



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Bretagne Loire

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1
Mention : Biologie

Ecole doctorale Vie Agro Santé

Elise MAHE

Préparée à l'unité de recherche UMR CNRS 6290
Institut de Génétique et Développement de Rennes
UFR Sciences de la Vie et de l'Environnement
Université de Rennes1

**Dynamique
chromatinienne lors
de l'activation des
enhancers au cours
de la différenciation
cellulaire**

**Thèse soutenue à Rennes
le 30 Mars 2016**

devant le jury composé de :

Ali HAMICHE

Directeur de Recherche CNRS, IGBMC –
Strasbourg / *rapporteur*

Salvatore SPICUGLIA

Chargé de Recherche Inserm, TAGC – Marseille /
rapporteur

Michaël WEBER

Chargé de recherche Inserm, IREBS – Strasbourg /
examineur

Michael PRIMIG

Directeur de Recherche Inserm, IRSET – Rennes/
examineur

Raphaël METIVIER

Chargé de recherche CNRS, IGDR – Rennes/ *co-
directeur de thèse*

Gilles SALBERT

Professeur d'Université, IGDR – Université de
Rennes 1 / *directeur de thèse*

A mon père,

REMERCIEMENTS

Je tiens dans un premier temps à remercier les Dr. **Ali Hamiche** et **Salvatore Spicuglia** d'avoir accepté d'être rapporteurs de mon manuscrit de thèse. Je remercie également les Dr. **Michaël Weber** et **Michael Primig** d'avoir accepté de participer à ce jury.

Je remercie **Claude Prigent**, directeur de l'Institut de Génétique et Développement de Rennes, de m'avoir accueillie au sein de son unité.

Je souhaite remercier vivement les membres de mon comité de thèse pour tous leurs précieux conseils et leur bonne humeur. **Denis Michel**, ce fut un plaisir de vous avoir comme tuteur, merci de votre considération et de vos nombreuses blagues ! **Frédéric Chalmel**, merci de m'avoir accompagné dès le début de cette thèse avec quelques conseils pour le concours, ta participation à ce comité et notre collaboration afin de sublimer mes résultats grâce à la PCA. **Marie-Dominique Galibert**, merci de votre intérêt pour mon travail et de vos conseils.

Gilles, tout d'abord merci de m'avoir accueillie dans votre équipe pour mon stage de M2 et ensuite pour la thèse. Ces mots ne sont que peu de choses face à la reconnaissance que j'ai envers vous. Je vous remercie de tout ce que vous m'avez appris, du temps considérable que vous m'avez accordé tout au long de cette thèse, de votre professionnalisme et de votre confiance. Merci également de m'avoir donné l'opportunité d'aller plusieurs fois à de très bons congrès qui resteront de beaux souvenirs. Merci d'avoir partagé votre bureau avec moi, on aura bien travaillé mais également beaucoup rigolé !

Raphaël, merci de m'avoir prise en stage il y a 4 ans et de m'avoir donné la passion de la chromatine et de la 5hmC ! Je te remercie pour tout ce que tu m'as appris, pour ta rigueur, pour tout le temps que tu as passé sur ce projet, pour tes conseils et bien évidemment pour ton aide précieuse pour la préparation et finalisation du manuscrit. Merci également de ta gentillesse et de ton calme à toute épreuve !

Maud, je te dois, et cette thèse te doit beaucoup. Tu m'as énormément aidé, du début à la fin. Je te remercie d'avoir suivi toutes mes petites manies, c'est pour cela que je pouvais faire entièrement confiance à ton travail. La chance nous a souvent tourné le dos mais tu persévérais avec moi. Merci infiniment de ta bonne humeur, de ta gentillesse et de ton soutien.

Christine, merci de ta bienveillance et de ta générosité. Quel plaisir de t'avoir eue comme collègue ! Merci également de toute l'aide que tu m'as apporté.

Gaëlle, merci pour ta bonne humeur et ton entrain sans faille, tes « A mardaï ! » si célèbres et ta douce voix qui chante dans le labo !

Thierry, je te remercie pour ton implication dans le projet 1 et pour la reprise du n°2, pour tes talents de cruciverbiste et ton humour piquant.

Stéphane, merci pour tout le travail que tu as donné pour les cinétiques, nous nous souviendrons longtemps de l'été 2015 à essayé les conditions de normalisations. Merci également pour toutes les conversations autour du rituel de la pomme au goûter.

Audrey, on se sera seulement croisées. Profites bien de ton séjour chez les Spartiates, il ne peut en ressortir que du bon.

Mes chers **Spartiates**, « le chercheur n'est rien sans son équipe » et cela se vérifie avec vous. Merci de votre implication dans mon projet de thèse que ce soit concrètement ou moralement, pour vos nombreux coups de main, petites attentions ou grands fous rires. Ce fût vraiment un plaisir de travailler avec vous pendant 4 ans et je trouve vraiment dur de devoir vous quitter...

En parlant de Spartiate... J'ai aussi une pensée pour **Aurélien**, merci également d'avoir participé à mon projet et de m'avoir accueillie ces quelques jours à Cambridge pour que je puisse voir le sort réservé à mes petits échantillons. Je te souhaite beaucoup de réussite dans tes projets à venir.

Un grand merci aux membres des équipes TREC et NAO, plus particulièrement à **Catherine** et **Pascale**, pour les discussions plus ou moins farfelues et animées le midi à la cafèt', les francs moments de rigolades et les potins, et pour les conseils cinématographiques... ou pas ! ; à **Denis H**, je te souhaite sincèrement de pouvoir enfin concrétiser ton projet professionnel ; à **Florence**, **Colette** et **Elisabeth** pour tous les repas que nous avons partagé ensemble.

Mes remerciements vont également à **Fred**. Ah, Fred ! Notre superman de la culture ! Un immense merci pour ton travail sur la plateforme de culture cellulaire, qui nous permet de travailler dans d'excellentes conditions, pour tout ce que tu m'as appris et pour tes implications dans mes différents projets notamment sur les clones sh avant mon arrivée en stage et plus tard, sur les clones MCF-7. Je ne peux évidemment pas achever ce paragraphe sans parler de ton autre talent : le chant. Tes chansons, parfois sorties d'un autre temps, ont souvent sauvé mes longs vendredi après-midi de cross-linking et je t'en remercie. J'en profite pour remercier **Charlotte**, **Gaëlle B.**, **Emmanuelle** et **Adrien** pour leur travail sur la plateforme qui a grandement participé à la qualité top niveau de notre salle de culture.

Merci également à **Michèle** et **Valérie**, pour leur travail de laverie et de ménage, toujours parfaitement réalisé.

Chers collègues de l'IRSET, merci pour l'ambiance que vous apportiez au fameux bâtiment 13. J'ai particulièrement apprécié l'entente entre nos équipes qui faisait que tout était simple : prêt de matériel, dépannage de réactifs, conseils sur les manip etc... Au moment où j'écris ces remerciements, vous êtes en plein préparatifs de votre déménagement, les labos et bureaux se vident (plus doucement pour certains que pour d'autres... !) et les cartons

commencent à s'entasser dans le couloir. Je suis contente d'avoir fait ma thèse durant cette grande époque dans le couloir HIP/SPARTE car le bâtiment ne sera plus le même sans vous...

Une pensée particulière va à mes collègues de thèse de Beaulieu. **Gwenneg, Floriane, Justine, Yonis, Kévin**. Merci à **Claire** pour les moments que nous avons partagés sur la fin de ta thèse, pour nos footings et nos conversations ragots. **Alexis**, merci de m'avoir fait autant rire avec toutes tes aventures de vélo, de pipetman et autres qui ne pouvaient arriver qu'à toi. Je garde un très bon souvenir de nos discussions scientifiques et personnelles. Je te remercie de tes conseils et du soutien que tu m'as apporté, ce fut un grand plaisir de faire ces trois années de thèse en même temps que toi. **Charly**, le petit dernier mais le plus musclé ! Merci pour ta bonne humeur et ton côté étourdi qui me fait sourire le soir quand je t'entends revenir sur tes pas en courant car tu as oublié quelque chose dans ton bureau ! Bon courage pour la fin de ta thèse mais je ne me fais pas de souci pour toi, tout devrait très bien se passer. Merci à chacun d'entre vous pour les moments que nous avons partagé, je vous souhaite de réussir et j'espère que nos chemins se recroiseront.

Aux autres collègues thésards, et plus particulièrement à **Mélanie** et **Pierre**, que je remercie pour m'avoir permis de monter avec eux à cheval pendant presque un an. Que d'aventures et de folles balades à dos de dada ! C'était vraiment ce qu'il me fallait pour m'aérer l'esprit. Je vous souhaite bon courage pour la fin de vos thèses et bonne continuation.

Ma chère **Pauline**, ça y est on y est presque à la fin de cette thèse. Merci de l'avoir partagé avec moi, merci de ton soutien, de nos soirées « décompression » et de ta positive attitude ! Je te souhaite énormément de courage pour ton dernier mois et beaucoup de choses positives pour la suite car tu le mérites amplement.

Je tiens tout particulièrement à remercier mon amie **Marie** de m'avoir soutenue tout au long de cette thèse, de s'être intéressé à mon travail, d'essayer de comprendre mon projet alors que ce n'était pas son domaine de connaissances. Je te remercie pour les moments de détente et surtout pour notre virée mémorable à Disney. Merci la Marie !

Je remercie également toute ma famille pour leur soutien et particulièrement mon **Papa** pour avoir toujours tout fait pour que je ne manque de rien et que je puisse réussir. Merci du fond du cœur Papa de croire en moi et de m'encourager jours après jours.

On dit toujours, le meilleur pour la fin... Mon **Renaud**... Je ne saurais te remercier assez pour tout ce que tu m'as apporté durant ces 4 années. Merci de ton soutien exceptionnel, particulièrement au cours des dernières semaines qui, je le sais, n'étaient pas évidentes pour toi non plus. Merci de me faire garder le sourire jours après jours. De nouvelles aventures nous attendent lorsque tu auras fini ta thèse, et j'ai hâte de les partager avec toi.

RESUME

La différenciation cellulaire implique une régulation transcriptionnelle coordonnée et finement contrôlée qui passe par le recrutement de facteurs de transcription (FT) cellulaires spécifiques sur des régions génomiques régulatrices appelées enhancers. Parmi ces FT, des protéines nommées « facteurs pionniers » (FP) lient la chromatine condensée et favorisent la transition des enhancers d'un état inactif vers un état « préparé » (étape de « priming »), facilitant ainsi la fixation d'autres FT et permettant l'activation de ces régions. L'engagement vers un lignage cellulaire particulier est donc associé à l'engagement des FP au niveau d'enhancers dont la structure chromatinienne subit des changements architecturaux associés à la mise en place de marques spécifiques. Celles-ci incluent, la monométhylation de la lysine 4 de l'histone H3 (H3K4me1), l'acétylation de la lysine 27 de l'histone H3 (H3K27ac) ou encore des modifications des résidus cytosine (5-méthylcytosine, 5mC ; 5-hydroxyméthylcytosine, 5hmC).

La 5hmC est un intermédiaire de la voie de déméthylation active : elle résulte de l'oxydation de la 5mC par les enzymes « Ten Eleven Translocation » (TET) et peut être à son tour oxydée en 5-formylcytosine (5fC) et 5-carboxylcytosine (5caC) qui sont ensuite remplacées par des cytosines via l'intervention du système « Base Excision Repair ». Cependant, du fait de sa stabilité et de sa capacité à lier des protéines particulières, la 5hmC pourrait également jouer un rôle spécifique. De précédents travaux ont d'ores et déjà mis en évidence un lien entre le recrutement des FPs et les modifications des cytosines. Néanmoins, l'implication des processus de méthylation/déméthylation dans la régulation spatio-temporelle des étapes de « priming » et d'activation des enhancers n'a pas encore été caractérisée.

Dans ce contexte, l'objectif de cette étude a été de définir le rôle des modifications de cytosines (5mC et 5hmC) lors de l'activation des enhancers liés par des FPs. Pour ceci, nous avons analysé d'une part, l'implication des processus de méthylation et déméthylation des cytosines sur le « priming » et l'activation des enhancers, en utilisant des inhibiteurs des ADN méthyltransférases ou des enzymes TET. D'autre part, nous avons entrepris d'identifier les dynamiques de « priming » et d'activation des enhancers à l'échelle du génome au cours de la différenciation neurale, en lien avec la présence de la 5hmC. Les résultats obtenus nous ont notamment permis de proposer un schéma d'activation des enhancers pour lequel l'enlèvement de la 5mC est nécessaire au « priming » de ces régions tandis que la 5hmC serait responsable de l'ouverture de la chromatine.

ABBREVIATIONS

(h)MeDIP (hydroxy)methylated DNA immunoprecipitation	DMOG dimethyloxalylglycine
3C chromosome conformation capture	DNMT DNA methyltransferases
4C Circularized Chromosome Conformation Capture	DPE downstream promoter element
5-azadC 5-aza-2'-déoxycytidine	DSIF DRB Sensitivity Inducing Factor
5C Carbon-Copy Chromosome Conformation Capture	ECC embryonal carcinoma cells
5caC 5-carboxylcytosine	EKLF Erythroid Krüppel-like factor
5fC 5-formylcytosine	ER estrogen receptor
5hmC 5-hydroxyméthylcytosine	eRNA enhancer RNA
5mC 5-méthylcytosine	ESC embryonic stem cells
AID/APOBEC apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like	FACT facilitates chromatin transcription
ANP32E Acidic nuclear phosphoprotein 32 family, member E	FAIRE Formaldehyde Isolation of Regulatory Elements
AR androgen receptor	FOXA1 Forkhead box 1
ARN Pol II ARN polymérase II	FP facteur pionnier
ARNm ARN messenger	FT facteur de transcription
ATAC-seq Assay for Transposase-Accessible Chromatin with highthroughput sequencing	GR glucocorticoid receptor
ATP Adénosine Triphosphate	GRO-seq global nuclear run-on sequencing
BER Base Excision Repair	GTF general transcription factors
BRE TFIIB-recognition element	HCF1 Host Cell Factor 1
C/EBP CCAAT/enhancer-binding protein	HDAC histone deacetylase
CAGE cap analysis of gene expression	HP1 heterochromatin protein 1
CBP CREB-binding protein	ICF immunodeficiency, centromeric region instability, facial anomalies
CHD chromodomain helicase DNA binding	IDAX Inhibitor of Dvl and Axin complex
ChIA-PET chromatin interaction analysis by paired-end tag sequencing	Inr Initiator
ChIP-seq chromatin immunoprecipitation followed by high throughput sequencing	ISWI imitation SWI
CpG dinucléotides CG	KDM2A Lysine-specific demethylase 2A
CTCF CCCTC-binding factor	LAD Lamin-associated domains
CTD carboxy-terminal domain	LINE Long Interspersed Nuclear Elements
DCE downstream core element	LTR Long Terminal Repeat
DMEM Dulbecco's Modified Eagle's Medium	MBD Methyl-CpG binding domain
	MLL1 Mixed Lineage Leukemia 1
	NAP-1 nucleosome assembly protein 1
	NCoR nuclear receptor co-repressor
	NDR nucleosome depleted regions
	NELF negative elongation factor
	NPC neural progenitor cells
	NuRD nucleosome remodeling and deacetylation
	O-GlcNAc O-linked N-acetylglucosamine

OGT O-GlcNAc transferase
pb paire de bases
PBX1 pre-B-cell leukemia homeobox 1
PcG Polycomb-group proteins
PGC primordial germ cells
PIC pre-initiation complex
PPAR γ Peroxisome proliferator-activated receptor gamma
PRC2 Polycomb Repressive Complex 2
P-TEFb Positive Transcription Elongation Factor b
py/pu pyrimidine/purine
SAM S-Adenosyl Methionine
SET1/COMPASS Complex proteins associated with SET1
SETD1A SET Domain Containing 1A
siRNA small interference RNA
SMRT silencing mediator for retinoid or thyroid-hormone receptors
Spt Suppressor of Ty
SRA SET- and RING-associated DNA-binding domain
STARR-seq self-transcribing active regulatory region sequencing
SWI/SNF switching defective/sucrose non-fermenting
TAD Topologically Associating Domains
TAF TBP-associated factors
TALE-HD three amino-acid loop extension homéodomain
TBP TATA box-binding protein
TDG Thymidine DNA Glycosylase
TET Ten-eleven translocation
TGF- β Transforming growth factor beta
TNF- α tumor necrosis factor
TSS site d'initiation de la transcription
UHRF1 ubiquitin-like, containing PHD and RING finger domain 1
ZF Zinc Finger

SOMMAIRE

SOMMAIRE

INTRODUCTION	1
I – LA REGULATION DE LA TRANSCRIPTION	1
A – Structure de la chromatine	1
1 – Les variants et chaperons d’histones	2
2 – Les modifications post-traductionnelles des histones	4
3 – Complexes de remodelage de la chromatine dépendants de l’ATP	4
4 – Interactions entre les différents acteurs régulant la structure chromatinienne	5
B – Eléments régulateurs de la transcription	6
1 – Séquences régulatrices	7
a – Promoteur basal	7
b – Promoteur proximal et enhancer	9
c – Insulateur	9
2 – Les facteurs de transcription	10
C – Principes de régulation au cours des différentes étapes de la transcription	11
1 – L’initiation de la transcription	11
a – Restructuration de la chromatine	12
b – Mise en place du complexe de pré-initiation de la transcription	12
c – Co-régulateurs de l’initiation	13
2 – L’élongation de la transcription	13
a – Elongation précoce	13
b – Elongation productive	14
c – Structure chromatinienne	15
D – Organisation du génome	15
1 – Territoires chromosomiques	16
2 – Domaines chromatinien : à l’échelle de la cellule	17
3 – Domaines chromatinien : à l’échelle du chromosome	18
II – LES MODIFICATIONS DES CYTOSINES	20
A – Méthylation de l’ADN	20
1 – Mise en place et maintien de la méthylation des cytosines	21
a – Les DNMTs	21
b – Adressage des DNMTs	22
2 – Méthylation des CpG et répression transcriptionnelle	23
a – Exemples de répressions exercées par la méthylation des CpG	23
b – Mécanismes de répression	24

c – Initiation ou maintenance de la méthylation ?	25
3 – La déméthylation	27
a – La méthylation : une modification dynamique	27
b – Mécanismes de déméthylation active	28
B – Dérivés oxydés	30
1 – Mise en place des dérivés oxydés	30
a – Les protéines TET	30
b – Adressage et substrats	31
2 – La 5-hydroxyméthylcytosine	32
3 – La 5-formylcytosine et la 5-carboxylcytosine	33
C – Implications des protéines TET et TDG dans la régulation transcriptionnelle et la structure chromatinienne	34
1 – Régulation par les protéines TET	35
a – Répression par les TET	35
b – Interactions avec les protéines OGT	36
c – TET, pluripotence, différenciation et reprogrammation	37
2 – Régulation par TDG	37
D – Protéines de reconnaissance des dinucléotides CG	39
1 – CpG non modifiés	39
2 – CpG méthylés	39
3 – CpG oxydés	40
III – LES ENHANCERS	42
A – Identification des régions enhancers	42
1 – Comparaison de séquences	42
2 – Caractéristiques chromatinienne des enhancers	43
a – Modifications des histones	43
b – Positionnement et nature des nucléosomes aux enhancers	45
c – Modifications des cytosines	45
3 – Recrutement de facteurs protéiques	46
a – Recherche de motifs de liaison pour des FT	46
b – Liaison par des facteurs co-activateurs	47
4 – Présence de l'ARN Pol II et transcription aux enhancers	48
5 – Cartographie des enhancers à l'échelle du génome	49
6 – Activité des enhancers	50
B – Sélection des enhancers	51
1 – Enhancers <i>de novo</i>	52
a – Facteurs pionniers	52

b – Liaison de facteurs additionnels	53
c – Instabilité des nucléosomes : implication du variant H2A.Z	54
2 – Enhancers pré-marqués dans les cellules indifférenciées	55
3 – Enhancers latents	56
C – Interactions avec les promoteurs	57
1 – Détection des interactions longues distances et données résultantes	57
2 – Facteurs impliqués dans les interactions entre promoteurs et enhancers	58
D – Super-enhancers	60
E – Enhancers et promoteurs: sont-ils si différents ?	61
IV – INTEGRATION DES TRAVAUX DE THESE DANS LE CONTEXTE SCIENTIFIQUE	63

RESULTATS

I – PARTIE 1 – IMPLICATION DES MODIFICATIONS DES CYTOSINES DANS LA STRUCTURATION DE LA CHROMATINE DES ENHANCERS AU COURS DE LA DIFFERENCIATION CELLULAIRE.	66
Article « Cytosine modifications modulate the chromatin architecture of transcriptional enhancers »	
Elise A Mahé, Thierry Madigou, Aurélien A Sérandour, Maud Bizot, Stéphane Avner, Frédéric Chalmel, Gaëlle Palierne, Raphaël Métivier and Gilles Salbert.	
II – PARTIE 2 – DYNAMIQUE CHROMATINIENNE DES ENHANCERS HYDROXYMETHYLES AU COURS DE LA DIFFERENCIATION NEURALE	123

DISCUSSION	147
-------------------	------------

BIBLIOGRAPHIE	153
----------------------	------------

ANNEXE

INTRODUCTION

I – LA REGULATION DE LA TRANSCRIPTION

Chez les eucaryotes, les processus fondamentaux tels que la division cellulaire, le métabolisme, le développement ou la différenciation cellulaire requièrent la présence et l'intervention de protéines spécifiques dont l'activité ou la présence peut être critique. Le contrôle de l'expression génique menant à la synthèse protéique est donc crucial afin d'assurer le bon déroulement des fonctions cellulaires. Un premier niveau de régulation est effectué lors de la transcription de l'ADN, support de l'information génétique, en ARN messager (ARNm). La transcription est un processus hautement régulé faisant intervenir de nombreux niveaux structurels et divers acteurs. Ainsi, dans la première partie de cette introduction, nous aborderons (i) les caractéristiques structurales de l'organisation de l'ADN au sein de la chromatine qui permettent de moduler la transcription, en nous focalisant sur les gènes transcrits par la Polymérase II (ii) les séquences et facteurs clés de cette régulation, (iii) le rôle des phases séquentielles de la transcription dans le contrôle de l'expression génique, et enfin (iv) l'implication de l'organisation génomique dans ces mécanismes, des niveaux supérieurs d'organisation à une organisation plus locale.

A – Structure de la chromatine

La régulation dynamique de la structure de la chromatine fournit un moyen de contrôle de l'expression des gènes particulièrement efficace. L'ADN structuré en une double hélice de 2 nm de diamètre subit une première étape de compaction afin de former l'unité structurale de la chromatine : le nucléosome. Cette structure de 11 nm de diamètre est composée d'un octamère de protéines appelées histones (H2A, H2B, H3, H4) entouré d'un fragment d'ADN de 146 paires de bases (pb) (Luger *et al.*, 1997). Les histones sont de petites protéines basiques caractérisées par un domaine C-terminal globulaire très conservé, appelé domaine « histone-fold », qui permet la dimérisation de ces protéines (Arents et Moudrianakis, 1995) ; leurs extrémités N-terminales ne sont quant à elles pas structurées. L'enchaînement des nucléosomes, associés ou non à l'histone linker H1, forme la chromatine et en facilite la condensation. De façon très générale, selon les niveaux de condensation de cette structure, on parle soit d'euchromatine, qui correspond à une structure globalement décondensée et

généralement associée à une activité transcriptionnelle, soit d'hétérochromatine, qui correspond quant à elle à des régions génomiques denses en nucléosomes communément associées à une répression transcriptionnelle. En effet, de par sa structure et sa composition, la chromatine constitue un environnement restrictif pour tout événement de modulation de l'activité du génome qui dépend de l'accessibilité de l'ADN à des protéines régulatrices. Son ouverture peut néanmoins être modulée par l'apposition de modifications au niveau de l'ADN ou des histones et/ou l'action de facteurs de remodelage ATP-dépendants (Adénosine Triphosphate). Du fait de leur importance pour les thématiques développées dans ce travail de thèse, les modifications de l'ADN feront l'objet d'une partie indépendante dans cette introduction, et ne seront donc pas abordées ici.

1 – Les variants et chaperons d'histones

Par opposition aux histones canoniques H2A, H2B, H3, H4 dont l'expression est restreinte à la phase S du cycle cellulaire, il existe des variants d'histones exprimés tout au long du cycle cellulaire. Ces variants d'histones diffèrent de leurs équivalents canoniques par leur séquence ou par la présence de domaine(s) additionnel(s) qui peuvent modifier la structure et la stabilité du nucléosome. Ainsi, la régulation spatio-temporelle de la composition des nucléosomes est un élément crucial de modulation de la structure du génome. Celle-ci peut être maintenue grâce à l'intervention de facteurs nommés chaperons d'histones. Ces protéines interagissent avec les différentes histones et sont impliquées dans leur stockage et leur transport, ainsi que dans l'assemblage et le désassemblage des nucléosomes. Une histone particulière peut être liée par différents chaperons (Goldberg, 2010) et réciproquement, une protéine chaperon peut lier plus d'un hétérodimère d'histones.

Un certain nombre de variants pouvant jouer un rôle dans les processus de régulation de la réplication, de la transcription ou de la réparation de l'ADN ont été mis en évidence. Parmi ceux-ci, les variants impliqués dans la régulation transcriptionnelle, tels que H2A.Z, H3.3, H2A.B et macroH2A, sont présentés ci-après.

H2A.Z est le variant d'histone le plus fortement conservé parmi les eucaryotes. Il diffère de H2A de seulement quelques acides-aminés et est enrichi à la fois au niveau de l'euchromatine et de l'hétérochromatine (Sarcinella *et al.*, 2007). Par son action sur la structure des nucléosomes, H2A.Z joue un rôle à la fois dans l'activation et la répression

transcriptionnelle, principalement en favorisant la liaison de complexes activateurs ou répresseurs (Hu *et al.*, 2013). En effet, Li et collègues ont montré que H2A.Z et le facteur pionnier FOXA2 régulent la déplétion en nucléosome et l'activation génique lors de la différenciation des cellules souches embryonnaires de Souris (ESC – Li *et al.*, 2012a). D'autre part, ce variant d'histone montre une distribution similaire à celle de la protéine Suz12, appartenant au groupe répresseur Polycomb (PcG), au niveau de gènes silencieux qui ont tendance à être dérprimés dans des cellules ES déplétées en H2A.Z (Creyghton *et al.*, 2008). L'incorporation du variant H2A.Z est notamment liée aux protéines chaperons « nucleosome assembly protein 1 » (NAP-1 - Park, 2005) et Chz1 (Luk *et al.*, 2007).

H3.3 ne diffère que de quatre ou cinq acides-aminés des histones canoniques H3.2 et H3.1 respectivement (Szenker *et al.*, 2011) et est incorporé par la protéine chaperon HIRA (Tagami *et al.*, 2004). Ce variant est associé à des régions transcriptionnellement actives (McKittrick *et al.*, 2004) mais s'accumule également au niveau de gènes réprimés (Goldberg *et al.*, 2010).

Le variant H2A.B, ou H2ABbd pour « Barr body-deficient », est seulement retrouvé chez les mammifères et sa présence au niveau d'une chromatine peu compacte est associée à une activation transcriptionnelle (Tolstorukov *et al.*, 2012). En effet, l'altération du « docking domain » de ce variant, modifiant les interactions histone-histone, est responsable de la formation de cet environnement chromatinien particulier (Doyen *et al.*, 2006a). Aucune protéine chaperon n'a pour l'instant été identifiée *in vivo*, cependant NAP-1 est capable de chaperonner H2A.B *in vitro* (Okuwaki *et al.*, 2005).

Enfin, macroH2A est un variant d'histone majoritairement retrouvé chez les vertébrés, composé d'un domaine histone associé à un domaine globulaire non histone formant une protéine d'approximativement trois fois la taille de l'histone canonique H2A (Pehrson et Fried, 1992). Localisé sur le chromosome X inactif et au niveau de larges domaines silencieux, macroH2A est cependant impliqué à la fois dans la répression et dans l'activation de la transcription (Gamble *et al.*, 2010).

On constate ainsi que chacun des dimères ou tétramères d'histone peut être modifié par l'incorporation d'un ou plusieurs variant d'histones. S'il est admis de manière générale que les variants d'histones n'altèrent pas seuls de façon radicale la structure des nucléosomes, il a été observé que les nucléosomes contenant les deux variants H3.3 et H2A.Z sont plus

Tableau1. Modifications post-traductionnelles des histones (d’après Bannister et Kouzarides, 2011).

Rme1: monomethylarginine; Rme2as: dimethylarginine (asymmetric); Rme2s: dimethylarginine (symmetric)

Modification	Résidus cibles	Enzymes d'apposition		Enzymes d'effacement	
Acétylation	Lysines	Histones Acétyl-transférases (HATs)	Type A: familles GNAT, MYST et CBP/p300	Histones déacétylases (HDACs)	Classes I et II: proches des enzymes de levure scRpd3 et scHda1, respectivement
	H4K5 et K12; certains résidus H3		Type B: homologie de séquence avec scHat1		Classe III: homologue de l'enzyme scSir2 de levure
					Classe IV: HDAC11
Phosphorylation	Sérines, thréonines et tyrosines	Kinases		Phosphatases	
Méthylation	Lysines	Lysine méthyl-transférases (HKMTs)	Exemple: SUV39H1 (H3K9)	Lysine déméthylases	Lysine-specific demethylase 1 (LSD1; uniquement sur des substrats mono- ou di-méthylés)
	Arginines	Arginine méthyl-transférases (PRMTs)	Type-I: Rme1 et Rme2as		Enzymes contenant un domain catalytique Jumonji
			Type-II: Rme1 et Rme2s		
Citrullination	Arginines	Peptidyl déiminase	PADI4		
O-N-acétylglucosaminylation	Sérines et thréonines des histones H2A, H2B et H4	O-GlcNAc transférase		β-N-acetylgluco-saminidase	
ADP ribosylation	Glutamate et arginines	Poly-ADP-ribose polymerase (PARP)		Poly-ADP-ribose glyco-hydrolase	
Ubiquitylation	Lysines	Ubiquitine-ligases E1, E2 et E3		Enzymes de dé-ubiquitylation	
Sumoylation	Lysines	Enzymes E1, E2 et E3			

instables (Jin et Felsenfeld, 2007). De plus, une hétérogénéité de mobilité est observée entre le tétramère H3/H4 et les dimères H2A/H2B en fonction du niveau d'activité transcriptionnelle des gènes. En effet, au niveau des gènes très fortement transcrits, l'entièreté de l'octamère d'histone peut être déplacée notamment grâce à l'échange d'histones telles que H3 en H3.3 (Mito *et al.*, 2005) ; ce qui est rarement le cas pour les gènes modérément transcrits, pour lesquels ce sont majoritairement les dimères H2A/H2B qui sont échangés (Kimura et Cook, 2001).

2 – Les modifications post-traductionnelles des histones

La structure de la chromatine peut être remodelée par des modifications post-traductionnelles des histones. Celles-ci ont principalement lieu au niveau de la queue des histones, même si certaines peuvent concerner des résidus situés au cœur même des nucléosomes. Les résidus tels que les lysines, sérines, thréonines ou encore arginines peuvent être modifiés par acétylation, phosphorylation, méthylation, citrullination, O-N-acétylglucosaminylation, ADP-rybosylation, ubiquitination ou encore sumoylation (Tableau 1; Bannister et Kouzarides, 2011). Ces modifications, en perturbant les interactions chimiques intra- ou inter-nucléosomales, affectent leur stabilité, entraînant ainsi soit l'ouverture soit la fermeture de la chromatine. Elles sont cependant très majoritairement réversibles grâce à l'action d'enzymes « d'effacement » par opposition aux enzymes « d'apposition » (Tableau 1; Bannister et Kouzarides, 2011) permettant ainsi l'acquisition d'une balance entre les propriétés activatrices et répressives de ces modifications.

3 – Complexes de remodelage de la chromatine dépendants de l'ATP

Les niveaux d'ouverture de la chromatine peuvent également être régulés par l'intervention de facteurs de remodelage dont l'activité dépend de l'ATP. En effet, ces complexes multi-protéiques, composés de 4 à 17 sous-unités et dont le poids moléculaire est généralement supérieur à 1 megadalton, déroulent, déplacent, échangent ou éjectent les nucléosomes en utilisant l'énergie de l'hydrolyse de l'ATP, permettant ainsi de moduler l'accessibilité de l'ADN. Quatre grandes familles de complexes de remodelage ATP-dépendants ont été mises en évidence chez la levure et présentent des homologues chez certains eucaryotes supérieurs (Drosophile, Homme) : les familles SWI/SNF (« switching defective/sucrose non-fermenting » – Cairns *et al.*, 1994, 1996 ; Côté *et al.*, 1994 ; Kwon *et*

al., 1994 ; Shi *et al.*, 2014), ISWI (« imitation SWI » – Ito *et al.*, 1997 ; LeRoy *et al.*, 1998 ; Tsukiyama *et al.*, 1995, 1999 ; Tsukiyama and Wu, 1995 ; Varga-Weisz *et al.*, 1997), NuRD (« nucleosome remodeling and deacetylation »)/Mi-2/CHD (« chromodomain helicase DNA binding » – Tong *et al.*, 1998 ; Xue *et al.*, 1998 ; Zhang *et al.*, 1998) et INO80 (« inositol requiring 80 » – Kanemaki *et al.*, 1999 ; Kobor *et al.*, 2004 ; Mizuguchi *et al.*, 2004 ; Shen *et al.*, 2000). Ces complexes contiennent tous une sous-unité ATPase/hélicase de type SWI2/SNF2 appartenant à la superfamille des hélicases de groupe 2 (Byrd et Raney, 2012). La classification des complexes de remodelage se base sur la présence de domaines additionnels au domaine commun ATPase, spécifiques pour chaque famille. Par son action sur la structure chromatinienne, la famille SWI/SNF est majoritairement impliquée dans l'activation de la transcription tandis que les complexes des familles ISWI et NuRD sont principalement responsables de la répression transcriptionnelle. Mais toutes les familles semblent en fait capables d'exercer les deux types de régulation (activation et répression – Havas *et al.*, 2001).

4 – Interactions entre les différents acteurs régulant la structure chromatinienne

Contrairement à ce qui était initialement suggéré, il est désormais admis et démontré que tous les facteurs régulant la structure chromatinienne, *i.e.* variants, chaperons, modifications post-traductionnelles des histones ainsi que les complexes de remodelage, n'agissent pas seuls mais plutôt de concert.

D'une part, la mise en place des modifications post-traductionnelles des histones ne sert pas uniquement à moduler directement la structure de la chromatine mais également à recruter des protéines ou des complexes possédant des activités enzymatiques spécifiques. En ce sens, l'acétylation des lysines influence notamment le recrutement des complexes de remodelage de la chromatine ATP-dépendants de type Snf2 car nombre d'entre eux contiennent un domaine de reconnaissance pour les histones acétylées (bromodomaine – Ferreira *et al.*, 2007). De plus, les activités de désacétylation d'histone et de remodelage peuvent être présentes dans le même complexe protéique comme c'est le cas pour NuRD qui inclut entre autres deux sous-unités histone désacétylases (HDAC1 et HDAC2) et une sous-unité ADN hélicase (CHD – Xue *et al.*, 1998).

D'autre part, un lien important entre les variants d'histones et les complexes de remodelage a également été mis en évidence. En effet, en ce qui concerne le variant H2A.Z, les protéines chaperons NAP-1 et Chz1 fournissent les variants nécessaires au complexe ATP-dépendant Swr1 (qui appartient à la famille INO80) pour l'échange de l'histone canonique H2A en H2A.Z (Kobor *et al.*, 2004 ; Mizuguchi *et al.*, 2004). De plus, il a été observé que l'extension d'une région acide contenue sur la surface d'H2A.Z stimule l'activité de remodelage d'un autre complexe ATP-dépendant, le complexe ISWI (Goldman *et al.*, 2010). Enfin, H2A.Z peut être elle-même échangée ou éjectée par deux acteurs : soit par l'enzyme de remodelage INO80 (Papamichos-Chronakis *et al.*, 2011), soit par le chaperon d'histone ANP32E (« Acidic nuclear phosphoprotein 32 family, member E » – Obri *et al.*, 2014). D'autres chaperons ou variants d'histone interagissent avec ces complexes comme Daxx, chaperon d'H3.3, qui coopère avec un membre de la famille SWI/SNF pour l'incorporation du variant au niveau de l'hétérochromatine péricentrique (Drané *et al.*, 2010) ; ou encore macroH2A qui bloque le remodelage par Swi/Snf (Doyen *et al.*, 2006b). De plus, macroH2A réprime l'acétylation des histones dépendante de P300 et Gal4-VP16 (Doyen *et al.*, 2006b) démontrant un lien entre modifications post-traductionnelles et variants d'histones.

Ces exemples illustrent l'importance de la coopération entre ces différents acteurs afin d'assurer la régulation transcriptionnelle. C'est ainsi que les enzymes responsables des modifications des histones collaborent avec les complexes de remodelage et les chaperons d'histone lors des étapes du processus de transcription afin de permettre l'accessibilité et la progression de l'ARN polymérase II (Pol II) sur l'ADN (*cf.* partie C–Étapes de la transcription).

B – Éléments régulateurs de la transcription

En plus de ces éléments de régulation inhérents à la structure de la chromatine, la transcription génique est contrôlée par le recrutement et l'action de protéines régulatrices nommées facteurs de transcription (FT) au niveau de séquences d'ADN spécifiques plus ou moins distantes du site d'initiation de la transcription (TSS).

Tableau 2. Positions, séquences et fréquences des éléments composants les promoteurs (d'après Heintzman et Ren, 2007; Roy *et al.*, 2015).

Inr: « Initiator », BREu: TFIIB-recognition element up, BREd: TFIIB-recognition element down, DPE: downstream promoter element, DCE: downstream core element. D = A, G ou T ; K = G ou T ; R = A ou G ; S = G ou C ; W = A ou T ; Y = C ou T.

Élément	Position approximative relative au TSS (+1)	Séquence consensus	Fréquence chez les vertébrés
TATA	-31 à -26	TATAWAAR	10-16 %
Inr	-2 à +4	YYANWYY	55 %
BREu	-35	SSRCGCC	12-62 %
BREd	-20	RTDKKKK	Non disponible
DPE	+28 à +32	RGWYV	48 %
DCE	+9, +18, +32	CTTC, CTGT, AGC	Non disponible

1 – Séquences régulatrices

Chez les eucaryotes, l'expression des gènes est généralement régulée grâce à l'engagement de plusieurs régions régulatrices situées en *cis*. Certaines de ces régions sont situées au niveau du site d'initiation de la transcription (promoteur basal) alors que d'autres sont plus distantes (promoteur proximal, enhancer et insulateur).

a – Promoteur basal

Le promoteur basal est très souvent considéré comme l'élément central permettant la transcription et, par définition, correspond à la séquence minimale nécessaire à l'initiation du processus de transcription. Il est localisé entre approximativement – 35 et + 35 nucléotides par rapport au TSS chez les métazoaires (Heintzman et Ren, 2007). Autrefois, le promoteur basal était considéré comme une entité générique pouvant fonctionner de manière universelle. Mais l'apport des nouvelles technologies ainsi que la variété associée à la complexité des modèles étudiés ont permis de mettre en évidence les diversités de structures et de fonction de cette région régulatrice.

Les premières études réalisées afin de caractériser le promoteur basal chez les métazoaires se sont inspirées de l'architecture des promoteurs chez les bactéries et les eucaryotes unicellulaires. C'est ainsi que, dans un premier temps, le promoteur basal était considéré comme constitué d'un ou de plusieurs motifs servant de sites de reconnaissance pour la liaison du complexe basal de transcription. Ces motifs correspondent à des séquences conservées au cours de l'évolution telles que la boîte TATA, le motif « Initiator » (Inr), l'élément de reconnaissance à TFIIB (BRE), le « downstream promoter element » (DPE) et le « downstream core element » (DCE – Tableau 2 – Heintzman et Ren, 2007 ; Roy *et al.*, 2015).

Le développement de nouvelles techniques et des études à l'échelle du génome a finalement démontré que 90% des phénomènes d'initiation de la transcription sont indépendants de la présence de la boîte TATA chez la Souris et l'Homme (Carninci *et al.*, 2006). L'utilisation de la technique de CAGE (« cap analysis of gene expression » – technique permettant de déterminer les TSS et leur utilisation à l'échelle du génome) a également permis de classer les promoteurs basaux en deux catégories d'après la répartition des TSS (Carninci *et al.*, 2006). La première correspond à des régions pour lesquelles la

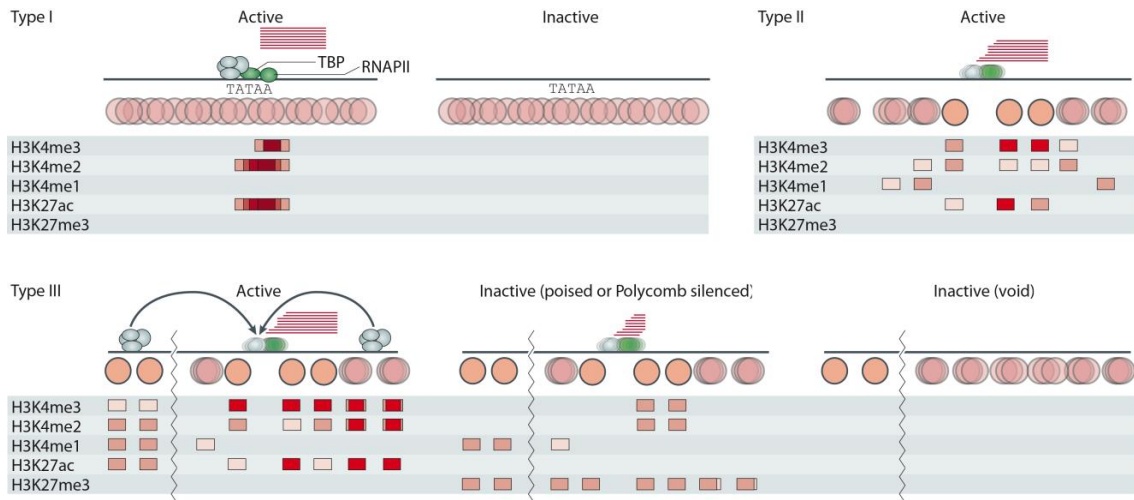


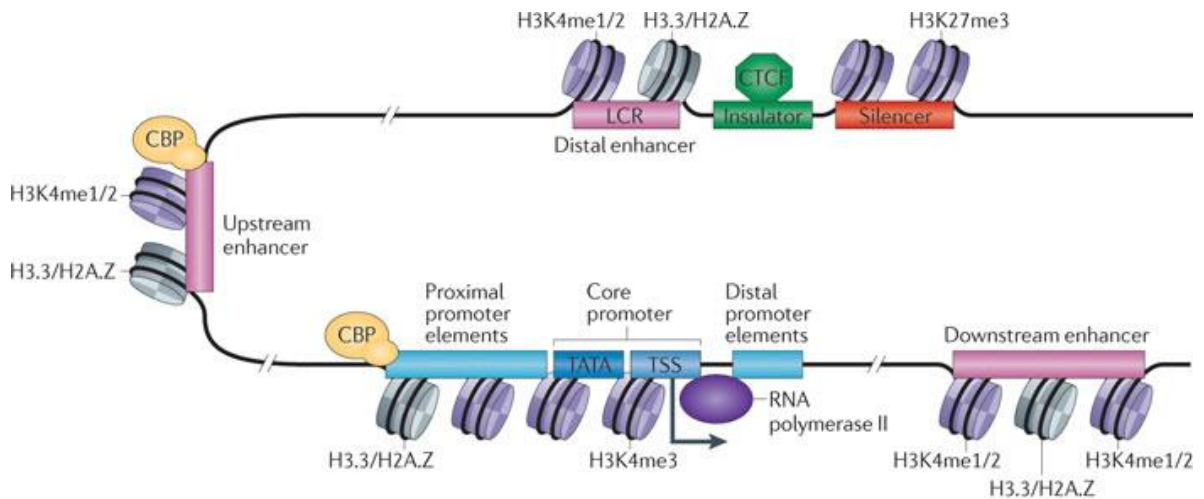
Figure 1. Caractéristiques des principales classes de promoteurs chez les Métazoaires (Lenhard *et al.*, 2012).

Chez les Métazoaires, la plupart des promoteurs peuvent être classés selon trois groupes généraux, établis à partir de la position du TSS et des nucléosomes ainsi que des marques épigénétiques associées. Les promoteurs de type I contrôlent le plus fréquemment l'expression spécifique de gènes dans les tissus périphériques différenciés chez l'adulte. Les promoteurs de type II sont actifs dans tous les types cellulaires. Les promoteurs de type III sont caractéristiques des gènes dont l'expression est régulée lors du développement et coordonnée à travers plusieurs cellules.

Dans ce schéma, l'ADN est représenté par la ligne noire horizontale ; les lignes rouges correspondent aux extrémités 5' des transcrits. Les complexes de facteurs de transcription sont en gris, tandis que les facteurs généraux de la transcription sont représentés en vert. Les parties floues correspondent au positionnement putatif de l'ARN polymérase II pour les initiations indépendantes de la boîte TATA au niveau des promoteurs « broad ». En dessous de chaque promoteur, les nucléosomes sont représentés par des traits rouges, plus ou moins flous selon le degré de positionnement des nucléosomes. Les modifications d'histones sont présentes en dessous des nucléosomes et l'intensité de couleur représente la prévalence de la modification (rouge correspondant au plus prévalent). Pour les promoteurs de type III, les scénarios avec l'ARN polymérase II « poised » et/ou la répression par les complexes Polycomb, ainsi que sans l'ARN polymérase II sont montrés.

transcription démarre au niveau d'un seul ou d'un petit nombre restreint de nucléotides. De façon intéressante, cette catégorie de promoteurs comprend, entre autres, les promoteurs à boîte TATA et correspond aux séquences d'initiation de gènes étroitement régulés tels que les gènes spécifiques d'un type cellulaire (Schug *et al.*, 2005). L'existence de cette classe de promoteurs, appelée « focused » ou « sharp » (promoteurs de type I – Figure 1) est apparemment une constante évolutive ; ils constituent le mode principal de régulation de la transcription chez les organismes les plus simples. Le second type de promoteurs inclut, quant à lui, de multiples sites d'initiations plus faibles, sur une région d'environ 50 à 100 nucléotides. Ceci coïncide avec la présence de régions denses en dinucléotides CG, *i.e.* des îlots CpG, et la transcription de gènes ubiquitaires (Schug *et al.*, 2005 ; Yamashita *et al.*, 2005). Chez les vertébrés, 70% des gènes codants sont régulés par des promoteurs appartenant à cette catégorie. Il est intéressant de noter qu'un sous-ensemble de ces régions peut encore définir un autre type de promoteur ayant les caractéristiques suivantes : (i) utilisation de plusieurs TSS, (ii) faible composition en trinuécléotides ATG afin d'assurer la production d'un seul produit protéique, (iii) absence de boîte TATA et, (iv) indépendance vis-à-vis de l'enrichissement en CpG (Lee *et al.*, 2005a).

Plus récemment, cette seconde catégorie de promoteurs, appelés « dispersed » ou « broad » a été scindée en deux sous-ensembles distinguant (i) les promoteurs de type II, régulant l'expression de gènes ubiquitaires, caractérisés par un petit îlot CpG qui chevauche le TSS, et (ii) les promoteurs de type III, régulant l'expression des gènes de développement, caractérisés par plusieurs îlots CpG qui s'étendent parfois jusqu'au corps de gène (Akalin *et al.*, 2009 – Figure 1). Cependant, une autre étude a quant à elle démontré que le positionnement des nucléosomes et les modifications post-traductionnelles des histones sont des critères plus adaptés à la distinction des différents types de promoteurs basaux que la présence d'îlots CpG (Rach *et al.*, 2011). En effet, les deux catégories de promoteurs, « sharp » et « broad » contiennent des îlots CpG, et tandis que les caractéristiques chromatinienne et le schéma d'initiation de la transcription sont conservés chez les métazoaires, les îlots CpG n'existent pas chez la *Drosophile*. Ceci suggère l'existence de caractéristiques additionnelles pouvant discriminer les différents types de promoteurs. Rach et collègues ont ainsi démontré qu'en plus de leur particularité de séquence, les différents types de promoteurs se distinguent également par l'organisation de leurs nucléosomes et leur structure chromatinienne (Rach *et al.*, 2011). Ainsi, au niveau des promoteurs « broad », les



Nature Reviews | Genetics

Figure 2. Séquences régulatrices de la transcription chez les Métazoaires (Ong et Corces., 2011).

Le promoteur est typiquement constitué d'une séquence promotrice basale (appelée ici « core promoter ») et d'une séquence promotrice proximale. La transcription génique peut être régulée par de multiples enhancers, ainsi que par des silencers et insulateurs, situés généralement à de longues distances des promoteurs. Les modifications d'histones et les variants associés à ces différentes séquences régulatrices sont indiqués.

nucléosomes ont une organisation périodique bien définie et sont enrichis en variant H2A.Z tandis que les promoteurs « sharp » ont une structure des nucléosomes moins organisée (Rach *et al.*, 2011). Dans leur état actif, les promoteurs de type I, II et III sont de plus caractérisés par la présence des marques de chromatine H3K4me3 et H3K27ac, qui sont particulièrement centrées sur le TSS pour les types I (Heintzman *et al.*, 2007). De façon intéressante, les promoteurs de type III peuvent porter à la fois les marques H3K4me3 et H3K27me3, respectivement associées à l'activation et à la répression, qualifiant ces promoteurs de « bivalents » (Alkaline *et al.*, 2009 ; Bernstein *et al.*, 2006). Cette bivalence permettrait de réprimer les gènes de développement tout en les gardant dans un état « prêt » (également appelé « poised ») pour l'activation (Bernstein *et al.*, 2006).

Quelles que soient leurs caractéristiques, les promoteurs basaux agissent en tant que plateformes de recrutement du complexe basal de transcription. Grâce à leur diversité de composition (motifs, îlots CpG ...), d'organisation et de structure des nucléosomes (variants, modifications de chromatine...), ces régions modulent l'étape initiale de la transcription et peuvent donc être considérées comme de vraies séquences régulatrices.

b – Promoteur proximal et enhancer

Par définition, le promoteur proximal est une séquence d'ADN située entre 100 et 200 pb en amont d'un TSS tandis qu'un enhancer correspond à une région génomique distale retrouvée en amont, en aval ou à l'intérieur du gène cible et qui peut effectuer son action régulatrice indépendamment de son orientation (Figure 2). Outre leur localisation plus ou moins proche du TSS, ces deux types de séquences régulatrices ont la même fonction : moduler les niveaux de transcription. En effet, les promoteurs proximaux ainsi que les enhancers contiennent des sites de reconnaissance permettant le recrutement de protéines stimulatrices ou inhibitrices de la transcription. De plus, ces séquences possèdent une activité spécifique du type cellulaire. Une description plus complète des régions enhancers est réalisée dans la partie III de cette introduction.

c – Insulateur

Les insulateurs sont des séquences qui permettent d'assurer une régulation supplémentaire en séparant les gènes différenciellement exprimés et en les protégeant de signaux inappropriés. Cette régulation peut s'effectuer de deux manières. La première

Tableau 3. Caractéristiques structurales des familles de facteurs de transcription.

Famille	Spécificité	Domaine de liaison à l'ADN
« Helix-loop-helix »	Deux hélices α séparées par une boucle et un tour	Hélices α
« Winged-Helix » (Forkhead)	Quatres hélices, deux feuilletts β et une structure « winged helix »	Hélice α et les « wings » flanquant l'hélice.
« Zinc finger »	Enchainement de domaines compacts stabilisés par la liaison de résidus cystéine et histidine à un ion Zn^{2+} central	Hélices formées par la partie C-terminale de chaque doigt de zinc
« Leucine-Zipper »	Domaine de dimérisation riche en Leucine	Région basique adjacente à la région riche en leucine

consiste à bloquer la communication entre un enhancer et un promoteur tandis que la seconde permet d'empêcher la propagation de l'hétérochromatine, dans ce cas, l'insulateur agit comme une « barrière » (Gaszner et Felsenfeld, 2006). La protéine CTCF (« CCCTC-binding factor ») a été décrite comme étant la principale protéine insulatrice chez les vertébrés. Elle est d'ailleurs très conservée et sert de marqueur afin de déterminer les régions insulatrices à l'échelle du génome (Barski *et al.*, 2007 ; Kim *et al.*, 2007).

2 – Les facteurs de transcription

La caractérisation des séquences régulatrices a conduit à l'identification de protéines qui se fixent de manière spécifique sur ces régions. L'utilisation des techniques d'analyse de mutants, d'empreinte et de gel retard ont ainsi mis en lumière une vaste famille de protéines présente chez tous les organismes vivants : celle des FT (Latchman, 1990). Ces protéines sont essentielles à la régulation de l'expression génique et jouent donc un rôle crucial pour le développement et les diverses fonctions cellulaires. Par définition, ces protéines incluent deux domaines essentiels à leur fonction régulatrice, l'un permettant la fixation à l'ADN et l'autre l'établissement d'interactions protéine-protéine, reliés tous deux par une région polypeptidique flexible.

Le domaine de liaison à l'ADN permet aux FT de reconnaître et de se lier spécifiquement à une séquence d'ADN donnée qui peut être située au niveau de régions *cis*-régulatrices des gènes cibles. La diversité de structure de ces domaines a permis de les classer en plusieurs sous-familles. Les plus connues et représentées de ces dernières sont les « Helix-loop-Helix », les « Winged-Helix » ou « Forkhead », les « Zinc Finger » ou encore les « leucine-Zipper » (Tableau 3). De par leur structure, certaines protéines se lient à l'ADN sous formes de monomères tandis que d'autres ont la capacité de former des homo- ou même des hétéro-dimères. Les FT n'agissent néanmoins pas seulement *via* leur domaine de liaison à l'ADN, mais également *via* des domaines d'interaction protéine-protéine permettant la reconnaissance d'autres facteurs ou de l'ARN Polymérase elle-même. Ces domaines sont également définis comme activateurs ou répresseurs selon les régulations mises en jeu.

La localisation des sites de liaison des FT par rapport à leurs gènes cibles et leurs rôles dans la régulation transcriptionnelle ont permis d'établir deux principales catégories de FT. La première comprend les protéines qui se fixent à proximité du TSS et qui participent à la

formation de la machinerie transcriptionnelle ainsi qu'à l'initiation de la transcription. Il s'agit des facteurs généraux de la transcription (GTF) TFIIA, TFIIB, TFIID, TFIIIE, TFIIF et TFIIH (Roeder, 1996). La seconde catégorie regroupe les FT qui se fixent aux autres séquences régulatrices et qui stimulent ou répriment la transcription du gène cible.

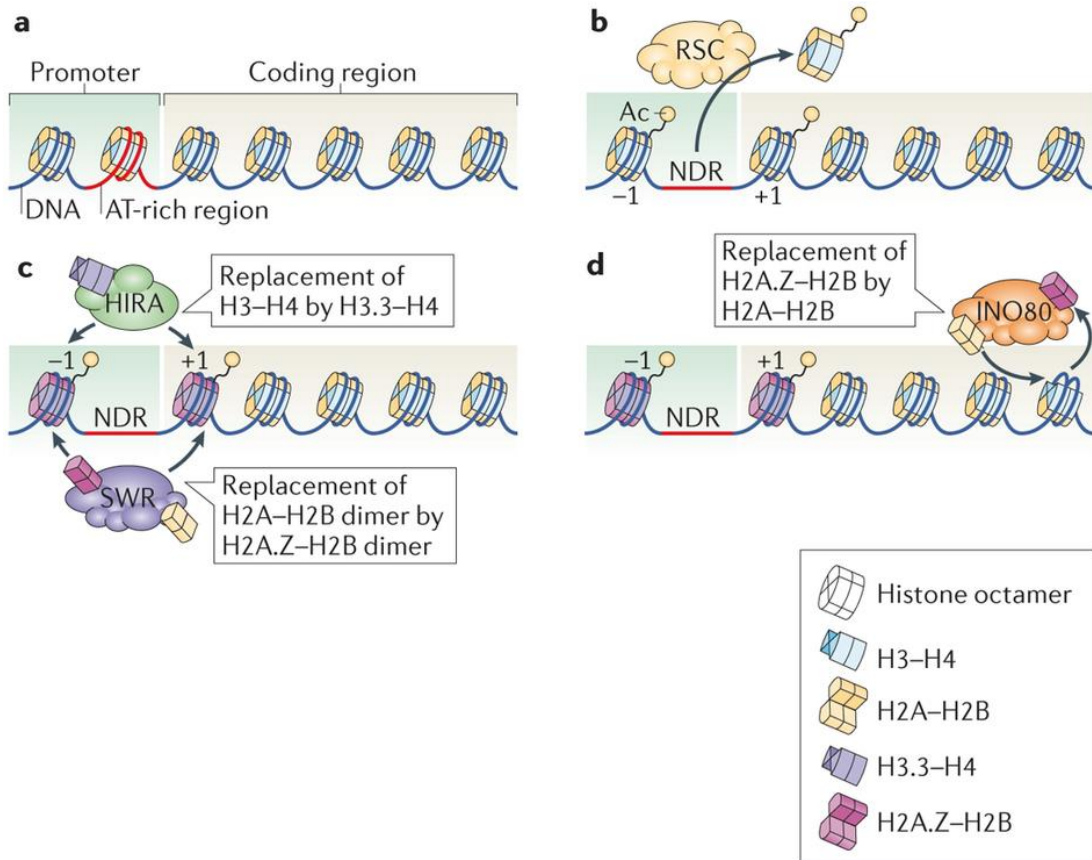
La diversité de structure et de composition des FT leur permettent ainsi d'établir différents modes de régulation (dont certains exemples seront décrits dans la partie suivante de cette introduction : C-Principes de régulation au cours des différentes étapes de la transcription) : d'une part, la stabilisation ou l'obstruction de la fixation de l'ARN Polymérase sur l'ADN et d'autre part, le recrutement de protéines co-activatrices ou co-répressives de la transcription.

C – Principes de régulation au cours des différentes étapes de la transcription

La transcription de l'ADN est un processus séquentiel qui se déroule en trois étapes principales : initiation, élongation et terminaison. Les processus les mieux décrits dans la littérature et étant considérés comme majeurs pour la régulation de l'expression génique se déroulent lors de l'activation transcriptionnelle. Celle-ci nécessite une restructuration chromatinienne, l'assemblage du complexe de pré-initiation et la transition vers l'étape d'élongation. Dans cette partie, nous aborderons donc les mécanismes de régulation ayant lieu lors de l'initiation et de l'élongation de la transcription chez les eucaryotes. Ces processus mettent en jeu une coordination fine des différents acteurs précédemment introduits (structure chromatinienne, séquences régulatrices et facteurs de transcription).

1 – L'initiation de la transcription

La première étape de la transcription, l'initiation, est mise en œuvre après une première restructuration de la chromatine qui permet et stabilise la liaison du complexe basal de transcription (GTF + ARN Pol II) au niveau du promoteur basal formant ainsi le complexe de pré-initiation de la transcription (PIC, pour « pre-initiation complex »), permettant d'initier la synthèse de l'ARN.



Nature Reviews | Molecular Cell Biology

Figure 3. Structure de la chromatine et initiation de la transcription (Venkatesh et Workman, 2015).

(a) Distribution des nucléosomes au niveau des gènes avant l'initiation de la transcription. L'ADN des promoteurs est généralement riche en nucléotides AT (ligne rouge), ce qui ne favorise pas la courbure de l'ADN autour du nucléosome. (b) De plus, le complexe de remodelage RSC enlève ces nucléosomes et crée une région appauvrie en nucléosomes (NDR) au niveau du promoteur. RSC permet également le maintien de la NDR de façon dépendante de l'acétylation des histones. (c) L'acétylation des histones cible le complexe SWR au niveau de la NDR, permettant le remplacement des dimères canoniques H2A-H2B par le variant H2A.Z-H2B au sein des nucléosomes encadrant cette région (positions -1 et +1). De plus, le remplacement de H3-H4 par le variant H3.3-H4, effectué par le chaperon d'histone HIRA, crée des nucléosomes incluant une double substitution des histones canoniques, plus instables et favorisant l'élongation de la transcription. (d) L'acétylation d'H2AZ au niveau du promoteur permet le maintien de nucléosomes incluant ce variant, en empêchant l'action de INO80. A l'inverse, une incorporation incorrecte d'H2A.Z au niveau de nucléosomes non acétylés dans la région codante est empêchée par l'action d'INO80.

a – Restructuration de la chromatine

Le recrutement de la machinerie transcriptionnelle nécessite la présence d'une structure chromatinienne favorable au niveau des promoteurs de gènes actifs. L'action de complexes de remodelage permet la création de régions peu enrichies en nucléosomes (NDR, pour « nucleosome depleted regions ») qui, cependant ne suffisent pas à elles seules à assurer l'initiation de la transcription. Les différents acteurs de régulation de la structure chromatinienne (échange/variants/modifications d'histones etc...) jouent donc ici un rôle essentiel pour la déstabilisation des nucléosomes entourant les NDR (Figure 3).

b – Mise en place du complexe de pré-initiation de la transcription

L'étape initiale de la formation du PIC est la reconnaissance du promoteur basal par la machinerie transcriptionnelle. Aux vues de la diversité de composition et de structure des promoteurs basaux, la formation du PIC peut s'effectuer de différentes manières *via* l'intervention de protéines spécifiques.

L'acteur principal de la mise en place du PIC est le facteur TFIID, composé d'une sous-unité de liaison à la TATA-box (TBP, pour « TATA box-binding protein ») et de sous-unités nommées TAF (pour « TBP-associated factors »). Certaines de ces sous-unités reconnaissent spécifiquement les éléments composant les promoteurs basaux : TBP se lie à la TATA box (Patikoglou *et al.*, 1999), TAF1 reconnaît le DCE (Lee *et al.*, 2005b) tandis que la reconnaissance de l'élément Inr nécessite à la fois TAF1 et TAF2 (Chalkley et Verrijzer, 1999), TAF6 et TAF9 lient le DPE (Burke et Kadonaga, 1997). Concernant les promoteurs sans boîte TATA, pour lesquels un consensus pyrimidine/purine (py/pu) favorable à l'initiation de la transcription définit le TSS (Carninci *et al.*, 2006), le complexe basal d'initiation peut être recruté *via* la reconnaissance de cette séquence py/pu par certaines sous-unités de TFIID. Ainsi la variabilité de formation du PIC, notamment due à la flexibilité de composition de TFIID, participe activement à la régulation transcriptionnelle de divers processus biologiques par la transcription de gènes ayant des fonctions cellulaires ou tissus spécifiques (Goodrich et Tjian, 2010). Il est à noter que les modifications chromatinienne influencent également le recrutement de TFIID (Jacobson *et al.*, 2000 ; Vermeulen *et al.*, 2007).

Après la liaison de TFIID sur le promoteur basal, il existe deux modes de recrutement des autres éléments du PIC : l'assemblage séquentiel des différents facteurs selon un ordre défini, ou le recrutement d'un complexe préformé contenant tous les GTFs (excepté TFIID), l'ARN Pol II ainsi que des co-régulateurs, c'est le modèle holoenzyme (Koleske et Young, 1994). La formation du PIC permet ensuite d'assurer l'ouverture de l'ADN pour permettre l'accès à l'ARN Pol II, grâce aux activités hélicases de TFIIH (Holstege *et al.*, 1996),

c – Co-régulateurs de l'initiation

Les niveaux de transcription peuvent être modulés par l'action de co-régulateurs recrutés par les facteurs de transcription présents au niveau des séquences régulatrices.

Le plus connu parmi les co-activateurs est le complexe Médiateur. Ce complexe, présent chez tous les eucaryotes et composé de 26 sous-unités, a été décrit comme permettant d'établir un pont entre les facteurs de régulation et la machinerie de transcription basale. En effet, grâce à des interactions protéine-protéine, le complexe Médiateur est capable d'interagir avec un grand nombre de FT (dont les GTFs) et avec l'ARN Pol II (Esnault *et al.*, 2008 ; Kim *et al.*, 1994). Cette capacité de reconnaissance est due à la composition du complexe Médiateur dont la nature des sous-unités peut varier afin de moduler sa fonction biologique. Le complexe Médiateur ne permet pas seulement d'augmenter le niveau de transcription d'un gène mais également de stabiliser la formation du PIC. Outre l'initiation de la transcription, il régule également le phénomène d'entrée et /ou de sortie de pause de l'ARN Pol II, ainsi que l'architecture chromatinienne en assurant le maintien des NDR et en participant à la stabilisation des boucles d'ADN (Allen et Taatjes, 2015).

2 – L'élongation de la transcription

a – Elongation précoce

Après synthèse de quelques nucléotides (~10), le passage de l'étape d'initiation à celle de l'élongation est assuré par la phosphorylation de la sérine 5 du domaine carboxy-terminal (CTD) de l'ARN Pol II par TFIIH (Dvir, 2002). Ceci entraîne la déstabilisation de l'interaction entre l'ARN Pol II et les autres composants du PIC, permettant ainsi de favoriser le processus d'échappement du promoteur et d'amorcer l'élongation. Cette modification du

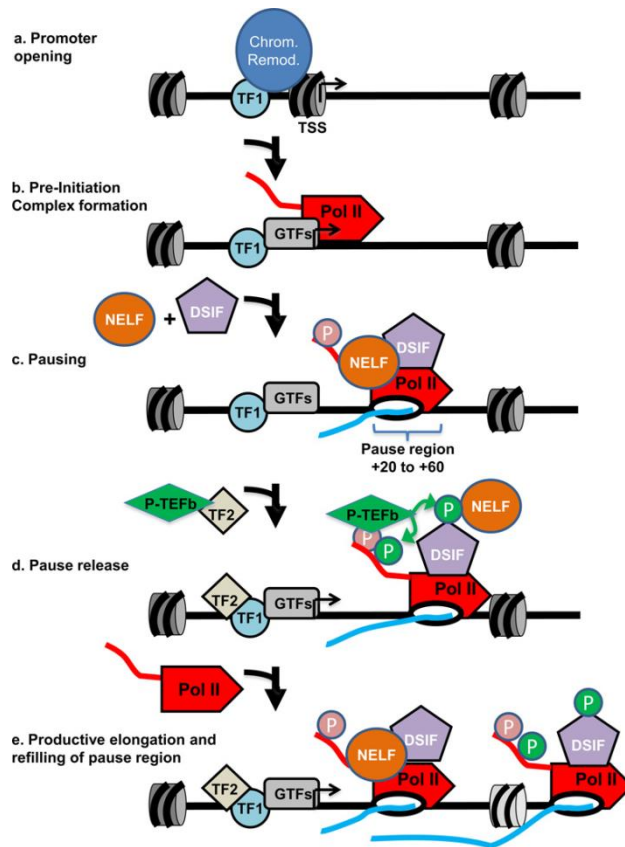


Figure 4. Mise en place et relargage de l'ARN Pol II en « pause » (Adelman et Lis, 2012).

La région promotrice d'un gène transcrit par l'ARN Pol II est représentée sur ce schéma, le site d'initiation de la transcription (TSS) étant indiqué par une flèche, les nucléosomes en gris et l'ARN Pol II étant illustrée par une « fusée » rouge. Le transcrit naissant est représenté en bleu et une accolade indique la région où la « pause » est effectuée. Les facteurs impliqués dans la mise en place et le relargage de l'ARN Pol II en « pause », tels que DSIF (pentagone violet), NELF (ovale orange) et P-TEFb (losange vert) sont indiqués. **(a)** L'ouverture du promoteur implique souvent la liaison de facteurs de transcription (ici, TF1 en rond bleu clair) qui recrutent des complexes de remodelage de la chromatine (ovale bleu) pour enlever les nucléosomes autour du TSS et rendre le promoteur accessible pour le recrutement de la machinerie transcriptionnelle. **(b)** La formation du complexe de pré-initiation implique le recrutement de facteurs généraux de la transcription (GTF, ovale gris) et de l'ARN Pol II également facilitée par la liaison de facteurs de transcription spécifiques. Cette étape précède l'initiation de la synthèse d'ARN. **(c)** La « pause » de Pol II se produit peu après l'initiation de la transcription et implique l'association des facteurs de « pause » DSIF et NELF. L'ARN Pol II en pause est phosphorylée sur son CTD (en rose). **(d)** La sortie de la « pause » est provoquée par le recrutement de la kinase P-TEFb, directement ou indirectement par un facteur de transcription (ici TF2, losange). P-TEFb phosphoryle le complexe DSIF/NELF pour libérer l'ARN Pol II en « pause » et cible également son CTD (en vert). La phosphorylation de DSIF/NELF dissocie NELF du complexe d'élongation et transforme DSIF en un facteur d'élongation positif qui s'associe avec l'ARN Pol II à travers le gène. **(e)** En présence de TF1 et TF2, l'échappement de la « pause » de l'ARN Pol II vers l'élongation productive est rapidement suivie par l'entrée d'une nouvelle ARN Pol II dans le site de « pause », permettant une production efficace d'ARN.

CTD permettrait également le recrutement de facteurs impliqués dans l'élongation précoce ou dans la mise en place de modifications d'histones.

