confirmed the increase in epigenetic heterogeneity in MM cells compared to NPCs. We found that the average PDR (i.e. genomic loci with a large number of reads that contain both unmethylated and methylated CpGs over the total number of reads) was significantly higher (p -value < 0.05 (DTK pairwise)) in MM samples, regardless of the disease stage, than in NPCs (Fig. 2c).

Overall, these results revealed that during disease initiation, growth and progression, malignant PCs accumulate randomness in DNA methylation at the expense of a more coherent methylation state, leading to a high degree of intratumoral epigenetic heterogeneity in MM patients.

Entropy changes in MM development

Since we have shown that stochastic DNA methylation gains lead to intratumor heterogeneity, we sought to analyze epiallelic dynamics during MM development and progression. We used an algorithm that allows the identification of epigenetic loci (eloci) that have a significant epiallele composition change (entropy change) between two states (e.g., normal vs cancer or diagnosis vs relapse). To compare global methylome changes across patient samples, the number of eloci was normalized according to the number of covered loci, referred to as eloci per million loci covered (EPM) (Additional file 3: Figure S7)[17]. We first examined the link between the degree of epiallelic shifting and outcome. For this purpose, patients were separated into two groups according to their EPM values. We found that MM patients with high EPM had reduced relapse-free survival (p-value = 0.02 (log-rank test); Fig. 3a). EPM was independent of age, sex and ploidy. Remarkably, the extent of entropy changes varied greatly between patients at diagnosis and persisted at relapse compared to NPCs (Fig. 3b, c). The genome-wide distribution of eloci showed a predominant location at CGIs (average 54% in MM vs NPCs). As a control, we determined eloci by comparing NPCs to NPCs (NPC eloci) and found, on average, 19% of eloci in CGIs (p-value (Wilcoxon) = 0.0018) and at promoter-TSS (average 15% in MM vs NPCs and average 3% in NPC eloci; p-value (Wilcoxon) = 0.0123) (Additional file 3: Figure S8). We observed a similar distribution of eloci in relapse samples compared to NPCs (p-value (Wilcoxon) = 0.75 (CGIs) and 0.82 (promoter-TSS), Additional file 3: Figure S8).

To gain insights into the biological functions of epigenetic perturbations, we performed gene ontology enrichment analysis for genes having eloci (MM diagnosis vs. NPCs) in promoter CGIs (Fig. 3d). Interestingly, associated genes showed marked enrichment for developmental regulators and transcription factor activity, while genes with promoter CGIs, including NPC eloci, were not enriched in any pathways or gene ontology terms. The identification of key signaling pathways rather than specific genes was not surprising considering the high DNA methylation variability across patients. Similar findings were reported by Easwaran *et al.* across different tumor types [19].

To further examine the impact of treatment on entropy changes, we assessed the epiallele shifting level during time to relapse by comparing diagnosis vs relapse paired samples (Fig. 3e). The degree of difference was highly variable from one patient to another; 24% of patients showed no epigenetic changes (EPM = 0) while 42% showed substantial changes (EPM > 1000) between diagnosis and relapse. We then focused our analysis on the evolution of extreme epiallele patterns (i.e., one predominant epiallele or a mixture of low represented epialleles) during two disease stages (i.e., before and after treatment). For this purpose, we compared epiallele shifts between NPCs and diagnosis samples and between paired samples at diagnosis and relapse. Interestingly, among the four possible epiallele pattern changes (Additional file 3: Figure S9), the selection pattern was significantly enriched at relapse compared to that at diagnosis (p-value (paired Wilcoxon) = 0.0059, Fig. 3f). Although this pattern involved a small number of eloci (median value = 4.2%), this finding suggests that treatment escape is associated with clonal selection at specific genomic loci.

Given the central role of chromatin in stabilizing gene expression and cellular states, we wanted to determine whether chromatin states are associated with genomic loci showing significant entropy changes in MM. To answer this question, we analyzed the distribution of loci using 127 cell/tissue types at 25-state chromatin state segmentation [20] (Additional file 3: Figure S10a). Hypomethylated loci were found in quiescent states while hypermethylated epiallele shifts were predominant in bivalent promoter-associated chromatin states (Additional file 3: Figure S10b). In addition, the frequency of these two subtypes of loci varied widely between patients (Additional file 3: Figure S11).

Abnormal bivalent promoter loci CGI methylation increases intra- and intertumor heterogeneity and targets Wnt and Ras/MAPK signaling

We found that loci were predominantly enriched (82.5%) in the bivalent promoters of embryonic stem cells (ESCs) (Additional file 3: Figures S10a and S12). In the embryonic system, bivalent promoter genes are not regulated by DNA methylation but rather by the simultaneous presence of the repressive mark H3K27me3 and the active transcription mark H3K4me3, which allow low basal transcription states that are dynamically inducible to ensure a balance between self-renewal and lineage commitment [21]. We compared the average change in the DNA methylation of loci between MM samples and NPCs. In patient M#17, who had the highest number of loci, we observed that 93% of the loci acquired extensive gains of DNA methylation (higher than 40%) during neoplastic transformation (Fig. 4a and Additional file 3: Figure S13). Methylation gains are maintained during progression (Fig. 4a). This perturbation was associated with a decrease in gene expression in diagnosis and relapse samples compared to NPCs, whereas upregulated genes were present mostly at the time of diagnosis in almost all patients (Fig. 4b and Additional file 3: Figure S14). To further examine the relationship between gene expression and entropy changes in MM patients, we compared the expression levels between bivalent genes containing hypermethylated loci in their promoter and bivalent genes without loci. We found that disrupted (i.e., containing hypermethylated loci) bivalent genes were less expressed but displayed greater interpatient variability (Fig. 4c; only genes with a mean expression level above 1 were taken into account for the coefficient of variation). Gene set enrichment analysis revealed that disrupted bivalent genes were enriched for important cancer-related pathways, including the Wnt and MAPK signaling pathways (Fig. 4d).

Myeloma cells are characterized by predominant stochastic methylation gains at bivalent promoters reflected by a high PDR and epipolymorphism, leading to intra-tumor heterogeneity (Fig. 4e). We then analyzed the consequences of intratumoral methylation variability on bivalent gene expression levels across the 20 MM samples. We separated the genes into two groups according to the PDR level of their bivalent promoter (lower or higher than the mean PDR) and calculated, in each group and for each patient, the odds ratio (OR) of the association between gene expression (FPM > 1 vs. ≤ 1) and bivalent promoter methylation (mean methylation < 25% vs. mean methylation > 75%; Fig. 4f and Additional file 6: Table S5). In all MM samples, promoters with a low PDR tended to maintain the expected opposite relationship between promoter methylation and transcription, whereas in

promoters with a high PDR, for the majority of patients, the link between methylation and expression did not remain significant, and for some patients, we observed a significant link, but the relationship was opposite to the expected result ($OR > 1$). For example, *LIPG*, which showed comparable methylation levels in two samples (0.61 in M#14 and 0.63 in M#08) coupled with opposite expression levels (FPM of 4.57 in M#14 and 0.29 in M#08), demonstrates the decoupling relationship between promoter methylation and gene expression. M#14 displayed a high promoter PDR (0.80), whereas M#08 displayed a low promoter PDR (0.29, Additional file 3: Figure S15). OR analysis clearly showed the contribution of intratumoral methylation heterogeneity to increased transcriptional variability in MM. These results are concordant with those obtained in chronic lymphocytic leukemia (CLL) [15].

Entropy changes towards hypomethylation preferentially occur in PMDs

Hypomethylated epiallele shifts are localized mainly in quiescent genomic regions (Additional file 3: Figure S10). We identified in most of the MM samples two subsets of hypomethylated loci compared to normal samples (Fig. 5a, b and Additional file 3: Figure S16). The first type included extensively demethylated loci (Fig. 5b, bottom) and the second type contained epipolymorphic loci emerging due to partial methylation loss (Fig. 5b, top). Genomic regions that have lost their methylated state, termed partially methylated domains (PMDs), were initially discovered in a fibroblast cell line [22]. Several studies have reported cancer and noncancer human primary cells with PMDs [23, 24, 25, 26, 27]. PMDs cover approximately 50 to 75% of the genome of the human primary cell types and tissues investigated, while roughly a quarter are shared which indicates that PMDs retain strong tissue and cell type specificity characteristics [26]. To perform a comprehensive analysis of PMDs in MM, we used MethylSeekR [28] with default parameters to analyze WGBS data from our cohort of MM samples and the available WGBS dataset. We detected PMDs in both NPC samples and in only five MM samples from our cohort (Additional file 3: Figure S17). The PMD structure was highly similar between normal and MM cells despite a very variable level of DNA methylation (Additional file 3: Figure S18a). The base overlap was greater than 80%, the median length distribution was approximately 51 kb and the mean genome coverage was approximately 65% of the genome (Additional file 3: Figure S18b). Given the very good overlap between NPC and MM PMDs, we subsequently studied PMDs determined from NPCs (referred to as PC-PMDs) in MM patients.

PC-PMD borders are not random and coincide with spatial genome organization, as indicated by the closeness of topological associated domain (TAD) borders to PC-PMD borders, which is larger than expected by chance ($p < 0.001$). In addition, PC-PMDs share key features with PMDs of various tumor types and normal tissues [24, 26, 27], including the correlation with lamina-associated domains (LADs) [29], late-replication timing and low gene density (Additional file 3: Figure S19a,b,c). Except for one sample, PC-PMDs that intersected with LADs displayed a long replication time and significantly low methylation levels (Additional file 3: Figure S19d,e). Importantly, the vast majority of hypomethylated loci (70%) were located within PC-PMDs.

Severe DNA methylation loss in PC-PMDs contributes to the redistribution of repressive histone marks and perturbations in CGIs/TSSs

We next examined the PC-PMDs methylation status across patients and the associated functional features. As expected, the DNA methylation level was lower within than outside of PC-PMDs (Fig. 5c) and variable between patients (Fig. 5d). In addition, this variability was associated with the PC-PMD length and replication time (Additional file 3: Figure S20). To study the relationship between DNA methylation and other epigenetic features, we integrated available WGBS data together with ChIP-seq data for histone marks. We found that the DNA methylation changes within PC-PMDs were associated with perturbations in both H3K9me3 and H3K27me3. Notably, the H3K9me3 deposit was associated with severe DNA methylation loss in long PC-PMDs (> 1 Mb) (Fig. 6a), as exemplified for the patient MM15548, with an unmethylated 3 Mb PC-PMD that was highly enriched with H3K9me3 (Fig. 6b). To our knowledge, such widespread H3K9me3 deposits associated with DNA methylation loss have so far been observed only in PMDs of cancer cell lines [26, 30, 31]. H3K27me3 was also perturbed inside PMDs. Notably, H3K27me3 enrichment was correlated with DNA methylation erosion (Fig. 6c). One notable example is the *DOCK3*-containing PC-PMD, which is enriched in H3K27me3 and showed H3K27me3 deposits in patients MM15548 and MM22965, who also exhibited DNA methylation erosion in this PMD; however, in another patient (MM23977) with a methylation level in this PMD comparable to that in NPCs, the H3K27me3 mark was absent (Fig. 6d). Taken together, these results show that perturbations in repressive histone marks are variable and depend on genomic regions, patients and DNA methylation levels.

We next investigated the impact of PC-PMD methylation erosion on regulatory regions. We found that the normal near-bimodal methylation state observed outside of PC-PMDs, was completely abolished inside PC-PMDs (Fig. 6e). As a result, promoter CGIs lost their hypomethylated state and gained intermediate methylation levels, while adjacent regions became less methylated to reach an intermediate degree of methylation. Notably, the acquisition of DNA methylation inside PC-PMDs resulted in a significant increase in the intermediate methylation state ($>20\%$ to $<80\%$) of CGIs at the expense of strictly methylated or unmethylated states in all MM patients examined (Fig. 6f), in agreement with results obtained in breast cancer [24]. We noted that this phenomenon was also observed, albeit to a lesser extent, in NPCs (Additional file 3: Figure S21a); however, the percentage of loci with a high level of discordant reads ($> 50\%$) was higher in all MM compared to normal samples (Additional file 3: Figure S21b) indicating that although PC-PMD methylation is perturbed in NPCs, CGI methylation patterns are more homogeneous (i.e., either fully methylated or unmethylated) (Additional file 3: Figure S21c).

Interestingly, a large proportion of bivalent promoter CGIs with disrupted DNA methylation was embedded within PC-PMDs (39% totally included and 54% with at least 30% of shared bases, as exemplified in Additional file 3: Figure S22). Furthermore, PC-PMD methylation loss was correlated with an overall epiallele shift increase (Spearman $p = 0.46$) (Additional file 3: Figure S23). These results support the notion that PC-PMD methylation loss may locally fuel epiallele shifts.

Given the interpatient heterogeneity of PMD methylation, we also investigated variations in gene expression. We found more gene expression variability inside PC-PMDs than outside of PC-PMDs; this difference was also maintained at relapse (Fig. 7a). Interestingly, we found that genes inside PC-PMDs were abundant within 186 kb of PC-PMD boundaries (Fig. 7b). We therefore focused our analysis on genes located near boundaries: Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis revealed specific enrichment in immune-related pathways including cytokine-cytokine receptor interaction, complement and coagulation cascades, autoimmune thyroid disease, natural killer cell-mediated cytotoxicity, antigen processing and presentation, regulation of autophagy, cell adhesion molecule and graft-versus-host disease (Fig. 7c). This enrichment pattern in PC-PMDs could provide an explanation for immune evasion mechanisms that occur during tumor progression [32, 33].

Taken together, these results show that MM displays a high epigenomic instability and great transcriptomic variability in PC-PMDs which might be beneficial for tumor cells.

Three-dimensional (3D) reorganization is favored in hypomethylated PMDs

Global DNA methylation loss impacts spatial genome organization in the nucleus [34]. We wondered whether severe DNA hypomethylation that occurs during MM development could reshape 3D chromatin architecture in MM cells. The chromatin fiber of eukaryotic genomes is folded at multiple levels, including large-scale genomic structures to form distinct chromatin compartments A and B, characterized by gene-dense transcriptionally active open chromatin and gene-sparse transcriptionally closed chromatin, respectively [35]. During cancer development the genome is reorganized, and as a consequence, genomic regions of compartment A switch to those of compartment B and vice versa. The proportion of compartment A/B switching varies according to tumor type [36, 37]. We investigated the relationships between PC-PMDs and compartment A/B switching. We determined compartment A/B boundaries at 20-kb resolution from Hi-C data in a lymphoblastoid B cell line (GM12878), with a DNA methylome similar to that of PCs [12], and in the myeloma cell lines U266 and RPMI8226. We found that PC-PMDs were more prone to switch than other genomic regions (Fig. 8a). As a representative example, *IGF1R*, which encodes a major mediator of growth and survival in MM, and is located astride compartments A and B in NPCs, switched entirely to compartment A in both myeloma cell lines (Fig. 8b).

Altogether, these results show that compartment switching might be favored by PMD instability.

Discussion

MM is a heterogeneous disease. However, until now, only studies on genetic abnormalities have demonstrated the presence of subclonal populations that encountered complex evolution during the course of a disease. A recent study highlighted the contribution of intratumor epigenetic heterogeneity to shape CLL evolution after therapy [38]. To assess intrapatient DNA methylation heterogeneity in MM, we used the eRRBS technique, which allows us to capture the methylation status of

individual CpGs in a single cell (more precisely in one single allele), with a large sequencing depth. Combining the notions of epipolymorphism and epiallele switching, we showed that the MM epigenome is characterized by high intratumor heterogeneity with a predominantly stochastic methylation pattern. Furthermore, a high methylation entropy change between malignant and NPCs is associated with a high risk of relapse.

We found that genes that undergo epiallele shifting are enriched in the control of developmental genes. Interestingly, in a model wherein hematopoietic progenitors are proposed to be the cells of origin in MM, an aberrant epigenetic program persisting through normal cell differentiation is implicated in tumor initiation [39]. Further analysis of this aberrant epigenetic program revealed strong enrichment of biological functions associated with developmental regulation (Additional file 3: Figure S24), suggesting that a disruption in developmental pathways could play a key role in the initiation of MM and increase susceptibility to oncogene transformation in response to environmental changes [40].

Entropy changes in bivalent promoters are associated with DNA methylation gains. These gains are stochastic, with an elevated PDR, which leads to a decoupling relationship between promoter methylation and transcription. These results are in line with those obtained in CLL [15]. Moreover, in CLL, single cell RNA-seq analysis showed that a high PDR is correlated with a “noisy” transcriptional landscape and an intermediate transcriptional state that interferes with complete silencing or high-level expression. Several studies have also shown the role of cellular heterogeneity and gene expression noise in overcoming drug resistance or metastatic barriers [41, 42, 43]. Together, these data suggest that increased stochastic methylation variation allows tumor cells to better adapt and find new trajectories in response to environmental changes.

A large majority of entropy changes between normal and malignant plasma cells associated with DNA methylation loss are found in regions that are already partially methylated in normal cells, indicating that PC-PMDs delimit genomic regions in which methylation information is not accurately maintained due to reduced energy consumption and channel capacity, as recently demonstrated for compartment B [44]. These large regions that cover approximately 70% of the genome are the major source of variability within MM patients. The PMDs of naive B cells display strong, progressive DNA methylation loss with differentiation into PCs [25]. During their maturation, PCs undergo successive rounds of division coupled with DNA methylation loss, thus supporting the idea that PMD hypomethylation is linked to mitotic dynamics [45, 46, 47].

In myeloma cells, PC-PMD hypomethylation is associated with other key epigenetic aberrations, such as promoter CGI hypermethylation. Indeed, we observed a loss of the hypomethylated state and gained intermediate methylation levels in regulatory regions within PC-PMDs. This disturbance has already been observed in NPCs; however, disordered methylation at promoters is lower in NPCs than in malignant cells, indicating a homogeneous methylation pattern at the promoter being either fully methylated or unmethylated. Approximately half of bivalent promoters with a high PDR in MM are associated with PC-PMDs. Therefore, the DNA methylation landscape of MM resembles that of the placenta, with stochastic methylated

gain in CGIs embedded in large hypomethylated regions, suggesting that, as in other cancers, myeloma cells coopt placental nuclear programming [48, 49]; this finding is of particular interest since placental and cancerous tissues share relevant features such as immune modulation, angiogenesis and tissue invasion [50]. Interestingly, the specific enrichment of genes in immune-related pathways was revealed when we focused our analysis on genes located near PC-PMD boundaries.

PC-PMDs are also linked to the redistribution of repressive marks. We showed that PC-PMD hypomethylation is associated with a disruption in the H3K9me3 and H3K27me3 marks and variable depending on genomic regions, patients and DNA methylation levels. Promoter CGI controlling developmental genes are silenced by the PRC2 complex, which is responsible for the deposition of H3K27me3 via its catalytic component, EZH2. According to the model of Reddington et al [51], we can hypothesize that DNA methylation prevents PRC2 from binding to inappropriate targets, and that global hypomethylation due to tumorigenesis drives H3K27me3 redistribution responsible for the derepression of target genes and the repression of new genes. This redistribution may partially explain the derepression of *HOXA9* [52] and the de novo bivalent promoters [53] observed in MM. H3K27me3 redistribution could also play an important role in chromatin decompaction, as has been shown in ESCs [54], and could explain the increase in open chromatin regions within the heterochromatin state of MM samples [55]. These data highlight that PMDs must be considered as a separate entity in genomic analyses.

Furthermore, we showed that DNA hypomethylation is variable within the myeloma cells of one individual and that methylation loss modifies H3K27me3 distribution across patients; we can hypothesize that H3K27me3 redistribution is heterogeneous among myeloma cells, leading to cell-to-cell epigenetic variability. Consequently, the tumoral mass in a MM patient would be composed of an admixture of myeloma cells with divergent epigenetic identities, in accordance with what has been demonstrated recently in CLL [56].

Conclusion

Altogether, our results show the importance of genome-wide epigenetic analysis and reveal marked PMD instability in MM patients at presentation. The perturbation in PMDs occurs at the epigenetic, transcriptomic and 3D organization levels and is responsible for interpatient variability. This disturbance in PMDs could explain, in part, some aberrant and heterogeneous phenotypes across MM samples [57, 58]. In addition, the lack of accurate global DNA methylation maintenance also drives intrapatient DNA methylation heterogeneity, which can contribute to intrapatient variability, allowing cell-to-cell diversity in transcriptional programs and opening multiple trajectories in response to therapy (Additional file 3: Figure S25).

Materials and Methods

Patients and samples

Myeloma PCs were derived from bone marrow samples collected at different stages of myeloma: 47 samples from 26 patients were collected at the SMM, diagnosis and relapse stages. Patients (15 women and 11 men, with a median age of 59 years) were monitored at the Intergroupe Francophone du Myélome (IFM) center, and all provided informed consent (Additional file 1: Table S1).

Sample preparation and nucleic acid purification

Myeloma cells were purified using nanobeads and an anti-CD138 antibody (RoboSep, Stem Cell Technologies). After immunomagnetic sorting, the purity of the plasma cell suspension was verified, and only samples with at least 85% of PCs were subjected to genomic analysis. The average cell purity of MM was > 99% (range 90-100%). Control normal plasma cells (NPC1 to NPC3) were obtained from the bone marrow samples of three patients suspected of having monoclonal gammopathy of undetermined significance (MGUS), and the absence of abnormal PCs was assessed by flow cytometry. The average CD138+ cell purity was > 75%. The DNA and RNA of CD138+ cells were purified using the Qiagen protocol and the quality and quantity of the nucleic acids were measured on the Nanodrop, Qubit and Agilent profiles and stored in the MM biobanks of the Nantes and Toulouse Hospitals.

eRRBS

eRRBS is an improvement of the reduced representation bisulfite sequencing (RRBS) protocol, resulting in an increase in CpG detection and coverage. eRRBS library preparations were performed by Integragen and adapted from the protocol described by Garrett-Bakelman *et al.* [59]. DNA was digested with the Msp1 enzyme, fragments between 150 bp and 400 bp were selected, and bisulfite conversions were processed. Libraries were sequenced on a HiSeqTM 4000 Illumina machine using 75 bp paired-end reads to an average depth of 50X per covered CpG. The average number of reads sequenced per patient was 55 216 285; and the average alignment rate of uniquely mapped reads was 60.81%, with an average of 2 712 252 CpGs per patients with a coverage of 10X and an average of 1 538 510 CpGs per patient with a coverage of 60X (Additional file 4: Table S3).

eRRBS analysis

The adaptor sequences were removed by Cutadapt (version 1.10) [60]. FastQC (version 0.11.4) was used for quality control of the Illumina paired-end sequencing data [61]. Bisulfite reads were aligned to the bisulfite-converted hg19 genome with the nondirectional model of Bismark alignment software (version 0.14.1) [62]. CpG methylation levels were obtained using the R package methylKit with the default settings: a minimum of 10 reads covering a CpG and at least 20 PHRED quality scores by CpG [63]. We calculated the epigenetic changes between two stages using methclone [17], an algorithm that detects loci of 4 adjacent CpGs (minimum depth of 60 reads), called epialleles (16 possible patterns according to CpG methylation).

