

THÈSE DE DOCTORAT DE

L'UNIVERSITE DE NANTES

ECOLE DOCTORALE N° 605

Biologie Santé

Spécialité : Bioinformatique

Par

Victor GABORIT

Mode D'action Moléculaire De La Dexaméthasone Dans Le Myélome Multiple

Thèse présentée et soutenue à Nantes, le 30/06/2021

Unité de recherche : CRCINA, INSERM UMR1232, CNRS ERL6001, Equipe 11

Rapporteurs avant soutenance :

Charles Dumontet

Praticien Hospitalier, CRCL UMR 1052 CNRS 5286, Université de Lyon

Denis Puthier

Maître de Conférence Universitaire, TAGC UMR_S 1090, Aix-Marseille Université

Composition du Jury :

Président : **Guillaume Fertin**

Professeur des Universités, LS2N CNRS UMR 6004
Université de Nantes

Examineurs : **Carl Hermann**

Maître de Conférence Universitaire, DKFZ, Université
d'Heidelberg

Vera Pancaldi

Chargée de Recherche, CRCT UMR 1037, Université
Toulouse III Paul Sabatier

Dir. de thèse : **Stéphane Minvielle**

Directeur de Recherche, CRCINA UMR1232, CNRS
ERL6001, CHU de Nantes

Co-dir. de thèse : **Florence Magrangeas**

Docteur en Sciences, CRCINA UMR1232, CNRS ERL6001,
CHU de Nantes

Co-dir. de thèse : **Jérémie Bourdon**

Professeur des Universités, LS2N CNRS UMR 6004,
Université de Nantes

Liste des Figures

Figure 1 : La différenciation du lignage B se distingue en deux grands phénomènes: la lymphopoïèse et l'immunopoïèse.....	17
Figure 2 : Le développement et la progression du Myélome Multiple: localisation et anomalies génétiques	18
Figure 3 : Principales translocations et gènes majoritairement impliqués dans le Myélome Multiple	20
Figure 4 : Proposition de modèles d'évolution clonale dans le Myélome Multiple	22
Figure 5 : La voie classique et la voie alternative NF- κ B	25
Figure 6 : La voie de signalisation RAS/MAPK	26
Figure 7 : Expression biphasique d'IRF4 selon le stade de différenciation de la cellule B ...	27
Figure 8 : Rôle d'IRF4 dans la survie et la prolifération de la cellule de myélome.....	28
Figure 9 : Domaines et composition des sous-unités histones H2A et B, H3 et H4	33
Figure 10 : Modulation de l'expression génique selon la présence ou l'absence de diverses marques d'histones.....	34
Figure 11 : Compaction et accessibilité de la chromatine.....	35
Figure 12 : Organisation structurale de la chromatine à différents niveaux	37
Figure 13 : Les GCs régulent un très grand nombre de fonctions à l'état physiologique chez l'homme	39
Figure 14 : Composition du récepteur nucléaire aux glucocorticoïdes	41
Figure 15 : Mécanismes d'action moléculaire du récepteur au glucocorticoïde sur la transcription des gènes	42
Figure 16 : Le GR régule la transcription de gènes cibles via différents mécanismes.....	43
Figure 17 : La tetramérisation du GR permet de nouvelles fonctions moléculaires	44
Figure 18 : Le récepteur aux glucocorticoïdes agit sur la médiation des interactions enhancers-promoteurs à travers les boucles de régulation.....	45
Figure 19 : Protocole de Chromatine ImmunoPrecipitation (ChIP) couplé au séquençage ..	48
Figure 20 : Accessibilité de la chromatine liée à la transposase Tn5 (ATAC-seq).....	49
Figure 21 : Technique de séquençage des ARN messagers RNA-seq.....	50
Figure 22 : Technique de séquençage HiChIP	51
Figure 23 : Méthode de séquençage NGS Illumina	52
Figure 24 : Immunoprécipitation des protéines suivie d'une spectrométrie de masse (RIME)	53
Figure 25 : Préparation de librairie scRNA-seq	55
Figure 26 : Composition des billes GEM pour le séquençage Multiome	55
Figure 27 : Dose de Dex induisant la mort cellulaire dans la lignée MM.1S	57
Figure 28 : Indexation de génome par transformée de Burrows-Wheeler	59
Figure 29 : Utilisation de la transformée de Burrows-Wheeler pour l'alignement des reads issus du séquençage.....	60
Figure 30 : Alignement de reads HiC et HiChIP par HiCpro	61
Figure 31 : La complexité de la librairie de séquençage est importante pour la qualité de l'analyse.....	62

Figure 32 : Détection des pics ChIP-seq et ATAC-seq et shift réalisés pour combiner les données reads du brin + et du brin -	64
Figure 33 : Modèles statistiques de Markov	65
Figure 34 : Application des HMM pour l'annotation fonctionnelle du génome	66
Figure 35 : Exemple de résultats produit par le pipeline MEME-ChIP	67
Figure 36 : Identification de super-enhancers à l'aide de données de ChIP-seq avec ROSE ..	68
Figure 37 : Définition des circuits de régulation des facteurs de transcription	68
Figure 38 : Composition en nucléosome selon la taille des fragments ATAC-seq séquencés ..	69
Figure 39 : Pipeline d'analyse différentielle de données RNA-seq	70
Figure 40 : Définition des <i>anchors</i> liés aux pics HiChIP et aux sites de restrictions d'enzymes	72
Figure 41 : Définition des domaines de régulation des gènes dans GREAT	73
Figure 42 : Pipeline d'analyse de données ChIP-seq et ATAC-seq mis en place	76
Figure 43 : Pipeline d'analyse de données HiChIP mis en place	77
Figure 44 : L'intégration des données de NGS ATAC-seq, ChIP-seq et RNA-seq déterminent le paysage chromatinien tout au long du génome	78
Figure 45 : Utilisation des données de ChIP-seq de marques d'histones pour définir un modèle à 10 états chromatiniens fonctionnels avec chromHMM	79
Figure 46 : Distribution Génomique des sites de fixation du GR dans la lignée MM.1S	81
Figure 47 : La fixation de la Dex sur le génome est pré-déterminée par l'acétylation et l'accessibilité des régions	82
Figure 48 : Intégration des données pan-génomiques sur le locus du gène <i>TSC22D3</i> pour illustrer le rôle de la fixation du GR sur la régulation de la transcription	83
Figure 49 : Détermination des super-enhancers et facteurs de transcription cœurs de la lignée MM.1S	84
Figure 50 : Recherche de motifs <i>de novo</i> pour les sites de fixation du GR selon les différents états chromatiniens	85
Figure 51 : Interaction du GR avec le facteur pionnier du lignage B IRF4	86
Figure 52 : GO biological process enrichis pour les sites co-localisant la fixation du GR et du facteur pionnier IRF4	86
Figure 53 : Schéma illustrant les différents termes de l'ontologie GO biological process enrichis pour l'association GR:IRF4 en fonction de la présence d'un motif de fixation sur l'ADN	87
Figure 54 : Footprinting différentiel pour les données ATAC-seq en condition Dex versus EtOH	88
Figure 55 : Le footprinting révèle l'ouverture de nouveaux sites GRE	89
Figure 56 : Heatmap illustrant l'activation de régions H3K27ac liées à la fixation du GR ..	91
Figure 57 : Distribution génomique des régions H3K27ac Dex-induites	92
Figure 58 : Rôle des régions H3K27ac Dex-induites dans la régulation des gènes cibles	92
Figure 59 : Détermination des super-enhancers liés à la Dex et gène clés de réponse à la Dex	93
Figure 60 : Exemple de fixation du GR sur le locus <i>TSC22D3</i> qui illustre le mécanisme de réponse à la Dex	94
Figure 61 : Schéma illustrant comment les boucles chromatiniennes formées par la marque d'histone H3K27ac peuvent être étudiées par la technique du HiChIP	95