Cependant, de nombreuses études ont démontré un ralentissement de l'ARN Pol II de 20 à 50 pb en aval du TSS. Très répandue chez les Métazoaires, cette phase de « pause » concerne particulièrement la transcription des gènes impliqués dans des voies de réponses à des signaux environnementaux ou liés au développement (Zeitlinger *et al.*, 2007). La mise en place de cette phase est dépendante de deux protéines, DSIF (« DRB Sensitivity Inducing Factor ») et NELF (« negative elongation factor »), qui forment un complexe stable avec l'ARN Pol II (Yamaguchi *et al.*, 1999). Précédemment jugée comme une étape sans importance fonctionnelle, la pause de l'élongation jouerait néanmoins plusieurs rôles dans le contrôle de la transcription. Elle pourrait permettre premièrement, d'assurer le maintien d'une chromatine permissive, et deuxièmement de déclencher rapidement la transcription en réponse à un stimulus, ceci ayant un intérêt notamment pour la transcription de gènes codant les protéines de choc thermique ou les protéines impliquées dans le développement. A la suite de la phase de « pause » deux voies sont alors possible : terminaison de la transcription par relargage de petits ARN ou élongation productive.

b – Elongation productive

La sortie de la phase de « pause » est assurée par le recrutement du complexe P-TEFb (pour « Positive Transcription Elongation Factor b » – Figure 4). P-TEFb phosphoryle d'une part, le complexe répresseur DSIF-NELF, entraînant la dissociation de NELF de l'ARN Pol II, et d'autre part, le CTD de l'ARN Pol II au niveau de la sérine 2. Ainsi, l'activité kinase de PTEFb permet de lever l'inhibition d'élongation et favorise le recrutement de divers complexes associés à la polymérase durant l'élongation (Ni *et al.*, 2008). D'autres protéines sont également impliquées dans la régulation de l'étape d'élongation, telles que c-myc et BRD4 qui recrutent P-TEFb au niveau de promoteurs (Eberhardy et Farnham, 2002 ; Yang *et al.*, 2005). Une fois l'élongation productive amorcée, le complexe contenant l'ARN Pol II est stabilisé et peut transcrire des dizaines voire des centaines de kilobases avant de se dissocier de l'ADN.

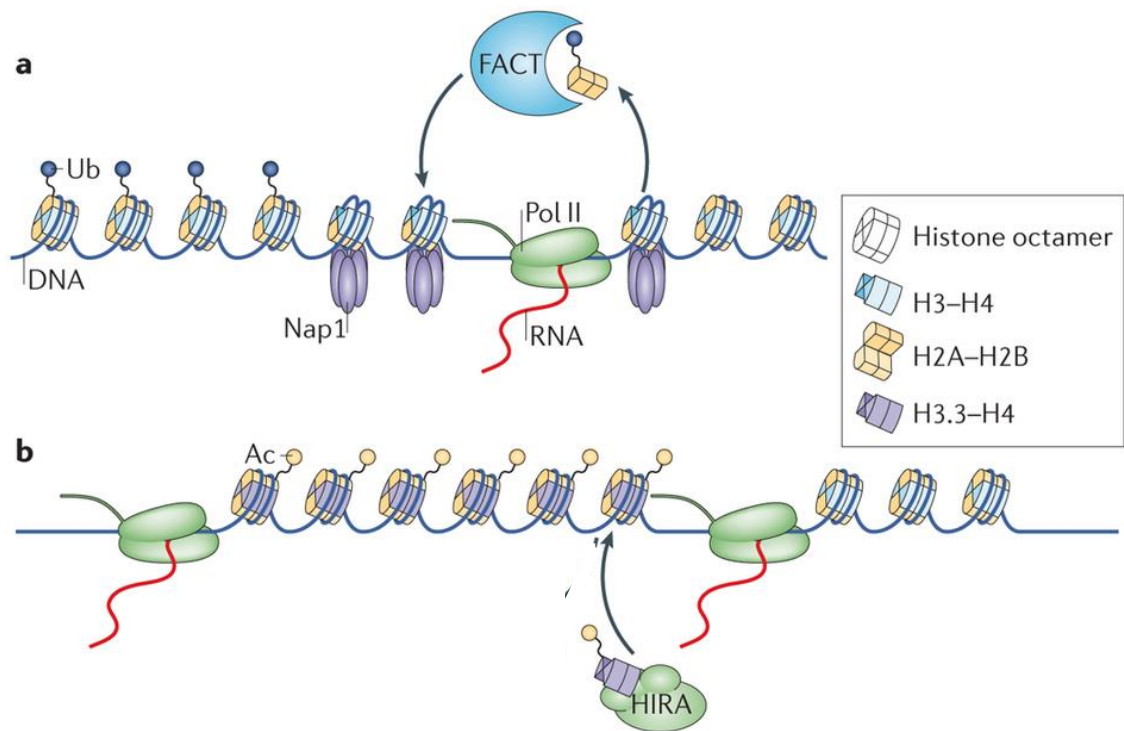


Figure 5. Structure de la chromatine et élongation de la transcription (Venkatesh et Workman, 2015).

(a) La passage de l'ARN polymérase en activité d'élongation est aidée par le chaperon d'histone FACT, qui cible les dimères H2A-H2B ubiquitylés (Ub) pour les enlever. Le nucléosome hexamérique résultant est stabilisé par l'histone chaperon Nap1 chez la levure. Ceci facilite les multiples cycles de transcription sur ce nucléosome sans destabilisation supplémentaire. (b) Bien que le remplacement du tétramère H3-H4 soit rare chez les eucaryotes, le remplacement d'H3.3 assuré par HIRA au niveau des corps de gènes est commun chez les Métazoaires. Ce processus est lié au taux de transcription. L'acétylation (Ac) des nucléosomes au niveau des corps de gènes facilite également les échanges d'histones par diverses chaperons, qui remplacent les nucléosomes existants avec des nucléosomes pré-acétylés, aidant efficacement le processus de transcription.

c – Structure chromatinienne

Néanmoins, afin de mener à bien l'élongation, l'ARN Pol II doit pouvoir parcourir « aisément » l'ADN. Cette étape implique donc une nouvelle restructuration de la chromatine et des nucléosomes par les différents acteurs précédemment introduits. Certaines protéines et complexes sont spécifiquement impliquées dans ce processus, tel que le complexe protéique FACT (« facilitates chromatin transcription ») qui participe à la désorganisation des nucléosomes induite par la présence de l'ARN Pol II, mais qui est également nécessaire au réassemblage ultérieur des nucléosomes (Belotserkovskaya *et al.*, 2003 ; Orphanides *et al.*, 1999). Le facteur général de transcription TFIIF, qui possède une activité hélicase, facilite quant à lui la transcription à travers la chromatine, et les protéines Spt (« Suppressor of Ty ») participent à la reconstruction de la chromatine durant l'élongation (Kulaeva *et al.*, 2007 – Figure 5).

D – Organisation du génome

Il est maintenant admis que la régulation transcriptionnelle n'implique pas seulement les séquences régulatrices et les acteurs moléculaires associés à ces séquences (modifications des histones, facteurs de transcription...). En effet, la combinaison d'un très grand nombre d'études a également permis de montrer l'importance de l'architecture nucléaire dans le contrôle des processus transcriptionnels. Ainsi, chez les eucaryotes, le noyau des cellules est organisé en de multiples domaines nucléaires. Parmi ces structures particulières, la lamina nucléaire, le nucléole, les PML et Cajal « bodies », ainsi que les « nuclear speckles » jouent un rôle dans la régulation transcriptionnelle (Spector, 2003). De plus, de nombreux complexes protéiques nécessaires à la transcription (usines de transcription) ou à sa répression (complexes PcG) montrent une distribution non aléatoire dans le nucléoplasme (Spector, 2003). A l'échelle même du génome, le matériel génétique est replié et compacté dans le noyau sous forme de structures particulières qui contribuent à la régulation de l'expression génique. Dans cette partie, nous ne développerons que ce dernier aspect tridimensionnel de l'architecture nucléaire dont l'importance structurelle et fonctionnelle a été démontrée grâce aux approches de microscopies photoniques et plus récemment, aux méthodes de « chromosome conformation capture » *i.e.* 3C et ses dérivés 4C/5C (« Circularized

Chromosome Conformation Capture/ Carbon-Copy Chromosome Conformation Capture ») et HiC (Miele et Dekker, 2009).

1 – Territoires chromosomiques

Différentes études sur l'architecture du noyau et l'observation des chromosomes réalisées entre la fin du 19^{ème} et le début du 20^{ème} siècle, ont menées aux premières suppositions d'un positionnement non aléatoire des chromosomes, sous la forme de territoires, lors de l'interphase. C'est ainsi qu'apparaît la notion de territoires chromosomiques (Boveri, 1909 ; Rabl, 1885) confirmée seulement près d'un siècle plus tard, à la fin des années 1970s (Stack *et al.*, 1977). Ces territoires sont aujourd'hui observés en routine grâce à la technique de FISH (pour « fluorescence in situ hybridization » – Bolzer *et al.*, 2005 ; Croft *et al.*, 1999). Cette ségrégation spatiale non aléatoire des chromosomes au sein du noyau interphasique serait en partie due à la densité de gènes de chaque chromosome. En effet, les chromosomes ayant une forte densité en gènes auraient tendance à avoir une localisation nucléaire centrale tandis que les chromosomes ayant une plus faible densité de gènes seraient localisés à la périphérie du noyau (Boyle *et al.*, 2001 ; Croft *et al.*, 1999). Cependant, cette classification selon la densité en gènes ne semble pas être une généralité. D'autres études suggèrent ainsi que la compartimentation observée serait due à la taille des chromosomes, l'activité transcriptionnelle ou encore la composition en nucléotides GC (Bolzer *et al.*, 2005 ; Goetze *et al.*, 2007 ; Hepperger *et al.*, 2008). De façon intéressante, dans un même type cellulaire, ici les lymphocytes, l'organisation tridimensionnelle des territoires chromosomiques est conservée chez les primates supérieurs (Tanabe *et al.*, 2002) alors qu'elle diffère entre divers types cellulaires (Mayer *et al.*, 2005 ; Parada *et al.*, 2004), suggérant que l'organisation spatiale du génome de façon tissu-spécifique joue un rôle dans l'établissement des territoires chromosomiques.

L'utilisation des différentes techniques de « chromosome conformation capture » a également permis de montrer que, d'une façon générale chez les métazoaires, l'organisation génomique discrimine les régions comportant une chromatine active des régions ayant une chromatine inactive (Simonis *et al.*, 2006). Ainsi, les régions dites actives ont une forte densité en gènes, une chromatine ouverte, accessible et activement transcrite, et tendent à s'associer les unes avec les autres, principalement par des interactions intra-chromosomique ;

tandis que les régions dites inactives se contactent principalement de façon inter-chromosomique (Lieberman-Aiden *et al.*, 2009 ; Simonis *et al.*, 2006 ; Yaffe et Tanay, 2011).

2 – Domaines chromatinien : à l'échelle de la cellule

La première classification des régions chromatinien a introduit les termes d'euchromatine et d'hétérochromatine afin de distinguer respectivement les régions actives des régions inactives. L'euchromatine définie dans le paragraphe précédent comme correspondant aux régions actives (*i.e.* ayant une forte activité transcriptionnelle) et dont la chromatine est peu condensée, est majoritairement retrouvée au centre du noyau tandis que l'hétérochromatine est associée à la périphérie nucléaire dans la plupart des types cellulaires et correspond aux régions chromatinien très condensées et généralement inactives. Elle est communément divisée en deux catégories : l'hétérochromatine constitutive et l'hétérochromatine facultative. L'hétérochromatine facultative correspond à des régions d'euchromatine qui deviennent silencieuses et condensées de façon spécifique du type cellulaire, alors que l'hétérochromatine constitutive reste condensée indépendamment du contexte cellulaire. Chez les métazoaires, l'hétérochromatine constitutive correspond à de larges domaines adjacents aux centromères (*i.e.* l'hétérochromatine péri-centrique) et aux télomères. Elle présente une faible densité de gènes mais est enrichie en répétitions, en ADN satellite et en éléments transposables silencieux. Elle est caractérisée par la présence des marques d'histones H3K9me2/me3 et par l'absence d'H3K27me3 et de marques d'activation (Ernst et Kellis, 2011 ; Zhu *et al.*, 2013 ; Fillion *et al.*, 2010). L'hétérochromatine constitutive est enrichie en protéines spécifiques telles que HP1 (pour « heterochromatin protein 1 ») qui reconnaît la méthylation d'H3K9 grâce à son chromodomaine et interagit avec une variété de protéines chromosomiques *via* son domaine « chromo-shadow », ce qui lui permet d'agir à la fois sur la structure chromatinien et sur la répression des gènes (Fillion *et al.*, 2010 ; Lachner *et al.*, 2001).

Contrairement à ce qui a été initialement suggéré, l'hétérochromatine constitutive n'est en fait un environnement restrictif ni à la fixation des facteurs de transcription, ni aux processus de transcription. En effet, certains FT peuvent accéder à des sites de liaison présentant des caractéristiques hétérochromatiniennes et peuvent même être déterminants pour l'établissement de telles régions (Bulut-Karslioglu *et al.*, 2012 ; Lundgren *et al.*, 2000). De plus l'hétérochromatine constitutive contient des gènes essentiels pour la viabilité et la

fertilité qui y sont activement transcrits (Dimitri *et al.*, 2009). La faible activité transcriptionnelle de l'hétérochromatine péri-nucléaire a été suggérée comme étant une conséquence de la répression génique observée au niveau de la lamina nucléaire (Reddy *et al.*, 2008). En effet, environ 40 % du génome des mammifères interagirait avec la lamina nucléaire formant ainsi de larges régions de tailles variables, pauvres en gènes, enrichies en H3K9me2 et H3K27me3. Ces domaines, appelés LAD (pour « lamin-associated domains ») sont généralement associés à une répression transcriptionnelle (Guelen *et al.*, 2008 ; Peric-Hupkes *et al.*, 2010). En accord avec les observations faites concernant la régulation transcriptionnelle au niveau de l'hétérochromatine constitutive, des études complémentaires ont montré que la relocalisation de *loci* à la périphérie nucléaire n'entraîne la répression que d'une partie d'entre eux (Finlan *et al.*, 2008 ; Kumaran et Spector, 2008). Ainsi, ce n'est probablement pas la localisation d'un gène au niveau de régions d'euchromatine ou d'hétérochromatine qui régule l'activité génique mais plutôt d'autres facteurs tels que les modifications d'histones ou d'ADN. En revanche d'autres études ont observé une relocalisation de certains *loci* de la périphérie vers le noyau associée à leur activation ou permettant leur activation ultérieure lors de certains événements de différenciation suggérant que les motifs d'interaction avec la périphérie sont en partie spécifiques du type cellulaire (Kosak *et al.*, 2002 ; Peric-Hupkes *et al.*, 2010 ; Williams *et al.*, 2006).

Cette classification euchromatine *versus* hétérochromatine reste cependant trop généraliste et simpliste. Le développement des techniques d'analyses à l'échelle du génome a permis d'affiner cette classification en fonction de l'enrichissement des modifications post-traductionnelles des histones, des protéines recrutées et des fonctions de régulations exercées par ces différentes régions (Ernst *et al.*, 2011 ; Filion *et al.*, 2010).

3 – Domaines chromatiniens : à l'échelle du chromosome

Récemment, trois études visant à analyser l'organisation spatiale (i) du génome de cellules ES murines et humaines ainsi que de fibroblastes humains (Dixon *et al.*, 2012), ou (ii) d'une région de 4,5 Mb autour du centre d'inactivation du chromosome X (Nora *et al.*, 2012), ou encore (iii) du génome d'embryons de *Drosophila* (Sexton *et al.*, 2012), ont mis en évidence une nouvelle organisation structurale du génome. En effet, l'utilisation de techniques de « chromosome conformation capture » a révélé la présence, dans le génome, de larges régions organisées en structures distinctes les unes des autres, favorisant les

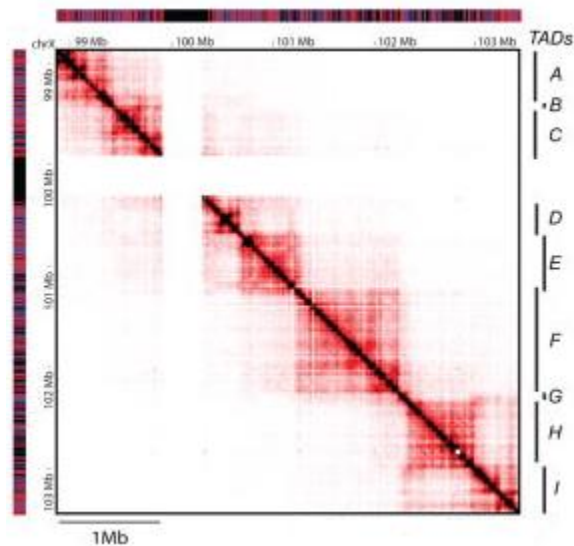


Figure 6. Exemple de carte d'interactions génomiques révélant les « topologically associated domains » (TAD – Nora *et al.*, 2012).

Données obtenues par la technique de 5C à partir de cellules ES murines non différenciées (E14), dont la fréquence est illustrée ici par l'intensité des points. Les contacts chromosomiques sont organisés en blocs génomiques distincts (TAD A-I). Une région contenant des duplications a été exclue de l'analyse des données 5C et est masquée dans cette représentation (partie blanche).

interactions chromatinienne en *cis* plutôt qu'en *trans*. Les cartes d'interactions obtenues montrent (i) une diagonale très marquée reflétant les contacts importants entre *loci* proches au niveau de l'ADN, (ii) différents carrés représentant les régions d'interactions fréquentes au sein d'un même domaine et (iii) des frontières nettes entre ces domaines d'interactions internes (Dixon *et al.*, 2012 ; Nora *et al.*, 2012 ; Sexton *et al.*, 2012 – Figure 6). Ces régions ont été qualifiées de TAD pour « Topologically Associating Domains ». La longueur de ces TAD est variable, entre 200 kb et 1 Mb chez les mammifères (Nora *et al.*, 2012), contre 10 à 100 kb chez la *Drosophile* (Sexton *et al.*, 2012). De façon intéressante, ces longueurs correspondent respectivement à la taille des clusters de co-expression définis chez l'Homme (Gierman *et al.*, 2007) et chez la *Drosophile* (Spellman *et al.*, 2002), suggérant un rôle de ces structures dans la régulation de l'expression des gènes. Dans ce sens, les TAD coïncident avec des régions comportant des signatures épigénomiques caractéristiques des différents domaines chromatinien, mais également avec des ensembles de gènes co-régulés (Nora *et al.*, 2012 ; Sexton *et al.*, 2012).

L'organisation génomique sous forme de TAD et l'emplacement de leurs frontières sont conservés entre différents types cellulaires ainsi qu'entre la Souris et l'Homme (Dixon *et al.*, 2012). Les frontières sont enrichies en gènes de ménage et en régions insultrices (Dixon *et al.*, 2012) liées par CTCF et le complexe cohésine, suggérant un rôle de ces protéines dans leur établissement.

L'importance des TAD dans la régulation transcriptionnelle est soulignée par deux observations. La première provient de l'étude de Nora *et al.*, montrant que lors de la différenciation des cellules ES, les interactions intra-TAD sont remodelées tandis que le TAD en lui-même n'est pas affecté (Nora *et al.*, 2012). Dans un deuxième temps, Lupianez *et al.* ont montré que la perturbation de la frontière d'un TAD par une délétion entraîne de nouvelles interactions promoteur-enhancer menant à une dérégulation de l'expression génique responsable de diverses pathologies (Lupianez *et al.*, 2015).

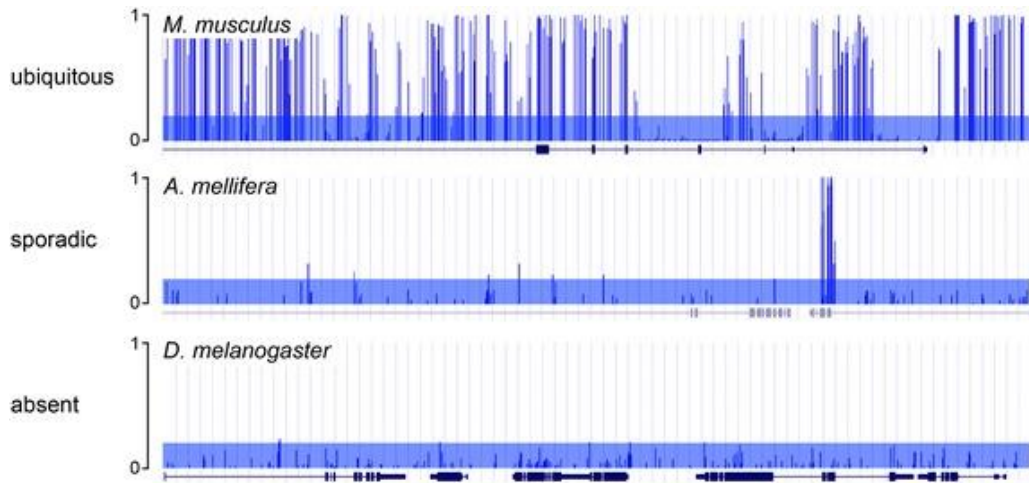


Figure 7. Trois catégories principales de méthylomes (Breiling et Lyko, 2015).

Les trois catégories de méthylation de l'ADN, ubiquitaire, sporadique et absente sont illustrées par trois exemples d'analyse de séquençage au bisulfite pan-génomique chez la Souris (panel du haut), l'Abeille (panel du milieu) et la Drosophile (panel du bas). Une fenêtre de 40 kb a été aléatoirement choisie pour chacune des 3 espèces. Les données montrées sont exprimées en ratios de méthylation pour chaque dinucléotide CpG. Les barres bleues transparentes indiquent le seuil de significativité (ratios de méthylation de 0,2) en dessous duquel les signaux observés sont dus à des artefacts de conversion au bisulfite).

II – LES MODIFICATIONS DES CYTOSINES

De nombreuses études ont identifié les modifications des cytosines comme étant des éléments clés de la régulation de la transcription. A l'heure actuelle, quatre formes de cytosines modifiées ont été détectées chez les mammifères. Intermédiaires réactionnels ou réelles marques épigénétiques, la mise en place, le maintien et l'effacement de ces modifications par des enzymes spécifiques sont des processus cruciaux assurant le développement et la différenciation cellulaire. De ce fait, leur dérégulation est fréquemment associée à des pathologies et aux processus cancéreux. De tels exemples étant trop nombreux dans la littérature pour être décrits dans cette partie, nous aborderons ici principalement les mécanismes moléculaires impliqués dans ces phénomènes fondamentaux. Néanmoins, de nombreuses revues présentent les différentes implications physiopathologiques des modifications des cytosines (*cf.* par exemple Mahé et Salbert, 2015 ; Robertson, 2005 ; Smith et Meissner, 2013).

A – Méthylation de l'ADN

Mise en évidence il y a 40 ans, la méthylation de l'ADN est la plus étudiée des modifications de cytosines. Dans cette partie, nous nous intéresserons plus particulièrement à la méthylation dans le contexte des dinucléotides Cytosine-Guanine chez les mammifères, pour lesquels le groupement méthyle est déposé sur le carbone 5 de la base cytosine. En effet, même si les cytosines méthylées sont également présentes hors dinucléotide CpG, celles-ci ne concernent qu'une proportion mineure du méthylome chez les mammifères (Lister *et al.*, 2009). De plus, l'existence de la méthylation au niveau des CpG peut être très variable entre certains organismes, démontrant notamment un lien avec l'évolution (Figure 7). Certains organismes comme *C. elegans* et *D. melanogaster* ou les levures *S. cerevisiae* et *S. pombe* ne présentent aucune méthylation des CpG (Raddatz *et al.*, 2013), signifiant que cette modification n'est pas essentielle pour la mise en place et le maintien des différents processus cellulaires chez ces organismes. Chez les insectes, les champignons et les plantes, le génome est majoritairement pauvre en CpG méthylés, excepté sur certaines régions telles que les transposons et/ou les corps de gènes (Feng *et al.*, 2010 ; Zemach *et al.*, 2010). A l'inverse, chez les mammifères, le génome est enrichi en CpG méthylés (plus de 80 % des CpG chez

l'Homme sont méthylés) ; les régions non méthylées correspondent majoritairement à des régions régulatrices actives (Lister *et al.*, 2009). La transition entre méthylation mosaïque chez les invertébrés et méthylation globale chez les vertébrés semble s'être effectuée de façon progressive avec l'évolution des organismes (Okamura *et al.*, 2010). Cependant, en favorisant le phénomène de déamination spontanée des cytosines méthylées, la méthylation a entraîné un taux important de mutations non réparées, appauvrissant ainsi le génome des vertébrés en dinucléotides CpG à l'échelle de l'évolution (Cooper et Krawczak, 1989).

1 – Mise en place et maintien de la méthylation des cytosines

La formation du groupement méthyle sur les cytosines à partir du donneur de méthyle SAM (« S-Adenosyl Methionine ») implique l'intervention d'ADN méthyltransférases (DNMT, pour « DNA methyltransferases »). A l'heure actuelle, quatre d'entre elles, responsables de la mise en place et/ou du maintien de la méthylation au niveau de l'ADN, ont été décrites dans la littérature. Plusieurs scénarios existent concernant l'adressage des DNMTs au niveau du génome révélant un lien entre modification de l'ADN, modifications des histones, FT et nature des nucléosomes.

a – Les DNMTs

Chez les mammifères, la méthylation *de novo* est assurée par les enzymes DNMT3A et DNMT3B capables de transférer un groupement méthyle sur des sondes non-méthylées et hémi-méthylées avec une activité équivalente (Okano *et al.*, 1998, 1999). Ces enzymes sont très fortement exprimées dans les cellules ES et leur expression est réduite après différenciation, suggérant une fonction redondante de ces protéines lors des étapes précoces de développement (Okano *et al.*, 1998, 1999). Cependant, d'autres études montrent que DNMT3A et DNMT3B ont deux profils d'expressions distincts gouvernant leurs fonctions respectives : DNMT3B est plus fréquemment exprimée dans les cellules totipotentes tandis que l'expression de DNMT3A est principalement ubiquitaire lors d'étapes plus tardives de l'embryogenèse ainsi que dans les cellules différenciées (Watanabe *et al.*, 2002 ; Xie *et al.*, 1999).

Dans les cellules germinales, une autre enzyme appartenant à la famille des DNMTs mais qui ne possède pas de domaine catalytique, DNMT3L, est nécessaire à l'établissement de l'empreinte parentale (Bourc'his *et al.*, 2011). De façon intéressante, *via* une interaction

directe, DNMT3L guide DNMT3A et DNMT3B au niveau de l'ADN et stimule leur activité méthyltransférase (Ooi *et al.*, 2007 ; Suetake *et al.*, 2004).

La nature palindromique des dinucléotides CpG permet de transmettre l'information épigénétique au brin nouvellement synthétisé lors de la réplication grâce à DNMT1. Considérée comme l'enzyme de maintenance de la méthylation, DNMT1 est une enzyme processive qui, par son affinité pour l'ADN hémiméthylé, se fixe sur le brin nouvellement synthétisé et méthyle les cytosines selon le modèle parental (Hermann *et al.*, 2004 ; Leonhardt *et al.*, 1992).

Initialement décrites comme ayant des fonctions singulières, plusieurs arguments suggèrent finalement que les enzymes DNMT3 et DNMT1 sont requises à la fois pour l'établissement et pour le maintien de la méthylation des cytosines dans le génome des mammifères (Jones et Liang, 2009 ; Li *et al.*, 2015). En effet, la méthylation des régions répétées nécessite l'action de DNMT3A/B ou de DNMT1, selon le type de régions, révélant une division du travail par ces enzymes (Li *et al.*, 2015). De plus, DNMT3A/B participent au maintien de la méthylation souvent incomplet de DNMT1 (Li *et al.*, 2015).

b – Adressage des DNMTs

L'adressage adéquat des DNMTs semble essentiel afin de réguler la mise en place de la méthylation au niveau du génome. Le recrutement de DNMT1 au niveau des sites de réplication est dépendant de la protéine UHRF1 (« ubiquitin-like, containing PHD and RING finger domain 1 ») qui se lie dans un premier temps à DNMT1 pour ensuite, grâce à un domaine de reconnaissance des CpG méthylés (le domaine SRA, « SET- and RING-associated DNA-binding domain »), guider le recrutement de l'enzyme sur les régions hémiméthylées (Bostick *et al.*, 2007 ; Sharif *et al.*, 2007).

Concernant le ciblage de la méthylation *de novo*, les enzymes DNMT3 sont capables de se fixer sur l'ADN grâce à un domaine N-terminal PWWP indispensable à leur activité (Ge *et al.*, 2004). Malgré quelques éléments (Wienholz *et al.*, 2010), le peu d'informations concernant un ciblage direct des DNMTs par la séquence ADN ainsi que la diversité de ces sites de ciblage suggèrent l'implication d'un facteur intermédiaire. Celui-ci peut correspondre à un FT dont le recrutement dépend de la séquence d'ADN et qui va à son tour, par interaction directe et même indirecte (impliquant un deuxième facteur intermédiaire) assurer

l'adressage des DNMTs (Brenner *et al.*, 2005 ; Hervouet *et al.*, 2009 ; Velasco *et al.*, 2010). L'élément intermédiaire peut également correspondre à une modification de la queue des histones. En effet, DNMT3L et DNMT3A sont capables de reconnaître les queues d'histones H3 qui ne sont pas méthylées sur la lysine 4 (Ooi *et al.*, 2007 ; Otani *et al.*, 2009 ; Zhang *et al.*, 2010a), ce qui concorde avec la corrélation négative observée entre la liaison des DNMT3 et la méthylation d'H3K4 (Ooi *et al.*, 2007 ; Zhang *et al.*, 2010a). D'autres études montrent que les histones méthyltransférases d'H3K9 guident la méthylation de l'ADN au niveau de certains *loci* dans les cellules ES ou lors de l'embryogénèse chez la Souris (Auclair *et al.*, 2015 ; Lehnertz *et al.*, 2003) et que l'une d'entre elles, SUV39H1 interagit avec DNMT3A et DNMT1 (Fuks *et al.*, 2003). De plus, DNMT3A est également capable de reconnaître une autre modification d'histone associée à la répression : H3K36me3 (Dhayalan *et al.*, 2010).

Cependant, comme souvent suggéré dans la littérature, si on considère que la méthylation dite globale du génome est une méthylation « par défaut », les DNMTs ne seraient pas ciblées spécifiquement sur certaines régions du génome mais se fixeraient sur toutes les régions qui leur seraient accessibles. Dans ce sens, les FT semblent protéger les séquences qu'ils fixent de la méthylation *de novo* car leur absence entraîne une hyperméthylation de ces régions (Brandeis *et al.*, 1994 ; Macleod *et al.*, 1994).

2 – Méthylation des CpG et répression transcriptionnelle

La méthylation de l'ADN est communément associée à la répression transcriptionnelle et de nombreux exemples attestent de cette fonction. Celle-ci peut être exercée de manière directe ou indirecte par la présence de la 5mC. Cependant, certaines études suggèrent que la méthylation ne serait pas responsable de l'initiation de la répression génique mais plutôt de sa maintenance.

a – Exemples de répressions exercées par la méthylation des CpG

Le rôle répresseur de la méthylation des cytosines a été particulièrement décrit au niveau de régions intergéniques et de certains promoteurs situés dans des régions denses en CpG appelées îlots CpG.

La méthylation intergénique réprime l'expression des éléments transposables ou rétroviraux (Walsh *et al.*, 1998). DNMT1 est responsable de la méthylation au niveau des

séquences LTR (« Long Terminal Repeat ») tandis que les DNMT3A/B sont impliquées dans la répression des séquences LINE (« Long Interspersed Nuclear Elements » - Li *et al.*, 2015). La répression des éléments mobiles semble être la fonction primaire de la méthylation et est même considérée par certains chercheurs comme la seule fonction de cette marque épigénétique. Néanmoins, celle-ci permet également d'assurer d'autres processus de répression à long terme tels que l'inactivation du second chromosome X (Mohandas *et al.*, 1981) et le phénomène d'empreinte génomique (Li *et al.*, 1993).

Les îlots CpG correspondraient à environ 50 à 70 % des promoteurs annotés chez l'Homme (Antequera et Bird, 1993), mais seulement 3 à 21 % des promoteurs compris dans ces îlots CpG, selon leur densité décroissante en CpG, seraient méthylés (Weber *et al.*, 2007). Un grand nombre d'études a participé à démontrer l'implication de la méthylation des îlots CpG dans la répression génique de manière dépendante du type cellulaire. Dans les cellules ES, la méthylation *de novo* des îlots CpG intervient notamment au niveau des promoteurs de gènes codant pour des facteurs développementaux (Fouse *et al.*, 2008). Lors du développement, de l'engagement vers un lignage ou dans les cellules somatiques, elle concerne les promoteurs de gènes des facteurs de pluripotence, de gènes spécifiques des cellules germinales et même de certains gènes spécifiques d'un lignage (Borgel *et al.*, 2010 ; Futscher *et al.*, 2002 ; Mohn *et al.*, 2008 ; Shen *et al.*, 2007 ; Weber *et al.*, 2007). Ainsi, la méthylation au niveau des promoteurs des îlots CpG semble participer à la régulation spatio-temporelle des gènes afin d'en empêcher l'expression anarchique.

b – Mécanismes de répression

Deux mécanismes principaux de répression des gènes associés à la méthylation des CpG ont été proposés. Le premier implique une répression directe par laquelle la présence de 5mC inhibe la fixation de protéines activatrices (Iguchi-Arigo et Schaffner, 1989 ; Watt et Molloy, 1988). Le second mécanisme, quant à lui, propose une répression indirecte *via* le recrutement de protéines possédant un domaine de liaison aux CpG méthylés (MBD pour « methyl-CpG binding domain » – voir partie C–Protéines de reconnaissance des dinucléotides CG) qui à leur tour vont recruter des corépresseurs capables de moduler la structure chromatinienne afin de réduire l'accessibilité à l'ADN (Boyes et Bird, 1991).

Parmi ces protéines, MBD1 participe à la répression de manière dépendante des HDAC mais intervient également dans la formation d'hétérochromatine lors de la réplication

en recrutant la méthyltransférase d'histone SETDB1 afin de méthyliser H3K9 (Fujita *et al.*, 1999 ; Sarraf et Stancheva, 2004).

MBD2 interagit avec des HDAC avec le répresseur Sin3A et dirige le complexe répresseur NuRD au niveau des régions méthylées (Boeke *et al.*, 2000 ; Ng *et al.*, 1999 ; Zhang *et al.*, 1999).

Deux études réalisées chez les mammifères et le Xénope ont montré que MeCP2, certainement la plus étudiée des protéines contenant un MBD, est associée à un complexe corépresseur composé de Sin3A et de HDAC (Nan *et al.*, 1998 ; Jones *et al.*, 1998). Cependant, du fait de l'instabilité d'interaction entre MeCP2 et Sin3A, une étude supplémentaire suggère qu'il ne s'agirait pas du mécanisme principal de répression utilisé par MeCP2, et que d'autres facteurs pourraient ainsi être impliqués (Klose et Bird, 2004). Plus récemment, deux études ont en effet observé l'interaction entre MeCP2 et le complexe NCoR/SMRT (« nuclear receptor co-repressor »/ « silencing mediator for retinoid or thyroid-hormone receptors ») proposant un nouveau mécanisme de répression *via* MeCP2 (Ebert *et al.*, 2013 ; Lyst *et al.*, 2013). Le complexe NCoR avait d'ailleurs déjà été associé à la répression en raison de son interaction avec une autre protéine de liaison aux CpG méthylés, Kaiso (Yoon *et al.*, 2003).

La densité en CpG et la « force » du promoteur sont des paramètres à prendre en compte concernant la répression transcriptionnelle. En effet, une faible densité en CpG méthylés ne participe à la répression qu'à court terme de promoteurs dits faibles, contrairement à une forte densité qui permet la formation et la diffusion d'une structure chromatinienne répressive même sur des promoteurs dits forts (Boyes et Bird, 1992 ; Curradi *et al.*, 2002). Malgré le fait que la répression transcriptionnelle associée à la méthylation des CpG soit connue depuis longtemps, des recherches sont encore nécessaires pour préciser les mécanismes impliqués dans cette régulation.

c – Initiation ou maintenance de la méthylation ?

En lien avec le paragraphe précédent, certaines études ont observé que les DNMT elles-mêmes sont capables d'interagir avec des protéines de modulation de la structure chromatinienne. En effet, Fuks et collègues ont montré que DNMT1 possède un domaine répresseur capable de lier HDAC1 et que les activités méthyltransférases et désacétylases de

ces protéines sont associées (Fuks *et al.*, 2000). Cependant, les données obtenues ne permettent pas de savoir laquelle de ces deux enzymes est responsable du ciblage au niveau du génome. Les auteurs suggèrent donc deux scénarios : le premier, communément admis et discuté plus haut, propose que les DNMT sont responsables du recrutement des HDACs ; tandis que le second propose que la méthylation serait consécutive aux modifications des histones (Fuks *et al.*, 2000). En support de cette seconde hypothèse, DNMT3B interagit avec des HDACs, une méthylase d'histone et une enzyme de remodelage de la chromatine (Geiman *et al.*, 2004). De façon intéressante, dans des lignées cellulaires mimant le syndrome ICF (« Immunodeficiency, centromeric region instability, facial anomalies ») pour lequel DNMT3B est muté principalement au niveau de son domaine catalytique, les activités désacétylase et méthylase d'histone, et ATPase sont maintenues (Geiman *et al.*, 2004). Cependant, ces résultats sont à modérer car dans les lignées ICF utilisées pour cette étude, l'activité méthyltransférase d'ADN n'est pas entièrement abolie due à une interaction entre DNMT3B et DNMT1.

Néanmoins, d'autres études ont directement montré un rôle de la méthylation des CpG dans le maintien de la répression plutôt que dans son initiation. L'une d'elles a étudié la dynamique de structure chromatinienne liée à la répression du gène *Oct3/4* lors de la différenciation des cellules ES ou de cellules de carcinome embryonnaire (ECC) de Souris (Feldman *et al.*, 2006). Ces travaux montrent que la méthylation *de novo* du promoteur intervient après, et est dépendante, de la méthylation d'H3K9 responsable de « l'hétérochromatinisation » de cette région. Malgré le fait qu'elles conservent la répression d'*Oct3/4*, les cellules DNMT3A/3B^{-/-} sont capables de lever cette répression au cours de la différenciation (Feldman *et al.*, 2006). De plus, dans des cellules murines, seule la répression et non la réactivation d'un transgène, est indépendante de la présence de CpG (Feng *et al.*, 2006). Ces différents résultats montrent que dans certains cas, la méthylation n'est pas à l'origine de l'initiation de la répression mais agit plutôt en tant que verrou à long terme de cet état répressif.

3 – La déméthylation

Longtemps considérée comme statique, la méthylation de l'ADN s'avère en fait résulter de processus très dynamiques. Le nombre croissant d'études démontrant la présence d'une telle dynamique au niveau des promoteurs et des régions régulatrices distantes révèle notamment l'importance du mécanisme de déméthylation active lors du développement et de la différenciation chez les mammifères. Par opposition à la déméthylation passive, qui résulte d'un défaut de maintien de la méthylation lors de la réplication, la déméthylation active est indépendante de la réplication. Il s'agit d'un processus complexe comportant plusieurs voies et faisant intervenir de nombreux facteurs.

a – La méthylation : une modification dynamique

L'aspect dynamique de la méthylation a dans un premier temps été constaté au cours des premières étapes du développement. Le génome des mammifères y subit une reprogrammation épigénétique qui comprend notamment deux vagues de déméthylation (Saitou *et al.*, 2012). La première apparaît après fertilisation et concerne chacun des génomes parentaux via une déméthylation passive. La deuxième vague de déméthylation est quant à elle active et se produit dans les cellules germinales primordiales (PGCs).

A une échelle plus locale, dans des cellules cancéreuses mammaires, quelques promoteurs testés montrent un recrutement cyclique d'ER α corrélé à une activité transcriptionnelle cyclique (Kangaspeska *et al.*, 2008 ; Métivier *et al.*, 2008). De façon intéressante, cette cyclicité est associée à une dynamique de méthylation/déméthylation du promoteur spécifiant le caractère dynamique de cette modification (Métivier *et al.*, 2008 ; Kangaspeska *et al.*, 2008). D'autres études ont également observé une déméthylation de certains promoteurs en lien avec l'activation génique lors de la différenciation neurale, même s'il s'agit là de rares exemples (Mohn *et al.*, 2008). La plupart des événements de déméthylation active observés au cours de la différenciation semblent néanmoins survenir dans des régions pauvres en CpG et correspondant à des régions régulatrices distantes telles que les enhancers (Meissner *et al.*, 2008 ; Schimdl *et al.*, 2009 ; Sérandour *et al.*, 2011).

De plus, en complément du dernier paragraphe de la partie II-1-b-Adressage des DNMTs, si la méthylation globale du génome est un paramètre « par défaut » et que les DNMTs ne méthylent que les cytosines qui leur sont accessibles, il est envisageable qu'au

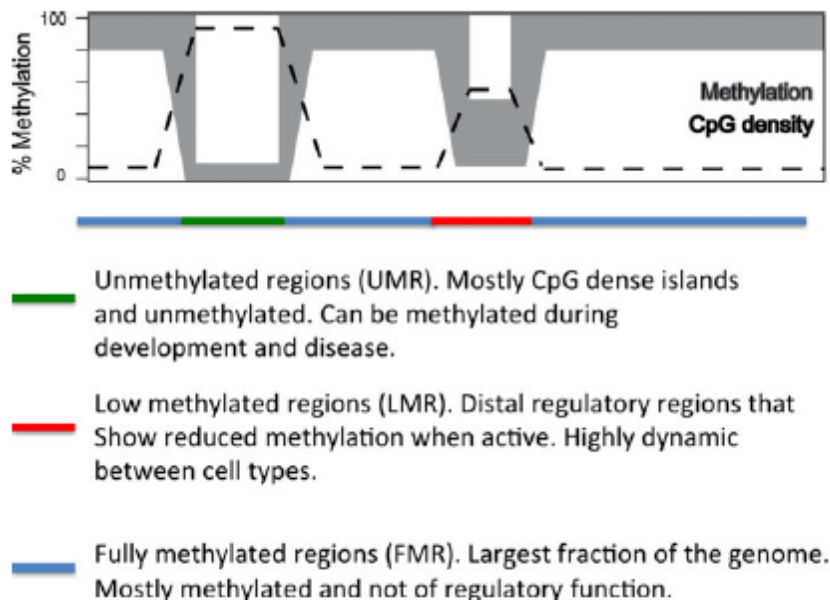


Figure 8. Représentation graphique de la distribution de la méthylation de l'ADN dans le génome des mammifères (Schübeler, 2015).

Les îlots CpG n'ont pas ou peu de méthylation et contiennent de nombreux CpG. Le reste du génome tend à être majoritairement méthylé; à l'exception des régions régulatrices actives qui présentent des niveaux moindres de méthylation et qui sont fonction de leur occupation par des facteurs de transcription. Cette variabilité de méthylation est illustrée dans ce schéma par l'épaisseur du graphe en gris représentant le pourcentage de méthylation. La densité en CpG est elle illustrée par la ligne en pointillés.

lieu d'être protégées de la méthylation, les régions détectées comme non méthylées soient des régions soumises à déméthylation. Dans ce sens, des études réalisées à partir de cellules humaines ou murines ont révélé que les CpG hypométhylés sont présents dans les îlots CpG ou au niveau de régions régulatrices actives (Lister *et al.*, 2009; Stadler *et al.*, 2011; Ziller *et al.*, 2013 – Figure 8). Ainsi, le faible niveau de méthylation observé sur les régions régulatrices actives corrèle avec les observations précédentes qui suggèrent l'implication d'un processus de déméthylation active de ces régions lors de l'engagement vers un programme transcriptionnel particulier (Kangaspeska *et al.*, 2008 ; Meissner *et al.*, 2008 ; Métivier *et al.*, 2008 ; Mohn *et al.*, 2008 ; Schimdl *et al.*, 2009 ; Sérandour *et al.*, 2011). Dans un tel scénario, les îlots CpG constitutivement non méthylés seraient pour leur part la cible d'une déméthylation constante.

b – Mécanismes de déméthylation active

A ce jour, les diverses propositions concernant les mécanismes de déméthylation active impliquent la succession de différentes réactions enzymatiques telles que la déamination et/ou l'oxydation de la 5mC suivie d'une étape de réparation par excision de base (BER – « Base Excision Repair »).

Plus précisément, une des voies de déméthylation active met en œuvre l'action d'une déaminase de cytosines appartenant à la famille des protéines AID/APOBEC (« apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like »). La thymine, résultante de la déamination de la 5mC, faisant apparaître un mésappariement T:G, est ensuite reconnue et éliminée par des enzymes de réparation appartenant à la voie de réparation BER telles que MBD4 (Hendrich *et al.*, 1999) ou TDG (« Thymidine DNA Glycosylase » – Cortellino *et al.*, 2011 – Figure 9). Ce mécanisme a dans un premier temps été observé chez le poisson zèbre, puis confirmé chez les mammifères dans les PGCs maternelles et lors de la reprogrammation vers la pluripotence (voir Fritz et Papavasiliou, 2010).

Une étude réalisée au sein du laboratoire a montré que la déméthylation cyclique du promoteur du gène *pS2* dans des cellules cancéreuses mammaires est associé au recrutement de DNMT3A/b, TDG et de protéines impliquées dans le mécanisme de BER (Métivier *et al.*, 2008). De plus, une diminution de l'expression de TDG entraîne une augmentation des niveaux de méthylation et réduit l'activité transcriptionnelle de *pS2*. Les auteurs proposent que la déméthylation observée résulte de la présence d'un processus de déamination au niveau

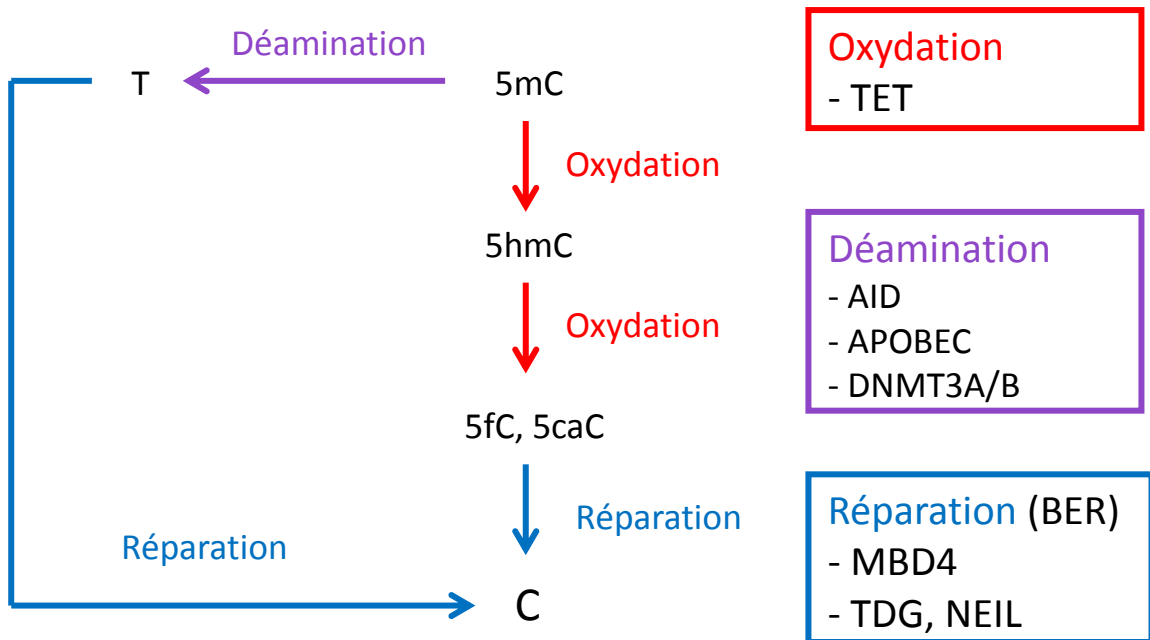


Figure 9. Mécanismes de déméthylation active des cytosines.

La déméthylation active des cytosines comporte deux voies principales: la voie de déamination et la voie d'oxydation. Dans la première voie de déméthylation, la 5-méthylcytosine (5mC) est déaminée par des enzymes telles que AID/APOBEC ou les DNMT. Les bases déaminées sont ensuite réparées par le mécanisme de réparation par excision de base (BER) impliquant des glycosylases spécifiques. Dans la deuxième voie de déméthylation active des cytosines, la 5mC peut être oxydée par les enzymes de la famille TET afin de former la 5-hydroxyméthylcytosine (5-hmC) qui peut, elle-même à son tour, subir une ou deux étapes d'oxydation pour être ensuite réparée par le système BER.

5mC : 5-méthylcytosine, 5hmC: 5-hydroxyméthylcytosine, 5fC: 5-formylcytosine, 5-caC: 5-carboxylcytosine, TET: « Ten eleven translocation », AID: « Activation induced deaminase », APOBEC: « apolipoprotein B mRNA editing enzyme catalytic subunit », DNMT3A/B: « DNA methyltransferases 3A/B », MBD4: « Merthyl-CpG binding domain protein 4 », TDG: « Thymine DNA glycosylase ».

de ce promoteur pouvant être réalisé par les protéines DNMT3A/b, comme observé *in vitro* en présence de faibles quantités de donneur de méthyle SAM (Métivier *et al.*, 2008 – Figure 9).

La deuxième voie classique de déméthylation implique plusieurs étapes d'oxydation à partir des cytosines méthylées, puis une excision des bases résultantes par des glycosylases, suivie d'une étape de réparation (Mahé et Salbert, 2015 – Figure 9). L'identification de ce processus qui a révolutionné le monde de l'épigénétique a été permise grâce à la (re-) découverte d'une forme oxydée de la 5mC nommée 5-hydroxyméthylcytosine (5hmC), générée par l'action enzymatique de TET1 (Ten-eleven translocation 1), et identifiée par homologie avec des protéines oxydant le groupement méthyle des thymines chez le trypanosome (Iyer *et al.*, 2009 ; Kriaucionis et Heinz, 2009 ; Tahiliani *et al.*, 2009). Des études supplémentaires ont ensuite montré que deux autres protéines de la famille TET, TET2 et TET3, possèdent également une activité oxydase (Ito *et al.*, 2010). L'oxydation itérative de la 5hmC par les protéines TET engendre la formation de deux dérivés additionnels : la 5-formylcytosine (5fC) et la 5-carboxylcytosine (5caC – He *et al.*, 2011 ; Ito *et al.*, 2011). La formation de ces produits oxydés est une étape essentielle dans le mécanisme de déméthylation active car les glycosylases TDG ou NEIL sont capables de reconnaître et d'exciser les bases 5fC et 5caC, ultérieurement remplacées par une cytosine (He *et al.*, 2011 ; Maiti et Drohat, 2011 ; Müller *et al.*, 2014). Des études *in vitro* proposent également que la voie de déméthylation par oxydation pourrait impliquer l'action d'une décarboxylase (Xu *et al.*, 2013) ou des enzymes DNMT3A/B en tant que déhydroxyméthylases (Chen *et al.*, 2012). Des études *in vivo* sont cependant nécessaires afin de pouvoir valider ces deux hypothèses.

La découverte et la caractérisation de différentes voies de déméthylation active ont permis une grande avancée dans l'identification et la compréhension de la dynamique de régulation du génome, même si de nombreuses questions restent posées. Malgré quelques exemples montrant l'implication de la déamination dans le processus de déméthylation active des cytosines, la voie oxydative concentre une grande partie des recherches actuelles, afin d'en déterminer les conséquences sur la régulation transcriptionnelle.

B – Dérivés oxydés

Les travaux amorcés en 2009 par Kriaucionis et Heinz ainsi que par le laboratoire d'Anjana Rao ont permis l'identification de trois nouvelles formes de cytosines modifiées : 5hmC, 5fC et 5caC (He *et al.*, 2011 ; Ito *et al.*, 2011 ; Kriaucionis et Heinz, 2009 ; Tahiliani *et al.*, 2009). Ces résidus sont le sujet d'intenses recherches et de nombreuses techniques ont été mises au point afin de pouvoir les détecter et étudier leur distribution (Mahé et Salbert, 2015). Une attention particulière est portée sur la 5hmC, car plus qu'un simple intermédiaire dans la voie de déméthylation active des cytosines, plusieurs éléments laissent à penser que cette « sixième base » de l'ADN pourrait avoir une fonction propre dans la régulation génique.

Ce sujet a fait l'objet de l'écriture d'une revue : « Mahé EA, Salbert G. DNA demethylation by TET proteins: a potential therapeutic target in cancer. *Epigenetic Diagnosis & Therapy*, 2015, 1, 49-59 », disponible en annexe de ce mémoire de thèse.

1 – Mise en place des dérivés oxydés

a – Les protéines TET

Chez les mammifères, les enzymes TET sont des protéines conservées au cours de l'évolution qui appartiennent à la famille des dioxygénases dépendantes du Fe^{2+} et du 2-OG (2-oxoglutarate). Avant la découverte de son implication dans le mécanisme de déméthylation, TET1 était connue pour sa fusion avec MLL1 (« Mixed lineage leukemia 1 ») dans la leucémie myéloïde aigue. Les protéines TET exercent leur activité enzymatique grâce à leur domaine catalytique mais seules TET1 et TET3 contiennent un domaine de liaison à l'ADN de type CXXC (Tahiliani *et al.*, 2009 ; Zhang *et al.*, 2010b). *Tet1* et *Tet2* sont fortement exprimés dans les cellules ES murines puis rapidement réprimés au cours de la différenciation (Szwagierczak *et al.*, 2010). L'expression de *Tet2* est ensuite restaurée tandis que *Tet3*, qui n'était pas exprimé à l'état indifférencié, est induit. Ainsi, les tissus différenciés expriment majoritairement *Tet2* et *Tet3* (Szwagierczak *et al.*, 2010).

La cartographie des sites de liaison de TET1 a été établie dans les cellules ES murines grâce à la technique de ChIP-seq (pour « chromatin immunoprecipitation », suivie par du séquençage à haut débit). Ces études ont montré que les régions recrutant TET1 sont

généralement situées dans les corps de gènes et au niveau de promoteurs de gènes importants pour la pluripotence et le développement des cellules ES ; ces régions sont hypométhylées et ont une forte densité en CpG (*i.e.* sont pour la plupart, des îlots CpG – Williams *et al.*, 2011 ; Wu *et al.*, 2011a ; Xu *et al.*, 2011). De même, TET2 et TET3 sont également enrichis au niveau des îlots CpG et des TSS dans une lignée rénale embryonnaire humaine (HEK293T – « Human Embryonic Kidney » – Deplus *et al.*, 2013). Cette localisation particulière suggère un rôle de ces protéines dans la régulation génique.

b – Adressage et substrats

La reconnaissance des protéines TET au niveau de l'ADN s'effectue grâce à la présence d'un domaine en doigt de zinc CXXC. Ce domaine est très conservé et, en règle générale, induit une forte spécificité de liaison pour les CpG non méthylés (Long *et al.*, 2013). Cependant, chacune des enzymes TET possède des propriétés uniques de reconnaissance de l'ADN conférées par ce domaine CXXC. En plus de la reconnaissance des CpG non méthylés, TET1 et TET3 sont capables de se lier aux cytosines méthylées (Xu *et al.*, 2011, 2012). En outre, le domaine CXXC de TET1 reconnaît les CpG hydroxyméthylés tandis que celui de TET3 a la capacité de se lier à des cytosines hors CpG (Xu *et al.*, 2011, 2012). L'absence de domaine CXXC chez TET2 provient d'une inversion chromosomique au cours de l'évolution séparant le domaine de liaison à l'ADN, devenu un nouveau gène codant pour IDAX (« inhibition of the Dv1 and axin complex » aussi connu sous le nom de CXXC4), du reste de la protéine (Iyer *et al.*, 2009). De façon intéressante, IDAX entraîne le recrutement de TET2 au niveau des promoteurs et des îlots CpG (Ko *et al.*, 2013). Pour TET2 comme pour les autres protéines TET, il est probable que la spécificité de liaison sur certaines régions génomiques ne soit pas seulement due au domaine CXXC mais également à une interaction avec d'autres facteurs permettant un ciblage particulier. Dans ce sens, NANOG recrute TET1 et TET2 au niveau de gènes impliqués dans le maintien de la pluripotence et l'engagement vers un lignage dans les cellules ES (Costa *et al.*, 2013). Les protéines TET interagissent avec de nombreuses protéines mais pour l'instant peu d'études décrivent la séquence de leur recrutement.

Plusieurs études ont révélé que les enzymes TET ont une préférence de substrat pour les cytosines méthylées comparées aux cytosines hydroxyméthylées ou formylées (Ito *et al.*, 2011 ; Hu *et al.*, 2015). Cependant, des analyses *in vitro* et informatiques centrées sur TET2

ont montré que la préférence de substrat n'est pas due à une spécificité de reconnaissance par le domaine catalytique, car TET2 interagit aussi bien avec la 5mC qu'avec les 5hmC ou 5fC. La préférence de substrat des TET serait donc plutôt une conséquence de l'efficacité et de la possibilité de réaliser la réaction catalytique en elle-même (Hu *et al.*, 2015 ; Lu *et al.*, 2016).

2 – La 5-hydroxyméthylcytosine

La 5hmC correspond au premier produit d'oxydation de la 5mC par les enzymes TET (Tahiliani *et al.*, 2009). Cette modification est présente dans de nombreux tissus et types cellulaires chez les mammifères (Globisch *et al.*, 2010). Elle représente entre 0,03% et 0,7% des cytosines, les niveaux les plus faibles étant observés dans la rate et les glandes endocrines tandis que les tissus neuraux contiennent les niveaux les plus élevés de 5hmC (Globisch *et al.*, 2010). De plus, le niveau de 5hmC est généralement anti-corrélé à la vitesse de prolifération cellulaire (Bachman *et al.*, 2014 ; Gackowski *et al.*, 2015). La distribution de la 5hmC a été étudiée dans de nombreuses lignées cellulaires telles que les cellules ES murines et humaines (Booth *et al.*, 2012 ; Ficz *et al.*, 2011 ; Pastor *et al.*, 2011 ; Stroud *et al.*, 2011 ; Szulwach *et al.*, 2011 ; Williams *et al.*, 2011 ; Wu *et al.*, 2011b ; Xu *et al.*, 2011 ; Yu *et al.*, 2012), les cellules de carcinome embryonnaire (Putiri *et al.*, 2014 ; Sérandour *et al.*, 2012), les préadipocytes et adipocytes (Sérandour *et al.*, 2012) ou dans les cellules du cervelet (Song *et al.*, 2011) et les lymphocytes B et plasmocytes humains (Caron *et al.*, 2015). La plupart de ces études s'accordent sur l'existence d'un enrichissement de la 5hmC au niveau du corps des gènes, notamment des exons, et des enhancers. Néanmoins, sa répartition au niveau des promoteurs et notamment des TSS est plus discutée. Dans certains cas, un enrichissement en 5hmC est observé au niveau des TSS (Booth *et al.*, 2012 ; Pastor *et al.*, 2011 ; Szulwach *et al.*, 2011 ; Williams *et al.*, 2011), tandis que dans d'autres cas, la 5hmC est déplétée des TSS et enrichie de manière bimodale autour du TSS (Song *et al.*, 2011 ; Szulwach *et al.*, 2011 ; Wu *et al.*, 2011b). D'une façon générale, la présence de 5hmC est anti-corrélée avec la densité en CpG (Booth *et al.*, 2012 ; Szulwach *et al.*, 2011 ; Wu *et al.*, 2011b ; Xu *et al.*, 2011), mais au niveau des TSS, son enrichissement est associée à un niveau intermédiaire ou élevé de densité en CpG (Pastor *et al.*, 2011 ; Williams *et al.*, 2011). En accord avec une distribution au niveau des régions régulatrices, la 5hmC est également associée aux sites de fixation de CTCF et de protéines impliquées dans le maintien de la pluripotence (Ficz *et al.*, 2011 ; Stroud *et al.*, 2011 ; Wu *et al.*, 2011b ; Yu *et al.*, 2012). Par conséquent, les différentes

cartographies suggèrent l'implication de la « sixième base » dans la régulation transcriptionnelle. Dans ce sens, ces études montrent que la 5hmC est présente au niveau de promoteurs de gènes transcriptionnellement inactifs mais que sa présence au niveau des corps de gènes est positivement corrélée à leur transcription. De plus, la 5hmC est une modification stable (Bachman *et al.*, 2014) spécifiquement reconnue par certaines protéines intervenant dans la régulation transcriptionnelle (Iurlaro *et al.*, 2013 ; Spruijt *et al.*, 2013 – Voir II-D– Protéines de reconnaissance des dinucléotides CG). Malgré le nombre important d'études réalisées sur la 5hmC, sa distribution génomique ne suffit pas à la compréhension de son rôle et implique la nécessité d'effectuer des études fonctionnelles.

3 – La 5-formylcytosine et la 5-carboxylcytosine

Les 5fC et 5caC sont les cytosines modifiées les moins abondantes, représentant moins de 0,002% de l'ensemble des cytosines, avec environ 10 fois moins de 5caC que de 5fC (Ito *et al.*, 2011). Elles sont détectées dans les cellules ES (Booth *et al.*, 2014 ; Ito *et al.*, 2011 ; Shen *et al.*, 2013 ; Song *et al.*, 2013). La 5fC a également été détectée dans la plupart des organes murins (Ito *et al.*, 2011). L'utilisation d'une méthode de détection par chromatographie liquide et spectrométrie de masse a également permis d'observer la 5caC au sein de tissus embryonnaires et adultes murins (Bachman *et al.*, 2015). Les niveaux faibles de ces dérivés oxydés nécessitent généralement la déplétion par « knockdown » (shRNA – « short hairpin RNA) ou « knockout » de l'enzyme TDG pour qu'ils puissent être détectés. De cette manière, les 5fC et 5caC ont été détectées au niveau des régions répétées et de plusieurs régions régulatrices proximales ou distales (Shen *et al.*, 2013 ; Song *et al.*, 2013). Notamment, la 5fC est particulièrement enrichie au niveau des enhancers « poised » et sa présence est associée à celle de P300 sur ces enhancers, suggérant un rôle de la 5fC et/ou de la déméthylation active dans la préparation des enhancers (Song *et al.*, 2013).

L'évolution des techniques de détection des dérivés de cytosines à la base près a permis à deux autres études de compléter l'analyse de leurs distributions au niveau du génome (Neri *et al.*, 2015 ; Raiber *et al.*, 2012). Ces travaux ont confirmé les précédentes cartographies des 5fC et 5caC mais ont de plus mis en évidence un enrichissement important au niveau de promoteurs actifs, hypométhylés, régulant des gènes à haut niveau d'expression

et dont le TSS porte la marque H3K4me3. Ces observations révèlent ainsi que la distribution des 5fC et 5caC ne concorde pas nécessairement avec celle de la 5hmC ou de la 5mC, comme cela a également été constaté dans d'autres études (Bachman *et al.*, 2015 ; Raiber *et al.*, 2012 ; Song *et al.*, 2013). De façon intéressante, l'utilisation d'un marquage par un isotope stable *in vivo* a révélé que la 5fC peut être une modification stable (Bachman *et al.*, 2015). Dans ce sens, la comparaison des niveaux de 5fC entre différents tissus montre qu'ils ne corrélaient pas avec ceux de ses précurseurs, 5mC et 5hmC (Bachman *et al.*, 2015).

Plusieurs études ont cherché à comprendre les conséquences de la présence de ces dérivés provenant de l'étape de super-oxydation. Lorsque les 5fC et 5caC ne sont pas excisées par la TDG, leur présence au moment de la réplication de l'ADN perturberait les processus associés et induirait des phénomènes de mutagenèse et d'instabilité génomique (Kamiya *et al.*, 2002; Shibutani *et al.*, 2014). Ces observations sont en opposition avec celles de Münzel et collègues (Münzel *et al.*, 2011). D'autres travaux ont cependant également montré que la présence de 5fC et de 5caC, mais pas celle de 5mC ni de 5hmC, réduit le taux de polymérisation par l'ARN Pol II lors de la transcription *in vitro* (Kellinger *et al.*, 2012). La détection de ces dérivés par l'ARN Pol II, induisant le phénomène de pause, suggère que les 5fC et 5caC agissent comme un niveau supplémentaire de régulation de la dynamique de transcription (Kellinger *et al.*, 2012).

C – Implications des protéines TET et TDG dans la régulation transcriptionnelle et la structure chromatinienne

Les protéines TET et TDG sont deux enzymes essentielles à la voie de déméthylation active des cytosines et leur présence ainsi que leur activité enzymatique sur des régions génomiques spécifiques doit pouvoir être finement régulée et modulée. Comme déjà abordé au cours cette introduction, les mécanismes de régulation nécessitent généralement la présence de co-régulateurs permettant d'affiner l'information transmise. Ainsi, de nombreuses études ont visé à identifier les facteurs interagissant avec les protéines TET et TDG, notamment les complexes de modification des histones et de remodelage de la chromatine, afin de caractériser leurs différentes fonctions lors de la régulation transcriptionnelle.

1 – Régulation par les protéines TET

De façon intéressante, la déplétion des protéines TET induit à la fois une augmentation et une diminution de l'expression génique, ce qui suggère une double fonction de ces protéines dans la régulation transcriptionnelle (Wu *et al.*, 2011a). Les études visant à caractériser ces différentes fonctions ont permis l'identification de nombreuses protéines associées aux TET et ont de plus révélé que l'implication de ces protéines dans la régulation génique peut s'effectuer de manière dépendante ou indépendante de leur activité catalytique.

a – Répression par les TET

Plusieurs études ont mis en évidence, dans différentes lignées cellulaires humaines ou murines, une interaction entre TET1 et la protéine SIN3A, appartenant à un complexe répresseur de transcription (Cartron *et al.*, 2013 ; Deplus *et al.*, 2013 ; Shi *et al.*, 2013 ; Williams *et al.*, 2013). Contrairement à TET2, TET3 est également co-purifiée avec SIN3A (Deplus *et al.*, 2013 ; Williams *et al.*, 2013). Dans les cellules ES, pour lesquelles TET1 et SIN3A sont recrutés sur des sites communs localisés au niveau d'un certain nombre de TSS, la déplétion de TET1 perturbe le recrutement de SIN3A et entraîne une surexpression des gènes cibles (Shi *et al.*, 2013 ; Williams *et al.*, 2013). Au contraire, dans des cellules de glioblastome, c'est la déplétion de SIN3A, par siRNA (« small interference RNA »), qui diminue le recrutement de TET1 sur des régions communes (Cartron *et al.*, 2013). La déplétion de SIN3A ou de TET1 induit une augmentation de la méthylation sur un gène cible et diminue son expression. Ainsi, dans les cellules ES, mais pas dans les cellules de glioblastome, l'association de TET1 et SIN3A a un effet répressif sur la transcription qui s'avère être indépendant de l'activité catalytique de l'enzyme TET (Williams *et al.*, 2013). Il serait donc possible que cette interaction soit cellule-spécifique, rendant cet effet répressif dépendant de l'état de différenciation cellulaire. En plus de SIN3A, des composants majeurs de complexes de répression sont co-purifiés avec TET1 tels que HDAC1/2, MTA3, and CHD4 renforçant le lien entre les protéines TET et la répression transcriptionnelle (Shi *et al.*, 2013). En effet, lors de la résolution de l'inflammation dans des cellules myéloïdes, TET2 recrute HDAC2 et réprime ainsi la transcription d'un médiateur inflammatoire *via* la déacétylation des histones (Zhang *et al.*, 2015).