Epipolymorphism and epiallele shifting

Epipolymorphism (*e*) was calculated for each locus of the four adjacent CpGs, to measure intratumor epigenetic heterogeneity, as previously described by Landan *et al.* [18]:

$$e = 1 - \sum_{i=1}^{16} p_i^2$$

where p_i is the proportion of the epiallele *i*. The minimal epipolymorphism is 0 (only one pattern represented); epipolymorphism cannot have a value greater than 1. The maximum value of epipolymorphism is 0.9375 (when all 16 patterns are equally distributed). To quantify the degree of epiallele pattern shift, methclone was

used to compute the entropy difference (ΔS) and thus compare the distributions of epialleles between different stages. The entropy difference value can range from 0 (no change) to -144 (maximum change). Loci are characterized as eloci when $\Delta S < -70$ (corresponding to a significant entropy shift). The methclone algorithm allows the discovery, quantification and ranking of subclonal selection based on epiallele shifts. This allows us to measure clonal evolution between disease stages and epigenetic heterogeneity.

To normalize and compare the number of eloci per patient, we computed the number of eloci per million loci sequenced (EPM) as previously described by Sheng Li et al. [17]:

$$EPM = \frac{10^6}{C} \times E$$

where E is the total number of eloci detected between the two stages and C is the total number of loci covered by both samples.

A locus between the NPC and diagnosis samples was defined as an elocus only if it was annotated as an elocus with the 3 NPCs. A hypermethylated or hypomethylated elocus was defined as a DNA methylation difference of at least 25%.

The PDR variable was defined as the percentage of epialleles exhibiting heterogeneous DNA methylation (i.e., not fully methylated or unmethylated).

WGBS

Data quality control and adaptor trimmed were performed with the Trimmomatic tool [64]. Read mapping was carried out with the methylCtools aligner [31]. methylCtools uses an alignment approach similar to that of Bismark by improving the handling of large amounts of data and the speed of alignment. CpG extraction and methylation analyses were carried out with methylCtools and the bsseq R package [65], allowing the analysis, management and storage of WGBS data. As with eRRBS data, we set a coverage threshold of a minimum of 10X for each CpG sequenced for our analyses. The dmrseq R package was used to visualize WGBS data and smooth methylation signals [66].

Detecting PMDs

PMDs were detected in WGBS samples using the R package MethylSeekR [28]. Prior to the detection of PMDs, CpGs overlapping with common single-nucleotide polymorphism (SNPs) were removed from the data (dbSNP137commonhg19, version 1.0.0). The distribution of α values was used to determine the presence or absence of PMDs in samples. When the α distribution was bimodal, the segmentPMDs function was run to identify PMDs on the genome by a hidden Markov model.

RNA sequencing (RNA-seq)

RNA-seq was performed using 200 ng of total RNA by GATC Biotech. Directional libraries were generated after mRNA selection by polyA selection using the UTP method. RNA-seq libraries were sequenced on a HiSeq 2500 Illumina machine using 100 bp paired-end reads. The average number of reads sequenced per sample was 68 454 433. The average alignment rate of uniquely mapped reads was 66% (corresponding to 48 034 921 reads). Read alignment was performed using the STAR aligner (version 2.4.0f1) [67] and human genome hg19 as the reference. FastQC

(version 0.11.4) [61] was used for basic quality control of the Illumina paired-end sequencing data. PCR duplicates were determined and removed using the Picard algorithm [68]. The number of reads mapped to each gene was calculated using HTSeq-count, part of the HTSeq framework [69], version 0.6.0. We then normalized the mapped read counts per million of mapped fragments (FPM) using the robust median ratio method with the DESeq2 R package [70].

Hi-C data analysis and compartments A/B

Hi-C datasets were downloaded for three cell lines: GM12878 (GSM1608505), RPMI-8226 (GSM2334832) and U266 (GSM2334834). HiC-Pro software (version 2.10.0) [71] was used to process Hi-C data, from raw-data to normalized contact maps. All reads were mapped to hg19 using Bowtie2 (global parameters: `-very-sensitive -L 30 -score-min L,-0.6,-0.2 -end-to-end -reorder`; local parameters: `-very-sensitive -L 20 -score-min L,-0.6,-0.2 -end-to-end -reorder`). Contact maps were generated at 20 kb resolution and normalized by the iterative correction and eigenvector decomposition (ICED) technique.

The 20-kb resolution intrachromosomal contact matrices generated by HiC-Pro were used as input to determine compartments A/B and to annotate and visualize interaction maps with R package HiTC version 1.22.1 [72]. Principal component analysis was used to separate chromatin into two compartments: compartment A, with higher gene density, and compartment B, with lower gene density. The determination of compartments A and B was estimated by the analysis of the eigenvectors of the genome contact matrix by the observed-expected method. On the basis that changes in the sign of the eigenvector of the contact matrix correspond to the limits of the genome compartments and taking into account gene density, the compartmentalization of the genome was defined.

TADs

Hi-C data were used to determine the TADs with the HiCExplorer tool, a set of programs used to process, normalize, analyze and visualize Hi-C data [73]. TADs were defined with `hicFindTADs` and `hicPlotTADs` was used to visualize the TADs. For the comparison of PMD and TAD borders, we generated a set of randomized PMDs with a size and genomic distributions similar to those of real PMDs. We then calculated the distance between TAD borders and, on the one hand, PMD borders, and, on the other hand, random PMD borders. A *t* test was carried out to compare the distributions of these distances.

Genomic annotation

RefSeq annotation and CpG islands were obtained from UCSC (<https://genome.ucsc.edu/>) using the February 2009 (GRCh37/hg19) assembly. CpG shores were defined as regions flanking 2 kb of CpG islands, and CpG shelves were defined as regions flanking 2 kb of CpGs shores.

HOMER was used to annotate loci and eloci, using the `annotatePeaks.pl` script [74], which determines the genomic type annotation occupied by the center of the loci. We used their basic annotation, including TSSs, transcription termination sites (TTSs), exons, introns and intergenic regions. We defined the promoters as regions

flanking 2 kb of the TSS, and we defined promoter CGIs as promoters that intersect with CpG islands.

All 127 reference epigenomes with 25 chromatin state segmentation annotations were downloaded from the NIH Roadmap Epigenomics Project (https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/imputed12marks/jointModel/final/;all\OT1\textquotedblleft25_imputed12marks_mnemonics.bed.gz\OT1\textquotedblrightfiles).

To measure the intersections between regions, bedtools was used (more precisely, the "intersect" option) [75].

Data from the BLUEPRINT Consortium were converted to hg19 coordinates using the liftOver tool [76].

Functional annotation

GREAT tools (version 3.0.0) was used to assign biological functions to noncoding genomic regions by analyzing the annotations of the nearby genes [77] under the following default gene regulatory domain definitions; basal promoter: 5 kb upstream, 1 kb downstream and extension up to 1 Mb. Enrichment statistics were computed using the binomial test and the hypergeometric gene-based test. Pathways were selected as significantly enriched if the false discovery rate (FDR q-value) was < 0.01.

The Database for Annotation, Visualization and Integrated Discovery (DAVID) web interface (version 6.7) was used to perform functional enrichment analysis from a list of genes, especially the KEGG pathways functional database [78]. Enrichment statistics were computed using the Fisher test. For the significance threshold, we have considered genes as greatly enriched if the annotation categories yielded a p-value less than 0.1 (DAVID default threshold)[79].

Survival analysis

Time to relapse data were available for 17 patients. For the relapse-free survival (RFS) analysis, the survival endpoint in this study was the time from diagnosis until relapse. The patients were divided into two groups by the median EPM value: low and high EPMS. Survival curves were estimated using the Kaplan-Meier method and compared with the log-rank test.

For all statistical tests in this study, a two-sided p-value of 0.05 was considered statistically significant. All statistical analyses were performed with software R 3.5.1 [80], in addition to the packages already mentioned in the text, packages data.table [81], viridis, ggbio, GenomicRanges, RColorBrewer, biovizBase, grid, gridBase, MASS, fields, KernSmooth, sp, org.Hs.eg.db, DBI and survival were used.

Abbreviation

MM: Multiple myeloma; PMDs : Partially methylated domains; CGI: CpG island; TSS: Transcription start site; PCs: Plasma cells; NPCs: Normal plasma cells; WGBS: Whole genome bisulfite sequencing; eRRBS: Enhanced reduced representation bisulfite sequencing; SMM: Smoldering multiple myeloma; PDR: Proportion of discordant reads; eloci: Epigenetic loci; EPM: Eloci per million loci covered; ESCs: embryonic stem cells; FPM: mapped read counts per million of mapped fragments; OR: Odd ratio; CLL: Chronic lymphocytic leukemia; TAD: Topological associated domain; LAD: lamina-associated domain; KEGG: Kyoto encyclopedia of genes and genomes; 3D: Three-dimensional; IFM: Intergroupe francophone du myélome; MGUS: Gammopathy of undetermined significance; RRBS: Reduced representation bisulfite sequencing; SNPs: Single-nucleotide polymorphism; RNA-seq: RNA sequencing; ICED: Iterative correction and eigenvector decomposition; TTS: Transcription termination site; FDR: False discovery rate; DAVID: Database for annotation, visualization and integrated discovery; RFS: Relapse-free survival.

Acknowledgements

This study makes use of data generated by the BLUEPRINT Consortium. A full list of the investigators who contributed to the generation of the data is available from www.blueprint-epigenome.eu. Funding for the project was provided by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement no 282510 – BLUEPRINT.

Author's contributions

S.M. and F.M. designed the study, collected, and analyzed the data and wrote the paper; J.D. and C.H. collected and analyzed the data and wrote the paper; M.D., E.D., N.R., H.A.-L., O.D., T.F., J.-P. M., R.E., M.P. and N.M. collected the data; C.G.-C., V.G. and L.C. analyzed the data.

Funding

This study was supported by the International Myeloma Foundation (IMF-R16103NN), the I-SITE Next (ANR-16-IDEX-0007) and the SIRIC ILIAD (INCa-DGOS-Inserm-12558).

Availability of data

WGBS, eRRBS and RNA-seq data have been submitted to the European Genome-phenome Archive (<https://www.ebi.ac.uk/ega>) under dataset accession EGAS00001004346, EGAS00001004348, EGAS00001004347, respectively.

WGBS Blueprint data of normal plasma cells and multiple myeloma patient are available from the EGA under the accessions EGAD00001002322 and EGAD00001002521 respectively. ChIP-Seq data of multiple myeloma patient are available from the EGA under the accessions EGAD00001002379.

Ethics approval and consent to participate

All the samples were obtained from patients enrolled in the different randomized clinical trials implemented by the Intergroupe Francophone du Myelome (IFM); IFM 99-02, IFM 99-06, IFM 2005-01, FIRST/IFM 2007-01, IFM 2007-02, Genomgus Study and IFM 2009 (www.clinicaltrials.gov) or routine practice in France. Samples and data were obtained after an informed consent was signed. All research procedures conformed to the principles of Helsinki Declaration.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Université de Nantes, CNRS, INSERM, CRCINA, F-44000 Nantes, France. ²Institut de Cancérologie de l'Ouest, Nantes-Saint Herblain, France. ³LS2N, CNRS, Université de Nantes, Nantes, France. ⁴Centre Hospitalier Universitaire, Nantes, France. ⁵Institut Universitaire du Cancer, CHU, Centre de Recherche en Cancérologie de Toulouse, INSERM 1037, Toulouse, France. ⁶Centre Hospitalier Universitaire, Rennes, France. ⁷Centre Hospitalier Universitaire, Lille, France. ⁸Research Group Genome Organization & Function, DKFZ, and BioQuant Heidelberg, 69120 Heidelberg, Germany. ⁹Health Data Science Unit, Medical Faculty Heidelberg and BioQuant, 69120 Heidelberg, Germany. ¹⁰Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Charitéplatz 1, 10117 Berlin, Germany Berlin Institute of Health (BIH), Center for Digital Health, Anna-Louisa-Karsch-Strasse 2, 10178 Berlin, Germany. ¹¹Dana-Farber Cancer Institute, Harvard Medical School, LeBow Institute for Myeloma Therapeutics and Jerome Lipper Center for Multiple Myeloma Research, Boston, MA, United States.

References

1. Leung-Hagsteeijn, C., Erdmann, N., Cheung, G., Keats, J.J., Stewart, A.K., Reece, D.E., Chung, K.C., Tiedemann, R.E.: Xbp1s-negative tumor b cells and pre-plasmablasts mediate therapeutic proteasome inhibitor resistance in multiple myeloma. *Cancer cell* **24**(3), 289–304 (2013)
2. Zhang, X.-D., Baladandayuthapani, V., Lin, H., Mulligan, G., Li, B., Esseltine, D.-L.W., Qi, L., Xu, J., Hunziker, W., Barlogie, B., *et al.*: Tight junction protein 1 modulates proteasome capacity and proteasome inhibitor sensitivity in multiple myeloma via egfr/jak1/stat3 signaling. *Cancer cell* **29**(5), 639–652 (2016)
3. Bolli, N., Avet-Loiseau, H., Wedge, D.C., Van Loo, P., Alexandrov, L.B., Martincorena, I., Dawson, K.J., Iorio, F., Nik-Zainal, S., Bignell, G.R., *et al.*: Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nature communications* **5** (2014)
4. Weinhold, N., Ashby, C., Rasche, L., Chavan, S.S., Stein, C., Stephens, O.W., Tytarenko, R., Bauer, M.A., Meissner, T., Deshpande, S., *et al.*: Clonal selection and double-hit events involving tumor suppressor genes underlie relapse in myeloma. *Blood* **128**(13), 1735–1744 (2016)
5. Magrangeas, F., Avet-Loiseau, H., Gouraud, W., Lodé, L., Decaux, O., Godmer, P., Garderet, L., Voillat, L., Facon, T., Stoppa, A.M., *et al.*: Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia* **27**(2), 473 (2013)
6. Keats, J.J., Chesi, M., Egan, J.B., Garbitt, V.M., Palmer, S.E., Braggio, E., Van Wier, S., Blackburn, P.R., Baker, A.S., Dispenzieri, A., *et al.*: Clonal competition with alternating dominance in multiple myeloma. *Blood* **120**(5), 1067–1076 (2012)
7. Egan, J.B., Shi, C.-X., Tembe, W., Christoforides, A., Kurdoglu, A., Sinari, S., Middha, S., Asmann, Y., Schmidt, J., Braggio, E., *et al.*: Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood, The Journal of the American Society of Hematology* **120**(5), 1060–1066 (2012)

8. Walker, B.A., Wardell, C.P., Melchor, L., Hulkki, S., Potter, N.E., Johnson, D.C., Fenwick, K., Kozarewa, I., Gonzalez, D., Lord, C.J., *et al.*: Intraclonal heterogeneity and distinct molecular mechanisms characterize the development of t (4; 14) and t (11; 14) myeloma. *Blood, The Journal of the American Society of Hematology* **120**(5), 1077–1086 (2012)
9. Maura, F., Bolli, N., Angelopoulos, N., Dawson, K.J., Leongamornlert, D., Martincorena, I., Mitchell, T.J., Fullam, A., Gonzalez, S., Szalat, R., *et al.*: Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Nature communications* **10**(1), 1–12 (2019)
10. Lawrence, M.S., Stojanov, P., Mermel, C.H., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., Getz, G.: Discovery and saturation analysis of cancer genes across 21 tumor types. *Nature* **505**(7484), 495 (2014)
11. Marusyk, A., Janiszewska, M., Polyak, K.: Intratumor heterogeneity: The rosetta stone of therapy resistance. *Cancer cell* **37**(4), 471–484 (2020)
12. Agirre, X., Castellano, G., Pascual, M., Heath, S., Kulis, M., Segura, V., Bergmann, A., Esteve, A., Merkel, A., Raineri, E., *et al.*: Whole-genome analysis in multiple myeloma reveals dna hypermethylation of b cell-specific enhancers. *Genome research* **25**(4), 478–487 (2015)
13. Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P., van Dijk, C.M., Tollenaar, R.A., *et al.*: Regions of focal dna hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature genetics* **44**(1), 40 (2012)
14. Hansen, K.D., Timp, W., Bravo, H.C., Sabuncian, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., *et al.*: Increased methylation variation in epigenetic domains across cancer types. *Nature genetics* **43**(8), 768–775 (2011)
15. Landau, D.A., Clement, K., Ziller, M.J., Boyle, P., Fan, J., Gu, H., Stevenson, K., Sougnez, C., Wang, L., Li, S., *et al.*: Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer cell* **26**(6), 813–825 (2014)
16. Li, S., Garrett-Bakelman, F.E., Chung, S.S., Sanders, M.A., Hricik, T., Rapaport, F., Patel, J., Dillon, R., Vijay, P., Brown, A.L., *et al.*: Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nature medicine* **22**(7), 792 (2016)
17. Li, S., Garrett-Bakelman, F., Perl, A.E., Luger, S.M., Zhang, C., To, B.L., Lewis, I.D., Brown, A.L., D'Andrea, R.J., Ross, M.E., *et al.*: Dynamic evolution of clonal epialleles revealed by methclone. *Genome biology* **15**(9), 472 (2014)
18. Landan, G., Cohen, N.M., Mukamel, Z., Bar, A., Molchadsky, A., Brosh, R., Horn-Saban, S., Zalcenstein, D.A., Goldfinger, N., Zundeleich, A., *et al.*: Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nature genetics* **44**(11), 1207–1214 (2012)
19. Easwaran, H., Johnstone, S.E., Van Neste, L., Ohm, J., Mosbrugger, T., Wang, Q., Aryee, M.J., Joyce, P., Ahuja, N., Weisenberger, D., *et al.*: A dna hypermethylation module for the stem/progenitor cell signature of cancer. *Genome research* **22**(5), 837–849 (2012)
20. Ernst, J., Kellis, M.: Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nature biotechnology* **33**(4), 364 (2015)
21. Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.-K., Koche, R.P., *et al.*: Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**(7153), 553–560 (2007)
22. Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., *et al.*: Human dna methylomes at base resolution show widespread epigenomic differences. *nature* **462**(7271), 315–322 (2009)
23. Schroeder, D.I., Blair, J.D., Lott, P., Yu, H.O.K., Hong, D., Cray, F., Ashwood, P., Walker, C., Korf, I., Robinson, W.P., *et al.*: The human placenta methylome. *Proceedings of the national academy of sciences* **110**(15), 6037–6042 (2013)
24. Brinkman, A.B., Nik-Zainal, S., Simmer, F., Rodriguez-Gonzalez, F.G., Smid, M., Alexandrov, L.B., Butler, A., Martin, S., Davies, H., Glodzik, D., *et al.*: Partially methylated domains are hypervariable in breast cancer and fuel widespread cpg island hypermethylation. *Nature communications* **10**(1), 1–10 (2019)
25. Durek, P., Nordström, K., Gasparoni, G., Salhab, A., Kressler, C., De Almeida, M., Bassler, K., Ulas, T., Schmidt, F., Xiong, J., *et al.*: Epigenomic profiling of human cd4+ t cells supports a linear differentiation model and highlights molecular regulators of memory development. *Immunity* **45**(5), 1148–1161 (2016)
26. Salhab, A., Nordström, K., Gasparoni, G., Kattler, K., Ebert, P., Ramirez, F., Arrigoni, L., Müller, F., Polansky, J.K., Cadenas, C., *et al.*: A comprehensive analysis of 195 dna methylomes reveals shared and cell-specific features of partially methylated domains. *Genome biology* **19**(1), 150 (2018)
27. Zhou, W., Dinh, H.Q., Ramjan, Z., Weisenberger, D.J., Nicolet, C.M., Shen, H., Laird, P.W., Berman, B.P.: Dna methylation loss in late-replicating domains is linked to mitotic cell division. *Nature genetics* **50**(4), 591–602 (2018)
28. Burger, L., Gaidatzis, D., Schübeler, D., Stadler, M.B.: Identification of active regulatory regions from dna methylation data. *Nucleic acids research* **41**(16), 155–155 (2013)
29. Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., *et al.*: Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**(7197), 948–951 (2008)
30. Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., *et al.*: Global dna hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome research* **22**(2), 246–258 (2012)
31. Hovestadt, V., Jones, D.T., Picelli, S., Wang, W., Kool, M., Northcott, P.A., Sultan, M., Stachurski, K., Ryzhova, M., Warnatz, H.-J., *et al.*: Decoding the regulatory landscape of medulloblastoma using dna methylation sequencing. *Nature* **510**(7506), 537–541 (2014)
32. Jung, H., Kim, H.S., Kim, J.Y., Sun, J.-M., Ahn, J.S., Ahn, M.-J., Park, K., Esteller, M., Lee, S.-H., Choi, J.K.: Dna methylation loss promotes immune evasion of tumours with high mutation and copy number load.