Figure 62 : Signal ChIP-seq H3K27ac des pics localisés hors et à l'intérieur des <i>anchors</i> des interactions HiChIP H3K27ac.....	96
Figure 63 : Caractérisation des boucles d'interactions H3K27ac HiChIP.....	97
Figure 64 : Annotation génomique des boucles d'interactions H3K27ac HiChIP.....	97
Figure 65 : La fixation du GR se fait majoritairement dans les anchors des boucles d'interactions chromatinienne H3K27ac.....	98
Figure 66 : Analyse différentielle des boucles d'interactions H3K27ac en condition Dex versus EtOH.....	99
Figure 67 : La fixation du GR est plus importante dans les boucles H3K27ac Dex-induites	99
Figure 68 : Annotation génomique des boucles d'interactions H3K27ac dynamiques.....	100
Figure 69 : Le GR se fixe dans les boucles H3K27ac Dex induites pour contrôler la transcription des gènes clés de réponse à la Dex.....	100
Figure 70 : Les régions H3K27ac Dex-induites sont associées aux boucles H3K27ac Dex-induites.....	101
Figure 71 : Les super-enhancers jouent également un rôle en contrôlant les boucles d'interactions chromatinienne et la transcription des gènes.....	101
Figure 72 : Les interactions H3K27ac permettent une organisation des enhancers en CLIQUES pour réguler la transcription des gènes.....	103
Figure 73 : Le rôle de l'enhancer majoritaire dans le recrutement d'enhancers secondaires pour contrôler la transcription des gènes.....	104
Figure 74 : Heatmap illustrant la fixation du CTCF en condition EtOH et Dex.....	105
Figure 75 : Le facteur CTCF permet de stabiliser les interactions stables et dynamiques des boucles H3K27ac Dex-induites.....	105
Figure 76 : Heatmap illustrant l'augmentation de l'accessibilité de la chromatine dans certaines régions ATAC-seq présentes dans les anchors des boucles H3K27ac Dex-induites...	106
Figure 77 : Recherche de motif <i>de novo</i> dans les régions ATAC Dex-induites et ATAC Dex-stables.....	107
Figure 78 : Volcano plot illustrant la répartition des régions ATAC Dex-induites sur le différentiel des boucles d'interactions H3K27ac.....	108
Figure 79 : Co-localisation des régions ATAC Dex-induites avec les super-enhancers et enhancers H3K27ac en condition Dex.....	108
Figure 80 : Clustering des cellules en réponse à la Dex ou non selon l'accessibilité de la chromatine.....	109
Figure 81 : Clustering des cellules en fonction de la phase du cycle cellulaire.....	110
Figure 82 : Estimation du ratio de régions accessibles (ROR) sur les régions ATAC Dex-induites.....	110
Figure 83 : Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique sur le gène <i>TSC22D3</i> (GILZ).....	111
Figure 84 : Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique sur le gène <i>FKBP5</i>	111
Figure 85 : Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique sur le gène <i>BCL2L11</i> (BIM).....	112

LISTE DES FIGURES

Figure 86 : Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique sur le gène <i>CXCR4</i>	112
Figure 87 : Estimation du ratio de cellules accessibles (RAC) basé sur les régions enhancers de 26 des 41 gènes Dex-up-régulés	113
Figure 88 : Identification des logFC élevés après application d'un modèle Gaussien.....	114
Figure 89 : Diagramme illustrant la médiane, le rang inter-quartile, la fraction par cellule (fpc) et l'entropie pour les cellules traitées en EtOH ou en Dex.....	114
Figure 90 : Heatmap montrant la co-expression des 51 gènes les plus différentiels entre EtOH et Dex.....	115
Figure 91 : Comparaison des IQR et des médianes des scDAGs par rapport à des gènes sélectionnés aléatoirement.....	116
Figure 92 : Distribution de la densité des coefficients de corrélation pour des gènes sélectionnés aléatoirement et les scDAGs	116
Figure 93 : Estimation du ratio de gènes répondant à la Dex (RRG)	117
Figure 94 : Boxplots illustrant la distribution des valeurs scRNA-seq pour des gènes d'intérêt	118
Figure 95 : UMAP single cell RNAseq	119
Figure 96 : UMAP révélant l'expression anti-corrélée de <i>BIM</i> et <i>CXCR4</i> dans les cellules	119
Figure 97 : Schéma récapitulant l'effet de la fixation du GR sur la régulation des gènes de réponse à la Dex.....	123

Liste des Tableaux

Table 1 : Principales anomalies génétiques primaires et secondaires retrouvées chez les patients atteints de Myélome Multiple	21
Table 2 : Table récapitulative des données générées et analysées	56
Table 3 : Annotation biologique et fonctionnelle du modèle à 10 états chromatinien réalisée pour la lignée MM.1S	80
Table 4 : Gènes induits contrôlés par au moins une interaction chromatinienne qui augmentent en réponse à la Dex et co-localisée avec un SE MM.1SDex.	102

Liste des Abréviations

ACC acetylCoA carboxylase

ADN acide desoxyribonucléique

Ag antigène

AID activation-induced desaminase

ALOX12B arachidonate 12-lipoxygenase type 12R

ARN acide ribonucléique

ATM ataxia-Telangiectasia mutated gene

ATR ATR Serine/Threonine kinase

BCL-6 B-cell Lymphoma 6

BIRC2 baculoviral containning repetition IAP protein IAP 2

BIRC3 baculoviral containning repetition IAP protein 3

BRAF B-RAF proto-oncogene

CARD11 caspase recruitment domain-containing protein 11

CCND1 cycline D1

CCND3 cycline D3

CDKN2 cyclin-dependant kinase inhibitor 2

CSH cellules souches hématopoïétiques

CYLD lysine-63 deubiquitinase

CYLD1 lysine 63 Deubiquinase

DBD DNA binding domain

Dex dexaméthasone

DIS3 exosom complex exonuclease

FAM46C polymerase Poly(A) Non Canonique

FAS fat acid synthase

FGFR3 fibroblast growth factor receptor 3

LISTE DES ABREVIATIONS

GR	récepteur aux glucocorticoïdes
GRBS	glucocorticoïd receptor binding site
GRE	glucocorticoïd response element
HLA-A	human leukocyte antigen A
HMM	Hidden Markov Model
IGH	chaînes lourdes des immunoglobulines
Igs	immunoglobulines
IKBKB	inhibitor of NF- κ B subunit kinase B
IL	interleukine
IRF8	interferon regulatory factor 8
ISRE	interferon signaling response element
KRAS	v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog
LB	lymphocyte B
LDB	ligand domain binding
LPL	lipoprotein lipase
LTB	lymphotoxine beta
MAF	facteur de transcription MAF Bzip
MAFB	facteur de transcription MAFB
MAGED1	melanoma-associated antigen D1
MAPK	mitogen-activated protein kinase
MGUS	monoclonal gammopathy of undetermined significance
MITF	microphthalmia-associated transcription factor
MM	myélome multiple
MMSET	multiple myeloma SET domain
MYC	proto-oncogene C-Myc
NF-κB	nuclear factor kappa B

LISTE DES ABREVIATIONS

NRAS neuroblastoma RAS oncogene

OL-I organes lymphoïdes primaires

PCL plasma cell Leukemia

PNRC1 prolin-rich nuclear receptor coactivator 1

PRDM1 PR-1 domain-containing with ZNF

RAF RAF-GTPase-Activated synaptic protein 1

RAS RAS-GTPase-Activated synaptic protein 1

RB1 retinoblastoma protein 1

SE super-enhancers

SMM smoldering multiple myeloma

STAT6 signal transduction and activator of transcription 6

TAD topologically-associated domain

TP53 tumoral protein 53

TRAF3 TNF receptor-associated factor 3

TRAF3IP1 TRAF3 interacting protein 1

TSS transcription start site

VLDL very low density protein

Table des Matières

Liste des Figures	3
Liste des Tableaux	7
Liste des Abréviations	8
1 Introduction	14
2 Le Myélome Multiple	15
2.1 Physiopathologie du Myélome Multiple	15
2.1.1 Epidémiologie	15
2.1.2 Manifestations cliniques du Myélome Multiple	15
2.1.3 Diagnostic de la pathologie	16
2.2 Développement et progression du Myélome Multiple	16
2.2.1 Différenciation du lignage B	16
2.2.2 Les différents stades du Myélome Multiple	17
2.2.3 Hétérogénéité de la pathologie	19
2.3 Voies de signalisation importantes et perturbées par les anomalies génétiques du Myélome Multiple	24
2.3.1 Voie de signalisation NF- κ B	24
2.3.2 Voie de signalisation RAS/RAF/MAPK	26
2.3.3 La différenciation des cellules plasmocytaires	27
2.3.4 Autres voies perturbées dans le Myélome Multiple	29
2.4 Traitements du Myélome Multiple	29
2.5 Rechute du Myélome Multiple	30
3 Mécanismes de la régulation d'expression des gènes	31
3.1 Transcription des gènes	31
3.2 Régions régulatrices de la transcription et facteurs de transcription	32
3.3 Modifications post-traductionnelles des histones	33
3.4 Accessibilité de la chromatine	35
3.5 Conformation 3D de la chromatine	36
4 Les Glucocorticoïdes	38
4.1 Rôle physiologique du Récepteur aux Glucocorticoïdes	38
4.1.1 Métabolisme	38
4.1.2 Inflammation	39
4.1.3 Utilisation dans diverses pathologies	40
4.2 Mode d'action du récepteur aux glucocorticoïdes	41
4.2.1 Composition et structure du récepteur aux glucocorticoïdes	41
4.2.2 Génomique appliquée au GR	41