Toutefois, d'autres études révèlent que la répression exercée par TET1 peut-être indirecte, c'est-à-dire dépendante de son activité enzymatique. En effet, la protéine MBD3

appartenant au complexe répresseur NuRD co-localise avec TET1 et la 5hmC dans les cellules ES, et participe à la régulation des gènes hydroxyméthylés (Yildirim *et al.*, 2011). TET1 est nécessaire au recrutement de MBD3, qui reconnaît les cytosines hydroxyméthylées, et la présence de MBD3 est nécessaire au maintien des niveaux de 5hmC sur ces régions révélant une co-dépendance entre ces deux facteurs (Yildirim *et al.*, 2011). De façon intéressante, malgré le fait que TET1 et EZH2 (une des sous-unités du complexe PRC2, « Polycomb Repressive Complex 2 », appartenant à la famille des PcG et possédant une activité méthyltransférase d'histones) ne semblent pas interagir, la présence de TET1 sur des gènes cibles des complexes PcG est nécessaire au recrutement de EZH2, suggérant que l'activité enzymatique de TET1 est impliquée dans le recrutement indirect de EZH2 (Williams *et al.*, 2013 ; Wu *et al.*, 2011a).

b – Interactions avec les protéines OGT

Les protéines OGT (pour « O-linked N-acetylglucosamine (O-GlcNAc) transferase ») possèdent un rôle crucial pour la viabilité et le développement des cellules ES. Ces enzymes assurent la mise en place d'une modification post-traductionnelle réversible, l'O-GlcNAcylation, sur des protéines diverses telles que des régulateurs transcriptionnels ou des histones. Ces protéines ont récemment suscité un intérêt particulier car elles co-localisent avec les protéines TET à l'échelle du génome, plus particulièrement au niveau des îlots CpG et des TSS (Deplus *et al.*, 2013 ; Vella *et al.*, 2013). Des études approfondies ont révélées que les TET interagissent directement avec OGT, indépendamment de leur activité catalytique, et que cette interaction est nécessaire au recrutement d'OGT sur la chromatine (Chen *et al.*, 2013 ; Deplus *et al.*, 2013). Les résultats de ces travaux semblent discriminer deux fonctions distinctes entre les protéines TET1 et les protéines TET2 et TET3 dans la répression et l'activation transcriptionnelle, respectivement. En effet, dans les cellules ES, les enzymes TET1 et OGT participent à la répression de gènes développementaux et ainsi au maintien de la pluripotence (Shi *et al.*, 2013). De plus, SIN3A co-immunoprécipite avec OGT et TET1 suggérant que ces protéines participent à la formation d'un complexe répresseur (Shi *et al.*, 2013 ; Vella *et al.*, 2013). Dans des lignées HEK293T, les enzymes TET2 et TET3, quant à elles, co-localisent avec OGT sur les TSS de gènes fortement exprimés (Deplus *et al.*, 2013). Cette association permet l'O-GlcNAcylation de l'histone H2B (Chen *et al.*, 2013) et régule la triméthylation d'H3K4 via l'O-GlcNAcylation du facteur HCF1 (« Host Cell Factor 1 »), ce qui permet le recrutement du complexe de méthyltransférase d'histone SET1/COMPASS

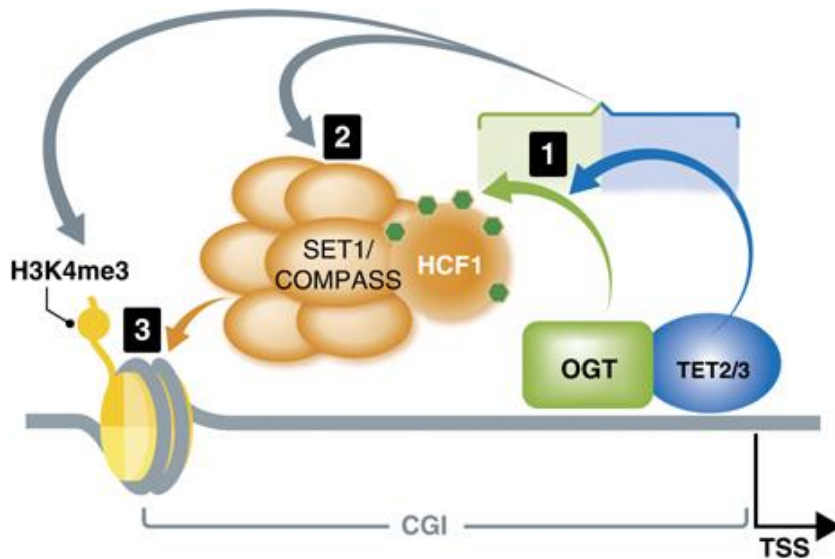


Figure 10. Modèle permettant de relier les enzymes de modifications de l'ADN (TET), une protéine « senseur » (OGT), et un complexe de modification des histones (SET1/COMPASS) (Deplus *et al.*, 2013).

Schéma hiérarchique d'une cascade d'évènements impliquant les activités des différentes protéines étudiées au niveau d'un îlot CpG (CGI) : (1) La première séquence d'évènements est la formation d'une interaction TET2/3-OGT, qui favorise la GlcNAcylation par OGT sur de nombreuses protéines, incluant HCF1; (2) Cette modification TET-dépendante de HCF1 est importante pour la formation du complexe SET1/COMPASS qui inclut HCF1; (3) Lors de la dernière étape, les protéines TET et l'activité d'OGT favorisent toutes deux la liaison de SETD1A au niveau de la chromatine, un événement nécessaire pour la mise en place d'H3K4me3 et l'activation transcriptionnelle ultérieure.

(« Complex proteins associated with SET1 » – Deplus *et al.*, 2013 – Figure 10), participant ainsi positivement à la régulation transcriptionnelle. Les protéines TET peuvent être elles-mêmes modifiées par OGT et cette modification jouerait plusieurs rôles : (i) l'O-GlcNAcylation de TET1 permettrait de stabiliser la liaison de TET1 sur la chromatine (Shi *et al.*, 2013 ; Vella *et al.*, 2013) ; (ii) la modification de TET3 est indispensable à sa relocalisation subcellulaire par OGT, ce qui n'est pas le cas de TET1 et TET2 (Zhang *et al.*, 2014) ; enfin, (iii) l'O-GlcNAcylation des protéines semblerait participer à la régulation des protéines TET en modulant leur taux de phosphorylation (Bauer *et al.*, 2015).

c – TET, pluripotence, différenciation et reprogrammation

La présence des protéines TET et de la 5hmC dans les cellules ES ainsi que leur(s) rôle(s) putatif(s) dans la régulation transcriptionnelle suggère leur implication dans les mécanismes de pluripotence, de différenciation et de reprogrammation. Dans ce sens, l'interaction de TET1 et TET2 avec le régulateur transcriptionnel PRDM14 induit la déméthylation et l'activation d'un groupe de gènes liés à la pluripotence, spécifiques de la lignée germinale ou soumis à l'empreinte (Okashita *et al.*, 2014). De plus, la différenciation adipocytaire ainsi que celle des monocytes en ostéoclastes est dépendante de l'interaction de TET1 et/ou TET2 avec PPAR γ (« Peroxisome proliferator-activated receptor gamma » – Fujiki *et al.*, 2013) ou PU.1 (de la Rica *et al.*, 2013), respectivement. Enfin, NANOG recrute TET1/2 afin d'augmenter l'expression d'un groupe de gènes clés impliqués dans la reprogrammation des cellules ES (Costa *et al.*, 2013).

2 – Régulation par TDG

Les enzymes TDG sont étudiées depuis longtemps pour leur rôle dans la réparation des mésappariements T/G et U/G de l'ADN double brin. Elles sont connues pour interagir avec plusieurs FT, les méthylases *de novo* DNMT3A et DNMT3B ainsi qu'avec des histones acétyltransférases ce qui évoque un rôle potentiel dans la régulation de la structure chromatinienne et de la transcription. Le rôle crucial de TDG n'a cependant été dévoilé qu'en 2011, lorsque deux études révélèrent que le KO de ces enzymes est létal à l'état embryonnaire chez la Souris (Cortázar *et al.*, 2011 ; Cortellino *et al.*, 2011). De façon intéressante, les trois quarts des gènes dont l'expression est altérée en absence de TDG sont réprimés, ce qui concorde avec une hyperméthylation d'un certain nombre de promoteurs. Les auteurs

proposent ainsi que le phénotype observé soit certainement lié à l'inactivation d'un processus associé à la régulation transcriptionnelle au cours du développement. Plusieurs fonctions de TDG ont alors été proposées. Premièrement, l'absence de TDG entraîne la perte de recrutement de la méthyltransférase d'H3K4, MLL1, et de l'histone acétyltransférase P300 sur un certain nombre de promoteurs (Cortázar *et al.*, 2011). Dans ce cas, TDG participerait directement ou indirectement à la modulation de la structure chromatinienne pour favoriser une chromatine de type « active ». Deuxièmement, l'hyperméthylation détectée au niveau des promoteurs touche majoritairement des gènes codant pour des TFs liés au développement, tels que les gènes Hox, et des gènes de réponses aux signaux hormonaux. En conséquence, TDG semble avoir deux fonctions, maintenir un état non méthylé et réguler la déméthylation initiée lors de la différenciation, ce qui concorde avec la découverte de l'implication de ces enzymes dans la dernière étape de déméthylation active des cytosines (Cortázar *et al.*, 2011 ; Cortellino *et al.*, 2011 ; He *et al.*, 2011 ; Maiti et Drohat, 2011). De plus, la distribution de TDG à l'échelle du génome correspond à celle de TET1 et des bases 5fC et 5caC, et son invalidation entraîne l'augmentation des 5fC et 5caC au niveau des TSS des gènes fortement exprimés (Neri *et al.*, 2015). L'ensemble de ces travaux permet de montrer que le phénotype léthal observé suite à l'absence de TDG est notamment dû à un défaut de régulation des modifications des cytosines impactant l'expression de gènes liés au développement. Une autre hypothèse, non exclusive des deux précédentes, prend en compte l'instabilité génomique résultante de la présence de 5fC et de 5caC non reconnues par le système de réplication (Kamiya *et al.*, 2002; Shibutani *et al.*, 2014).

Les enzymes TET et TDG possèdent toutes deux diverses fonctions mais leur rôle principal est de maintenir la stabilité épigénétique grâce à leur activité catalytique. Les différentes interactions observées avec leurs cofacteurs suggèrent la formation de larges complexes de régulation sur les régions régulatrices. En accord avec le postulat que la méthylation globale du génome est déposée par défaut et que les régions régulatrices nécessitent la présence d'un mécanisme de déméthylation pour garder ou obtenir leur état d'activation, la déplétion de TET1 et TET2 conduit à une augmentation de la méthylation, déposée par DNMT1 et DNMT3A sur les promoteurs de gènes fortement exprimés en conditions normales (Neri *et al.*, 2015). Ainsi les cycles de méthylation/déméthylation proposés il y a quelques années sur les promoteurs de gènes régulés par les œstrogènes (Kangaspeska *et al.*, 2008 ; Métivier *et al.*, 2008) semblent à présent concerner toutes les

régions génomiques dont le niveau de méthylation doit être modulé afin d'assurer la conservation ou l'établissement d'un programme transcriptionnel particulier.

D – Protéines de reconnaissance des dinucléotides CG

1 – CpG non modifiés

In vitro, le domaine ZF-CXXC (« Zinc Finger-CXXC ») est capable de se lier sur des cytosines non méthylées (Voo *et al.*, 2000), et il s'avère que les protéines comportant ce domaine sont principalement localisées au niveau des îlots CpG. Parmi ces protéines, nombreuses sont celles dont la fonction est liée directement ou indirectement à la structure de la chromatine (Long *et al.*, 2013). En effet, des protéines appartenant à des complexes de méthylation/déméthylation des histones (CFP1, MLL1/2, KDM2A/B – « Lysine-specific demethylase 2A/B ») ainsi que les enzymes TET1 et TET3 contiennent un domaine CXXC. Il semblerait toutefois que ce domaine ne soit pas nécessairement responsable du ciblage de ces protéines sur les îlots CpG. Néanmoins, la présence de cytosines non modifiées favorise leur fixation et en conséquence l'établissement d'un environnement chromatinien particulier.

2 – CpG méthylés

Comme précédemment mentionné, une catégorie particulière de protéines dites de liaison à l'ADN méthylé est capable de reconnaître spécifiquement cette modification. Ces protéines sont classées en différents groupes selon leur domaine de reconnaissance aux CpG méthylés. D'une part, les protéines appartenant à la famille des MBD telles que MeCP2, MBD1, 2 et 4 (Hendrich et Bird, 1998 ; Lewis *et al.*, 1992 ; Meehan *et al.*, 1989), dont la liaison, qui nécessite généralement la présence d'une forte densité en CpG, participe indirectement à la régulation exercée par la méthylation en recrutant des complexes co-régulateurs. D'autre part, des protéines à doigts de zinc telles que Kaiso et Klf2, 4 et 5 (Spruijt *et al.*, 2013 ; Yoon *et al.*, 2003) ainsi que UHRF1 et son domaine SRA (Sharif *et al.*, 2007) composent les deux autres familles majeures de protéines de reconnaissance à l'ADN méthylé.

3 – CpG oxydés

L'évolution des techniques de spectrométrie de masse a permis le développement de protocoles permettant d'identifier les facteurs associés avec une sonde nucléotidique d'intérêt. Deux études réalisées en 2013 dans les laboratoires de Vermeulen et de Reik ont notamment visé à déterminer les protéines interagissant avec les cytosines modifiées (Iurlaro *et al.*, 2013 ; Spruijt *et al.*, 2013). Spruijt et collègues ont effectué leurs travaux en cellules ES, NPC (pour « neural progenitor cells ») et tissus adultes de cerveaux murins avec une sonde comportant un enchainement de quatre trinuécléotides ACG, tandis qu'en cellules ES, Iurlaro et collègues ont utilisé deux sondes différentes correspondant à deux promoteurs dont l'un est dans un contexte d'îlot CpG. D'une manière générale, les résultats obtenus montrent qu'il existe peu de protéines retenues sur les sondes qui soient communes aux différentes modifications des cytosines. Les protéines identifiées en complexe avec la sonde hydroxyméthylée dans les cellules ES et les NPC sont peu nombreuses mais suggèrent tout de même une double fonctionnalité de la 5hmC. La détection de glycosylases et d'hélicases soutient un rôle de la 5hmC dans la voie de déméthylation active, tandis que l'interaction de MeCP2 et UHRF1, préalablement démontrée avec la sixième base de l'ADN (Frauer *et al.*, 2011 ; Mellén *et al.*, 2012), suggère son implication en tant que marque épigénétique. Au cours du développement, la composition des complexes interagissant avec les 5mC et les 5hmC sont très dynamiques, un fait en partie dû à l'expression spécifique de certaines protéines dans certains types cellulaires uniquement (Spruijt *et al.*, 2013). Par exemple, UHRF2 n'est identifiée comme associée à la 5hmC qu'à partir des NPC car elle n'est pas exprimée dans les cellules ES. Une analyse approfondie de l'interaction entre UHRF2 et 5hmC révèle que UHRF2 stimule de façon transitoire l'activité enzymatique séquentielle de TET1 (Spruijt *et al.*, 2013). De plus, les protéines identifiées avec la sonde hydroxyméthylée sont particulièrement plus nombreuses que celles identifiées avec la sonde méthylée dans le tissu adulte pour lequel les niveaux endogènes de 5hmC sont les plus hauts, suggérant un rôle particulier de la 5hmC dans le cerveau. Contrairement à des travaux précédents révélant une préférence de liaison de la protéine MBD3 sur des cytosines hydroxyméthylées comparées à des cytosines méthylées *in vitro* (Yildirim *et al.*, 2011), l'inverse a été observé dans ces deux études (Iurlaro *et al.*, 2013 ; Spruijt *et al.*, 2013) suggérant un impact très important de la séquence de la sonde utilisée sur les résultats obtenus.

L'analyse des protéines identifiées à partir des sondes contenant des cytosines formylées a révélé la présence de protéines ayant des fonctions liées à la réparation de l'ADN, en accord avec un rôle de la 5fC comme intermédiaire de déméthylation (Iurlaro *et al.*, 2013 ; Spruijt *et al.*, 2013). Cependant, un plus grand nombre de protéines interagissent uniquement avec ces sondes comparé aux autres sondes modifiées (Iurlaro *et al.*, 2013). Ces protéines spécifiques de la 5fC possèdent des activités liées à la transcription et à la régulation de la chromatine qui semblent être majoritairement répressives dans le contexte d'un îlot CpG. Ainsi, ces résultats suggèrent que la 5fC peut également agir sur la régulation génique et que la nature activatrice ou répressive de cette action dépend de la densité en CpG du promoteur (Iurlaro *et al.*, 2013).

Le nombre limité de protéines communes associées aux différentes modifications des cytosines tend à suggérer que ces bases ont des fonctions distinctes. Cependant, un point commun peut venir du fait qu'elles excluent le même type de facteurs tels que certaines protéines contenant un domaine CXXC (Spruijt *et al.*, 2013). Néanmoins, il est important de préciser que les associations présentées par Spruijt et Iurlaro, qu'ils nomment « interactors », n'impliquent pas forcément une interaction directe entre la cytosine modifiée et les protéines identifiées. De plus, les sondes utilisées par Spruijt et collègues contiennent une séquence arbitraire non représentative de régions régulatrices du génome. De ce fait, il serait dangereux de généraliser les résultats obtenus avant la réalisation d'études complémentaires.

III – LES ENHANCERS

Les enhancers sont des régions régulatrices dites distantes car elles sont généralement situées entre quelques kilobases voire des milliers de kilobases des gènes dont elles modulent la transcription. Un certain nombre de critères permettent de caractériser et d'identifier ces séquences d'environ 200 à 300 pb de longueur dont une des caractéristiques est la présence de nombreux sites de fixations pour des FT. Les enhancers sont responsables de la régulation spatio-temporelle des gènes nécessaire à un contrôle précis des processus de développement, différenciation et homéostasie cellulaires. Leur sélection est donc une étape cruciale afin qu'ils puissent exercer leurs fonctions spécifiques d'un type cellulaire ou d'un tissu, en interagissant avec les promoteurs des gènes ciblés.

A – Identification des régions enhancers

Les enhancers possèdent de nombreuses caractéristiques, qu'elles soient chromatinienne, protéique ou transcriptomique, qui ont permis de définir et d'étudier ces régions régulatrices à l'échelle du génome. La combinaison de plusieurs de ces critères permet d'accroître la précision de détection des enhancers et de distinguer leurs différents états d'activation.

1 – Comparaison de séquences

Les premières méthodes d'identification des enhancers *in silico* étaient basées sur la comparaison de séquences entre différents organismes plus ou moins éloignés à l'échelle de l'évolution. En effet, cette approche repose sur le postulat que les séquences des éléments régulateurs sont sous sélection négative afin de ne pas induire de changement de séquence délétère pour l'organisme. Plusieurs études ont entrepris d'utiliser cette approche en comparant les séquences génomiques de l'Homme et de différents organismes tels que la Souris, le Poisson-globe, le Rat et la Grenouille (Nobrega *et al.*, 2003 ; Pennacchio *et al.*, 2006 ; Wolfe *et al.*, 2005). Pour ces travaux, l'alignement de séquences a permis d'identifier des enhancers putatifs fortement conservés, dont l'activité d'un certain nombre a été confirmée par transgénèse. Des outils statistiques tels que PhastCons ont été développés afin d'exprimer la conservation phylogénétique d'une région génomique sous forme d'un score

(Siepel *et al.*, 2005). Les analyses basées sur la comparaison génomique comportent néanmoins plusieurs limites. D'une part, la conservation seule est insuffisante pour caractériser l'activité spatio-temporelle des enhancers ; d'autre part, certaines régions régulatrices ne sont pas ou peu « contraintes » au cours de l'évolution et ne seront donc pas prises en compte (Blow *et al.*, 2010 ; Encode project, Birney *et al.*, 2007).

2 – Caractéristiques chromatinienne des enhancers

a – Modifications des histones

Les modifications de la queue des histones des nucléosomes présents au niveau des régions promotrices et insulatrices sont majoritairement invariantes entre les différents types cellulaires. Par opposition, les modifications observées aux enhancers sont généralement propres à chaque type cellulaire, coïncidant avec le rôle de ces régions lors du développement ou de la différenciation, même s'il existe un certain nombre d'enhancers ubiquitaires (Heintzman *et al.*, 2009).

Les modifications post-traductionnelles aux enhancers s'étendent généralement sur de larges régions et la mise en place ainsi que l'effacement de ces marques sont très dynamiques durant l'établissement de programmes transcriptionnels visant à réguler la destinée cellulaire. Le développement des techniques de ChIP à l'échelle du génome (ChIP-chip et ChIP-seq) ainsi que la disponibilité d'un grand nombre d'anticorps dirigés contre une quantité croissante de modifications ont permis d'identifier des caractéristiques chromatinienne propres aux régions enhancers. Dans cette partie, seules les marques associées aux enhancers les plus courantes sont présentées mais des modifications supplémentaires ont été utilisées dans certaines études afin d'affiner leur caractérisation (Ernst *et al.*, 2011 ; Hon *et al.*, 2009 ; Wang *et al.*, 2008).

La première modification des histones associée aux enhancers est la méthylation d'H3K4. Tandis que la présence d'H3K4me1 et H3K4me2 aux enhancers fait l'unanimité, celle d'H3K4me3 est controversée. En effet, certaines études ont montré que les enhancers présentent la marque H3K4me1 et/ou H3K4me2 sans H3K4me3, ce qui permettrait de distinguer les enhancers des promoteurs (Heintzman *et al.*, 2007, 2009 ; Encode project, Birney *et al.*, 2007 ; Koch *et al.*, 2007). D'autres travaux ont cependant observé les trois

formes de méthylation au niveau de ces régions régulatrices, même si le ratio H3K4me1/H3K4me3 reste plus élevé aux enhancers qu'aux promoteurs (Barski *et al.*, 2007 ; Pekowska *et al.*, 2011 ; Wang *et al.*, 2008). Pekowska et collègues ont de plus montré que la présence d'H3K4me3, acquis lors de la différenciation, est liée à l'activité des enhancers dans les cellules T et les lymphocytes B, contrairement à l'enrichissement d'H3K4me1 dont l'enrichissement est indépendant de l'activité (Pekowska *et al.*, 2011).

L'acétylation d'H3K27 est une autre marque permettant de caractériser les enhancers, et plus particulièrement les enhancers actifs (Creyghton *et al.*, 2010 ; Heintzman *et al.*, 2009 ; Rada-Iglesias *et al.*, 2011). En effet, l'expression des gènes est corrélée à la présence, sur des régions enhancers voisines, des marques H3K4me1 et H3K27ac (Creyghton *et al.*, 2010 ; Hawkins *et al.*, 2011 ; Heintzman *et al.*, 2009). D'une façon plus générale, l'acétylation aux enhancers semble liée à l'activation génique (Roh *et al.*, 2005). L'étude de la dynamique de modifications des histones au niveau des enhancers en lien avec l'expression des gènes associés a permis de distinguer, dans un premier temps, deux types d'enhancers : (i) les enhancers marqués par H3K4me1 et H3K27ac, associés à la régulation transcriptionnelle spécifique d'un type cellulaire, et (ii) les enhancers enrichis seulement en H3K4me1. Ceux-ci sont dits « poised » car ils présentent un potentiel d'activation pour l'expression de gènes développementaux lors du possible engagement de la cellule dans des voies de différenciation (Cui *et al.*, 2009 ; Creyghton *et al.*, 2010 ; Hawkins *et al.*, 2011 ; Heintzman *et al.*, 2009). De façon intéressante, deux études ont par la suite montré la présence d'une autre modification des histones, H3K27me3, au niveau de certains enhancers H3K4me1 positifs/H3K27ac négatifs (Rada-Iglesias *et al.*, 2011, Zentner *et al.*, 2011). Ces enhancers, pouvant être également enrichis en H3K9me3, sont observés dans les cellules ES humaines et murines et deviennent actifs lors de la différenciation, par le remplacement d'H3K27me3 en H3K27ac. Ceci assure un contrôle efficace de la régulation des gènes impliqués dans les étapes précoces de développement, en réponse à un stimulus ou à un signal développemental (Rada-Iglesias *et al.*, 2011, Zentner *et al.*, 2011). De ce fait, les enhancers qui ne montrent qu'un enrichissement en H3K4me1 (sans H3K27me3) sont désormais qualifiés de « primed » et semblent réguler des processus cellulaires plus généraux (Zentner *et al.*, 2011). Il est également intéressant de noter que les niveaux d'H3K4me1 observés aux enhancers « primed » et « poised » sont plus faibles que ceux observés au niveau des enhancers actifs et

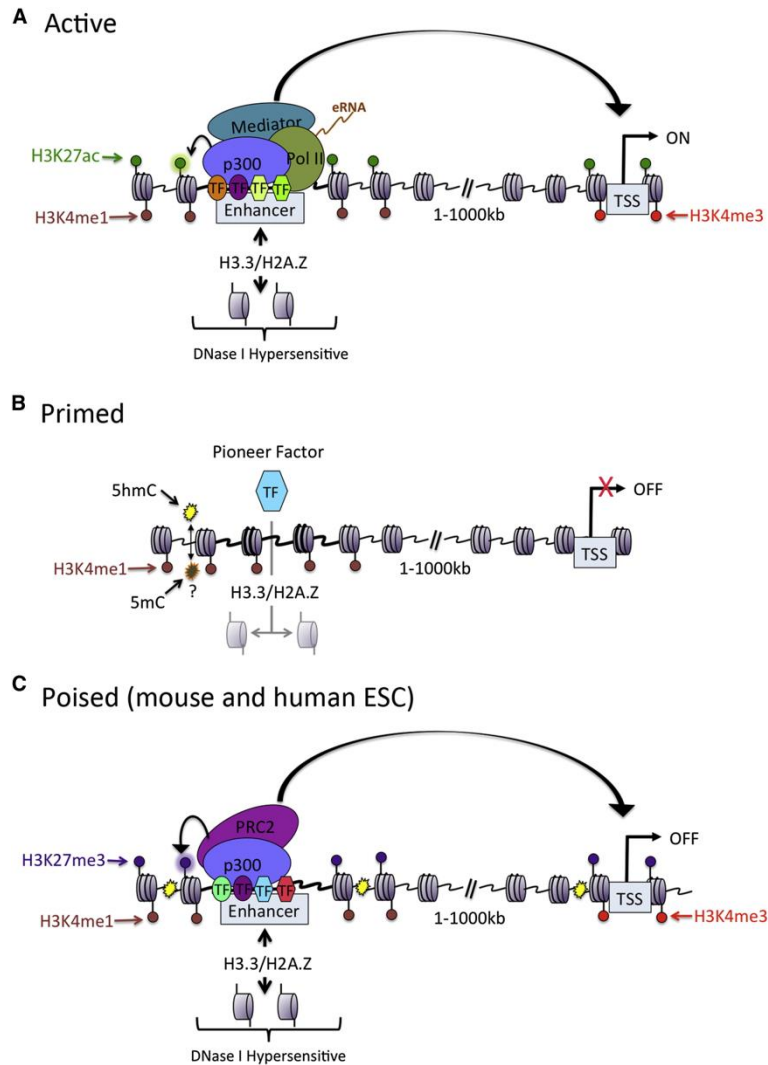


Figure 11. Caractéristiques des enhancers actifs, « primed » et « poised ».
(Calo et Wysocka, 2013).

(A) Représentation schématique des principales caractéristiques chromatinienne des enhancers actifs. Les enhancers sont associés à l'incorporation de nucléosomes hypermobiles composés des variants d'histone H3.3/H2A.Z, et à la mobilisation de facteurs de transcription qui peuvent à leur tour recruter des protéines co-activatrices qui peuvent modifier et remodeler les nucléosomes. L'ensemble de ces processus se reflète par l'hypersensibilité de cette région génomique à la digestion à la DNase1. H3K4me1 et H3K27ac sont les modifications d'histones prédominantes des nucléosomes encadrant ces enhancers qui peuvent contacter les promoteurs de leurs gènes cibles par l'intermédiaire de boucles de chromatine. (B) Préalablement à l'état d'activation, les enhancers existent à l'état « primed », caractérisé par la présence d'H3K4me1. D'autres caractéristiques qui ont été associées au « priming » des enhancers sont la présence de facteurs pionniers, de nucléosomes hypermobiles H3.3/H2A.Z, d'une hypométhylation de l'ADN. (C) Représentation schématique des caractéristiques des enhancers « poised » identifiés dans les cellules ES humaines et murines. Ces enhancers sont marqués par la présence d'H3K27me3 et sont associés à PRC2. Ils sont liés par des facteurs de transcription et des coactivateurs, et communiquent avec leurs promoteurs cibles.

qu'ils diminuent ensuite après différenciation (Creyghton *et al.*, 2010 ; Rada-Iglesias *et al.*, 2011, Zentner *et al.*, 2011).

Ainsi, en plus de différencier les enhancers des autres régions, les modifications des histones permettent de définir la dynamique d'expression de gènes essentiels lors la différenciation cellulaire (Figure 11).

b – Positionnement et nature des nucléosomes aux enhancers

Les régions correspondantes aux enhancers actifs ou « poised » sont caractérisées par une diminution de leur densité en nucléosomes (He *et al.*, 2010), ce qui permet de les identifier par les techniques de FAIRE (« Formaldehyde Isolation of Regulatory Elements »), d'empreinte à la DNaseI ou d'ATAC-seq (« Assay for Transposase-Accessible Chromatin with highthroughput sequencing » – Birney *et al.*, 2007; Buenrostro *et al.*, 2013 ; Song *et al.*, 2011). Il semblerait cependant qu'au lieu d'être affranchies en nucléosomes, les séquences correspondantes aux sites de liaisons des FT seraient associées à des nucléosomes instables composés des variants H2A.Z et H3.3, ayant une forte conservation de la séquence ADN associée avec une composition en nucléotides plus riche en A/T que les nucléosomes encadrant les sites de reconnaissance des FT, pour leur part moins mobiles (He *et al.*, 2010 ; Jin et Felsenfeld, 2007 ; Jin *et al.*, 2009 ; Mito *et al.*, 2007). Plusieurs études ont par la suite proposé que l'incorporation d'H2A.Z et la dynamique de remplacement des histones participerait à faciliter l'accessibilité des FT aux séquences régulatrices (Hu *et al.*, 2013 ; Li *et al.*, 2012a ; Mito *et al.*, 2007 – Figure 11).

c – Modifications des cytosines

L'analyse du méthylome et de l'hydroxyméthylome des cellules ES a montré une distribution particulière des cytosines modifiées au niveau des régions enhancers. En effet, les séquences enhancers ont pour particularité d'avoir une faible densité en CpG et d'être hypométhylées (Lister *et al.*, 2009 ; Stadler *et al.*, 2011). Ceci concorde avec l'anti-corrélation observée entre méthylation de l'ADN et modifications des histones associées à l'activité des enhancers, ou liaison des FT (Neph *et al.*, 2012; Stadler *et al.*, 2011; Schmidl *et al.*, 2009), ainsi qu'avec le rôle répresseur de la 5mC. A l'inverse, en lien avec la mise en œuvre du processus de déméthylation active au niveau des régions régulatrices, la 5hmC est spécifiquement enrichie aux enhancers et est associée à la présence d'H3K4me1 et

d'H3K27ac au cours de la différenciation, marquant ainsi les enhancers de façon spécifique des types cellulaires (Sérandour *et al.*, 2012 ; Stroud *et al.*, 2011). Les modifications des cytosines et plus particulièrement l'enrichissement en 5hmC, constituent ainsi des critères supplémentaires d'identification des enhancers.

3 – Recrutement de facteurs protéiques

a – Recherche de motifs de liaison pour des FT

Plusieurs méthodes, non exclusives les unes des autres, permettent de prédire la nature « enhancer » de régions génomiques en analysant leur densité en sites de liaison pour des FT (Aerts, 2012 ; Wang *et al.*, 2013). La détection de ces sites comporte une difficulté particulière du fait que les FT sont capables de reconnaître leur motif de liaison avec un certain degré de divergence de séquence. La création de matrices appelées PWM (pour « position weight matrix ») permet de connaître la probabilité d'observer chaque nucléotide A, C, G et T à chacune des positions du motif de liaison d'un FT donné. Ces matrices sont disponibles dans des bases de données telles que TRANSFAC, JASPAR ou MEME (Bailey *et al.*, 2009 ; Matys *et al.*, 2003 ; Portales-Casamar *et al.*, 2009) et sont utilisées par différents outils afin d'identifier les motifs présents dans des séquences données (Aerts, 2012). Cependant, ces analyses génèrent un grand nombre de faux positifs. Ainsi, afin de caractériser des enhancers, la recherche de motifs *via* les PWM est généralement associée à une recherche de groupes de sites de liaison et/ou à une comparaison entre différents organismes. En effet, au niveau des régions régulatrices, les sites de reconnaissances des FT sont particulièrement regroupés. La détection de ces combinaisons de motifs permet alors d'affiner la localisation des séquences régulatrices même si cette technique est peu sensible et basée sur un seuil arbitraire (Aerts, 2012 ; Bailey et Noble, 2003; Johansson *et al.*, 2003). Les prédictions obtenues par PWM peuvent également être filtrées selon leur conservation phylogénétique car les régions régulatrices sont majoritairement conservées au cours de l'évolution. Cependant, par cette méthode, il est possible de manquer des sites spécifiques d'un type cellulaire ou d'un organisme, ou peu conservés (Aerts, 2012).

Pour l'instant, ces protocoles *in silico* de détection d'enhancers montrent plusieurs inconvénients et nécessitent donc d'être associés à l'examen de paramètres additionnels comme l'identification des marques chromatinienne associées à ces régions putatives.

L'inconvénient majeur, inhérent à la technique, est l'absence d'information concernant la dynamique d'activation de ces régions.

b – Liaison par des facteurs co-activateurs

La diversité des FT pouvant se lier au niveau des enhancers rend l'identification de ces régions à partir de la liaison de protéines activatrices beaucoup trop complexe. Une alternative consiste à examiner les sites de fixation des facteurs co-activateurs tels que P300 ou le complexe Médiateur.

Les protéines de la famille CBP (« CREB-binding protein »)-P300 sont des acétyltransférases exprimées dans de nombreux tissus, décrites comme des enzymes co-activatrices essentielles pour le développement embryonnaire (Yao *et al.*, 1998). Heintzman et collègues ont analysé le potentiel prédictif de la présence de P300 pour la détermination des régions enhancers dans des cellules humaines (Heintzman *et al.*, 2007). Les sites de liaison de P300 s'avérèrent associés à plusieurs caractéristiques des enhancers dont la localisation distale par rapport aux promoteurs, la sensibilité à la DNaseI, la conservation de séquence et la présence d'H3K4me1. Quelques enhancers putatifs identifiés grâce à la liaison de P300 ont ensuite été validés par l'expression de gènes rapporteurs dans des Souris transgéniques confirmant que la liaison de ce co-activateur permet d'identifier précisément des enhancers actifs (Heintzman *et al.*, 2007). La comparaison et la validation des prédictions obtenues à partir de trois tissus du cerveau d'embryons de Souris démontrent qu'en plus de diminuer les faux positifs comparé aux analyses de conservation de séquences, la liaison de P300 sur une région donnée reflète son activité spécifiquement liée au type cellulaire dans lequel elle a été détectée (Visel *et al.*, 2009). Cependant, dans les cellules ES humaines, les enhancers « poised » portant les marques H3K4me1 et H3K27me3 sans H3K27ac, sont également liés par P300 (Rada-Iglesias *et al.*, 2011). Ce nouveau paramètre d'identification des enhancers a ensuite été utilisé dans diverses études et a notamment permis l'identification d'enhancers peu conservés dans le cœur (Blow *et al.*, 2010). De façon intéressante, le recrutement de CBP/P300 est induit après stimulation inflammatoire dans les macrophages ou après dépolarisation des cellules corticales et permet ainsi de caractériser spécifiquement les enhancers inductibles impliqués dans la régulation associée à un stimulus (Ghisletti *et al.*, 2010 ; Kim *et al.*, 2010).

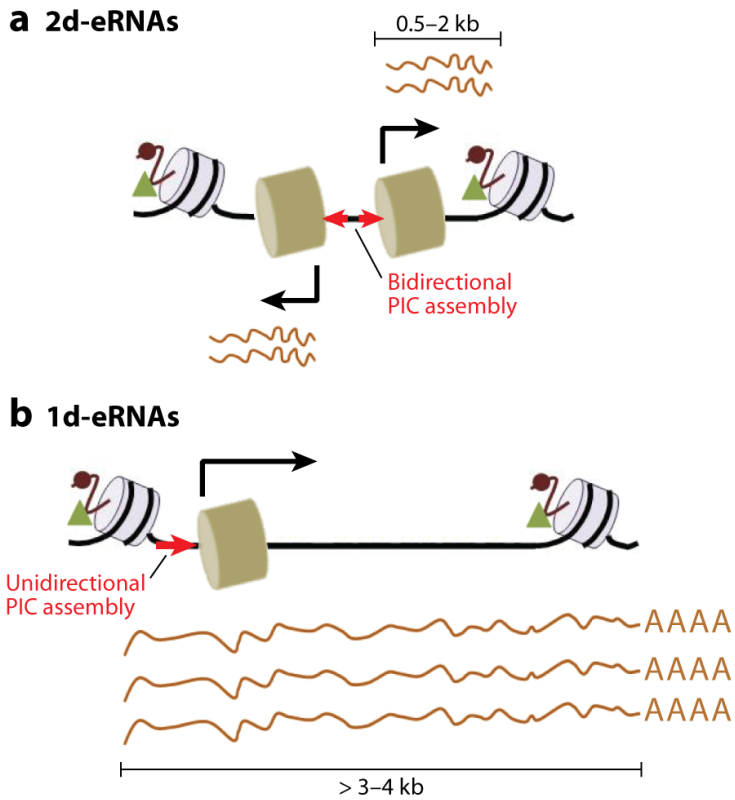


Figure 12. Schématisation des deux types d'ARN transcrits par les enhancers (eRNA) (Natoli et Andrau, 2012).

(a) eRNA bidirectionnels, relativement courts, et non polyadénylés (2d-eRNAs). (b) eRNA unidirectionnels, long et polyadénylés (1d-eRNAs). Sur ces schémas, l'ARN Pol II est illustrée par un cylindre marron, tandis que les cylindres gris représentent les nucléosomes.

D'autres études ont tiré avantage de la large distribution du complexe Médiateur (Heintzman *et al.*, 2007, 2009 ; Kagey *et al.*, 2010). L'association de ce complexe avec des FT essentiels au programme transcriptionnel des cellules ES (appelés « Masters » FT) a par ailleurs permis d'identifier des régions composées de groupes d'enhancers appelées super-enhancers (voir le point D de cette partie – Hnisz *et al.*, 2013 ; Lovén *et al.*, 2013 ; Whyte *et al.*, 2013).

L'identification d'enhancers à partir du recrutement de co-activateurs transcriptionnels s'est révélée être efficace et très spécifique. Cependant, dans les cellules ES, cette méthode ne permet pas de distinguer les enhancers « poised » des enhancers actifs (Rada-Iglesias *et al.*, 2011). De plus, les prédictions obtenues avec un seul type de co-activateur ne prendront pas en compte les enhancers uniquement liés par d'autres co-activateurs (Heintzman *et al.*, 2007).

4 – Présence de l'ARN Pol II et transcription aux enhancers

De façon inattendue, plusieurs études réalisées à l'échelle du génome murin à partir de macrophages, de neurones primaires et de cellules T ont montré que les enhancers peuvent également être caractérisés par le recrutement de l'ARN Pol II et de GTFs (De Santa *et al.*, 2010 ; Kim *et al.*, 2010 ; Koch *et al.*, 2011). De plus, ce recrutement est associé à la synthèse d'ARN nommées eRNA (pour « enhancer RNA ») car transcrits à partir de ces séquences régulatrices (De Santa *et al.*, 2010 ; Kim *et al.*, 2010 ; Koch *et al.*, 2011). En outre, la transcription des enhancers est dépendante de stimulus et est positivement corrélée à l'expression, de façon spécifique du type cellulaire, des gènes voisins (De Santa *et al.*, 2010 ; Kim *et al.*, 2010). D'après ces différentes études, ces longs ARN non-codants peuvent être classés en deux catégories : (i) les ARN unidirectionnels, polyadénylés mais très instables, d'une taille supérieure à 3 kb, et (ii) les ARN bidirectionnels, non polyadénylés, mesurant généralement moins de 2 kb de longueur et formant la catégorie la plus fréquemment observée (De Santa *et al.*, 2010 ; Kim *et al.*, 2010 ; Koch *et al.*, 2011 – Figure 12). Ainsi, le recrutement de l'ARN Pol II et la synthèse d'ARN non-codants bidirectionnels, hors régions promotrices, constituent de nouveaux paramètres permettant de spécifier l'emplacement d'enhancers actifs. Ceci a été facilité par la mise au point des techniques de GRO-seq (pour « global nuclear run-on sequencing ») qui permet de détecter la localisation et la quantité

d'ARN naissants à l'échelle du génome (Core *et al.*, 2008 ; Melgar *et al.*, 2011) ou celle de CAGE (Andersson *et al.*, 2014 ; Kodzius *et al.*, 2006).

La synthèse de transcrits à partir des enhancers est maintenant admise comme étant un phénomène courant et observé dans de nombreux types cellulaires, Cependant, les fonctions exactes des eRNA ne sont pas encore élucidées, même s'il existe des indications d'un possible rôle dans la régulation transcriptionnelle. Il est néanmoins possible de classer les rôles putatifs de la synthèse de eRNA en trois catégories : (i) aucune fonction biologique, les eRNA constitueraient seulement du bruit, dû à la fixation aléatoire de l'ARN Pol II sur des régions accessibles, (ii) rôle indirect, l'activation des enhancers dans ce cas ne serait liée qu'à la fixation de l'ARN Pol II permettant de recruter divers facteurs de régulations, la synthèse des eRNA ne serait donc qu'un « effet secondaire », (iii) rôle direct, les eRNA pourraient agir sur la transcription en *cis* ou en *trans*, tout comme certains ARN longs non-codants (Natoli et Andrau, 2012). De façon intéressante, plusieurs études ont évoqué un lien entre eRNA et formation de boucles chromosomiques entre enhancers et promoteurs. En effet, le « knockdown » d'eRNA provenant d'enhancers liés par le récepteur aux œstrogènes diminue les interactions promoteurs-enhancers (Li *et al.*, 2013). *A contrario*, Hah et collègues ont montré que l'inhibition de l'élongation de l'ARN Pol II ne perturbe pas la formation de boucles de chromatine. Dans ce cas, la corrélation entre la synthèse des eRNA et les boucles de chromatine signifierait que leur transcription dépend de la formation de boucle plutôt que l'inverse (Hah *et al.*, 2013). Cette hypothèse est corroborée par les travaux de Lin *et al.* durant le développement des cellules B (Lin *et al.*, 2012). Des travaux récents ont cependant observé une concomitance de la synthèse des eRNA et l'interaction entre le promoteur et la région du contrôle de locus du gène codant pour la β -globine (« Locus control region », LCR – Kim *et al.*, 2015), renforçant la nécessité d'effectuer des études supplémentaires afin de définir le lien entre ces deux phénomènes.

5 – Cartographie des enhancers à l'échelle du génome

L'utilisation combinée des différentes caractéristiques des enhancers énoncées précédemment augmente la robustesse de leur détection. De ce fait, de nombreuses études ont associé plusieurs paramètres afin de définir le répertoire d'enhancers existant au sein de différentes lignées cellulaires humaines et murines. Certaines de ces études, visant à annoter

globalement le génome humain, proviennent de consortiums (Roadmap Epigenomics *et al.*, 2015; Encode Project Consortium *et al.*, 2012), tandis que d'autres utilisent les données de ces consortiums afin d'annoter les enhancers (Andersson *et al.*, 2014). A plus petite échelle de collaborations, certains travaux ont combiné modifications des histones et recrutement de l'ARN Pol II (Bonn *et al.*, 2012 ; Shen *et al.*, 2012 ; Heintzman *et al.*, 2007), ou présence d'H2A.Z (Wang *et al.*, 2008), ou encore sensibilité à la DNaseI et recrutement de P300 (Heintzman *et al.*, 2007, 2009). Hon et collègues ont pour leur part effectué une première sélection d'enhancers par un fort niveau d'H3K4me1 et un faible niveau d'H3K4me3 qu'ils ont ensuite trié en fonction de la présence d'H2A.Z, de 20 modifications d'histones, de la protéine CTCF et de l'ARN Pol II (Hon *et al.*, 2009). Ces différentes études fournissent ainsi un catalogue des enhancers essentiel pour appréhender les fonctions régulatrices de ces régions.

6 – Activité des enhancers

Certains paramètres décrits précédemment permettent de distinguer les enhancers actifs des enhancers inactifs. Cependant, cette détection reste indirecte et ne permet pas de mesurer réellement le niveau d'activité d'un enhancer. Plusieurs études ont donc cherché à valider les données obtenues en ChIP par des tests de gènes rapporteurs. Deux études réalisées dans le même laboratoire ont notamment utilisé des Souris transgéniques afin de vérifier avec succès l'expression spatio-temporelle de gènes exogènes (dans ce cas, le gène LacZ) placés sous contrôle d'enhancers déterminés par ChIP-seq (Blow *et al.*, 2010 ; Visel *et al.*, 2009). Un catalogue d'enhancers validés par cette approche chez l'Homme ou la Souris est disponible sur le « Vista Enhancer Browser » (<http://enhancer.lbl.gov/> - Visel *et al.*, 2007).

Cette technique utilisant un gène rapporteur ne permet cependant pas d'analyser l'activité d'enhancers à grande échelle. Afin de palier à ce problème, Arnold et collègues ont développé une méthode nommée STARR-seq (pour « self-transcribing active regulatory region sequencing ») permettant d'identifier et de quantifier l'activité de régions enhancers à partir d'un ensemble de fragments aléatoires d'ADN (Arnold *et al.*, 2013). Pour cela, les fragments d'ADN sont insérés en aval d'un promoteur minimal au sein d'un vecteur. L'activité potentielle d'enhancer de la séquence ADN sera ainsi déterminée par sa capacité à activer sa propre transcription après transfection en cellules (Arnold *et al.*, 2013). Plus

récemment, Vanhille et collègues ont associé une étape préalable de capture de régions à la technique du STARR-seq, donnant lieu au CapStarr-seq (Vanhille *et al.*, 2015). Brièvement, des régions d'intérêts ont été déterminées par leur enrichissement en FT d'intérêts et leur sensibilité à la DNaseI afin de réaliser une puce ciblant ces régions. Une banque d'ADN génomique a ensuite été hybridée sur la puce afin d'amplifier les régions d'intérêts avant de les soumettre au protocole STARR-seq (Vanhille *et al.*, 2015). Le développement de telles technologies a permis d'augmenter la quantité d'enhancers potentiels testés à partir de différents génomes et de diminuer le coût technique tout en assurant une détection efficace de l'activité de ces régions régulatrices.

B – Sélection des enhancers

La sélection des enhancers permet de préparer la chromatine en initiant un changement dans le positionnement des nucléosomes, l'ouverture de la chromatine et la mise en place de modifications d'histones. L'activité régulatrice des enhancers s'effectue notamment grâce à la présence au sein de leur séquence de nombreux sites de fixations pour des FT ; ces régions sont ainsi couramment qualifiées de « plateformes de liaison de FT ». En effet, la mobilisation de co-activateurs ne possédant pas de domaines de liaison à l'ADN nécessite la fixation préalable de FT dont le recrutement dépend de signaux contrôlant une réponse cellulaire spécifique. Ces FT sont généralement largement exprimés et possèdent des sites de fixation soit communs à différents types cellulaires, soit spécifiques d'un type cellulaire. Ainsi, la sélection d'enhancers cellule-spécifiques reflète une combinatoire complexe entre le nombre (i) d'enhancers putatifs (de 300 000 à 1 million dans le génome humain - Encode Project Consortium *et al.*, 2012), (ii) de types cellulaires dans un organisme, (iii) de FT exprimés, et (iv) de sites potentiels de liaison pour ces FT, et nécessite probablement la présence de facteurs additionnels spécifiques d'un type cellulaire. La diversité et complexité des systèmes étudiés ainsi que certaines contradictions concernant la séquence d'interventions de certains acteurs a pour le moment empêché l'émergence d'un modèle général décrivant la dynamique des événements et facteurs impliqués dans la sélection des enhancers.

1 – Enhancers *de novo*

Les processus de développement, différenciation et de réponse à des stimuli environnementaux impliquent la sélection *de novo* d'enhancers inactifs qui vont s'activer suite à un signal particulier afin d'assurer la transition vers un lignage ou un état cellulaire particuliers. Cependant, la structure condensée adoptée par la chromatine au niveau de régions inactives entrave la sélection de ces nouveaux sites de régulation, car la plupart des FT sont incapables de reconnaître individuellement leurs sites de fixations dans de telles conditions. Même si la fixation coopérative des FT permet la reconnaissance des nucléosomes (Adams et Workman, 1995), la plupart des études portent sur la compréhension de la sélection des sites pour lesquels la fixation des FT est séquentielle. Plus particulièrement, ces travaux visent à caractériser le rôle des facteurs pionniers et du variant d'histone H2A.Z dans la création de sites pauvres en nucléosomes permettant l'activation d'enhancers *de novo* (Figure 13).

a – Facteurs pionniers

Par définition, les facteurs pionniers (FP) correspondent à des FT exprimés spécifiquement dans certains types cellulaires et capables de reconnaître et d'accéder à leurs sites de liaison présents au sein de chromatine condensée. Cette particularité a notamment été observée *in vitro* pour les facteurs FoxA1 (« Forkhead box 1 », aussi connu sous le nom de HNF3 α) et GATA-4, deux FT connus pour leur rôle dans l'induction et le développement du foie (Cirillo *et al.*, 2002). Le qualificatif de pionnier provient du fait que ces protéines sont présentes au niveau des enhancers avant le recrutement d'autres facteurs et avant l'activation de ces mêmes enhancers (Cirillo et Zaret, 1999). Ils ont pour fonction d'établir un état chromatinien permissif assurant l'accessibilité des régions régulatrices à d'autres FT. En effet, la liaison de FoxA1 ou de GATA-4 induit une ouverture locale de la chromatine indépendamment de l'implication de complexes de remodelage (Cirillo *et al.*, 2002). FoxA1 possède une structure particulière qui induirait localement des perturbations entre les nucléosomes : en plus de son domaine de liaison à l'ADN, ce FP possède un domaine C-terminal capable de se lier au cœur des histones (Cirillo *et al.*, 2002). Outre la perturbation directe de la structure chromatinienne, certains pionniers comme AP-1 ou PAX-5 recrutent des complexes de remodelage pour induire une réorganisation des nucléosomes ou conserver une structure chromatinienne ouverte (Biddie *et al.*, 2011 ; McManus *et al.*, 2011). De plus, en lien avec leur fonction de pionniers, FoxA1, PU.1 et Pax5 sont capables d'induire

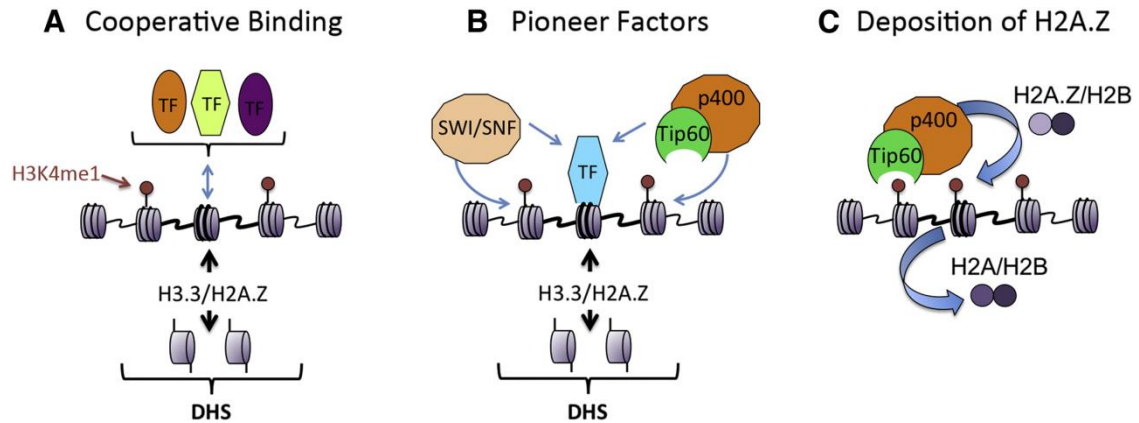


Figure 13. Modèles de sélection des enhancers (Calo et Wysocka, 2013).

(A) La liaison coopérative des facteurs de transcription peut passer outre la barrière des nucléosomes et initier des événements de régulation au niveau de la chromatine. En particulier la monométhylation de l'H3K4 et l'incorporation de variants d'histones, créant des régions accessibles à la digestion par la DNase1 (DHS). (B) Les régions enhancers peuvent devenir accessibles par l'action de facteurs pionniers, qui s'associent directement à l'ADN du nucléosome, précédant et généralement permettant le recrutement d'autres facteurs de transcription via le recrutement de protéines possédant des activités de remodelage de la chromatine. (C) L'incorporation du variant d'histone H2A.Z, par le complexe TIP60/p400, crée une région dont les nucléosomes sont hypermobiles, ce qui pourrait faciliter la liaison de facteurs de transcription.

indirectement la mise en place de modifications des histones telles que la mono- ou la diméthylation d'H3K4 caractéristiques des enhancers (Ghisletti *et al.*, 2010 ; Heinz *et al.*, 2010 ; McManus *et al.*, 2011 ; Sérandour *et al.*, 2011). Cependant, une autre étude concernant FoxA1 a démontré que la fixation de ce FP requiert la présence préalable d'H3K4me2 afin d'établir un programme de régulation spécifique d'un type cellulaire (Lupien *et al.*, 2008), ce qui induit un doute sur la nature pionnière de FoxA1 au niveau de ces sites, si on s'en tient à la définition première d'un facteur pionnier.

Le recrutement de FP au niveau de régions régulatrices du génome permet ainsi l'initiation de la sélection des enhancers afin de contrôler le destin cellulaire d'un tissu. Les FP jouent donc un rôle crucial dans de nombreux phénomènes cellulaires. FoxA1 et les facteurs GATA régulent notamment l'induction hormonale au sein des cellules différenciées exprimant le récepteur aux œstrogènes (ER – Carroll *et al.*, 2005 ; Eeckhoutte *et al.*, 2007 ; Hurtado *et al.*, 2011) ou aux androgènes (AR – Gao *et al.*, 2003 ; Wang *et al.*, 2007), tandis que AP-1 participe à la régulation de la transcription par le récepteur aux glucocorticoïdes (GR – Biddie *et al.*, 2011). Dans des cellules cancéreuses mammaires, la protéine PBX1 (« pre-B-cell leukemia homeobox 1 »), incluse dans la famille de FT de type TALE-HD (pour « three amino-acid loop extension homéodomain »), favorise la progression cancéreuse en modulant les sites de recrutement d'ER (Magnani *et al.*, 2011). Le FT hématopoïétique PU.1, faisant partie des FP les plus étudiés, est impliqué dans la régulation de l'engagement des cellules souches hématopoïétiques vers divers lignages comme les macrophages, les neutrophiles et les lymphocytes B et T (Scott *et al.*, 1994). Enfin, les membres de la famille C/EBP (« CCAAT/enhancer-binding protein ») agissent comme pionnier du PPAR γ lors de l'adipogénèse (Sierbaek *et al.*, 2011). C/EBP α agit également en tant que facteur pionnier de PU.1 afin d'activer des enhancers *de novo* pour des cellules myéloïdes (van Oevelen *et al.*, 2015).

b – Liaison de facteurs additionnels

L'achèvement de la sélection des enhancers *de novo* s'effectue notamment grâce au recrutement de FT spécifiques d'un type cellulaire et/ou dont l'expression dépend d'un signal reçu par la cellule. En effet, les sites de reconnaissance pour les FP sont souvent adjacents à des séquences de liaison pour d'autres FT dont l'expression est dépendante du type cellulaire. Une coopérativité de liaison entre ces facteurs et les FP au niveau des enhancers permettrait

de conserver et d'étendre la région de chromatine ouverte (Boyes & Felsenfeld, 1996). Elle permettrait surtout d'entériner le caractère cellule- ou tissu-spécifique de l'enhancer sélectionné et d'assurer une régulation adaptée de l'expression des gènes associés (Heinz *et al.*, 2010). En effet, le FP PU.1 est associé à des facteurs spécifiques des macrophages dans les macrophages, et à des facteurs spécifiques des cellules B dans les cellules B (Heinz *et al.*, 2010). Le recrutement de ces facteurs au niveau des enhancers permet par la suite de guider de façon appropriée la fixation de FT dont l'expression est induite par un signal ou un stimulus (Heinz *et al.*, 2013). Par exemple, la protéine Smad3 qui est impliquée dans la signalisation TGF- β (« Transforming growth factor beta ») est associée à différents facteurs spécifiques d'un lignage en fonction des cellules dans lesquelles son action est requise : elle est ainsi retrouvée associée à Oct4 dans les cellules ES, à Myod1 dans les myotubes et à PU.1 dans les cellules pro-B (Mullen *et al.*, 2011). En outre, deux études récentes ont montré que l'environnement tissulaire est un facteur capable de diriger la sélection des enhancers vers les différentes populations de macrophages (Gosselin *et al.*, 2014 ; Lavin *et al.*, 2014).

De façon intéressante, l'expression des gènes ubiquitaires est également régulée par des enhancers spécifiques du type cellulaire, liés par des facteurs spécifiquement exprimés dans le lignage correspondant, donnant lieu à changement d'enhancers lors de l'engagement vers un autre lignage (Kieffer-kwon *et al.*, 2013 ; Zhang *et al.*, 2013). Cette « course de relais » proposée par les auteurs permettrait de réguler finement l'expression de ces protéines constitutivement exprimées et de les mettre sous le contrôle de facteurs spécifiques du type cellulaire pour lequel elles sont exprimées (Kieffer-kwon *et al.*, 2013).

c – Instabilité des nucléosomes : implication du variant H2A.Z

Longtemps focalisées à comprendre le rôle des FT, les recherches concernant la sélection des enhancers tendent à s'orienter vers l'implication de la structure chromatinienne dans ces événements. Il semblerait que plutôt que d'être une conséquence de l'enchaînement des processus menant à un destin cellulaire, celle-ci constituerait un déterminant majeur de la sélection des enhancers. En effet, les cellules ES possèdent une chromatine « un peu plus relâchée » que d'autres types cellulaires, générée notamment par une importante dynamique de fixation de facteurs structuraux, tels que HP1, l'histone linker H1 ou les histones H2B et H3, assurant la plasticité de ces cellules pluripotentes (Meshorer *et al.*, 2006). Un changement local dans l'organisation des nucléosomes est notamment observé au niveau des enhancers

lors de la différenciation de ces cellules (West *et al.*, 2014). Comme décrit précédemment, l'incorporation des variants d'histones H2A.Z et H3.3 induit la formation de nucléosomes instables (Jin et Felsenfeld, 2007). Le variant H2A.Z est détecté de façon ubiquitaire à l'échelle du génome et est associé aux différentes méthylation d'H3K4 au niveau d'éléments distaux (Ku *et al.*, 2012). Hu et collègues ont notamment établi un lien entre la présence d'H2A.Z et la fixation de FT dans les cellules ES murines (Hu *et al.* 2013). En effet, H2A.Z est essentiel au recrutement d'Oct4 et des complexes de méthylation des histones MLL et PRC2 suggérant que ce variant d'histone constituerait un élément antérieur à la modification des histones (Hu *et al.* 2013). De plus, lors de la stimulation œstrogénique, la présence d'H2A.Z induit le recrutement de l'ARN Pol II et la synthèse d'eRNA, abolis en son absence, révélant H2A.Z comme un élément régulateur de l'activité des enhancers (Brunelle *et al.*, 2015). Ces différentes études soutiennent l'idée que le dépôt d'H2A.Z constitue un élément précoce de la sélection des enhancers et l'on peut se demander au final, qui de H2A.Z, des modifications d'histones ou des FP assure en premier la sélection des enhancers ? Ces trois éléments semblent déterminants et il est somme toute possible qu'en fonction des conditions un enchainement soit favorisé plutôt qu'un autre, qui le sera à son tour à un autre moment ou dans une autre voie de régulation. Des études supplémentaires sont donc nécessaires afin de répondre à cette question.

2 – Enhancers pré-marqués dans les cellules indifférenciées

Plusieurs études ont constaté la présence d'un certain nombre d'enhancers dits pré-marqués dans les cellules indifférenciées car portant les marques de chromatine H3K4me1/2, étant liés par des FT et proches de gènes exprimés dans les phases ultérieures de différenciation.

En effet, dans les cellules ES, la liaison de FT sur certains enhancers régulant l'expression de gènes inactifs dans des cellules pluripotentes, est nécessaire afin d'assurer leur préparation et ainsi leur activation lors de la différenciation. Ainsi, le facteur de pluripotence Sox2 a été observé comme se liant à un enhancer spécifique des cellules pro-B (Liber *et al.*, 2010) ou à un enhancer régulant la lignée neurale (Bergsland *et al.*, 2011) dans les cellules ES. Dans les deux cas, la liaison de Sox2 est associée à une modification épigénétique locale (di- ou tri-méthylation d'H3K4). De façon intéressante, l'entrée dans un programme de différenciation induit un relai dans la liaison des facteurs de transcription. Sox2 est ainsi

remplacé par Sox4 dans les cellules pro-B (Liber *et al.*, 2010) ou par Sox3 dans les progéniteurs neuraux puis Sox11 lors de la différenciation en neurones (Bergsland *et al.*, 2011). Dans les cellules ES, la présence de Sox2 sur un enhancer spécifique des cellules pro-B est associée à celle de Foxd3, jouant un rôle répresseur sur la transcription afin d'éviter une expression inappropriée du gène régulé par cet enhancer (Liber *et al.*, 2010). La nécessité de conserver la séquence du site de liaison Fox intacte pour l'activation de l'enhancer suggère que la fixation de Foxd3, perdue lors de la différenciation, pourrait être relayée par un autre facteur de la famille Forkhead (Liber *et al.*, 2010). Ceci a par exemple été démontré lors de la différenciation des cellules ES en fibroblastes embryonnaires de Souris où Foxd3 peut être remplacé par Foxa1 (Xu *et al.*, 2009). Dans les cellules ES, plusieurs enhancers régulant des gènes spécifiques de tissus montrent un appauvrissement en CpG méthylés et la présence de Foxd3 est nécessaire au maintien de cette hypométhylation au niveau de l'enhancer du gène *Alb1* spécifiquement exprimé dans le foie, assurant la mise en place d'une chromatine permissive (Xu *et al.*, 2007, 2009).

Deux autres études ont plus récemment montré la présence d'enhancers pré-marqués hors cellules ES. Samstein et collègues ont tout d'abord montré que Foxp3, qui contrôle la différenciation et la fonction des cellules T régulatrices, n'est en fait pas responsable de la sélection d'enhancers *de novo*, mais se fixe sur des enhancers pré-accessibles dans les cellules précurseurs (Samstein *et al.*, 2012). D'autre part, PU.1 marque des enhancers pré-existants dans les précurseurs myéloïdes ou les cellules B qui seront ensuite hyper-activés par la liaison de C/EBP α lors de la différenciation ou de la transdifférenciation en macrophage, respectivement (van Oevelen *et al.*, 2015).

L'existence d'enhancers pré-marqués dans les cellules ES ou dans certains précurseurs spécifiques d'un lignage révèle l'établissement précoce, bien en amont de l'activation des enhancers, d'un programme de spécification des enhancers à l'échelle du génome.

3 – Enhancers latents

Un troisième mode de sélection des enhancers concerne les enhancers dits « latents ». Ces derniers ne sont pas sélectionnés dans des cellules différenciées (absence de modifications des histones caractéristiques des enhancers et de liaison de FT) mais le deviennent suite à un stimulus (Ostuni *et al.*, 2013). Ainsi, la stimulation de macrophages entraîne le recrutement collaboratif (non séquentiel, par opposition à la sélection des

enhancers *de novo*) des FT dont l'action dépend des signaux et de FT spécifiques au type cellulaire (Kaikkonen *et al.*, 2013 ; Ostuni *et al.*, 2013). De façon intéressante, la mise en place des marques H3K4me1/2 sur ces enhancers nécessite la mise en place du processus de transcription de l'enhancer (et non l'eRNA en lui-même – Kaikkonen *et al.*, 2013). Après stimulation, ces enhancers ne retrouvent pas leur précédent état de latence mais persistent afin d'assurer une réponse rapide au prochain événement de stimulation (Ostuni *et al.*, 2013). Les enhancers latents permettent donc, suite à un stimulus, d'étendre le nombre de régulateurs tout en maintenant une régulation spécifique du type cellulaire.

C – Interactions avec les promoteurs

La fonction régulatrice des enhancers sur la transcription ne peut s'exercer que si ces régions interagissent avec les promoteurs des gènes qu'ils régulent. La distance, souvent importante entre les enhancers et les promoteurs a amené les chercheurs à soulever plusieurs hypothèses concernant le ou les processus menant à un rapprochement physique de ces deux entités. Parmi les différentes propositions, outre le mécanisme de « balayage » pour lequel, après fixation au niveau des enhancers, les FT et autres protéines régulatrices parcourent l'ADN jusqu'à rencontrer leur promoteur cible (Blackwood et Kadonaga, 1998), c'est le modèle de la formation d'une boucle entre l'enhancer et le promoteur qui semble le plus pertinent. En effet, le développement de techniques permettant l'étude des interactions longues distances permet aujourd'hui de valider cette hypothèse.

1 – Détection des interactions longues distances et données résultantes

La confirmation et l'observation des interactions entre différentes régions régulatrices à l'échelle du génome a été obtenue grâce aux techniques de « chromosome conformation capture » (Miele et Dekker, 2009) et de ChIA-PET (pour « chromatin interaction analysis by paired-end tag sequencing » – Fullwood *et al.*, 2009). Bien que n'indiquant pas la fonctionnalité des interactions observées, l'utilisation de techniques supplémentaires (telles que le ChIP pour des facteurs et marques d'activités, ou les gènes rapporteurs) a confirmé l'importance des interactions entre promoteurs et enhancers dans la régulation de la mise en place et de l'activité de programmes transcriptionnels spécifiques (Chepelev *et al.*, 2012 ;

Zhang *et al.*, 2013). Le développement de ces techniques et leur application dans différentes lignées cellulaires et organismes a permis d'obtenir une quantité importante de données qui permettent de préciser certaines régulations spécifiques et leurs dynamiques ainsi que de déceler de nouvelles caractéristiques des contacts entre promoteurs et enhancers (Jin *et al.*, 2013 ; Kieffer-Kwon *et al.*, 2013 ; Sanyal *et al.*, 2012). Certaines des études réalisées dans différentes lignées humaines et murines ont d'ores et déjà montré qu'environ 90 % des enhancers potentiels n'interagissent qu'avec un seul promoteur, et que 2 à 9 % d'entre eux interagissent avec de multiples promoteurs (Chepelev *et al.*, 2012 Kieffer-Kwon *et al.*, 2013 ; Sanyal *et al.*, 2012, Zhnag *et al.*, 2013). Montavon et collègues ont notamment décrit la nécessité d'avoir l'interaction d'un nombre précis de régions régulatrices (appelé « archipelago ») entre elles et avec le gène *Hoxd13* pour obtenir une formation correcte des doigts chez la Souris (Montavon *et al.*, 2011). 25 % des promoteurs des cellules T humaines contactent au minimum deux enhancers mais l'expression génique ne corrèle pas toujours avec le nombre d'enhancers liés (Chepelev *et al.*, 2012), ce qui concorde avec l'observation de nombreux enhancers redondants impliqués dans le contrôle de l'expression dorso-ventrale ou lors de la formation du mésendoderme chez la Drosophile (Cannavo *et al.*, 2016 ; Hong *et al.*, 2008). Les interactions promoteurs-promoteurs concernent des gènes co-exprimés, ce qui est également le cas des gènes qui contactent le même enhancer (Chepelev *et al.*, 2012 ; Kieffer-Kwon *et al.*, 2013). De plus, la détermination des interactions longues distances a également permis d'observer que la plupart des enhancers n'interagissent pas avec le promoteur le plus proche et que de façon générale ces interactions engendrent des réseaux très complexes (Kieffer-Kwon *et al.*, 2013 ; Quintin *et al.*, 2014 ; Sanyal *et al.*, 2012 ; Zhang *et al.*, 2013). Enfin, une interaction entre les enhancers de réponse au TNF- α (« tumor necrosis factor ») et leurs promoteurs cibles est observée antérieurement à la signalisation par le TNF- α . Ceci suggère que la structuration du génome dans l'espace n'est pas nécessairement dépendante de la liaison de FT mais qu'elle jouerait plutôt un rôle dans la sélection des enhancers et l'activation des gènes (Jin *et al.*, 2013).

2 – Facteurs impliqués dans les interactions entre promoteurs et enhancers

Les mécanismes de mise en place et de maintien des interactions entre les enhancers et les gènes ne sont pas encore entièrement révélés. Cependant, l'implication de certains facteurs comme CTCF ou le complexe cohésine a été mise en évidence dans ces processus.

Précédemment introduites dans la première partie de cette introduction pour leur enrichissement au niveau des régions insulatrices et aux frontières des TAD, les protéines CTCF, par la formation de dimères, sont également impliquées dans la formation de boucles entre les enhancers et les promoteurs (Handoko *et al.*, 2011 ; Splinter *et al.*, 2006). Il est intéressant de noter que l'orientation même des sites de liaison CTCF est un critère déterminant pour la formation de boucles (de Wit *et al.*, 2015 ; Guo *et al.*, 2015 ; Rao *et al.*, 2014). L'inversion des sites CTCF, qui assurent leur fonction sous forme de sites convergents, conduit à la déstabilisation des interactions longues-distance existantes et à leur réorganisation, entraînant de ce fait des changements transcriptionnels (de Wit *et al.*, 2015 ; Guo *et al.*, 2015).

Le complexe cohésine, tout d'abord connu pour son rôle dans le maintien de la cohésion des chromatides sœurs, participe également à la stabilisation des interactions longues distances entre promoteurs et enhancers (Schaaf *et al.*, 2013). Ce rôle a notamment été mis en évidence par le fait que la cohésine co-localise avec CTCF à l'échelle du génome et que ce recrutement dépend de CTCF (Degner *et al.*, 2011 ; Parelho *et al.*, 2008 ; Rubio *et al.*, 2008 ; Wendt *et al.*, 2008). D'autres études ont ensuite démontré l'importance de ces deux facteurs dans l'établissement de boucles, participant ainsi à l'organisation chromatinienne et à la régulation transcriptionnelle spécifique à chaque lignée cellulaire (Hou *et al.*, 2010 ; Mishiro *et al.*, 2009). La cohésine participe également à ces interactions longues distances sur des sites dépourvus de CTCF, en interagissant physiquement et fonctionnellement avec le complexe Médiateur dans les cellules ES (Kagey *et al.*, 2010).