- Nature communications **10**(1), 1–12 (2019)
33. Ryu, D., Kim, S.J., Hong, Y., Jo, A., Kim, N., Kim, H.-J., Lee, H.-O., Kim, K., Park, W.-Y.: Alterations in the transcriptional programs of myeloma cells and the microenvironment during extramedullary progression affect proliferation and immune evasion. *Clinical Cancer Research* **26**(4), 935–944 (2020)
 34. Madakashira, B.P., Sadler, K.C.: Dna methylation, nuclear organization, and cancer. *Frontiers in genetics* **8**, 76 (2017)
 35. Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., *et al.*: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science* **326**(5950), 289–293 (2009)
 36. Barutcu, A.R., Lajoie, B.R., McCord, R.P., Tye, C.E., Hong, D., Messier, T.L., Browne, G., van Wijnen, A.J., Lian, J.B., Stein, J.L., *et al.*: Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome biology* **16**(1), 214 (2015)
 37. Wu, P., Li, T., Li, R., Jia, L., Zhu, P., Liu, Y., Chen, Q., Tang, D., Yu, Y., Li, C.: 3d genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. *Nature communications* **8**(1), 1–11 (2017)
 38. Gaiti, F., Chaligne, R., Gu, H., Brand, R.M., Kothen-Hill, S., Schulman, R.C., Grigorev, K., Risso, D., Kim, K.-T., Pastore, A., *et al.*: Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* **569**(7757), 576–580 (2019)
 39. Vicente-Dueñas, C., Romero-Camarero, I., González-Herrero, I., Alonso-Escudero, E., Abollo-Jiménez, F., Jiang, X., Gutierrez, N.C., Orfao, A., Marín, N., Villar, L.M., *et al.*: A novel molecular mechanism involved in multiple myeloma development revealed by targeting mafb to haematopoietic progenitors. *The EMBO journal* **31**(18), 3704–3717 (2012)
 40. Vaz, M., Hwang, S.Y., Kagiampakis, I., Phallen, J., Patil, A., O'Hagan, H.M., Murphy, L., Zahnow, C.A., Gabrielson, E., Velculescu, V.E., *et al.*: Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by kras mutations. *Cancer cell* **32**(3), 360–376 (2017)
 41. Nguyen, A., Yoshida, M., Goodarzi, H., Tavazoie, S.F.: Highly variable cancer subpopulations that exhibit enhanced transcriptome variability and metastatic fitness. *Nature communications* **7**(1), 1–13 (2016)
 42. Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., *et al.*: Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**(7658), 431–435 (2017)
 43. Farquhar, K.S., Charlebois, D.A., Szenk, M., Cohen, J., Nevozhay, D., Balázs, G.: Role of network-mediated stochasticity in mammalian drug resistance. *Nature communications* **10**(1), 1–14 (2019)
 44. Jenkinson, G., Pujadas, E., Goutsias, J., Feinberg, A.P.: Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nature Genetics* **49**(5), 719–729 (2017)
 45. Barwick, B.G., Scharer, C.D., Bally, A.P., Boss, J.M.: Plasma cell differentiation is coupled to division-dependent dna hypomethylation and gene regulation. *Nature immunology* **17**(10), 1216 (2016)
 46. Lai, A.Y., Mav, D., Shah, R., Grimm, S.A., Phadke, D., Hatz, K., Melnick, A., Geigerman, C., Sobol, S.E., Jaye, D.L., *et al.*: Dna methylation profiling in human b cells reveals immune regulatory elements and epigenetic plasticity at alu elements during b-cell activation. *Genome research* **23**(12), 2030–2041 (2013)
 47. Caron, G., Hussein, M., Kulis, M., Delaloy, C., Chatonnet, F., Pignarre, A., Avner, S., Lemarié, M., Mahé, E.A., Verdaguer-Dot, N., *et al.*: Cell-cycle-dependent reconfiguration of the dna methylome during terminal differentiation of human b cells into plasma cells. *Cell reports* **13**(5), 1059–1071 (2015)
 48. Smith, Z.D., Shi, J., Gu, H., Donaghey, J., Clement, K., Cacchiarelli, D., Gnirke, A., Michor, F., Meissner, A.: Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* **549**(7673), 543–547 (2017)
 49. Lorincz, M.C., Schübeler, D.: Evidence for converging dna methylation pathways in placenta and cancer. *Developmental cell* **43**(3), 257–258 (2017)
 50. Novakovic, B., Saffery, R.: Placental pseudo-malignancy from a dna methylation perspective: unanswered questions and future directions. *Frontiers in genetics* **4**, 285 (2013)
 51. Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H., *et al.*: Redistribution of h3k27me3 upon dna hypomethylation results in de-repression of polycomb target genes. *Genome biology* **14**(3), 25 (2013)
 52. Chapman, M.A., Lawrence, M.S., Keats, J.J., Cibulskis, K., Sougnez, C., Schinzel, A.C., Harview, C.L., Brunet, J.-P., Ahmann, G.J., Adli, M., *et al.*: Initial genome sequencing and analysis of multiple myeloma. *Nature* **471**(7339), 467–472 (2011)
 53. Agarwal, P., Alzrigat, M., Párraga, A.A., Enroth, S., Singh, U., Ungerstedt, J., Österborg, A., Brown, P.J., Ma, A., Jin, J., *et al.*: Genome-wide profiling of histone h3 lysine 27 and lysine 4 trimethylation in multiple myeloma reveals the importance of polycomb gene targeting and highlights ezh2 as a potential therapeutic target. *Oncotarget* **7**(6), 6809 (2016)
 54. McLaughlin, K.A., Flyamer, I.M., Thomson, J.P., Mjoseng, H.K., Shukla, R., Williamson, I., Grimes, G.R., Illingworth, R.S., Adams, I.R., Pennings, S., *et al.*: Dna methylation directs polycomb-dependent 3d genome re-organisation in naive pluripotency. *bioRxiv*, 527309 (2019)
 55. Jin, Y., Chen, K., De Paepe, A., Hellqvist, E., Krstic, A.D., Metang, L., Gustafsson, C., Davis, R.E., Levy, Y.M., Surapaneni, R., *et al.*: Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood, The Journal of the American Society of Hematology* **131**(19), 2138–2150 (2018)
 56. Pastore, A., Gaiti, F., Lu, S.X., Brand, R.M., Kulm, S., Chaligne, R., Gu, H., Huang, K.Y., Stamenova, E.K., Béguelin, W., *et al.*: Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in cll. *Nature communications* **10**(1), 1–11 (2019)
 57. Bataille, R., Jégou, G., Robillard, N., Barillé-Nion, S., Harousseau, J.-L., Moreau, P., Amiot, M., Pellat-Deceunynck, C.: The phenotype of normal, reactive and malignant plasma cells. identification of many

- and multiple myelomas" and of new targets for myeloma therapy. *Haematologica* **91**(9), 1234–1240 (2006)
58. Capp, J.-P., Bataille, R.: Multiple myeloma exemplifies a model of cancer based on tissue disruption as the initiator event. *Frontiers in oncology* **8**, 355 (2018)
 59. Garrett-Bakelman, F.E., Sheridan, C.K., Kacmarczyk, T.J., Ishii, J., Betel, D., Alonso, A., Mason, C.E., Figueroa, M.E., Melnick, A.M.: Enhanced reduced representation bisulfite sequencing for assessment of dna methylation at base pair resolution. *Journal of visualized experiments: JoVE* (96) (2015)
 60. Martin, M.: Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**(1), 10 (2011)
 61. Andrews, S., et al.: FastQC: a quality control tool for high throughput sequence data (2010)
 62. Krueger, F., Andrews, S.R.: Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *bioinformatics* **27**(11), 1571–1572 (2011)
 63. Akalin, A., Kormaksson, M., Li, S., Garrett-Bakelman, F.E., Figueroa, M.E., Melnick, A., Mason, C.E.: methylkit: a comprehensive r package for the analysis of genome-wide dna methylation profiles. *Genome biology* **13**(10), 87 (2012)
 64. Bolger, A.M., Lohse, M., Usadel, B.: Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* **30**(15), 2114–2120 (2014)
 65. Hansen, K.D., Langmead, B., Irizarry, R.A.: Bsmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome biology* **13**(10), 83 (2012)
 66. Korthauer, K., Chakraborty, S., Benjamini, Y., Irizarry, R.A.: Detection and accurate false discovery rate control of differentially methylated regions from whole genome bisulfite sequencing. *Biostatistics* **20**(3), 367–383 (2019)
 67. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingeras, T.R.: Star: ultrafast universal rna-seq aligner. *Bioinformatics* **29**(1), 15–21 (2013)
 68. Picard toolkit. Broad Institute (2018)
 69. Anders, S., Pyl, P.T., Huber, W.: Htseq—a python framework to work with high-throughput sequencing data. *Bioinformatics* **31**(2), 166–169 (2015)
 70. Love, M.I., Huber, W., Anders, S.: Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome biology* **15**(12), 550 (2014)
 71. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., Barillot, E.: Hic-pro: an optimized and flexible pipeline for hi-c data processing. *Genome biology* **16**(1), 259 (2015)
 72. Servant, N., Lajoie, B.R., Nora, E.P., Giorgetti, L., Chen, C.-J., Heard, E., Dekker, J., Barillot, E.: Hitc: exploration of high-throughput 'c' experiments. *Bioinformatics* **28**(21), 2843–2844 (2012)
 73. Ramírez, F., Bhardwaj, V., Arrighi, L., Lam, K.C., Grüning, B.A., Villaveces, J., Habermann, B., Akhtar, A., Manke, T.: High-resolution tads reveal dna sequences underlying genome organization in flies. *Nature communications* **9**(1), 1–15 (2018)
 74. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., Glass, C.K.: Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Molecular cell* **38**(4), 576–589 (2010)
 75. Quinlan, A.R., Hall, I.M.: Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**(6), 841–842 (2010)
 76. Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haeussler, M., et al.: The ucsc genome browser database: 2015 update. *Nucleic acids research* **43**(D1), 670–681 (2015)
 77. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., Bejerano, G.: Great improves functional interpretation of cis-regulatory regions. *Nature biotechnology* **28**(5), 495–501 (2010)
 78. Huang, D.W., Sherman, B.T., Lempicki, R.A.: Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* **37**(1), 1–13 (2009)
 79. Sherman, B.T., Lempicki, R.A., et al.: Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature protocols* **4**(1), 44 (2009)
 80. Team, R.C., et al.: R: A language and environment for statistical computing (2013)
 81. Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Short, T., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., et al.: Package 'data.table'. Extension of 'data.frame' (2015)

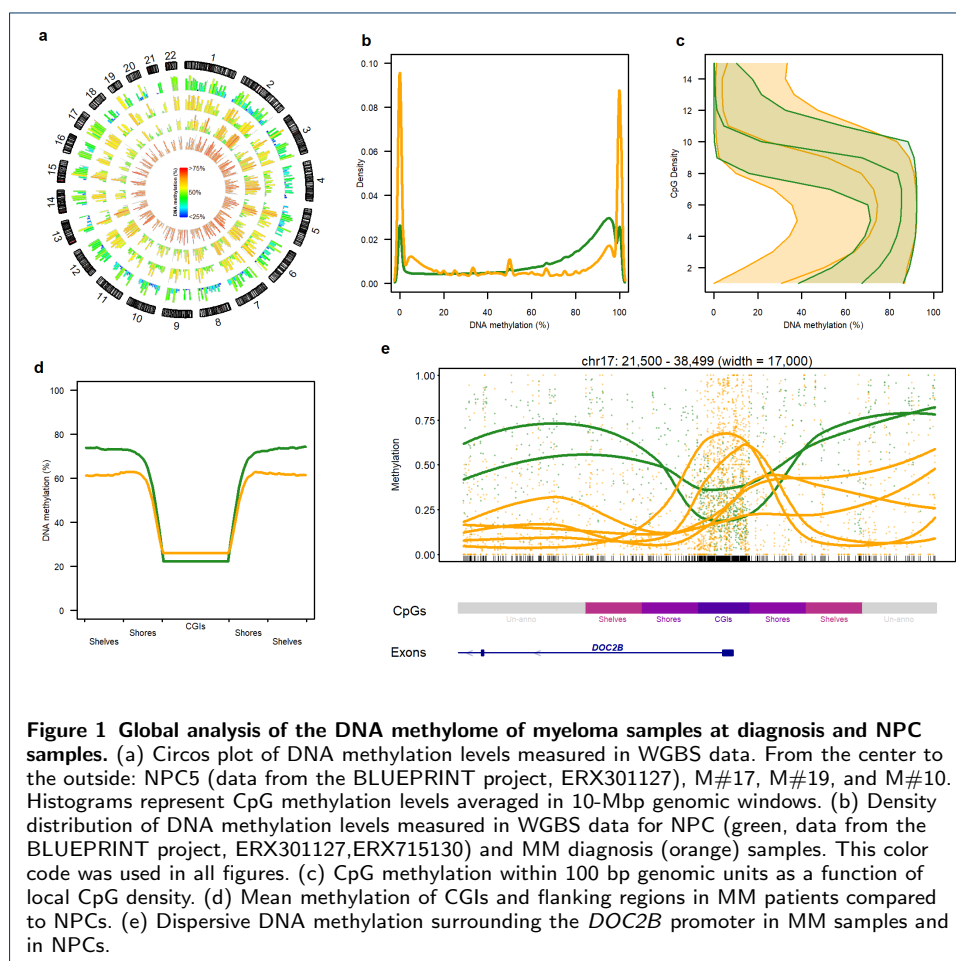


Figure 1 Global analysis of the DNA methylome of myeloma samples at diagnosis and NPC samples. (a) Circos plot of DNA methylation levels measured in WGBS data. From the center to the outside: NPC5 (data from the BLUEPRINT project, ERX301127), M#17, M#19, and M#10. Histograms represent CpG methylation levels averaged in 10-Mbp genomic windows. (b) Density distribution of DNA methylation levels measured in WGBS data for NPC (green, data from the BLUEPRINT project, ERX301127, ERX715130) and MM diagnosis (orange) samples. This color code was used in all figures. (c) CpG methylation within 100 bp genomic units as a function of local CpG density. (d) Mean methylation of CGIs and flanking regions in MM patients compared to NPCs. (e) Dispersive DNA methylation surrounding the *DOC2B* promoter in MM samples and in NPCs.

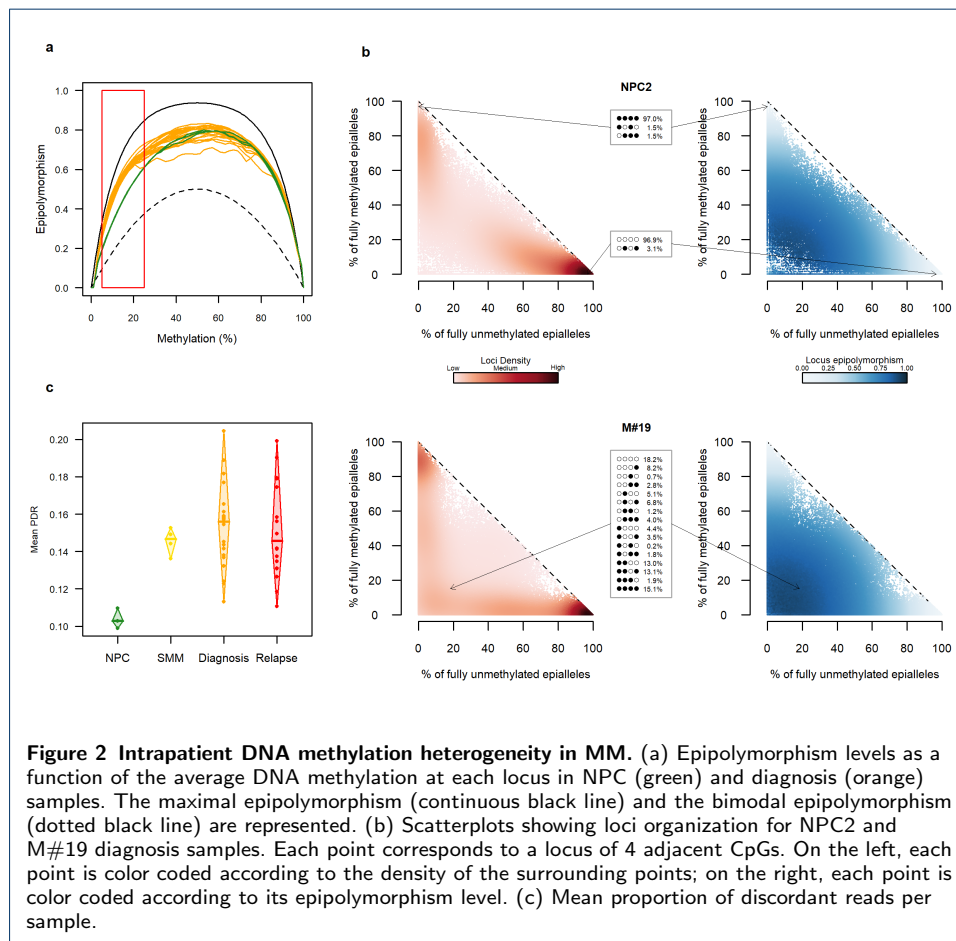
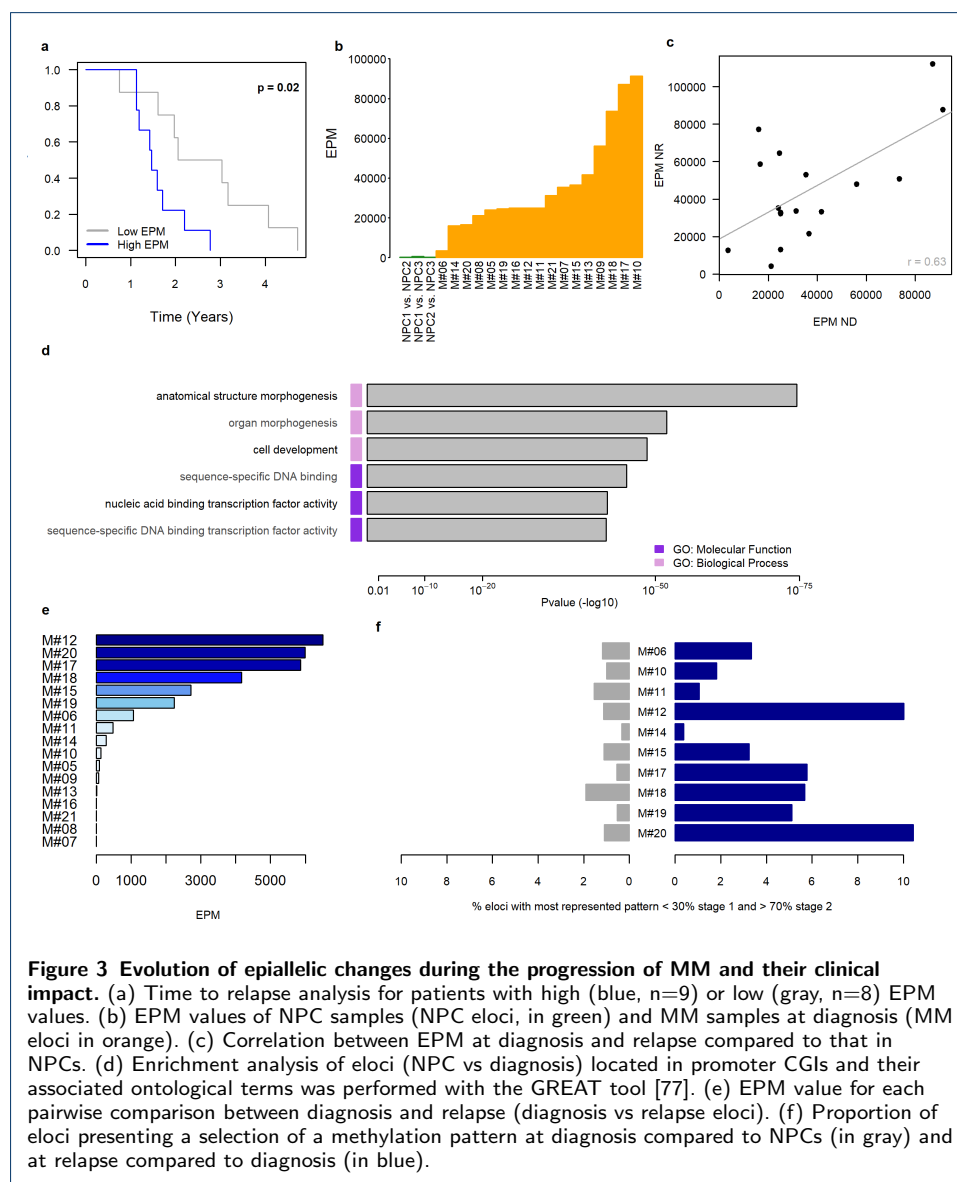


Figure 2 Intrapatient DNA methylation heterogeneity in MM. (a) Epipolymorphism levels as a function of the average DNA methylation at each locus in NPC (green) and diagnosis (orange) samples. The maximal epipolymorphism (continuous black line) and the bimodal epipolymorphism (dotted black line) are represented. (b) Scatterplots showing loci organization for NPC2 and M#19 diagnosis samples. Each point corresponds to a locus of 4 adjacent CpGs. On the left, each point is color coded according to the density of the surrounding points; on the right, each point is color coded according to its epipolymorphism level. (c) Mean proportion of discordant reads per sample.