5	Sujet de thèse	46
6	Techniques et méthodes d'analyse	47
6.1	Matériel biologique utilisé	47
6.2	Techniques de biologie moléculaire	47
6.2.1	Immunoprécipitation de protéines	47
6.2.2	Techniques de détection d'accessibilité de la chromatine	48
6.2.3	Analyse des niveaux d'expression ARN	49
6.2.4	Analyse des interactions ADN:ADN liées à la présence d'une protéine	50
6.2.5	Techniques de séquençage nouvelle génération	52
6.2.6	Immunoprécipitation de protéines suivie de spectrométrie de masse	53
6.2.7	Nouvelles approches en cellule unique	54
6.2.8	Génération des données	56
6.3	Analyses computationnelles	57
6.3.1	Préparation des données pour alignement	57
6.3.2	Transformée de Burrows-Wheeler pour l'alignement de reads	58
6.3.3	Analyse post-alignement des données de ChIP-seq	63
6.3.4	Analyse post-alignement des données d'ATAC-seq	69
6.3.5	Analyse différentielle des données RNA-seq	70
6.3.6	Analyse post-alignement des données de HiChIP-seq	71
6.3.7	Analyses bioinformatiques diverses	72
6.3.8	Analyses spécifiques au séquençage en cellule unique	74
6.3.9	Pipelines d'analyse bioinformatique	74
7	Résultats	78
7.1	Le paysage chromatinien détermine la fixation du GR sur le génome	78
7.1.1	Annotation fonctionnelle de la lignée cellulaire MM.1S	79
7.1.2	La fixation du GR dépend du paysage chromatinien de la cellule	81
7.1.3	Le GR interagit avec un facteur pionnier du lignage B: IRF4	85
7.1.4	La fixation du GR ouvre des nouveaux sites de fixation sur le génome	88
7.2	Le GR active des régions régulatrices Dex-spécifiques	89
7.3	Action génomique du GR dans la régulation des boucles chromatinienne	95
7.3.1	Caractérisation des interactions chromatinienne	95
7.3.2	La fixation du GR dans les boucles pré-existante induit une régulation de ces boucles	98
7.3.3	Rôle des régions H3K27ac Dex-induites au niveau des boucles de régulation	101
7.3.4	Le CTCF maintient les boucles chromatinienne stables	104
7.3.5	La régulation des boucles chromatinienne est permise par l'ouverture de régions GR-spécifiques dans les anchors de ces interactions	106
7.4	Hétérogénéité de l'accessibilité de la chromatine en réponse à la Dex	109
7.5	Hétérogénéité transcriptomique en réponse à la Dex	113

TABLE DES MATIERES

8 Discussion et conclusion	120
8.1 Discussion	120
8.2 Conclusion	124
Références	125
Annexes	136

1 Introduction

Le myélome multiple (MM) est une hémopathie maligne caractérisée par une prolifération incontrôlée de plasmocytes anormaux dans la moelle osseuse. Le développement et la progression du MM se fait via l'acquisition d'anomalies génétiques diverses chez les patients, au sein même de la population tumorale ce qui en fait une maladie très hétérogène. Ainsi, les anomalies génétiques identifiées dans cette pathologie vont impacter de nombreuses voies de signalisation importantes pour le plasmocyte qui vont lui permettre de proliférer. A l'heure actuelle et malgré l'émergence de nouveaux traitements, cette pathologie demeure toujours incurable.

Les glucocorticoïdes (GCs) sont largement utilisés dans le traitement du MM, essentiellement en combinaison avec d'autres molécules. Leurs effets se font au travers de la liaison avec le récepteur aux GCs (GR) qui, une fois transloqué dans le noyau, se fixe sur son site de reconnaissance (GRE) pour activer ou réprimer la transcription de ses gènes cibles. Parmi les GCs, la dexaméthasone (Dex) est utilisée dans toutes les phases de traitement du Myélome (induction, consolidation et maintenance). Cependant et malgré un bénéfice indéniable sur la maladie, tous les patients ne répondent pas à la Dex de la même manière. De plus, alors que l'efficacité des GCs contribue largement à l'apoptose spécifique des lymphocytes, les mécanismes d'actions moléculaires de la Dex restent à ce jour encore méconnus.

Afin de comprendre les mécanismes d'actions moléculaires de la Dex, l'étude seule des régions codantes de l'ADN n'est pas suffisante pour comprendre la régulation de l'expression des gènes. L'évolution des technologies de séquençage au cours de ces dernières années a donc permis l'étude plus approfondie de l'ensemble des mécanismes épigénétiques de régulations de la transcription des gènes au sein de la cellule et permet de mieux définir le rôle de l'ADN non codant dans cette régulation.

Dans ce contexte précis, nous avons réalisé des analyses pan-génomiques intégrant différentes données épigénétiques telles que l'accessibilité de la chromatine, la fixation de facteurs de transcription et les modifications d'histones couplés à des analyses de conformation de la chromatine afin de mieux comprendre le mode d'action moléculaire de la Dex dans le Myélome mais également de comprendre les mécanismes de résistance mis en place par la cellule.

Je présente dans cette thèse la pathologie étudiée (Myélome Multiple) ainsi que les différents mécanismes de régulations épigénétiques de la transcription des gènes avant d'aborder le mécanisme d'action des glucocorticoïdes, majoritairement utilisés pour le traitement des leucémies et du myélome. Dans un second temps, je décris les analyses génomiques et les méthodes bioinformatiques associées que j'ai mises en place pour réaliser les différentes études de cette thèse. Enfin les résultats obtenus au cours de cette thèse seront présentés.

2 Le Myélome Multiple

2.1 Physiopathologie du Myélome Multiple

2.1.1 Epidémiologie

Le Myélome est une hémopathie maligne caractérisée par la prolifération de plasmocytes anormaux dans la moelle osseuse. Bien que de nombreux facteurs génétiques et environnementaux aient été identifiés ces dernières années, les causes du myélome multiple restent à ce jour mal définies (Li et al. (2016); Mitchell et al. (2016)).

Le Myélome Multiple représente 1.8% des cancers totaux ce qui en fait une maladie orpheline. Cependant, cette pathologie représente 20% de la mortalité annuelle des maladies hématologiques dans le monde. Elle représente environ 10% des cancers hématologiques (Siegel et al., 2019).

Le principal facteur lié à cette maladie est l'âge, elle touche généralement les personnes âgées de plus de 70 ans (Palumbo et al., 2011). Cette pathologie a une plus grande incidence sur les hommes que les femmes et chez les populations afro-américaines que chez les populations blanches (Altekruse et al., 2009). Malgré l'émergence de nouveaux traitements qui ont permis d'augmenter la survie d'environ 5 ans chez les patients, le MM reste aujourd'hui une maladie incurable (Siegel et al., 2016).