L'établissement d'un niveau supérieur d'organisation de la chromatine au niveau du LCR régulant l'expression du gène codant pour la β -globine nécessite la présence du FT EKLF (« Erythroid Krüppel-like factor » – Drissen *et al.*, 2004) tandis que la formation d'une boucle entre le LCR et le promoteur de ce même gène requiert les facteurs GATA-1 et FOG1 (Vakoc *et al.*, 2005). La fréquence des interactions entre le LCR du T helper type 2 (T(H)2) et les gènes codant pour des cytokines dépendent quant à elles des facteurs GATA3 et STAT6 (Spilianakis et Flavell, 2004). Ces différentes études montrent l'implication des FT dans l'organisation tridimensionnelle de la régulation de la transcription, notamment par la stabilisation des interactions entre promoteurs et enhancers.

Comme discuté dans la partie III-A-3-Présence de l'ARN Pol II et transcription aux enhancers, la participation des eRNA à l'établissement des boucles de chromatine nécessite encore quelques investigations avant d'être validée et généralisée.

D – Super-enhancers

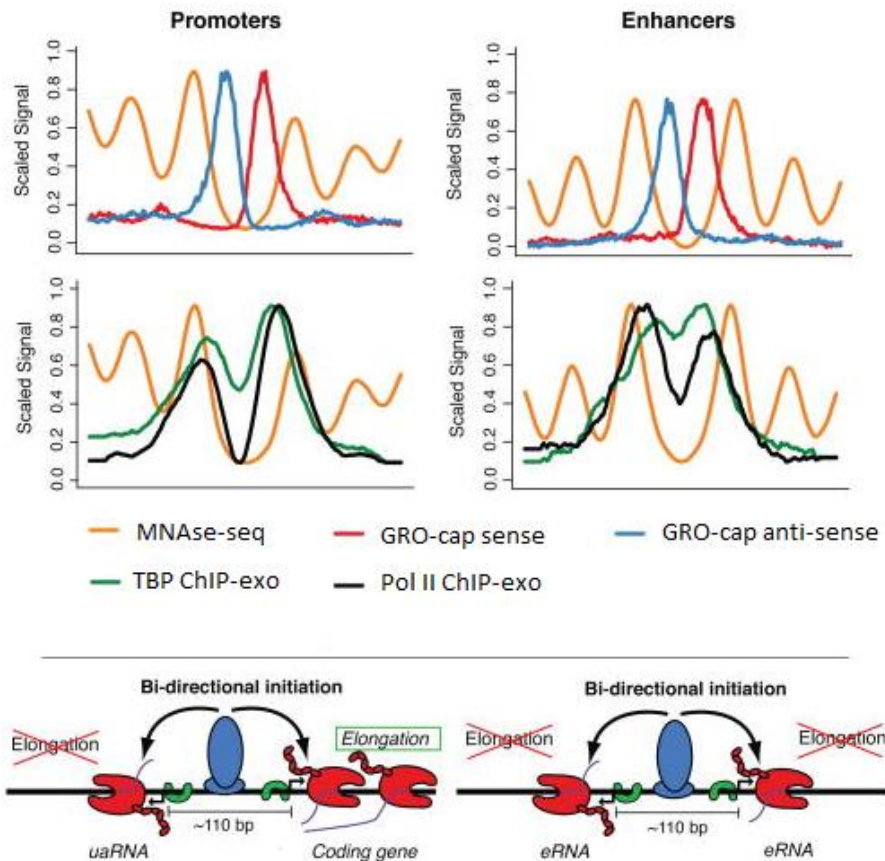
Il semble ici important de mentionner un sujet, de prime abord inédit, mais qui finalement ne semble pas être si révolutionnaire et provoque le débat : les super-enhancers. Le terme fut introduit en 2013 par deux études indépendantes. Dans la première étude, les auteurs étudient l'importance du complexe Médiateur et des « Masters » FT (OCT4, SOX2 et NANOG) recrutés sur les enhancers des ESC de Souris (Whyte *et al.*, 2013), tandis que la seconde porte sur l'effet spécifique de l'inhibition de certains gènes en réponse à l'inhibition d'un facteur ubiquitaire (BRD4) dans des cellules de myélome (Lovén *et al.*, 2013). Ces deux investigations ont mené à la description et à la définition d'un nouveau type d'élément régulateur du génome, les super-enhancers. Correspondant à environ 3% des enhancers totaux, ils sont décrits comme des clusters d'enhancers fortement enrichis en « Masters » FT et en Médiateur (les super-enhancers lient environ 40% des complexes Médiateur se trouvant sur l'ensemble des enhancers). Les super-enhancers, comparés aux enhancers canoniques, sont plus larges, ont une forte densité de liaison de FT essentiels au destin cellulaire et ont un niveau plus élevé d'H3K27ac (Lovén *et al.*, 2013 ; Whyte *et al.*, 2013). Les gènes qui leurs sont associés sont en général plus fortement exprimés que les gènes associés aux enhancers, leur expression corrèle davantage avec des gènes spécifiques d'un type cellulaire et est notamment plus sensible aux variations de concentrations de FT (Lovén *et al.*, 2013 ; Whyte *et al.*, 2013). Ces régions régulatrices, dont l'implication semble essentielle dans le maintien de l'identité cellulaire, sont également retrouvées dans les cellules pro-B, les myotubes, les cellules T helper, les macrophages murins (Whyte *et al.*, 2013), ainsi que dans plusieurs sortes de tumeurs pour lesquelles elles sont associées à des oncogènes (Lovén *et al.*, 2013). Une étude supplémentaire réalisée sur 86 échantillons de tissus et cellules humains révèle l'importance des régions super-enhancers dans la régulation génique dans le cas de certaines pathologies (Hnisz *et al.*, 2013). En effet, l'enrichissement en SNP (« single nucleotide polymorphism ») associé à des pathologies est significativement plus important, dans les cellules atteintes, au niveau des super-enhancers que dans les enhancers « canoniques ». De plus, les super-enhancers acquis par les cellules cancéreuses lors de la tumorigénèse sont

notablement associés à des oncogènes ou des gènes caractéristiques des cancers (Hnisz *et al.*, 2013).

Par ces différents exemples probants, ces trois études montrent l'importance de ces nouvelles régions régulatrices (sans pourtant réaliser d'études fonctionnelles) pour la régulation transcriptionnelle des gènes impliqués dans le maintien de l'identité cellulaire, proposant également de nouvelles pistes pour la compréhension de différentes pathologies. Cependant, plusieurs points sont sources de débat. Parmi ceux-ci, discutés par Pott et Lieb en 2015 dans une perspective pour le journal *Nature Genetics*, se pose le problème de la notion de « nouveauté » car en effet, les super-enhancers ressemblent à des structures préalablement décrites comme les LCR, les enhancers « stretch », les CORE (pour « clusters of open regulatory regions » - Pott et Lieb, 2015) ou encore les « archipelago » (Montavon *et al.*, 2011). Effectivement, certains de ces super-enhancers chevauchent des LCR et d'autres domaines préalablement cartographiés (Hnisz *et al.*, 2013). S'ensuit la difficulté de concilier les différents protocoles de détection des super-enhancers (natures des FT, seuils...), menant à une définition bien trop flexible de ces régions (Pott et Lieb, 2015). L'importance des données biologiques fournies par ces études (Hnisz *et al.*, 2013 ; Lovén *et al.*, 2013 ; Whyte *et al.*, 2013) n'est pour autant pas remise en cause même si elles nécessitent davantage d'expériences fonctionnelles. Néanmoins, il semble désormais important de définir une fois pour toute le terme de super-enhancer et de ne l'utiliser que pour des cas réellement appropriés.

E – Enhancers et promoteurs: sont-ils si différents ?

Initialement présentés dans la première partie de cette introduction concernant les régions régulatrices (I-B-1-Séquences régulatrices), les promoteurs et les enhancers sont depuis très longtemps définis comme deux entités distinctes. Nous avons ici déjà regroupé les notions de promoteurs proximaux et d'enhancers selon leur fonction, qui est de moduler les niveaux de transcription par le biais de liaison de FT, ne les différenciant que par leur distance au TSS. La comparaison des données acquises grâce au développement des techniques d'étude à l'échelle du génome a ainsi révélée l'existence de nombreux points communs entre les promoteurs dans leur totalité (basal + proximal) et les enhancers. Ces nouveaux éléments impliquent donc une révision de notre vision de ces séquences régulatrices.



Un paramètre essentiel de la similitude entre promoteurs et enhancers provient de la découverte de la présence de l'ARN Pol II et des GTFs associés à un processus de transcription au niveau des enhancers (De Santa *et al.*, 2010 ; Kim *et al.*, 2010 ; Koch *et al.*, 2011). De façon intéressante, les caractéristiques des eRNA correspondent à celle des ARN anti-sens synthétisés aux promoteurs (Andersson *et al.*, 2014 ; Core *et al.*, 2008). Au niveau chromatinien, les enhancers et promoteurs sont tous deux caractérisés par (i) une région symétrique faiblement enrichie en nucléosome encadrée par deux nucléosomes stables, (ii) la présence d'H3K27ac dans leur état actif, (iii) la présence d'H3K27me3 dans leur état « poised » caractéristique d'un domaine bivalent H3K4me1/H3K27me3 pour les enhancers *versus* H3K4me3/H3K27me3 pour les promoteurs (Bernstein *et al.*, 2006 ; Core *et al.*, 2014 ; Ernst *et al.*, 2011, Heintzman *et al.*, 2009 ; Rada-Iglesias *et al.*, 2011, Zentner *et al.*, 2011 – Figure 14). C'est justement le ratio H3K4me1/H3K4me3 qui distingue les enhancers des promoteurs au niveau chromatinien, *i.e.* fort et faible respectivement. Cependant plusieurs études ont observé la présence de niveaux élevés d'H3K4me3 aux enhancers positivement corrélés à l'activité transcriptionnelle (Core *et al.*, 2014 ; Koch *et al.*, 2011 ; Pekowska *et al.*, 2011). Il semblerait donc que, plutôt que de discriminer des fonctions différentes, le ratio H3K4me1/H3K4me3 permettrait de discerner des niveaux d'activité transcriptionnelle qui en général, mais pas constamment, sont plus faibles aux enhancers d'où leur ratio H3K4me1/H3K4me3 souvent plus élevé (Core *et al.*, 2014).

De leurs côtés, un certain nombre de promoteurs impliqués dans des interactions promoteur-promoteur peuvent influencer l'activité des promoteurs avec lesquels ils interagissent rappelant la fonction originelle des enhancers (Li *et al.*, 2012b). De plus, une étude récente a mis en évidence, dans les macrophages, la présence d'un groupe de TSS anti-sens comportant une signature chromatinienne semblable aux enhancers : présence d'H3K4me1, H3K27ac, et recrutement de P300 et PU.1 (Scruggs *et al.*, 2015).

Ces premiers arguments appuient le fait que les enhancers et les promoteurs possèdent des caractéristiques structurales et fonctionnelles communes (Figure 14 ; Core *et al.*, 2014), leurs différences résultant principalement de leur niveau d'activité et de la nature des FT qu'ils recrutent. Les données présentées ici ainsi que des arguments supplémentaires sont plus amplement discutés par Andersson *et al.* et par Kim et Shiekhattar dans deux publications récentes (Andersson *et al.*, 2015 ; Kim et Shiekhattar, 2015).

IV INTEGRATION DES TRAVAUX DE THESE DANS LE CONTEXTE SCIENTIFIQUE

L'étude des mécanismes régissant la régulation de la transcription et permettant ainsi d'assurer le développement et la différenciation cellulaire profite désormais des nombreuses avancées techniques réalisées au cours de ces dernières années. En particulier, le développement des techniques de séquençage de nouvelle génération a joué un rôle primordial dans la récente progression des connaissances dans ce domaine.

De nombreuses études ont ainsi tiré profit de la cartographie à l'échelle du génome des modifications post-traductionnelles des histones en caractérisant les éléments clés intervenant dans la régulation transcriptionnelle, *i.e.* les promoteurs, les enhancers/silencers et les insulateurs. L'utilisation et la comparaison d'échantillons provenant de différents stades de développement ou de différenciation ont permis d'observer que contrairement aux régions promotrices et insulatrices, les enhancers sont des séquences régulatrices spécifiques d'un type cellulaire ou d'un tissu. Ces régions ont donc un rôle crucial dans l'établissement d'un répertoire transcriptionnel particulier. L'étude de leur sélection ainsi que de leur activation a dévoilé l'aspect dynamique de ces processus en association avec (i) la variabilité des modifications d'histones en fonction du type cellulaire, permettant de définir l'état d'activation de ces régions (Hawkins *et al.*, 2011 ; Heintzman *et al.*, 2009) ; (ii) le recrutement des différentes protéines telles que les facteurs pionniers, les co-régulateurs ou encore l'ARN Pol II (Bonn *et al.*, 2012 ; Heinz *et al.*, 2015) ; et (iii) le placement des nucléosomes (He *et al.* 2010 ; West *et al.*, 2014). Dans cette perspective, la découverte probablement la plus marquante de ces dernières années concerne la méthylation de l'ADN, préalablement considérée comme une marque stable. En effet, des études comparant le méthylome des cellules ES pluripotentes avec des cellules ou tissus différenciés ont mis en évidence des régions différentiellement méthylées correspondant à des régions régulatrices distantes impliquées dans le contrôle de l'expression de gènes associés au développement et à la différenciation (Meissner *et al.*, 2008 ; Song *et al.*, 2008). De plus, le fait que les régions régulatrices actives soient majoritairement hypométhylées (Lister *et al.*, 2009 ; Stadler *et al.*, 2011) appuie les données montrant que les dynamiques de méthylation/déméthylation observées aux enhancers sont corrélées à leur activation (Schmidl *et al.*, 2009 ; Sérandour *et al.*, 2011).

Un élément clé de la compréhension des mécanismes impliqués dans la dynamique de méthylation fut la (re-)découverte de la 5hmC et des enzymes TETs (Kriaucionis et Heinz, 2009 ; Tahiliani *et al.*, 2009). En effet, par oxydations itératives de la 5mC en 5hmC puis en 5fC et/ou en 5caC, les enzymes TETs associées à des glycosylases sont les acteurs majeurs de la déméthylation active des cytosines (He *et al.*, 2011 ; Ito *et al.*, 2011 ; Maiti et Drohat, 2011). Etonnamment, selon les tissus, les quantités de 5hmC détectées sont des dizaines voire des centaines de fois supérieures à celles des autres dérivés oxydés (Globisch *et al.*, 2010 ; Ito *et al.*, 2011). De plus, cette modification est spécifiquement enrichie au niveau des corps de gènes et des enhancers (Stroud *et al.*, 2011). Ces deux résultats suggèrent ainsi qu'hormis son rôle d'intermédiaire de déméthylation, la 5hmC pourrait être une modification relativement stable impliquée dans la régulation génique. En ce sens, de précédentes études réalisées au laboratoire ont établi l'existence de régions différenciellement hydroxyméthylées lors de la différenciation de cellules de carcinomes embryonnaires de Souris en progéniteurs neuraux (cellules P19 – Sérandour *et al.*, 2012). En accord avec l'hypothèse d'un rôle de la déméthylation et plus particulièrement de l'hydroxyméthylation dans la régulation spécifique exercée par les régions distales, les séquences enrichies en 5hmC lors de cette étude correspondent à des enhancers activés au cours de la différenciation des P19 et associés à des gènes spécifiquement exprimés dans les tissus neuraux. De même, lors de la différenciation adipocytaire des cellules 3T3-L1, la 5hmC est spécifiquement enrichie au niveau d'enhancers associés à des gènes exprimés dans les adipocytes (Sérandour *et al.*, 2012).

L'activation des enhancers étant liée au recrutement de FT, il est également intéressant de noter que plusieurs études observent une association entre la présence des FT et les modifications différentielles des CpG. En outre, Stadler et collègues proposent que l'acquisition de régions peu méthylées dépend de la fixation des FT (Stadler *et al.*, 2011). Ceci concorde avec des résultats précédemment obtenus au laboratoire montrant que le recrutement du FP FOXA1 est associé à une déméthylation des enhancers lors de la différenciation des P19 (Sérandour *et al.*, 2011). Un autre FP, PU.1, est également associé à des changements des niveaux de méthylation lors de la différenciation des monocytes en ostéoclastes (de la Rica *et al.*, 2013). De plus, en accord avec la présence de la 5hmC au niveau de sites de liaisons de FT (Ficz *et al.*, 2011 ; Stroud *et al.*, 2011), les régions qui deviennent hydroxyméthylées lors de la différenciation des cellules P19 montrent un enrichissement particulier pour des motifs de liaison de protéines de types TALE-HD :

MEIS1 et PBX1 (Sérandour *et al.*, 2012). Ces deux protéines ont un rôle crucial dans le développement et la différenciation neurale et certaines études ont proposé que ces protéines pourraient agir en tant que FP (Heidt *et al.*, 2007; Magnani *et al.*, 2011 ; Qin *et al.*, 2004 ; Yamada *et al.*, 2013).

Ces différents travaux suggèrent donc que, au même titre que les modifications d'histones et la nature des nucléosomes présents aux enhancers, les modifications des cytosines participent à la sélection et à l'activation de ces régions régulatrices distantes. Cependant ces différentes observations reliant déméthylation, hydroxyméthylation des cytosines et FP restent corrélatives et ne permettent pas en tant que telles d'élucider la séquence d'évènements menant à l'activation des enhancers.

Dans ce contexte, mes travaux de thèse ont eu pour objectif de caractériser les relations existant entre les modifications des cytosines aux enhancers et la fonctionnalité de ces derniers au cours de la différenciation neurale des cellules P19. Des expériences réalisées en cinétique sur quelques enhancers choisis ont révélé que la présence de 5hmC précède la liaison de MEIS1 et même dans certains cas, l'activation des enhancers au cours de cette différenciation (Sérandour *et al.*, 2012). Ces travaux suggèrent ainsi que l'hydroxyméthylation des cytosines serait un événement précoce de l'activation des enhancers mis en place avant le recrutement des FT. Afin de tester cette hypothèse, nous avons dans un premier temps analysé l'implication des processus de méthylation et de déméthylation des cytosines sur la liaison des FP et la structure chromatinienne des enhancers lors de la différenciation neurale des cellules P19. Les résultats de cette étude font actuellement l'objet d'un premier article scientifique en cours de soumission. En parallèle, nous avons entrepris d'analyser à l'échelle du génome les dynamiques chromatinienne associées aux enhancers hydroxyméthylés au cours des premières 48 heures de différenciation des cellules P19. L'analyse bioinformatique des données issues de ces expériences de cinétique a nécessité la mise au point d'une nouvelle méthode de normalisation des données de séquençage, présentée ici sous forme d'un manuscrit en cours de préparation. Les travaux décrits ci-après ont permis de préciser les rôles respectifs de la 5mC et de la 5hmC dans la régulation de l'activité des enhancers.

RESULTATS | PARTIE 1

PARTIE I – IMPLICATION DES MODIFICATIONS DES CYTOSINES DANS LA STRUCTURATION DE LA CHROMATINE DES ENHANCERS AU COURS DE LA DIFFERENCIATION CELLULAIRE.

OBJECTIF

Dans le but d'établir un lien fonctionnel entre modifications des cytosines et activation des enhancers, ces travaux ont tout d'abord visé à examiner l'influence de ces modifications sur la liaison de FP dont les sites de liaisons sont spécifiquement enrichis au niveau d'enhancers hydroxyméthylés au cours de la différenciation des cellules P19 (Sérandour *et al.*, 2012). Nous avons ensuite étudié l'impact de la perturbation des niveaux de méthylation et d'hydroxyméthylation sur la structure chromatinienne lors de l'activation des enhancers, *via* l'utilisation d'inhibiteurs chimiques des ADN méthyltransférases et des TET. L'objectif des expériences décrites dans cette première partie était de déterminer si les processus de méthylation/déméthylation ont un rôle direct ou indirect dans les différentes étapes menant à l'activation d'enhancers liés par des FP.

RESULTATS

Les résultats des ces travaux sont exposés dans l'article ci-après intitulé « Cytosine modifications modulate the chromatin architecture of transcriptional enhancers », en cours de soumission.

Brièvement, cette étude a permis d'observer que les enhancers liés par des FP sont enrichis en 5hmC et que les modifications des cytosines favorisent la liaison des protéines TALE-HD *in vitro*. De plus, la méthylation et l'hydroxyméthylation de l'ADN semblent nécessaires pour assurer l'établissement d'un programme transcriptionnel associé à la différenciation neurale. En effet, l'utilisation d'inhibiteurs de DNMT ou de TET provoque un retard de l'expression des marqueurs de progéniteurs neuraux lors de la différenciation des cellules P19. Les conséquences de la perturbation des quantités de 5mC et/ou 5hmC par ces inhibiteurs ont été étudiées sur quelques enhancers liés par des FP. Les résultats obtenus ont permis de démontrer que (i) l'enlèvement de la 5mC est indispensable au « priming » des enhancers, indépendamment des signaux de différenciation, (ii) la présence de 5hmC est nécessaire à l'ouverture de la chromatine, ce qui facilite le recrutement ultérieur des FT, lui-même favorisé par la présence des modifications des cytosines, et induit l'activation des enhancers.

CONCLUSION

Les résultats obtenus lors de cette étude confirment l'importance des dynamiques de méthylation/déméthylation des cytosines au cours la différenciation neurale et révèlent l'implication à la fois de la méthylation et de l'hydroxyméthylation de l'ADN dans la modulation de la structure chromatinienne au cours de l'activation des enhancers.

CONTRIBUTIONS AUX DIFFERENTS RESULTATS

Contributions expérimentales

- Organisation et participation à la préparation des cultures cellulaires pour les expériences de cinétiques à l'échelle du génome.
- Expériences de MeDIP et hMeDIP (immunoprécipitation d'ADN méthylé ou hydroxyméthylé) en cinétique avant séquençage.
- En collaboration avec la technicienne du laboratoire : production et purification des protéines recombinantes HOXA9, PBX1 et MEIS1 ; mise au point du protocole de pull-down et réalisation des expériences ; mise au point et expériences de gel-retard non-radioactif.
- Mise au point des conditions de traitement des cellules P19 au DMOG (dimethyloxallylglycine).
- Pour les expériences en conditions DMOG : culture et préparation des cellules ; RT-qPCR ; MeDIP et hMeDIP ; ChIP-qPCR H3K4me1, H3K27me3, FOXA1 et MEIS1.
- Pour les expériences en conditions 5-azadC (5-aza-2'-déoxycytidine) : hMeDIP. Toutes les autres expériences en conditions 5-azadC ont été menées par un ingénieur du laboratoire.
- Participation aux analyses bioinformatiques.

Contributions pour l'élaboration de l'article

- Participation à l'analyse des données et à l'écriture de l'article.
- Réalisation de toutes les figures excepté la figure supplémentaire 12.

Cytosine modifications modulate the chromatin architecture of transcriptional enhancers

Elise A Mahé^{1,2}, Thierry Madigou^{1,2}, Aurélien A Sérandour³, Maud Bizot^{1,2}, Stéphane Avner^{1,2},
Frédéric Chalmel⁴, Gaëlle Palierne^{1,2}, Raphaël Métivier^{1,2} and Gilles Salbert^{1,2,§}

¹ CNRS UMR6290, Equipe SP@RTE, Institut de Génétique et Développement de Rennes,
Campus de Beaulieu, 35042 Rennes cedex, France

² Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France

³ EMBL, Meyerhofstrasse 1, Heidelberg 69117, Germany

⁴ Inserm U1085-Irset, Université de Rennes 1, F-35042 Rennes, France

[§] Corresponding author: gilles.salbert@univ-rennes1.fr

Running title: 5(h)mC roles at enhancers

Key words: 5mC, 5hmC, enhancer, FOXA1, MEIS1, PBX1

Abstract

The dynamics of cytosine modifications such as 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) have been correlated with activation of transcriptional enhancers. However, their functional relationship with the regulation of enhancer activity by pioneer transcription factors remains elusive. To gain insights into their function, we perturbed formation of cytosine modifications by inhibiting DNA methyltransferases and Ten Eleven Translocation proteins and interrogated chromatin characteristics of enhancers bound by the pioneer factors FOXA1, MEIS1 and PBX1 in mouse embryonal carcinoma cells. Results suggest that 5mC removal is required for enhancer priming whereas 5hmC is essential for chromatin opening allowing further recruitment of transcription factors.

Introduction

Cell differentiation relies on the engagement of short genomic regions called transcriptional enhancers and acting at a distance from their target genes. Enhancers are endowed with the capacity to bind numerous cell-specific and general transcription factors (TFs) in a temporal sequence which is partly dictated by the timely controlled expression of TFs and/or their regulated binding to chromatin through signaling pathways (Heinz et al., 2015; Spitz and Furlong, 2012). In their inactive state, enhancers can be targeted by so called “pioneer factors” (PFs) which bind condensed chromatin and favor transition of enhancers to a primed state facilitating binding of additional TFs and further activation (Iwafuchi-Doi and Zaret, 2014). Several TFs, including FOXA1, PU.1, PBX1, C/EBP β , AP-1 and MEIS1 have been qualified as PFs and play major roles in pluripotent cell differentiation during development and in adulthood (Biddie et al., 2011; Ghisletti et al., 2010; Heidt et al., 2007; Lupien et al., 2008; Magnani et al., 2011; Siersbæk et al., 2011). Hence, PF engagement during lineage

commitment is associated with changes in enhancer chromatin architecture. In agreement with this view, binding of PU.1 to enhancer chromatin during monocyte-to-osteoclast differentiation impacts cytosine methylation (5mC) at CpGs through the recruitment of either DNA methyltransferases (DNMTs) at active enhancers or Ten Eleven Translocation (TET) enzymes, which catalyze 5mC oxidation into 5-hydroxymethylcytosine (5hmC - Tahiliani et al., 2009), at inactive enhancers (de la Rica et al., 2013). 5hmC is an intermediate in a pathway of active DNA demethylation generating further oxidative derivatives (5-formylcytosine, 5fC; and 5-carboxylcytosine, 5caC) subsequently replaced by cytosines through action of the Base Excision Repair (BER) machinery (He et al., 2011; Maiti and Drohat, 2011). However, due to the relative stability of 5hmC and its ability to bind specific proteins (Bachman et al., 2014; Iurlaro et al., 2013; Spruijt et al., 2013), 5hmCpGs may act as a signaling module just as 5mCpGs do (Bird, 2011). In this respect, selected enhancers targeted by the three amino-acid loop extension-homeodomain (TALE-HD) proteins MEIS1 and PBX1 were found to be marked by high levels of 5hmC prior to TALE-HD binding during neural differentiation of mouse P19 embryonal carcinoma cells (ECCs) into neural progenitors (Sérandour et al., 2012), suggesting that DNA hydroxymethylation at enhancers could favor MEIS1 and PBX1 engagement. Moreover, consistent with data showing that TF binding to regulatory regions lowers their CpG methylation levels (Stadler et al., 2011), we previously observed that FOXA1 recruitment to enhancers is associated with DNA demethylation during P19 ECC differentiation (Sérandour et al., 2011). Nonetheless, it remains unclear which steps of enhancer priming and activation are regulated through DNA demethylation. In order to gain insights into the role of cytosine modifications at PF-bound enhancers, we analyzed the impact of DNA methylation/demethylation processes on the chromatin structure of FOXA1 and TALE-HD enhancers. Through chemical inhibition of

DNMTs and TETs, we show here that 5mC inhibits enhancer priming whereas 5hmC favors chromatin opening, a process which likely facilitates TALE-HD binding.

Results and Discussion

Pioneer factor-bound enhancers are enriched with 5hmC

Primed enhancers are characterized by the presence of histone H3 monomethylated on lysine 4 (H3K4me1) and the absence of acetylated H3K27 (H3K27ac), whereas active enhancers present both marks (reviewed in Calo and Wysocka, 2013). We previously observed that, during retinoic acid (RA)-induced differentiation of P19 ECCs into neural progenitors, differentially hydroxymethylated regions (DhMRs) are detected at enhancers either being silenced for down DhMRs or undergoing activation for up DhMRs (Sérandour et al., 2012). In order to obtain a dynamic view of these processes at the full sets of up and down DhMRs, we first analyzed the kinetics of H3K4me1 and H3K27ac histone modifications genome-wide by Chromatin Immunoprecipitation followed by high-throughput sequencing (ChIP-seq), together with those of 5hmC and 5mC by hydroxymethylated or methylated DNA Immunoprecipitation (hMeDIP-seq and MeDIP-seq respectively). In addition, chromatin opening was assessed by Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE - Giresi et al., 2007). For both types of DhMRs, the levels of H3K4me1 and chromatin opening followed the gradual changes observed for 5hmC, whereas 5mC levels showed only mild fluctuations over time (Fig. 1A for up DhMRs and Supplemental Fig. 1A for down DhMRs). Although changes in H3K27ac levels paralleled those of 5hmC at up DhMRs, only 12.5% of hydroxymethylated regions overlapped H3K27ac peaks, whereas 58.4% and 21.6% overlapped H3K4me1 and FAIRE peaks respectively (data not shown). These observations suggested a strong correlation between 5mC oxidation and enhancer priming during ECC

differentiation. We next mapped FOXA1, MEIS1 and PBX1, three PFs induced during RA-mediated differentiation of P19 cells and controlling their commitment to the neural lineage (Dong et al., 2012; Qin et al., 2004; Yamada et al., 2013). Consistent with an enrichment in PF binding motifs at enhancers gaining 5hmC during differentiation (Sérandour et al., 2012), PF binding peaked at up DhMRs (Fig. 1B) but not at down DhMRs (Supplemental Fig. 1B), and heatmaps clearly identified PF-bound regions being enriched in 5hmC (Fig. 1C). In addition, these three PFs shared a significant fraction of their binding sites (Supplemental Fig. 2A), indicating that FOXA1 and TALE-HD proteins target a common subset of enhancers. Collectively, these data suggest that the recruitment of these critical regulators at enhancers could influence or be sensitive to cytosine modifications.

Cytosine modifications increased TALE-HD binding in vitro

TALE-HDs bind DNA through the formation of MEIS1/PBX1 heterodimers or through dimerization with HOX proteins (Reviewed in Moens and Selleri, 2006). Accordingly, analysis of the proteins associated with PBX1-bound chromatin in RA-differentiated P19 cells by ChIP coupled to mass spectrometry (ChIP-MS) revealed the presence of several HOX and TALE-HD proteins, including MEIS1 (Supplemental Fig. 2B and Supplemental Table1). To address the possible influence of cytosine modifications on the binding of TALE-HD proteins to their recognition sites *in vitro*, pull-down analyses of cell extracts from differentiated P19 cells were then performed using probes containing the TGATTTACGAC PBX1/HOXA9 motif (Shen et al., 1997) in which cytosines from the CpG dinucleotide were differentially modified (either methylated or hydroxymethylated), as well as probes containing the MEIS1/PBX1 motif TGATTGACAG (Chang et al., 1997) modified at underlined positions. For each probe, the specific binding of MEIS1/2 and PBX1/2/3 was first verified by supershift experiments

(Supplemental Fig. 2C). Pull-down assays revealed that MEIS1 and PBX1/2/3/4 were better retained by the PBX1/HOXA9 probes containing specific combinations of modified cytosines (such as hemi-methylated, fully hydroxy- or methylated, or both methylated and hydroxymethylated probes) compared to the probe containing unmodified cytosines (Fig. 1D). Likewise, MEIS1 and PBX1/2/3/4 proteins were better retained, although less markedly, by the PBX1/MEIS1 probes containing modified cytosines (Supplemental Fig. 2D). These results suggest that MEIS1 and PBX1/2/3/4 bind preferentially to sequences containing modified cytosines *in vitro*. This is consistent with data from Spruijt et al. (2013) indicating that MEIS1 and PBX1 are preferentially retained by fully methylated CpGs and extends their findings by showing that TALE-HDs also prefer CpGs where 5mC is combined to 5hmC on the other strand rather than to an unmodified cytosine (Fig. 1D).

Cytosine modifications are required for the correct timing of neural progenitor differentiation

Since cytosine modifications could influence the binding of PFs to DNA *in vitro*, we next tested the impact of perturbing the methylation and demethylation processes on PF-bound enhancers in P19 cells by using subtoxic concentrations of either 5-aza-2'-deoxycytidine (5-azadC), an intercalating hypo-methylating agent which forms adducts with DNMTs (Jüttermann et al., 1994), or dimethyloxalylglycine (DMOG), a cell-penetrating derivative of N-Oxalylglycine (NOG), which inhibits Fe^{2+} /2-OG dioxygenases such as TET enzymes (Rose et al., 2011). As assessed by dot blot analysis, 5-azadC treatment reduced global 5mC levels in both undifferentiated and differentiated P19 cells and also limited 5hmC formation in RA-treated cells (Supplemental Fig. 3A). Treatment of P19 ECCs with DMOG did not impact global 5mC levels but inhibited significantly the increase in 5hmC observed during RA-induced differentiation (Supplemental Fig. 3B). We next evaluated the impact of the

perturbation of 5mC and 5hmC levels on the production of mRNAs encoding pluripotency factors (OCT4 and NANOG), neural progenitor markers (TSHZ1, PAX6 and IRX3) and cell fate determination TFs (FOXA1, MEIS1 and PBX1) during RA-induced P19 cell differentiation (Supplemental Fig. 4A). Although RA-mediated changes in expression of pluripotency factors and *Meis1* were not dramatically affected by 5-azadC and DMOG treatments, *Foxa1* and *Pbx1* mRNA levels did show sensitivity to the drugs. However, such sensitivity was not observed at the protein level (Supplemental Fig. 4B). Notably, induction of *Pax6* and *Irx3* expression by RA was significantly delayed by 5-azadC and DMOG treatments (Supplemental Fig. 4A), suggesting that both DNMT and TET inhibition, although not affecting the repression of pluripotency factor genes, delayed differentiation of neural progenitors. This is consistent with the observed delayed induction of certain marker genes during neural differentiation of TET2^{-/-} mouse embryonic stem cells (mESCs - Hon et al., 2014). Since both RA-induced loss of pluripotency and PF induction are not affected by the drugs while neural differentiation is delayed, these results suggest that our experimental conditions are suitable for the study of the role of DNA modifications at PF-bound enhancers.

Cytosine modifications control enhancer priming

In order to determine the role of cytosine modifications in enhancer priming and activation, we next analyzed the effects of 5-azadC or DMOG treatments on the levels of 5mC, 5hmC, H3K4me1, H3K27ac, chromatin opening (FAIRE), MEIS1 and FOXA1 recruitment at enhancers selected according to genome-wide DIP- and ChIP-seq data: (i) gain of 5hmC during differentiation, (ii) enrichment in H3K4me1 and H3K27ac, and (iii) recruitment of MEIS1 and/or FOXA1. Hence, 15 enhancers were selected with five bound by MEIS1 but not by FOXA1 (enhancers M1 to M5), five recruiting both MEIS1 and FOXA1 (enhancers MF1 to

MF5), and five bound by FOXA1 but not by MEIS1 (enhancers F1 to F5). As a control, we selected an enhancer enriched in H3K4me1 and H3K27ac and bound by both FOXA1 and MEIS1, but having no CpG in 400 bp flanking FOXA1 and MEIS1 ChIP-seq peak summits (enhancer “No CpG”). Notably, the chromatin features of this control enhancer were not affected by 5-azadC or DMOG treatments (Supplemental Fig. 5A-B). Since DMOG can also inhibit histone demethylases of the jumonji family which target methylated H3K9 and H3K27 (Rose et al., 2011), we controlled whether DMOG treatment increased histone methylation, potentially leading to inhibition of enhancers, and found that H3K27me3 levels did not increase in DMOG-treated cells for any of the selected enhancers (Supplemental Fig. 5C). These data indicate that, in our conditions, the observed effects of the DMOG treatment mainly reflected TET inhibition.

Chromatin features of the 15 selected enhancers were analyzed by (h)MeDIP-, ChIP- and FAIRE-qPCR experiments. Results of these experiments are depicted in the histograms of Supplemental Fig. 6 and 7 and were analyzed by principal component analysis (PCA) with data expressed as fold change between control and treated samples from either undifferentiated or differentiated P19 cells (detailed results for PCA and hierarchical clustering on principal components (HCPC) are given in supplemental Fig. 8 and 9). Due to the low sensitivity of the MeDIP technique to detect slight changes in 5mC levels, we also interrogated the cytosine modification status of CpGs included in a CCGG *MspI* site whenever present in the selected enhancers (Supplemental Fig. 10). Results obtained with undifferentiated P19 cells showed that inhibition of both DNMTs (Fig. 2A) and TETs (Fig. 2B) mainly affected 5hmC and H3K4me1 levels. Remarkably, under DNMT inhibition, HCPC identified two clusters: one characterized by a decrease in 5hmC at sites having significant levels of 5hmC in control condition (Supplemental Fig. 11, MF1, F4, M5 and F2), consistent

with a decrease in the substrate for TET-mediated oxidation; and another one with the regions showing an increase in H3K4me1 indicating that 5mC prevents monomethylation of H3K4 (Fig. 2A). This observation reveals that H3K4 monomethylation is triggered by cytosine demethylation rather than hydroxymethylation itself and does not necessarily require RA-induced factors. The anti-correlation between 5mC and H3K4 monomethylation could also be observed in TET inhibition condition. Indeed, a number of enhancers showed a decrease in 5hmC, likely leading to an increase in their 5mC content, and a decrease in H3K4me1 levels (Fig. 2B). This is consistent with data showing (i) that, in MCF-7 cells, binding of the H3K4 methyltransferase MLL1 to the *TFF1* enhancer depends on DNA hypomethylation (Jeong et al., 2014), and (ii) that 5mC repels MLL1 *in vitro* (Spruijt et al., 2013).

Importantly, data obtained with differentiated P19 cells revealed a positive correlation between the levels of 5hmC and chromatin opening (Fig. 2C-D). Since inhibition of 5mC formation by 5-azadC did not impact chromatin opening (FAIRE) in undifferentiated cells while increasing H3K4me1 levels (Fig. 2A), chromatin opening could be directly triggered by the presence of 5hmC itself and not by the loss of 5mC. This hypothesis is supported by data indicating that nucleosomes formed *in vitro* on a hydroxymethylated DNA template are unstable due to a lower interaction of DNA with H2A-H2B dimers (Mendonca et al., 2014). Interestingly, TET inhibition by DMOG led to both a decrease and increase in 5hmC levels (Fig. 2D), a phenomenon which has also been observed in human ECCs transfected with TET siRNAs (Putiri et al., 2014). The latter event (*i.e.* 5hmC increase when TETs are inhibited) could occur at enhancers where, in control conditions, 5hmC superoxidation in 5fC and 5caC takes place at a high rate, most probably because TET enzymes have a higher residency time at these particular enhancers. Based on their higher ability to gain 5hmC in the presence of DMOG, enhancers bound by FOXA1 seemed to be

more prone to superoxidation than MEIS1-only enhancers (Fig. 2D). In agreement with this hypothesis, clustering according to FOXA1 and 5hmC levels of FOXA1-bound enhancers genome-wide revealed that 5hmC and TET levels were inversely correlated for 3 clusters out of 4 (Fig. 3), clearly suggesting that high TET residency may indeed be associated with high 5hmC superoxidation. Hence, FOXA1 binding to enhancers might favor the engagement of TET enzymes involved in 5hmC oxidation. Accordingly, it was recently demonstrated that FOXA1 interacts with TET2 in prostate cancer cells (Takayama et al., 2015), and that 5hmC oxidation in human ECCs is mainly dependent on TET2 (Putiri et al., 2014). While recruitment of FOXA1 at a majority of enhancers was neither significantly impacted by TET and DNMT inhibition nor globally correlated to variations in any of the analyzed chromatin features, MEIS1 binding exhibited a significant decrease in 5-azadC condition and had a tendency to follow changes in chromatin opening levels in DMOG treated cells (Fig. 2 C-D). These observations are consistent with a model (Supplemental Fig. 12) in which FOXA1 engagement is mainly insensitive to chromatin attributes, as expected for a true PF, whereas MEIS1 binding can be stabilized by 5mC (in agreement with *in vitro* data shown in Fig. 1D and Supplemental Fig. 2D) and facilitated by 5hmC-promoted chromatin opening. Accordingly, the rapid increase in 5hmC levels at MEIS1 binding regions was systematically followed by chromatin opening and MEIS1 recruitment (Fig 4), and a significant fraction of these sites (Fig. 4, Clusters 1+2, 1,032 regions out of 3,366) showed chromatin opening before MEIS1 engagement. In turn, binding of MEIS1 would favor acetylation of H3K27, as suggested by the correlation between these two events (Fig. 2 and Supplemental Fig. 6 and 7).

Collectively, our data further validate the concept that DNA methylation is required for the proper regulation of enhancer priming and activation during cell differentiation and

provide a rationale for understanding how non-promoter DNA methylation can participate in the cell-specific control of gene activity.

Materials and Methods

Complete Materials and Methods are available in Supplemental Material section. All sequencing data are available at NCBI GEO (GSMXXXXX).

Cell culture

P19.6 mouse embryonal carcinoma cells were grown and differentiated as described (Sérandour et al., 2012). To inhibit DNA methyltransferases, cells were plated in a medium containing 20 nM of 5-aza-2-deoxycytidine (5-azadC, Sigma) for 24 hours. The medium was then renewed and cells were cultured in the presence of 20 nM 5-azadC for 48 hours with or without RA treatment. Inhibition of TET proteins was performed by treating P19 cells with 0.5 mM of dimethyl-oxalyl-glycine (DMOG, Cayman Chemical). Due to DMOG instability, the medium was renewed every 24 hours.

Chromatin and DNA immunoprecipitation

ChIP experiments were carried out as already described (Sérandour et al, 2011). ChIP followed by high-throughput sequencing (ChIP-seq) experiments were performed as in Schmidt et al., 2009. DNA immunoprecipitation protocols are available on the epigenesys website (<http://www.epigenesys.eu/en/protocols>): “Methylated DNA Immunoprecipitation (MeDIP)” (Protocol 33 - Michaël Weber and Dirk Schübeler) and “Hydroxymethylated DNA

Immunoprecipitation (hMeDIP) in mammalian cells” (Protocol 64 - Elise Mahé, Gilles Salbert).

Formaldehyde Isolation of Regulatory Element (FAIRE)

Formaldehyde assisted isolation of regulatory elements (FAIRE) was performed as previously described (Eeckhoutte et al., 2009).

Pull down assay

Differentially modified 20 bp synthetic oligonucleotides (Sigma), biotinylated on the forward strand and containing either the PBX1/HOXA9 (TGATTACG) or the PBX1/MEIS1 (TGATTGACAG) recognition sites, were used as probes for pull down assays. For each probe, cytosines from each CR dinucleotide (R=A or G), were unmodified (C), methylated (M) or hydroxymethylated (H). DNA/protein complexes were purified with streptavidin-coupled magnetic beads (Invitrogen) after incubation in binding buffer (Hepes 20 mM pH7.9, 50 mM KCl, 1 mM DTT, 2.5 mM MgCl₂, 5 µg/µL BSA, 10% Glycerol) with 10, 20 or 40 µg of extracts from P19 cell treated during 48 hours with RA. Protein binding was then analyzed by Western blot. Signals were quantified using the Quantity One™ 1-D Analysis Software from Bio-Rad.

Real-time PCR (qPCR)

Real-time PCR was performed using SYBR® Green Master Mix (BioRad, France) on a Bio-Rad CFX96 machine with 39 rounds of amplification. All primers were purchased from Sigma and are listed in Supplemental Material and Methods section.

Statistical analyses

Statistical analyses were performed using Prism software. Statistical significance was determined using Mann and Whitney test.

Principal component analysis and clustering

The R software, with the FactoMineR package (Lê *et al.*, 2008), was used to perform principal component analysis (PCA) and hierarchical clustering.

Acknowledgments

We thank C. Le Péron for critical reading of the manuscript. We also thank the sequencing facilities 'Human & Environmental Genomics' platform of Biogenouest Génomique in Rennes. This work was supported by the CNRS, the University of Rennes I and the Association de la Recherche contre le Cancer. E.M. was supported by a PhD fellowship from the Ministère de l'Enseignement Supérieur et de la Recherche.

References

- Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 2014. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* **6**: 1049-1055.
- Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, Miranda TB, Sung MH, Trump S, Lightman SL et al. 2011. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* **43**: 145-155.
- Bird A. 2011. The dinucleotide CG as a genomic signaling module. *J Mol Biol* **409**: 47-53.
- Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* **49**: 825-837.
- Chang CP, Jacobs Y, Nakamura T, Jenkins NA, Copeland NG, Cleary ML. 1997. Meis proteins are major in vivo DNA binding partners for wild-type but not chimeric Pbx proteins. *Mol Cell Biol* **17**: 5679-5687.
- de la Rica L, Rodríguez-Ubreva J, García M, Islam AB, Urquiza JM, Hernando H, Christensen J, Helin K, Gómez-Vaquero C, Ballestar E. 2013. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol* **14**: R99.
- Dong D, Meng L, Yu Q, Tan G, Ding M, Tan Y. 2012. Stable expression of FoxA1 promotes pluripotent P19 embryonal carcinoma cells to be neural stem-like cells. *Gene Expr* **15**: 153-162.
- Eeckhoutte J, Lupien M, Meyer CA, Verzi MP, Shivdasani RA, Liu XS, Brown M. 2009. Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res* **19**: 372-380.
- Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL et al. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* **32**: 317-328.
- Giresi PG, Kim J, McDaniel RM, Iyer VR, Lieb JD. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res* **17**: 877-885.

- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L et al. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**: 1303-1307.
- Heidt AB, Rojas A, Harris IS, Black BL. 2007. Determinants of myogenic specificity within MyoD are required for noncanonical E box binding. *Mol Cell Biol* **27**: 5910-5920.
- Heinz S, Romanoski CE, Benner C, Glass CK. 2015. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16**: 144-154.
- Hon GC, Song CX, Du T, Jin F, Selvaraj S, Lee AY, Yen CA, Ye Z, Mao SQ, Wang BA et al. 2014. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* **56**: 286-297.
- Iurlaro M, Ficiz G, Oxley D, Raiber EA, Bachman M, Booth MJ, Andrews S, Balasubramanian S, Reik W. 2013. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol* **14**: R119.
- Iwafuchi-Doi M, Zaret KS. 2014. Pioneer transcription factors in cell reprogramming. *Genes Dev* **28**: 2679-2692.
- Jeong KW, Andreu-Vieyra C, You JS, Jones PA, Stallcup MR. 2014. Establishment of active chromatin structure at enhancer elements by mixed-lineage leukemia 1 to initiate estrogen-dependent gene expression. *Nucleic Acids Res* **42**: 2245-2256.
- Jüttermann R, Li E, Jaenisch R. 1994. Toxicity of 5-aza-2'-deoxycytidine to mammalian cells is mediated primarily by covalent trapping of DNA methyltransferase rather than DNA demethylation. *Proc Natl Acad Sci U S A* **91**: 11797-11801.
- Lê S, Josse J, Husson F. 2008. FactoMineR: An R package for multivariate analysis. *J Stat Software* **25**: 1-18.
- Lupien M, Eeckhoutte J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**: 958-970.
- Magnani L, Ballantyne EB, Zhang X, Lupien M. 2011. PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS Genet* **7**: e1002368.
- Maiti A, Drohat AC. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* **286**: 35334-35338.

- Mendonca A, Chang EH, Liu W, Yuan C. 2014. Hydroxymethylation of DNA influences nucleosomal conformation and stability in vitro. *Biochim Biophys Acta* **1839**: 1323-1329.
- Moens CB, Selleri L. 2006. Hox cofactors in vertebrate development. *Dev Biol* **291**: 193-206.
- Putiri EL, Tiedemann RL, Thompson JJ, Liu C, Ho T, Choi JH, Robertson KD. 2014. Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. *Genome Biol* **15**: R81.
- Qin P, Habermusch JM, Zhang Z, Soprano KJ, Soprano DR. 2004. Pre-B cell leukemia transcription factor (PBX) proteins are important mediators for retinoic acid-dependent endodermal and neuronal differentiation of mouse embryonal carcinoma P19 cells. *J Biol Chem* **279**: 16263-16271.
- Rose NR, McDonough MA, King ON, Kawamura A, Schofield CJ. 2011. Inhibition of 2-oxoglutarate dependent oxygenases. *Chem Soc Rev* **40**: 4364-4397.
- Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* **48**: 240-248.
- Sérandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Métivier R et al. 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* **21**: 555-565.
- Sérandour AA, Avner S, Oger F, Bizot M, Percevault F, Lucchetti-Miganeh C, Palierne G, Gheeraert C, Barloy-Hubler F, Péron CL et al. 2012. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* **40**: 8255-8265.
- Shen WF, Rozenfeld S, Lawrence HJ, Largman C. 1997. The Abd-B-like Hox homeodomain proteins can be subdivided by the ability to form complexes with Pbx1a on a novel DNA target. *J Biol Chem* **272**: 8198-8206.
- Siersbæk R, Nielsen R, John S, Sung MH, Baek S, Loft A, Hager GL, Mandrup S. 2011. Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *EMBO J* **30**: 1459-1472.
- Spitz F, Furlong EE. 2012. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet* **13**: 613-626.

- Spruijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Münzel M, Wagner M, Müller M, Khan F et al. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**: 1146-1159.
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490-495.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L et al. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**: 930-935.
- Takayama K, Misawa A, Suzuki T, Takagi K, Hayashizaki Y, Fujimura T, Homma Y, Takahashi S, Urano T, Inoue S. 2015. TET2 repression by androgen hormone regulates global hydroxymethylation status and prostate cancer progression. *Nat Commun* **25**: 8219.
- Yamada T, Urano-Tashiro Y, Tanaka S, Akiyama H, Tashiro F. 2013. Involvement of crosstalk between Oct4 and Meis1a in neural cell fate decision. *PLoS One* **8**: e56997.

Figure Legends

Figure 1. *De novo* hydroxymethylated enhancers bind pioneer factors driving ECC commitment to the neural lineage. (A) Average kinetic profiles of 5hmC, 5mC, H3K4me1, H3K27ac and chromatin opening (FAIRE) at 20,492 regions gaining 5hmC (up DhMRs) at 48 hours of P19 ECC differentiation. (B) Average profiles of FOXA1, PBX1 and MEIS1 binding at up DhMRs at 48 hours of differentiation. (C) Heatmap representations of 5hmC signal at MEIS1, PBX1 and FOXA1 enriched regions in 48 hours RA-treated P19 cells. (D) MEIS1 and PBX1/2/3/4 *in vitro* interaction with differentially modified synthetic PBX1/HOXA9 probes. Results are expressed according to the strongest binding equivalent to 100 (C: unmodified cytosine; M: methylated cytosine; H: hydroxymethylated cytosine – first letter: upper strand; second letter: lower strand). Statistical significance of differential binding to modified probes compared to the unmodified probe was tested with a Mann and Whitney test (MEIS1 protein: * for $p < 0.05$, ** for $p < 0.01$; PBX1/2/3/4: # for $p < 0.05$, ## for $p < 0.01$). $\alpha\alpha$ corresponds to $p < 0.01$ between M-C and M-H conditions.

Figure 2. Cytosine modifications regulate enhancer priming. Dendrograms and heatmaps from hierarchical clustering of enhancer features (left dendrogram) and the 15 selected enhancers studied (top dendrogram). Upper panel (A and B): undifferentiated P19 ECCs. Lower panel (C and D): differentiated neural progenitors. The color scale bars at the bottom of each column indicate $\log_2$ -transformed fold-change between treated and control cells with a scale of 0.5 for 5-azadC (A and C) and 0.25 for DMOG (B and D). Variations in MEIS1 and FOXA1 recruitment are represented below heatmaps C and D, but were not used for hierarchical clustering.

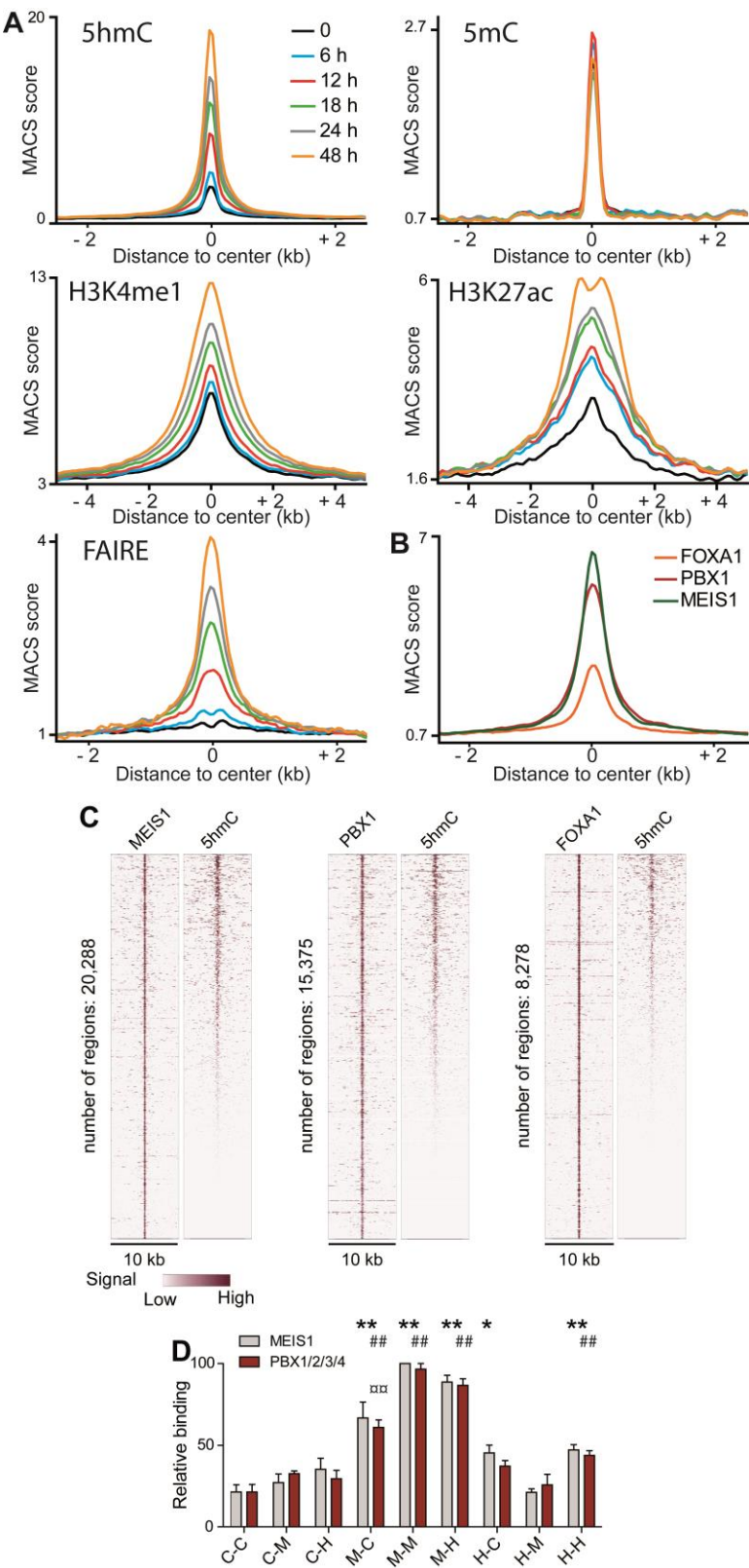
Figure 3. High levels of 5hmC correlate with low levels of TETs at FOXA1-bound enhancers.

(A) Heatmaps of FOXA1 and 5hmC signals in clusterized FOXA1 binding regions. (B) Average profiles of 5hmC, TET1 and TET2 recruitment at the clusters C2, C3, C4 and C5 shown in (A).

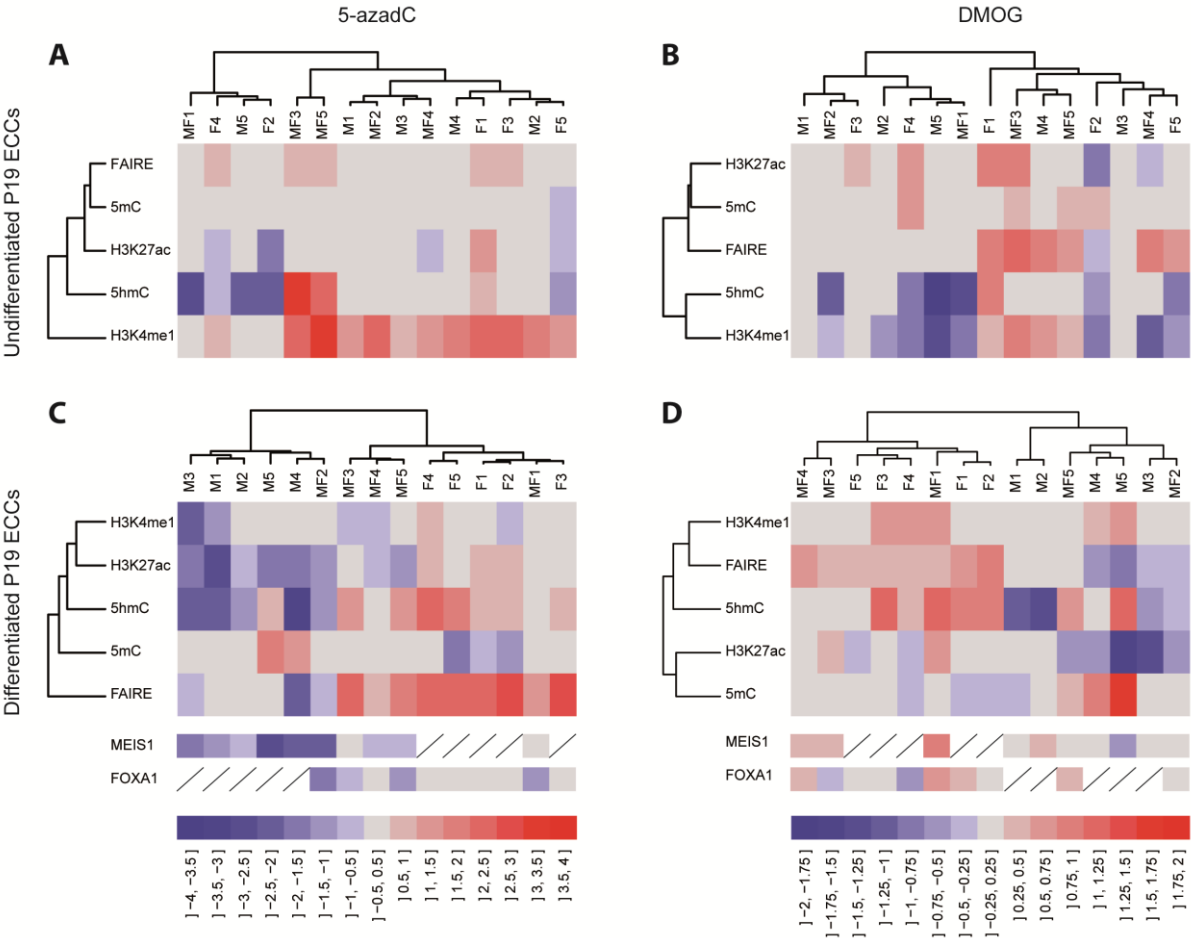
Figure 4. Gain of 5hmC is followed by chromatin opening and MEIS1 binding. (A) Heatmaps

of MEIS1 and FAIRE differential signals (ΔS) during P19 cell differentiation in clusterized MEIS1 binding regions. Blue: decrease; Red: increase. (B) Average profiles of dynamics of 5hmC, chromatin opening and MEIS1 recruitment at clusters shown in (A).

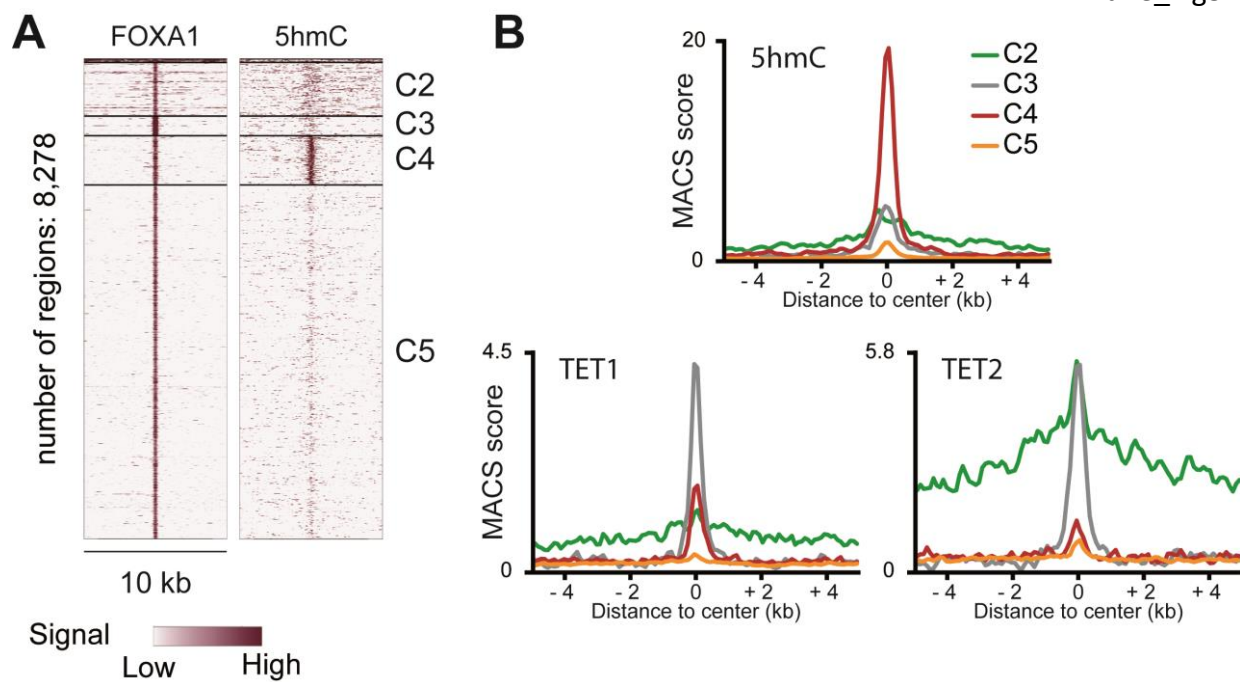
Mahé_Fig1



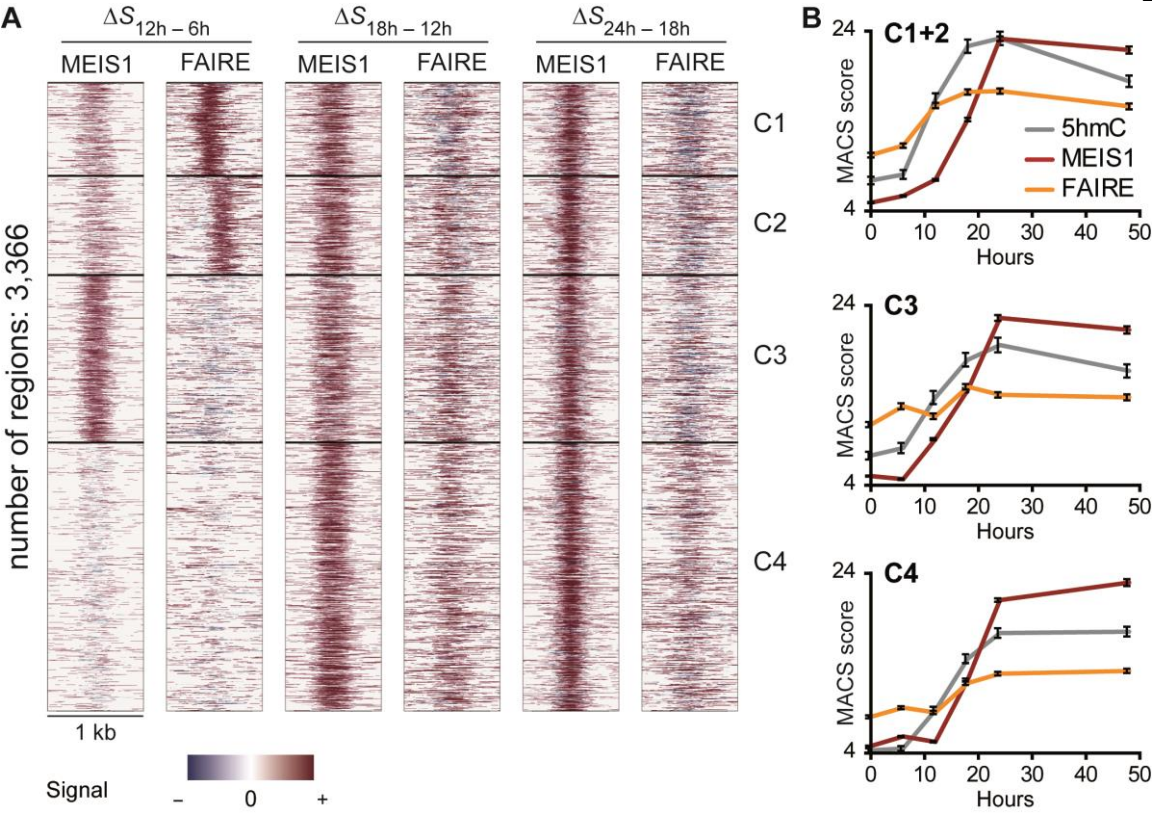
Mahé_Fig2



Mahé_Fig3



Mahé_Fig4



Cytosine modifications modulate the chromatin architecture of transcriptional enhancers

Elise A Mahé, Thierry Madigou, Aurélien A Sérandour, Maud Bizot, Stéphane Avner, Frédéric Chalmel, Gaëlle Palierne, Raphaël Métivier and Gilles Salbert

Supplemental Material

- Supplemental Figure 1 is related to Figure 1 in the text and shows the dynamics of chromatin features for down dhMRs.
- Supplemental Figure 2 is also related to Figure 1 and further support an influence of the epigenome on the binding of pioneer factors.
- Supplemental Figure 3 displays the results of experiments performed to verify 5-azadC and DMOG efficiency.
- Supplemental Figure 4 shows the analysis of gene expression (mRNA and protein) of several transcription factors implicated in P19 cell differentiation.
- Supplemental Figure 5 depicts controls for 5-azadC and DMOG specificity.
- Supplemental Figure 6 is related to Figure 2 A and C in the text and contains all the results obtained for DNMT inhibition on the 15 selected enhancers from (h)MeDIP-, ChIP- and FAIRE-qPCR experiments.
- Supplemental Figure 7 is related to Figure 2 B and D in the text and contains all the results obtained for TET inhibition on the 15 selected enhancers from (h)MeDIP-, ChIP- and FAIRE-qPCR experiments.
- Supplemental Figure 8 is related to Figure 2 A and B in the text and shows the detailed results of PCA and HCPC for undifferentiated P19 ECCs treated with 5-azadC or DMOG.

- Supplemental Figure 9 is related to Figure 2 C and D in the text and shows the detailed results of PCA and HCPC for differentiated P19 ECCs treated with 5-azadC or DMOG.
- Supplemental Figure 10 displays the results obtained for Epimark® 5-hmC and 5-mC analysis as additional evidences to complete (h)MeDIP experiments.
- Supplemental Figure 11 represents the basal levels of enrichment of each enhancer feature for the 15 enhancers studied.
- Supplemental Figure 12 proposes a model for cytosine modification-promoted events during enhancer priming and activation.
- Supplemental Table 1 presents details of the proteomic data from the CHIP-MS experiments.
- Supplemental Materials and Methods section contains the list of antibodies and oligonucleotides used for this study, as wells as protocols for the sequencing preparation and bioinformatics, the CHIP-MS, Chemiluminescent EMSA, dot blot assay, RNA preparation and reverse transcription, Epimark® 5-hmC and 5-mC Analysis, supplemental information for Principal component analysis and clustering.
- Supplemental references are related to the Supplemental Materials and Methods section.

Legends to supplemental Figures

Supplemental Figure 1. Dynamics of chromatin features at down DhMRs during P19 cells differentiation. (A) Average kinetic profiles of 5hmC, 5mC, H3K4me1, H3K27ac and chromatin opening (FAIRE) at 1,712 regions losing 5hmC (down DhMRs) at 48 hours of P19 ECC differentiation. (B) Average profiles of FOXA1, PBX1 and MEIS1 binding to down DhMRs at 48 hours of differentiation.

Supplemental Figure 2. TALE-HDs are recruited to common sites and bind preferentially to sequences containing modified cytosines *in vitro*. (A) Heatmap representations of PBX1 signal at MEIS1-bound regions (left panel) and of PBX1 and MEIS1 signals at FOXA1-bound regions (right panel) at 48 hours of P19 ECC differentiation. (B) Part of a STRING (Search Tool for the Retrieval of Interacting Genes; <http://string-db.org/>; Szklarczyk et al., 2015) schematic representation of interactions between proteins retrieved from PBX1 ChIP-MS data. Pink lines represent known physical interaction and grey lines functional relationships. (C) Validation of MEIS1/2 and PBX1/2/3 binding to differentially modified MEIS1/PBX1 probes by supershift experiments. Ab: antibody; NE: P19 nuclear extract; HA: Human influenza virus hemagglutinin (negative control). Validation was also achieved for PBX1/HOXA9 probes (data not shown). (D) MEIS1 and PBX1/2/3/4 *in vitro* interaction with differentially modified synthetic MEIS1/PBX1 probes. Results are expressed according to the strongest binding equivalent to 100 (C: unmodified cytosine; M: methylated cytosine; H: hydroxymethylated cytosine – first letter: upper strand; second letter: lower strand). Statistical significance of differential binding to modified probes compared to the

unmodified probe was tested with a Mann and Whitney test (MEIS1 protein: * for $p<0.05$, ** for $p<0.01$; PBX1/2/3/4: # for $p<0.05$, ## for $p<0.01$).