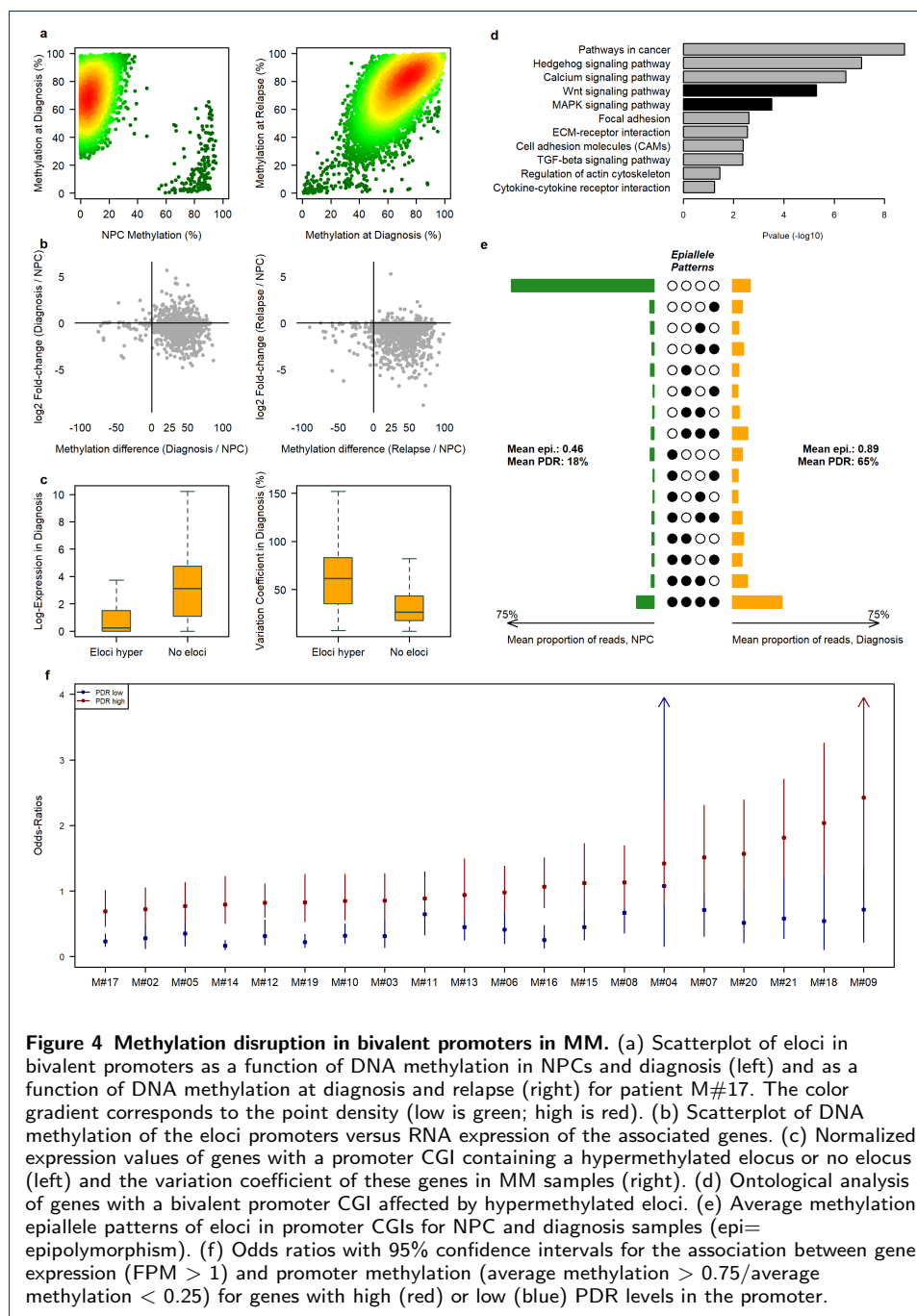


Figure 4 Methylation disruption in bivalent promoters in MM. (a) Scatterplot of eloci in bivalent promoters as a function of DNA methylation in NPCs and diagnosis (left) and as a function of DNA methylation at diagnosis and relapse (right) for patient M#17. The color gradient corresponds to the point density (low is green; high is red). (b) Scatterplot of DNA methylation of the eloci promoters versus RNA expression of the associated genes. (c) Normalized expression values of genes with a promoter CGI containing a hypermethylated elocus or no elocus (left) and the variation coefficient of these genes in MM samples (right). (d) Ontological analysis of genes with a bivalent promoter CGI affected by hypermethylated eloci. (e) Average methylation epiallele patterns of eloci in promoter CGIs for NPC and diagnosis samples (epi = epipolymorphism). (f) Odds ratios with 95% confidence intervals for the association between gene expression (FPM > 1) and promoter methylation (average methylation > 0.75/average methylation < 0.25) for genes with high (red) or low (blue) PDR levels in the promoter.

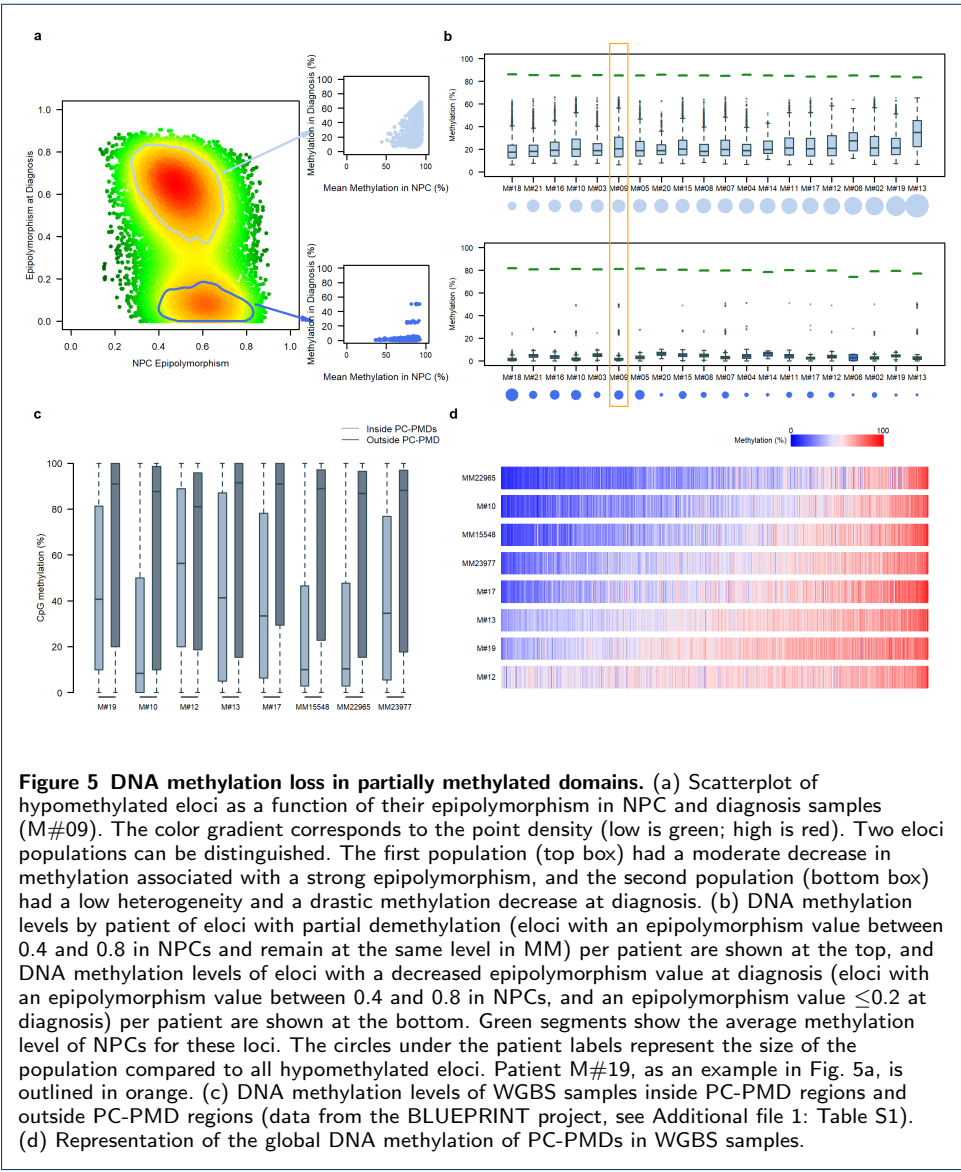
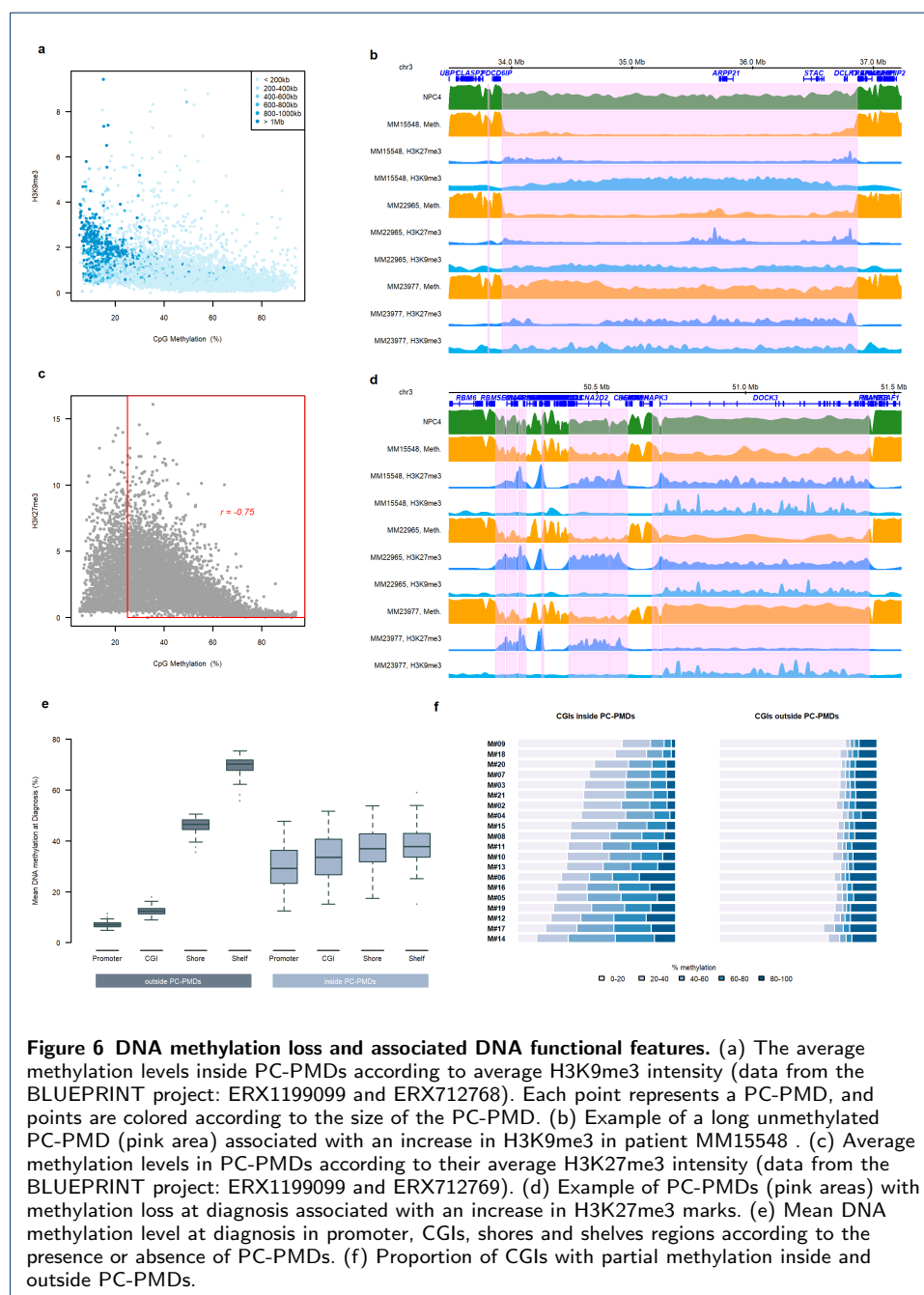
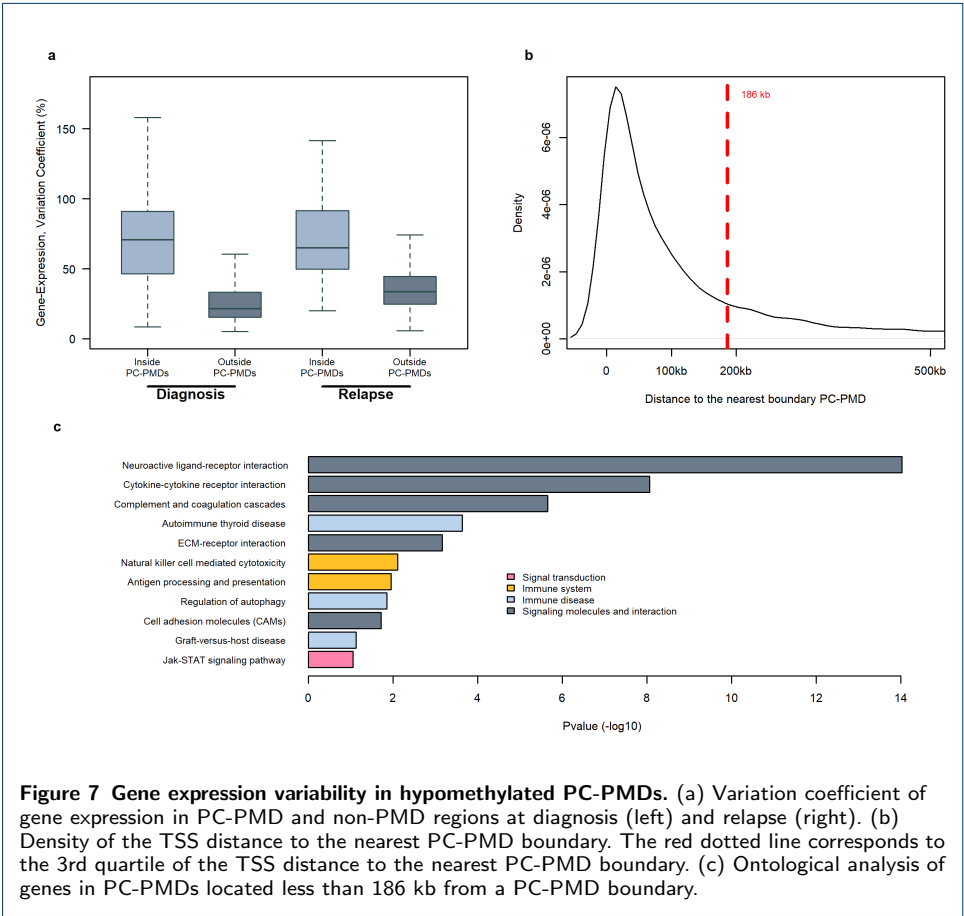
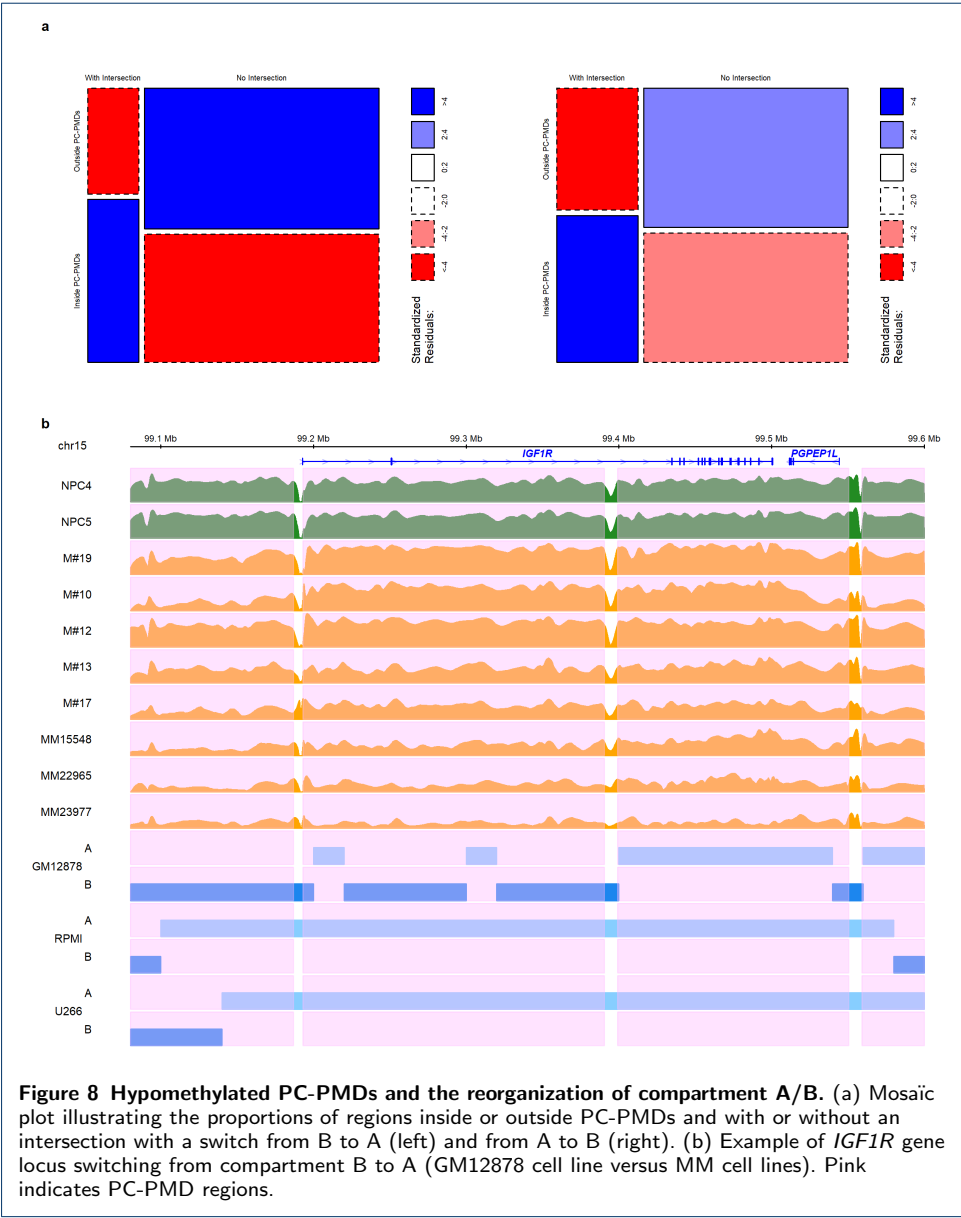


Figure 5 DNA methylation loss in partially methylated domains. (a) Scatterplot of hypomethylated loci as a function of their epipolymorphism in NPC and diagnosis samples (M#09). The color gradient corresponds to the point density (low is green; high is red). Two loci populations can be distinguished. The first population (top box) had a moderate decrease in methylation associated with a strong epipolymorphism, and the second population (bottom box) had a low heterogeneity and a drastic methylation decrease at diagnosis. (b) DNA methylation levels by patient of loci with partial demethylation (loci with an epipolymorphism value between 0.4 and 0.8 in NPCs and remain at the same level in MM) per patient are shown at the top, and DNA methylation levels of loci with a decreased epipolymorphism value at diagnosis (loci with an epipolymorphism value between 0.4 and 0.8 in NPCs, and an epipolymorphism value ≤ 0.2 at diagnosis) per patient are shown at the bottom. Green segments show the average methylation level of NPCs for these loci. The circles under the patient labels represent the size of the population compared to all hypomethylated loci. Patient M#19, as an example in Fig. 5a, is outlined in orange. (c) DNA methylation levels of WGBS samples inside PC-PMD regions and outside PC-PMD regions (data from the BLUEPRINT project, see Additional file 1: Table S1). (d) Representation of the global DNA methylation of PC-PMDs in WGBS samples.







Additional Files

Additional file 1 — Table S1

Data resource.

Additional file 2 — Table S2

Distribution of CpG sites within 100 bp genomic units in the human genome.

Additional file 3 — Figures S1–S25

Supplementary figures.

Additional file 4 — Table S3

Description of sequencing metrics for eRRBS data.

Additional file 5 — Table S4

Average number of CpGs covered by eRRBS with read depths greater than 10X and 60X, given by genomic features.

Additional file 6 — Table S5

Odds ratios and 95% confidence intervals for association between gene expression and promoter methylation according to low or high PDR levels in the promoter.

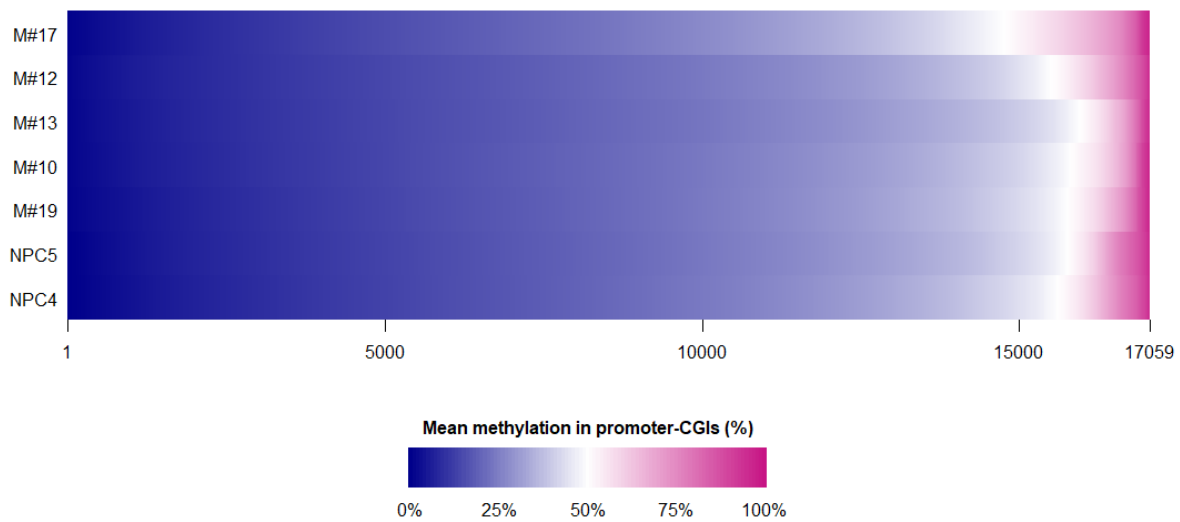


Figure S1 | Mean methylation levels in promoter-CGIs across MM patients and NPCs.

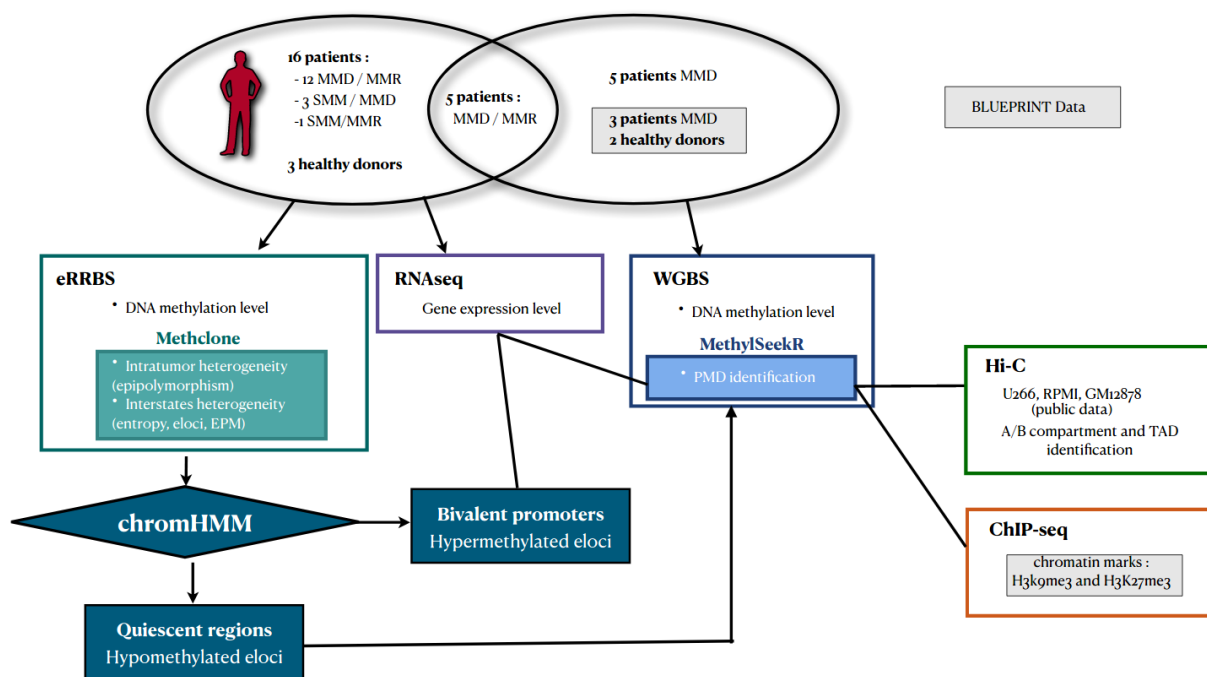


Figure S2 | Diagram of the analytical process.