2.1.2 Manifestations cliniques du Myélome Multiple

Il existe de nombreux symptômes liés à cette pathologie, le principal étant les douleurs osseuses chez le patient (Rajkumar and Kyle, 2005). Les os sont constitués de deux grandes classes de cellules: les ostéoblastes qui permettent l'ossification et les ostéoclastes qui détruisent les ostéoblastes et donc les os. Dans la pathologie, les cellules myélomateuses localisées dans les os sécrètent de nombreuses cytokines qui vont avoir pour effet de stimuler les ostéoclastes et d'inhiber les ostéoblastes, ce qui va causer petit à petit de nombreuses lésions osseuses chez le patient.

Un autre symptôme qui est retrouvé chez environ 10% des patients atteints de Myélome est l'hypercalcémie. Ce symptôme est bien évidemment directement lié à la résorption osseuse qui provoque une augmentation du taux de calcium libre dans le sang. L'hypercalcémie peut également être la cause de perte de poids, de nausées, de vomissements et de troubles du rythme cardiaque. On retrouve également chez 20 à 40% des patients atteints de myélome des insuffisances rénales (Dimopoulos et al., 2008). Ces insuffisances peuvent atteindre des formes graves qui nécessitent des greffes et des dialyses pour 10% des cas. Les insuffisances rénales peuvent être causées par production de chaînes légères et de fragments d'immunoglobulines (Igs) par les plasmocytes malins qui vont s'accumuler au niveau des reins et bloquer leur fonction de filtration du sang. Il faut également savoir que l'hypercalcémie est la deuxième cause d'insuffisance rénale.

Enfin, l'anémie due à une diminution des globules rouges, des infections dues à une diminution des globules blancs (neutropénie) et des saignements dus à une diminution du nombre de plaquettes

(thrombopénie) sont également des symptômes du myélome dû à l'envahissement de la moelle osseuse par les cellules tumorales (Mittelman, 2003).

2.1.3 Diagnostic de la pathologie

Le MM est précédé de différents stades asymptomatiques (monoclonal gammopathy of undetermined significance (MGUS) et smoldering multiple myeloma (SMM); cf. 2.2.2) rendant son diagnostic tardif. Le principal diagnostic repose donc sur les lésions provoquées aux organes ou durant un diagnostic fortuit qui fait apparaître le pic d'Ig monoclonal lors d'une électrophorèse des protéines.

Une fois le pic d'Igs détecté chez le patient, le diagnostic est complété avec des analyses sanguines et urinaires, une biopsie de la moelle osseuse (Myélogramme) et des tests d'imageries (myélogramme). Pour différencier le stade symptomatique du myélome des stades asymptomatiques, l'International Myeloma Working Group a défini plusieurs critères (critères CRAB): présence des symptômes (hypercalcémie, atteintes rénales, anémie et lésions osseuses), présence de plasmocytes monoclonaux à plus de 10% dans la moelle, un ratio de chaînes légères des Igs libres/liées supérieur à 100 ou encore au moins deux lésions focales osseuses de plus de 5mm.

2.2 Développement et progression du Myélome Multiple

2.2.1 Différenciation du lignage B

Le lignage B est issu des cellules souches hématopoïétiques (CSH). Deux grands phénomènes dirigent la différenciation de la cellule B: la Lymphopoïèse ainsi que l'Immunopoïèse.

Durant la Lymphopoïèse, la cellule souche se différencie en quatre étapes: elle va passer par le stade cellule pro-B, pré-B avant de devenir un lymphocyte B (LB) immature, puis un LB mature naïf. Cette différenciation se fait à l'intérieur des tissus lymphoïdes foetaux (lors du développement embryonnaire) et plus généralement dans la moelle osseuse (organes lymphoïdes primaires, OL-1). Au cours de cette différenciation cellulaire, la cellule va subir des réarrangements au niveau des locus des gènes codant pour la chaîne lourde μ et de la pseudo chaîne légère des Igs. Ainsi, la cellule pré-B exprimera à sa membrane un précurseur des Igs formé d'une chaîne lourde μ et d'une chaîne légère de substitution (pseudo-chaîne légère). Le LB immature exprimera uniquement à sa membrane des IgM et subit une sélection négative qui va provoquer l'apoptose des cellules reconnaissant un antigène du soi. Le LB mature naïf exprimera à la fois des IgD et des IgM. Dans cet état cellulaire, la cellule est dite naïve car elle n'a pas encore eu de rencontre avec l'antigène (Ag) (Figure1A).

Lorsque le lymphocyte B mature naïf rencontre l'Ag, l'immunopoïèse va débiter et le lymphocyte va pouvoir donner deux types de cellules: les plasmocytes sécréteurs d'Igs et les cellules B mémoires. Une fois sorties des centres germinatifs, les cellules B mémoires vont circuler dans le sang et seront capables d'être réactivées rapidement lors d'une nouvelle rencontre avec l'antigène.

Les plasmocytes produiront les anticorps spécifiques des antigènes, une partie migrera dans la moelle osseuse pour se différencier en plasmocytes à longue durée de vie. Dans les conditions normales, ces cellules sont incapables de proliférer et ne sécrètent qu'un seul type d'anticorps (Kumar et al., 2017).

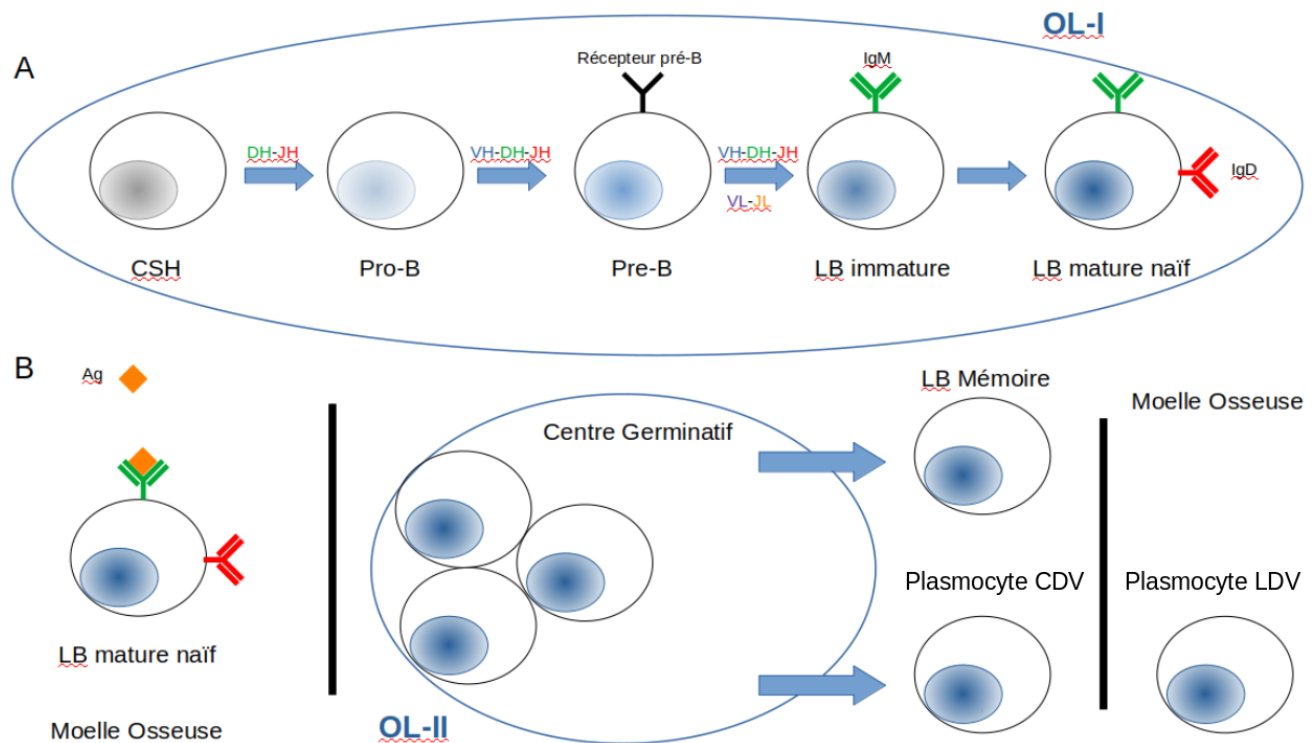


Figure 1: **La différenciation du lignage B se distingue en deux grands phénomènes: la lymphopoïèse et l'immunopoïèse.** (A) La lymphopoïèse permet la différenciation des CSH en LB matures naïfs au sein des OL-I; (B) L'immunopoïèse permet la différenciation des LB matures en plasmocytes dans les OL-II. ; Plasmocyte CDV = Plasmocytes à courte durée de vie; Plasmocytes LDV = Plasmocytes à longue durée de vie.