Supplemental Figure 3. Efficiency of 5-azadC and DMOG treatments. (A-B) Global 5hmC and 5mC levels of three DNA quantities (300, 150 or 75 ng) from P19 cells either undifferentiated (-RA) or differentiated during 48 hours with RA (+RA) and treated with or without 5-azadC (A – $n=6$) or DMOG (B – $n=9$) were evaluated by dot-blot. Quantified signals ($n=6$) for either 5mC or 5hmC were normalized to total DNA signals obtained by methylene blue staining. Statistical significance was tested with a Mann and Whitney test (5-azadC: # for $p<0.05$, ## for $p<0.01$; DMOG: *** for $p<0.005$).

Supplemental Figure 4. Perturbing cytosine modification levels delays differentiation of neural progenitors. (A) Kinetic analysis of the expression of *Oct4*, *Nanog*, *Tshz1*, *Pax6*, *Irx3*, *FoxA1*, *Meis1* and *Pbx1* genes by RT-qPCR ($n=6$) during the first 48 hours of RA-induced differentiation of P19 ECCs (control: black lines, DMOG: red lines, 5-azadC: dashed grey lines). Statistical significance was tested with a Mann and Whitney test (DMOG: * for $p<0.05$, ** for $p<0.01$; 5-azadC: # for $p<0.05$, ## for $p<0.01$). (B) Western blot quantification of MEIS1, PBX1 and FOXA1 proteins for 5-azadC (left panel) or DMOG (right panel) treated samples compared to control.

Supplemental Figure 5. Specificity of 5-azadC and DMOG treatments. (A-B) Impact of 5-azadC (A) or DMOG (B) treatments on a control site having no CpG and bound by MEIS1 and FOXA1 proteins. Histograms correspond to results obtained from P19 cells undifferentiated (- RA) or differentiated during 48 hours with RA (+RA) with or without inhibitor treatment.

Analysis of the opening of chromatin was performed by Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE). (C) ChIP-qPCR analysis of H3K27me3 enrichment at the 15 enhancers of interest in DMOG treated cells versus control cells (n=6).

Supplemental Figure 6. Impact of DNMT inhibition on characteristic features of enhancers. Histograms correspond to results obtained from undifferentiated (- RA) or differentiated P19 cells (+RA, 48 hours) with or without 5-azadC treatment (5-mC: n=9; 5hmC, H3K4me1, H3K27ac and FAIRE: n=6; MEIS1 and FOXA1: n=9). Statistical significance was tested with Mann and Whitney (# for $p<0.05$, ## for $p<0.01$, ### for $p<0.005$). (A) Enhancers bound by MEIS1 but not FOXA1 (M1 to M5). (B) Enhancers bound by both MEIS1 and FOXA1 (MF1 to MF5). (C) Enhancers bound by FOXA1 but not MEIS1 (F1 to F5).

Supplemental Figure 7. Impact of TET inhibition on characteristic features of enhancers. Histograms correspond to results obtained from undifferentiated (- RA) or differentiated P19 cells (+RA, 48 hours) with or without DMOG treatment (5hmC: n=9, excepted 48 hours n=15; 5-mC: n=6; H3K4me1: n=6, excepted 48 hours n=9; H3K27ac: n=12; FAIRE, MEIS1 and FOXA1: n=9). Statistical significance was tested with Mann and Whitney (* for $p<0.05$, ** for $p<0.01$, *** for $p<0.005$). (A) Enhancers bound by MEIS1 but not FOXA1 (M1 to M5). (B) Enhancers bound by both MEIS1 and FOXA1 (MF1 to MF5). (C) Enhancers bound by FOXA1 but not MEIS1 (F1 to F5).

Supplemental Figure 8. Principal Component Analysis (PCA) and Hierarchical Clustering on Principal Component (HCPC) of the impact of DNMT or TET inhibition on undifferentiated P19 ECCs. PCA and HCPC results obtained from the $\log_2$ -transformed fold-change between 5-

azadC (A-D)- or DMOG (E-H)-treated *versus* control cells for 5hmC, 5mC, H3K4me1, H3K27ac and FAIRE at the 15 enhancers studied. (A) and (E) Scree plots of information rate of the principal components. Most of the variance in the data set is contained in the first two (A) or four (E) principal components. (B) and (F) Individuals factor maps showing the coordinates of the individuals on the two first principal components. (C) and (G) Correlation loading plots of the variables with the components. The eigenvalues of the correlation matrix are symbolized as vectors representing traits that influence the principal components 1 and 2. (D) and (H) Hierarchical Clustering on Principal Components map.

Supplemental Figure 9. Principal Component Analysis (PCA) and Hierarchical Clustering on Principal Component (HCPC) of the impact of DNMT or TET inhibition on differentiated P19 ECCs. PCA and HCPC results obtained from the log₂-transformed fold-change between 5-azadC (A-D)- or DMOG (E-H)-treated *versus* control cells for 5hmC, 5mC, H3K4me1, H3K27ac and FAIRE at the 15 enhancers studied. (A) and (E) Scree plots of information rate of the principal components. Most of the variance in the data set is contained in the first three principal components. (B) and (F) Individual factor map showing the coordinates of the individuals on the two first principal components. (C) and (G) Correlation loading plots of the variables with the components. The eigenvalues of the correlation matrix are symbolized as vectors representing traits that influence the principal components 1 and 2. (D) and (H) Hierarchical Clustering on Principal Components map.

Supplemental Figure 10. 5-azadC and DMOG treatments affect the methylation or hydroxymethylation status of selected CpGs in eight enhancers of interest. Epimark® 5-hmC and 5-mC analysis of the impact of 5-azadC and DMOG treatments of undifferentiated or

differentiated P19 cells, in the context of CCGG sequences. Results are expressed as percentages for each cytosine form.

Supplemental Figure 11. Basal levels of enhancer features at the 15 selected enhancers. Fold enrichment compared to a negative control of 5hmC, 5mC, H3K4me1, H3K27ac and chromatin opening (FAIRE) in undifferentiated P19 cells for the five MEIS1 sites (M), the five MEIS1-FOXA1 sites (MF), the five FOXA1 sites (F) and an enhancer having no CpG (“No CpG”). Red lines correspond to the thresholds determined for each mark and above which enrichment was considered as significant. Thresholds were set at: 4 fold enrichment compare to negative control for 5hmC, 2 fold enrichment for 5mC and 3 fold enrichment for H3K4me1 and H3K27ac. No threshold was set for chromatin opening as this feature corresponds to levels of opening rather than presence or absence of a specific mark.

Supplemental Figure 12. Model for cytosine modification-promoted events during enhancer priming and activation. RA-mediated neural differentiation of P19 ECCs is associated with loss of expression of the pluripotency factors OCT4 and NANOG and the sequential induction of FOXA1 and MEIS1 pioneer factors. At an early stage of differentiation, FOXA1 binding to inactive enhancers associates with 5mC oxidation which indirectly promotes monomethylation of H3K4 by histone methyltransferases (HMTs), while 5hmC formation promotes chromatin opening. Both processes allow the establishment of a “primed enhancer” chromatin state. Later during differentiation (*i.e.* when MEIS1 levels raise) MEIS1 recruitment to primed enhancers is favored by the presence of modified cytosines either fully methylated or methylated on one DNA strand and hydroxymethylated on the other

strand and correlates with H3K27 acetylation by histones acetyl-transferases (HATs), resulting in enhancer activation.

Supplemental Table 1. Proteomic data for HOX and TALE-HD proteins found associated with PBX1 on chromatin. Exclusive Unique Peptide Count and percentage of coverage for ChIP-MS data are shown for TALE-HD and HOX proteins. Results are presented for three replicates of PBX1 immunoprecipitation compared to three replicates of IgG negative control.

Supplemental Materials and Methods*Antibodies*

	Target	Reference
Immunoprecipitations	5hmC	Diagenode, CS-HMC-100
	5mC	Diagenode, Mab-081-500
	PBX1	Abnova, H00005087-M01
	H3K4me1	Abcam, ab8895
	H3K27ac	Abcam, ab4729
	MEIS1	Abcam, ab19867
	FOXA1	Abcam, ab5089
	H3K27me3	Upstate, 07-449
Blots	5hmC	Active Motif, #39769
	5mC	Diagenode, Mab-081-500
	MEIS1	Abcam, ab124686
	PBX1/2/3/4	Santa Cruz, sc-25411X
	FOXA1	Abcam, ab5089
	Sp1	Santa Cruz, sc-59
	Actin	Sigma, A1978
Supershift	MEIS1/2	Santa Cruz, sc-10596X
	PBX1/2/3	Santa Cruz, sc-888
	HA	Santa Cruz, sc-805

Oligonucleotides

Oligonucleotides were designed under Primer 3 interface (<http://frodo.wi.mit.edu/primer3/>; Untergasser et al., 2012)

Primer type	Name	Forward primer	Reverse primer
RT-qPCR	RPS28	GGTGACGTGCTCACCTATT	TTCCGTGGGCTAAGTAGTGG
	OCT4	TTGCAGCTCAGCCTTAAGAA	TGGTCTCCAGACTCCACCTC
	NANOG	AAGCAGAAGATGCGGACTGT	GCACTTCATCCTTTGGTTTTG
	FOXA1	GAAGATGGAAGGGCATGAAA	GCCTGAGTTCATGTTGCTGA
	MEIS1	CAGAAAAAGCAGTTGGCACA	TGCTGACCGTCCATTACAAA
	PBX1	GCCAATATTTATGCTGCCAAA	AATCCCCATTGAGTGAAGTC
	TSHZ1	CCGAGGAGGAAGCAGCAG	TGGGCTTCTTTGATCTCTGTC
	IRX3	ACAGCACCTGCTGGGACTT	TTCTCCACTTCCAAGGCACT
	PAX6	ATGGCAAACAACCTGCCTAT	CATGGGCTGACTGTTTCATGT
	p21	CACTTCTTACCTGGGGGTGA	AGCTGGCCTTAGAGGTGACA
	GATA4	TCCAGCCTGAACATCTACCC	GTGTGAAGGGGTGAAAAGGA
	p53	TGGAAGACTCCAGTGGGAAC	TCTTCTGTACGGCGGTCTCT
	NESTIN	CCAAAGAGGTGTCCGATCAT	TCATCAGCAAACCCATCAGA
ChIP-qPCR	Negative control	TCCTTTGACCTCCACACACA	GCAACAGAAGATGCGATGAA
	M1	AGAGCGAACCGAGAAATCAA	AGAGCCTCTATTTGCCACGA
	M2	GACCTGAGCGCACAGTAACA	ATGAAGCTGCCTTTGAGGA
	M3	CCGCATTATGAACACACTCG	GCTCTCCCAGGGGTAATAGG
	M4	ACATTCCCCAGACAAGCAAG	CACTGACTGTCATCCCCTCA
	M5	CCGAGTGTTCTCCACCTAA	ATTTGCATACAGGCGGAGTC
	MF1	TGAGGTGCTCAGTGGAGATG	CTCCAGAGTGAGCGTTGACA
	MF2	CAGTCCCATCGATCACTTACA	TTCAAAGGCTTGCTTGACCT
	MF3	GCAGCCACTGAAGAGACAAA	TGGTTGTGGGCTTGTTTGA
	MF4	GACAAGCAGAACAGAACGGG	TGTCAGTCCCCCTTTCATCT
	MF5	TGCTTTACTCTCTGCAGCCT	ATATGGCCCTGTCAGTTCCC
	F1	GAAAAGCAAAGCAAAGGGGC	GAAGGCGACACTGCTCTGTA
	F2	CACATGCTAAGCCAGGAAGC	CCTGAGTTTGACCCCTGGAT
	F3	GCCGATGAGTAAACAGCTGA	CCGTTTCAAGCTGGTGTACA
	F4	CAGGTGGCTGAGACTGAGTA	TGGGAAATTCAGGTTGGCC
	F5	TGAAGTCCTCCAGTGGCTTT	CCCCAAGGCAAGAGTAAATACA
	No CpG	ACTATCTATCAAGGGAGCAAACA	TGAGCAGTAAAACCAGGTCTG
Epimark® 5-hmC and 5-mC Analysis	M2	CACATGTCCCTTTTGCTCCT	AGAGCGAGCACCTCACAAAA
	M3	CCGCATTATGAACACACTCG	GCTCTCCCAGGGGTAATAGG
	M5	TCTTTCAGCGGGACTGAGAT	GGTTGATTGCCAGCTTTGTT
	F4	ATGACCAGTATGCACCACCA	TCTTCCCTGACCCTACATGG

Sequencing preparation and bioinformatics

The (h)MeDIP, ChIP, FAIRE and input libraries were prepared using the TruSeq ChIP Sample Prep Kit (Illumina, ref. IP-202-1012). The libraries were sequenced by HiSeq following the manufacturer's protocol. Alignment of reads, conversion of extracted reads to .wig signal files and peak calling were performed as previously described (Sérandour et al., 2012). Blacklisted regions available from UCSC (<http://genome.ucsc.edu>) were removed from each .wig signal file. CGI coordinates were also obtained from UCSC. Heatmaps were generated by the cistrome web-platform (<http://cistrome.org/ap/>; Liu et al., 2011). TET1 sequencing data are from the NCBI GEO database reference GSM941681.

ChIP-MS

ChIP followed by Mass Spectrometry analysis (ChIP-MS) was performed as in Mohammed et al., 2013. Nuclear fraction from (1×10^7) P19 cells cross-linked by formaldehyde was extracted and sonicated. Immunoprecipitation was performed with 100 μ L of magnetic beads (Dyna) prebound with 15-20 μ g antibody. After several washes of the beads, the bead-bound proteins were digested with trypsin. The beads were then incubated overnight at 37°C to allow the salvage of the digested peptides which were then analyzed by nano-LC-MS/MS. Mass spectrometry (MS) was performed using an LTQ Velos-Orbitrap MS (Thermo Scientific) coupled to an Ultimate RSLCnano-LC system (Dionex). Results were analyzed with the Scaffold Viewer software with Protein Threshold and Peptide Threshold set at 90%.

Chemiluminescent EMSA

EMSA experiments were performed with nuclear extracts (NE) prepared from P19 cells treated during 48 hours with 1 μ M RA. Two micrograms of NE were incubated with binding

buffer (Hepes 20 mM pH7.9, 10 mM KCl, 1 mM DTT, 2.5 mM MgCl₂, 10% Glycerol), 1µg poly(dI-dC) and 150 fmol of the biotinylated probes used in pull down assays. For supershift assays, NE were first pre-incubated with the antibody of interest. EMSAs were performed using the “Gelshift™ Chemiluminescent EMSA” kit (Active Motif) following the manufacturer’s instructions.

Dot Blot Assay

Genomic DNA was extracted using a DNA extraction kit (DNeasy Blood and Tissue kit; QIAGEN) with a supplementary step of RNase digestion. 300 ng, 150 ng and 75 ng of each sample DNA was spotted to a nylon membrane (Hybond N, GE Healthcare) previously soaked in 2X SSPE solution (0.3 M NaCl, 20 mM sodium phosphate, 2 mM EDTA) and inserted in the dot blot apparatus (SCIE-PLAS). DNA was crosslinked to the membrane by 15 min exposure to UV light before blotting. After revelation, membranes were stained with methylene blue to control the amount of DNA spotted. Signals were quantified using the Quantity One™ 1-D Analysis Software from Bio-Rad.

RNA preparation and reverse transcription

RNA was isolated by TRIzol™ extraction (Invitrogen) and reverse-transcribed using M-MLV reverse transcriptase (Invitrogen) and Pd(N)₆ random hexamers (Amersham Pharmacia Biosciences).

EpiMark® 5-hmC and 5-mC Analysis

Methylation or hydroxymethylation status of selected CpGs included in a *MspI* CCGG restriction site was verified using the EpiMark® 5-hmC and 5-mC Analysis kit (New England

Biolabs) according to the manufacturer's protocol. Briefly, genomic DNA was first treated by T4 beta-glucosyltransferase (T4-BGT) in order to add a glucose moiety to the hydroxyl group of 5hmC, and then digested by either *MspI* or *HpaII* restriction enzymes. Both enzymes recognize a CCGG sequence but *HpaII* cleaves only unmodified sites whereas *MspI* cleaves 5mC or 5hmC but not 5ghmC. Approximate quantification of cytosine modifications was obtained by real-time PCR. Specific oligonucleotides were designed whenever those designed for ChIP-qPCR were not appropriate. Analysis and calculations were performed according to the kit protocol.

Principal component analysis and clustering

Briefly, chromatin mark variations across several genomic regions were investigated by PCA where variables were genomic regions and individuals were chromatin marks. Data were sequentially scaled to unit variance and then projected on orthogonal components. The resulting Eigenvalues (based on both the covariance and correlation matrices) were used to retain the first components accounting for over 90% of the variance in the data. Finally, the coordinates of the individuals and variables on the selected principal components were used for the hierarchical classification.

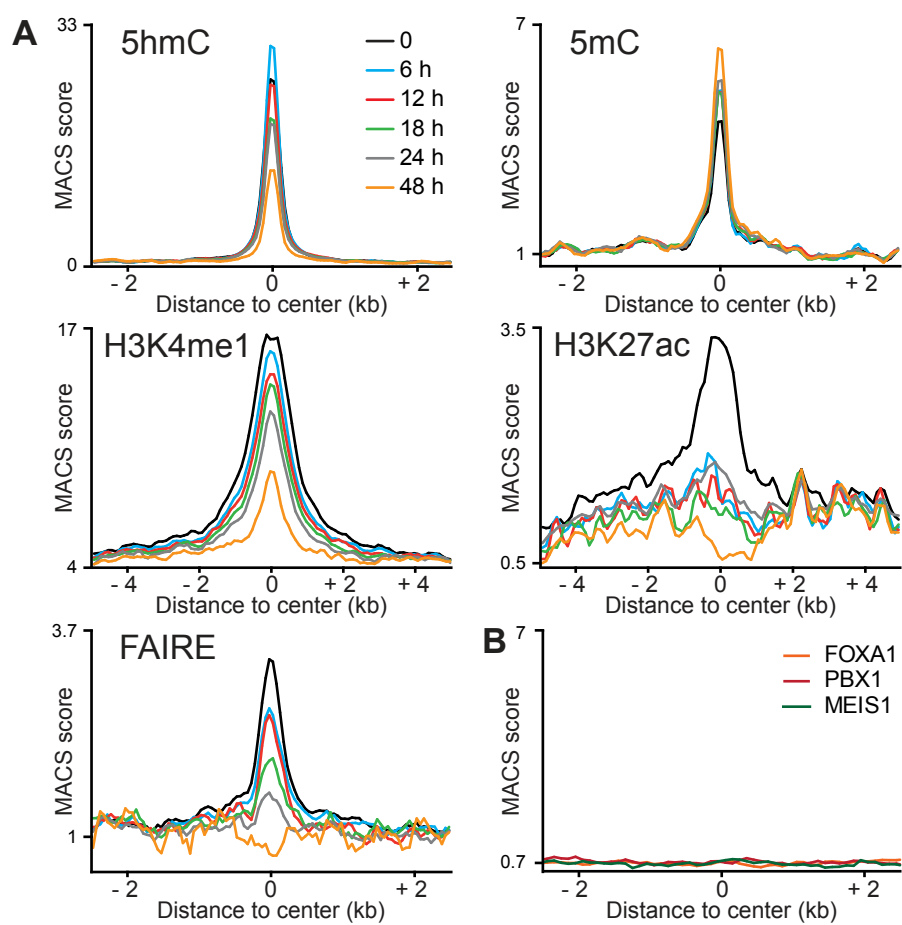
Supplemental references

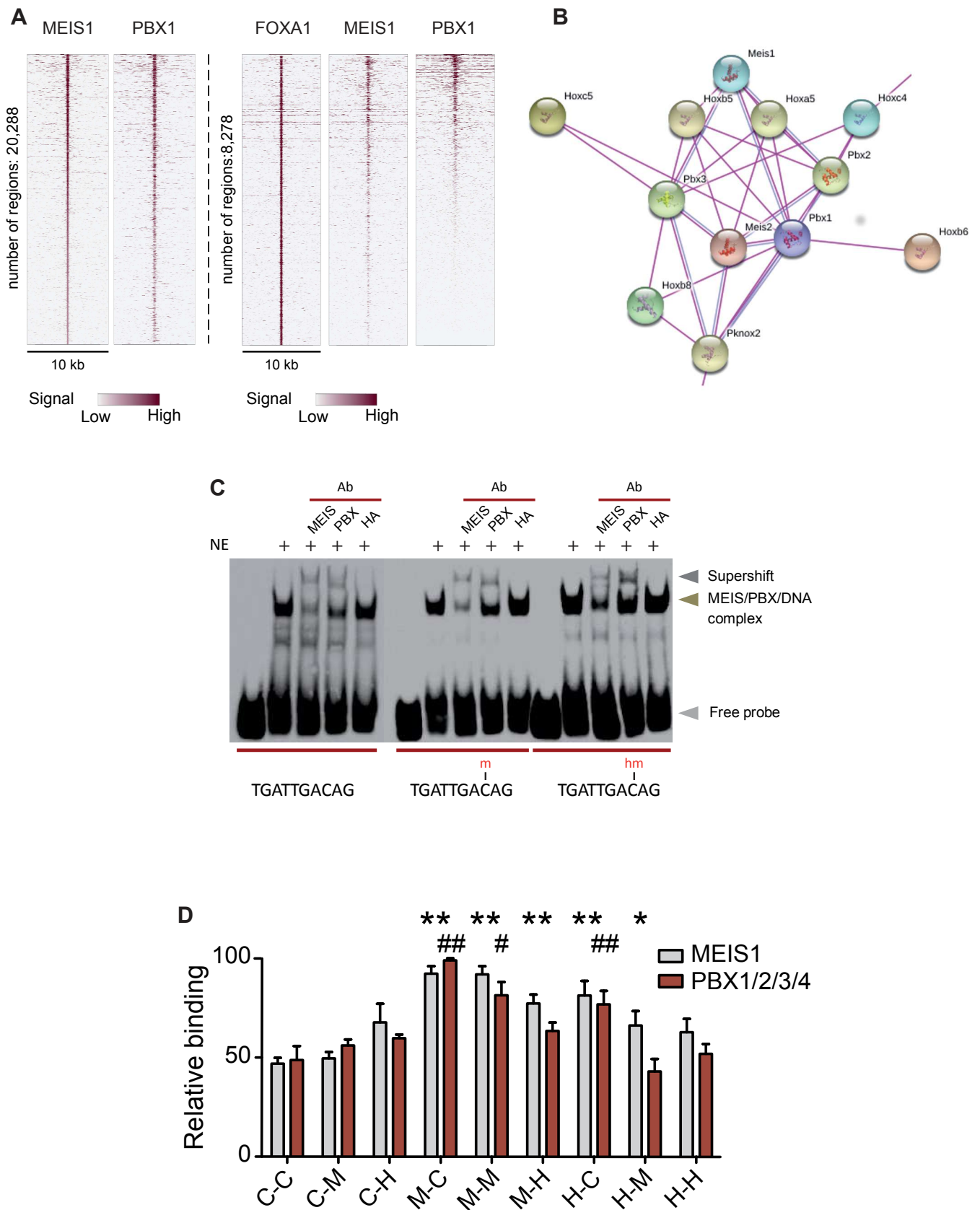
Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y et al. 2011. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* **12**: R83.

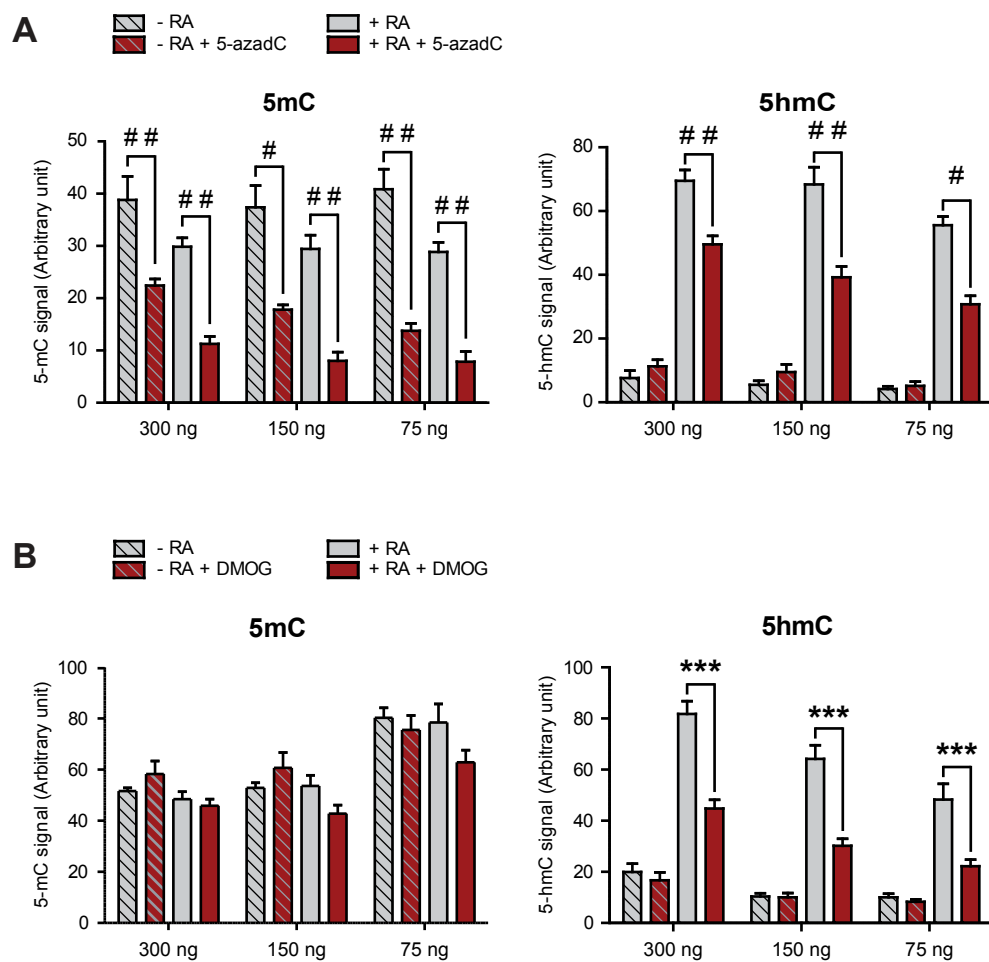
Mohammed H, D'Santos C, Serandour AA, Ali HR, Brown GD, Atkins A, Rueda OM, Holmes KA, Theodorou V, Robinson JL et al. 2013. Endogenous purification reveals GREB1 as a key estrogen receptor regulatory factor. *Cell Rep* **21**: 342-349.

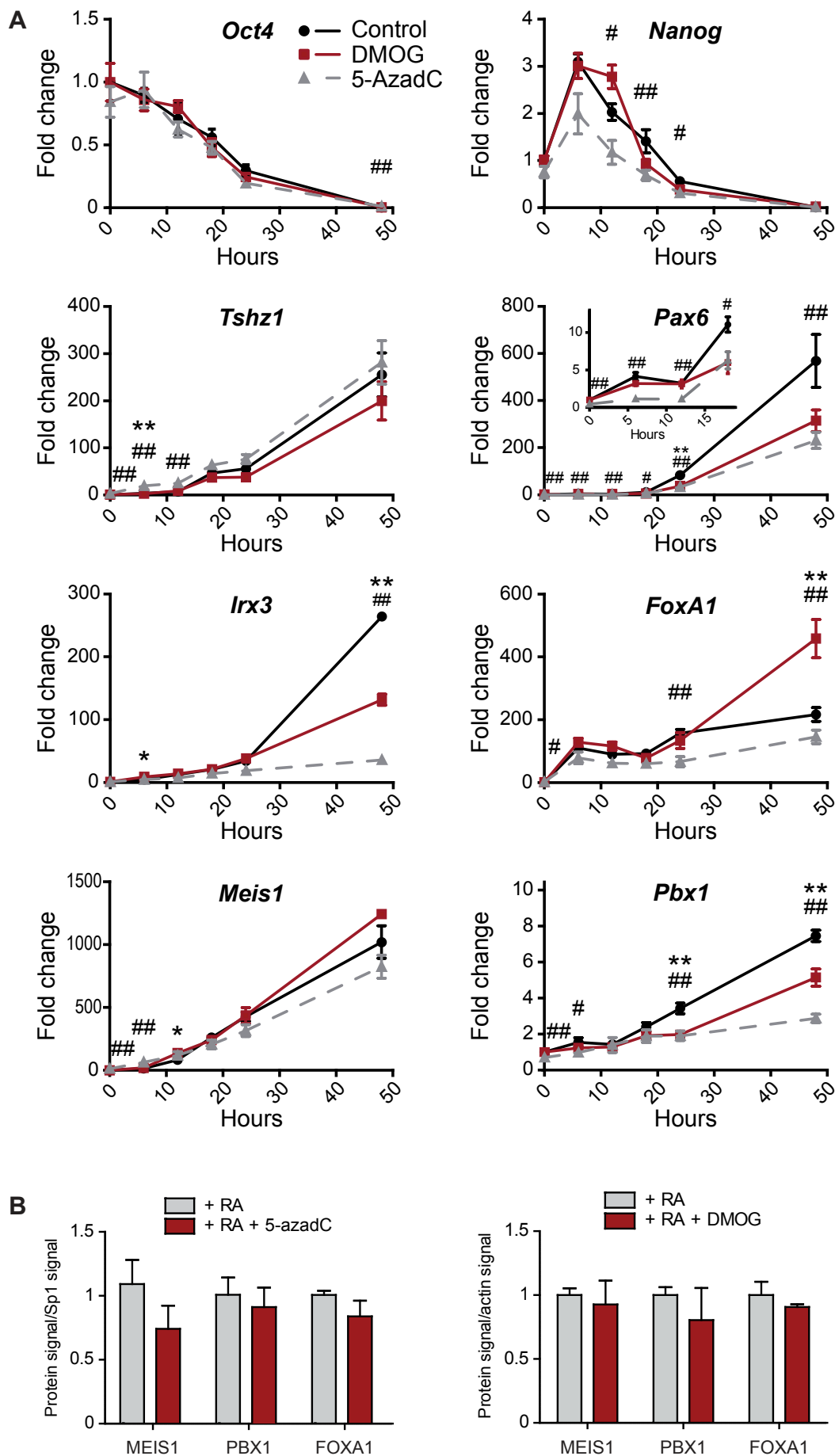
Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP et al. 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **43**: D447-452.

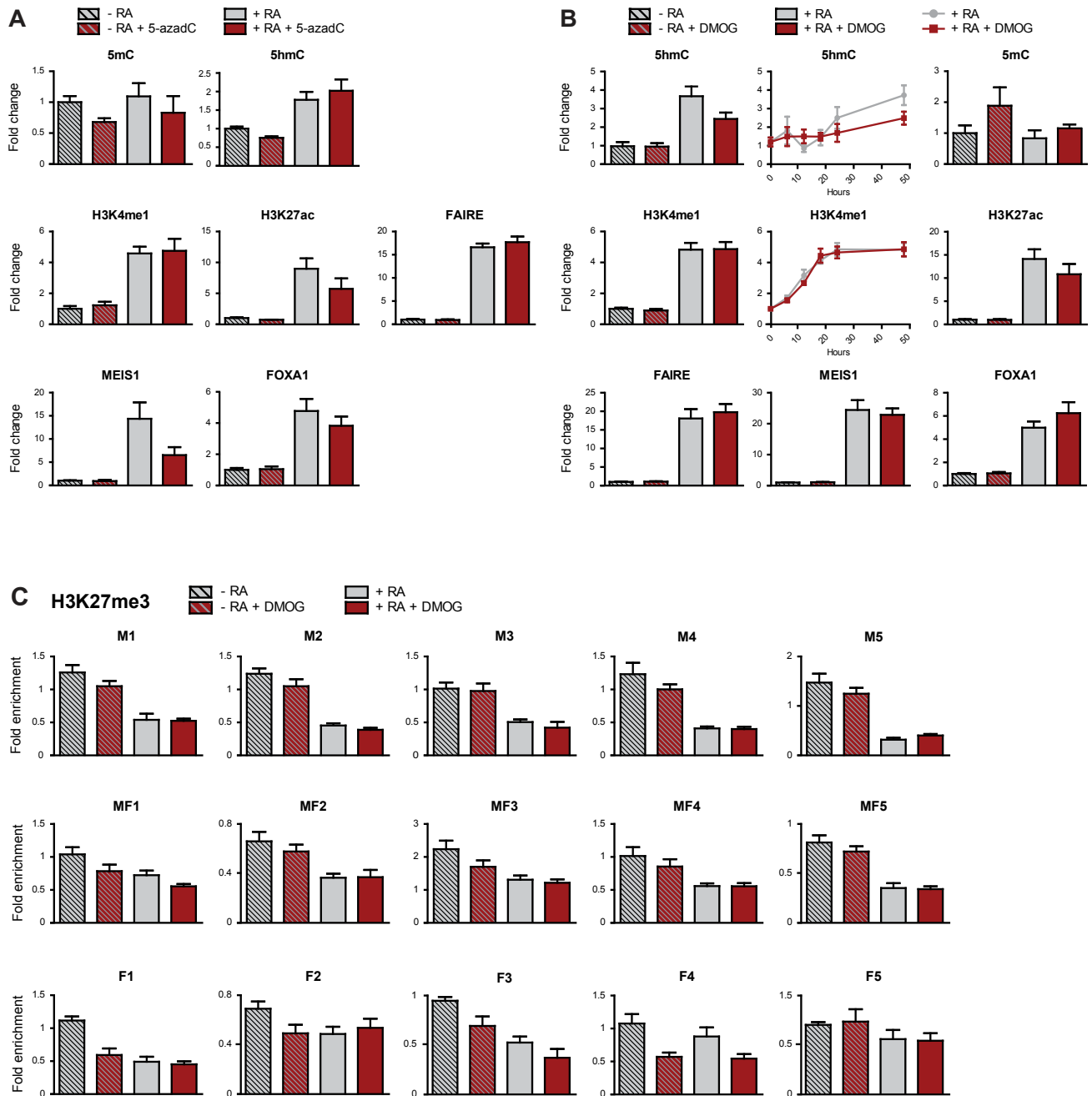
Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG. 2012. Primer3--new capabilities and interfaces. *Nucleic Acids Res* **40**: e115.

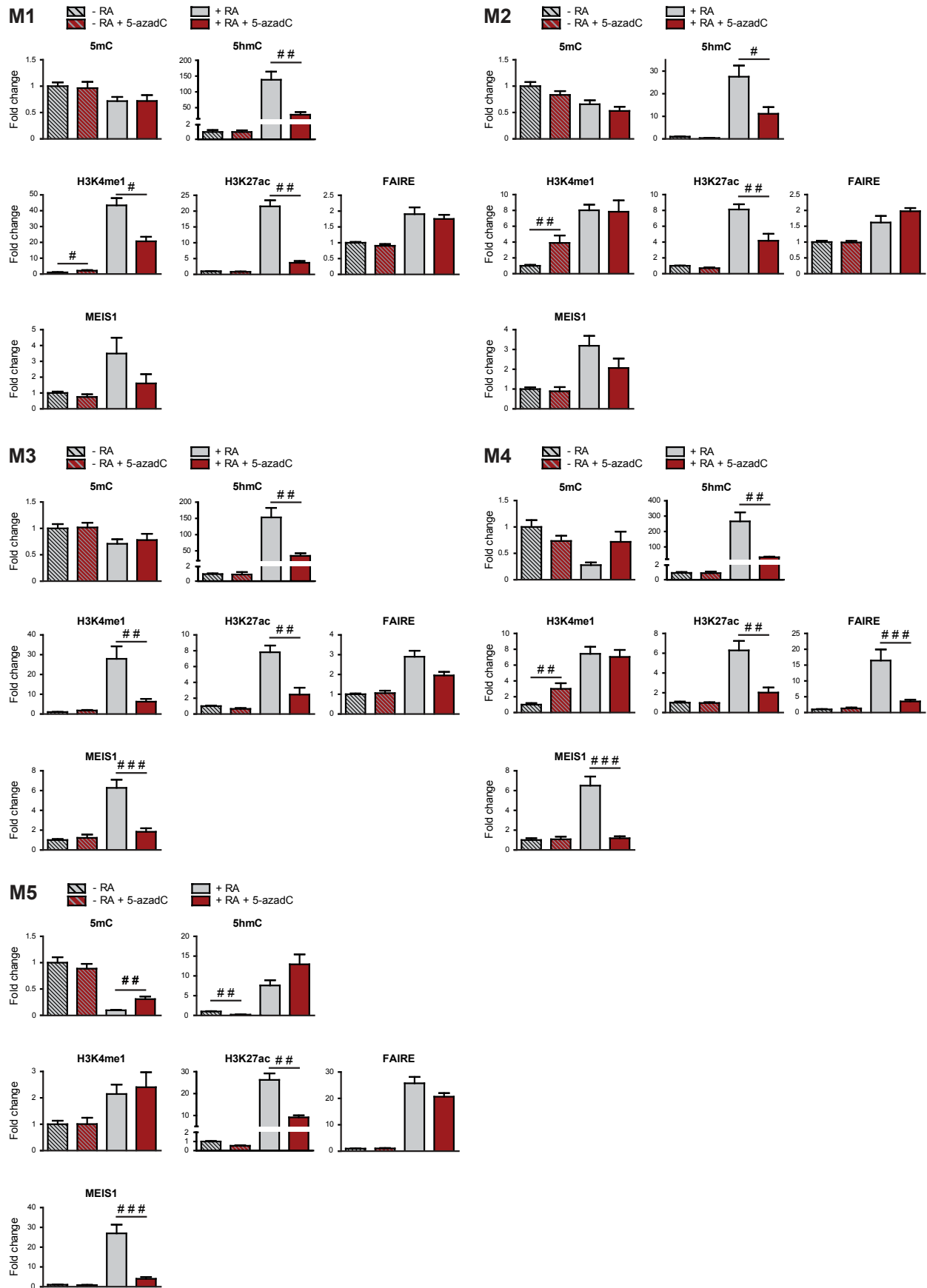


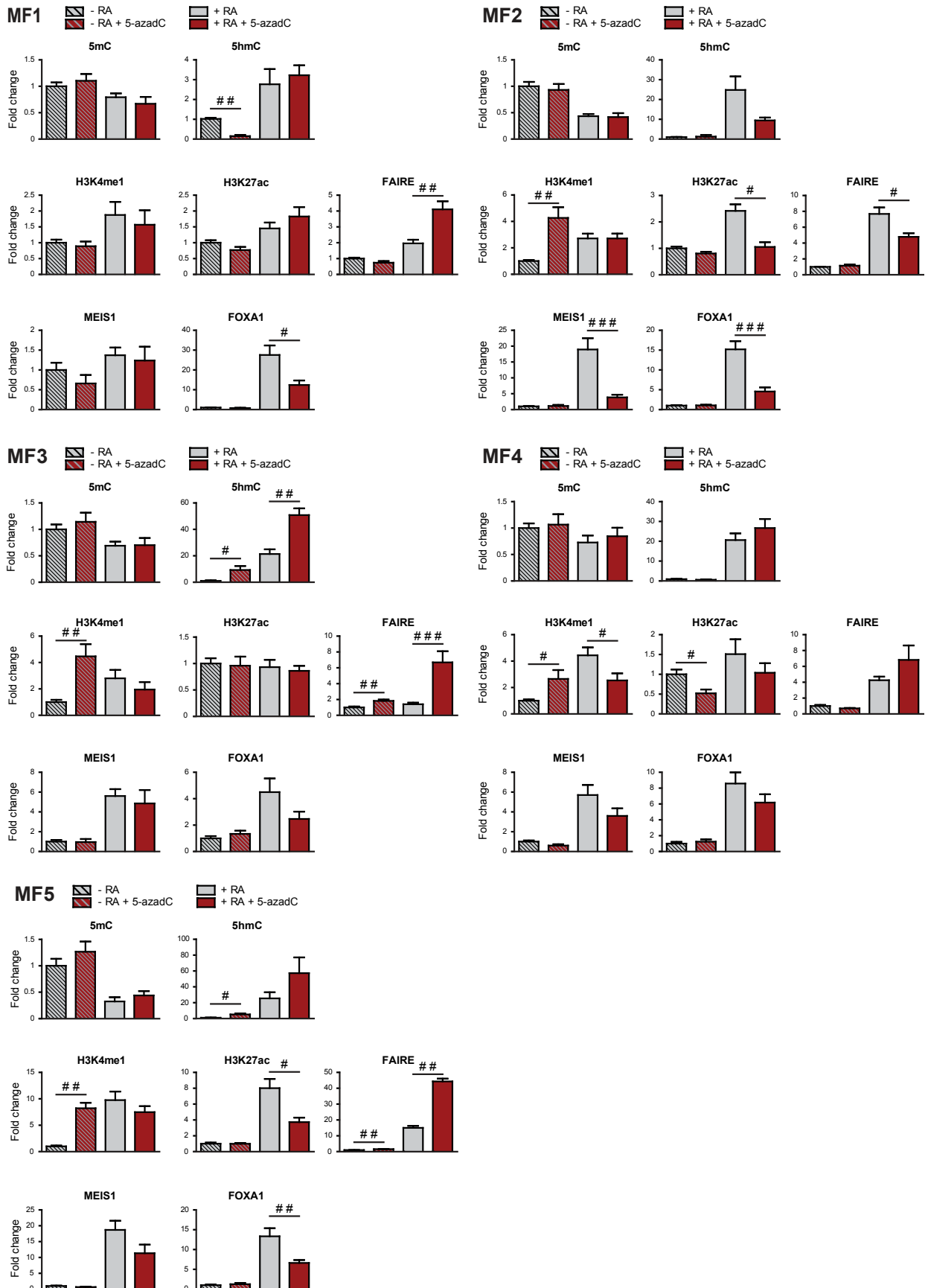


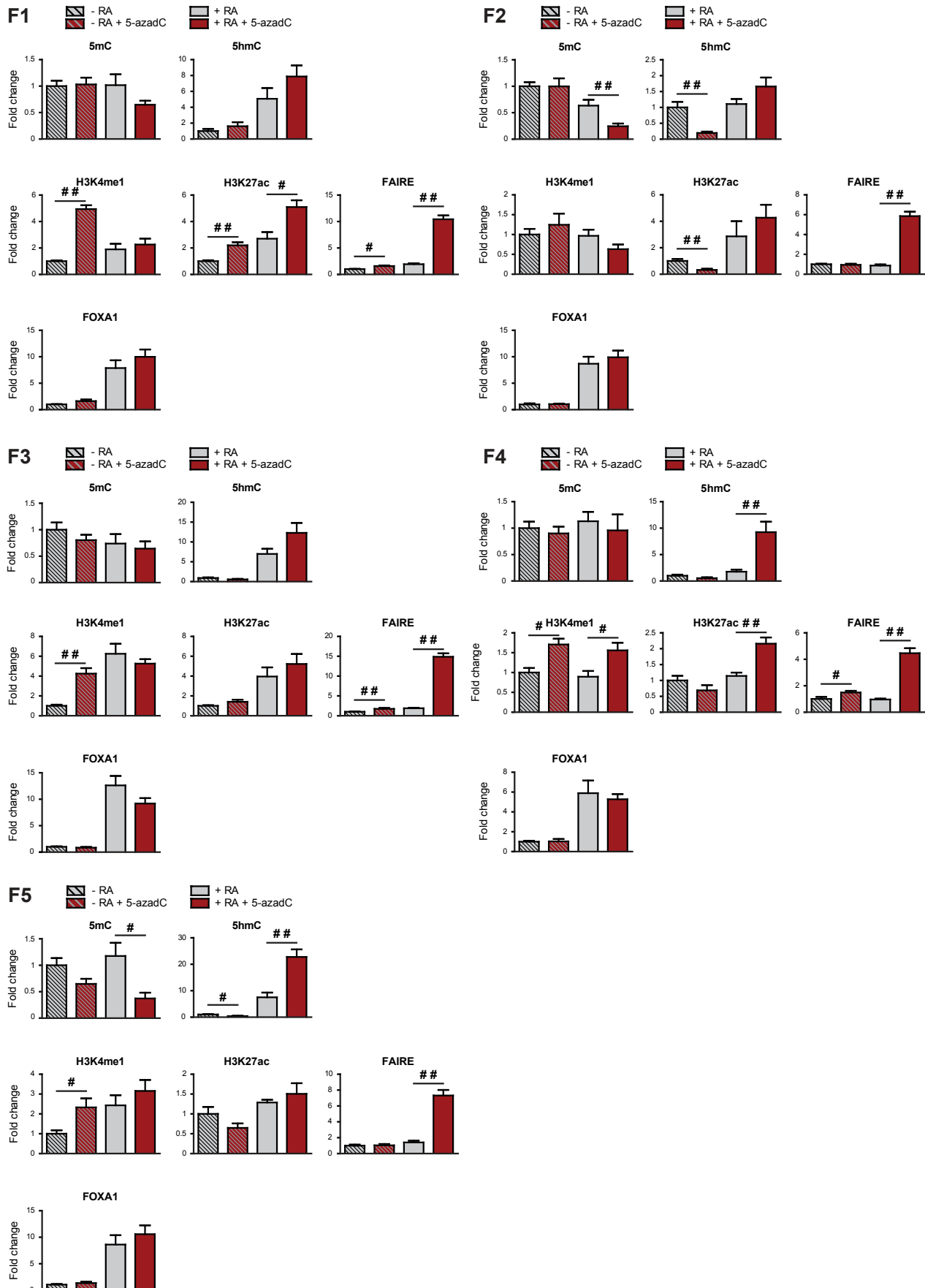




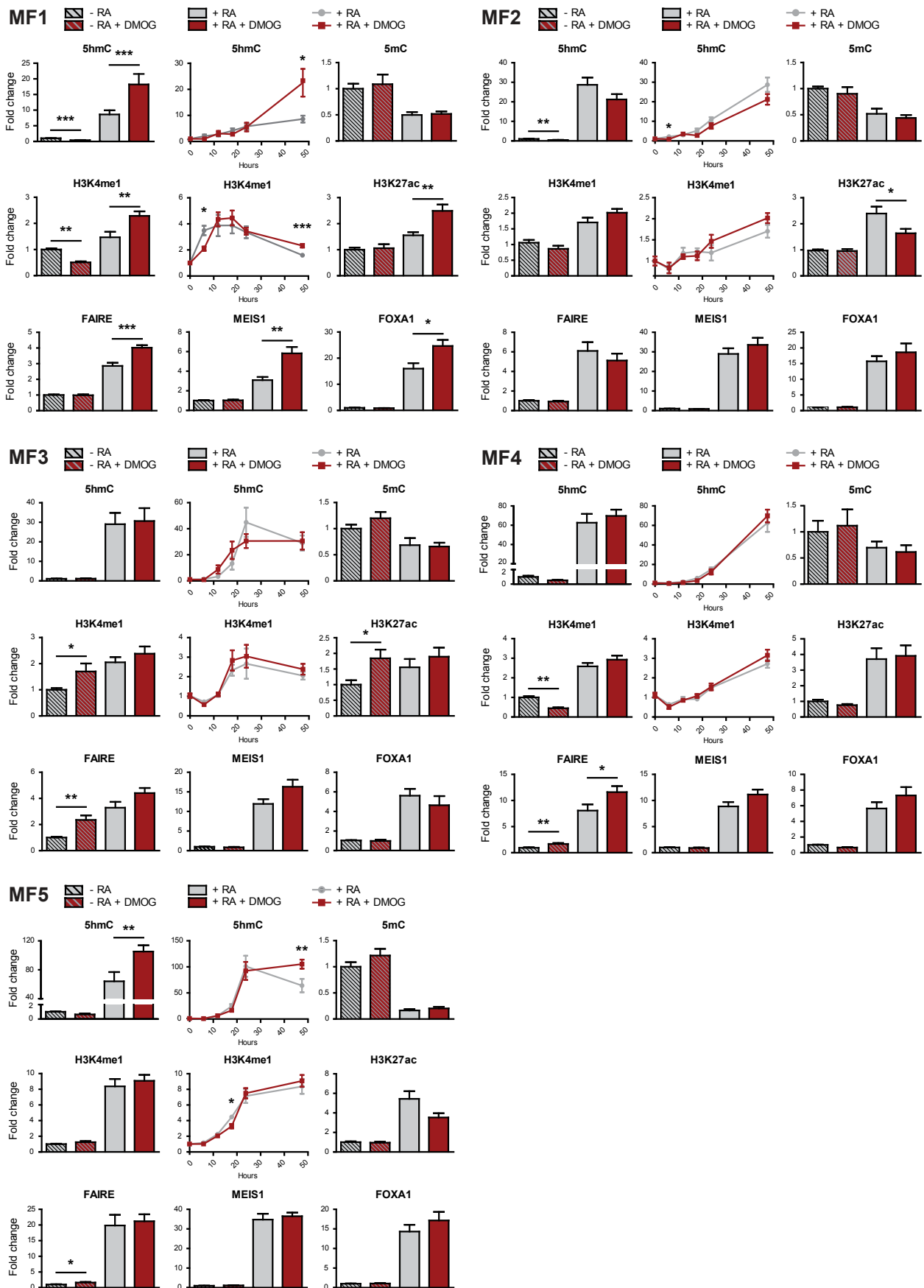


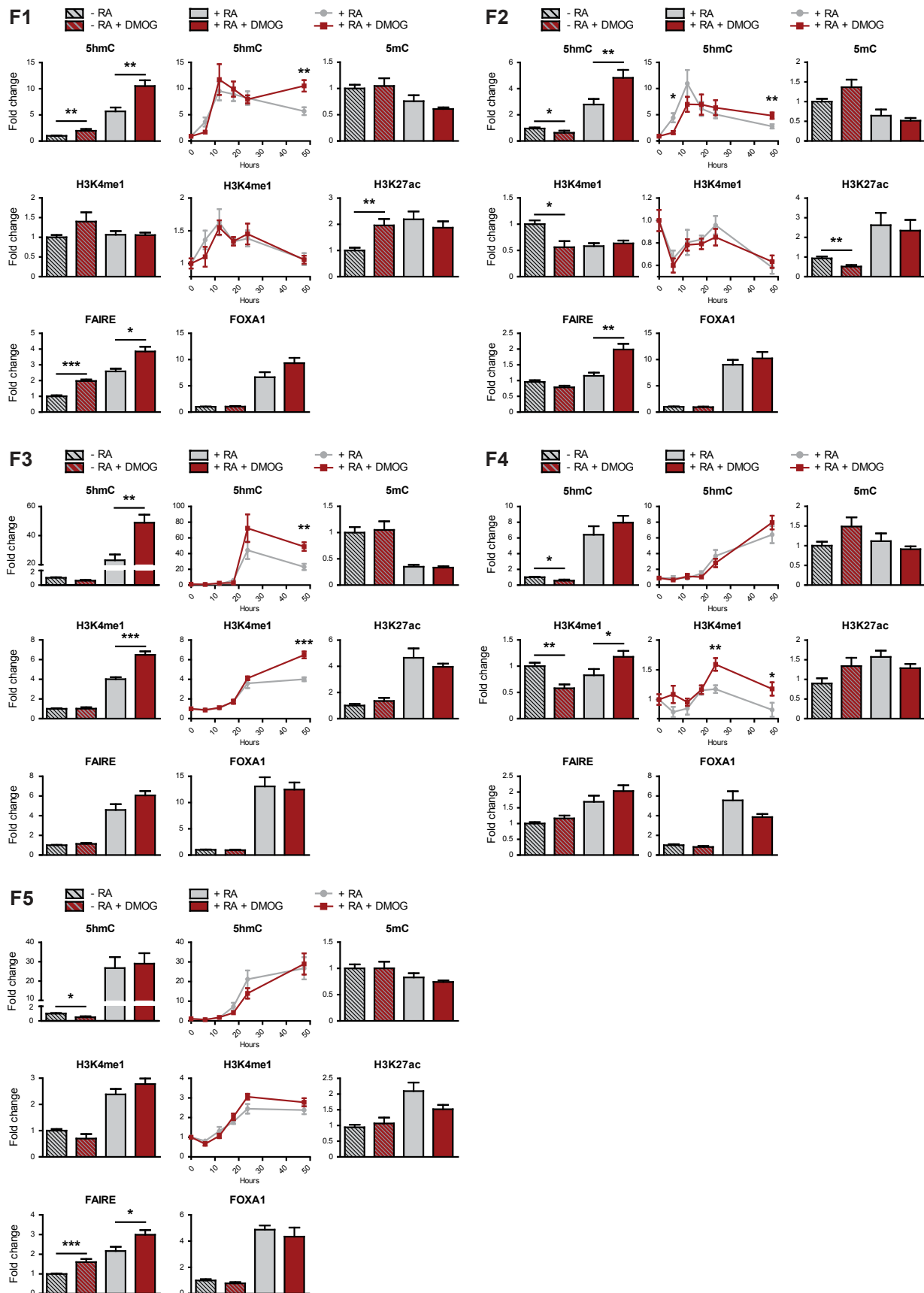


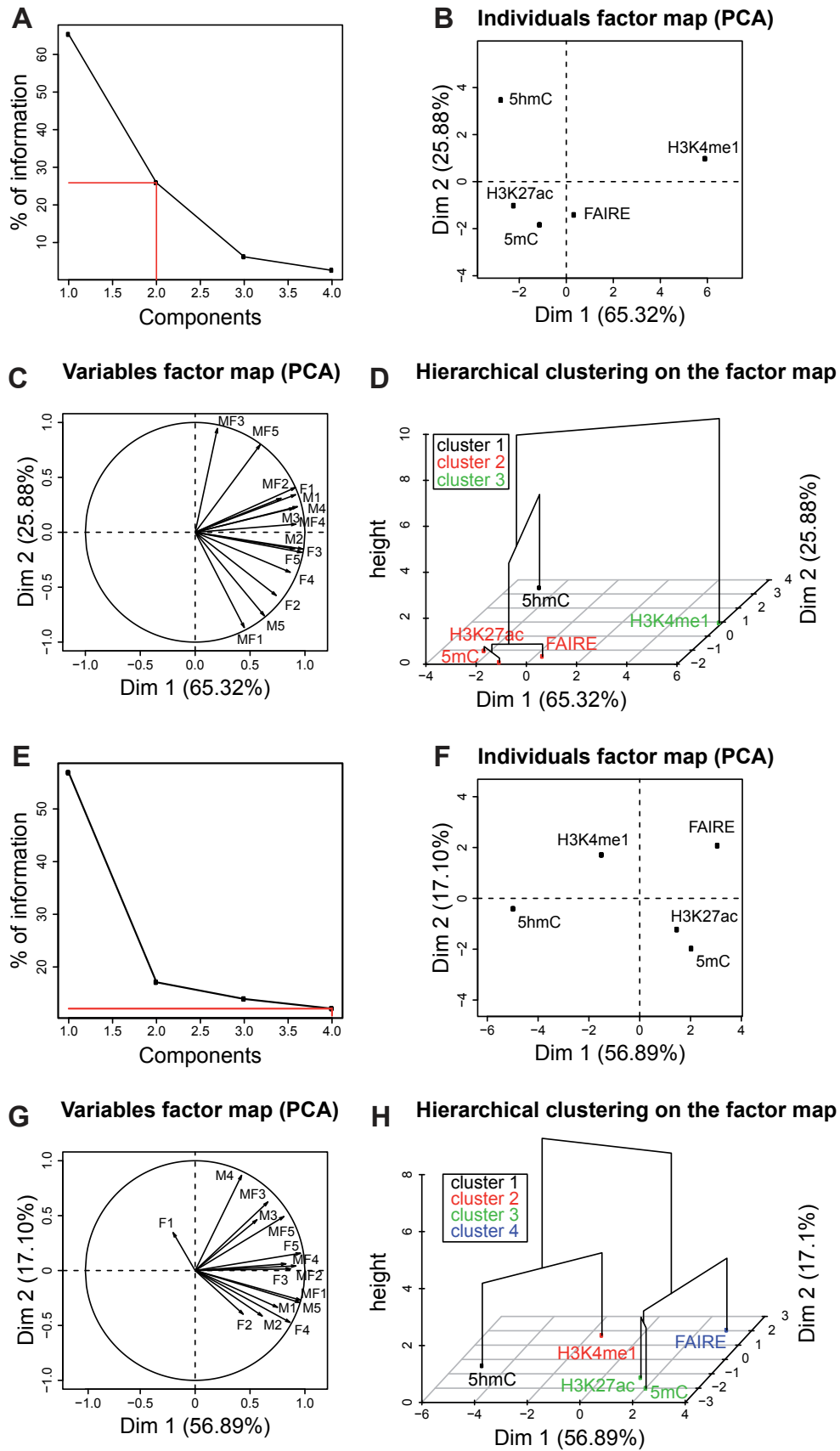


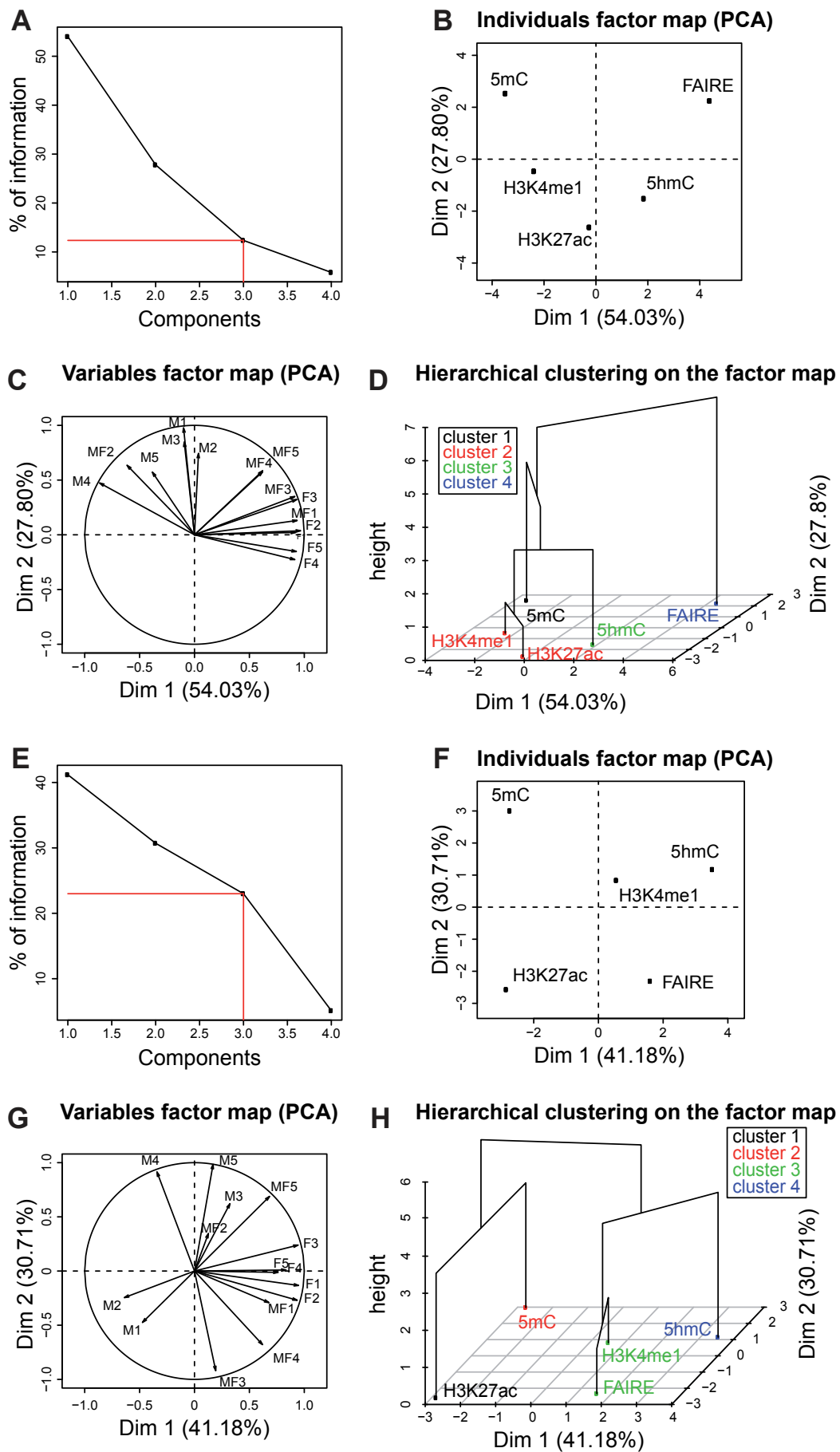


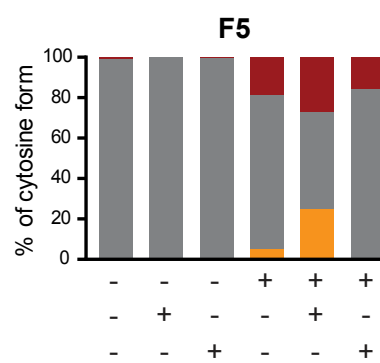
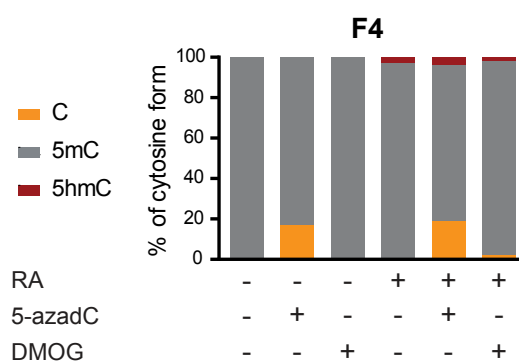
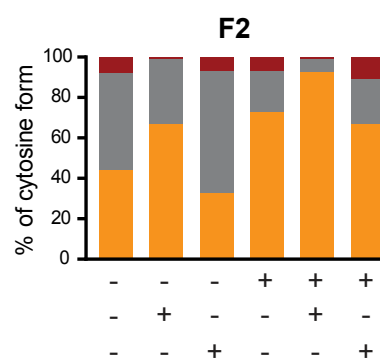
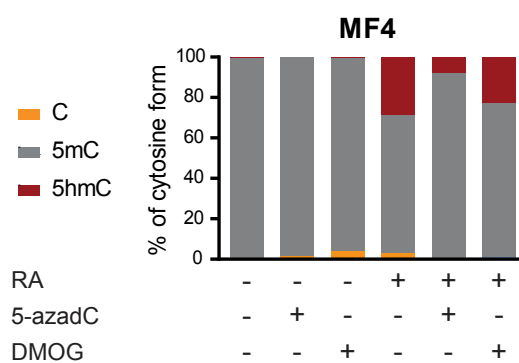
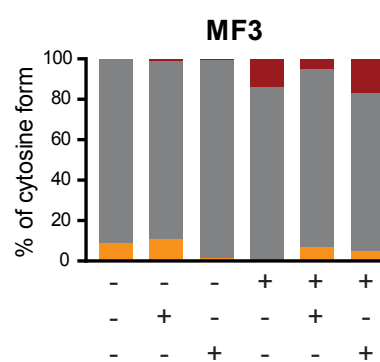
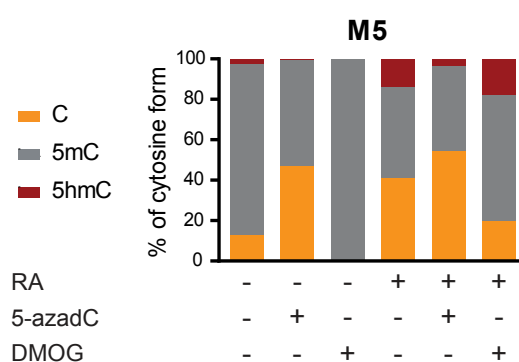
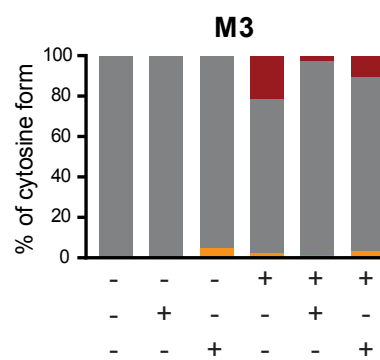
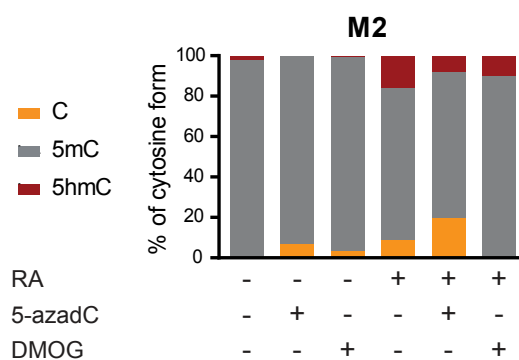


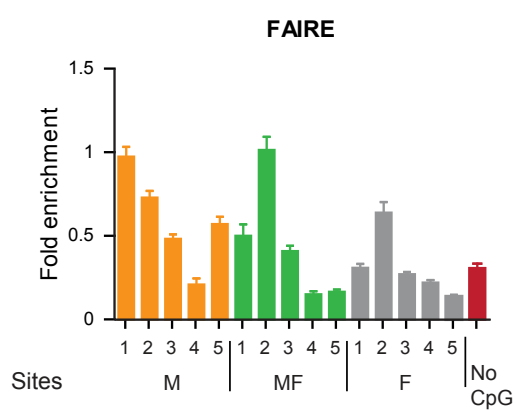
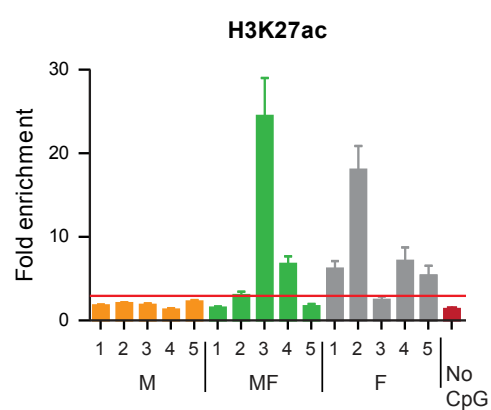
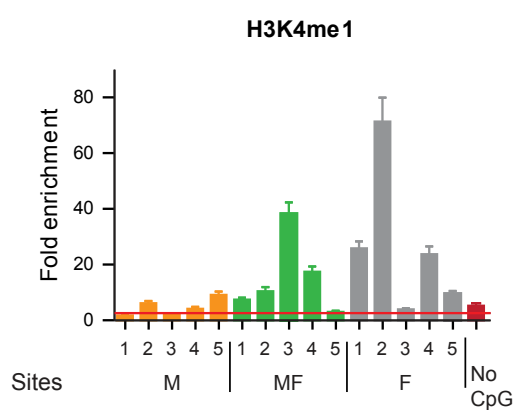
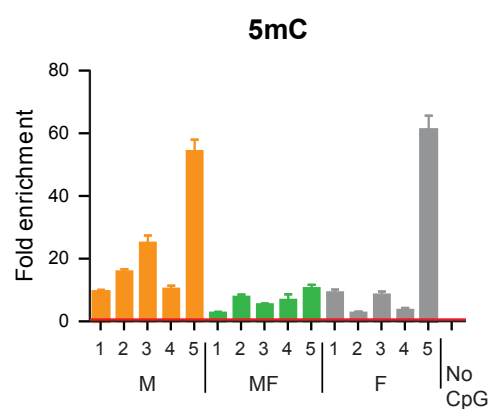
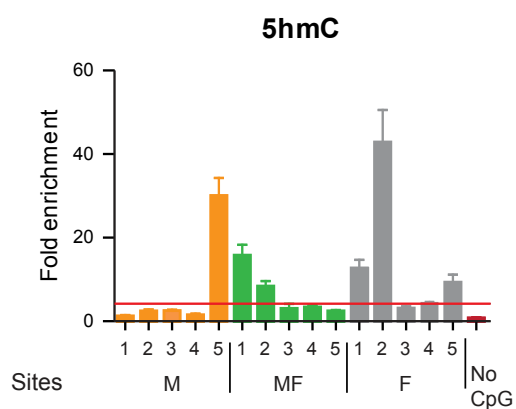


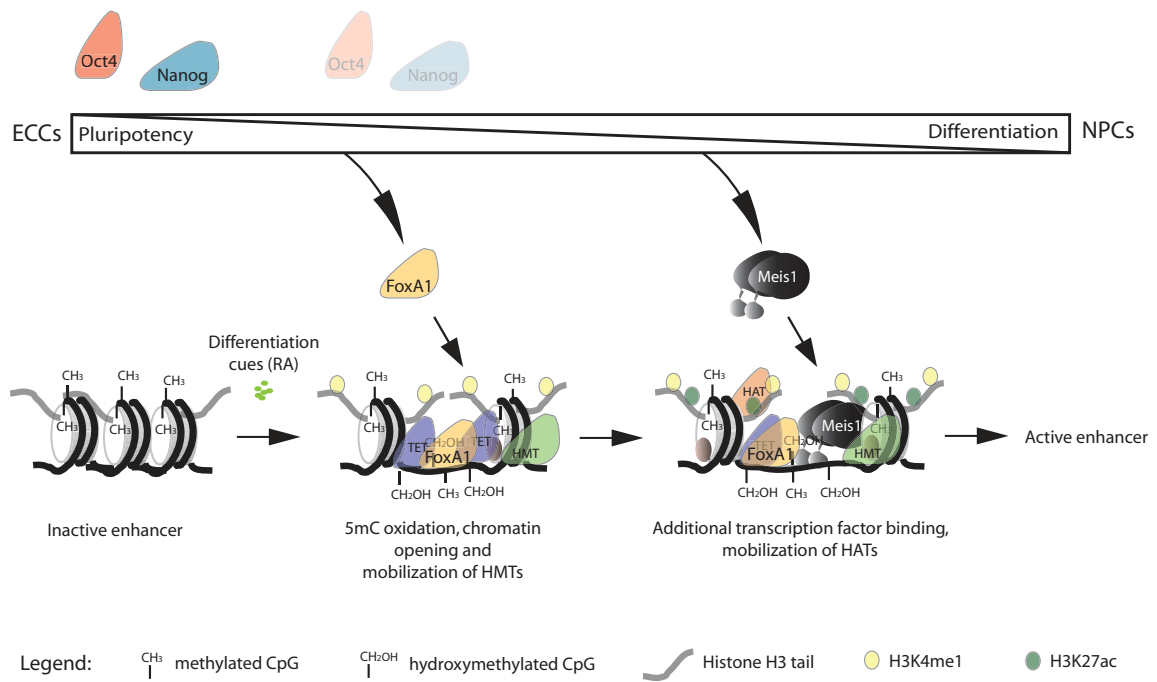












ChIP-MS_PBX1_mouse

Identified Proteins	Accession Number	Molecular Weight	Sample	Exclusive Unique Peptide Count	% Coverage
Pre-B-cell leukemia transcription factor 1 OS=Mus musculus GN=Pbx1 PE=1 SV=2	PBX1_MOUSE	47 kDa	PBX1_rep1	19	48%
			PBX1_rep2	15	40%
			PBX1_rep3	20	51%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Pre-B-cell leukemia transcription factor 2 OS=Mus musculus GN=Pbx2 PE=1 SV=1	PBX2_MOUSE	46 kDa	PBX1_rep1	14	62%
			PBX1_rep2	12	52%
			PBX1_rep3	15	63%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Meis1 OS=Mus musculus GN=Meis1 PE=1 SV=1	MEIS1_MOUSE	43 kDa	PBX1_rep1	15	54%
			PBX1_rep2	13	44%
			PBX1_rep3	14	49%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Pre-B-cell leukemia transcription factor 3 OS=Mus musculus GN=Pbx3 PE=2 SV=1	PBX3_MOUSE	47 kDa	PBX1_rep1	8	50%
			PBX1_rep2	7	31%
			PBX1_rep3	9	50%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Hox-A5 OS=Mus musculus GN=Hoxa5 PE=1 SV=1	HXA5_MOUSE	29 kDa	PBX1_rep1	10	49%
			PBX1_rep2	5	38%
			PBX1_rep3	10	49%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Meis2 OS=Mus musculus GN=Meis2 PE=1 SV=2	MEIS2_MOUSE	52 kDa	PBX1_rep1	4	25%
			PBX1_rep2	4	25%
			PBX1_rep3	5	30%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-

Mahé_Suppl Table1, continued

Homeobox protein PKNOX2 OS=Mus musculus GN=Pknnox2 PE=2 SV=1	PKNX2_MOUSE	52 kDa	PBX1_rep1	2	3.6%
			PBX1_rep2	2	3.6%
			PBX1_rep3	3	11%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Hox-B5 OS=Mus musculus GN=Hoxb5 PE=2 SV=3	HXB5_MOUSE	29 kDa	PBX1_rep1	3	25%
			PBX1_rep2	2	15%
			PBX1_rep3	3	25%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Hox-B6 OS=Mus musculus GN=Hoxb6 PE=2 SV=2	HXB6_MOUSE	25 kDa	PBX1_rep1	1	12%
			PBX1_rep2	1	8%
			PBX1_rep3	2	19%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Hox-B8 OS=Mus musculus GN=Hoxb8 PE=2 SV=2	HXB8_MOUSE (+1)	28 kDa	PBX1_rep1	0	3.3%
			PBX1_rep2	3	13%
			PBX1_rep3	1	8.2%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Hox-C5 OS=Mus musculus GN=Hoxc5 PE=2 SV=3	HXC5_MOUSE	25 kDa	PBX1_rep1	1	21%
			PBX1_rep2	1	15%
			PBX1_rep3	2	27%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-
Homeobox protein Hox-C4 OS=Mus musculus GN=Hoxc4 PE=2 SV=1	HXC4_MOUSE	30 kDa	PBX1_rep1	2	16%
			PBX1_rep2	1	13%
			PBX1_rep3	1	2.7%
			IgG_rep1	0	-
			IgG_rep2	0	-
			IgG_rep3	0	-

RESULTATS | PARTIE 2

PARTIE II – DYNAMIQUE CHROMATINIENNE DES ENHANCERS HYDROXYMETHYLES AU COURS DE LA DIFFERENCIATION NEURALE

OBJECTIF

Actuellement, la majorité des études visant à définir les programmes transcriptionnels impliqués dans la différenciation cellulaire comparent des données obtenues au stade indifférencié avec des données correspondant au stade différencié. Cependant, des expériences de cinétiques réalisées lors de la différenciation des cellules P19 sur quelques enhancers choisis ont révélés l'existence de plusieurs dynamiques d'activation de ces régions (Sérandour *et al.*, 2012). Ainsi, l'établissement de tels programmes d'expression semble mettre en œuvre des dynamiques diverses d'engagement des enhancers lors de la différenciation de cellules pluripotentes en progéniteurs neuraux. Afin d'identifier ces différentes dynamiques à l'échelle du génome, nous avons entrepris de cartographier les marques caractéristiques des enhancers en cinétique au cours des 48 premières heures de différenciation des cellules P19. Ces travaux, en cours de réalisation, ont pour objectif principal de décrypter les premières étapes de la différenciation neurale en abordant différents aspects de cette dynamique. L'intégration de la dynamique chromatinienne avec l'enrichissement en motif pour des FT au sein des régions identifiées et les gènes potentiellement associés vont permettre notamment de caractériser:

- Les enhancers activés de façon transitoire au cours des premières heures de différenciation pouvant amener à l'identification des gènes importants pour enclencher le processus de différenciation.
- Les enhancers activés plus tardivement. Dans ce cas, sont-ils déjà dans un état « primed » dans les cellules indifférenciées ou le deviennent-ils au cours de la différenciation, et si oui quand?
- Les enhancers qui s'inactivent au début de la différenciation
- Les régions enrichies en H3K27me3 afin de déterminer si l'état « poised » est général ou concerne particulièrement certains groupes d'enhancers.
- L'implication de la 5hmC dans les différentes dynamiques d'activation des enhancers en observant sa stabilité ainsi que les corrélations avec les autres modifications.
- La contribution des FP aux mécanismes de différenciation.