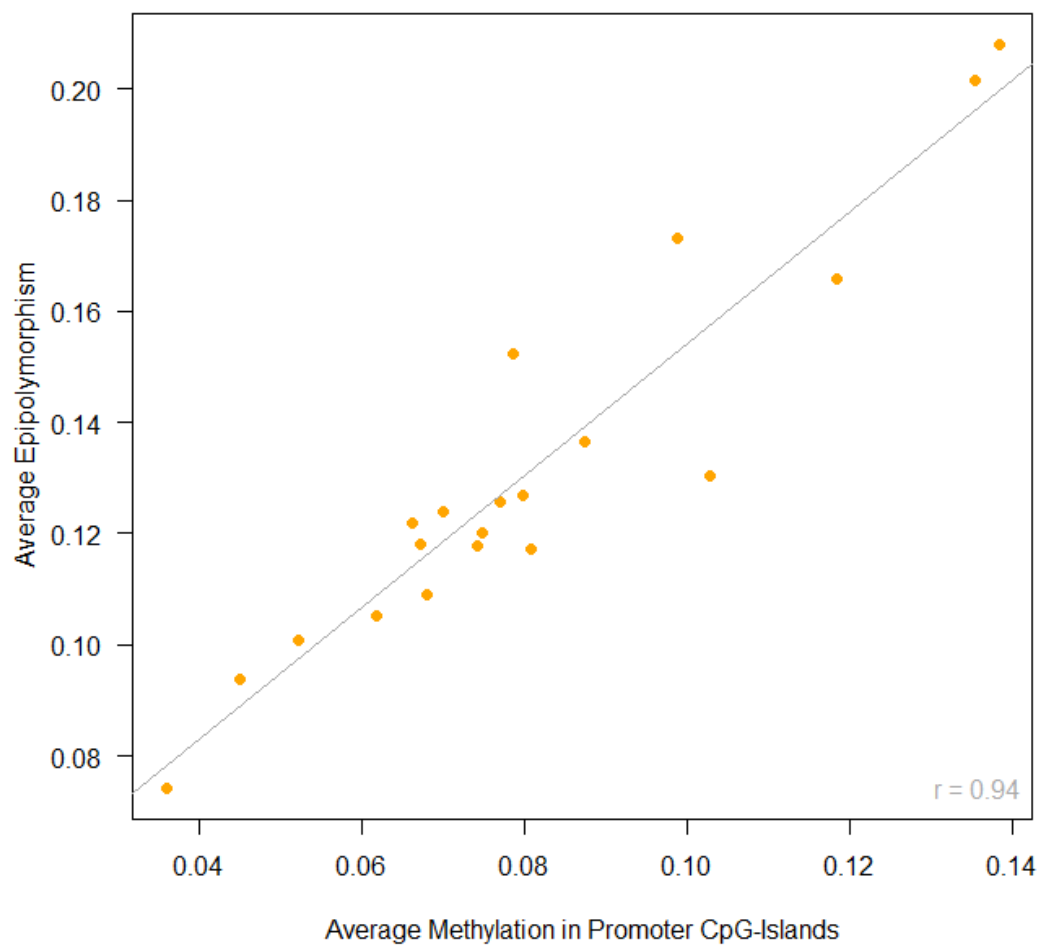
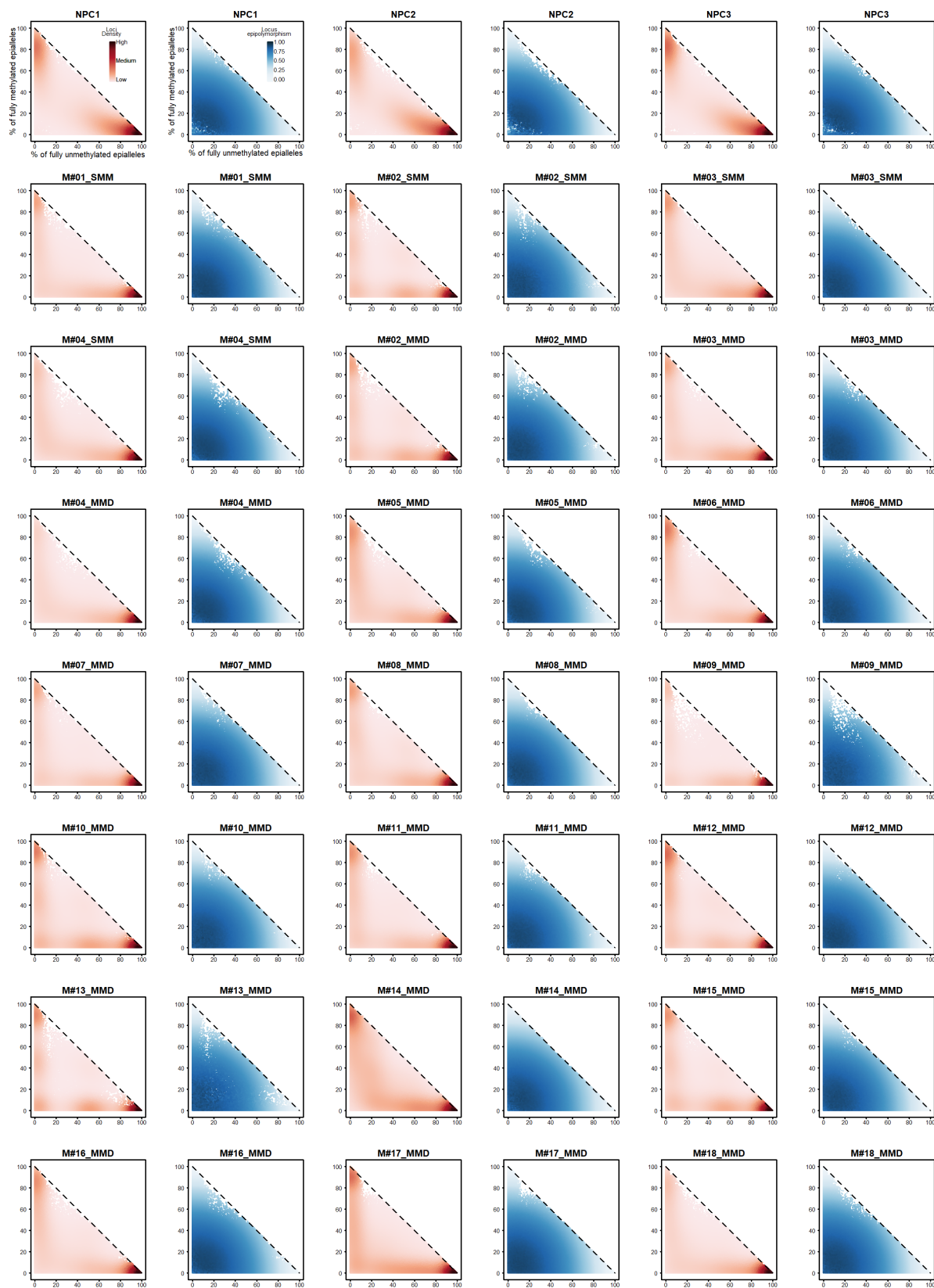


Figure S5 | Correlation between average promoter CGI methylation and epipolymorphism at diagnosis per sample.



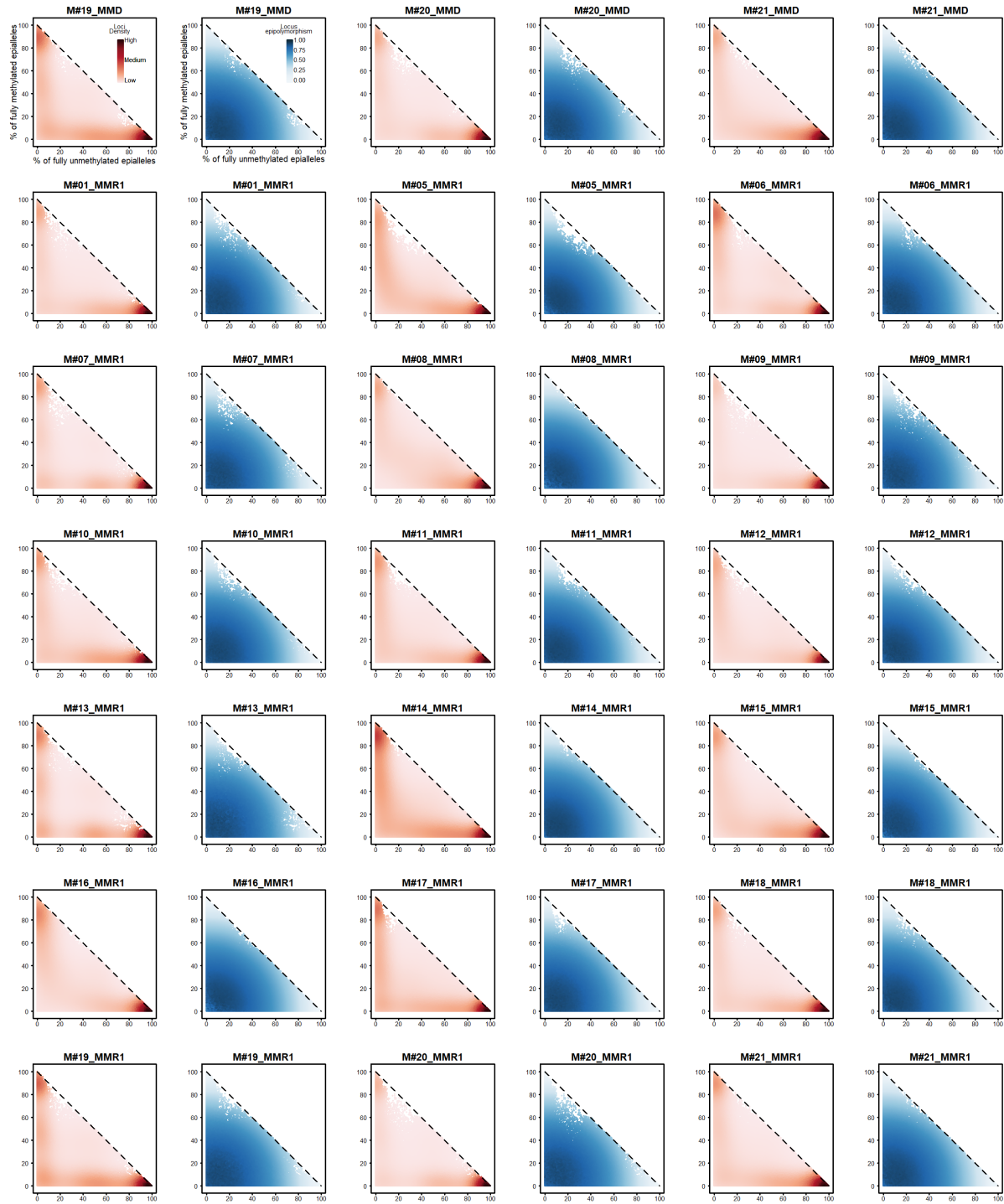


Figure S6 | Scatterplots showing loci organization as in Figure 2b for NPCs and all paired samples (SMM/Diagnosis/Relapse). Each point corresponds to a locus of 4 adjacent CpGs. Shown on the left is a representation where each point is color coded according to the density of the surrounding points; on the right each point is color coded according to its level of epipolymorphism.

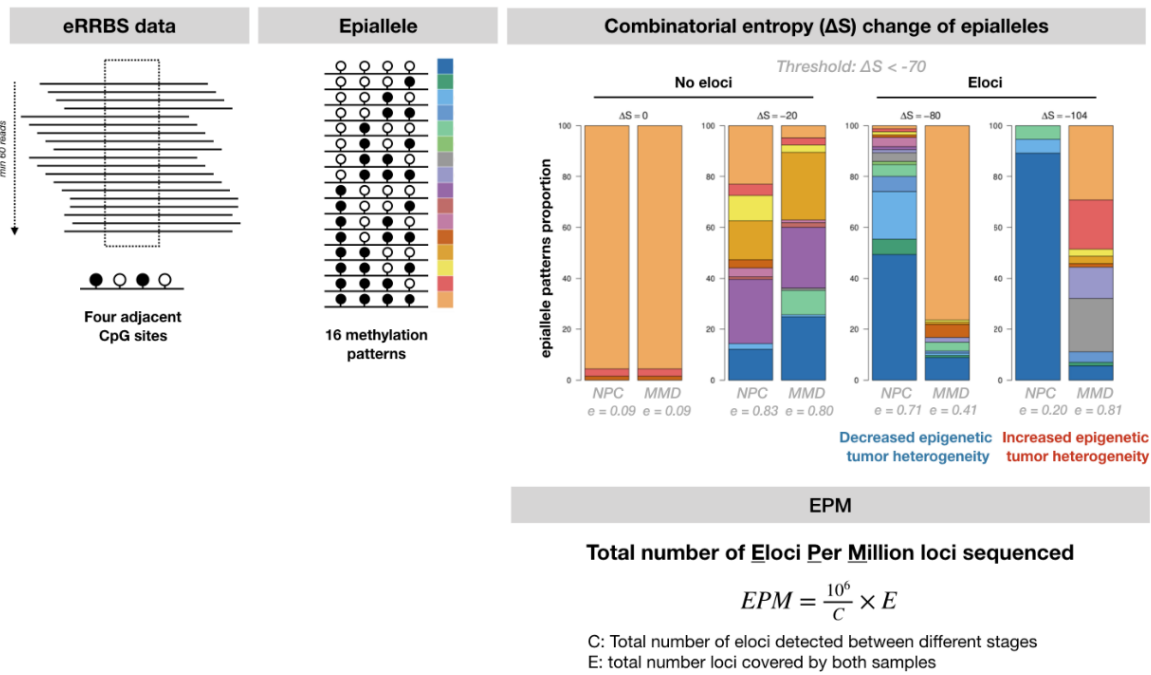


Figure S7 | Detection of epiallelic changes. From a combinatorial entropy value (threshold: $\Delta S < -70$), Methclone determines loci with a significant change in epiallelic composition (= eloci). Filled black circle: methylated CpG; empty circle: unmethylated CpG.

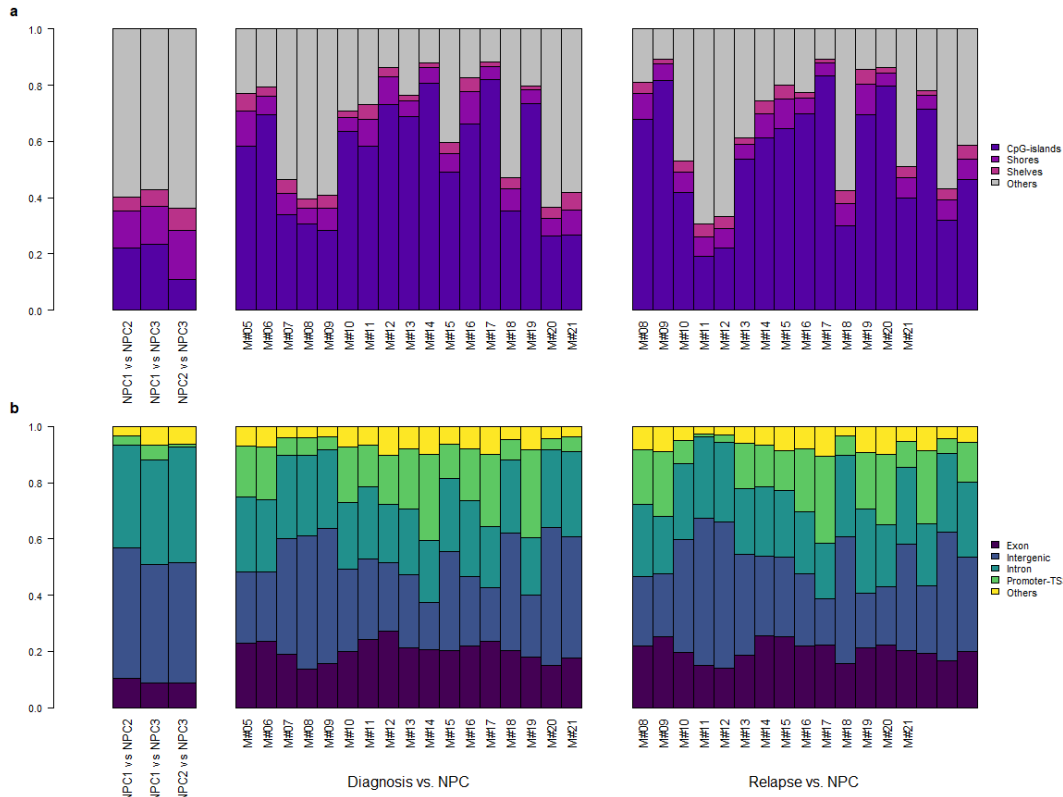


Figure S8 | Genomic distribution of NPCs versus NPCs eloci (to the left), NPCs versus diagnosis eloci (at the middle) and NPCs versus relapse eloci (to the right); in CGIs and adjacent regions (a) and in the main genomic regions: promoter TSS, exons, introns and intergenic regions (b). The distribution of NPCs versus NPCs eloci is shown as a control.

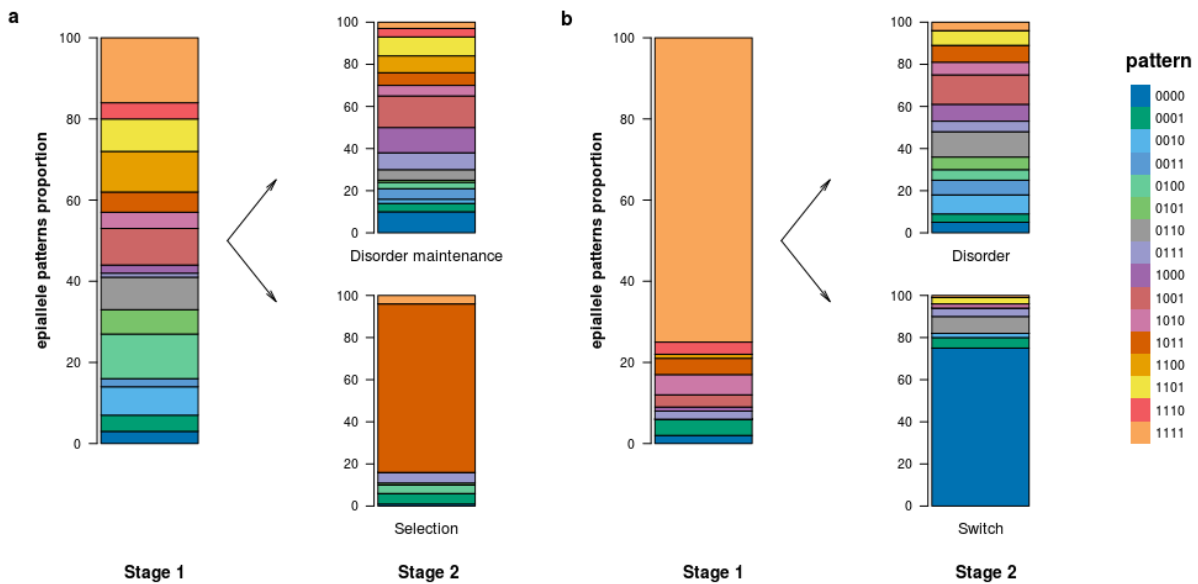


Figure S9 | Illustration of the four extreme changes in epiallele patterns between two stages: (a) Stage 1, heterogeneous epialleles; stage 2 (top), a similar heterogeneous pattern; stage 2 (bottom), selection of one pattern. (b) Stage 1, major fully methylated pattern; stage 2 (top), heterogeneous pattern; stage 2 (bottom), major fully unmethylated pattern. Legend for methylation patterns: 1 = methylated CpG and 0 = unmethylated CpG.

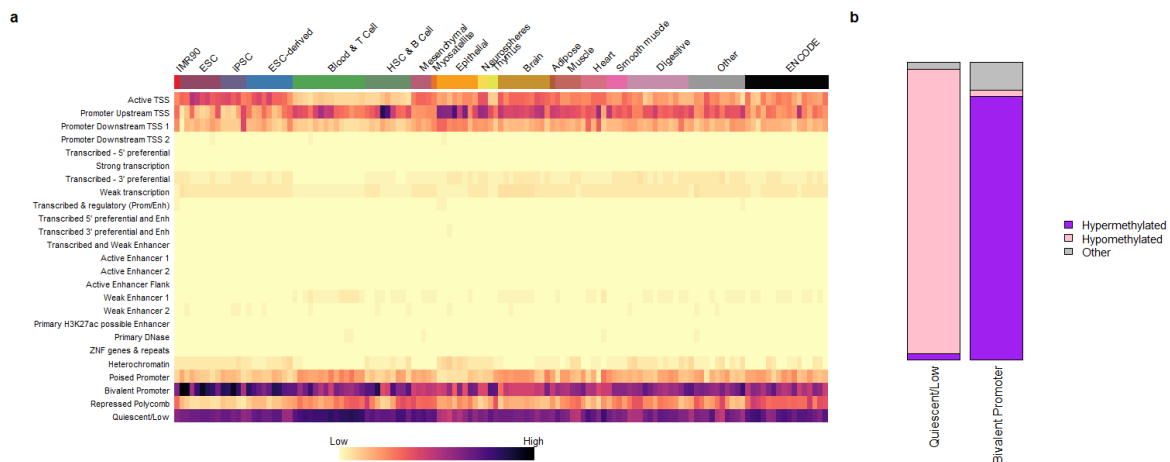


Figure S10 | Eloci (NPC vs diagnosis) annotation according to the chromatin states defined by ChromHMM. (a) Eloci annotation between NPC and MM at diagnosis samples according to the 25 chromatin states defined by ChromHMM for 127 epigenomes available on UCSC. (b) Proportion of hypo- and hypermethylated eloci in bivalent promoter and quiescent regions defined by ChromHMM.