2.2.2 Les différents stades du Myélome Multiple

Différents stades ont été décrits dans le cadre de cette pathologie: (i) le stade MGUS qui représente le stade pré-tumoral de la maladie, (ii) le stade SMM qui correspond à une augmentation des plasmocytes malins dans la moelle osseuse, sans apparition de signes cliniques, (iii) le stade myélome avéré (MM) qui correspond à une apparition des signes cliniques du myélome et enfin, pour une minorité de patients, (iv) le stade leucémie plasmocytaire (PCL) où les anomalies génétiques ont permis à la cellule de s'affranchir de son micro-environnement et de rejoindre la circulation sanguine (Figure 2).

Le MGUS est le stade précurseur du myélome multiple. Ce stade est dit asymptomatique et est généralement associé avec l'âge des patients. D'un point de vue biologique, le stade MGUS est caractérisé par un taux d'infiltration des plasmocytes monoclonaux inférieur à 10% dans la moelle osseuse et un taux d'Igs sériques monoclonales inférieur à 30g/L. Ce stade est relativement stable, sa progression annuelle vers un MM est de 1%. Du fait de l'absence de symptômes, la quasi-totalité des MGUS ne sont pas diagnostiqués chez les patients. D'un point de vue génétique, le passage de la cellule B post-centre germinatif vers le stade MGUS est identifié par des transformations primaires sur les gènes des Igs. Ces translocations vont permettre à des gènes de se retrouver sous contrôle d'enhancers du locus *IGH*. Ces translocations vont donner au clone un avantage prolifératif. L'ensemble de ces anomalies génétiques est répertorié dans la Table 1. Cependant le stade MGUS n'est pas une transition obligatoire dans la progression du Myélome, de nombreux Myélomes diagnostiqués chez les patients ne passent pas systématiquement par un stade MGUS.

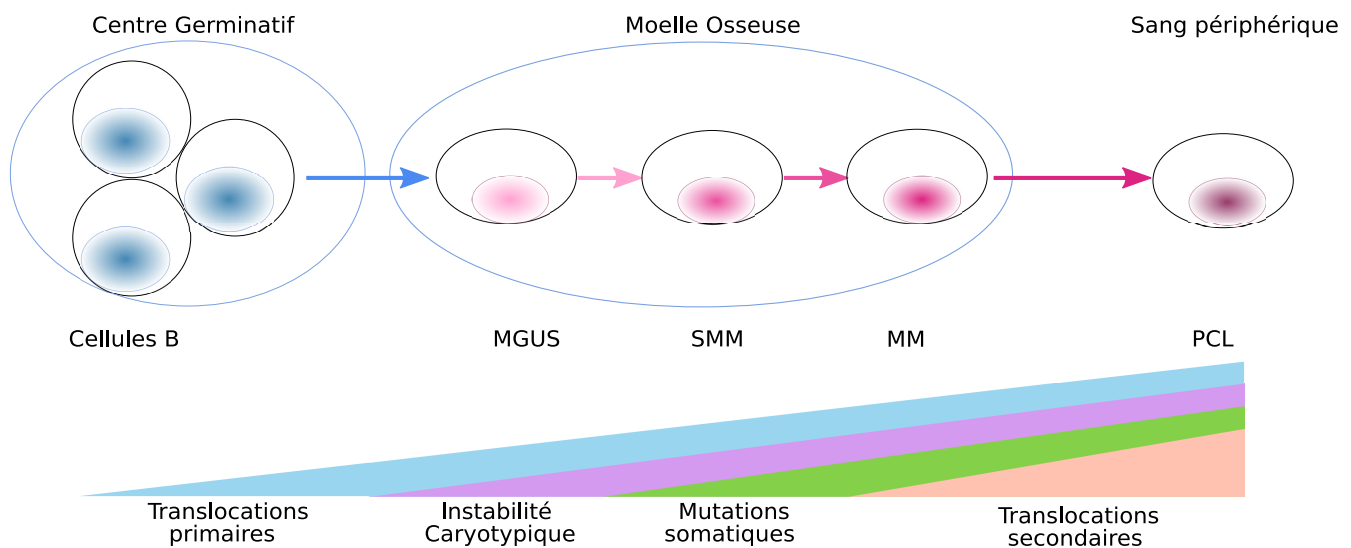


Figure 2: **Le développement et la progression du Myélome Multiple: localisation et anomalies génétiques.** Illustration de l'évolution de la cellule B post-centre germinatif vers les différents stades de MM en lien avec l'apparition d'anomalies génétiques.

Le SMM est donc la phase intermédiaire entre le MGUS et le MM. Ce qui le diffère du MGUS est le taux d'Igs monoclonaux sériques qui est supérieur à 30g/L. Le taux d'infiltration des plasmocytes est également nettement plus important (de 10% à 60%). Néanmoins, en raison de l'absence de symptôme clinique, ce stade est difficilement détectable même si une prise en charge précoce de la pathologie améliore le pronostic des patients. La différence entre les stades MGUS/SMM d'une part et MM d'autre part est donc surtout expliquée par l'absence de signes cliniques ainsi que le pourcentage d'envahissement de la moelle osseuse par les cellules plasmocytaires.

2.2.3 Hétérogénéité de la pathologie

Hétérogénéité inter-patient

Le Myélome est une maladie très hétérogène, ce qui explique une évolution différente de la pathologie entre les patients. Plusieurs événements ont été décrits à différents stades de la maladie. Ainsi, il existe divers événements génétiques qui seront plus ou moins précoces. Deux grands groupes d'anomalies sont catégorisés: les événements dits primaires et les secondaires. Ces différentes anomalies ne sont pas retrouvées de façon uniforme chez les patients atteints de Myélome.

Pour ce qui est des événements génétiques primaires, ils peuvent se distinguer en deux grandes catégories: les tumeurs hyper-diploïdes et les non hyper-diploïdes (Table 1).

Les tumeurs hyperdiploïdes sont caractérisées par un nombre aberrant de copies de différents chromosomes (3, 5, 7, 9, 11, 15, 19 et/ou 21) que l'on appelle trisomie. Ainsi, un caryotype d'une personne saine présente 46 chromosomes (22 paires d'autosomes + 2 chromosomes dits sexuels, les gonosomes) alors qu'une cellule de personne souffrant d'un myélome et présentant une tumeur dite hyper-diploïde peut présenter un caryotype allant de 47 jusqu'à 74 chromosomes. On considère aujourd'hui que ce type de tumeur concerne 50% des patients atteints de MGUS, SMM et MM.

Les tumeurs dites "non hyper-diploïdes" sont généralement caractérisées par des translocations génétiques aberrantes impliquant le locus du gène des Igs. En effet, lors du développement des cellules B, l'ADN codant pour les régions hypervariables de la chaîne lourde des Igs subit une hyper-mutation somatique pour produire des Ac de grande spécificité. La maturité de ces Ac est ensuite augmentée par la commutation de classe qui va permettre à la cellule d'exprimer une Ig de différents isotypes. Ces deux mécanismes (hyper-mutation somatique et commutation de classe) nécessitent d'une part l'expression de l'activation-induced desaminase (AID) et sont d'autre part régulés par des cassures de l'ADN double brin localisées dans des loci codant pour des Ig. Si ces cassures induites par l'AID sont en temps normal réparées localement, il est possible qu'une jointure de l'ADN s'effectue avec une autre cassure de l'ADN double brin localisée beaucoup plus loin dans le génome ou sur un autre chromosome ce qui va créer une translocation génétique (Morgan et al., 2012).

Ainsi, les tumeurs non hyper-diploïdes sont caractérisées par des translocations affectant les gènes codants pour les chaînes lourdes des Igs dans la majorité des cas. En ce qui concerne le Myélome, 90% de ces translocations affectent en effet le chromosome 14 (Egan et al. (2012), Figure 3; Table 1).