MATERIELS ET METHODES

Culture cellulaire

Les cellules de carcinome embryonnaire de souris P19.6 ont été cultivées dans du milieu DMEM (« Dulbecco's Modified Eagle's Medium ») contenant 10 % de sérum de veau fœtal (GIBCO) et différenciées avec 1 μ M d'acide rétinoïque « all *trans* » durant 6, 12, 18, 24 ou 48 heures.

Immunoprécipitation de la chromatine et de l'ADN

Les expériences de ChIP ont été réalisées comme précédemment décrit par Schmidt *et al.*, 2009. Les anticorps utilisés ciblent les marques de chromatine H3K4me1 (ab8895, Abcam), H3K27ac (ab4729, Abcam), H3K27me3 (07-449, Upstate), H3K9me3 (07-442, Millipore) ou les FT FOXA1 (ab5089, Abcam) et MEIS1 (ab19867, Abcam). Les protocoles concernant l'immunoprécipitation d'ADN méthylé ou hydroxyméthylé sont disponibles sur le site internet d'epigenesys (<http://www.epigenesys.eu/en/protocols>): « Methylated DNA Immunoprecipitation (MeDIP) » (Protocole 33 – Michaël Weber et Dirk Schübeler) et « Hydroxymethylated DNA Immunoprecipitation (hMeDIP) in mammalian cells » (Protocole 64 – Elise Mahé, Gilles Salbert). Les anticorps utilisés pour ces expériences ciblant respectivement la 5mC (Mab-081-500) et la 5hmC (CS-HMC-100) proviennent de chez Diagénodé.

Formaldehyde Isolation of Regulatory Element (FAIRE)

Les expériences de FAIRE ont été effectuées comme préalablement décrit (Eeckhoute *et al.*, 2009).

Séquençage

Les banques d'(h)MeDIP, ChIP et FAIRE ont été préparées avec le TruSeq ChIP Sample Prep Kit (Illumina, ref. IP-202-1012), puis séquencées par le système HiSeq selon les instructions du fabricant. L'alignement des reads, la conversion des reads extraits en fichiers signaux (.wig) et le « peak calling » ont été réalisés comme précédemment décrit (Sérandour *et al.*,

2012). Les régions « blacklistées » fournies par l'UCSC (« University of California, Santa Cruz » – <http://genome.ucsc.edu>) ont été supprimées de chaque fichier signal.

Identification des régions enhancer

L'identification de l'ensemble des régions enhancer « primed » ($5\text{hmC} \cap \text{H3K4me1} - \text{H3K27ac}$) ou actifs ($5\text{hmC} \cap \text{H3K4me1} \cap \text{H3K27ac}$) présentes au cours de la cinétique de différenciation des cellules P19 a été réalisée en effectuant, pour chaque temps (0, 6, 12, 18, 24 et 48 heures) l'union et/ou l'exclusion des marques d'intérêts ($5\text{hmC} \cap \text{H3K4me1}$ (– or $\cap \text{H3K27ac}$)). Nous avons par la suite effectué l'union de ces régions correspondant aux enhancers à tous les temps puis retiré les régions promotrices (situées entre – 5000 et + 1000 pb des TSS annotés) de ces fichiers. En parallèle, nous avons généré des fichiers contenant des régions choisies aléatoirement dans le génome (nommées par la suite « random ») de même taille et nombre que nos régions d'intérêt.

« Clusterisation »

Afin de générer des clusters de régions génomiques qui présentent une dynamique de 5hmC similaire, nous avons effectué une « clusterisation » par l'algorithme des *k-means*, notamment disponible sur la plateforme cistrome (<http://cistrome.org/ap/>; Liu *et al.*, 2011). Contrairement aux marques histones, la 5hmC est une marque localement peu étendue. Afin que les clusters ne soient pas construits sur la base de la distribution du signal pour chaque région génomique, nous avons utilisé l'algorithme sur des régions de 100 pb, centrées sur la médiane de l'ensemble de ces régions. De plus, l'intérêt des études réalisées en cinétiques étant d'étudier les différentes dynamiques de comportements entre régions, on ne s'intéresse pas à l'intensité du signal en lui-même, mais plutôt aux changements d'intensité de ce signal au cours de la cinétique. Nous avons donc utilisé ici la dérivée du signal, qui permet d'indiquer si le signal augmente ou diminue entre deux temps de cinétiques. Il est possible de générer cette dérivée en soustrayant, pour chaque position du signal, les deux valeurs de signaux obtenus à deux temps de la cinétique : $\Delta \text{Signal}_{t2-t1} = \text{Signal}_{t2} - \text{Signal}_{t1}$. Un signal différentiel aura donc une valeur positive si le signal augmente entre les deux temps, ou un signal négatif s'il diminue.

RESULTATS

Afin de caractériser les premières étapes de la différenciation neurale, nous avons entrepris d'analyser la dynamique chromatinienne associée aux enhancers. Pour cela, nous avons choisi de cartographier à l'échelle du génome, (i) la présence des modifications de cytosines 5mC et 5hmC, (ii) les marques de chromatines caractéristiques des enhancers H3K4me1 (enhancer « primed » et/ou actif) et H3K27ac (enhancer actif) ainsi que H3K27me3 et H3K9me3 (enhancer « poised »), (iii) l'ouverture de la chromatine, et (iv) le recrutement des FP FOXA1 et MEIS1. Nous avons ainsi réalisé des expériences de (h)MeDIP, ChIP et FAIRE en cinétique à partir d'échantillons prélevés à 6 temps différents de différenciation des cellules P19 (0, 6, 12, 18, 24 et 48 heures), suivie par un séquençage à haut débit.

A réception des données de séquençage, les statistiques concernant le nombre de reads mappés sur le génome ont révélé une hétérogénéité de la quantité de matériel séquencé, particulièrement importante pour H3K27ac (de 6 à 34 millions de reads ; cf. Tableau 4). L'analyse comparative des données obtenues aux divers temps ne peut se réaliser sans une normalisation préalable de ces données hétérogènes. C'est ainsi que nous avons développé et mis en place une nouvelle méthode de normalisation des données de séquençage présentée dans l'article « Calibrating and cleaning NGS signals from time-course studies help identify different dynamical behaviours of chromatin regions » en cours de préparation.

Une fois les données correctement normalisées, nous avons entrepris, dans un premier temps, d'identifier les enhancers activés au cours de la différenciation ($5\text{hmC}^+ \text{H3K4me1}^+ \text{H3K27ac}^+$; Figure 15) ainsi que ceux qui sont « primed » ($5\text{hmC}^+ \text{H3K4me1}^+ \text{H3K27ac}^-$; Figure 16). Ces enhancers ont par la suite été regroupés (« clusterisés ») en fonction de leurs variations en 5hmC au cours de la différenciation neurale (Figures 15A et 16A). L'établissement de profils moyens des dynamiques de 5hmC, H3K4me1, H3K27ac et FAIRE révèle que l'identification de clusters à partir des variations de 5hmC permet d'obtenir des sous-ensembles de régions génomiques possédant des dynamiques particulières des autres marques caractéristiques des enhancers (Figures 15B et 16B). De plus, les variations de chacune des marques ont tendance à être corrélées à celles de la 5hmC. Les profils moyens obtenus pour les niveaux d'ouverture de la chromatine (FAIRE) semblent peu dynamiques à l'échelle d'un cluster. Cependant, la visualisation des données dans IGB (« Integrated genome

browser ») à l'échelle de quelques régions montrent l'existence de dynamiques d'ouverture de la chromatine hétérogènes au niveau de régions regroupées à l'intérieur d'un même cluster, suggérant l'existence de comportements/mécanismes distincts. Il serait éventuellement intéressant de sous-clusteriser ces clusters en fonction des données du FAIRE. Afin de s'assurer de la spécificité des profils observés au niveau des clusters, nous avons également généré ces profils dans les régions « random » et n'observons aucun signal significatif (Figure 17).

CONCLUSION

Les premières données obtenues lors de ces travaux démontrent l'importance de l'utilisation d'une méthode appropriée de normalisation. Les analyses en cours devraient maintenant permettre de caractériser de façon exhaustive les remaniements chromatinien associés aux premières étapes de la différenciation des cellules P19.

CONTRIBUTIONS AUX DIFFERENTS RESULTATS

Contributions expérimentales

- Organisation et participation à la préparation des cultures cellulaires pour les expériences de cinétiques à l'échelle du génome.
- Expériences de MeDIP et hMeDIP en cinétique avant séquençage.
- Concernant le développement d'une nouvelle méthode de normalisation : la partie purement informatique a été entièrement réalisée par l'ingénieur bioinformaticien de l'équipe. J'ai, pour ma part, guidé l'évolution des scripts en testant les différentes hypothèses et méthodes avec nos données biologiques.
- Traitements et analyses bioinformatiques des données : application de la normalisation, « peak-calling », identification des régions enhancers « primed » ou actifs, clusterisation, profils moyens.

Contributions à l'élaboration de l'article

- Participation aux corrections de l'article en cours de préparation.

Tableau 4. Statistiques de séquençage.

5hmC	Nombre de read uniques mappés
T0	7,038.10 ⁶
T6H	7,060.10 ⁶
T12H	7,169.10 ⁶
T18H	8,610.10 ⁶
T24H	6,824.10 ⁶
T48H	6,236.10 ⁶

5mC	Nombre de read uniques mappés
T0	13,170.10 ⁶
T6H	10,756.10 ⁶
T12H	10,981.10 ⁶
T18H	10,348.10 ⁶
T24H	8,745.10 ⁶
T48H	9,458.10 ⁶

H3K4me1	Nombre de read uniques mappés
T0	19,638.10 ⁶
T6H	19,421.10 ⁶
T12H	17,113.10 ⁶
T18H	19,921.10 ⁶
T24H	20,198.10 ⁶
T48H	20,404.10 ⁶

H3K27ac	Nombre de read uniques mappés
T0	14,726.10 ⁶
T6H	9,891.10 ⁶
T12H	6,530.10 ⁶
T18H	8,094.10 ⁶
T24H	34,129.10 ⁶
T48H	17,035.10 ⁶

H3K27me3	Nombre de read uniques mappés
T0	20,623.10 ⁶
T6H	16,119.10 ⁶
T12H	19,405.10 ⁶
T18H	22,540.10 ⁶
T24H	18,963.10 ⁶
T48H	17,567.10 ⁶

H3K9me3	Nombre de read uniques mappés
T0	14,098.10 ⁶
T6H	7,770.10 ⁶
T12H	18,733.10 ⁶
T18H	20,411.10 ⁶
T24H	22,012.10 ⁶
T48H	18,031.10 ⁶

FAIRE	Nombre de read uniques mappés
T0	24,868.10 ⁶
T6H	19,354.10 ⁶
T12H	16,165.10 ⁶
T18H	18,149.10 ⁶
T24H	17,772.10 ⁶
T48H	17,411.10 ⁶

FOXA1	Nombre de read uniques mappés
T0	14,027.10 ⁶
T6H	12,345.10 ⁶
T12H	16,104.10 ⁶
T18H	11,010.10 ⁶
T24H	16,808.10 ⁶
T48H	15,903.10 ⁶

MEIS1	Nombre de read uniques mappés
T0	16,468.10 ⁶
T6H	14,109.10 ⁶
T12H	12,277.10 ⁶
T18H	17,854.10 ⁶
T24H	17,320.10 ⁶
T48H	16,793.10 ⁶

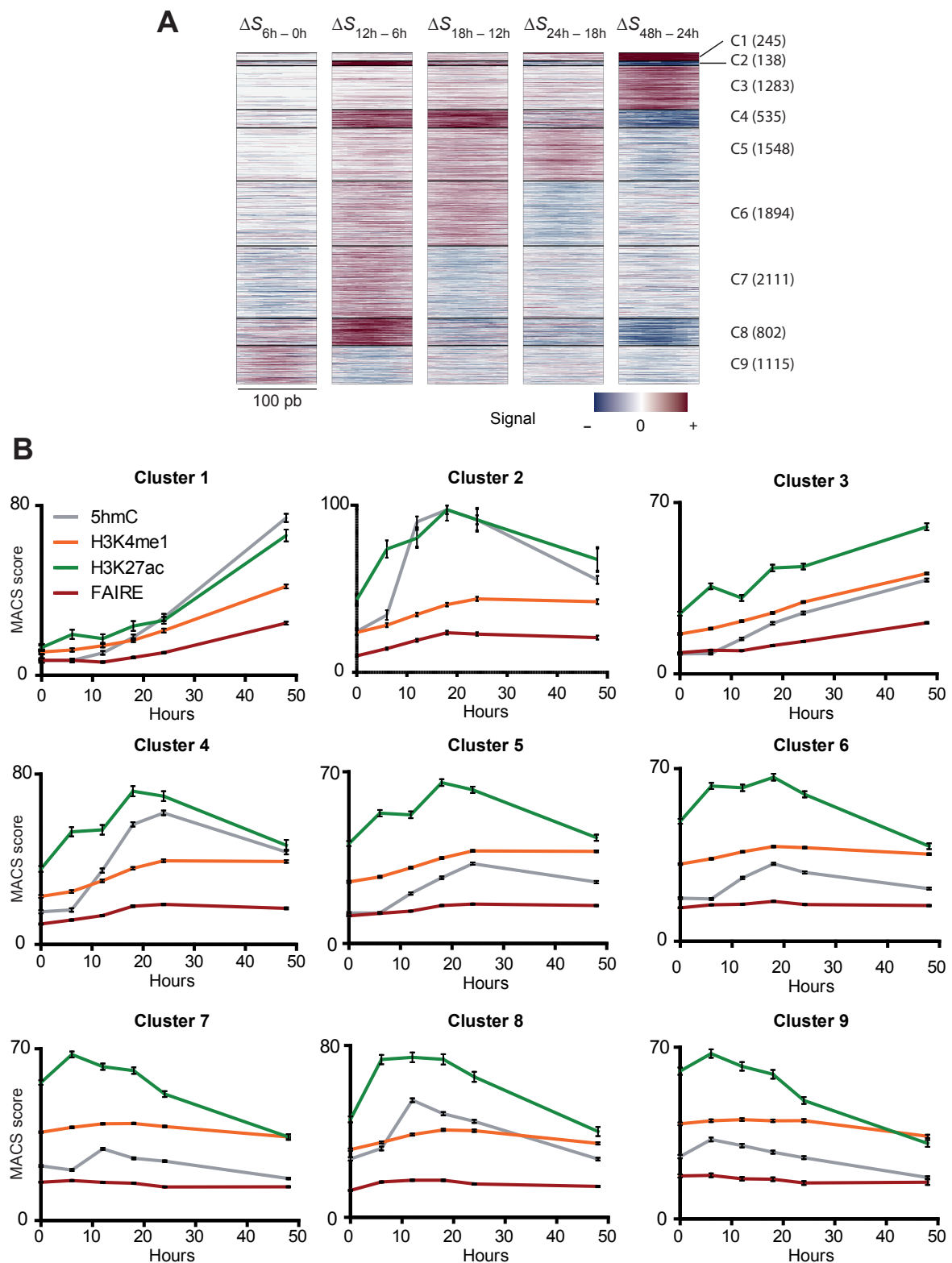


Figure 15. Dynamique chromatinienne des enhancers actifs au cours de la différenciation des cellules P19. (A) « Heatmap » résultant de la « clusterisation » des enhancers actifs *i.e.* 5hmC⁺ H3K4me1⁺ H3K27ac⁺, à partir du signal différentiel 5hmC. Le nom de chaque cluster est indiqué à droite de la « heatmap » : C=cluster ; suivi d'un numéro, avec indiqué entre parenthèses le nombre de régions génomiques pour chaque cluster. (B) Profils moyens d'enrichissement en signaux 5hmC, H3K4me1, H3K27ac et FAIRE au cours des 48 premières heures de différenciation des cellules P19, au niveau des régions des clusters 1 à 9.

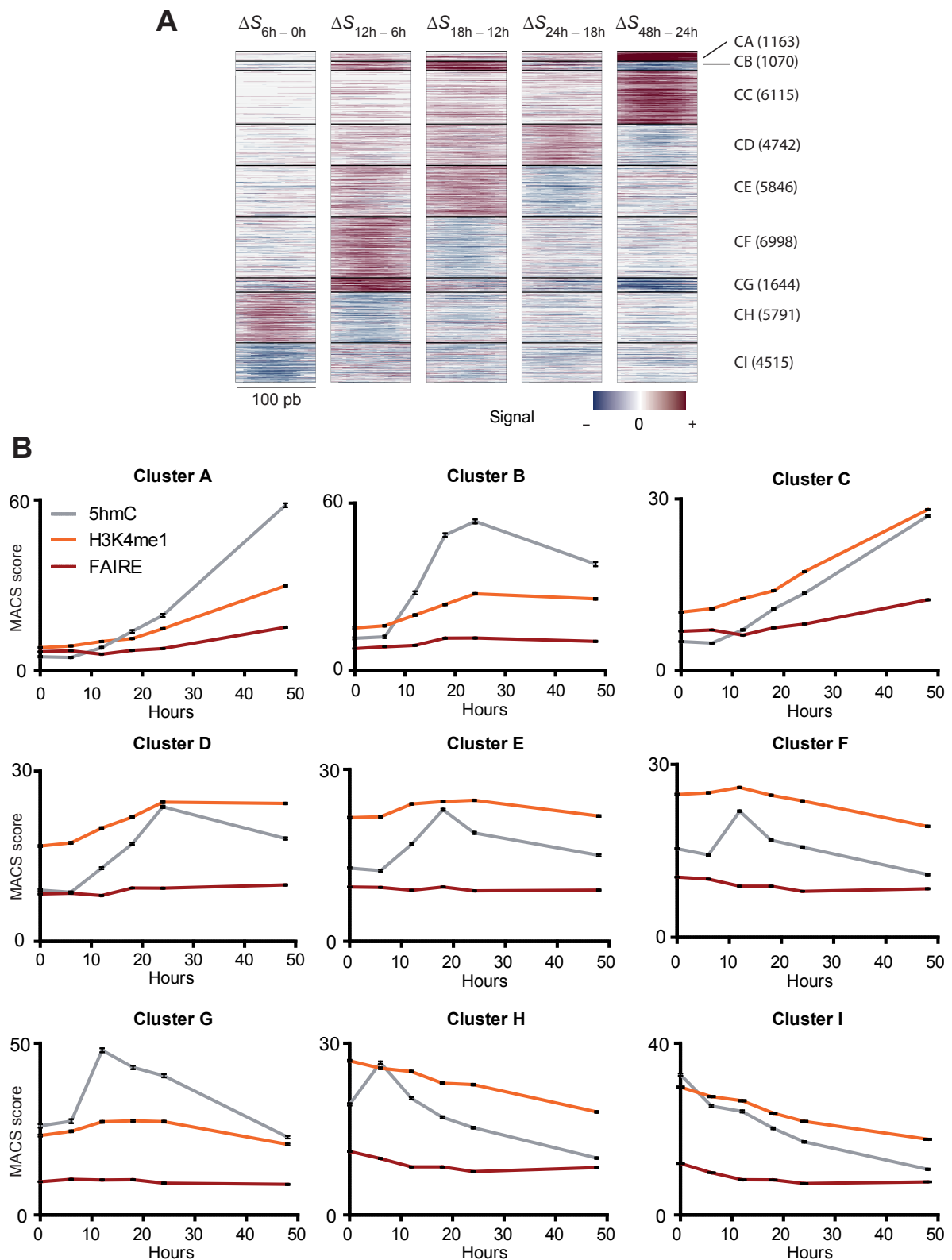


Figure 16. Dynamique chromatinienne des enhancers « primed » au cours de la différenciation des cellules P19. (A) « Heatmap » résultant de la « clusterisation » des enhancers « primed » *i.e.* 5hmC⁺ H3K4me1⁺ H3K27ac⁻, à partir du signal différentiel 5hmC. Le nom de chaque cluster est indiqué à droite de la « heatmap » : C=cluster ; suivi d’une lettre, avec indiqué entre parenthèses le nombre de régions génomiques pour chaque cluster. (B) Profils moyens d’enrichissement en signaux 5hmC, H3K4me1 et FAIRE au cours des 48 premières heures de différenciation des cellules P19, au niveau des régions des clusters A à I.

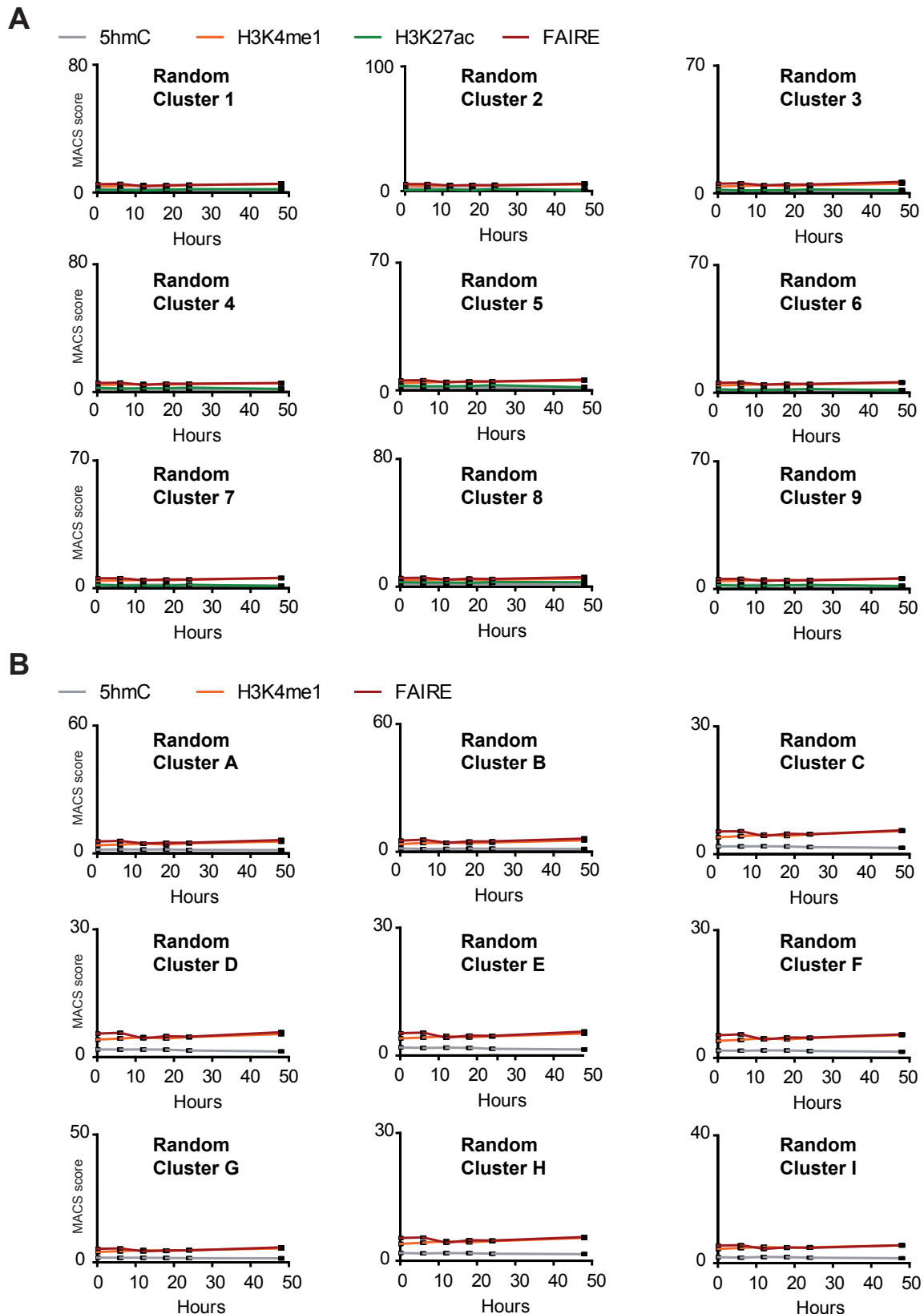


Figure 17. Contrôle de la spécificité des profils dynamiques observés.

Profils moyens d'enrichissement en signaux 5hmC, H3K4me1, FAIRE et H3K27ac au cours des 48 premières heures de différenciation des cellules P19, au niveau des régions random correspondant aux clusters d'enhancers actifs (1 à 9) (A) ou d'enhancers « primed » (A à I) (B).

Calibrating and cleaning NGS signals from time-course studies help identify different dynamical behaviours of chromatin regions

Abstract

Deciphering transcription regulatory principles requires the investigation of spatio-temporal correlations between genome-wide binding of transcription factors, presence of various epigenetic marks, and enrichments in specific genomic annotation features or DNA sequences. Hence, a deep mechanistic understanding of these principles begs for conducting large-scale kinetic studies using chromatin immuno-precipitation of the molecules of interest coupled to sequencing (ChIP-seq) of relevant biological samples at various time points. Yet, generally, ChIP-seq signals are not readily comparable, for they depend upon the quantity of material effectively available and the amount of noise within each sample, which may vary considerably. We propose here several bioinformatic protocols to allow a direct comparison of kinetic ChIP-seq signals and help identify clusters of genomic regions exhibiting different dynamical behaviours. Such protocols are required to resolve the temporal sequence of interactions involved in finely tuned processes occurring at the level of chromatin and will promote the implementation of currently under-represented kinetic ChIP-seq studies.

Introduction

Biologists currently use ChIP-seq technology to obtain comprehensive genome-wide maps of the distribution of proteins and/or epigenetic marks like cytosine modifications in DNA and histone post-translational modifications. Still, gaining a deeper understanding of the molecular mechanisms actually involved in a biological process may require generating several such maps in various conditions, or at various time points. However, the signals

obtained using ChIP-seq may vary greatly in quantity or quality sample to sample, depending for instance on the quantity of purified material used to generate sequencing libraries or on heterogeneous noise levels, possibly due to non-controlled events occurring during the implementation of these multistep protocols. Hence, signals obtained by ChIP-seq can only be compared once they have been normalized.

Various normalization methods exist^{1,2}; some are linear, preserving the shape of the signal, and some are non-linear. The method most commonly used applies a linear normalization with respect to the number of reads mapped to the reference genome and involves multiplying all signals by factors that even out their number of mapped reads (NMR). However, the NMR method exhibits limitations whenever signal distributions differ². Another methodology suggests adding exogenous DNA during the ChIP-seq procedure in order to provide an independent reference for calibration³, but this cannot rescue noisy samples. Other approaches are based on levelling the signal obtained within noisy background regions^{4,5,6,7}. The difficulty with the latter methods lies in the determination of those background regions. Such approaches have been assessed by evaluating the relevance of the genomic regions effectively determined by peak calling algorithms², which identify the coordinates of the peaks present within a genomic signal.

Yet, the sets of regions identified by peak calling do not constitute the only data impacted by the choice of the normalization method. Indeed, mean profiles of signals computed within given genomic regions, in different conditions or at various time points, will not generally be comparable if the signals are not properly normalized. Likewise, time-course studies, represented using graphs displaying the mean signal values of various factors at several time points, will also be impacted by the normalization method chosen. As an example, data from a time-course experiment conducted in our laboratory were originally

normalized with the NMR method, but produced biased graphs. Artefacts were clearly evidenced by the similar variations of signal observed in specific and random regions (Fig.1A-1B). It is noteworthy that the artefacts we observed affected signals with the lowest initial number of mapped reads (*i.e.* multiplied by the highest factor α - Table 1). Hence, multiplying by a large α applies to noise as well as signal, thus generating artefact signals. But the NMR normalization, because it accounts for noisy reads whose proportion may be heterogeneous among samples, impacts not only noise level but also genuine signals, as illustrated in Fig. 1C-1D, in cases where signal to noise ratios vary among samples. This suggests that the NMR normalization method is inadequate for such data.

Time-course studies actually constitute an ideal framework for testing normalization methods. In this paper, we present various protocols that have been employed to (i) normalize heterogeneous ChIP-seq signals into a set of comparable signals, allowing to (ii) draw calibrated mean profiles and kinetic curves.

Normalization of ChIP-seq signals

To remove artefacts from kinetic data, not only should high signals be calibrated, but noise levels also need to be made comparable. We thus propose a protocol (a) to fine-tune the multiplier factor α used to calibrate ChIP-seq signals by comparing only regions exempt of noise. Independently of this protocol, it is necessary to devise methods to deal with noise heterogeneity. Two methods are suggested here: (b) flattening the noise using a non-linear transformation of the signal, or (c) removing the noise.

(a) As calibration requires comparable signal distributions, a simple solution to calibrating several signals consists in selecting a subset of peaks corresponding only to genomic regions with a very high signal to noise ratio (thereafter called “noise-free”

regions). In signals with steady high peaks, *i.e.* related to housekeeping genes or stable levels of histone marks for instance, one may assume that the highest peaks have a very high signal to noise ratio and remain at similar levels on average. Obviously, this is not true for induced or repressed signals, which may be related to genes involved for example in developmental processes. An automatic procedure was designed to determine the value of threshold τ above which the signal may be considered as being noise-free (see Methods). Henceforth, to normalize *Signal1* to the level of *Signal2* using noise-free calibration (NFC), *Signal1* may be multiplied by:

$$\alpha_{\text{NFC}} = \frac{\bar{u}_{\text{high2}}}{\bar{u}_{\text{high1}}} \Big|_{\forall u > \tau},$$

where u is the signal value and $\bar{u}_{\text{high2}}$ the mean value of *signal2* at all genomic positions for which $u > \tau$. The artefacts exhibited by Figure 1C for instance were caused by heterogeneous signal to noise ratios between time points. Applying factor α_{NFC} instead of using the standard NMR factor $\alpha_{\text{NMR}} = \bar{u}_2/\bar{u}_1$ allowed us, by calibrating the signals independently of their inherent noise, to produce the graph displayed in Fig.3C.

(b) In order to minimize the noise while still linearly multiplying the highest signal values by a factor $\alpha = \alpha_{\text{NFC}}$, we defined and applied a non-linear transformation $s_{\alpha,c}(u)$, that combines the *softplus* and *sigmoid* functions (SSF) to transform all signal values u into values u' , given the factor α and an upper-noise threshold c , according to the expression:

$$u' = s_{\alpha,c}(u) = \ln \left[1 + e^{\alpha(u-pc)} \right] + \frac{\alpha pc}{1 + e^{-ku+qc}}, \quad u \in \mathbb{R},$$

where p , q and k are constants whose values determine the slope of the curve. It can be shown that $s_{\alpha,c}(u)$ flattens low signal values up to an arbitrarily fixed upper-noise threshold c while still multiplying high signal values by a factor α_{NFC} . The graph representing function

$s_{\alpha,c}(u)$ is displayed in Fig.2A and an example of signal transformed using this expression is illustrated in Fig.2B-2C. Although this function may be applied successfully to flatten the noise and multiply linearly the highest values, one must remain aware that intermediate values of the data are somehow transformed non-linearly and this fact can be of importance when generating mean profiles or kinetic graphs.

(c) The second approach to dealing with noise heterogeneity attempts to remove the noise before recalibrating the peaks. To our knowledge, noise removal¹⁰ has never been applied to process ChIP-seq signals. Here, we made use of a sound peak-calling algorithm¹¹ to retain only peaks (genomic regions) with significant signal, and remove the noise located in other regions (*i.e.* cleaning the signal). The peaks identified by the peak-calling algorithm cover a minimal number of consecutive positions whose signal lies above a predefined threshold τ . The minimal width necessary to define a peak has been set empirically to 40 bp (or 4 consecutive ChIP-seq positions for signals mapped every 10 bp). The identified regions were extended 110 bp (empirical value) on each side of the identified peaks, so as to keep the genuine low signals lying on the sides, which would otherwise be eliminated since they fall below the computed threshold. Only the original signal present within these identified and extended genomic regions was retained to generate a noise-free signal file (Supplemental Fig.1). This solution will hereafter be designated HPNR.

Displaying calibrated mean profiles and kinetic curves

Once data have been normalized, studying the dynamical behaviour of several marks within a single set or several clusters of genomic regions can be achieved by displaying kinetic curves representing the enrichments in transcription factors or epigenetic marks within

these clusters at various time points (Fig. 3A). The curves are constructed by performing the calculations in a specific order because the signal files are very large and insufficient planning may lead to carrying out inefficient computations in terms of time or memory. To maximize efficiency, we considered one signal at a time, and performed all the computations relative to that signal before releasing it from memory and processing another signal. Our algorithm thus concentrates on a single mark, taken at a single time point, at a time: it reads the signal, and then considers in turn all submitted clusters of genomic regions. Every region of every cluster is extended with respect to its median in both directions by a width of L bp (a parameter supplied by the user), all signal values covering this extended region are retrieved, and their mean or max value is computed. For each cluster, the mean of all these values is calculated together with its associated standard error, and all points are collected within text files, which are then used to draw the kinetic curves. To achieve this, an application, developed in Java and Python, (i) collects from the user, via a graphical interface, all the parameters and input files (signals, clusters) required, (ii) calculates the mean signal values and associated standard errors for each mark, at every time point, within each cluster of genomic regions, (iii) creates files storing all data, and eventually (iv) plots the kinetic curves.

Discussion

To our knowledge, few kinetic studies^{14,15} involving ChIP-seq experiments have been implemented, presumably because the ChIP-seq technology is only semi-quantitative, and thus requires crucial calibration steps to allow signals to be compared. The methods described here have been used to normalize our signals and successfully provided meaningful kinetic graphs. Yet, these methods and their empirically chosen parameters

should be applied onto other data sets for validation. We believe efforts to fine-tune the calibration of ChIP-seq signals need to be pursued, and should eventually help reveal temporal sequences governing regulatory processes.

Methods

The thresholds τ defining the noise-free and noise-removal levels of a signal have been evaluated automatically by constructing the histogram $h(u)$ of a signal (Supplemental Fig.2). The method is based on the assumption that if only noise were present, a continuous decrease in the number of occurrences $h(u)$ with increasing values u of the signal should be observed. On the opposite, the presence of a genuine signal symptomatically causes fluctuations in the histogram. These consecutively make derivatives change sign, implying that they can be used to determine thresholds. The subsequent derivatives are defined recursively by:

$$d^n h(u) = \frac{d^{n-1} h(u+1) - d^{n-1} h(u)}{u+1-u} = d^{n-1} h(u+1) - d^{n-1} h(u),$$

where $d^0 h(u) = h(u)$ is the number of genomic positions possessing signal value u , and $d^n h(u)$ is the n^{th} derivative of $h(u)$ defined recursively using the above expression. Now, the point u at which a given n^{th} -derivative changes sign:

$$d^n h(u) = 0$$

may be used to define a meaningful threshold. Empirically, we found that the third derivative changed sign at a value suitable to define the noise-free threshold, while the fourth derivative changed sign at a value suitable to define the noise removal threshold.

References

1. Bailey et al., Practical guidelines for the comprehensive analysis of ChIP-seq data, *PLOS Comput. Biol.* 2013, 9:11.
2. Angelini et al., Is this the right normalization? A diagnostic tool for ChIP-seq normalization, *BMC Bioinformatics* 2015, 16:150.
3. Orlando et al., Quantitative ChIP-seq normalization reveals global modulation of the epigenome, *Cell Reports* 2014, 9:1163.
4. Xu et al., A signal noise model for significance analysis of ChIP-seq with négative control, *Nat. Biotechnol.* 2008, 26:1199.
5. Ji et al., An integrated software system foa analysing ChIP-chip and ChIP-seq data, *Nat. Biotechnol.* 2008, 26:1293.
6. Liang & Keles, Normalization on ChIP-seq data with control, *BMC Bioinformatics* 2012, 13:199.
7. Diaz et al., Normalization, bias correction, and peak calling for ChIP-seq, *Stat. Appl. Gen. Mol. Biol.* 2012, 11.
8. Mahé et al., Cytosine modifications modulate the chromatin architecture of transcriptional enhancers, submitted.
9. Nicol et al., The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets, *Bioinformatics* 2009, 25:2730.
10. Rudin et al., Non-linear total variation based noise removal algorithms, *Physica D* 1992, 60:259.
11. Serandour et al., Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-asociated enhancers, *Nucleic Acid Res.* 2012, 40:8255.
12. Saeed et al., TM4: a free, open-source system for microarray data management and analysis, *Biotechniques* 2003, 34:374.
13. Liu et al., Cistrome: an integrative platform for transcriptional regulation studies, *Genome Biology* 2011, 12.
14. Lickwar et al., Genome-wide protein DNA binding dynamics suggest a molecular clutch for transcription factor function, *Nature* 2012, 484:251.
15. Siersbæk et al., Extensive chromatin remodelling and establishment of transcription factor hotspots during early adipogenesis, *The EMBO Journal*, 2011, 30:1459.

Figure Legends

Table 1. Number of reads in ChIP-seq data of Mark A, provided for every time point.

Figure 1. Kinetic curves showing the average signal value of mark A and B at various time points within several clusters of genomic regions. (A) All kinetic curves of mark A seem to exhibit artefacts at times t_2 and t_3 . (B) To verify this hypothesis, random regions were generated and kinetic curves, showing the average signal value of mark A at various time points within those regions, were constructed. The curves, by exhibiting peaks at times t_2 and t_3 while there should be no enrichment, validated the presence of artefacts. (C) All kinetic curves of mark B seem to exhibit artefacts at time t_3 . (D) Random regions do not exhibit peaks at times t_3 , indicating that there is no artefactual noise enrichment, but rather the use of a faulty normalization procedure.

Figure 2. Graphs and genomic maps illustrating properties of the *softplus-sigmoid* function $s(x)$. (A) The graphs show that function $s(x)$ is linear for high values of x , compresses the signal for low values of x and smoothly joins both extremities. The graphs have been generated for various values of parameters α and c (red: $\alpha=5$, $c=6$; green: $\alpha=1.2$, $c=6$; blue: $\alpha=1$, $c=20$). The smoothness of the joining is determined by the predefined values of constants: here, values $k=0$, $p=1.0$, $q=0.5$ were used. Graphs were produced on web site: <https://rechneronline.de/function-graphs/>. (B) Normalization is required to compare the original signal of mark A at time point t_2 shown in third row with signals at other time points. Normalization with respect to the number of mapped reads (NMR) consisted here in multiplying the signal by a factor 5.18, resulting in signal shown on second row. Softplus-

sigmoid function (SSF) normalization was also applied and resulted in signal shown on first row. (C) Zoomed view of the baselines of the above three rows. Notice how noise is multiplied by NMR normalization, and compressed by the SSF method.

Figure 3. Kinetic curves showing the average signal value of marks *A* and *B* at various time points within several clusters of genomic regions. (A) No strong artefact seems to occur within the kinetic curves of mark *A*. (B) Random regions were generated and exhibit no artefacts when displaying their average signal value of mark *A* at various time points. (C) No strong artefact seems to occur within the kinetic curves of mark *B*. (D) Random regions were generated and exhibit no artefacts when displaying their average signal value of mark *B* at various time points.

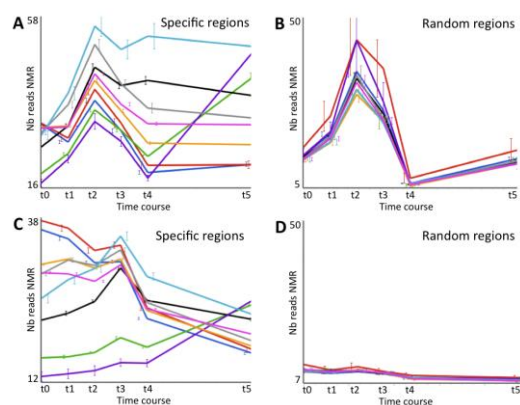
Suppl_Fig1. Integrated Genome Browser view illustrating the cleaning of a ChIP-seq signal. Top track shows the original signal, while lowest track displays the same signal after cleaning, where its noise, low and narrow peaks have been removed.

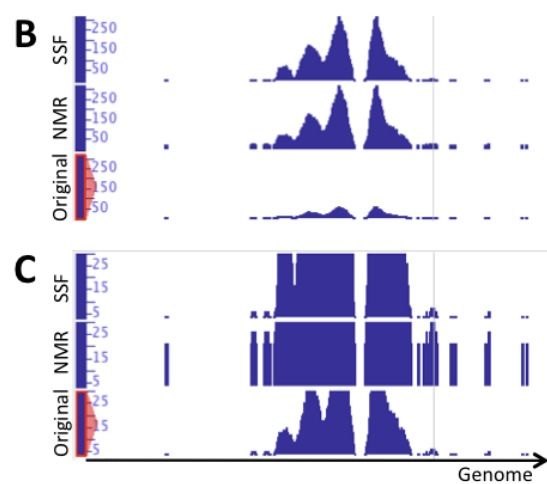
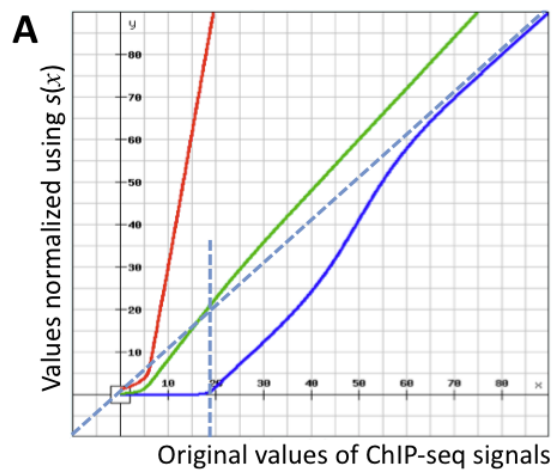
Suppl_Fig2. Graphical representation of a ChIP-seq signal histogram, giving the number of genomic positions exhibiting each signal value. The curve is zoomed to show the fluctuations present in the number of positions for high values, suggestive of genuine signal (in contrast to noise), and detectable using subsequent derivatives.

Mahé_Table1

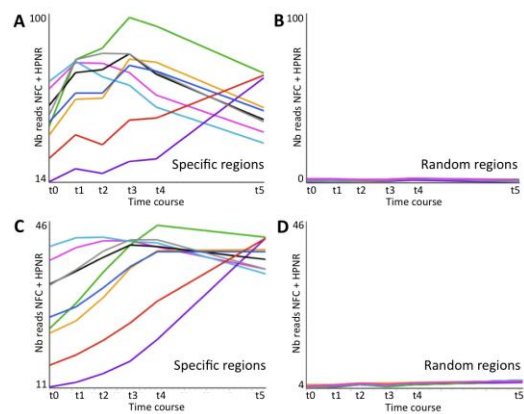
Time	Number of reads	Multiplier factor α
t0	14,726,361	2.31
t1	9,891,325	3.40
t2	6,530,492	5.18
t3	8,093,645	4.17
t4	34,128,883	1
t5	17,035,163	1.97

Mahé_Fig1

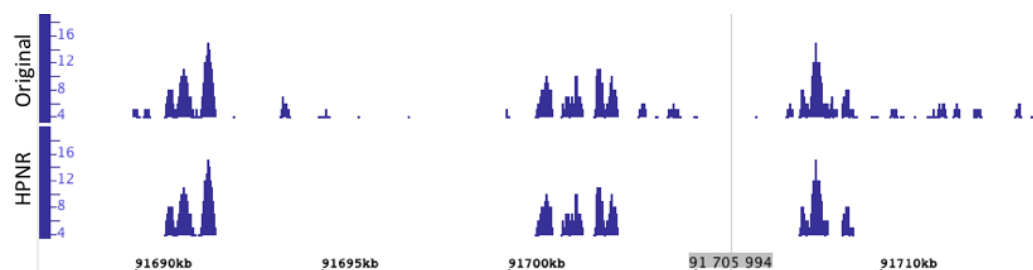


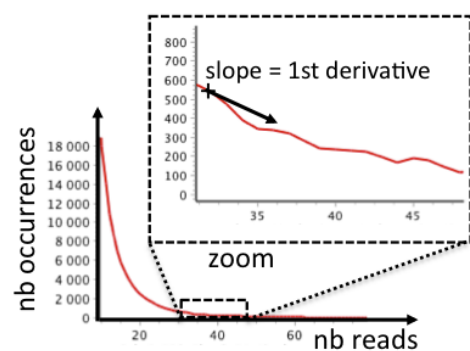


Mahé_Fig3



Mahé_Suppl_Fig1





DISCUSSION

Les travaux accomplis lors de cette thèse ont révélé plusieurs aspects de la dynamique d'activation des enhancers au cours de la différenciation cellulaire. Les résultats obtenus dans la première partie de ce travail nous ont notamment permis de proposer un modèle d'activation des enhancers liés par des FP, impliquant les modifications des cytosines. Ainsi, au cours de la différenciation neurale des cellules P19, la liaison du pionnier FOXA1 est concomitante à la déméthylation des enhancers qui, elle-même, induit indirectement la monométhylation d'H3K4, tandis qu'en parallèle, la formation de 5hmC favorise l'ouverture de la chromatine. Cette première série d'événements provoque la mise en place d'un état chromatinien particulier qui va permettre la liaison ultérieure de MEIS1, favorisée par la présence de modifications des cytosines, et l'acétylation d'H3K27 reflétant l'activation des enhancers. Par conséquent, la 5mC et son effacement, ainsi que la présence de 5hmC sont des paramètres cruciaux pour le « priming » et l'activation des enhancers. L'implication de la 5hmC dans la dynamique d'activation des enhancers est aussi suggérée par les résultats décrits dans la deuxième partie de ces travaux. En effet, des dynamiques différentes de 5hmC au niveau d'enhancers « primed » ou activés lors de la différenciation des cellules P19 sont associées à des dynamiques particulières des autres marques caractéristiques des enhancers *i.e.* H3K4me1, ouverture de la chromatine et H3K27ac. Ainsi, afin d'assurer une mise en place appropriée du répertoire transcriptionnel nécessaire à la différenciation, l'établissement d'une structure chromatinienne particulière semble être étroitement liée aux modifications des cytosines, suggérant qu'elles pourraient influencer l'activité de complexes de remodelage.

Cependant, s'il est admis que la méthylation des cytosines est capable de recruter des complexes de remodelage de la chromatine *via* la liaison de protéines contenant un domaine de liaison à l'ADN méthylé, permettant ainsi une restructuration de la chromatine, peu de choses sont connues concernant un tel recrutement par la 5hmC. Les principales données disponibles concernent le complexe NuRD et sa sous-unité MBD3. En effet, deux études ont observé que MBD3/NuRD colocalise avec TET1 et 5hmC (Baubec *et al.*, 2013 ; Yildirim *et al.*, 2011). Néanmoins, tandis que Yildirim et collègues observent une préférence de liaison de MBD3 sur des sondes contenant des cytosines hydroxyméthylées comparées à des sondes méthylées *in vitro*, Baubec et collègues constatent que dans des cellules triple « KO » pour les DNMT, la liaison de MBD3 n'est pas perturbée au niveau des régions communes avec TET1 et enrichies en 5hmC. Ceci suggère donc que le recrutement de MBD3 est indépendant de l'hydroxyméthylation. De façon intéressante, le ChIP-MS dirigé contre PBX1 dans les cellules P19 différenciées en progéniteurs neuraux nous a également permis d'observer un

enrichissement spécifique de MBD3 et de plusieurs sous-unités du complexe NuRD au niveau de la chromatine liée par PBX1. Dans la mesure où la chromatine liée par PBX1 est particulièrement riche en 5hmC, ces différentes données appuient un lien entre hydroxyméthylation des cytosines et recrutement du complexe NuRD. De la même façon, ces expériences de ChIP-MS ont identifié les sous-unités du complexe FACT, SSRP1 et SPT16, présentes au sein des complexes chromatinien contenant PBX1. Ceci est cohérent avec l'étude récemment publiée par Shakya et collègues, mettant en évidence l'engagement du complexe FACT et son rôle dans l'éviction des nucléosomes au niveau d'un enhancer lié par Oct4 dans les cellules ES (Shakya *et al.*, 2015). Il est intéressant de noter que l'utilisation du DMOG dans leur système d'étude induit une diminution du recrutement de FACT, suggérant que celui-ci dépend de la déméthylation d'H3K9 et/ou de l'hydroxyméthylation des cytosines (Shakya *et al.*, 2015). De plus, l'activité chaperon de FACT, *i.e.* l'éviction des dimères H2A/H2B des nucléosomes, peut être augmentée par la poly-ADP ribosylation de sa sous-unité SPT16 (Dinant *et al.*, 2013) par PARP1, une protéine également détectée par le ChIP-MS ciblant PBX1 et liant spécifiquement la 5hmC dans les expériences de « pull-down » réalisées par le groupe de M. Vermeulen (Spruijt *et al.*, 2013). Enfin, TET2 et TET3 dirigent OGT au niveau de la chromatine, entraînant ainsi la GlcNAcylation de la sérine 112 d'H2B (Chen *et al.*, 2013 ; Deplus *et al.*, 2013). Etant donné que cette modification favorise la mono-ubiquitination de la lysine 20 d'H2B (Fujiki *et al.*, 2011), qui à son tour favorise l'accumulation de FACT (Fleming *et al.*, 2008), il semble vraisemblable que l'engagement des protéines TET au niveau des enhancers puisse mener au recrutement du complexe FACT.

Outre les complexes capables d'induire un remodelage de la chromatine, la restructuration des nucléosomes dépend également de la dynamique d'incorporation des variants d'histones. Ces processus semblent également être associés aux modifications des cytosines. En effet, des études ont montré que chez les plantes et la Souris, il existe une anti-corrélation entre la présence de 5mC et de H2A.Z (Conerly *et al.*, 2010 ; Zilberman *et al.*, 2008). De surcroît, ces deux marques sont mutuellement antagonistes : la méthylation de l'ADN exclut H2A.Z, tandis que l'incorporation d'H2A.Z empêche la mise en place du groupement méthyle (Zilberman *et al.*, 2008) ; et la réactivation génique induite par traitement à la 5-azadC dépend de l'incorporation d'H2A.Z au niveau des promoteurs (Yang *et al.*, 2012). Si la nature des nucléosomes est sensible à la méthylation, il est probable qu'elle soit également influencée par l'hydroxyméthylation. Les résultats obtenus grâce à l'inhibition des protéines TET dans les cellules P19 supportent un tel rôle de la 5hmC et les données acquises

par le CHIP-MS PBX1, nous amènent à proposer une hypothèse supplémentaire quant aux mécanismes liés à ces processus. En effet, parmi les protéines retrouvées enrichies au niveau de la chromatine liée par PBX1, nous avons observé la présence d'ANP32E, récemment montrée comme étant une nouvelle protéine chaperon d'H2A.Z chez l'Homme et responsable de l'éviction d'H2A.Z des nucléosomes spécifiquement au niveau des enhancers et des insulateurs (Obri *et al.*, 2014). Des données préliminaires obtenues au laboratoire dans des cellules P19 indifférenciées montrent cependant des résultats opposés à ceux de Conerly et Zilberman: (i) H2A.Z est présent au niveau d'enhancers méthylés, tandis que dans les progéniteurs neuraux (ii) les enhancers qui deviennent enrichis en 5hmC ont tendance à « perdre » H2A.Z, alors que (iii) les nucléosomes présents au sein d'enhancers qui s'inactivent acquièrent des niveaux plus importants en H2A.Z. Ainsi, dans le cas des enhancers, le faible taux de CpG n'est sans doute pas suffisant pour exclure H2A.Z lorsqu'ils sont méthylés. Ces observations renforcent alors l'hypothèse de l'implication d'une protéine chaperonne assurant l'éviction de ce variant lors de la différenciation. L'hypothèse d'un lien entre la 5hmC et la dynamique d'ouverture de la chromatine est de plus étayée par le fait que les nucléosomes présents au niveau des sites de liaisons de CTCF sont plus sensibles à une digestion à la nucléase micrococcale (Teif *et al.*, 2014). De plus, des nucléosomes formés *in vitro* à partir d'un ADN hydroxyméthylé sont instables, ceci étant dû à une diminution de l'interaction de l'ADN avec les dimères H2A et H2B (Mendonca *et al.*, 2014).

La mise en évidence du processus de déméthylation active et des dérivés oxydés a révélé un aspect dynamique des régions régulatrices du génome jusqu'alors insoupçonné. Au niveau des enhancers, la déméthylation active des cytosines permet l'engagement de ces régions pour promouvoir un programme transcriptionnel particulier. Comme nos résultats le montrent, l'enlèvement de la 5mC et la formation de 5hmc sont tous deux essentiels au « priming » des enhancers. Cette dernière observation renforce l'hypothèse que la 5hmC est une réelle marque épigénétique. En effet, son enrichissement non aléatoire dans le génome et les observations d'une possible association avec la régulation transcriptionnelle, ont été les deux premiers arguments en faveur d'un tel rôle de la « sixième base ». Par la suite, des études ont démontré que la majorité des 5hmC provenant de l'ADN génomique de cellules en culture ou *in vivo* sont stables (Bachman *et al.*, 2014). De plus, une comparaison effectuée entre différents tissus révèle que les niveaux de 5hmC ne corrèlent pas avec ceux de leur précurseur 5mC (Gackowski *et al.*, 2015). Au niveau des régions denses en CpG, la déméthylation permettrait de conserver l'état d'hypométhylation observé sur ces régions. La

concomitance des liaisons de TET1 et de TDG, ainsi que l'augmentation des dérivés oxydés 5fC et 5caC en absence de TDG au niveau des promoteurs de gènes fortement exprimés, renforcent cette hypothèse (Neri *et al.*, 2015). Ainsi, les régions promotrices seraient soumises à de nombreux cycles rapides de méthylation/déméthylation (Kangaspeska *et al.*, 2008 ; Métivier *et al.*, 2008) tandis que le cycle de déméthylation au niveau des enhancers pourrait être majoritairement incomplet, permettant la formation de modifications stables.

Cependant, les résultats que nous avons obtenus lors de l'inhibition des enzymes TET dans les cellules P19, montrent que certains enhancers semblent également être soumis à des cycles complets de déméthylation. En effet, sur ces sites, on observe à la fois la présence de l'étape d'oxydation ($5mC \rightarrow 5hmC$) et celle de l'étape de superoxydation des cytosines ($5hmC \rightarrow 5fC/5caC$). Dans ce cas, si la nature de la région régulatrice et surtout la densité en CpG ne permettent pas d'expliquer pourquoi certaines régions sont soumises à la superoxydation tandis que d'autres s'arrêtent à l'étape d'oxydation, d'autres paramètres doivent être pris en compte. Nos résultats suggèrent ici l'implication de la combinaison de deux facteurs : (i) un FP/FT particulier serait responsable du recrutement de (ii) une enzyme TET particulière. En effet, comme abordé dans la partie introductive de ce manuscrit, afin d'assurer un adressage précis de telles enzymes, il est possible que leur recrutement aux enhancers soit guidé par la présence d'autres protéines. De façon intéressante, les sites enclins à la superoxydation sont des sites liés par le FP FOXA1. Ainsi au niveau des enhancers, il semblerait que la liaison de FOXA1 assure le recrutement des TET. A l'heure actuelle, au niveau fonctionnel, excepté leurs schémas d'expression lors de la différenciation, peu d'éléments distinguent les trois enzymes TET. Sachant que toutes trois possèdent la même activité catalytique *i.e.* oxyder et superoxyder les modifications des cytosines, on peut se poser la question de la nécessité de la présence de trois enzymes. Outre, la redondance d'activité pour assurer le déroulement correct du cycle de déméthylation en cas de défaut d'une autre protéine, ou encore l'interaction spécifique avec d'autres facteurs, nous proposons que l'existence de ces trois enzymes pourrait être le reflet de particularités fonctionnelles vis-à-vis de la catalyse de l'oxydation des 5mC comme suggéré par Putiri *et al.* (2014), Hu *et al.* (2015) et Lu *et al.* (2016). C'est notamment le cas pour TET2, qui est principalement responsable de l'oxydation des 5hmC dans les ECC humaines (Putiri *et al.*, 2014). L'existence d'une interaction entre FOXA1 et TET2 dans des cellules cancéreuses de prostate (Takayama *et al.*, 2015) renforce l'hypothèse d'une implication de ce facteur pionnier dans le contrôle de la superoxydation des cytosines aux enhancers. Un tel mécanisme pourrait ainsi

expliquer le lien entre la liaison des FT et la mise en place d'un état d'hypométhylation observée au niveau des enhancers lors de la différenciation (Sérandour *et al.*, 2011 ; Stadler *et al.*, 2009).

Contrairement à Yu et collègues qui n'observent ni méthylation ni hydroxyméthylation des cytosines au niveau même des sites de liaison des FT mais plutôt autour (*i.e.* à une distance moyenne de 50 bp par rapport au motif lié par les FT - Yu *et al.*, 2012), des données préalablement obtenues au laboratoires montrent un enrichissement des sites de liaisons des protéines TALE-HD MEIS1 et PBX1 au niveau d'enhancers hydroxyméthylés lors de la différenciation des cellules P19 (Sérandour *et al.*, 2012). Il faut cependant prendre en compte le fait que Yu *et al.* utilisent le TAB-seq et ont donc une résolution à la base près, ce que ne permet pas d'obtenir l'hMeDIP-seq utilisé pour cartographier la 5hmC dans les cellules P19 (Sérandour *et al.*, 2012). De plus, ces observations concernent deux types cellulaires différents possédant des combinatoires de FT et un hydroxyméthylome qui leur sont propres. Néanmoins, en utilisant des sondes correspondant aux sites de reconnaissance des hétérodimères PBX1/HOXA9 ou PBX1/MEIS1 différenciellement modifiées au niveau des cytosines, nous avons montré que la fixation des protéines TALE-HD est favorisée par la présence de modifications des cytosines *in vitro*. Deux atouts majeurs de ces expériences proviennent (i) de l'utilisation de sondes comportant des séquences retrouvées dans le génome, donc non arbitraires, et (ii) de la présence de combinaisons de modifications asymétriques (par exemple : hémi-méthylé, ou méthylé sur un brin et hydroxyméthylé sur l'autre) comme c'est le cas dans le génome des cellules ES (Yu *et al.*, 2012). Si la liaison d'un certain nombre de FT tels que C/EBP β , Tcf3, Zfp57, Egr1 et WT1 semble être directement influencée par la présence de dérivés oxydés des cytosines au sein de leur séquences cibles (Liu *et al.*, 2013 ; Golla *et al.*, 2014 ; Hashimoto *et al.*, 2014 ; Sayeed *et al.*, 2015), des données récemment obtenues au laboratoire avec une résolution au CpG près suggèrent que l'hydroxyméthylation des cytosines ne cible pas préférentiellement les motifs liés par les FT (Sérandour *et al.*, 2016, soumis). Ces données vont dans le sens d'un rôle de la déméthylation active des cytosines dans l'ouverture de la chromatine plutôt que dans la régulation directe des interactions entre FT et ADN.

L'ensemble des données de la littérature pointe vers un rôle particulièrement important des TET et de l'oxydation des 5mC dans la régulation de la transcription et tout particulièrement dans le contrôle de l'activité des enhancers. Néanmoins, deux études

réalisées dans les cellules ES (Hon *et al.* 2014 ; Lu *et al.*, 2014) ont montré que l'ablation des TET n'a qu'une influence limitée sur les marques d'histones associées aux enhancers et que seule une petite fraction d'enhancers gagne la 5mC dans ces conditions. Il est alors vraisemblable, ainsi que suggéré par Hon *et al.* (2014) et comme nous le montrons dans la partie I des résultats, que la déméthylation active des cytosines des enhancers soit impliquée dans la régulation fine dans le temps et l'espace de l'expression des gènes, notamment au cours du développement, sans être cependant absolument requise pour leur activation. Cet impact modéré pourrait éventuellement être expliqué par l'intervention de multiples enhancers dans le contrôle de l'expression d'un gène donné, dont certains, dépourvus de CpG, ne nécessiteraient pas le recrutement des TET. La redondance des enhancers est un phénomène déjà observé dans d'autres contextes (Attanasio *et al.*, 2013 ; Quintin *et al.*, 2014) et qui permettrait d'assurer la robustesse des programmes d'expression génique. Nos résultats suggèrent néanmoins une implication fonctionnelle de l'hydroxyméthylation des cytosines dans les phénomènes locaux de remodelage de la chromatine. La multiplicité des mécanismes de régulation des complexes de remodelage permet là-encore certainement d'entretenir la robustesse de ces phénomènes et la suppression d'un seul de ces mécanismes, en l'occurrence la déméthylation des cytosines, pourrait n'avoir qu'un impact limité. Il n'en reste pas moins important cependant de continuer à décrypter les processus mis en œuvre. Dans ce contexte, nous envisageons, à court terme, d'étudier les relations entre l'hydroxyméthylation des cytosines et l'activité de complexes de remodelage tels que NuRD, FACT ou encore du chaperon d'H2A.Z ANP32E.

BIBLIOGRAPHIE

- Adams CC, Workman JL. 1995. Binding of disparate transcriptional activators to nucleosomal DNA is inherently cooperative. *Mol Cell Biol* 15(3): 1405-1421.
- Adelman K, Lis JT. 2012. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat Rev Genet* 13(10): 720-731.
- Aerts S. 2012. Computational strategies for the genome-wide identification of cis-regulatory elements and transcriptional targets. *Curr Top Dev Biol* 98: 121-145.
- Akalin A, Fredman D, Arner E, Dong X, Bryne JC, Suzuki H, Daub CO, Hayashizaki Y, Lenhard B. 2009. Transcriptional features of genomic regulatory blocks. *Genome Biol* 10(4): R38.
- Allen BL, Taatjes DJ. 2015. The Mediator complex: a central integrator of transcription. *Nat Rev Mol Cell Biol* 16(3): 155-166.
- Andersson R, Gebhard C, Miguel-Escalada I, Hoof I, Bornholdt J, Boyd M, Chen Y, Zhao X, Schmidl C, Suzuki T, Ntini E, Arner E, Valen E, Li K, Schwarzfischer L, Glatz D, Raithel J, Lilje B, Rapin N, Bagger FO, Jørgensen M, Andersen PR, Bertin N, Rackham O, Burroughs AM, Baillie JK, Ishizu Y, Shimizu Y, Furuhashi E, Maeda S, Negishi Y, Mungall CJ, Meehan TF, Lassmann T, Itoh M, Kawaji H, Kondo N, Kawai J, Lennartsson A, Daub CO, Heutink P, Hume DA, Jensen TH, Suzuki H, Hayashizaki Y, Müller F; FANTOM Consortium, Forrest AR, Carninci P, Rehli M, Sandelin A. 2014. An atlas of active enhancers across human cell types and tissues. *Nature* 507(7493): 455-461.
- Andersson R, Sandelin A, Danko CG. 2015. A unified architecture of transcriptional regulatory elements. *Trends Genet* 31(8): 426-433.
- Antequera F, Bird A. 1993. Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci U S A* 90(24): 11995-11999.
- Arents G, Moudrianakis EN. 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A* 92(24): 11170-11174.
- Arnold CD, Gerlach D, Stelzer C, Boryń LM, Rath M, Stark A. 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* 339(6123): 1074-1077.
- Attanasio C, Nord AS, Zhu Y, Blow MJ, Li Z, Liberton DK, Morrison H, Plajzer-Frick I, Holt A, Hosseini R, Phouanavong S, Akiyama JA, Shoukry M, Afzal V, Rubin EM, FitzPatrick DR, Ren B, Hallgrímsson B, Pennacchio LA, Visel A. 2013. Fine tuning of craniofacial morphology by distant-acting enhancers. *Science* 342(6157): 1241006.
- Auclair G, Borgel J, Sanz LA, Vallet J, Guibert S, Dumas M, Cavelier P, Girardot M, Forné T, Feil R, Weber M. 2015. EHMT2 directs DNA methylation for efficient gene silencing in mouse embryos. *Genome Res.* Nov 17.
- Bachman M, Uribe-Lewis S, Yang X, Burgess HE, Iurlaro M, Reik W, Murrell A, Balasubramanian S. 2015. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat Chem Biol* 11(8): 555-557.
- Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 2014. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* 6(12): 1049-55.

- Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37(Web Server issue): W202-208.
- Bailey TL, Noble WS. 2003. Searching for statistically significant regulatory modules. *Bioinformatics* 19 Suppl 2: ii16-25.
- Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Res* 21(3): 381-395.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. 2007. High-resolution profiling of histone methylations in the human genome. *Cell* 129(4): 823-837.
- Baubec T, Ivánek R, Lienert F, Schübeler D. 2013. Methylation-dependent and -independent genomic targeting principles of the MBD protein family. *Cell* 153(2): 480-492.
- Bauer C, Göbel K, Nagaraj N, Colantuoni C, Wang M, Müller U, Kremmer E, Rottach A, Leonhardt H. 2015. Phosphorylation of TET proteins is regulated via O-GlcNAcylation by the O-linked N-acetylglucosamine transferase (OGT). *J Biol Chem* 290(8): 4801-4812.
- Belotserkovskaya R, Oh S, Bondarenko VA, Orphanides G, Studitsky VM, Reinberg D. 2003. FACT facilitates transcription-dependent nucleosome alteration. *Science* 301(5636): 1090-1093.
- Bergsland M, Ramsköld D, Zaouter C, Klum S, Sandberg R, Muhr J. 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev* 25(23): 2453-2464.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125(2): 315-326.
- Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, Miranda TB, Sung MH, Trump S, Lightman SL, Vinson C, Stamatoyannopoulos JA, Hager GL. 2011. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* 43(1): 145-155.
- Blackwood EM, Kadonaga JT. 1998. Going the distance: a current view of enhancer action. *Science* 281(5373): 60-3.
- Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Bristow J, Ren B, Black BL, Rubin EM, Visel A, Pennacchio LA. 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat Genet* 42(9): 806-810.
- Boeke J, Ammerpohl O, Kegel S, Moehren U, Renkawitz R. 2000. The minimal repression domain of MBD2b overlaps with the methyl-CpG-binding domain and binds directly to Sin3A. *J Biol Chem* 275(45): 34963-34967.
- Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR, Cremer T. 2005. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. *PLoS Biol* 3(5): e157.
- Bonn S, Zinzen RP, Girardot C, Gustafson EH, Perez-Gonzalez A, Delhomme N, Ghavi-Helm Y, Wilczyński B, Riddell A, Furlong EE. 2012. Tissue-specific analysis of chromatin state identifies temporal signatures of enhancer activity during embryonic development. *Nat Genet* 44(2): 148-156.

- Booth MJ, Branco MR, Ficiz G, Oxley D, Krueger F, Reik W, Balasubramanian S. 2012. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336(6083): 934-937.
- Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S. 2014. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem* 6(5): 435-440.
- Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M. 2010. Targets and dynamics of promoter DNA methylation during early mouse development. *Nat Genet* 42(12): 1093-1100.
- Bostick M, Kim JK, Estève PO, Clark A, Pradhan S, Jacobsen SE. 2007. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317(5845): 1760-1764.
- Bourc'his D, Xu GL, Lin CS, Bollman B, Bestor TH. 2001. Dnmt3L and the establishment of maternal genomic imprints. *Science* 294(5551): 2536-2539.
- Boveri T. 1909. Die Blastomerenkerne von *Ascaris megalocephala* und die Theorie der Chromosomenindividualität. *Arch Zellforsch* 3: 181–268.
- Boyes J, Bird A. 1991. DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64(6): 1123-1134.
- Boyes J, Bird A. 1992. Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. *EMBO J* 11(1): 327-333.
- Boyes J, Felsenfeld G. 1996. Tissue-specific factors additively increase the probability of the all-or-none formation of a hypersensitive site. *EMBO J* 15(10): 2496-2507.
- Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum Mol Genet* 10(3): 211-219.
- Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H. 1994. Sp1 elements protect a CpG island from de novo methylation. *Nature* 371(6496): 435-438.
- Breiling A, Lyko F. 2015. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* 8: 24.
- Brenner C, Deplus R, Didelot C, Lorient A, Viré E, De Smet C, Gutierrez A, Danovi D, Bernard D, Boon T, Pelicci PG, Amati B, Kouzarides T, de Launoit Y, Di Croce L, Fuks F. 2005. Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J* 24(2): 336-346.
- Brunelle M, Nordell Markovits A, Rodrigue S, Lupien M, Jacques PÉ, Gévry N. 2015. The histone variant H2A.Z is an important regulator of enhancer activity. *Nucleic Acids Res* 43(20): 9742-9756.
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10(12): 1213-1218.

- Bulut-Karslioglu A, Perrera V, Scaranaro M, de la Rosa-Velazquez IA, van de Nobelen S, Shukeir N, Popow J, Gerle B, Opravil S, Pagani M, Meidhof S, Brabletz T, Manke T, Lachner M, Jenuwein T. 2012. A transcription factor-based mechanism for mouse heterochromatin formation. *Nat Struct Mol Biol* 19(10): 1023-1030.
- Burke TW, Kadonaga JT. 1997. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by TAFII60 of *Drosophila*. *Genes Dev* 11(22): 3020-3031.
- Byrd AK, Raney KD. 2012. Superfamily 2 helicases. *Front Biosci (Landmark Ed)*. 17:2070-2088.
- Cairns BR, Kim YJ, Sayre MH, Laurent BC, Kornberg RD. 1994. A multisubunit complex containing the SWI1/ADR6, SWI2/SNF2, SWI3, SNF5, and SNF6 gene products isolated from yeast. *Proc Natl Acad Sci U S A* 91(5): 1950-1954.
- Cairns BR, Lorch Y, Li Y, Zhang M, Lacomis L, Erdjument-Bromage H, Tempst P, Du J, Laurent B, Kornberg RD. 1996. RSC, an essential, abundant chromatin-remodeling complex. *Cell* 87(7): 1249-1260.
- Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Mol Cell* 49(5): 825-837.
- Cannavò E, Khoueiry P, Garfield DA, Geeleher P, Zichner T, Gustafson EH, Ciglar L, Korbel JO, Furlong EE. 2016. Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr Biol* 26(1): 38-51.
- Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engström PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y. 2006. Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38(6): 626-635.
- Caron G, Hussein M, Kulis M, Delaloy C, Chatonnet F, Pignarre A, Avner S, Lemarié M, Mahé EA, Verdaguer-Dot N, Queirós AC, Tarte K, Martín-Subero JL, Salbert G, Fest T. 2015. Cell-Cycle-Dependent Reconfiguration of the DNA Methylome during Terminal Differentiation of Human B Cells into Plasma Cells. *Cell Rep* 13(5): 1059-1071.
- Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* 122(1): 33-43.
- Cartron PF, Nadaradjane A, Lepape F, Lalier L, Gardie B, Vallette FM. 2013. Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer* 4(5-6): 235-241.
- Chalkley GE, Verrijzer CP. 1999. DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator. *EMBO J* 18(17): 4835-4845.
- Chen CC, Wang KY, Shen CK. 2012. The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J Biol Chem* 287(40): 33116-33121.
- Chen Q, Chen Y, Bian C, Fujiki R, Yu X. 2013. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493(7433): 561-564.

- Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. 2012. Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res* 22(3): 490-503.
- Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 9(2): 279-289.
- Cirillo LA, Zaret KS. 1999. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell* 4(6): 961-969.
- Conerly ML, Teves SS, Diolaiti D, Ulrich M, Eisenman RN, Henikoff S. 2010. Changes in H2A.Z occupancy and DNA methylation during B-cell lymphomagenesis. *Genome Res* 20(10): 1383-1390.
- Cooper DN, Krawczak M. 1989. Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum Genet* 83(2): 181-188.
- Core LJ, Martins AL, Danko CG, Waters CT, Siepel A, Lis JT. 2014. Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* 46(12): 1311-1320.
- Core LJ, Waterfall JJ, Lis JT. 2008. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* 322(5909): 1845-1848.
- Cortázar D, Kunz C, Selfridge J, Lettieri T, Saito Y, MacDougall E, Wirz A, Schuermann D, Jacobs AL, Siegrist F, Steinacher R, Jiricny J, Bird A, Schär P. 2011. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* Feb 470(7334): 419-423.
- Cortellino S, Xu J, Sannai M, Moore R, Caretti E, Cigliano A, Le Coz M, Devarajan K, Wessels A, Soprano D, Abramowitz LK, Bartolomei MS, Rambow F, Bassi MR, Bruno T, Fanciulli M, Renner C, Klein-Szanto AJ, Matsumoto Y, Kobi D, Davidson I, Alberti C, Larue L, Bellacosa A. 2011. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146(1): 67-79.
- Costa Y, Ding J, Theunissen TW, Faiola F, Hore TA, Shliha PV, Fidalgo M, Saunders A, Lawrence M, Dietmann S, Das S, Levasseur DN, Li Z, Xu M, Reik W, Silva JC, Wang J. 2013. NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495(7441): 370-374.
- Côté J, Quinn J, Workman JL, Peterson CL. 1994. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF protein complex. *Science* 265(5168): 53-60.
- Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, Hanna J, Lodato MA, Frampton GM, Sharp PA, Boyer LA, Young RA, Jaenisch R. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* 107(50): 21931-21936.
- Creyghton MP, Markoulaki S, Levine SS, Hanna J, Lodato MA, Sha K, Young RA, Jaenisch R, Boyer LA. 2008. H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. *Cell* 135(4): 649-661.
- Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *J Cell Biol* 145(6): 1119-1131.

- Cui K, Zang C, Roh TY, Schones DE, Childs RW, Peng W, Zhao K. 2009. Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation. *Cell Stem Cell* 4(1): 80-93.
- Curradi M, Izzo A, Badaracco G, Landsberger N. 2002. Molecular mechanisms of gene silencing mediated by DNA methylation. *Mol Cell Biol* 22(9): 3157-3173.
- de la Rica L, Rodríguez-Ubreva J, García M, Islam AB, Urquiza JM, Hernando H, Christensen J, Helin K, Gómez-Vaquero C, Ballestar E. 2013. PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol* 14(9): R99.
- De Santa F, Barozzi I, Mietton F, Ghisletti S, Polletti S, Tusi BK, Muller H, Ragoussis J, Wei CL, Natoli G. 2010. A large fraction of extragenic RNA pol II transcription sites overlap enhancers. *PLoS Biol* 8(5): e1000384.
- de Wit E, Vos ES, Holwerda SJ, Valdes-Quezada C, Verstegen MJ, Teunissen H, Splinter E, Wijchers PJ, Krijger PH, de Laat W. 2015. CTCF Binding Polarity Determines Chromatin Looping. *Mol Cell* 60(4): 676-684.
- Degner SC, Verma-Gaur J, Wong TP, Bossen C, Iverson GM, Torkamani A, Vettermann C, Lin YC, Ju Z, Schulz D, Murre CS, Birshtein BK, Schork NJ, Schlissel MS, Riblet R, Murre C, Feeney AJ. 2011. CCCTC-binding factor (CTCF) and cohesin influence the genomic architecture of the Igh locus and antisense transcription in pro-B cells. *Proc Natl Acad Sci U S A* 108(23): 9566-9571.
- Deplus R, Delatte B, Schwinn MK, Defrance M, Méndez J, Murphy N, Dawson MA, Volkmar M, Putmans P, Calonne E, Shih AH, Levine RL, Bernard O, Mercher T, Solary E, Urh M, Daniels DL, Fuks F. 2013. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* 32(5): 645-55.
- Dhayalan A, Rajavelu A, Rathert P, Tamas R, Jurkowska RZ, Ragozin S, Jeltsch A. 2010. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J Biol Chem* 285(34): 26114-26120.
- Dimitri P, Caizzi R, Giordano E, Carmela Accardo M, Lattanzi G, Biamonti G. 2009. Constitutive heterochromatin: a surprising variety of expressed sequences. *Chromosoma* 118(4): 419-435.
- Dinant C, Ampatzidis-Michailidis G, Lans H, Tresini M, Lagarou A, Grosbart M, Theil AF, van Cappellen WA, Kimura H, Bartek J, Fousteri M, Houtsmuller AB, Vermeulen W, Marteijn JA. 2013. Enhanced chromatin dynamics by FACT promotes transcriptional restart after UV-induced DNA damage. *Mol Cell* 51(4): 469-479.
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485(7398): 376-380.
- Doyen CM, Montel F, Gautier T, Menoni H, Claudet C, Delacour-Larose M, Angelov D, Hamiche A, Bednar J, Faivre-Moskalenko C, Bouvet P, Dimitrov S. 2006a. Dissection of the unusual structural and functional properties of the variant H2A.Bbd nucleosome. *EMBO J* 25(18): 4234-4244.
- Doyen CM, An W, Angelov D, Bondarenko V, Mietton F, Studitsky VM, Hamiche A, Roeder RG, Bouvet P, Dimitrov S. 2006b. Mechanism of polymerase II transcription repression by the histone variant macroH2A. *Mol Cell Biol* 26(3): 1156-1164.

- Drané P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. 2010. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev* 24(12): 1253-1265.
- Drissen R, Palstra RJ, Gillemans N, Splinter E, Grosveld F, Philipsen S, de Laat W. 2004. The active spatial organization of the beta-globin locus requires the transcription factor EKLF. *Genes Dev* 18(20): 2485-2490.
- Dvir A. 2002. Promoter escape by RNA polymerase II. *Biochim Biophys Acta* 1577(2): 208-223.
- Eberhardy SR, Farnham PJ. 2002. Myc recruits P-TEFb to mediate the final step in the transcriptional activation of the cad promoter. *J Biol Chem* 277(42): 40156-40162.
- Ebert DH, Gabel HW, Robinson ND, Kastan NR, Hu LS, Cohen S, Navarro AJ, Lyst MJ, Ekiert R, Bird AP, Greenberg ME. 2013. Activity-dependent phosphorylation of MeCP2 threonine 308 regulates interaction with NCoR. *Nature* 499(7458): 341-345.
- Eeckhoutte J, Keeton EK, Lupien M, Krum SA, Carroll JS, Brown M. 2007. Positive cross-regulatory loop ties GATA-3 to estrogen receptor alpha expression in breast cancer. *Cancer Res* 67(13): 6477-6483.
- Eeckhoutte J, Lupien M, Meyer CA, Verzi MP, Shivdasani RA, Liu XS, Brown M. 2009. Cell-type selective chromatin remodeling defines the active subset of FOXA1-bound enhancers. *Genome Res* 19(3): 372-380.
- ENCODE Project Consortium, Birney, E. *et al.* 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* 447(7146): 799-816.
- ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414): 57-74.
- Ernst J, Kellis M. 2010. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 28(8): 817-825.
- Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 473(7345): 43-49.
- Esnault C, Ghavi-Helm Y, Brun S, Soutourina J, Van Berkum N, Boschiero C, Holstege F, Werner M. 2008. Mediator-dependent recruitment of TFIIH modules in preinitiation complex. *Mol Cell* 31(3): 337-346.
- Feldman N, Gerson A, Fang J, Li E, Zhang Y, Shinkai Y, Cedar H, Bergman Y. 2006. G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat Cell Biol* 8(2): 188-194.
- Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, Hetzel J, Jain J, Strauss SH, Halpern ME, Ukumadu C, Sadler KC, Pradhan S, Pellegrini M, Jacobsen SE. 2010. Conservation and divergence of methylation patterning in plants and animals. *Proc Natl Acad Sci U S A* 107(19): 8689-8694.
- Feng YQ, Desprat R, Fu H, Olivier E, Lin CM, Lobell A, Gowda SN, Aladjem MI, Bouhassira EE. 2006. DNA methylation supports intrinsic epigenetic memory in mammalian cells. *PLoS Genet* 2(4): e65.
- Ferreira H, Flaus A, Owen-Hughes T. 2007. Histone modifications influence the action of Snf2 family remodelling enzymes by different mechanisms. *J. Mol. Biol.* 374(3): 563-579.

- Ficz G, Branco MR, Seisenberger S, Santos F, Krueger F, Hore TA, Marques CJ, Andrews S, Reik W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473(7347): 398-402.
- Filion GJ, van Bommel JG, Braunschweig U, Talhout W, Kind J, Ward LD, Brugman W, de Castro IJ, Kerkhoven RM, Bussemaker HJ, van Steensel B. 2010. Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* 143(2): 212-224.
- Finlan LE, Sproul D, Thomson I, Boyle S, Kerr E, Perry P, Ylstra B, Chubb JR, Bickmore WA. 2008. Recruitment to the nuclear periphery can alter expression of genes in human cells. *PLoS Genet* 4(3): e1000039.
- Fleming AB, Kao CF, Hillyer C, Pikaart M, Osley MA. 2008. H2B ubiquitylation plays a role in nucleosome dynamics during transcription elongation. *Mol Cell* 31(1): 57-66.
- Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G. 2008. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. *Cell Stem Cell* 2(2): 160-169.
- Frauer C, Hoffmann T, Bultmann S, Casa V, Cardoso MC, Antes I, Leonhardt H. 2011. Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* 6(6): e21306.
- Fritz EL, Papavasiliou FN. 2010. Cytidine deaminases: AIDing DNA demethylation? *Genes Dev* 24(19): 2107-2114.
- Fujiki K, Shinoda A, Kano F, Sato R, Shirahige K, Murata M. 2013. PPAR γ -induced PARylation promotes local DNA demethylation by production of 5-hydroxymethylcytosine. *Nat Commun* 4: 2262.
- Fujiki R, Hashiba W, Sekine H, Yokoyama A, Chikanishi T, Ito S, Imai Y, Kim J, He HH, Igarashi K, Kanno J, Ohtake F, Kitagawa H, Roeder RG, Brown M, Kato S. 2011. GlcNAcylation of histone H2B facilitates its monoubiquitination. *Nature* 480(7378): 557-560.
- Fujita N, Takebayashi S, Okumura K, Kudo S, Chiba T, Saya H, Nakao M. 1999. Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. *Mol Cell Biol* 19(9): 6415-6426.
- Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. 2000. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat Genet* 24(1): 88-91.
- Fuks F, Hurd PJ, Deplus R, Kouzarides T. 2003. The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res* 31(9): 2305-2312.
- Fullwood MJ, Liu MH, Pan YF, Liu J, Xu H, Mohamed YB, Orlov YL, Velkov S, Ho A, Mei PH, Chew EG, Huang PY, Welboren WJ, Han Y, Ooi HS, Ariyaratne PN, Vega VB, Luo Y, Tan PY, Choy PY, Wansa KD, Zhao B, Lim KS, Leow SC, Yow JS, Joseph R, Li H, Desai KV, Thomsen JS, Lee YK, Karuturi RK, Herve T, Bourque G, Stunnenberg HG, Ruan X, Cacheux-Rataboul V, Sung WK, Liu ET, Wei CL, Cheung E, Ruan Y. 2009. An oestrogen-receptor- α -bound human chromatin interactome. *Nature* 462(7269): 58-64.
- Futscher BW, Oshiro MM, Wozniak RJ, Holtan N, Hanigan CL, Duan H, Domann FE. 2002. Role for DNA methylation in the control of cell type specific maspin expression. *Nat Genet* 31(2): 175-179.

- Gackowski D, Zarakowska E, Starczak M, Modrzejewska M, Olinski R. 2015. Tissue-Specific Differences in DNA Modifications (5-Hydroxymethylcytosine, 5-Formylcytosine, 5-Carboxylcytosine and 5-Hydroxymethyluracil) and Their Interrelationships. *PLoS One* 10(12): e0144859.
- Gamble MJ, Frizzell KM, Yang C, Krishnakumar R, Kraus WL. 2010. The histone variant macroH2A1 marks repressed autosomal chromatin, but protects a subset of its target genes from silencing. *Genes Dev* 24(1): 21-32.
- Gao N, Zhang J, Rao MA, Case TC, Mirosevich J, Wang Y, Jin R, Gupta A, Rennie PS, Matusik RJ. 2003. The role of hepatocyte nuclear factor-3 alpha (Forkhead Box A1) and androgen receptor in transcriptional regulation of prostatic genes. *Mol Endocrinol* 17(8): 1484-1507.
- Gaszner M, Felsenfeld G. 2006. Insulators: exploiting transcriptional and epigenetic mechanisms. *Nat Rev Genet* (9): 703-713.
- Ge YZ, Pu MT, Gowher H, Wu HP, Ding JP, Jeltsch A, Xu GL. 2004. Chromatin targeting of de novo DNA methyltransferases by the PWWP domain. *J Biol Chem* 279(24): 25447-25454.
- Geiman TM, Sankpal UT, Robertson AK, Zhao Y, Zhao Y, Robertson KD. 2004. DNMT3B interacts with hSNF2H chromatin remodeling enzyme, HDACs 1 and 2, and components of the histone methylation system. *Biochem Biophys Res Commun* 318(2): 544-555.
- Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL, Ragoussis J, Natoli G. 2010. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* 32(3): 317-328.
- Gierman HJ, Indemans MH, Koster J, Goetze S, Seppen J, Geerts D, van Driel R, Versteeg R. 2007. Domain-wide regulation of gene expression in the human genome. *Genome Res* 17(9): 1286-1295.
- Globisch D, Münzel M, Müller M, Michalakakis S, Wagner M, Koch S, Brückl T, Biel M, Carell T. 2010. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5(12): e15367.
- Goetze S, Mateos-Langerak J, Gierman HJ, de Leeuw W, Giromus O, Indemans MH, Koster J, Ondrej V, Versteeg R, van Driel R. 2007. The three-dimensional structure of human interphase chromosomes is related to the transcriptome map. *Mol Cell Biol* 27(12): 4475-4487.
- Goldberg AD, Banaszynski LA, Noh KM, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, Wen D, Chagier A, DeKolver RC, Miller JC, Lee YL, Boydston EA, Holmes MC, Gregory PD, Greally JM, Rafii S, Yang C, Scambler PJ, Garrick D, Gibbons RJ, Higgs DR, Cristea IM, Urnov FD, Zheng D, Allis CD. 2010. Distinct factors control histone variant H3.3 localization at specific genomic regions. *Cell* 140(5): 678-691.
- Goldman JA, Garlick JD, Kingston RE. 2010. Chromatin remodeling by imitation switch (ISWI) class ATP-dependent remodelers is stimulated by histone variant H2A.Z. *J Biol Chem* 285(7): 4645-4651.
- Golla JP, Zhao J, Mann IK, Sayeed SK, Mandal A, Rose RB, Vinson C. 2014. Carboxylation of cytosine (5caC) in the CG dinucleotide in the E-box motif (CGCAG|GTG) increases binding of the Tcf3|Ascl1 helix-loop-helix heterodimer 10-fold. *Biochem Biophys Res Commun* 449(2): 248-255.
- Goodrich JA, Tjian R. 2010. Unexpected roles for core promoter recognition factors in cell-type-specific transcription and gene regulation. *Nat Rev Genet* 11(8): 549-558.

- Gosselin D, Link VM, Romanoski CE, Fonseca GJ, Eichenfield DZ, Spann NJ, Stender JD, Chun HB, Garner H, Geissmann F, Glass CK. 2014. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell* 159(6): 1327-1340.
- Guo Y, Xu Q, Canzio D, Shou J, Li J, Gorkin DU, Jung I, Wu H, Zhai Y, Tang Y, Lu Y, Wu Y, Jia Z, Li W, Zhang MQ, Ren B, Krainer AR, Maniatis T, Wu Q. 2015. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* 162(4): 900-910.
- Hah N, Murakami S, Nagari A, Danko CG, Kraus WL. 2013. Enhancer transcripts mark active estrogen receptor binding sites. *Genome Res* 23(8): 1210-1223.
- Handoko L, Xu H, Li G, Ngan CY, Chew E, Schnapp M, Lee CW, Ye C, Ping JL, Mulawadi F, Wong E, Sheng J, Zhang Y, Poh T, Chan CS, Kunarso G, Shahab A, Bourque G, Cacheux-Rataboul V, Sung WK, Ruan Y, Wei CL. 2011. CTCF-mediated functional chromatin interactome in pluripotent cells. *Nat Genet* 43(7): 630-638.
- Hashimoto H, Olanrewaju YO, Zheng Y, Wilson GG, Zhang X, Cheng X. 2014. Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes Dev* 28(20): 2304-2313.
- Havas K1, Whitehouse I, Owen-Hughes T. 2001. ATP-dependent chromatin remodeling activities. *Cell Mol Life Sci* 58(5-6): 673-682.
- Hawkins RD, Hon GC, Yang C, Antosiewicz-Bourget JE, Lee LK, Ngo QM, Klugman S, Ching KA, Edsall LE, Ye Z, Kuan S, Yu P, Liu H, Zhang X, Green RD, Lobanenko VV, Stewart R, Thomson JA, Ren B. 2011. Dynamic chromatin states in human ES cells reveal potential regulatory sequences and genes involved in pluripotency. *Cell Res* 21(10): 1393-1409.
- He HH, Meyer CA, Shin H, Bailey ST, Wei G, Wang Q, Zhang Y, Xu K, Ni M, Lupien M, Mieczkowski P, Lieb JD, Zhao K, Brown M, Liu XS. 2010. Nucleosome dynamics define transcriptional enhancers. *Nat Genet* 42(4): 343-347.
- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, Dai Q, Song CX, Zhang K, He C, Xu GL. 2011. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333(6047): 1303-1307.
- Heidt AB, Rojas A, Harris IS, Black BL. 2007. Determinants of myogenic specificity within MyoD are required for noncanonical E box binding. *Mol Cell Biol* 27(16): 5910-520.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenko VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459(7243): 108-12.
- Heintzman ND, Ren B. 2007. The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome. *Cell Mol Life Sci* 64(4): 386-400.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 39(3): 311-318.

- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38(4): 576-589.
- Heinz S, Romanoski CE, Benner C, Glass CK. 2015. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* 16(3): 144-154.
- Hendrich B, Bird A. 1998. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* 18(11): 6538-6547.
- Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. 1999. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401(6750): 301-304.
- Hepperger C, Mannes A, Merz J, Peters J, Dietzel S. 2008. Three-dimensional positioning of genes in mouse cell nuclei. *Chromosoma* 117(6): 535-551.
- Hermann A, Goyal R, Jeltsch A. 2004. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J Biol Chem* 279(46): 48350-48359.
- Hervouet E, Vallette FM, Cartron PF. 2009. Dnmt3/transcription factor interactions as crucial players in targeted DNA methylation. *Epigenetics* 4(7): 487-499.
- Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA. 2013. Super-enhancers in the control of cell identity and disease. *Cell* 155(4): 934-947.
- Holstege FC, van der Vliet PC, Timmers HT. 1996. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIE and IIH. *EMBO J* 15(7): 1666-1677.
- Hon G, Wang W, Ren B. 2009. Discovery and annotation of functional chromatin signatures in the human genome. *PLoS Comput Biol* 5(11): e1000566.
- Hon GC, Song CX, Du T, Jin F, Selvaraj S, Lee AY, Yen CA, Ye Z, Mao SQ, Wang BA, Kuan S, Edsall LE, Zhao BS, Xu GL, He C, Ren B. 2014. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* 56(2): 286-297.
- Hong JW, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty. *Science* 321(5894): 1314.
- Hu G, Cui K, Northrup D, Liu C, Wang C, Tang Q, Ge K, Levens D, Crane-Robinson C, Zhao K. 2013. H2A.Z facilitates access of active and repressive complexes to chromatin in embryonic stem cell self-renewal and differentiation. *Cell Stem Cell* 12(2): 180-192.
- Hu L, Lu J, Cheng J, Rao Q, Li Z, Hou H, Lou Z, Zhang L, Li W, Gong W, Liu M, Sun C, Yin X, Li J, Tan X, Wang P, Wang Y, Fang D, Cui Q, Yang P, He C, Jiang H, Luo C, Xu Y. 2015. Structural insight into substrate preference for TET-mediated oxidation. *Nature* 527(7576): 118-122.
- Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D, Carroll JS. 2011. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* 43(1): 27-33.
- Iguchi-Arigo SM, Schaffner W. 1989. CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev* 3(5): 612-619.

- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466(7310): 1129-1133.
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. 2011. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333(6047): 1300-1303.
- Ito T, Bulger M, Pazin MJ, Kobayashi R, Kadonaga JT. 1997. ACF, an ISWI-containing and ATP-utilizing chromatin assembly and remodeling factor. *Cell* 90(1): 145-155.
- Iurlaro M, Ficiz G, Oxley D, Raiber EA, Bachman M, Booth MJ, Andrews S, Balasubramanian S, Reik W. 2013. A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol* 14(10): R119.
- Iyer LM, Tahiliani M, Rao A, Aravind L. 2009. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8(11): 1698-1710.
- Jacobson RH, Ladurner AG, King DS, Tjian R. 2000. Structure and function of a human TAFII250 double bromodomain module. *Science* 288(5470): 1422-1425.
- Jin C, Felsenfeld G. 2007. Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* 21(12): 1519-1529.
- Jin C, Zang C, Wei G, Cui K, Peng W, Zhao K, Felsenfeld G. 2009. H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat Genet* 41(8): 941-945.
- Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, Yen CA, Schmitt AD, Espinoza CA, Ren B. 2013. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* 503(7475): 290-294.
- Johansson O, Alkema W, Wasserman WW, Lagergren J. 2003. Identification of functional clusters of transcription factor binding motifs in genome sequences: the MSCAN algorithm. *Bioinformatics* 19 Suppl 1: i169-176.
- Jones PA, Liang G. 2009. Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet* 10(11): 805-811.
- Jones PL, Veenstra GJ, Wade PA, Vermaak D, Kass SU, Landsberger N, Strouboulis J, Wolffe AP. 1998. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* 19(2): 187-191.
- Kagey MH, Newman JJ, Bilodeau S, Zhan Y, Orlando DA, van Berkum NL, Ebmeier CC, Goossens J, Rahl PB, Levine SS, Taatjes DJ, Dekker J, Young RA. 2010. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 467(7314): 430-435.
- Kaikkonen MU, Spann NJ, Heinz S, Romanoski CE, Allison KA, Stender JD, Chun HB, Tough DF, Prinjha RK, Benner C, Glass CK. 2013. Remodeling of the enhancer landscape during macrophage activation is coupled to enhancer transcription. *Mol Cell* 51(3): 310-325.
- Kamiya H, Tsuchiya H, Karino N, Ueno Y, Matsuda A, Harashima H. 2002. Mutagenicity of 5-formylcytosine, an oxidation product of 5-methylcytosine, in DNA in mammalian cells. *J Biochem* 132(4): 551-555.

- Kanemaki M, Kurokawa Y, Matsu-ura T, Makino Y, Masani A, Okazaki K, Morishita T, Tamura TA. 1999. TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J Biol Chem* 274(32): 22437-22444.
- Kangaspeska S, Stride B, Métivier R, Polycarpou-Schwarz M, Ibberson D, Carmouche RP, Benes V, Gannon F, Reid G. 2008. Transient cyclical methylation of promoter DNA. *Nature* 452(7183): 112-115.
- Kellinger MW, Song CX, Chong J, Lu XY, He C, Wang D. 2012. 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol* 19(8): 831-833.
- Kieffer-Kwon KR, Tang Z, Mathe E, Qian J, Sung MH, Li G, Resch W, Baek S, Pruett N, Grøntved L, Vian L, Nelson S, Zare H, Hakim O, Reyon D, Yamane A, Nakahashi H, Kovalchuk AL, Zou J, Joung JK, Sartorelli V, Wei CL, Ruan X, Hager GL, Ruan Y, Casellas R. 2013. Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. *Cell* 155(7): 1507-1520.
- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* 128(6): 1231-1245.
- Kim TK, Hemberg M, Gray JM, Costa AM, Bear DM, Wu J, Harmin DA, Laptevich M, Barbara-Haley K, Kuersten S, Markenscoff-Papadimitriou E, Kuhl D, Bito H, Worley PF, Kreiman G, Greenberg ME. 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465(7295): 182-187.
- Kim TK, Shiekhata R. 2015. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* 162(5): 948-959.
- Kim YW, Lee S, Yun J, Kim A. 2015. Chromatin looping and eRNA transcription precede the transcriptional activation of gene in the β -globin locus. *Biosci Rep* 35(2). pii: e00179.
- Kim, Y. J., Bjorklund, S., Li, Y., Sayre, M. H. & Kornberg, R. D. 1994. A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II. *Cell* 77(4): 599-608.
- Kimura H, Cook PR. 2001. Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J Cell Biol* 153(7): 1341-1353.
- Klose RJ, Bird AP. 2004. MeCP2 behaves as an elongated monomer that does not stably associate with the Sin3a chromatin remodeling complex. *J Biol Chem* 279(45): 46490-46496.
- Ko M, An J, Bandukwala HS, Chavez L, Aijö T, Pastor WA, Segal MF, Li H, Koh KP, Lähdesmäki H, Hogan PG, Aravind L, Rao A. 2013. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497(7447): 122-126.
- Kobor MS, Venkatasubrahmanyam S, Meneghini MD, Gin JW, Jennings JL, Link AJ, Madhani HD, Rine J. 2004. A protein complex containing the conserved Swi2/Snf2-related ATPase Swr1p deposits histone variant H2A.Z into euchromatin. *PLoS Biol* 2(5):E131.

- Koch CM, Andrews RM, Flicek P, Dillon SC, Karaöz U, Clelland GK, Wilcox S, Beare DM, Fowler JC, Couttet P, James KD, Lefebvre GC, Bruce AW, Dovey OM, Ellis PD, Dhami P, Langford CF, Weng Z, Birney E, Carter NP, Vetric D, Dunham I. 2007. The landscape of histone modifications across 1% of the human genome in five human cell lines. *Genome Res* 17(6): 691-707.
- Koch F, Fenouil R, Gut M, Cauchy P, Albert TK, Zacarias-Cabeza J, Spicuglia S, de la Chapelle AL, Heidemann M, Hintermair C, Eick D, Gut I, Ferrier P, Andrau JC. 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nat Struct Mol Biol* 18(8): 956-963.
- Kodzius R, Kojima M, Nishiyori H, Nakamura M, Fukuda S, Tagami M, Sasaki D, Imamura K, Kai C, Harbers M, Hayashizaki Y, Carninci P. 2006. CAGE: cap analysis of gene expression. *Nat Methods* 3(3): 211-222.
- Koleske AJ, Young RA. 1994. An RNA polymerase II holoenzyme responsive to activators. *Nature* 368(6470): 466-469.
- Kosak ST, Skok JA, Medina KL, Riblet R, Le Beau MM, Fisher AG, Singh H. 2002. Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296(5565): 158-62.
- Kriaucionis S, Heintz N. 2009. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324(5929): 929-930.
- Ku M, Jaffe JD, Koche RP, Rheinbay E, Endoh M, Koseki H, Carr SA, Bernstein BE. 2012. H2A.Z landscapes and dual modifications in pluripotent and multipotent stem cells underlie complex genome regulatory functions. *Genome Biol* 13(10): R85.
- Kulaeva OI, Gaykalova DA, Studitsky VM. 2007. Transcription through chromatin by RNA polymerase II: histone displacement and exchange. *Mutat Res* 618(1-2): 116-129.
- Kumaran RI, Spector DL. 2008. A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence. *J Cell Biol* 180(1): 51-65.
- Kwon H, Imbalzano AN, Khavari PA, Kingston RE, Green MR. 1994. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* 370(6489): 477-481.
- Lachner M, O'Carroll D, Rea S, Mechtler K, Jenuwein T. 2001. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410(6824): 116-120.
- Latchman DS. 1990. Eukaryotic transcription factors. *Biochem J* 270(2): 281-2819.
- Lavin Y, Winter D, Blecher-Gonen R, David E, Keren-Shaul H, Merad M, Jung S, Amit I. 2014. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell* 159(6): 1312-1326.
- Lee DH, Gershenzon N, Gupta M, Ioshikhes IP, Reinberg D, Lewis BA. 2005b. Functional characterization of core promoter elements: the downstream core element is recognized by TAF1. *Mol Cell Biol* 25(21): 9674-9686.
- Lee MP, Howcroft K, Kotekar A, Yang HH, Buetow KH, Singer DS. 2005a. ATG deserts define a novel core promoter subclass. *Genome Res* 15(9): 1189-1197.

- Lehnertz B, Ueda Y, Derijck AA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters AH. 2003. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Curr Biol* 13(14): 1192-1200.
- Lenhard B, Sandelin A, Carninci P. 2012. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet* 13(4): 233-245.
- Leonhardt H, Page AW, Weier HU, Bestor TH. 1992. A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* 71(5): 865-873.
- LeRoy G, Orphanides G, Lane WS, Reinberg D. 1998. Requirement of RSF and FACT for transcription of chromatin templates in vitro. *Science* 282(5395): 1900-1904.
- Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A. 1992. Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69(6): 905-914.
- Li E, Beard C, Jaenisch R. 1993. Role for DNA methylation in genomic imprinting. *Nature* 366(6453): 362-365.
- Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, Zhang Z, Landt S, Raha D, Euskirchen G, Wei CL, Ge W, Wang H, Davis C, Fisher-Aylor KI, Mortazavi A, Gerstein M, Gingeras T, Wold B, Sun Y, Fullwood MJ, Cheung E, Liu E, Sung WK, Snyder M, Ruan Y. 2012b. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* 148(1-2): 84-98.
- Li W, Notani D, Ma Q, Tanasa B, Nunez E, Chen AY, Merkurjev D, Zhang J, Ohgi K, Song X, Oh S, Kim HS, Glass CK, Rosenfeld MG. 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* 498(7455): 516-520.
- Li Z, Dai H, Martos SN, Xu B, Gao Y, Li T, Zhu G, Schones DE, Wang Z. 2015. Distinct roles of DNMT1-dependent and DNMT1-independent methylation patterns in the genome of mouse embryonic stem cells. *Genome Biol* 16: 115.
- Li Z, Gadue P, Chen K, Jiao Y, Tuteja G, Schug J, Li W, Kaestner KH. 2012a. Foxa2 and H2A.Z mediate nucleosome depletion during embryonic stem cell differentiation. *Cell* 151(7): 1608-1616.
- Liber D, Domaschitz R, Holmqvist PH, Mazzarella L, Georgiou A, Leleu M, Fisher AG, Labosky PA, Dillon N. 2010. Epigenetic priming of a pre-B cell-specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. *Cell Stem Cell* 7(1): 114-126.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, Sandstrom R, Bernstein B, Bender MA, Groudine M, Gnirke A, Stamatoyannopoulos J, Mirny LA, Lander ES, Dekker J. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326(5950): 289-293.
- Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, Murre C. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol* 13(12): 1196-1204.

- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsall L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462(7271): 315-322.
- Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, Shin H, Wong SS, Ma J, Lei Y, Pape UJ, Poidinger M, Chen Y, Yeung K, Brown M, Turpaz Y, Liu XS. 2011. Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol* 12(8): R83.
- Liu Y, Olanrewaju YO, Zhang X, Cheng X. 2013. DNA recognition of 5-carboxylcytosine by a Zfp57 mutant at an atomic resolution of 0.97 Å. *Biochemistry* 52(51): 9310-9317.
- Long HK, Blackledge NP, Klose RJ. 2013. ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem Soc Trans* 41(3): 727-740.
- Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153(2): 320-334.
- Lu F, Liu Y, Jiang L, Yamaguchi S, Zhang Y. 2014. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* 28(19): 2103-2119.
- Lu J, Hu L, Cheng J, Fang D, Wang C, Yu K, Jiang H, Cui Q, Xu Y, Luo C. 2016. A computational investigation on the substrate preference of ten-eleven-translocation 2 (TET2). *Phys Chem Chem Phys* Jan 22. [Epub ahead of print]
- Luger K, Mäder AW, Richmond RK, Sargent DF, Richmond TJ. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648): 251-260.
- Luk E, Vu ND, Patteson K, Mizuguchi G, Wu WH, Ranjan A, Backus J, Sen S, Lewis M, Bai Y, Wu C. 2007. Chz1, a nuclear chaperone for histone H2AZ. *Mol Cell* 25(3): 357-368.
- Lundgren M, Chow CM, Sabbattini P, Georgiou A, Minaee S, Dillon N. 2000. Transcription factor dosage affects changes in higher order chromatin structure associated with activation of a heterochromatic gene. *Cell* 103(5): 733-743.
- Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, Horn D, Kayserili H, Opitz JM, Laxova R, Santos-Simarro F, Gilbert-Dussardier B, Wittler L, Borschiwer M, Haas SA, Osterwalder M, Franke M, Timmermann B, Hecht J, Spielmann M, Visel A, Mundlos S. 2015. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161(5): 1012-1025.
- Lupien M, Eeckhoutte J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. 2008. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* 132(6): 958-970.
- Lyst MJ, Ekiert R, Ebert DH, Merusi C, Nowak J, Selfridge J, Guy J, Kastan NR, Robinson ND, de Lima Alves F, Rappsilber J, Greenberg ME, Bird A. 2013. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci* 16(7): 898-902.
- Macleod D, Charlton J, Mullins J, Bird AP. 1994. Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. *Genes Dev* 8(19): 2282-2292.

- Magnani L, Ballantyne EB, Zhang X, Lupien M. 2011. PBX1 genomic pioneer function drives ER α signaling underlying progression in breast cancer. *PLoS Genet* 7(11): e1002368.
- Mahé EA, Salbert G. 2015. DNA demethylation by TET proteins: a potential therapeutic target in cancer. *Epigenetic Diagnosis & Therapy* 1: 49-59.
- Maiti A, Drohat AC. 2011. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem* 286(41): 35334-35338.
- Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Münch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. 2003. TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31(1): 374-378.
- Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF. 2010. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 466(7303): 253-257.
- Mayer R, Brero A, von Hase J, Schroeder T, Cremer T, Dietzel S. 2005. Common themes and cell type specific variations of higher order chromatin arrangements in the mouse. *BMC Cell Biol* 6: 44.
- McKittrick E, Gafken PR, Ahmad K, Henikoff S. 2004. Histone H3.3 is enriched in covalent modifications associated with active chromatin. *Proc Natl Acad Sci U S A* 101(6): 1525-1530.
- McManus S, Ebert A, Salvagiotto G, Medvedovic J, Sun Q, Tamir I, Jaritz M, Tagoh H, Busslinger M. 2011. The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells. *EMBO J* 30(12): 2388-2404.
- Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP. 1989. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58(3): 499-507.
- Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES. 2008. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454(7205): 766-770.
- Melgar MF, Collins FS, Sethupathy P. 2011. Discovery of active enhancers through bidirectional expression of short transcripts. *Genome Biol* 12(11): R113.
- Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N. 2012. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151(7): 1417-1430.
- Mendonca A, Chang EH, Liu W, Yuan C. 2014. Hydroxymethylation of DNA influences nucleosomal conformation and stability in vitro. *Biochim Biophys Acta* 1839(11): 1323-1329.
- Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. 2006. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* 10(1): 105-116.

- Métivier R, Gallais R, Tiffocche C, Le Péron C, Jurkowska RZ, Carmouche RP, Ibberson D, Barath P, Demay F, Reid G, Benes V, Jeltsch A, Gannon F, Salbert G. 2008. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452(7183): 45-50.
- Miele A, Dekker J. 2009. Mapping cis- and trans- chromatin interaction networks using chromosome conformation capture (3C). *Methods Mol Biol* 464: 105-21.
- Mishiro T, Ishihara K, Hino S, Tsutsumi S, Aburatani H, Shirahige K, Kinoshita Y, Nakao M. 2009. Architectural roles of multiple chromatin insulators at the human apolipoprotein gene cluster. *EMBO J* 28(9): 1234-1245.
- Mito Y, Henikoff JG, Henikoff S. 2005. Genome-scale profiling of histone H3.3 replacement patterns. *Nat Genet* 37(10): 1090-1097.
- Mizuguchi G, Shen X, Landry J, Wu WH, Sen S, Wu C. 2004. ATP-driven exchange of histone H2AZ variant catalyzed by SWR1 chromatin remodeling complex. *Science* 303(5656): 343-348.
- Mohandas T, Sparkes RS, Shapiro LJ. 1981. Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* 211(4480): 393-396.
- Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schübeler D. 2008. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol Cell* 30(6): 755-66.
- Montavon T, Soshnikova N, Mascres B, Joye E, Thevenet L, Splinter E, de Laat W, Spitz F, Duboule D. 2011. A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147(5): 1132-1145.
- Mullen AC, Orlando DA, Newman JJ, Lovén J, Kumar RM, Bilodeau S, Reddy J, Guenther MG, DeKoter RP, Young RA. 2011. Master transcription factors determine cell-type-specific responses to TGF- β signaling. *Cell* 147(3): 565-576.
- Müller U, Bauer C, Siegl M, Rottach A, Leonhardt H. 2014. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res* 42(13): 8592-8604.
- Münzel M, Lischke U, Stathis D, Pfaffeneder T, Gnerlich FA, Deiml CA, Koch SC, Karaghiosoff K, Carell T. 2011. Improved synthesis and mutagenicity of oligonucleotides containing 5-hydroxymethylcytosine, 5-formylcytosine and 5-carboxylcytosine. *Chemistry* 17(49): 13782-13788.
- Nan X, Ng HH, Johnson CA, Laherty CD, Turner BM, Eisenman RN, Bird A. 1998. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393(6683): 386-389.
- Natoli G, Andrau JC. 2012. Noncoding transcription at enhancers: general principles and functional models. *Annu Rev Genet* 46: 1-19.
- Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK, Maurano MT, Humbert R, Rynes E, Wang H, Vong S, Lee K, Bates D, Diegel M, Roach V, Dunn D, Neri J, Schafer A, Hansen RS, Kuttyavin T, Giste E, Weaver M, Canfield T, Sabo P, Zhang M, Balasundaram G, Byron R, MacCoss MJ, Akey JM, Bender MA, Groudine M, Kaul R, Stamatoyannopoulos JA. 2012. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489(7414): 83-90.

- Neri F, Incarnato D, Krepelova A, Rapelli S, Anselmi F, Parlato C, Medana C, Dal Bello F, Oliviero S. 2015. Single-Base Resolution Analysis of 5-Formyl and 5-Carboxyl Cytosine Reveals Promoter DNA Methylation Dynamics. *Cell Rep* pii: S2211-1247(15)00009-1.
- Ng HH, Zhang Y, Hendrich B, Johnson CA, Turner BM, Erdjument-Bromage H, Tempst P, Reinberg D, Bird A. 1999. MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat Genet* 23(1): 58-61.
- Ni Z, Saunders A, Fuda NJ, Yao J, Suarez JR, Webb WW, Lis JT. 2008. P-TEFb is critical for the maturation of RNA polymerase II into productive elongation in vivo. *Mol Cell Biol* 28(3): 1161-1170.
- Nobrega MA, Ovcharenko I, Afzal V, Rubin EM. 2003. Scanning human gene deserts for long-range enhancers. *Science* 302(5644): 413.
- Nora EP, Lajoie BR, Schulz EG, Giorgetti L, Okamoto I, Servant N, Piolot T, van Berkum NL, Meisig J, Sedat J, Gribnau J, Barillot E, Blüthgen N, Dekker J, Heard E. 2012. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485(7398): 381-385.
- Obri A, Ouarrhni K, Papin C, Diebold ML, Padmanabhan K, Marek M, Stoll I, Roy L, Reilly PT, Mak TW, Dimitrov S, Romier C, Hamiche A. 2014. ANP32E is a histone chaperone that removes H2A.Z from chromatin. *Nature* 505(7485): 648-653.
- Okamura K, Matsumoto KA, Nakai K. 2010. Gradual transition from mosaic to global DNA methylation patterns during deuterostome evolution. *BMC Bioinformatics* 11 Suppl 7:S2.
- Okano M, Xie S, Li E. 1998. Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat Genet* 19(3): 219-220.
- Okano M1, Bell DW, Haber DA, Li E. 1999. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99(3): 247-257.
- Okashita N, Kumaki Y, Ebi K, Nishi M, Okamoto Y, Nakayama M, Hashimoto S, Nakamura T, Sugawara K, Kojima N, Takada T, Okano M, Seki Y. 2014. PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development* 141(2): 269-280.
- Okuwaki M, Kato K, Shimahara H, Tate S, Nagata K. 2005. Assembly and disassembly of nucleosome core particles containing histone variants by human nucleosome assembly protein I. *Mol Cell Biol* 25(23): 10639-10651.
- Ong CT, Corces VG. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet* 12(4):283-93.
- Ooi SK, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH. 2007. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448(7154): 714-717.
- Orphanides G, Wu WH, Lane WS, Hampsey M, Reinberg D. 1999. The chromatin-specific transcription elongation factor FACT comprises human SPT16 and SSRP1 proteins. *Nature* 400(6741): 284-288.

- Ostuni R, Piccolo V, Barozzi I, Polletti S, Termanini A, Bonifacio S, Curina A, Prosperini E, Ghisletti S, Natoli G. 2013. Latent enhancers activated by stimulation in differentiated cells. *Cell* 152(1-2): 157-171.
- Otani J, Nankumo T, Arita K, Inamoto S, Ariyoshi M, Shirakawa M. 2009. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX-DNMT3-DNMT3L domain. *EMBO Rep* 10(11): 1235-1241.
- Papamichos-Chronakis M, Watanabe S, Rando OJ, Peterson CL. 2011. Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. *Cell* 144(2): 200-213.
- Parada LA, McQueen PG, Misteli T. 2004. Tissue-specific spatial organization of genomes. *Genome Biol* 5(7): R44.
- Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, Cobb BS, Yokomori K, Dillon N, Aragon L, Fisher AG, Merkschlager M. 2008. Cohesins functionally associate with CTCF on mammalian chromosome arms. *Cell* 132(3): 422-433.
- Park YJ, Chodaparambil JV, Bao Y, McBryant SJ, Luger K. 2005. Nucleosome assembly protein 1 exchanges histone H2A-H2B dimers and assists nucleosome sliding. *J Biol Chem* 280(3): 1817-1825.
- Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R, Ko M, McLoughlin EM, Brudno Y, Mahapatra S, Kapranov P, Tahiliani M, Daley GQ, Liu XS, Ecker JR, Milos PM, Agarwal S, Rao A. 2011. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473(7347): 394-397.
- Patikoglou GA, Kim JL, Sun L, Yang SH, Kodadek T, Burley SK. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev* 13(24): 3217-3230.
- Pehrson JR, Fried VA. 1992. MacroH2A, a core histone containing a large nonhistone region. *Science* 257(5075): 1398-1400.
- Pekowska A, Benoukraf T, Zacarias-Cabeza J, Belhocine M, Koch F, Holota H, Imbert J, Andrau JC, Ferrier P, Spicuglia S. 2011. H3K4 tri-methylation provides an epigenetic signature of active enhancers. *EMBO J* 30(20): 4198-4210.
- Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M, Minovitsky S, Dubchak I, Holt A, Lewis KD, Plajzer-Frick I, Akiyama J, De Val S, Afzal V, Black BL, Couronne O, Eisen MB, Visel A, Rubin EM. 2006. In vivo enhancer analysis of human conserved non-coding sequences. *Nature* 444(7118): 499-502.
- Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SW, Solovei I, Brugman W, Gräf S, Flicek P, Kerkhoven RM, van Lohuizen M, Reinders M, Wessels L, van Steensel B. 2010. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. *Mol Cell* 38(4): 603-613.
- Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E, Yusuf D, Lenhard B, Wasserman WW, Sandelin A. 2010. JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* 38(Database issue): D105-110.
- Pott S, Lieb JD. 2015. What are super-enhancers? *Nat Genet* 47(1): 8-12.
- Pradhan M, Estève PO, Chin HG, Samaranayake M, Kim GD, Pradhan S. 2008. CXXC domain of human DNMT1 is essential for enzymatic activity. *Biochemistry* 47(38): 10000-10009.

- Putiri EL, Tiedemann RL, Thompson JJ, Liu C, Ho T, Choi JH, Robertson KD. 2014. Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. *Genome Biol* 15(6): R81.
- Qin P, Haberbush JM, Zhang Z, Soprano KJ, Soprano DR. 2004. Pre-B cell leukemia transcription factor (PBX) proteins are important mediators for retinoic acid-dependent endodermal and neuronal differentiation of mouse embryonal carcinoma P19 cells. *J Biol Chem* 279(16): 16263-16271.
- Quintin J, Le Péron C, Palierne G, Bizot M, Cunha S, Sérandour AA, Avner S, Henry C, Percevault F, Belaud-Rotureau MA, Huet S, Watrin E, Eeckhoutte J, Legagneux V, Salbert G, Métivier R. 2014. Dynamic estrogen receptor interactomes control estrogen-responsive trefoil Factor (TFF) locus cell-specific activities. *Mol Cell Biol* 34(13): 2418-2436.
- Rabl C. 1885. Über Zelltheilung. *Morph Jb* 10: 214–330.
- Rach EA, Winter DR, Benjamin AM, Corcoran DL, Ni T, Zhu J, Ohler U. 2011. Transcription initiation patterns indicate divergent strategies for gene regulation at the chromatin level. *PLoS Genet* 7(1): e1001274.
- Rada-Iglesias A, Bajpai R, Swigut T, Bruggmann SA, Flynn RA, Wysocka J. 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470(7333): 279-283.
- Raddatz G, Guzzardo PM, Olova N, Fantappiè MR, Rampp M, Schaefer M, Reik W, Hannon GJ, Lyko F. 2013. Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc Natl Acad Sci U S A* 110(21): 8627-8631.
- Raiber EA, Beraldi D, Ficiz G, Burgess HE, Branco MR, Murat P, Oxley D, Booth MJ, Reik W, Balasubramanian S. 2012. Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol* 13(8): R69.
- Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159(7): 1665-1680.
- Reddy KL, Zullo JM, Bertolino E, Singh H. 2008. Transcriptional repression mediated by repositioning of genes to the nuclear lamina. *Nature* 452(7184): 243-247.
- Roadmap Epigenomics Consortium. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* 518(7539): 317-330.
- Robertson KD. 2005. DNA methylation and human disease. *Nat Rev Genet* 6(8): 597-610.
- Roeder RG. 1996. The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 21(9): 327-335.
- Roh TY, Cuddapah S, Zhao K. 2005. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev* 19(5): 542-552.
- Roy AL, Singer DS. 2015. Core promoters in transcription: old problem, new insights. *Trends Biochem Sci* 40(3): 165-171.

- Rubio ED, Reiss DJ, Welcsh PL, Disteché CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci U S A* 105(24): 8309-8314.
- Saitou M, Kagiwada S, Kurimoto K. 2012. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* 139(1): 15-31.
- Samstein RM, Arvey A, Josefowicz SZ, Peng X, Reynolds A, Sandstrom R, Neph S, Sabo P, Kim JM, Liao W, Li MO, Leslie C, Stamatoyannopoulos JA, Rudensky AY. 2012. Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. *Cell* 151(1): 153-166.
- Sanyal A, Lajoie BR, Jain G, Dekker J. 2012. The long-range interaction landscape of gene promoters. *Nature* 489(7414): 109-113.
- Sarcinella E, Zuzarte PC, Lau PN, Draker R, Cheung P. 2007. Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. *Mol Cell Biol* 27(18): 6457-6468.
- Sarraf SA, Stancheva I. 2004. Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol Cell* 15(4): 595-605.
- Sayed SK, Zhao J, Sathyanarayana BK, Golla JP, Vinson C. 2015. C/EBP β (CEBPB) protein binding to the C/EBP|CRE DNA 8-mer TTGC|GTCA is inhibited by 5hmC and enhanced by 5mC, 5fC, and 5caC in the CG dinucleotide. *Biochim Biophys Acta* 1849(6): 583-589.
- Schaaf CA, Kwak H, Koenig A, Misulovin Z, Gohara DW, Watson A, Zhou Y, Lis JT, Dorsett D. 2013. Genome-wide control of RNA polymerase II activity by cohesin. *PLoS Genet* 9(3): e1003382.
- Schmidl C, Klug M, Boeld TJ, Andreesen R, Hoffmann P, Edinger M, Rehli M. 2009. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome Res* 19(7): 1165-1174.
- Schmidt D, Wilson MD, Spyrou C, Brown GD, Hadfield J, Odom DT. 2009. ChIP-seq: using high-throughput sequencing to discover protein-DNA interactions. *Methods* 48(3): 240-248.
- Schübeler D. 2015. ESCI award lecture: regulation, function and biomarker potential of DNA methylation. *Eur J Clin Invest* 45(3): 288-293.
- Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ Jr. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol* 6(4): R33.
- Scott EW, Simon MC, Anastasi J, Singh H. 1994. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* 265(5178): 1573-1577.
- Scruggs BS, Gilchrist DA, Nechaev S, Muse GW, Burkholder A, Fargo DC, Adelman K. 2015. Bidirectional Transcription Arises from Two Distinct Hubs of Transcription Factor Binding and Active Chromatin. *Mol Cell* 58(6): 1101-1112.
- Sérandour AA, Avner S, Oger F, Bizot M, Percevault F, Lucchetti-Miganeh C, Palierne G, Gheeraert C, Barloy-Hubler F, Péron CL, Madigou T, Durand E, Froguel P, Staels B, Lefebvre P, Métivier R, Eeckhoutte J, Salbert G. 2012. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* 40(17): 8255-8265.

- Sérandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Métivier R, Salbert G, Eeckhoute J. 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* 21(4): 555-565.
- Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. 2012. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148(3): 458-472.
- Shakya A, Callister C, Goren A, Yosef N, Garg N, Khoddami V, Nix D, Regev A, Tantin D. 2015. Pluripotency transcription factor Oct4 mediates stepwise nucleosome demethylation and depletion. *Mol Cell Biol* 35(6): 1014-25.
- Sharif J, Muto M, Takebayashi S, Suetake I, Iwamatsu A, Endo TA, Shinga J, Mizutani-Koseki Y, Toyoda T, Okamura K, Tajima S, Mitsuya K, Okano M, Koseki H. 2007. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450(7171): 908-912.
- Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, Shu J, Chen X, Waterland RA, Issa JP. 2007. Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet* 3(10): 2023-2036.
- Shen L, Wu H, Diep D, Yamaguchi S, D'Alessio AC, Fung HL, Zhang K, Zhang Y. 2013. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153(3): 692-706.
- Shen X, Mizuguchi G, Hamiche A, Wu C. 2000. A chromatin remodelling complex involved in transcription and DNA processing. *Nature* 406(6795): 541-544.
- Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, Wagner U, Dixon J, Lee L, Lobanenkov VV, Ren B. 2012. A map of the cis-regulatory sequences in the mouse genome. *Nature* 488(7409): 116-120.
- Shi FT, Kim H, Lu W, He Q, Liu D, Goodell MA, Wan M, Songyang Z. 2013. Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J Biol Chem* 288(29): 20776-20784.
- Shi J, Zheng M, Ye Y, Li M, Chen X, Hu X1, Sun J, Zhang X, Jiang C. 2014. *Drosophila* Brahma complex remodels nucleosome organizations in multiple aspects. *Nucleic Acids Res* 42(15): 9730-9739.
- Shibutani T, Ito S, Toda M, Kanao R, Collins LB, Shibata M, Urabe M, Koseki H, Masuda Y, Swenberg JA, Masutani C, Hanaoka F, Iwai S, Kuraoka I. 2014. Guanine- 5-carboxylcytosine base pairs mimic mismatches during DNA replication. *Sci Rep* 4:5220.
- Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ, Miller W, Haussler D. 2005. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 15(8): 1034-1050.
- Siersbæk R, Nielsen R, John S, Sung MH, Baek S, Loft A, Hager GL, Mandrup S. 2011. Extensive chromatin remodelling and establishment of transcription factor 'hotspots' during early adipogenesis. *EMBO J* 30(8): 1459-1472.
- Simonis M, Klous P, Splinter E, Moshkin Y, Willemsen R, de Wit E, van Steensel B, de Laat W. 2006. Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* 38(11): 1348-1354.

- Smith ZD, Meissner A. 2013. DNA methylation: roles in mammalian development. *Nat Rev Genet* 14(3): 204-220.
- Song CX, Szulwach KE, Dai Q, Fu Y, Mao SQ, Lin L, Street C, Li Y, Poidevin M, Wu H, Gao J, Liu P, Li L, Xu GL, Jin P, He C. 2013. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153(3): 678-691.
- Song CX, Szulwach KE, Fu Y, Dai Q, Yi C, Li X, Li Y, Chen CH, Zhang W, Jian X, Wang J, Zhang L, Looney TJ, Zhang B, Godley LA, Hicks LM, Lahn BT, Jin P, He C. 2011. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* 29(1): 68-72.
- Song F, Mahmood S, Ghosh S, Liang P, Smiraglia DJ, Nagase H, Held WA. 2009. Tissue specific differentially methylated regions (TDMR): Changes in DNA methylation during development. *Genomics* 93(2): 130-139.
- Spector DL. 2003. The dynamics of chromosome organization and gene regulation. *Annu Rev Biochem* 72: 573-608.
- Spellman PT, Rubin GM. 2002. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J Biol* 1(1): 5.
- Spilianakis CG, Flavell RA. 2004. Long-range intrachromosomal interactions in the T helper type 2 cytokine locus. *Nat Immunol* 5(10): 1017-1027.
- Sprijt CG, Gnerlich F, Smits AH, Pfaffeneder T, Jansen PW, Bauer C, Münzel M, Wagner M, Müller M, Khan F, Eberl HC, Mensinga A, Brinkman AB, Lephikov K, Müller U, Walter J, Boelens R, van Ingen H, Leonhardt H, Carell T, Vermeulen M. 2013. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152(5): 1146-1159.
- Stack SM, Brown DB, Dewey WC. 1977. Visualization of interphase chromosomes. *J Cell Sci* 26: 281-299.
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schübeler D. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480(7378): 490-495.
- Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 2011. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* 12(6): R54.
- Suetake I, Shinozaki F, Miyagawa J, Takeshima H, Tajima S. 2004. DNMT3L stimulates the DNA methylation activity of Dnmt3a and Dnmt3b through a direct interaction. *J Biol Chem* 279(26): 27816-27823.
- Szenker E, Ray-Gallet D, Almouzni G. 2011. The double face of the histone variant H3.3. *Cell Res* 21(3): 421-434.
- Szulwach KE, Li X, Li Y, Song CX, Han JW, Kim S, Namburi S, Hermetz K, Kim JJ, Rudd MK, Yoon YS, Ren B, He C, Jin P. 2011. Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* 7(6): e1002154.
- Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. 2010. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 38(19): e181.

- Tagami H, Ray-Gallet D, Almouzni G, Nakatani Y. 2004. Histone H3.1 and H3.3 complexes mediate nucleosome assembly pathways dependent or independent of DNA synthesis. *Cell* 116(1): 51-61.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. 2009. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324(5929): 930-935.
- Takayama K, Misawa A, Suzuki T, Takagi K, Hayashizaki Y, Fujimura T, Homma Y, Takahashi S, Urano T, Inoue S. 2015. TET2 repression by androgen hormone regulates global hydroxymethylation status and prostate cancer progression. *Nat Commun* 6: 8219.
- Tanabe H1, Müller S, Neusser M, von Hase J, Calcagno E, Cremer M, Solovei I, Cremer C, Cremer T. 2002. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proc Natl Acad Sci U S A* 99(7): 4424-4429.
- Teif VB, Beshnova DA, Vainshtein Y, Marth C, Mallm JP, Höfer T, Rippe K. 2014. Nucleosome repositioning links DNA (de)methylation and differential CTCF binding during stem cell development. *Genome Res* 24(8):1285-1295.
- Tolstorukov MY, Goldman JA, Gilbert C, Ogryzko V, Kingston RE, Park PJ. 2012. Histone variant H2A.Bbd is associated with active transcription and mRNA processing in human cells. *Mol Cell* 47(4): 596-5607.
- Tong JK, Hassig CA, Schnitzler GR, Kingston RE, Schreiber SL. 1998. Chromatin deacetylation by an ATP-dependent nucleosome remodelling complex. *Nature* 395(6705): 917-921.
- Tsukiyama T, Daniel C, Tamkun J, Wu C. 1995. ISWI, a member of the SWI2/SNF2 ATPase family, encodes the 140 kDa subunit of the nucleosome remodeling factor. *Cell* 83(6): 1021-1026.
- Tsukiyama T, Palmer J, Landel CC, Shiloach J, Wu C. 1999. Characterization of the imitation switch subfamily of ATP-dependent chromatin-remodeling factors in *Saccharomyces cerevisiae*. *Genes Dev* 13(6): 686-697.
- Tsukiyama T, Wu C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* 83(6): 1011-1020.
- Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 17(3): 453-462.
- van Oevelen C, Collombet S, Vicent G, Hoogenkamp M, Lepoivre C, Badeaux A, Bussmann L, Sardina JL, Thieffry D, Beato M, Shi Y, Bonifer C, Graf T. 2015. C/EBP α Activates Pre-existing and De Novo Macrophage Enhancers during Induced Pre-B Cell Transdifferentiation and Myelopoiesis. *Stem Cell Reports* 5(2): 232-247.
- Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LT, Fernandez N, Ballester B, Andrau JC, Spicuglia S. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* 6: 6905.
- Varga-Weisz PD, Wilm M, Bonte E, Dumas K, Mann M, Becker PB. 1997. Chromatin-remodelling factor CHRAC contains the ATPases ISWI and topoisomerase II. *Nature* 388(6642): 598-602.

- Velasco G, Hubé F, Rollin J, Neuillet D, Philippe C, Bouzinba-Segard H, Galvani A, Viegas-Péquignot E, Francastel C. 2010. Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc Natl Acad Sci U S A* 107(20): 9281-9286.
- Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, Roberto A, Christensen J, Bonaldi T, Helin K, Pasini D. 2013. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* 49(4): 645-656.
- Venkatesh S, Workman JL. 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nat Rev Mol Cell Biol* 16(3): 178-189.
- Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, Baltissen MP, Stunnenberg HG, Mann M, Timmers HT. 2007. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 131(1): 58-69.
- Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C, Chen F, Afzal V, Ren B, Rubin EM, Pennacchio LA. 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* 457(7231): 854-858.
- Visel A, Minovitsky S, Dubchak I, Pennacchio LA. 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res* 35(Database issue): D88-92.
- Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. 2000. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol Cell Biol* 20(6): 2108-2121.
- Walsh CP, Chaillet JR, Bestor TH. 1998. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* 20(2): 116-117.
- Wang C, Zhang MQ, Zhang Z. 2013. Computational identification of active enhancers in model organisms. *Genomics Proteomics Bioinformatics* 11(3): 142-150.
- Wang Q, Li W, Liu XS, Carroll JS, Jänne OA, Keeton EK, Chinnaiyan AM, Pienta KJ, Brown M. 2007. A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol Cell* 27(3): 380-392.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ, Zhao K. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat Genet* 40(7): 897-903.
- Watanabe D, Suetake I, Tada T, Tajima S. 2002. Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mech Dev* 118(1-2): 187-190.
- Watt F, Molloy PL. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev* 2(9): 1136-1143.
- Weber M, Hellmann I, Stadler MB, Ramos L, Pääbo S, Rebhan M, Schübeler D. 2007. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat Genet* 39(4): 457-466.

- Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters JM. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* 451(7180): 796-801.
- West JA, Cook A, Alver BH, Stadtfeld M, Deaton AM, Hochedlinger K, Park PJ, Tolstorukov MY, Kingston RE. 2014. Nucleosomal occupancy changes locally over key regulatory regions during cell differentiation and reprogramming. *Nat Commun* 5: 4719.
- Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153(2): 307-319.
- Wienholz BL, Kareta MS, Moarefi AH, Gordon CA, Ginno PA, Chédin F. 2010. DNMT3L modulates significant and distinct flanking sequence preference for DNA methylation by DNMT3A and DNMT3B in vivo. *PLoS Genet* 6(9): e1001106.
- Williams K, Christensen J, Pedersen MT, Johansen JV, Cloos PA, Rappsilber J, Helin K. 2011. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473(7347): 343-348.
- Williams RR, Azuara V, Perry P, Sauer S, Dvorkina M, Jørgensen H, Roix J, McQueen P, Misteli T, Merckenschlager M, Fisher AG. 2006. Neural induction promotes large-scale chromatin reorganisation of the Mash1 locus. *J Cell Sci* 119(Pt 1): 132-140.
- Wu H, D'Alessio AC, Ito S, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011b. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* 25(7): 679-684.
- Wu H, D'Alessio AC, Ito S, Xia K, Wang Z, Cui K, Zhao K, Sun YE, Zhang Y. 2011a. Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473(7347): 389-393.
- Xie S, Wang Z, Okano M, Nogami M, Li Y, He WW, Okumura K, Li E. 1999. Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene* 236(1): 87-95.
- Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, Zaret KS, Weissman IL, Smale ST. 2007. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci U S A* 104(30): 12377-12382.
- Xu J, Watts JA, Pope SD, Gadue P, Kamps M, Plath K, Zaret KS, Smale ST. 2009. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev* 23(24): 2824-2838.
- Xu S, Li W, Zhu J, Wang R, Li Z, Xu GL, Ding J. 2013. Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyl-uracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res* 23(11): 1296-1309.
- Xu Y, Wu F, Tan L, Kong L, Xiong L, Deng J, Barbera AJ, Zheng L, Zhang H, Huang S, Min J, Nicholson T, Chen T, Xu G, Shi Y, Zhang K, Shi YG. 2011. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* 42(4): 451-464.