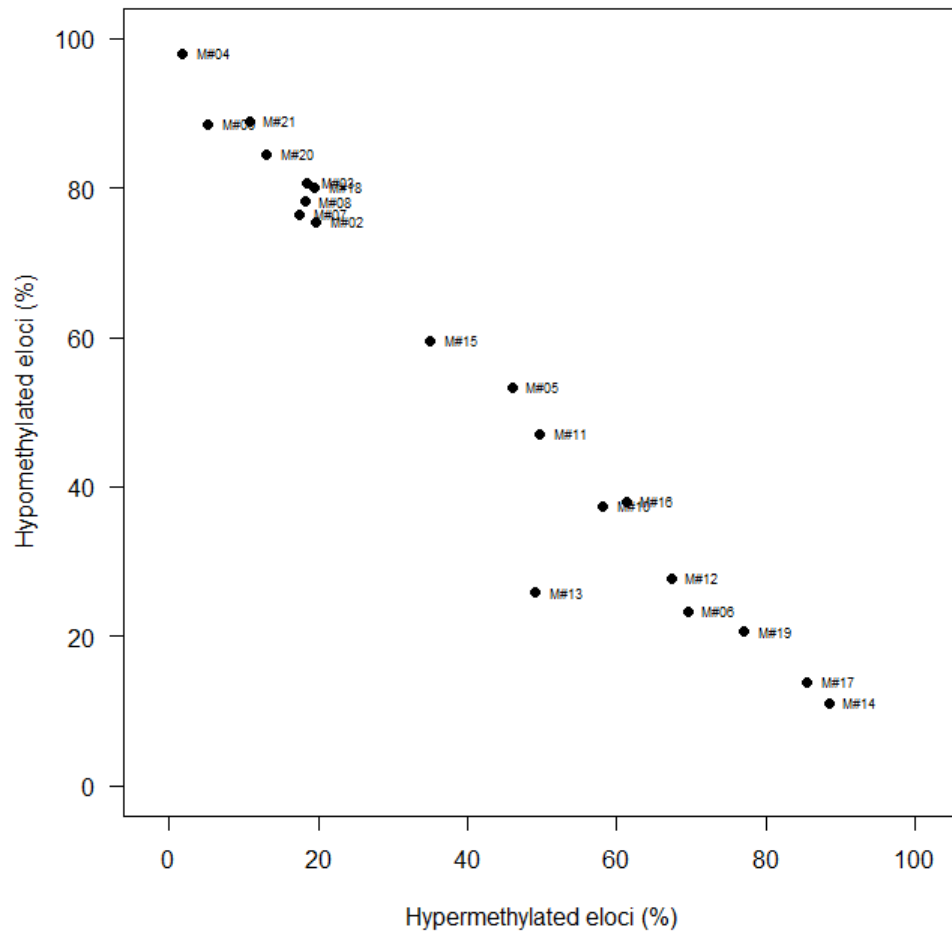


Figure S11 | Distribution of MM samples at diagnosis according to their percentage of hypo- and hypermethylated eloci (NPC vs diagnosis).

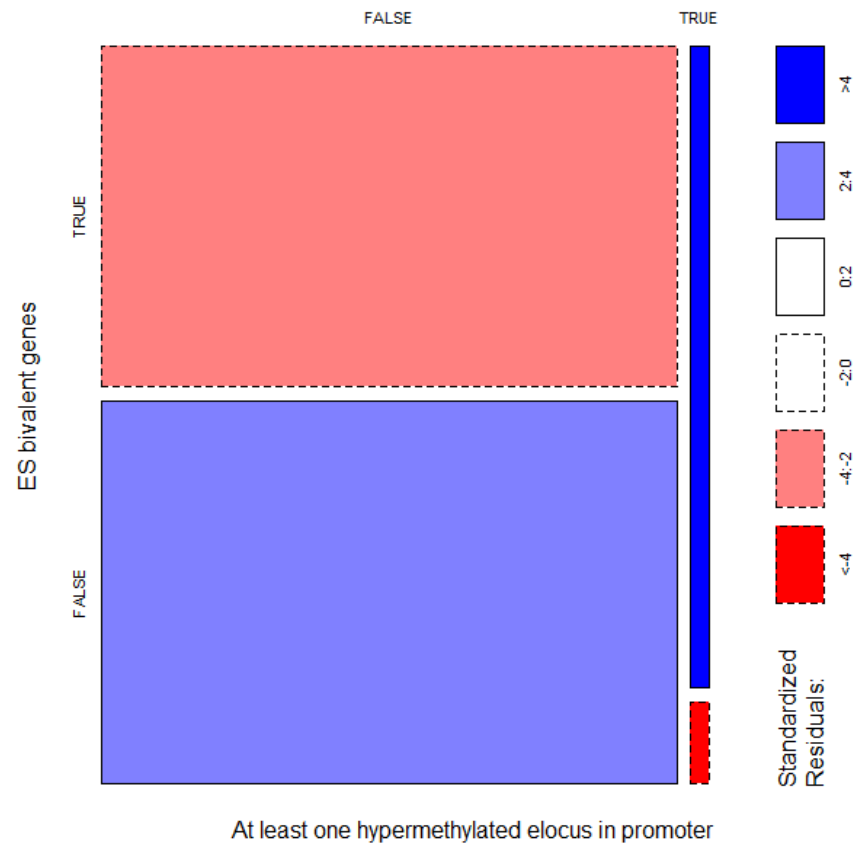


Figure S12 | Enrichment of hypermethylated eloci in the bivalent promoters of embryonic stem cells.

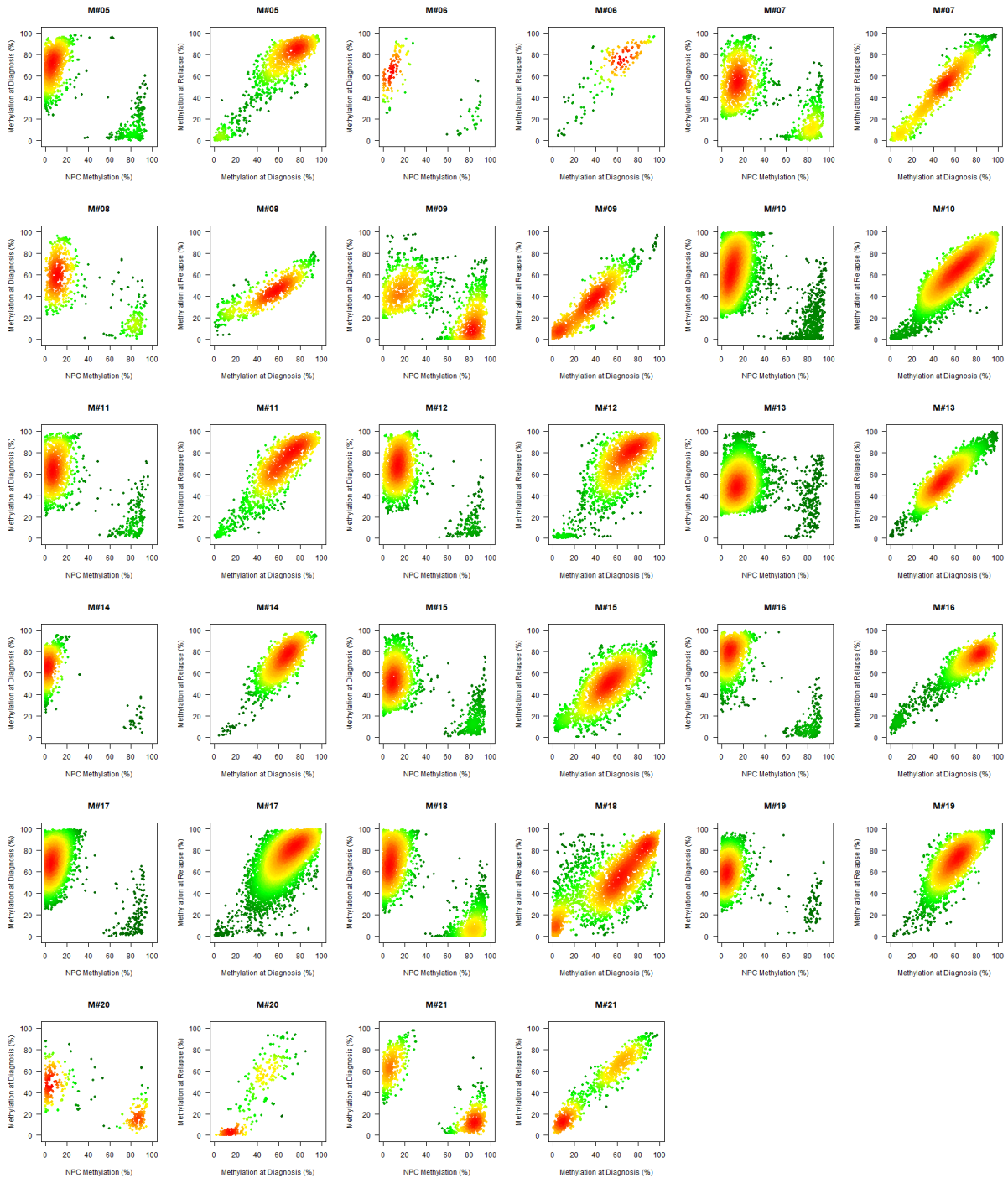


Figure S13 | Methylation disruption in bivalent promoters in MM. Scatterplot of eloci in bivalent promoters as a function of DNA methylation in NPCs and diagnosis samples (left figure) and as a function of DNA methylation at diagnosis and relapse (right figure) for all patients. The color gradient corresponds to the point density (low is green; high is red).

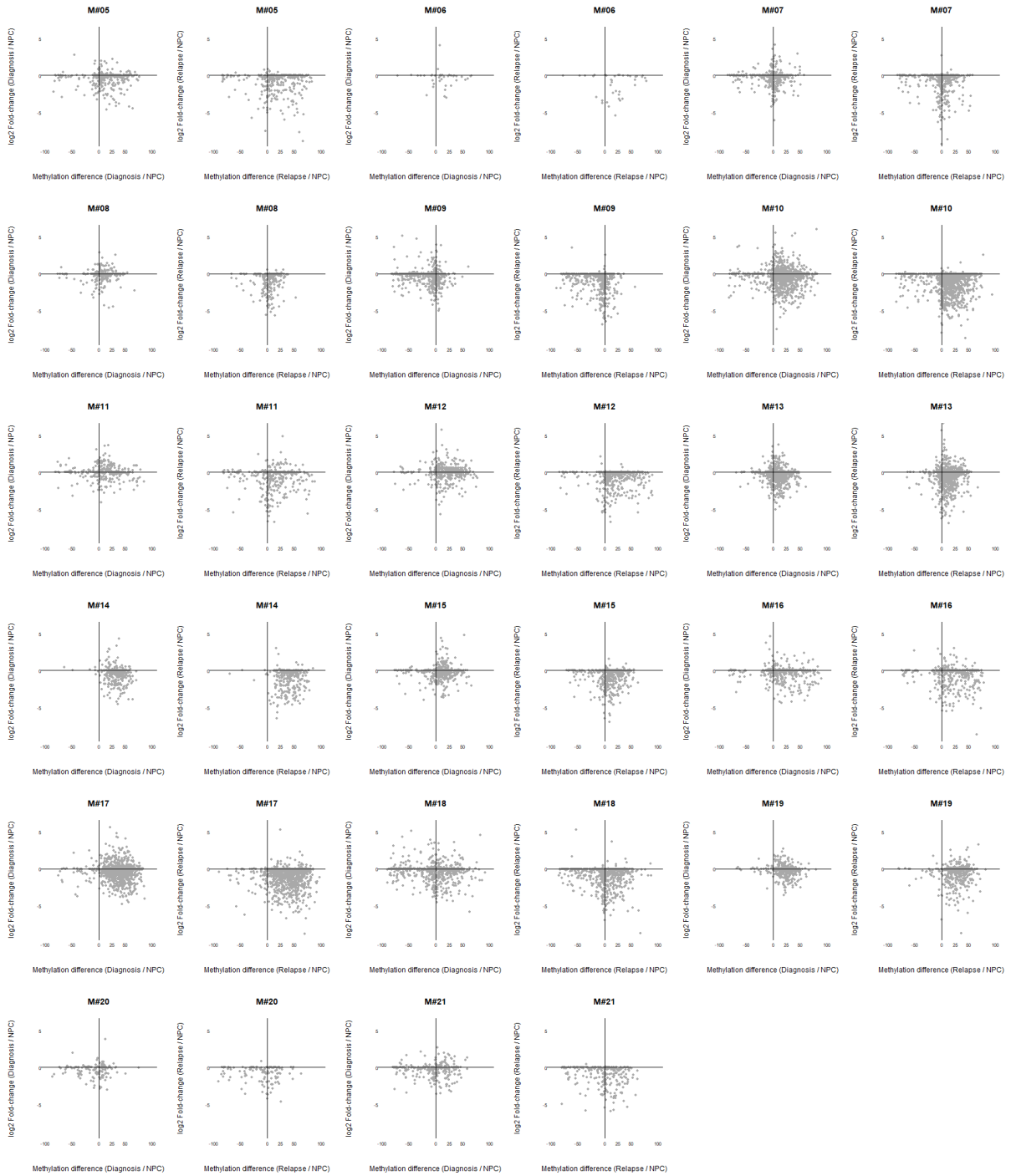


Figure S14 | Expression evolution of genes with eloci in bivalent promoters. Scatterplot of genes depending on the methylation level in the promoter and expression levels (left: diagnosis/NPC; right: relapse/NPC).

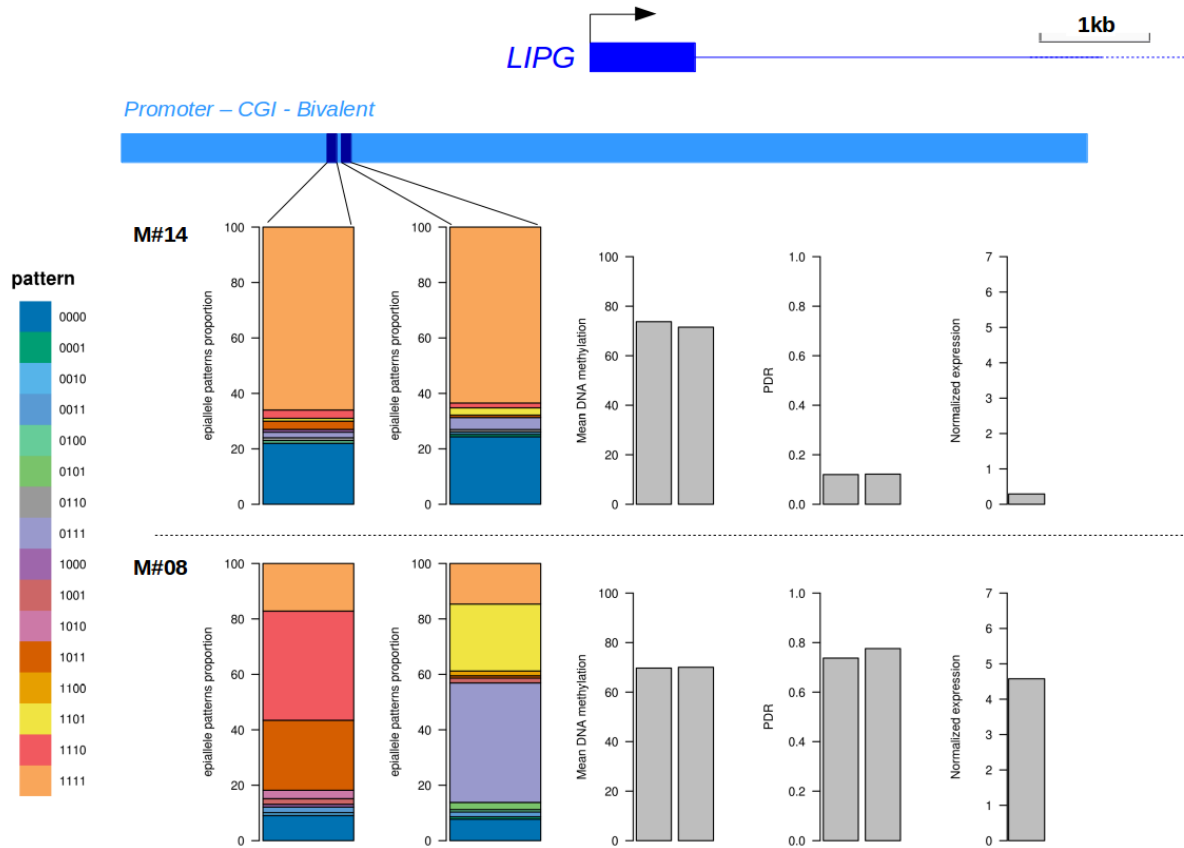


Figure S15 | Example of the decoupling relationship between promoter methylation and gene expression. The promoter region of *LIPG* shows comparable methylation levels in two samples (M#14 and M#08) but different PDR and expression levels (right). Epiallele patterns of two loci are shown in the illustration (left). Legend for methylation patterns: 1 = methylated CpG and 0 = unmethylated CpG.

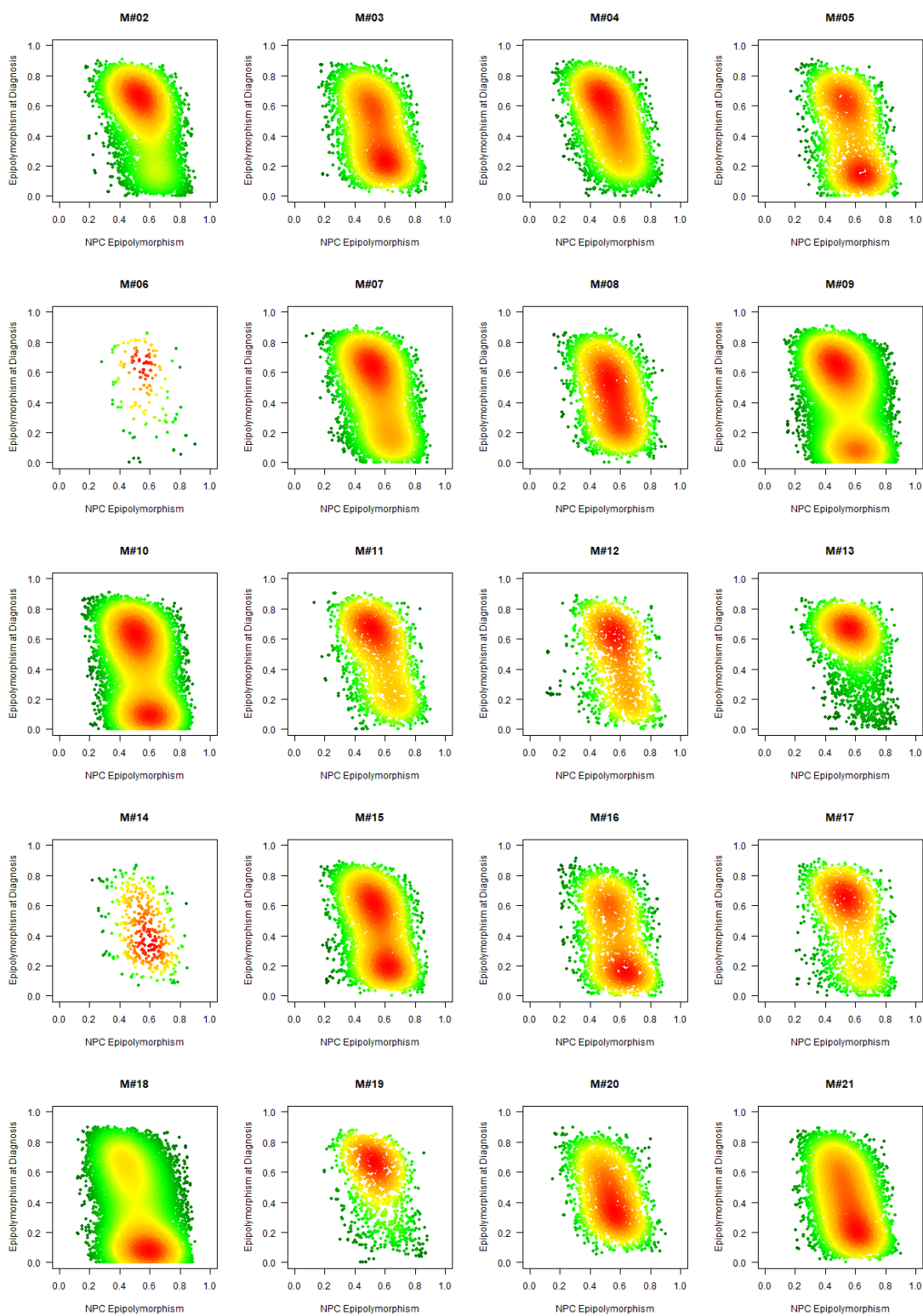


Figure S16 | Characterization of hypomethylated eloci epipolymorphism evolution. Scatterplot of eloci as a function of their epipolymorphism in NPC and diagnosis samples for all patients. The color gradient corresponds to the point density (low is green; high is red).