Dans le Myélome, les translocations primaires concernent essentiellement trois gènes:

- **La translocation t(11;14):** Ces translocations sur ce groupe de gènes sont retrouvées dans environ 20 à 25% des MM. Le gène cycline D1 (CCND1) est impliqué dans la régulation du cycle cellulaire et tout particulièrement dans la transition de la phase G1 à la phase S.
- **La translocation t(4;14):** Cette translocation impliquant le gène *FGFR3* ainsi que le gène *MMSET* est retrouvée dans environ 15% dans le MM. Cette translocation est de mauvais pronostic chez les patients atteints de myélome. Dans 70% des cas, on constate une sur-expression de *FGFR3* accompagnant la surexpression de *MMSET* (Xie and Chng, 2014). *MMSET* est une enzyme responsable du dépôt d'une di-méthylation sur résidu lysine 36

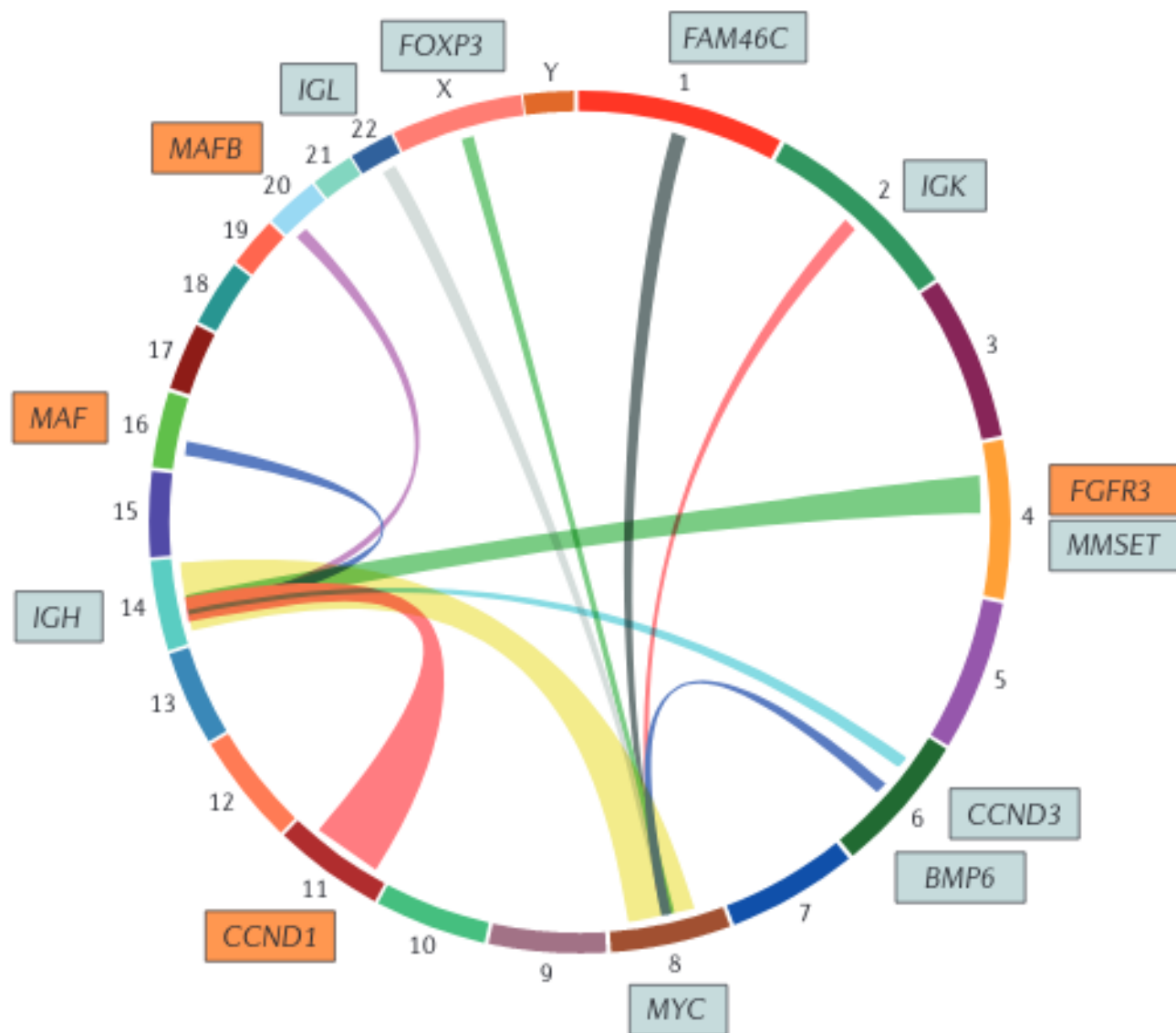


Figure 3: **Principales translocations et gènes majoritairement impliqués dans le Myélome Multiple**, Figure tirée de Manier et al. (2017).

de l'histone H3 (H3K36me2) qui permet de réguler la structure de la chromatine et donc la transcription des gènes (Popovic et al., 2014). L'inhibition de l'expression empêche la prolifération et induit l'apoptose des cellules de Myélome Multiple.

- **La translocation t(16;14):** On retrouve cette translocation dans 10% des MM (Kuehl and Bergsagel, 2002), elle implique le gène codant pour le facteur de transcription c-MAF qui est surexprimé dans de nombreux Myélomes.

Un certain nombre d'évènements génétiques secondaires vont conférer à la cellule tumorale un avantage sélectif responsable de la progression et même de la rechute du patient.

Dans 50% des Myélomes avancés, on identifie des anomalies génétiques qui impliquent le gène *MYC* (et plus occasionnellement *LMYC* et *NMYC*). Ces anomalies génétiques sont retrouvées dans 60% des tumeurs hyperdiploïdes et sont également de mauvais pronostic. De nombreuses voies de signalisation sont également concernées par ces anomalies génétiques (cf 2.3) notamment la voie des RAS/MAPK perturbée chez environ 40% des patients (cf 2.3.2), la voie NF- κ B (cf 2.3.1) qui est dérégulée chez 20% des patients (Demchenko et al., 2010) ou bien encore les différentes voies de réponses aux dommages de l'ADN (Chapman et al., 2011). Il a également été montré que ces évènements génétiques secondaires impliquent de nombreuses délétions ou amplifications dans le génome (ou CNVs).

Finalement, le séquençage haut-débit et particulièrement une étude sur une cohorte de 38 patients (Chapman et al. (2011), Magrangeas et al. (2013), Morgan et al. (2012)) sur lesquels ont été réalisés des séquençages de type Whole Genome (WGS) et/ou Whole Exome (WES) ont permis l'identification de 10 gènes qui ont une fréquence de mutation significativement importante chez les patients atteints de Myélome Multiple: *NRAS*, *KRAS*, *FAM46C*, *DIS3*, *TP53*, *CCND1*, *PNRC1*, *ALOX12B*, *HLA-A* et *MAGED1*.

Évènements génétiques	Anomalies génétiques	Fréquence
Évènements primaires		
Translocations	t(11;14), CCND1	15%
	t(4;14),MMSET/FGFR3	15%
	t(6;14),CCND3	2%
	t(14;16), MAF	5%
	t(14;20), MAFB	1%
	Hyperdiploïdies	50%
Aneuploïdies	del13q, potentiellement RB1	40%
Évènements secondaires		
Translocations	MYC	15%
Gain/perte bras chromosomiques	Gain 1q, MYC	40%
	Perte 1p, CDKN2 ou FAM46C	30%
	Perte 17p, TP53	10%
Mutations poncutelles	Voie des MAPK, KRAS, NRAS, BRAF	45%
	Voie NF- κ B, TRAF3, CYLD, LDB	15%
	Voie de réparation de l'ADN, TP53, ATM, ATR	10%

Table 1: Principales anomalies génétiques primaires et secondaires retrouvées chez les patients atteints de Myélome Multiple.

Hétérogénéité intra-patient

A la fin des années 70, les premiers modèles d'évolution clonale Darwiniens dans les cancers ont émergé. Ces modèles présentent une évolution de la tumeur où les clones cellulaires acquièrent au fur et à mesure des altérations génétiques différentes qui vont conférer à certains clones un avantage prolifératif favorisant son émergence, la croissance et la survie de la tumeur.