- Xu Y, Xu C, Kato A, Tempel W, Abreu JG, Bian C, Hu Y, Hu D, Zhao B, Cerovina T, Diao J, Wu F, He HH, Cui Q, Clark E, Ma C, Barbara A, Veenstra GJ, Xu G, Kaiser UB, Liu XS, Sugrue SP, He X, Min J, Kato Y, Shi YG. 2012. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* 151(6): 1200-1213.
- Xue Y, Wong J, Moreno GT, Young MK, Côté J, Wang W. 1998. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* 2(6): 851-861.
- Yaffe E, Tanay A. 2011. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nat Genet* 43(11): 1059-1065.
- Yamada T, Urano-Tashiro Y, Tanaka S, Akiyama H, Tashiro F. 2013. Involvement of crosstalk between Oct4 and Meis1a in neural cell fate decision. *PLoS One* 8(2): e56997.
- Yamaguchi Y, Takagi T, Wada T, Yano K, Furuya A, Sugimoto S, Hasegawa J, Handa H. 1999. NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation. *Cell* 97(1): 41-51.
- Yamashita R, Suzuki Y, Sugano S, Nakai K. 2005. Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity. *Gene* 350(2): 129-136.
- Yang Z, Yik JH, Chen R, He N, Jang MK, Ozato K, Zhou Q. 2005. Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* 19(4): 535-545.
- Yao TP, Oh SP, Fuchs M, Zhou ND, Ch'ng LE, Newsome D, Bronson RT, Li E, Livingston DM, Eckner R. 1998. Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* 93(3): 361-372.
- Yildirim O, Li R, Hung JH, Chen PB, Dong X, Ee LS, Weng Z, Rando OJ, Fazzio TG. 2011. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147(7): 1498-1510.
- Yoon HG, Chan DW, Reynolds AB, Qin J, Wong J. 2003. N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol Cell* 12(3): 723-734.
- Yu M, Hon GC, Szulwach KE, Song CX, Zhang L, Kim A, Li X, Dai Q, Shen Y, Park B, Min JH, Jin P, Ren B, He C. 2012. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 149(6): 1368-1380.
- Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, et al. 2007. RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo. *Nat Genet* 39(12): 1512-1516.
- Zemach A, McDaniel IE, Silva P, Zilberman D. 2010. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328(5980): 916-919.
- Zentner GE, Tesar PJ, Scacheri PC. 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res* 21(8): 1273-1283.
- Zhang H, Zhang X, Clark E, Mulcahey M, Huang S, Shi YG. 2010b. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res* 20(12): 1390-1393.

- Zhang Q, Liu X, Gao W, Li P, Hou J, Li J, Wong J. 2014. Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked β -N-acetylglucosamine transferase (OGT). *J Biol Chem* 289(9): 5986-5996.
- Zhang Q, Zhao K, Shen Q, Han Y, Gu Y, Li X, Zhao D, Liu Y, Wang C, Zhang X, Su X, Liu J, Ge W, Levine RL, Li N, Cao X. 2015. Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. *Nature* 525(7569): 389-393.
- Zhang Y, Jurkowska R, Soeroes S, Rajavelu A, Dhayalan A, Bock I, Rathert P, Brandt O, Reinhardt R, Fischle W, Jeltsch A. 2010a. Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res* 38(13): 4246-4253.
- Zhang Y, LeRoy G, Seelig HP, Lane WS, Reinberg D. 1998. The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. *Cell* 95(2): 279-289.
- Zhang Y, Ng HH, Erdjument-Bromage H, Tempst P, Bird A, Reinberg D. 1999. Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* 13(15): 1924-1935.
- Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E, Mulawadi FH, Sung WK, Nicolis S, Ahituv N, Ruan Y, Wei CL. 2013. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. *Nature* 504(7479): 306-310.
- Zhu J, Adli M, Zou JY, Verstappen G, Coyne M, Zhang X, Durham T, Miri M, Deshpande V, De Jager PL, Bennett DA, Houmard JA, Muoio DM, Onder TT, Camahort R, Cowan CA, Meissner A, Epstein CB, Shores N, Bernstein BE. 2013. Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152(3): 642-654.
- Zilberman D, Coleman-Derr D, Ballinger T, Henikoff S. 2008. Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* 456(7218): 125-129.
- Ziller MJ, Gu H, Müller F, Donaghey J, Tsai LT, Kohlbacher O, De Jager PL, Rosen ED, Bennett DA, Bernstein BE, Gnirke A, Meissner A. 2013. Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500(7463): 477-481.

ANNEXE

DNA Demethylation by TET Proteins: A Potential Therapeutic Target in Cancer

Elise A. Mahé^{1,2} and Gilles Salbert^{*,1,2,3}

¹*Equipe SP@RTE, CNRS UMR6290, Institut de Génétique et Développement de Rennes, Campus de Beaulieu, 35042 Rennes Cedex, France*

²*Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France*



Gilles Salbert

Abstract: DNA methylation in unique genomic regions like enhancers and promoters contributes to cell identity by silencing gene expression. Hence, establishment of cell-specific repertoires of active genes during cell differentiation or in response to extracellular signals may require erasure of this epigenetic mark. By their ability to hydroxylate the methyl group of 5-methylcytosine (5mC), Ten-eleven translocation (TET) proteins are part of the cellular machinery erasing DNA methylation through molecular mechanisms involving iterative oxidation of 5mC, base excision and repair. Here we review recent discoveries on TET functions and their role in regulating the genome activity to establish and maintain cell identity. We further describe how the TET-mediated processes are affected in cancer cells and propose possible strategies for the correction of these cancer-associated defects.

Keywords: 5-hydroxymethylcytosine, 5hmC, cancer, DNA methylation, TET.

INTRODUCTION

TET proteins have been in the limelight since the (re)discovery by two different laboratories of a new cytosine modification, named as 5-hydroxymethylcytosine (5hmC). Kriaucionis and Heintz identified 5hmC in Purkinje neurons and in the brain [1], while Anjana Rao's laboratory observed this "sixth base" in mouse embryonic stem cells (mESCs) [2], after identification of TET1 as a homolog of the trypanosoma proteins JBP1 and JBP2 which are capable of oxidizing the 5-methyl group of thymine [3]. Rao's group further demonstrated that TET1 was able to convert 5mC to 5hmC in HEK293 cells [2]. These two seminal studies paved the way for new directions of research on DNA methylation and brought new insights on how DNA demethylation could occur.

TET-MEDIATED ACTIVE DNA DEMETHYLATION IN CELL FATE DETERMINATION AND MAINTENANCE

TET Proteins and their Hydroxylase Activity

Metazoan TET family of proteins includes three members (TET1, TET2 and TET3) resulting from a triplication of an ancestral TET gene in jawed vertebrates [2, 3]. All three of them contain a carboxy-terminal catalytic domain (CD) typical of Fe(II)- and α -ketoglutarate (α -KG, or 2-oxoglutarate)-dependent dioxygenases holding a Cysteine-rich domain followed by a double-stranded β -helix (DSBH) region [2, 3]. TET1 and TET3 also contain an amino-terminal CXXC zinc finger domain which functions

as a DNA binding domain [4, 5] and which is absent in TET2 due to a chromosomal inversion during evolution, hence separating the CXXC domain-containing exon from the catalytic region (Fig. 1A). The ancestral CXXC domain became a new gene which encodes IDAX (inhibition of the Dvl and axin complex; also known as CXXC4).

Two recent structural studies of TET2 [6] and the *Naegleria* Tet-like (NgTet1) [7] dioxygenases allowed to further understand TET activity. These structures revealed how TETs specifically recognize CpGs and use of a base-flipping mechanism to insert 5mC into the catalytic cavity [6, 7]. In addition, analysis of TET2 also showed that (i) the Cysteine-rich domain is essential for the catalytic activity of TET proteins, (ii) the catalytic domain is not responsible for TET genomic localization, and (iii) TET2 is able to recognize C- and 5mC-containing DNA with the same affinity [6]. The oxygenase activity first observed for TET1 was also demonstrated for mouse TET2 and TET3 [8] and TET proteins were further shown to generate two additional products by iterative oxidation of 5hmC: 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), in an enzymatic activity-dependent manner [9, 10]. Several studies testing for global 5hmC levels in genomic DNA indicated wide differences in 5hmC enrichment between different cell lines and tissues [11], with the highest levels in neuronal cells and central nervous system [1, 2, 12-14], and the lowest levels in cancer cells [15]. The oxidative products 5fC and 5caC have been detected in mESCs [9, 16-18] and in male and female pronuclei [19], while only 5fC was present at detectable levels in major mouse organs [9].

These findings on TET oxidative activities and the detection of 5mC oxidative products opened up the way for reinvestigating the mechanisms of DNA demethylation. DNA methylation is a well described modification which is involved in cellular differentiation and gene regulation as it mediates transcriptional repression [20]. Transfer of a methyl

*Address correspondence to this author at the UMR6290 CNRS, Université Rennes 1, Campus de Beaulieu, B13, Avenue du Général Leclerc, 35042 Rennes Cedex, France; Tel: +33 223 236 625; Fax: +33 223 236 794; E-mail: gilles.salbert@univ-rennes1.fr

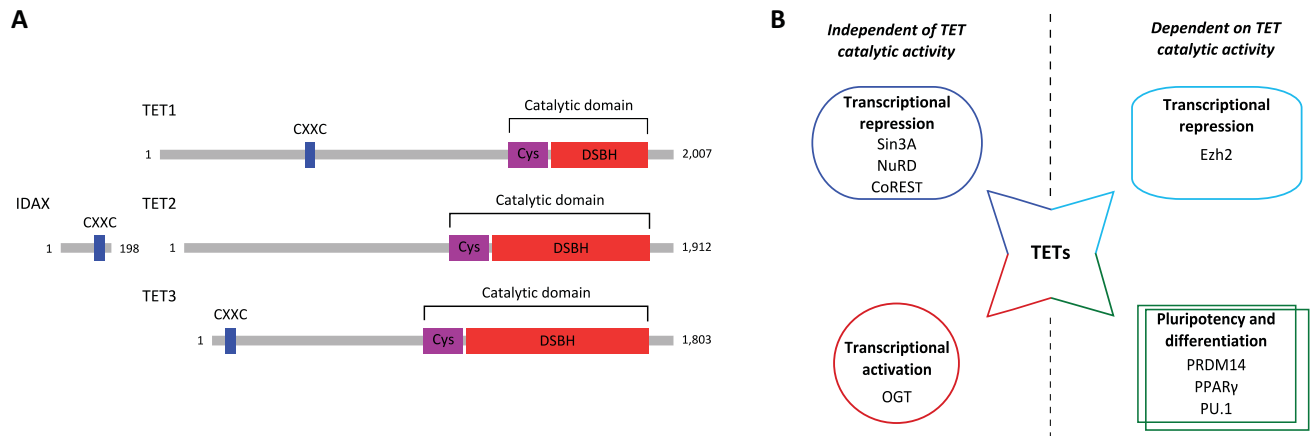


Fig. (1). Domains of TET protein family members, interactions and implications in regulation. (A) Mouse TET family members (TET1, TET2, TET3) are composed of a C-terminal catalytic domain containing a Cysteine-rich region (Cys) and a double-stranded β -helix (DSBH) region characteristic of Fe(II)- and α -ketoglutarate-dependent dioxygenases. In their N-terminal region, TET1 and TET3 comprise a CXXC zinc finger domain which allows for DNA recognition. TET2 lost this domain during evolution as a result of a chromosomal inversion leading to the formation of a new gene, IDAX, containing the CXXC region. Human TET proteins have the same domain organisation. (B) TET enzymes have been shown to interact with a number of proteins or complexes thus regulating several processes independently or dependently of their catalytic activity.

group from S-Adenosyl Methionine (SAM) to cytosine relies on the intervention of DNA Methyltransferases (DNMTs): DNMT1, DNMT3a and DNMT3b. While DNMT1, through its high affinity for hemimethylated CpGs, plays a role in the maintenance of methylation during DNA replication, DNMT3a and DNMT3b proteins are implicated in *de novo* methylation of DNA during cellular differentiation [21]. TET-mediated hydroxymethylation of cytosines was shown to be implicated in the two pathways of 5mC removal: "passive" and "active" DNA demethylation. "Passive" DNA demethylation results from the loss of cytosine modification through replication and was observed during pre-implantation of mouse embryos [19, 22] and in mouse primordial germ cells [23]. A number of studies suggest that 5hmC could favor passive DNA demethylation through incapacitation of the methylation maintenance machinery. Indeed, DNMT1 poorly methylates DNA in the context of 5hmCpG templates, and the DNMT1 partner Ubiquitin-like, containing PHD and RING finger domains, 1 (UHRF1) has a lower affinity for 5hmCpGs than for 5mCpGs [24-26], although this latter result can be questioned [27]. Hence, a complete picture of the molecular events occurring during DNA replication is still missing to understand the role of 5hmC in "passive" DNA demethylation. Oxidation of 5mC is also implicated in "active" demethylation (*i.e.* replication independent) as the oxidation products of 5hmC (5fC and 5caC) can be excised by the Base Excision Repair (BER) machinery through the activity of Thymine DNA glycosylase (TDG) [10, 28] or NEIL glycosylases [29]. Direct decarboxylation of 5caC to unmodified cytosine was also observed in ESC lysates [30], with no clue on a putative decarboxylase until the resolution of the crystal structure of isoorotate decarboxylases (IDCases) and biochemical analysis indicated that these enzymes may mediate 5caC decarboxylation to unmodified cytosine *in vitro* [31]. Another active DNA demethylation pathway was proposed *via* deamination of 5hmC by the activation-induced deaminase (AID)/ apolipoprotein B mRNA-editing enzyme

complex (APOBEC) family of cytidine deaminases [32], leading to the formation of 5-hydroxymethyluracil (5hmU) further excised either by TDG [33] or methyl-CpG binding domain 4 (MBD4) [34]. However, deamination of 5hmC was not observed in another study [35] and a recent work demonstrated that 5hmU detected in mouse ESCs genomic DNA is generated by oxidation of thymine through TET proteins and not by deamination of 5hmC [36]. These data shed light on a new TET-dependent oxidation process and on a putative role for the 5hmU base.

Methods for Mapping TET Oxidative Products

The initial discovery of 5hmC and its derivatives led to a flurry of new techniques allowing to map these cytosine modifications in the genome. DNA immunoprecipitation (DIP) with 5mC antibodies (MeDIP) has proved useful for 5mC detection and mapping genome-wide [37]. Similarly, the development of antibodies against 5hmC, 5caC and 5fC has allowed to study the distribution of 5mC derivatives by DIP [5, 17, 38-43]. In addition, treatment of DNA with sodium bisulphite (BS) converts 5hmC to cytosine 5-methylenesulphonate (CMS) that can be efficiently immunoprecipitated [44]. However, due to a possible CpG density bias, antibody-based methods against cytosine modifications have been criticized. Consequently, alternative methods were developed with many of them using the ability of β -glucosyltransferase (β -GT) from T-even bacteriophage to transfer glucose from a UDP-glucose donor to genomic 5hmC [13]. β -GT-based protocols such as GLIB (glucosylation, periodate oxidation, biotinylation) [44, 45], hMe-Seal [14] or fC-Seal [16] allowed to map 5hmC or 5fC respectively genome-wide.

Despite their efficiency, the above mentioned techniques generate low resolution 5hmC maps precluding the precise identification of the modified cytosines and thus single nucleotide resolution methods have been developed. In single-molecule, real-time (SMRT) sequencing, the presence

of modified bases in the DNA template slows down polymerisation and hence C, 5mC and 5hmC can be discriminated by specific kinetic signatures provided that 5hmC is first modified by β -GT [46]. High resolution mapping of 5hmC can also be achieved through the use of restriction enzymes based on their specific requirements for site recognition: Aba-seq relies on the recognition of glucosylated 5hmC (5gmC) by the restriction enzyme AbaSI which cuts DNA between 11 and 13 nucleotides downstream bound 5gmC [47], and reduced representation 5-hydroxymethylcytosine profiling (RRHP) identifies 5hmC at single-nucleotide resolution using a combination of both MspI restriction enzyme and β -GT [48]. Although being highly reproducible, RRHP restricts detection of 5hmC to cytosines present in the context of a CCGG MspI site (*i.e.* about 14% of all CpGs in the mouse genome).

Additional single nucleotide resolution methods rely on the oxidative action of sodium bisulfite coupled to sequencing. Even if conventional bisulfite sequencing (BS-seq) does not discriminate between 5mC and 5hmC [49], or between C, 5fC and 5caC, protective or deprotective chemical modifications before bisulfite modification can be used to selectively identify Cs, 5mCs and all forms of 5mC oxidative products. Genome-wide mapping of 5hmC at single base resolution can thus be achieved either through oxidative bisulfite sequencing (oxBS-seq) [50, 51] or by TET-assisted bisulfite sequencing (TAB-seq) [52]. Base-resolution mapping of 5fC requires either the conversion of 5fC to 5hmC such as in reduced bisulfite sequencing (redBS-seq) [18], or by protection from deamination in chemically assisted bisulfite sequencing (fCAB-seq) [16]. Protection from BS-mediated deamination is also efficient for mapping 5caC [53]. Finally, in methylase-assisted bisulfite sequencing (MAB-seq), unmodified cytosines are methylated by the MssI methylase prior to BS modification, allowing to discriminate 5fCs and 5caCs from unmodified Cs [54]. All these methods have been developed in a relatively short period of time and further attention will be required to ascertain that they generate reproducible datasets. However, invaluable information on the genome-wide distribution of 5mC oxidative products in different cell types has been already obtained through their implementation.

Genomic Distribution of TET1 and 5mC Oxidative Products

The development of methods to map 5hmC, 5caC and 5fC genome-wide coupled to the mapping of TET proteins by immunoprecipitation of chromatin (ChIP) rapidly expanded our knowledge on the distribution of these marks. TET1 distribution was studied in the genome of mouse ESCs where its binding sites were enriched in gene-rich euchromatic regions [55]. TET1 notably accumulates at gene bodies [5, 40], and at promoters which have a high CpG density and are hypomethylated [5, 38, 40, 55]. The distribution of 5hmC was investigated in numerous cell lines such as mouse and human ESCs [5, 38-41, 44, 52, 56], embryonal carcinoma cells (ECCs) [43, 57], 3T3-L1-derived adipocytes [43], and in cerebellar cell types [14, 58]. Globally, 5hmC is observed along all autosomal chromosome arms but is almost excluded from sex chromosomes [41, 56, 58] and from metaphase

pericentromeric heterochromatin [56]. 5hmC was found enriched in gene bodies and in exons [5, 14, 38-41, 56, 58], whereas it seems to be depleted from Transcription Start Sites (TSSs) [14, 38], although two independent studies observed 5hmC coinciding with TET1 binding sites at TSSs in mESCs [40, 44]. The 5hmC mark was also found to be enriched at promoters [5, 39, 44, 52, 57] and at distal regions [14, 52] such as enhancers [38, 41, 43, 44, 56, 57] or insulators [38]. Accordingly, 5hmC accumulates at binding sites for CTCF and proteins involved in pluripotency [38, 39, 41, 52, 56].

Two recent studies managed to map 5fC/5caC or 5fC modifications [16, 17]. Shen *et al.* observed 5fC and 5caC at major satellite repeats, like 5mC but not 5hmC [17]. Due to the low levels of these two cytosine modifications, the authors depleted TDG in mESCs and found 5fC and 5caC enrichment at several proximal and distal regulatory elements, thus suggesting that TET-TDG-dependent demethylation targets transcriptional enhancers [17]. In addition, Song *et al.* were able to detect 5fC enrichment at poised enhancers compared to other genomic regions. Data obtained in *Tdg* null mESCs further suggested a role of 5fC in priming regulatory elements as it correlates with the binding of P300 at poised enhancers [16].

Tet Proteins and Transcriptional Regulation

Mapping of TET1 and 5hmC genome-wide revealed that they can have common and distinct genomic locations and suggested their involvement in transcriptional regulation. Indeed, 5hmC is enriched within gene bodies of actively transcribed genes, whereas it accumulates specifically at TSSs of transcriptionally inactive genes [5, 14, 38, 39, 41, 56, 58, 59]. These data revealed that 5hmC may have a dual function as it can be associated either with activation or repression of target genes. This observation is also true for TET1 as this protein associates with both actively transcribed (marked with H3K4me3, H3K4me1 and H3K36me3) housekeeping genes as well as repressed bivalent promoters (marked with H3K4me3 and H3K27me3) involved in cellular differentiation [5, 55] and where 5hmC is enriched [44]. This dual function was confirmed by depletion of TET1 which led to both up- and down-regulation of genes [5, 39, 40, 55, 57]. The presence of 5hmC at bivalent promoters and at genes induced during differentiation could therefore reveal a link between this cytosine modification and a poised chromatin state [44, 57].

The putative role of TET1 in promoting either activation or repression of gene expression led to a search for TET-interacting proteins, notably chromatin remodelers or histone modifiers complexes (Fig. 1B). Supporting its role in the repression of transcription, TET1 was found to directly interact with the Sin3A co-repressor complex in an enzymatic activity-independent manner [40, 60-62], and to associate with Nucleosome Remodeling Deacetylase (NurD) and RE1-Silencing Transcription factor corepressor 1 (CoREST) subunits [62]. However, in other studies, TET-mediated repression seems to be indirect, recruiting either the Methyl-CpG-binding domain protein 3/ NuRD (MBD3/NuRD) complex [63], or EZH2 (a core subunit of Polycomb Repressive Complex 2; PRC2) [55] through its oxidative activity. Interestingly, TET proteins were shown to

co-localize with O-linked N-acetylglucosamine (O-GlcNAc) transferase (OGT) genome-wide, notably at CpG islands and at TSSs, in regions depleted of cytosine modifications [60, 64]. Accordingly, TETs interact with OGT *via* their DSBH carboxy-terminal catalytic domain, independently of their catalytic activity, an interaction that plays a role in targeting OGT to specific genomic sites [60, 62, 64-67]. OGT transfers O-GlcNAc to the hydroxyl group of serine or threonine of various proteins such as transcriptional regulators and histones [68]. TET2 targeting of OGT to chromatin enables histone glycosylation, thus participating to TET2-dependent transcriptional regulation [65]. In addition, a TET2/3-OGT complex regulates H3K4 trimethylation since GlcNAcylated Host cell factor C1 (HCF1) stabilizes binding of SET1/COMPASS and SETD1A H3K4 tri-methylases to chromatin [60]. Other studies showed that TET proteins are themselves modified by OGT [62, 66, 67] but it remains unclear whether GlcNAcylation of TETs directly regulates their catalytic activity. All these studies allow for a better comprehension of TET- and 5hmC-mediated transcriptional regulation but further analyses are needed to decipher more precisely the molecular mechanisms involved.

TET Proteins in Pluripotency, Differentiation and Reprogramming

Tet1 and *Tet2* are highly expressed in mESCs [8, 13] and are rapidly downregulated in parallel to a decrease in hydroxymethylation upon differentiation [2, 13, 39]. Interestingly, *Tet2* expression is restored at day 8 of Embryoid Bodies (EBs) while *Tet3*, which was not expressed at the undifferentiated state, is activated upon differentiation [13]. As an increase in 5hmC at day 8 of EBs was also observed in this study, the authors suggested that 5hmC level in undifferentiated ESCs is dependent on TET1 whereas it depends on both TET2 and TET3 at day 8 of EBs [13].

The presence of TET proteins and 5hmC in mESCs and their potential role in regulating gene expression led to studies investigating their possible implication in pluripotency and differentiation. This was in part achieved by depletion of endogenous TET proteins but results obtained from independent laboratories strongly differ. After TET1 depletion, Ito *et al.* observed morphological anomalies, decrease in ESC growth rate and in expression of pluripotency factors [8]. This downregulation was also observed by Ficiz *et al.*, in correlation with a decrease in 5hmC levels at promoters of ESC-specific genes and an increased methylation [39], and by Wu *et al.* with an upregulation of genes implicated in developmental processes [55]. In addition, some changes in gene expression after TET1 depletion correlated with changes occurring during differentiation, when TET1 is downregulated [40]. These results shed light on a role of TET1 in both ESCs maintenance and repression of key regulatory developmental genes. Although similar results were obtained by Dawlaty *et al.* with Tet triple knock-out (TKO) mice [69], the same group demonstrated that TET1 was dispensable for maintaining pluripotency and proper development in *Tet1* mutant mESCs and *Tet1*-null mice [70]. These results are in accordance with other studies which showed no major

changes in ESC morphology or growth rate nor in expression of key pluripotency factors [5, 40, 71]. However, in agreement with a proposed involvement of TET proteins in pluripotency and differentiation, TET1 and TET2 have been found to be regulated by the OCT4-SOX2 complex [71, 72] and to interact with PRDM14 transcriptional regulator to induce DNA demethylation and activation of a set of pluripotency-related and germline specific genes, as well as imprinted loci in ESCs [73]. The involvement of TET proteins in differentiation was also widely studied during neurogenesis as brain tissues are highly enriched in 5hmC. Neuronal differentiation correlates with an increase in 5hmC levels and this differentiation process relies on TET proteins both during mouse brain and *Xenopus* eye and neural development [74-77]. Moreover, through interaction with peroxisome proliferator-activated receptor γ (PPAR γ) or PU.1, TET1 and/or TET2 proteins were shown to regulate adipocyte [78] and monocyte-to-osteoclast [79] differentiation respectively (Fig. 1B). Finally, TET2 was shown to play a role in erythropoiesis in human and zebrafish [80, 81], and all three TET proteins have been implicated in oligodendrocyte differentiation [82].

Epigenetic reprogramming includes both *in vivo* and *in vitro* reprogramming and consists in erasure and remodeling of the epigenetic landscape (for review, see [83]). *In vivo*, the early zygote undergoes a first wave of demethylation which occurs immediately after fertilization. Interestingly, 5hmC was found to be enriched in the paternal pronucleus coinciding with the loss of 5mC and this oxidation was shown to depend on maternal TET3 [84-86]. TET3 is highly expressed in oocytes and zygotes while its expression decreases in 2-cell embryos in agreement with the fact that 5hmC seems to accumulate only during the first cell cycle but remains detectable at later cell-stage embryos [84, 86]. At one with the studies of Iqbal *et al.* and Wossidlo *et al.*, Inoue and Zhang observed a replication-dependent dilution of 5hmC in pre-implantation embryos [22]. They further detected appearance of 5fC and 5caC at later stages, also diluted in a replication-dependent manner [19]. Moreover, 5mC and 5hmC chromosome staining also revealed a parental asymmetry in their distribution in three different species (mouse, bovine and rabbit), suggesting the existence of an evolutionary conserved mechanism controlling this asymmetry [84, 86]. Such asymmetry could be due to 5mC protection by PGC7 (or Dppa3/Stella) [86] in the maternal pronucleus either directly, by inhibition of TET3 catalytic activity through the interaction of PGC7 with TET catalytic domain [87], or indirectly through binding of PGC7 to H3K9me2 containing chromatin [88]. Notably, a second wave of demethylation is observed in the primordial germ cells of the developing embryo where TET functions were shown to be critical [89].

TET proteins were also shown to be implicated in *in vitro* reprogramming. Indeed, a study revealed that NANOG, which is a direct TET1 target and whose expression is also regulated by TET1 [8], can associate with either TET1 or TET2 to enhance the efficiency of somatic cell reprogramming [90]. Doege *et al.* observed a role of both Poly [ADP-ribose] polymerase 1 (PARP1) and TET2 in the epigenetic reprogramming of mouse embryonic fibroblasts (MEFs) under the influence of OSKM (OCT4, SOX2, KLF4 and MYC) reprogramming cocktail [91], while Gao *et al.*

demonstrated that TET1 facilitates pluripotent stem cell induction and succeeded in replacing OCT4 by TET1 in OSKM cocktail [92]. Collectively, these studies established a prominent role for TET proteins in cell fate determination and maintenance.

ALTERATIONS OF THE TET-DEPENDENT PROCESSES IN CANCER CELLS

Cancer arises from the proliferation of cells which accumulate genetic, epigenetic and metabolic alterations, all of which contribute to the malignant phenotype. Not surprisingly, the TET pathway of active DNA demethylation is targeted by all three types of alteration in cancer cells. Indeed, whereas TET2 mutations have been specifically associated with lymphomas, a general silencing of TET expression is observed in malignant tissues, TET transcripts are identified targets of miRNAs in cancer cells, and TET activity is suppressed by cancer-enriched metabolites. As a result, 5hmC levels are fairly low in cancer cells and this may in turn induce aberrant gene regulation (Fig. 2A).

TET Mutations

TET1 was initially discovered as a translocation partner of the *Mixed Lineage Leukemia-1 (MLL1)* gene in acute myeloid leukemia (AML) [93, 94]. The resulting fusion protein is devoid of both the SET domain of MLL1 and a large part of the active site of TET1 and thus is a catalytically dead protein which could nonetheless bind to DNA through a CXXC domain. Many mutations invalidating the enzymatic activity of TET2 have since been described in hematological tumors and especially in AML with a prevalence of 12 to 24% depending on studies [95, 96] and in chronic monomyelocytic leukemia (CMML), with up to 50% of patients carrying *TET2* mutations [96], conditions

which are associated with decreased 5hmC levels [97]. Although not sufficient to drive cancer in mouse [98, 99], *TET2* mutations are early events in myeloid malignancies [95, 100] and associate with poor prognosis in human [101]. In addition, they often co-occur with mutations in ZRSR2 and SFRS2 splicing factors [100, 102], as well as in the chromatin regulator EZH2 [103], suggesting a cooperation between these mutations in driving cancer progression. Interestingly, the *DNMT3A* gene is also often mutated in leukemia, but exclusively of *TET2* mutations [100]. This could decrease 5mC levels globally or in particular genomic regions and hence indirectly contribute to lowering 5hmC in cancer cells devoid of *TET* mutations. Despite the high prevalence of *TET2* mutations in leukemia, few studies reported *TET* somatic mutations in other types of cancer such as melanoma [104], colon cancer [105] and clear-cell renal cell carcinoma [106]. Accordingly, analysis of the Sanger Institute catalogue of somatic mutations in cancer (COSMIC) [107] clearly indicates that *TET* mutations in solid cancers are rare events (Fig. 2B), suggesting that the low 5hmC levels detected in such cancers [15, 108] are not due to loss-of-function mutations of *TET* genes.

Silencing of TET Expression

Although often mutated in leukemia, wild-type *TET2* can be found in patients showing low levels of 5hmC [97]. In solid tumors, 5hmC levels are usually lower than in the corresponding normal tissues and this is strongly correlated with downregulation of *TET* gene expression [15, 109] rather than with *TET* mutations. In melanoma samples for instance, all three TETs have reduced expression, and overexpression of active TET2 in melanoma cells not only restores a normal 5hmC genomic pattern but also reduces melanoma growth and invasion [109]. This is in favor of the existence of mechanisms targeting *TET* gene expression in cancer cells

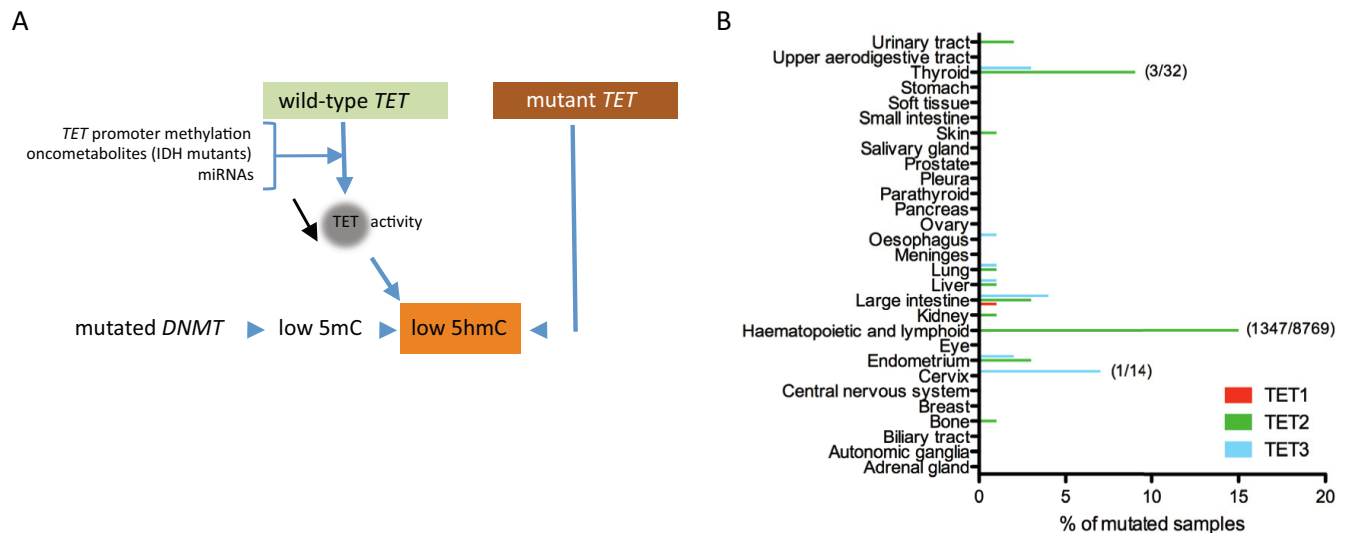


Fig. (2). Mechanisms leading to low 5hmC levels in cancer cells. (A) As frequently described for *TET2* in myeloid malignancies, low levels of 5hmC can directly result from mutations in *TET* genes. Alternatively, in cases in which TET enzymes are not mutated, low hydroxymethylation of cytosines reflect low expression/activity of TETs or could result from decreased levels of 5mC in cells with mutant *DNMT*. (B) Frequency of TET mutation in tumors originating from different tissues. The mutation data were obtained from the Sanger Institute Catalogue Of Somatic Mutations In Cancer web site, <http://www.sanger.ac.uk/cosmic>. The number of mutated samples and the total numbers of samples examined are indicated for highly occurring mutations.

and leading to low levels of proteins. One such mechanism could directly impact on *TET* promoters. Indeed, the *TET2* promoter has been found to be hypermethylated in acute lymphoblastic leukemias (ALLs) as well as in gliomas [110, 111]. This could be triggered by loss of binding of crucial transcription factors since such proteins are believed to be responsible for the definition of the methylome at regulatory regions such as transcriptional enhancers [112]. However, knowledge on transcriptional regulation of *TET* genes is still limited as only the pluripotency factor OCT4 has been shown to bind to and regulate *TET1* and *TET2* promoters [72]. In order to define new strategies for upregulating *TET*s in cancer, it will be important to obtain first a full picture of their transcriptional regulation. Interestingly, DNA methylation of the *TET1* promoter depends on expression of the high mobility group protein HMGA2 in breast cancer cells [113]. Indeed, silencing of HMGA2 not only decreases DNA methylation of the *TET1* promoter but also restores an autoregulatory loop involving *TET1* binding to its own promoter region [113].

A second mechanism regulating *TET* expression acts at the level of mRNA stability. A series of miRNAs can directly target 3'-UTRs of *TET1*, *TET2* and *TET3* mRNAs and thus contribute to *TET* silencing in cancer. For instance, miR-767 directly targets *TET1* and *TET3* 3'-UTRs and is responsible for downregulation of the corresponding proteins in melanoma [114]. In cancer cell lines, miR-29a,b,c target both *TET1,2,3* and *TET3* 3'-UTRs [115, 116]. The metastatic potential of breast cancer cells partly depends on miR-22 targeting of *TET1,2,3*, inhibiting the *TET*-mediated demethylation of the promoter of miR-200, an antimetastatic miRNA [117]. By screening a library of miRNAs, Lu and Collaborators [118] discovered that all three *TET* mRNAs are reduced by a collection of miRNAs (*i.e.* miR-125b, miR-29b,c, miR-101 and miR-7) overexpressed in *TET2*-unmutated AML. Most importantly, overexpression of these miRNAs affects normal hematopoiesis inducing phenotypes which are rescued by *TET2* co-expression [118]. Collectively, these studies indicate that similarities in their 3'-UTRs allow all three *TET* mRNAs to be targeted by a common set of miRNAs overexpressed in cancer.

Finally, *TET* levels can be regulated at the post-translational level. In ES cells, *TET2* is recruited to DNA by a direct interaction with the CXXC-containing proteins IDAX (CXXC4) and CXXC5 [119]. The physical association of *TET2* with DNA-bound IDAX then leads to a caspase-dependent degradation of *TET2* [119]. However, it is fairly unknown whether IDAX is overexpressed in tumors, except in colon cancer [120]. Interrogation of the Gene expression across normal and tumor tissue database (GENT - <http://medicalgenome.kribb.re.kr/GENT>) [121] indicated that IDAX is overexpressed also in kidney whereas CXXC5 is highly overexpressed in ovary and to a lesser extent in blood, colon, stomach and uterus cancers. A screen for effects of various protease inhibitors on *TET* protein stability revealed that calpeptin, a specific inhibitor of the calpain family of Ca^{2+} -dependent proteases, similarly extended the half-life of *TET1,2,3* in mouse ES cells [122]. This study further demonstrated that calpain1 (CAPN1) and calpain2 (CAPN2) degrade *TET1,2,3*. Analogous to CXXC4 and CXXC5, interrogation of the GENT database revealed overexpression of CAPN1 and CAPN2 in solid tumors from

various tissues, including liver, vagina, intestine, pancreas, uterus, bladder and testis. Hence, the low levels of *TET*s in cancer cells might finally result from a complex addition of defects in transcriptional, post-transcriptional and post-translational mechanisms.

TET Regulation by Oncometabolites

As other Fe(II)-dependent dioxygenases, *TET*s require α -KG (or 2-oxoglutarate) to generate a reactive Fe(IV)-oxo intermediate needed for 5mC oxidation. This cofactor is generated from isocitrate by isocitrate-dehydrogenases (IDHs) in the tricarboxylic acid (TCA) cycle but can also be produced from glutamate by glutamate dehydrogenase (GDH). In AML and glioma, mostly exclusive mutations in *IDH1* (IDH1/R132) and *IDH2* (IDH2/R140 and IDH2/R170) occur at high frequency and preclude isocitrate transformation to α -KG [123-126]. These mutations thus lower α -KG levels in cancer cells. In addition, they provide new function to the mutated enzyme since they can reduce α -KG to 2-hydroxyglutarate (2HG), a metabolite that is highly enriched in tumor cells compared to normal cells [127]. With the demonstration that 2HG is able to transform cells [128, 129], cancer specific gain-of-function mutations of *IDH1* and *IDH2* are thus responsible for the accumulation of an oncometabolite. Although 2HG might exert different effects in the cell, one of its prominent action is to inhibit α -KG-dependent dioxygenases. Indeed, 2HG can inhibit histone demethylases of the jumonji subgroup, *TET* enzymes and HIF prolyl hydroxylases, by occupying their α -KG binding sites [130-132]. In a similar way, fumarate and succinate act as competitive inhibitors of α -KG-dependent dioxygenases [133] and accumulate in cancers with fumarate hydratase (FH) and succinate dehydrogenase (SDH) loss-of-function mutations [134-136].

Targeting TET-Dependent Active DNA Demethylation in Cancer

The extensive accumulation of evidence for the implication of *TET* dysfunction in cancer suggests that protocols that could restore specific patterns of high 5mC levels in cancer cells could be beneficial to patients. Consistent with this idea, *TET* overexpression inhibits melanoma tumor growth [109] and breast cancer cell invasion [137] in mouse xenograft models. Depending on the molecular alterations targeting *TET*-dependent processes in individual patients, different strategies could be suggested. For instance, patients with low *TET* levels might benefit from being treated with compounds increasing *TET* expression, stability or activity; whereas compounds specifically targeting mutant IDH might help patients suffering from oncometabolite accumulation. In the latter case, targeting homodimerization interface rather than the active site of the enzymes is likely to provide specific inhibitors of the mutated forms. Such compounds were successfully isolated in screens for IDH2/R140Q and IDH1/R132H inhibitors (AGI-6780 and AGI-5198 compounds respectively) [138, 139]. Both compounds efficiently reduce 2HG content and induce differentiation of cancer cells [138, 139], however their effects on 5mC levels remain to be explored.

TET expression might also be targeted indirectly by compounds known to affect calpain as well as HMGA2. Calpain inhibition has been shown to increase TET levels in ES cells [122] and could prove useful in cancer cells overexpressing calpain. In addition, 5hmC levels raise in CAPN1-/- ES cells, indicating that elevation in TET protein levels result in higher 5hmC genomic content [122]. Calpain is involved in multiple aspects of cell biology linked to cancer progression like: (i) enhanced mobility through modulation of metalloproteinase activity, focal adhesion turnover and invadopodia formation; and (ii) cell cycle regulation (for review, see [140]). However, the proapoptotic roles of calpain [141-143] suggest that calpain inhibition would be more suited for the treatment of tumors in which the apoptotic pathway is deficient. Interestingly, calpain and HDAC inhibition synergize in reducing growth of breast cancer cell lines [144]. HDAC inhibition might provide another way to potentially restore high TET levels in cancer cells. HDAC inhibitors (HDACi) are known to arrest growth of many cancer cell lines and exert their effects through different downstream targets, including miRNAs reducing HMGA2 levels [145-147], a potent driver of leukemia [148].

Finally, vitamin C promotes recycling of Fe(II) and as such sustains TET activity [149-152]. Although the potential of antioxidants in cancer prevention is debated, dietary vitamin C intake is significantly associated with reduced mortality in breast cancer cases [153], and reduces the risk of colorectal adenoma [154]. Hence, it is anticipated that treating cancer cells with vitamin C, HDACi and/or calpain inhibitors could significantly enhance TET activity, an hypothesis that needs to be tested in the future.

CONCLUDING REMARKS

Restoring high 5hmC levels in cancer cells through the use of compounds targeting TET activity may be of interest for the definition of new anti-cancer therapies. However, the above-mentioned strategies are based on the ability of TET to influence cell differentiation but may not prove to be efficient in killing cancer cells. Although cancer cells have low 5hmC levels, the remaining 5hmC could be implicated in the maintenance of cancer cell characteristics. Indeed, 5hmC is enriched at enhancers regulating cell-specific gene expression in embryonal carcinoma cells [43], and cell identity is believed to rely heavily on the sets of used enhancers. Considering that cytosine hydroxymethylation may participate in enhancer activation, TET inhibition with the cell permeable compound octyl- α -KG [132], specifically delivered to tumor cells, might prove efficient in inducing cell-death, alone or in combination with pro-apoptotic drugs. Defining the exact role of DNA hydroxymethylation in enhancer activation in normal and cancer cells will undoubtedly help design new enhancer therapies for the benefit of patients.

ABBREVIATIONS

2HG	= 2-hydroxyglutarate
5caC	= 5-carboxylcytosine
5fC	= 5-formylcytosine
5gmC	= glucosylated 5hmC

5hmC	= 5-hydroxymethylcytosine
5hmU	= 5-hydroxymethyluracil
5mC	= 5-methylcytosine
α -kG	= α -ketoglutarate
β -GT	= β -glucosyltransferase
AID	= Activation-induced deaminase
ALL	= Acute lymphoblastic leukemia
AML	= Acute myeloid leukemia
APOBEC	= Apolipoprotein B mRNA-editing enzyme complex
BER	= Base excision repair
BS-seq	= Bisulfite sequencing
CAB-seq	= Chemical modification-assisted bisulfite sequencing to map 5caC
CAPN1/2	= Calpain1/2
CD	= Catalytic domain
ChIP	= Chromatin immunoprecipitation
CMML	= Chronic monomyelocytic leukemia
CMS	= Cytosine 5-methylenesulphonate
CoREST	= RE1-Silencing transcription factor corepressor 1
COSMIC	= Sanger Institute catalogue of somatic mutations in cancer
DIP	= DNA immunoprecipitation
DNMT	= DNA Methyltransferase
DSBH	= Double-stranded β -helix
EBs	= Embryoid bodies
ECCs	= Embryonal carcinoma cells
(m)ESCs	= (Mouse) embryonic stem cells
fC-Seal	= Selective chemical labeling method to map 5fC
fCAB-seq	= Chemically modification-assisted bisulfite sequencing method to map 5fC
FH	= Fumarate hydratase
GDH	= Glutamate dehydrogenase
GENT	= Gene expression across normal and tumor tissue database
GLIB	= Glucosylation, periodate oxidation, biotinylation
HCF1	= Host cell factor C1
HDACi	= HDAC inhibitors
hMe-Seal	= Selective chemical labeling method to map 5hmC
IDAX	= Inhibition of the Dvl and axin complex
IDCases	= Isoorotate decarboxylases
IDH	= Isocitrate-dehydrogenase

MAB-seq	= Methylase-assisted bisulfite sequencing
MBD4	= Methyl-CpG binding domain 4
MeDIP	= Methylated DNA immunoprecipitation
MEFs	= Mouse embryonic fibroblasts
MLL1	= Mixed Lineage Leukemia-1
NgTet1	= Naegleria Tet-like
NurD	= Nucleosome remodeling deacetylase
O-GlcNAc	= O-linked N-acetylglucosamine
OGT	= O-linked N-acetylglucosamine transferase
OSKM	= OCT4, SOX2, KLF4 and MYC reprogramming cocktail
oxBS-seq	= Oxidative bisulfite sequencing
PARP1	= Poly [ADP-ribose] polymerase 1
PPAR γ	= Peroxisome proliferator-activated receptor γ
PRC2	= Polycomb repressive complex 2
redBS-seq	= Reduced bisulfite sequencing
RRHP	= Reduced representation 5-hydroxymethylcytosine profiling
SAM	= S-adenosyl methionine
SDH	= Succinate dehydrogenase
SMRT	= Single-molecule, real-time
TAB-seq	= TET-assisted bisulfite sequencing
TCA	= Tricarboxylic acid
TDG	= Thymine DNA glycosylase
TET	= Ten-eleven translocation
TKO	= Triple knock-out
TSS	= Transcription start site
UHRF1	= Ubiquitin-like, containing PHD and RING finger domains, 1

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

We thank Thierry Madigou and Raphaël Métivier for critical reading of the manuscript. E.M. was supported by a PhD fellowship from the Ministère de l'enseignement supérieur et de la recherche. G.S. was supported by the CNRS and the University of Rennes 1.

REFERENCES

- [1] Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 2009; 324: 929-30
- [2] Tahiliani M, Koh KP, Shen Y, *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009; 324: 930-5
- [3] Iyer LM, Tahiliani M, Rao A, Aravind L. Prediction of novel enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell cycle* 2009; 8: 1698-710
- [4] Zhang H, Zhang X, Clark E, *et al.* TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res* 2010; 20: 1390-3
- [5] Xu Y, Wu F, Tan L, *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* 2011; 42: 451-64
- [6] Hu L, Li Z, Cheng J, *et al.* Crystal structure of TET2-DNA complex: insight into TET-mediated 5mC oxidation. *Cell* 2013; 19: 1545-55
- [7] Hashimoto H, Pais JE, Zhang X, *et al.* Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature* 2014; 506: 391-5
- [8] Ito S, D'Alessio AC, Taranova OV, *et al.* Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 2010; 466: 1129-33
- [9] Ito S, Shen L, Dai Q, *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 2011; 333: 1300-3
- [10] He YF, Li BZ, Li Z, *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 2011; 333: 1303-7
- [11] Globisch D, Münzel M, Müller M, *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 2010; 5: e15367
- [12] Münzel M, Globisch D, Brückl T, *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl* 2010; 49: 5375-7
- [13] Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 2010; 38: e181
- [14] Song CX, Szulwach KE, Fu Y, *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol* 2011; 29: 68-72
- [15] Yang H, Liu Y, Bai F, *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* 2013; 32: 663-9
- [16] Song CX, Szulwach KE, Dai Q, *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 2013; 153: 678-91
- [17] Shen L, Wu H, Diep D, *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 2013; 153: 692-706
- [18] Booth MJ, Marsico G, Bachman M, Beraldi D, Balasubramanian S. Quantitative sequencing of 5-formylcytosine in DNA at single-base resolution. *Nat Chem* 2014; 6: 435-40
- [19] Inoue A, Shen L, Dai Q, He C, Zhang Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res* 2011; 21: 1670-6
- [20] Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* 2006; 31: 89-97
- [21] Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999; 99: 247-57
- [22] Inoue A, Zhang Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 2011; 334: 194
- [23] Hackett JA, Sengupta R, Zyllicz JJ, *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* 2013; 339: 448-52
- [24] Valinluck V and Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res* 2007; 67: 946-50
- [25] Hashimoto H, Liu Y, Upadhyay AK, *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res* 2012; 40: 4841-9
- [26] Otani J, Kimura H, Sharif J, *et al.* Cell cycle-dependent turnover of 5-hydroxymethyl cytosine in mouse embryonic stem cells. *PLoS One* 2013; 8: e82961
- [27] Frauer C, Hoffmann T, Bultmann S, *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS One* 2011; 6: e21306
- [28] Maiti A, Drohat AC. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for

- active demethylation of CpG sites. *J Biol Chem* 2011; 286: 35334-8
- [29] Müller U, Bauer C, Siegl M, Rottach A, Leonhardt H. TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res* 2014; 4: 8592-604
- [30] Schiesser S, Hackner B, Pfaffeneder T, *et al.* Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew Chem Int Ed Engl* 2012; 51: 6516-20
- [31] Xu S, Li W, Zhu J, *et al.* Crystal structures of isoorotate decarboxylases reveal a novel catalytic mechanism of 5-carboxyluracil decarboxylation and shed light on the search for DNA decarboxylase. *Cell Res* 2013; 23: 1296-309
- [32] Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 2011; 145: 423-34
- [33] Cortellino S, Xu J, Sannai M, *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 2011; 146: 67-79
- [34] Hashimoto H, Zhang X, Cheng X. Excision of thymine and 5-hydroxymethyluracil by the MBD4 DNA glycosylase domain: structural basis and implications for active DNA demethylation. *Nucleic Acids Res* 2012; 40: 8276-84
- [35] Nabel CS, Jia H, Ye Y, *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat Chem Biol* 2012; 8: 751-8
- [36] Pfaffeneder T, Spada F, Wagner M, *et al.* Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat Chem Biol* 2014; 10: 574-81
- [37] Weber M, Davies JJ, Wittig D, *et al.* Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet* 2005; 37: 853-62
- [38] Wu H, D'Alessio AC, Ito S, *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev* 2011; 25: 679-84
- [39] Ficiz G, Branco MR, Seisenberger S, *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 2011; 473: 398-402
- [40] Williams K, Christensen J, Pedersen MT, *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 2011; 473: 343-8
- [41] Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol* 2011; 12: R54
- [42] Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res* 2011; 39: 5015-24
- [43] Sérandour AA, Avner S, Oger F, *et al.* Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res* 2012; 40: 8255-65
- [44] Pastor WA, Pape UJ, Huang Y, *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 2011; 473: 394-7
- [45] Pastor WA, Huang Y, Henderson HR, Agarwal S, Rao A. The GLIB technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat Protoc* 2012; 7: 1909-17
- [46] Flusberg BA, Webster DR, Lee JH, *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat Methods* 2010; 7: 461-5
- [47] Sun Z, Terragni J, Borgaro JG, *et al.* High-resolution enzymatic mapping of genomic 5-hydroxymethylcytosine in mouse embryonic stem cells. *Cell Rep* 2013; 3: 567-76
- [48] Petterson A, Chung T, Tan D, Sun X, Jia XY. RRRP: a tag-based approach for 5-hydroxymethylcytosine mapping at single-site resolution. *Genome Biol* 2014; 15: 456
- [49] Huang Y, Pastor WA, Shen Y, *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One* 2010; 5: e8888
- [50] Booth MJ, Branco MR, Ficiz G, *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 2012; 336: 934-7
- [51] Booth MJ, Ost TW, Beraldi D, *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat Protoc* 2013; 8: 1841-51
- [52] Yu M, Hon GC, Szulwach KE, *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* 2012; 149: 1368-80
- [53] Lu X, Song CX, Szulwach K, *et al.* Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J Am Chem Soc* 2013; 135: 9315-7
- [54] Wu H, Wu X, Shen L, Zhang Y. Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. *Nat Biotechnol* 2014. doi: 10.1038/nbt.3073
- [55] Wu H, D'Alessio AC, Ito S, *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 2011; 473: 389-93
- [56] Szulwach KE, Li X, Li Y, *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet* 2011; 7: e1002154
- [57] Putiri EL, Tiedemann RL, Thompson JJ, *et al.* Distinct and overlapping control of 5-methylcytosine and 5-hydroxymethylcytosine by the TET proteins in human cancer cells. *Genome Biol* 2014; 15: R81
- [58] Mellén M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 2012; 151: 1417-30
- [59] Huang Y, Chavez L, Chang X, *et al.* Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* 2014; 111: 1361-6
- [60] Deplus R, Delatte B, Schwinn MK, *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* 2013; 32: 645-55
- [61] Cartron PF, Nadaradjane A, Lepape F, Lallier L, Gardie B, Vallette FM. Identification of TET1 Partners That Control Its DNA-Demethylating Function. *Genes Cancer* 2013; 4: 235-41
- [62] Shi FT, Kim H, Lu W, *et al.* Ten-eleven translocation 1 (Tet1) is regulated by O-linked N-acetylglucosamine transferase (Ogt) for target gene repression in mouse embryonic stem cells. *J Biol Chem* 2013; 288: 20776-84
- [63] Yildirim O, Li R, Hung JH, *et al.* Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 2011; 147: 1498-510
- [64] Vella P, Scelfo A, Jammula S, *et al.* Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol Cell* 2013; 49: 645-56
- [65] Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 2013; 493: 561-4
- [66] Ito R, Katsura S, Shimada H, *et al.* TET3-OGT interaction increases the stability and the presence of OGT in chromatin. *Genes Cells* 2014; 19: 52-65
- [67] Zhang Q, Liu X, Gao W, *et al.* Differential regulation of the ten-eleven translocation (TET) family of dioxygenases by O-linked β -N-acetylglucosamine transferase (OGT). *J Biol Chem* 2014; 289: 5986-96
- [68] Hart GW, Housley MP, Slawson C. Cycling of O-linked beta-N-acetylglucosamine on nucleocytoplasmic proteins. *Nature* 2007; 446: 1017-22
- [69] Dawlaty MM, Breiling A, Le T, *et al.* Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev Cell* 2014; 29: 102-11
- [70] Dawlaty MM, Ganz K, Powell BE, *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 2011; 9: 166-75
- [71] Koh KP, Yabuuchi A, Rao S, *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 2011; 8: 200-13
- [72] Wu Y, Guo Z, Liu Y, *et al.* Oct4 and the small molecule inhibitor, SC1, regulates Tet2 expression in mouse embryonic stem cells. *Mol Biol Rep* 2013; 40: 2897-906
- [73] Okashita N, Kumaki Y, Ebi K, *et al.* PRDM14 promotes active DNA demethylation through the ten-eleven translocation (TET)-mediated base excision repair pathway in embryonic stem cells. *Development* 2014; 141: 269-80
- [74] Xu Y, Xu C, Kato A, *et al.* Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* 2012; 151: 1200-13

- [75] Hahn MA, Qiu R, Wu X, *et al.* Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep* 2013; 3: 291-300
- [76] Zhang RR, Cui QY, Murai K, *et al.* Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* 2013; 13: 237-45
- [77] Colquitt BM, Allen WE, Barnea G, Lomvardas S. Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc Natl Acad Sci U S A* 2013; 110: 14682-7
- [78] Fujiki K, Shinoda A, Kano F, Sato R, Shirahige K, Murata M. PPAR γ -induced PARylation promotes local DNA demethylation by production of 5-hydroxymethylcytosine. *Nat Commun* 2013; 4: 2262
- [79] de la Rica L, Rodríguez-Ubrea J, García M, *et al.* PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol* 2013; 14: R99
- [80] Madzo J, Liu H, Rodriguez A, *et al.* Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis. *Cell Rep* 2014; 6: 231-44
- [81] Ge L, Zhang RP, Wan F, *et al.* TET2 plays an essential role in erythropoiesis by regulating lineage-specific genes *via* DNA oxidative demethylation in a zebrafish model.
- [82] Zhao X, Dai J, Ma Y, Mi Y, Cui D, Ju G, Macklin WB, Jin W. Dynamics of ten-eleven translocation hydroxylase family proteins and 5-hydroxymethylcytosine in oligodendrocyte differentiation. *Glia* 2014; 62: 914-26
- [83] Hill PW, Amouroux R, Hajkova P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: An emerging complex story. *Genomics* 2014. pii: S0888-7543(14)00154-2
- [84] Iqbal K, Jin SG, Pfeifer GP, Szabó PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc Natl Acad Sci U S A* 2011; 108: 3642-7
- [85] Gu TP, Guo F, Yang H, Wu HP, Xu GF, Liu W, Xie ZG, Shi L, He X, Jin SG, Iqbal K, Shi YG, Deng Z, Szabó PE, Pfeifer GP, Li J, Xu GL. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 2011; 477: 606-10
- [86] Wossidlo M, Nakamura T, Lepikhov K, *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun* 2011; 2: 241
- [87] Bian C, Yu X. PGC7 suppresses TET3 for protecting DNA methylation. *Nucleic Acids Res* 2014; 42: 2893-905
- [88] Nakamura T, Liu YJ, Nakashima H, *et al.* PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 2012; 486: 415-9
- [89] Yamaguchi S, Shen L, Liu Y, Sandler D, Zhang Y. Role of Tet1 in erasure of genomic imprinting. *Nature* 2013; 504: 460-4
- [90] Costa Y, Ding J, Theunissen TW, *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 2013; 495: 370-4
- [91] Doege CA, Inoue K, Yamashita T, Rhee DB, Travis S, Fujita R, Guarnieri P, Bhagat G, Vanti WB, Shih A, Levine RL, Nik S, Chen EI, Abeliovich A. Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* 2012; 488: 652-5
- [92] Gao Y, Chen J, Li K, *et al.* Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* 2013; 12: 453-69
- [93] Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. LCX, leukemia-associated protein with a CXXC domain is fused to MLL in myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res* 2002; 62: 4075-80
- [94] Lorschach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 2003; 17: 637-41
- [95] Delhommeau F, Dupont S, Della Valle V, *et al.* Mutations in TET2 in myeloid cancers. *N Engl J Med* 2009; 360: 2289-301
- [96] Abdel-Wahab O, Mullally A, Hedvat C, *et al.* Genetic characterization of TET1, TET2, and TET3 mutations in myeloid malignancies. *Blood* 2009; 114: 144-7
- [97] Ko M, Huang Y, Jankowska AM, *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 2010; 468: 839-43
- [98] Moran-Crusio K, Reavie L, Shih A, *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 2011; 20: 11-24
- [99] Quivoron C, Couronné L, Della Valle V, *et al.* TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* 2011; 20: 25-38
- [100] Papaemmanuil E, Gerstung M, Malcovati L, *et al.* Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 2013; 122: 3616-27
- [101] Chou WC, Chou SC, Liu CY, *et al.* TET2 mutation is an unfavorable prognostic factor in acute myeloid leukemia patients with intermediate-risk cytogenetics. *Blood* 2011; 118: 3803-10
- [102] Haferlach T, Nagata Y, Grossmann V, *et al.* Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 2014; 28: 241-7
- [103] Muto T, Sashida G, Oshima M, *et al.* Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. *J Exp Med* 2013; 210: 2627-39
- [104] Song F, Amos CI, Lee JE, *et al.* Identification of a melanoma susceptibility locus and somatic mutation in TET2. *Carcinogenesis* 2014; 35: 2097-101
- [105] Seshagiri S, Stawiski EW, Durinck S, *et al.* Recurrent R-spondin fusions in colon cancer. *Nature* 2012; 488: 660-4
- [106] Sato Y, Yoshizato T, Shiraishi Y, *et al.* Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* 2013; 45: 860-7
- [107] Bamford S, Dawson E, Forbes S, *et al.* The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* 2004; 91: 355-8
- [108] Jin SG, Jiang Y, Qiu R, *et al.* 5-hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 2011; 71: 7360-5
- [109] Lian CG, Xu Y, Ceol C, *et al.* Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* 2012; 150: 1135-46
- [110] Kim YH, Pierscianek D, Mittelbronn M, *et al.* TET2 promoter methylation in low-grade diffuse gliomas lacking IDH1/2 mutations. *J Clin Pathol* 2011; 64: 850-2
- [111] Musialik E, Bujko M, Wypych A, Matysiak M, Siedlecki JA. TET2 promoter DNA methylation and expression analysis in pediatric B-cell acute lymphoblastic leukemia. *Hematol Rep* 2014; 6: 5333
- [112] Stadler MB, Murr R, Burger L, *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 2011; 480: 490-5
- [113] Sun M, Song CX, Huang H, *et al.* HMGA2/TET1/HOXA9 signaling pathway regulates breast cancer growth and metastasis. *Proc Natl Acad Sci USA* 2013; 110: 9920-5
- [114] Lorient A, Van Tongelen A, Blanco J, *et al.* A novel cancer-germline transcript carrying pro-metastatic miR-105 and TET-targeting miR-767 induced by DNA hypomethylation in tumors. *Epigenetics* 2014; 9: 1163-71
- [115] Morita S, Horii T, Kimura M, Ochiya T, Tajima S, Hatada I. miR-29 represses the activities of DNA methyltransferases and DNA demethylases. *Int J Mol Sci* 2013; 14: 14647-58
- [116] Zhang P, Huang B, Xu X, Sessa WC. Ten-eleven translocation (Tet) and thymidine DNA glycosylase (TDG), components of the demethylation pathway, are direct targets of miRNA-29a. *Biochem Biophys Res Commun* 2013; 437: 368-73
- [117] Song SJ, Poliseno L, Song MS, *et al.* MicroRNA-antagonism regulates breast cancer stemness and metastasis *via* TET-family-dependent chromatin remodeling. *Cell* 2013; 154: 311-24
- [118] Cheng J, Guo S, Chen S, *et al.* An extensive network of TET2-targeting MicroRNAs regulates malignant hematopoiesis. *Cell Rep* 2013; 5: 471-81
- [119] Ko M, An J, Bandukwala HS, *et al.* Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 2013; 497: 122-6
- [120] Nguyen AV, Albers CG, Holcombe RF. Differentiation of tubular and villous adenomas based on Wnt pathway-related gene expression profiles. *Int J Mol Med* 2010; 26: 121-5
- [121] Shin G, Kang T-W, Yang S, *et al.* GENT: gene expression database of normal and tumor tissue. *Cancer Inform* 2011; 10: 149-57

- [122] Wang Y, Zhang Y. Regulation of TET protein stability by calpains. *Cell Rep* 2013; 6: 278-284
- [123] Bleeker FE, Lamba S, Leenstra S, *et al.* IDH1 mutations at residue p.R132 (IDH1(R132)) occur frequently in high-grade glioma but not in other solid tumors. *Hum Mutat* 2009; 30: 7-11
- [124] Yan H, Parsons DW, Jin G, *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* 2009; 360: 765-73
- [125] Mardis ER, Ding L, Dooling DJ, *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 2009; 361: 1058-66
- [126] Paschka P, Schlenk RF, Gaidzik VI, *et al.* IDH1 and IDH2 mutations are frequent genetic alterations in acute myeloid leukemia and confer adverse prognosis in cytogenetically normal acute myeloid leukemia with NPM1 mutation without FLT3 internal tandem duplication. *J Clin Oncol* 2010; 28: 3636-43
- [127] Dang L, White DW, Gross S, *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 2009; 462: 739-44
- [128] Koivunen P, Lee S, Duncan CG, *et al.* Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* 2012; 483: 484-8
- [129] Losman JA, Looper RE, Koivunen P, *et al.* (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* 2013; 339: 1621-5
- [130] Chowdhury R, Yeoh KK, Tian YM, *et al.* The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep* 2011; 12: 463-9
- [131] Lu C, Ward PS, Kapoor GS, *et al.* IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* 2012; 483: 474-8
- [132] Xu W, Yang H, Liu Y, *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* 2011; 19: 17-30
- [133] Xiao M, Yang H, Xu W, *et al.* Inhibition of α -KG-dependent histone and DNA demethylases by fumarate and succinate accumulated in mutation of FH and SDH tumor suppressors. *Genes Dev* 2012; 26: 1326-38.
- [134] Pollard PJ, Brière JJ, Alam NA, *et al.* Accumulation of Krebs cycle intermediate and over-expression of HIF1 α in tumors which result from germline FH and SDH mutations. *Hum Mol Genet* 2005; 14: 2231-9
- [135] Bayley JP, Launonen V, Tomlinson IP. The FH mutation database: an online database of fumarate hydratase mutations involved in the MCUL (HLRCC) tumor syndrome and congenital fumarase deficiency. *BMC Med Genet* 2008; 9: 20
- [136] Bardella C, Pollard PJ, Tomlinson I. SDH mutations in cancer. *Biochem Biophys Acta* 2011; 1807: 1432-43
- [137] Hsu CH, Peng KL, Kang ML, *et al.* TET1 suppresses cancer invasion by activation the tissue inhibitors of metalloproteinases. *Cell Rep* 2012; 2: 568-79
- [138] Wang F, Travins J, DeLaBarre B, *et al.* Targeted inhibition of mutant IDH2 in leukemia cells induces cellular differentiation. *Science* 2013; 340: 622-6
- [139] Rohle D, Popovici-Muller J, Palaskas N, *et al.* An inhibitor of mutant IDH1 delays growth and promotes differentiation of glioma cells. *Science* 2013; 340: 626-30
- [140] Storr SJ, Carragher NO, Frame MC, *et al.* The calpain system and cancer. *Nat Rev Cancer* 2011; 11: 364-74
- [141] Toyota H, Yanase N, Yoshimoto T, Moriyama M, Sudo T, Mizuguchi J. Calpain-induced Bax-cleavage product is a more potent inducer of apoptotic cell death than wild-type Bax. *Cancer Lett* 2003; 189: 221-30
- [142] Fettucciari K, Fettriconi I, Mannucci R, *et al.* Group B Streptococcus induces macrophage apoptosis by calpain activation. *J Immunol* 2006; 176: 7542-56
- [143] Gafni J, Cong X, Chen SF, Gibson BW, Ellerby LM. Calpain-1 cleaves and activates caspase-7. *J Biol Chem* 2009; 284: 25441-9
- [144] Mataga MA, Rosenthal S, Heerboth S, *et al.* Anti-breast cancer effects of histone deacetylase inhibitors and calpain inhibitor. *Anticancer Res* 2012; 32: 2523-9
- [145] Lee S, Jung JW, Park SB, *et al.* Histone deacetylase regulates high mobility group A2-targeting microRNAs in human cord blood-derived multipotent stem cell aging. *Cell Mol Life Sci* 2011; 68: 325-36
- [146] Di Fazio P, Montalbano R, Neureiter D, *et al.* Downregulation of HMGA2 by the pan-deacetylase inhibitor panobinostat is dependent on hsa-let-7b expression in liver cancer cell lines. *Exp Cell Res* 2012; 318: 1832-43
- [147] Zhou H, Guo W, Zhao Y, *et al.* MicroRNA-26a acts as a tumor suppressor inhibiting gallbladder cancer cell proliferation directly targeting HMGA2. *Int J Oncol* 2014; 44: 2050-8
- [148] Efanov A, Zanasi N, Coppola V, *et al.* Human HMGA2 protein overexpressed in mice induces precursor T-cell lymphoblastic leukemia. *Blood Cancer* 2014; 4: e227
- [149] Minor EA, Court BL, Young JJ, Wang G. Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J Biol Chem* 2013; 288: 13669-74
- [150] Yin R, Mao SQ, Zhao B, *et al.* Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J Am Chem Soc* 2013; 135: 10396-403
- [151] Blaschke K, Ebata KT, Karimi MM, *et al.* Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 2013; 500: 222-6
- [152] Chen J, Guo L, Zhang L, *et al.* Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat Genet* 2013; 45: 1504-9
- [153] Harris HR, Orsini N, Wolk A. Vitamin C and survival among women with breast cancer: a meta-analysis. *Eur J Cancer* 2014; 50: 1223-31
- [154] Xu X, Yu E, Liu L, *et al.* Dietary intake of vitamins A, C, and E and the risk of colorectal adenoma: a meta-analysis of observational studies. *Eur J Cancer Prev* 2013; 22: 529-39.

ANNEXE

VU :

Le Directeur de Thèse
(Nom et Prénom)

VU :

Le Responsable de l'École Doctorale

VU pour autorisation de soutenance

Rennes, le

Le Président de l'Université de Rennes 1

Guy CATHELINEAU

VU après soutenance pour autorisation de publication :

Le Président de Jury,
(Nom et Prénom)