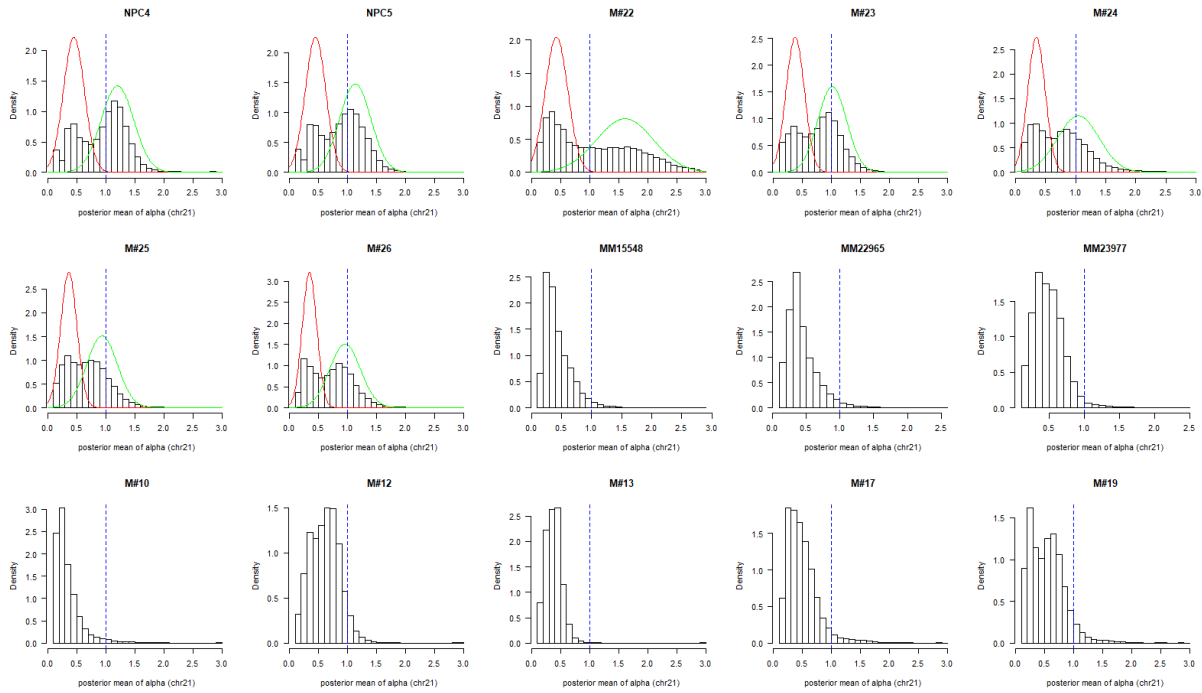


Figure S17 | Distribution of the α -value for chromosome 21. This distribution is similar for all chromosomes. The α -value characterizes the distribution of methylation levels in genomic windows of 100 CpGs. If the distribution of α -values is bimodal or has a large fraction of α value greater than or equal to 1 (blue dotted line), then this is evidence of the presence of PMDs. In the presence of a bimodal distribution, it is possible to calculate the PMDs on the genome with a hidden Markov model, whose two adjusted Gaussian distributions are represented in green and red.

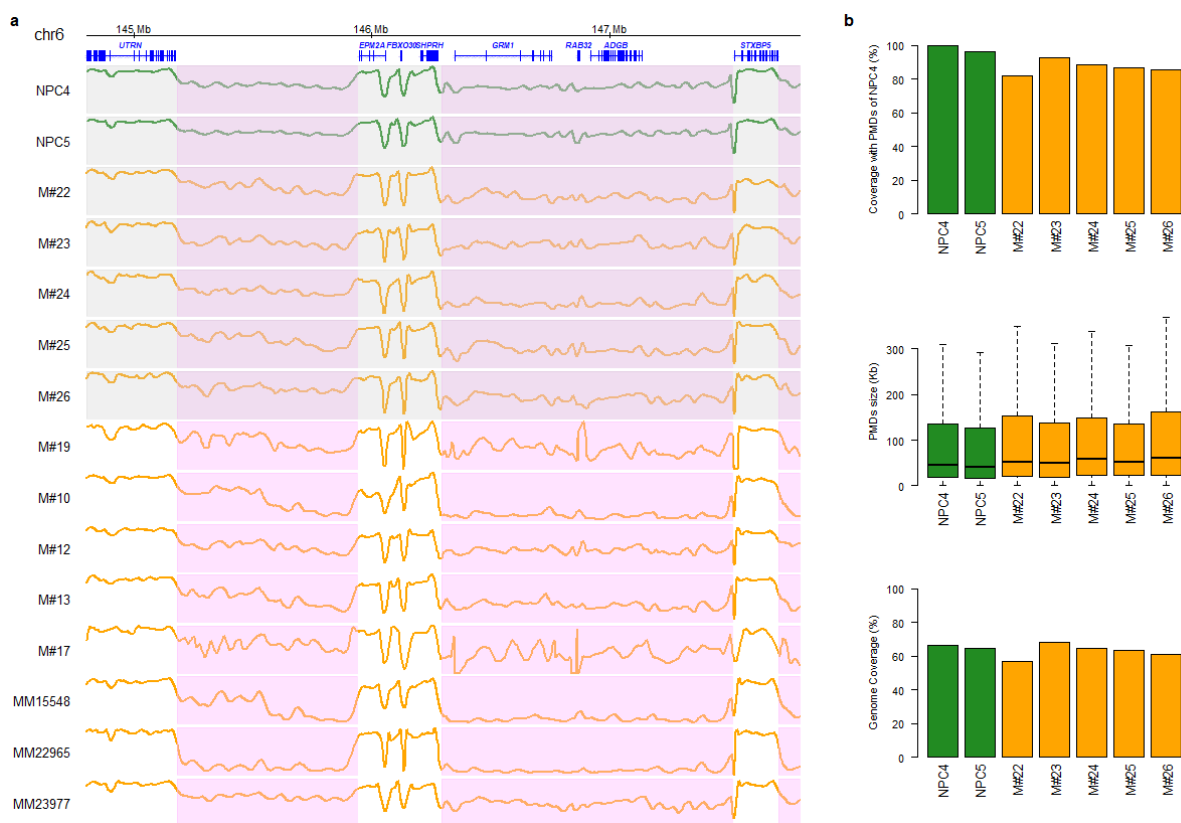


Figure S18 | Comparison of NPC and MM PMDs. (a) WGBS DNA methylation visualization in a genomic region encompassing two PMDs (pink area) in sample NPC4. Green lanes correspond to NPC methylation profiles, and orange lanes correspond to MM methylation profiles at diagnosis. The shaded profiles correspond to samples in which PMDs could be detected. (b) General characteristics of PMDs.

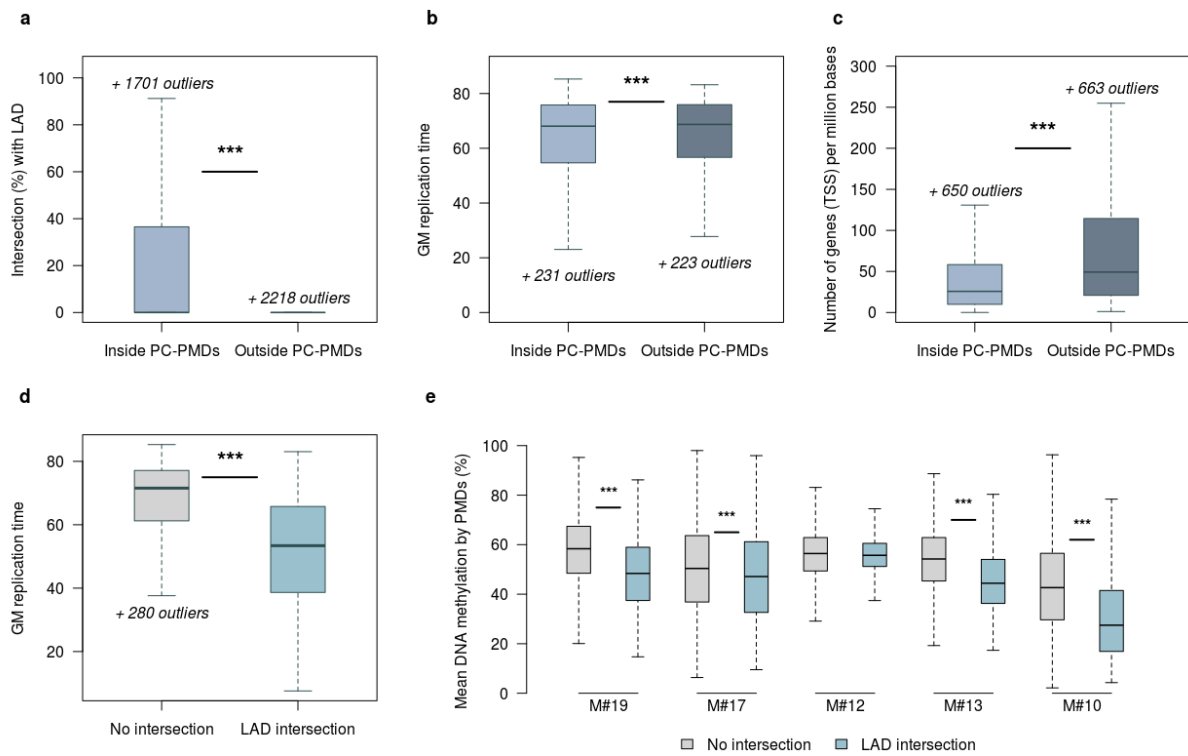


Figure S19 | Characteristics of PC-PMDs. (a) Percentage of intersections between LADs [30] and regions inside and outside PC-PMDs. (b) Replication timing of the GM12878 cell line (ENCODE data) inside and outside PC-PMDs. (c) Number of genes per million bases inside and outside PC-PMDs. (d) Replication timing of the GM12878 cell line in PC-PMDs intersecting (or not) with an LAD. (e) Mean DNA methylation level in PC-PMDs intersecting (or not) with an LAD per patient.

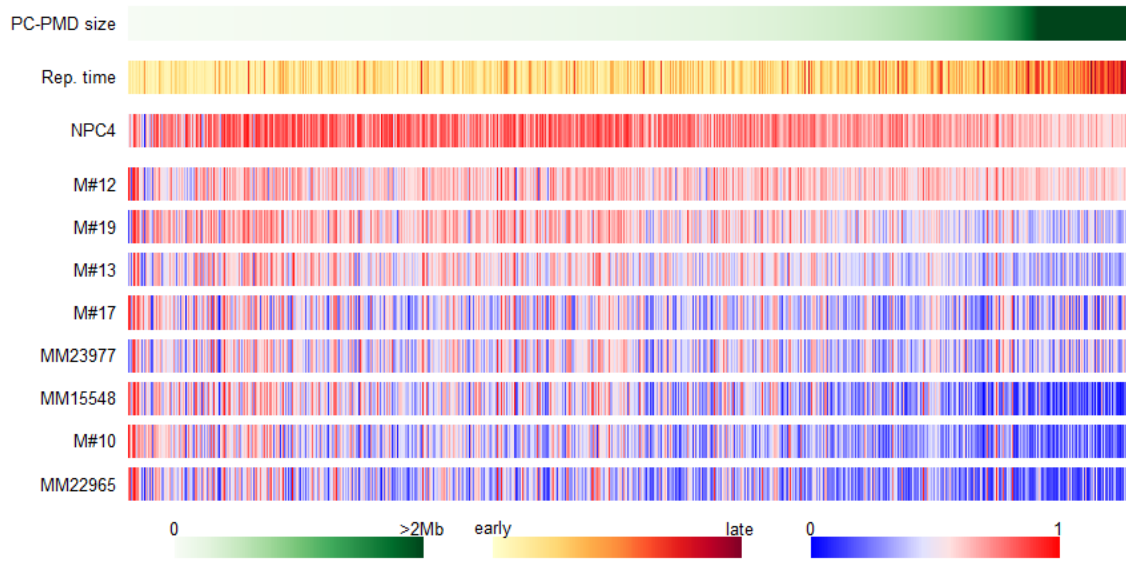


Figure S20 | Heatmap of PC-PMDs, sorted by increasing size, associated with GM12878 replication timing (ENCODE data) and DNA methylation levels in WGBS samples (NPC and diagnosis).

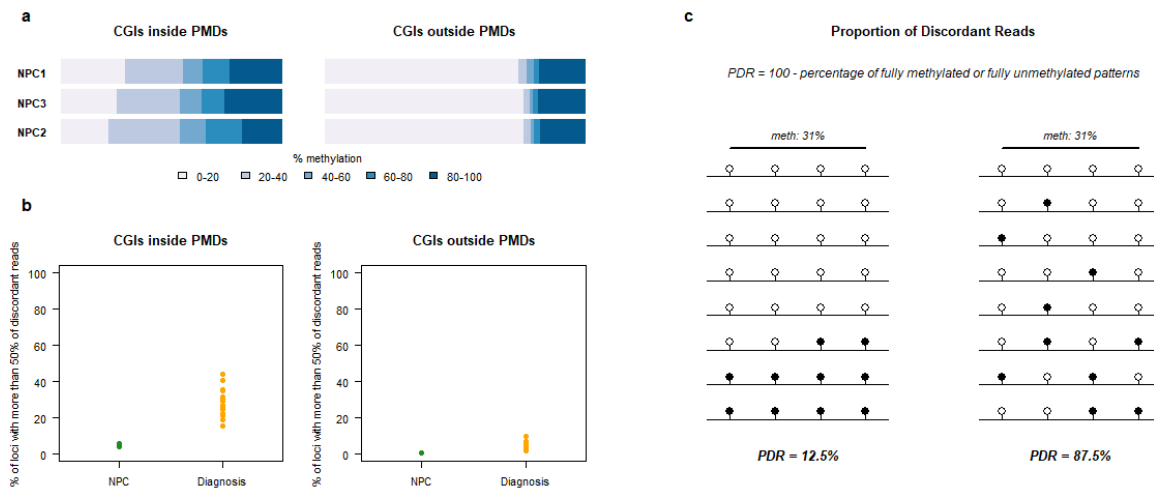


Figure S21 | Homogeneous CGI methylation perturbation in NPCs (a) Proportion of CGIs with partial methylation inside and outside PC-PMDs (b) Proportion of loci with more than 50% discordant reads inside and outside PC-PMDs. (c) PDR definition; for the same average level of methylation, different PDR values are possible.

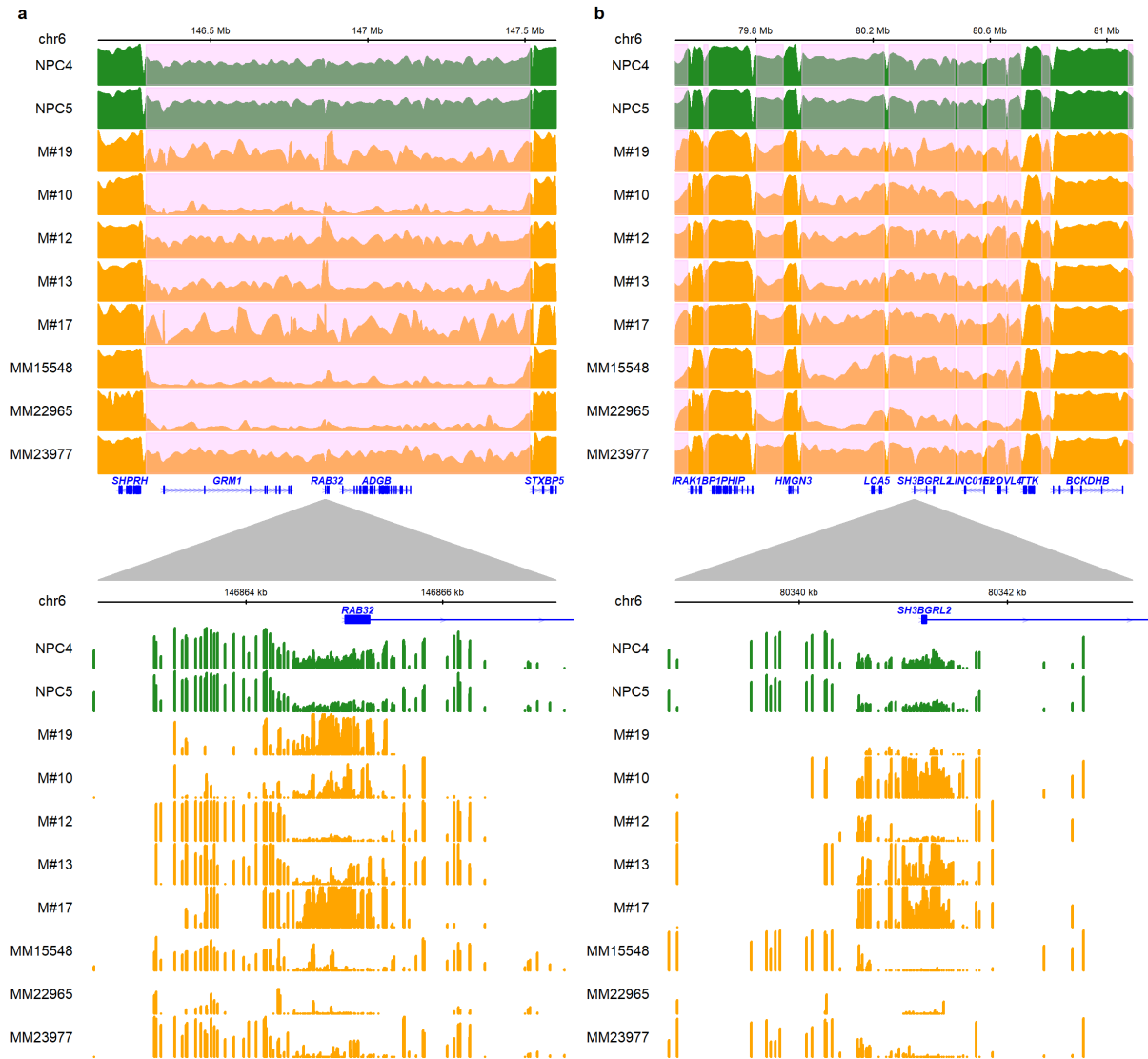


Figure S22 | Examples of genes with bivalent promoter CGIs with disrupted DNA methylation regions embedded within large partially methylated regions (pink area): *RAB32* (a) and *SH3BGRL2* (b).

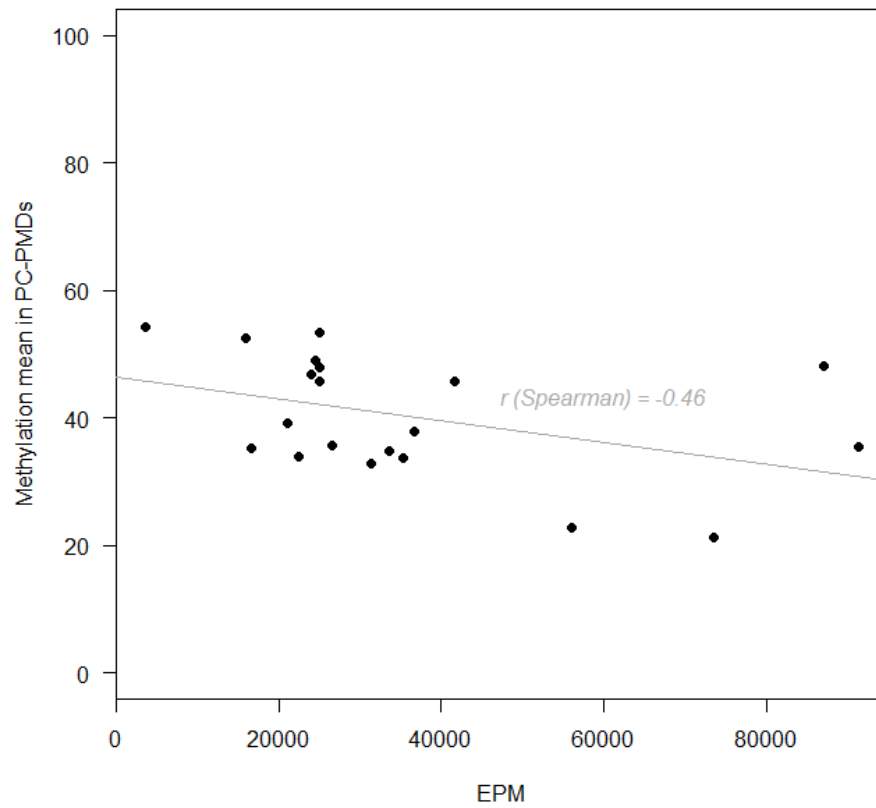


Figure S23 | Correlation between EPM and DNA methylation in PC-PMDs.

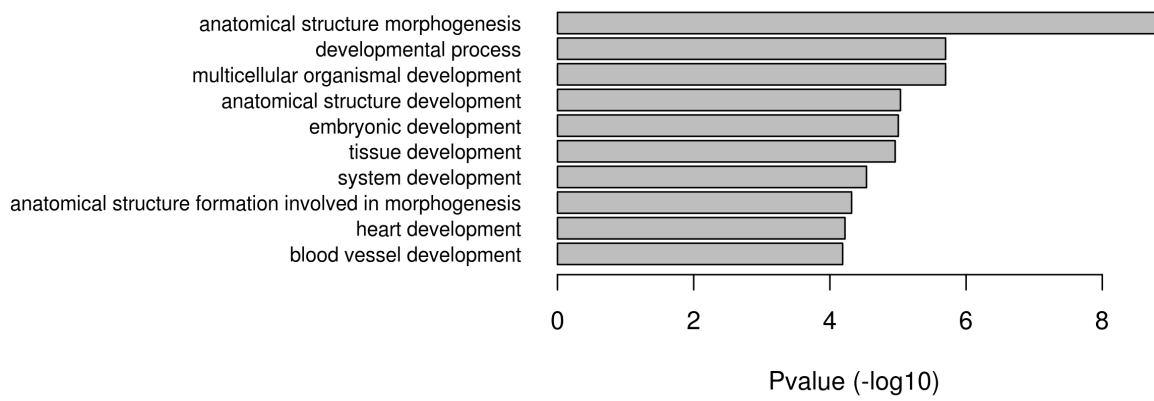


Figure S24 | Gene set enrichment analysis of genes from the list in the supplementary Table 7 [42].