Ce modèle d'évolution clonale a également été testé pour le Myélome (Manier et al., 2017) où les cellules propagatrices du myélome sont en réalité un mélange de sous clones génétiques acquérant au fur et à mesure de nouvelles anomalies génétiques permettant au clone d'avoir un avantage sélectif permettant l'évolution de la maladie (Figure 4). Chez les patients atteints de myélome, la pression thérapeutique entraîne une réorganisation et une diversification des populations cellulaires de la tumeur et on observe, en effet, l'émergence de clones jusqu'alors minoritaires qui deviennent dominants à la rechute. Ces clones plus agressifs accélèrent la progression de la maladie. (Margeas et al., 2013).

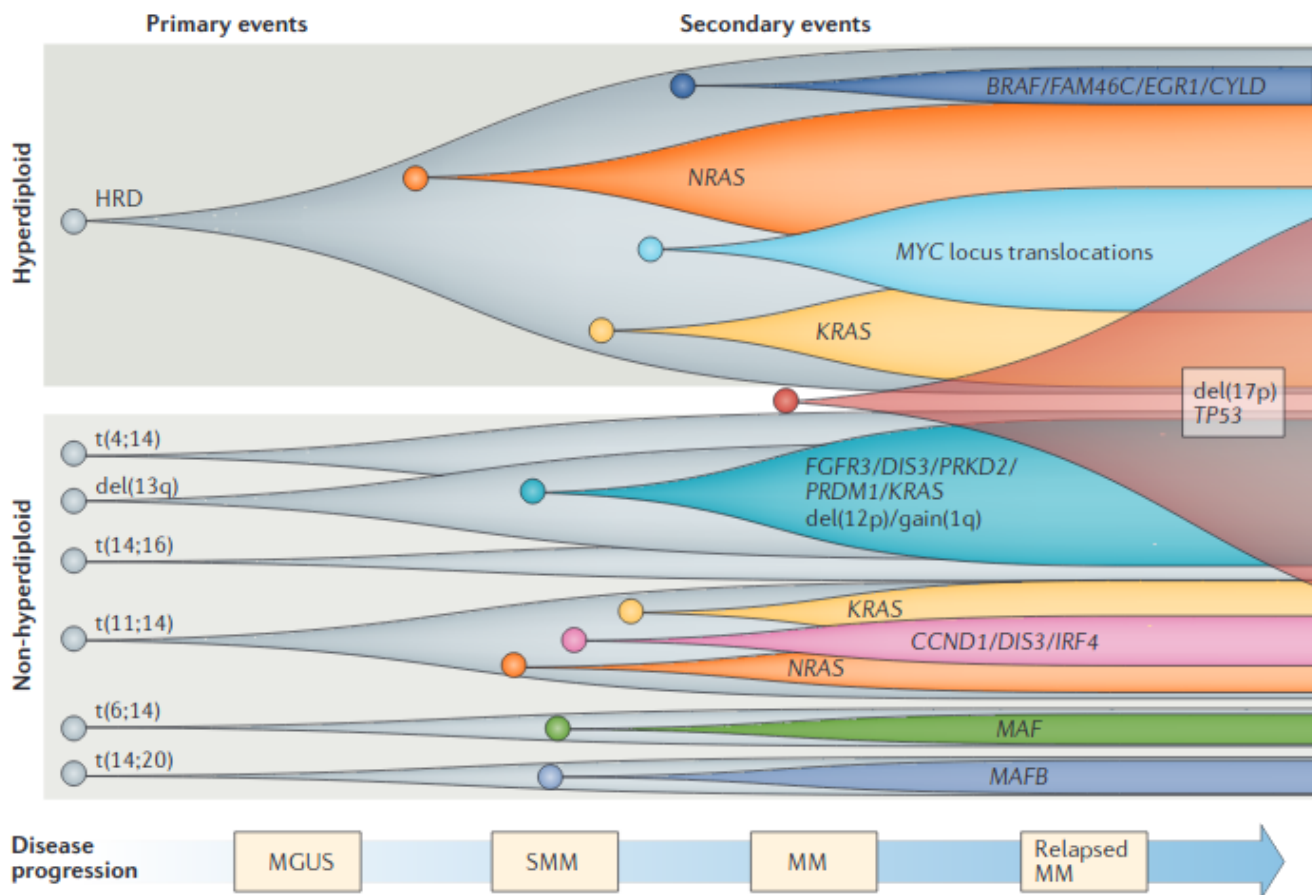


Figure 4: **Proposition de modèles d'évolution clonale dans le Myélome Multiple**, Figure tirée de Manier et al. (2017).

Les évolutions des technologies de séquençage et des analyses génomiques associées ont donc permis

de montrer qu'outre l'hétérogénéité intra-patient, il existe une hétérogénéité au sein même de la tumeur. C'est notamment le cas d'une étude d'analyses génomiques sur des échantillons prélevés à différents stades de progression du MM (Keats et al., 2012). Cette étude propose deux modèles d'évolution clonale: un modèle où l'évolution se fait de façon linéaire et un autre où l'évolution se fait de façon hétérogène avec différentes ramifications des sous populations clonales.

Pour appuyer ces hypothèses d'hétérogénéité intra-patient, d'autres études de séquençage de génomes entiers ont été réalisées (Egan et al., 2012) sur 4 tumeurs à différents stades de progression de la maladie (diagnostic, première rechute, seconde rechute et stade terminal). Cette étude a révélé la détection de variants nucléotidiques présents dans certains sous-clones et illustrant cette hétérogénéité et l'évolution clonale pour le MM. L'analyse de ces variants a pu mettre en évidence une grande diversité évolutive au fur et à mesure de la progression de la maladie et une pression sélective des clones suite au traitement. Ils ont ainsi montré chez un patient présentant un clone majoritaire au diagnostic, que ce clone était éliminé par un premier traitement mais qu'un clone jusque là minoritaire émergeait lors de la rechute de ce patient. Ce clone était résistant au premier traitement. Lorsque ce patient à la rechute recevait un second traitement, le clone résistant à la rechute était éliminé mais on constatait la réémergence du clone majoritaire au diagnostic qui lui était plutôt résistant à ce second traitement.

Par la suite, des études sur de plus grandes cohortes de patients ont été réalisées allant dans le sens de ce modèle d'évolution clonale. C'est notamment le cas d'une étude effectuée au sein de notre laboratoire (Magrangeas et al., 2013) qui comparait des échantillons de patients à la fois au diagnostic et à la rechute et qui a proposé deux types d'évolutions sous clonales dans le cas du Myélome Multiple: 2/3 des patients suivaient un modèle linéaire caractérisé par un sous clone au diagnostic qui fait acquisition de lésions génétiques et qui évolue vers un nouveau sous clone dominant à la rechute. Le reste des patients (1/3) présentait un modèle d'évolution clonale non linéaire dans lequel un sous clone majoritaire était éliminé par le traitement ce qui favorisait par la suite l'apparition d'un sous clone mineur lors de la rechute du patient. Cette étude a également révélé un lien entre ce modèle non linéaire et l'utilisation du bortezomib, un traitement efficace pour éliminer le clone majoritaire mais qui permettrait également la persistance de sous clones favorisant la rechute du patient.

D'autres études ont également montré la présence de populations sous clonales (Lohr et al. (2014) et Walker et al. (2015)), et ont respectivement identifié au moins 3 sous clones (jusqu'à 7 chez certains patients) et au moins 5 sous clones (jusqu'à 10 chez certains patients). Ces études ont mis en évidence des mutations génétiques qui étaient clonales (retrouvées dans tous les clones) comme *DIS3*, *KRAS* et d'autres sous clonales (dans moins de 10% de la population clonale) comme *BRAF* (Lohr et al., 2014). D'autres mutations génétiques importantes sur les gènes *FAM46C* et *TRAF3* sont également plus fréquemment observées dans des populations sous-clonales (Walker et al., 2015). Cette étude a également montré des mutations génétiques à divers niveaux de la voie des MAP-kinases dans des sous clones distincts.

Enfin la complexité génétique dépend également du stade de progression du myélome. Il est souvent observé que l'hétérogénéité intra-patient est plus réduite au stade MGUS comparé aux stades SMM et MM qui eux présentent déjà une forte hétérogénéité tumorale.