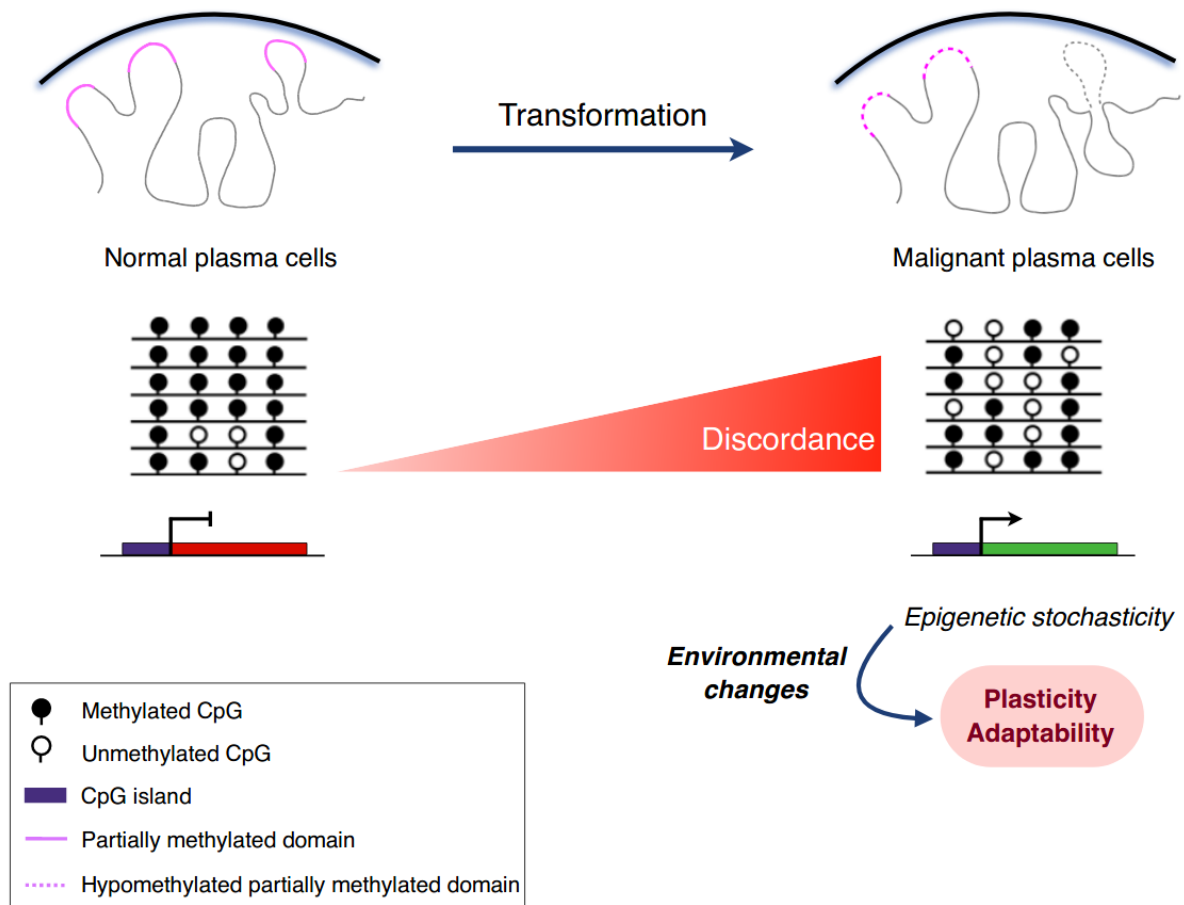


Figure S25 | A proposed model of the role of PMD instability in MM onset and development. Malignant transformation is accompanied by increased hypomethylation of PMDs which contributes to epigenetic, transcriptomic and 3D organization heterogeneity across patients. The lack of accurate global DNA methylation maintenance also drives inpatient DNA methylation heterogeneity which contributes to a decoupling relationship between promoter methylation and gene expression (genes with methylated promoter but not repressed) and transcriptomic variability. We therefore hypothesize that this epigenetic stochasticity could provide a selective advantage to the tumor cell by increasing its plasticity and adaptation to environmental changes.

Table S1

ID patient	Sexe	Collection date	Age at collection	Stage of the disease	Frontline treatment	Date of death/Date of last news	% Purif Tumor Fraction	eRRBS	WGBS	RNAseq
M#01_SMM	F	8/19/2008	69	SMM			100	y	n	n
M#01_R1	F	1/12/2012	72	Relapse*		9/3/2012	100	y	n	y
M#02_SMM	F	4/20/2010	54	SMM			100	y	n	n
M#02_D	F	2/20/2012	56	Diagnosis		12/24/2013	99	y	n	y
M#03_SMM	M	9/9/2010	64	SMM			96	y	n	n
M#03_D	M	2/23/2011	64	Diagnosis		10/17/2011	94	y	n	y
M#04_SMM	F	11/3/2010	64	SMM			99	y	n	n
M#04_D	F	1/18/2011	64	Diagnosis		6/10/2013	98	y	n	y
M#05_D	M	3/27/2007	55	Diagnosis	VAD		92	y	n	y
M#05_R1	M	12/26/2007	55	Relapse		1/15/2008	100	y	n	y
M#06_D	F	6/4/2007	64	Diagnosis	VAD		95	y	n	y
M#06_R1	F	2/21/2012	68	Relapse		7/26/2013	94	y	n	y
M#07_D	F	12/6/2007	58	Diagnosis	VTD		97	y	n	y
M#07_R1	F	1/20/2009	59	Relapse		11/22/2010	93	y	n	y
M#08_D	F	12/17/2007	38	Diagnosis	VD		97	y	n	y
M#08_R1	F	12/29/2010	41	Relapse		11/7/2013	98	y	n	y
M#09_D	M	1/7/2008	52	Diagnosis	VD		99	y	n	y
M#09_R1	M	10/14/2010	55	Relapse		9/18/2013	98	y	n	y
M#10_D	M	4/22/2008	55	Diagnosis	VAD		100	y	y	y
M#10_R1	M	6/10/2009	56	Relapse		5/21/2010	100	y	n	y
M#11_D	M	10/14/2008	60	Diagnosis	VD		100	y	n	y
M#11_R1	M	3/16/2010	62	Relapse		8/26/2011	95	y	n	y
M#12_D	F	12/11/2008	57	Diagnosis	VD		99	y	y	y
M#12_R1	F	12/2/2010	59	Relapse		1/21/2011	99	y	n	y
M#13_D	M	1/5/2009	46	Diagnosis	VD		100	y	y	y
M#13_R1	M	9/22/2010	48	Relapse		10/25/2010	99	y	n	y
M#14_D	F	1/29/2009	83	Diagnosis	MPT		87	y	n	y
M#14_R1	F	4/2/2012	86	Relapse		7/1/2013	100	y	n	y
M#15_D	M	11/3/2009	72	Diagnosis	VD		97	y	n	y
M#15_R1	M	1/16/2012	74	Relapse		4/9/2013	99	y	n	y
M#16_D	F	7/6/2010	68	Diagnosis	LD		100	y	n	y
M#16_R1	F	2/13/2012	69	Relapse		1/10/2014	94	y	n	y
M#17_D	F	11/4/2010	66	Diagnosis	VTD		100	y	y	y
M#17_R1	F	1/11/2012	67	Relapse		10/28/2012	100	y	n	y
M#18_D	F	11/25/2010	50	Diagnosis	VTD		97	y	n	y
M#18_R1	F	5/15/2012	51	Relapse		4/3/2013	100	y	n	y
M#19_D	M	9/27/2004	54	Diagnosis	VAD		99	y	y	y
M#19_R1	M	10/22/2008	58	Relapse			99	y	n	y
M#20_D	M	10/14/2008	73	Diagnosis	MPT		99	y	n	y
M#20_R1	M	11/4/2010	75	Relapse			100	y	n	y
M#21_D	M	3/10/2009	57	Diagnosis	VD		NA	y	n	y
M#21_R1	M	10/12/2010	59	Relapse			98	y	n	y
M#22_D	F	7/29/2003	51	Diagnosis	VAD	10/18/2012	95	n	y	n
M#23_D	M	4/11/2007	56	Diagnosis	VAD	10/13/2011	99	n	y	n
M#24_D	F	11/4/2008	60	Diagnosis	VD	11/17/2013	90	n	y	n
M#25_D	F	9/24/2008	54	Diagnosis	VD	3/18/2016	98	n	y	n
M#26_D	F	2/8/2011	45	Diagnosis		11/7/2017	90	n	y	n

* purified myeloma cells at diagnosis (02/02/2009) were not available

VAD: vincristine-doxorubicin-dexamethasone

VTD: bortezomib-thalidomide-dexamethasone

VD: bortezomib-dexamethasone

MPT: melphalan-prednisone-thalidomide

LD: lenalidomide-dexamethasone

Table S1

Healthy donor

ID	Sexe	Collection date	Age at collection	purity	eRRBS	RNAseq	WGBS
NPC1	F	22/03/12	64	86	y	y	n
NPC2	M	26/04/12	44	74	y	y	n
NPC3	M	29/01/13	47	82	y	y	n

Data from the BLUEPRINT Project, Accession codes

ID	WGBS	H3K9me3	H3K27me3
NPC4	ERX715130	NA	NA
NPC5	ERX301127	NA	NA
MM15548	ERX1199099	ERX712768	ERX712769
MM22965	ERX983226	ERX712721	ERX712712
MM23977	ERX1199098	ERX1007374	ERX1007375

Table S2

CpGs	NPC4*	NPC5*	M#19_D	M#10_D	M#12_D	M#13_D	M#17_D
1	7 412 460	7 496 640	7 731 487	7 643 814	7 654 704	7 688 824	7 672 711
2	3 088 267	3 101 886	3 253 559	3 227 908	3 234 065	3 245 044	3 235 723
3	1 338 245	1 339 674	1 424 619	1 418 547	1 420 231	1 424 339	1 419 027
4	637 362	637 073	688 628	686 982	687 379	689 244	686 564
5	318 966	318 365	351 238	350 494	350 759	351 659	350 345
6	167 601	167 402	188 929	188 552	188 737	189 334	188 376
7	92 601	92 586	107 627	107 452	107 627	108 002	107 375
8	53 738	53 734	64 525	64 479	64 732	64 941	64 407
9	34 346	33 894	40 811	40 923	41 050	41 115	40 838
10	23 846	23 677	27 478	27 540	27 543	27 653	27 452
11	17 892	17 855	20 264	20 314	20 337	20 375	20 239
12	14 162	14 076	15 961	15 978	15 981	15 988	15 942
13	10 974	10 887	12 617	12 634	12 644	12 638	12 612
14	8 270	8 367	10 208	10 228	10 242	10 251	10 237
>15	16 591	12 819	24 813	24 994	25 041	25 047	24 872

* indicates BLUEPRINT data

Table S3

ID patient	reads before trimming	reads after trimming	Uniquely aligned read count	Mapping efficiency	number of CpG >10X	number of CpG >60X
NPC1	55 567 561	54 027 218	34 663 936	64,20%	2 884 414	1 590 129
NPC2	49 150 116	48 209 441	30 711 256	63,70%	3 547 422	1 684 430
NPC3	51 116 901	50 362 568	32 886 482	65,30%	3 136 733	1 832 548
M#01_SMM	61 428 280	59 207 771	35 683 810	60,30%	2 523 053	1 543 912
M#02_SMM	62 504 131	57 985 061	33 884 808	58,40%	2 456 102	1 466 201
M#03_SMM	79 049 347	74 400 738	44 635 678	60,00%	2 867 303	1 691 093
M#04_SMM	55 666 213	51 991 090	30 807 429	59,30%	2 533 662	1 412 362
M#02_MMD	60 222 298	57 408 485	35 502 670	61,80%	2 787 716	1 687 971
M#03_MMD	61 453 332	57 349 448	34 221 080	59,70%	2 791 587	1 541 208
M#04_MMD	61 804 658	57 604 774	33 850 985	58,80%	2 700 718	1 498 271
M#05_MMD	54 629 384	51 885 205	33 076 074	63,70%	2 812 572	1 633 624
M#06_MMD	46 681 613	43 397 131	27 605 068	63,60%	2 699 457	1 442 788
M#07_MMD	69 469 974	65 468 086	42 120 566	64,30%	2 533 579	1 622 291
M#08_MMD	54 779 690	51 624 695	31 338 473	60,70%	2 441 333	1 456 238
M#09_MMD	41 337 053	39 151 584	23 348 924	59,60%	2 506 353	1 276 636
M#10_MMD	65 235 719	64 210 095	41 602 689	64,80%	3 491 223	2 024 878
M#11_MMD	57 417 003	56 070 614	37 123 603	66,20%	3 414 290	1 774 615
M#12_MMD	61 568 981	60 638 420	37 912 403	62,50%	3 232 145	1 872 579
M#13_MMD	57 745 574	54 768 857	32 962 736	60,20%	2 494 458	1 515 263
M#14_MMD	58 739 195	54 754 077	34 582 565	63,20%	2 471 051	1 512 614
M#15_MMD	66 892 182	63 816 252	39 587 129	62,00%	2 668 928	1 647 606
M#16_MMD	58 496 231	54 521 562	34 270 563	62,90%	2 719 319	1 494 593
M#17_MMD	57 606 174	55 598 676	35 527 994	63,90%	3 454 934	1 595 117
M#18_MMD	50 666 404	52 470 367	33 321 457	66,50%	3 356 743	1 736 359
M#19_MMD	57 118 881	53 508 937	33 090 485	61,80%	2 603 705	1 471 932
M#20_MMD	61 573 436	56 817 558	34 625 464	60,90%	2 392 815	1 453 640
M#21_MMD	60 212 170	57 101 372	35 877 159	62,80%	2 848 308	1 638 977
M#01_MMR1	52 739 274	49 708 346	29 824 514	59,90%	2 528 667	1 396 918
M#05_MMR1	50 177 115	46 500 488	28 649 068	61,60%	2 711 140	1 344 075
M#06_MMR1	53 912 442	51 174 246	33 066 619	64,60%	2 969 700	1 622 632
M#07_MMR1	59 033 111	56 016 048	35 634 947	63,60%	2 690 526	1 663 667
M#08_MMR1	58 193 387	55 234 338	33 738 536	61,10%	2 538 953	1 515 398
M#09_MMR1	51 550 876	48 716 249	28 981 223	59,50%	2 021 000	1 210 669
M#10_MMR1	57 273 549	56 007 161	36 579 637	65,30%	3 567 294	1 793 641
M#11_MMR1	52 939 643	51 879 778	34 858 886	67,20%	3 384 119	1 708 211
M#12_MMR1	59 998 081	58 026 302	36 772 293	63,40%	2 918 094	1 671 235
M#13_MMR1	68 925 567	63 526 054	37 248 398	58,60%	2 849 316	1 563 961
M#14_MMR1	61 150 842	57 482 751	35 359 012	61,50%	2 779 211	1 545 898
M#15_MMR1	56 237 707	52 979 111	32 166 568	60,70%	2 343 772	1 406 428
M#16_MMR1	60 726 394	57 950 251	36 874 091	63,60%	2 648 115	1 557 057
M#17_MMR1	53 734 759	52 470 367	33 250 786	63,40%	3 264 431	1 534 358
M#18_MMR1	51 615 146	50 791 299	33 049 147	65,10%	3 344 265	1 679 669
M#19_MMR1	53 839 425	51 066 242	31 474 438	61,60%	2 623 120	1 485 630
M#20_MMR1	51 773 512	47 988 803	27 841 483	58,00%	2 455 181	1 333 596
M#21_MMR1	67 197 181	63 763 682	39 928 897	62,60%	2 692 334	1 618 269

Table S4

Genomic feature	NPC samples			
	10X		60X	
	Mean	Standard dev	Mean	Standard dev
CpG island	1 685 434,00	116 179,07	1 064 617,67	70 321,87
CpG shores	325 652,67	44 758,19	146 179,00	12 743,51
CpG shelves	112 869,33	15 411,05	46 133,00	4 687,59
3'UTR	32 901,33	3 223,80	15 991,33	1 450,87
5'UTR	119 799,33	7 462,21	81 664,67	4 564,02
exon	402 853,00	27 705,84	228 018,00	23 546,72
intergenic	803 420,33	102 291,63	363 954,00	27 996,67
intron	1 008 993,00	120 620,80	483 153,00	35 097,96
non-coding	46 739,00	4 326,29	26 285,67	2 415,76
promoter-TSS	714 015,33	63 273,34	450 953,33	27 160,03
TTS	60 801,67	6 363,25	31 704,67	2 523,74

Genomic feature	MM diagnosis samples			
	10X		60X	
	Mean	Standard dev	Mean	Standard dev
CpG island	1 554 465,12	127 211,94	990 139,12	81 569,69
CpG shores	280 287,12	50 914,35	140 120,29	21 169,81
CpG shelves	95 902,00	19 193,92	44 915,06	7 691,76
3'UTR	28 837,82	4 291,79	15 426,59	1 964,95
5'UTR	111 975,53	7 771,64	75 485,53	5 284,99
exon	358 430,47	38 070,83	213 453,59	26 642,16
intergenic	693 104,00	127 340,43	345 887,24	49 789,38
intron	884 785,12	146 022,29	459 475,71	57 743,30
non-coding	41 552,47	5 010,40	24 221,00	3 021,85
promoter-TSS	658 834,71	60 672,78	418 531,18	32 241,87
TTS	54 315,94	7 621,16	30 608,65	3 678,23

Genomic feature	MM relapse samples			
	10X		60X	
	Mean	Standard dev	Mean	Standard dev
CpG island	1 538 411,79	129 962,61	960 684,68	78 413,74
CpG shores	272 954,53	50 931,76	131 989,26	16 614,27
CpG shelves	92 846,05	19 258,51	41 911,58	5 804,47
3'UTR	28 113,68	4 270,86	14 675,84	1 582,77
5'UTR	111 102,74	8 083,07	73 751,37	5 232,89
exon	351 867,16	39 218,37	203 409,58	24 139,25
intergenic	673 110,00	126 902,01	327 037,26	36 799,38
intron	863 460,89	145 348,57	437 271,89	44 388,35
non-coding	40 760,74	5 047,77	23 130,53	2 636,54
promoter-TSS	653 542,42	60 438,48	407 221,95	31 255,89
TTS	53 199,47	7 502,65	29 154,32	3 193,59

Table S5

MMD	PDR Low			PDR High		
	OR	IC lower	IC upper	OR	IC lower	IC upper
M#17	0.2308	0.1540	0.3436	0.6884	0.4622	1.0099
M#02	0.2810	0.1212	0.6467	0.7217	0.4917	1.0489
M#05	0.3512	0.1601	0.7443	0.7722	0.5191	1.1290
M#14	0.1640	0.1074	0.2467	0.7945	0.5026	1.2262
M#12	0.3152	0.1748	0.5608	0.8178	0.5989	1.1097
M#19	0.2175	0.1368	0.3417	0.8274	0.5320	1.2584
M#10	0.3188	0.2020	0.4979	0.8458	0.5584	1.2559
M#03	0.3103	0.1382	0.6805	0.8529	0.5646	1.2690
M#11	0.6451	0.3298	1.2976	0.8856	0.6348	1.2257
M#13	0.4477	0.2505	0.7913	0.9381	0.5715	1.4888
M#06	0.4086	0.2004	0.8241	0.9785	0.6841	1.3841
M#16	0.2526	0.1274	0.4781	1.0641	0.7430	1.5083
M#15	0.4510	0.2512	0.8055	1.1206	0.7101	1.7266
M#08	0.6692	0.3582	1.2615	1.1293	0.7413	1.6910
M#04	1.0777	0.1543	11.9275	1.4193	0.8189	2.3778
M#07	0.7098	0.3093	1.6649	1.5157	0.9754	2.3123
M#20	0.5143	0.2081	1.2470	1.5674	1.0084	2.3919
M#21	0.5818	0.2755	1.2345	1.8145	1.2013	2.7058
M#18	0.5437	0.1077	2.5296	2.0390	1.2548	3.2622
M#09	0.7167	0.2208	2.2658	2.4235	1.3827	4.1388

Titre : L'hétérogénéité de l'épigénétique intra et inter-patient dans le cas du myélome multiple.

Mots clés : Myélome multiple, épigénétique, méthylation, épipolymorphisme.

Résumé : L'évolution du cancer dépend d'une certaine hétérogénéité épigénétique et génétique. Dans le cas du myélome multiple (MM), la diversité sous-clonale et l'évolution tumorale ont été principalement étudiées sous un angle génétique. Dans notre étude, nous avons analysé l'hétérogénéité de la méthylation de l'ADN de patients atteints de MM, en utilisant la notion d'épipolymorphisme et la mesure des changements épialléliques entre deux états de la maladie. Nous avons montré que le MM est caractérisé par une augmentation stochastique / désordonnée de la méthylation de l'ADN au niveau des régions régulatrices. De forts changements épialléliques à l'initiation de la maladie sont associés à un mauvais pronostic et à la déméthylation de domaines partiellement méthylés (PMDs) sur le génome.

Ces PMDs représentent la principale source d'hétérogénéité de la méthylation de l'ADN dans le cas du MM, et sont associés à d'autres aberrations clés dans le cancer telles que l'hyperméthylation des régions régulatrices et la redistribution de la marque d'histone H3K27me3. De plus, la perturbation de la méthylation dans ces régions régulatrices est également associée à une faible expression des gènes et à une grande variabilité transcriptomique. Nous proposons que la forte déméthylation de l'ADN observée dans le MM soit responsable de fortes instabilités épigénétiques et transcriptomiques permettant aux cellules tumorales de s'adapter aux changements environnementaux et d'adopter de nouvelles trajectoires.

Title: Intra and inter-patient epigenetic heterogeneity in multiple myeloma.

Keywords : multiple myeloma, epigenetic, DNA methylation, epipolymorphism.

Abstract : Cancer evolution depends on epigenetic and genetic diversities. Historically, in multiple myeloma (MM), subclonal diversity and tumor evolution have been mostly investigated through the genetic angle. In our study, we analyzed DNA methylation heterogeneity in MM patients, using the notion of epipolymorphism and epiallele switching between two disease states. We showed that MM is characterized by a continuous accumulation of stochastic methylation at promoters. High epiallele entropy is associated with poor outcome and predominantly depends on the demethylation of partially methylated domains (PMDs).

These PMDs which represent the major source of DNA methylation heterogeneity in MM are linked to other key epigenetic aberrations such as CGI/TSS hypermethylation and H3K27me3 redistribution. In addition, methylation disorder is associated to low gene expression and high transcriptomic variability. We proposed that strong DNA methylation loss in MM is responsible for high epigenetic and transcriptomic instabilities allowing tumor cells to fit environmental changes by finding new trajectories.