2.3 Voies de signalisation importantes et perturbées par les anomalies génétiques du Myélome Multiple

2.3.1 Voie de signalisation NF- κ B

La voie NF- κ B est activée de façon normale au cours de la réponse immunitaire et dans la réponse au stress cellulaire. Cinq sous-unités de cette voie ont été décrites chez les mammifères : RelA (p65), RelB, cRel, p50 (p105) et p52 (p100). Chacune de ces sous-unités possède un domaine Rel-homologue en N-terminal qui va permettre la fixation spécifique sur les sites κ B de l'acide desoxyribonucléique (ADN). De plus, RelA, RelB et cRel ont également un domaine d'activation de la transcription du côté C-terminal. Cela indique que seules les sous-unités RelA, RelB et cRel peuvent jouer un rôle en tant qu'activateur de la transcription. Les sous-unités p50 et p52 doivent donc interagir (hétéro-dimérisation) avec d'autres facteurs activateurs pour pouvoir effectuer une régulation activatrice de leurs gènes cibles. Un hétéro-dimère p50:p52 est donc généralement associé à une répression de la transcription s'il se fixe sans co-facteur.

L'activation de cette voie va conduire à l'export d'homo et d'hétéro-dimères vers le noyau afin de pouvoir se fixer sur l'ADN et ainsi réguler la transcription de gènes (Figure 5). Il existe deux mécanismes activateurs de la voie NF- κ B. Le premier mécanisme, dit voie canonique, est une réponse à des signaux pro-inflammatoires qui va induire la dégradation d'un inhibiteur de NF- κ B ($I\kappa$ B). C'est un mécanisme essentiel pour la rapidité de la réponse immune qui va permettre la translocation dans le noyau des dimères suivants : RelA:p50, cRel:p50, RelA:RelA, cRel:cRel. La seconde voie, dite non canonique, permet l'organogenèse des organes lymphoïdes secondaires, le développement des cellules B et leur survie, ce sont les dimères RelB:p52, p52:p52 et p50:p52 qui seront exportés vers le noyau. Enfin une activation simultanée de ces deux mécanismes peut également conduire à la formation de dimères hybrides comme RelA:p52. En tout, 14 dimères NF- κ B différents ont été observés lorsque les deux mécanismes sont actifs dans le lymphome (Zhao et al., 2014).

Dans le cas du Myélome Multiple, la voie NF- κ B est perturbée chez environ 20% des patients. Des anomalies génétiques sur les régulateurs négatifs de la protéine NIK (Figure 5): *TRAF3*, *TRAF2* et *CIAP1/2* ou des mutations activatrices induisant un haut niveau d'expression du récepteur membranaire *CD40* ont pour conséquence une activation constitutive des deux voies NF- κ B. Ces anomalies vont permettre à la cellule de s'affranchir de son micro-environnement et de migrer vers le sang périphérique. Des anomalies perturbant significativement d'autres gènes tels que *CYLD1* (voie canonique) et *LTB* (voie non canonique) ont récemment été mis en évidence. D'autres gènes fréquemment touchés par des anomalies dans le Myélome ont été répertoriés: *IKBKB*, *BIRC2*, *BIRC3* (voie canonique et non canonique), *CARD11* (voie canonique) et *TRAF3IP1* (voie non canonique). La perturbation de cette voie de signalisation n'est cependant pas un marqueur pronostic des patients diagnostiqués MM.

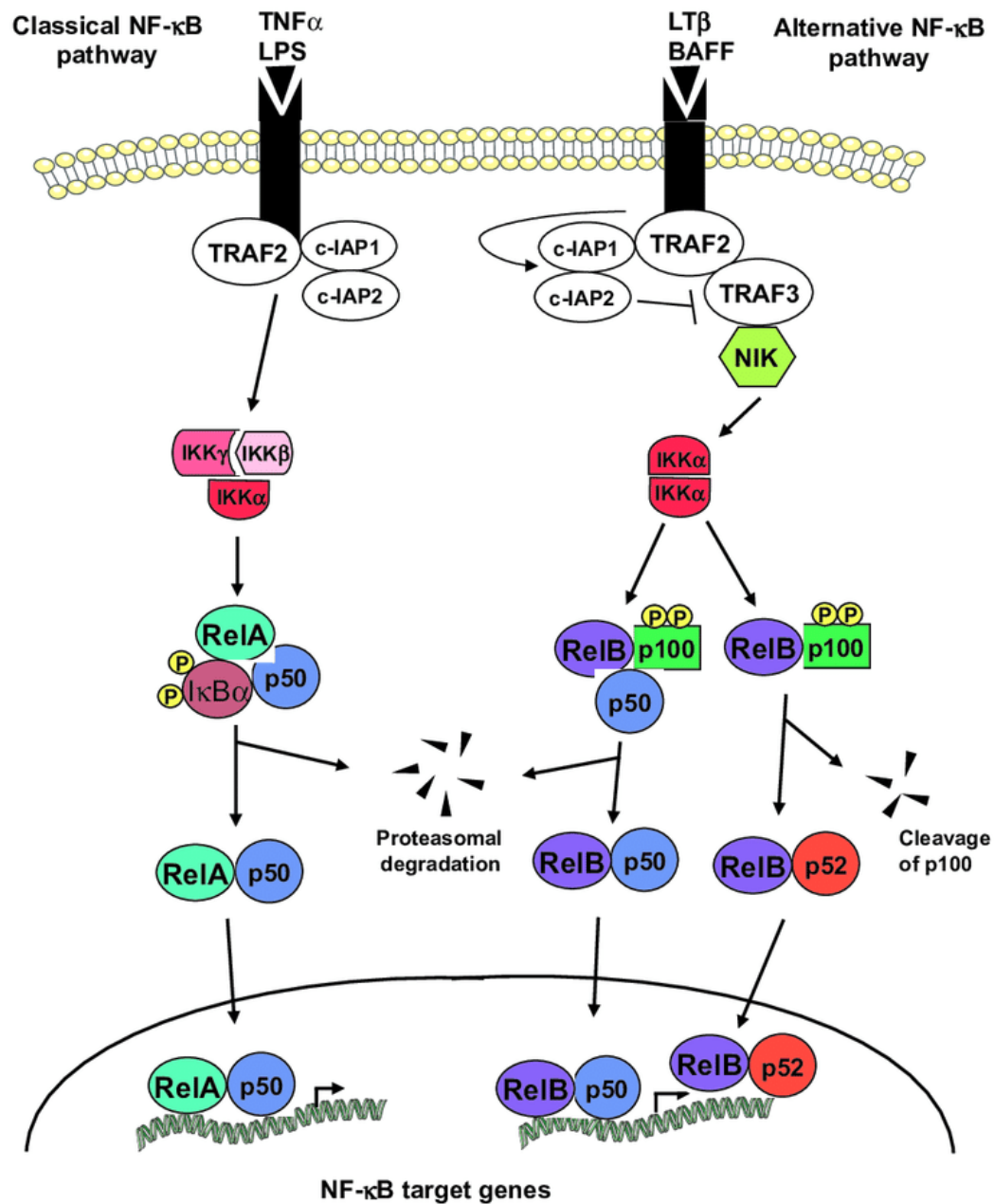


Figure 5: La voie classique et la voie alternative NF- κ B, Figure adaptée d'après Collares (2016).

2.3.2 Voie de signalisation RAS/RAF/MAPK

La famille des MAPK est composée de nombreux membres dont l'activation s'effectue par une série de phosphorylations. RAF est par exemple une MAP kinase tri-phosphorylée qui peut être activée par les membres de la famille RAS. Une cascade de phosphorylation est ensuite déclenchée: RAF va à son tour activer une MAP kinase di-phosphorylée qui va elle même activer une MAP kinase mono-phosphorylée (Figure 6).

Cette voie de signalisation est particulièrement critique pour la cellule car elle permet de réguler à la fois la différenciation, la prolifération, la survie, la réparation des dommages à l'ADN ou bien encore le cycle cellulaire.

Dans le cas du Myélome, les mutations sur cette voie de signalisation sont retrouvés dans près de 50% des patients et sont généralement associées à une prolifération cellulaire et donc à une progression de la maladie plutôt qu'à son apparition.

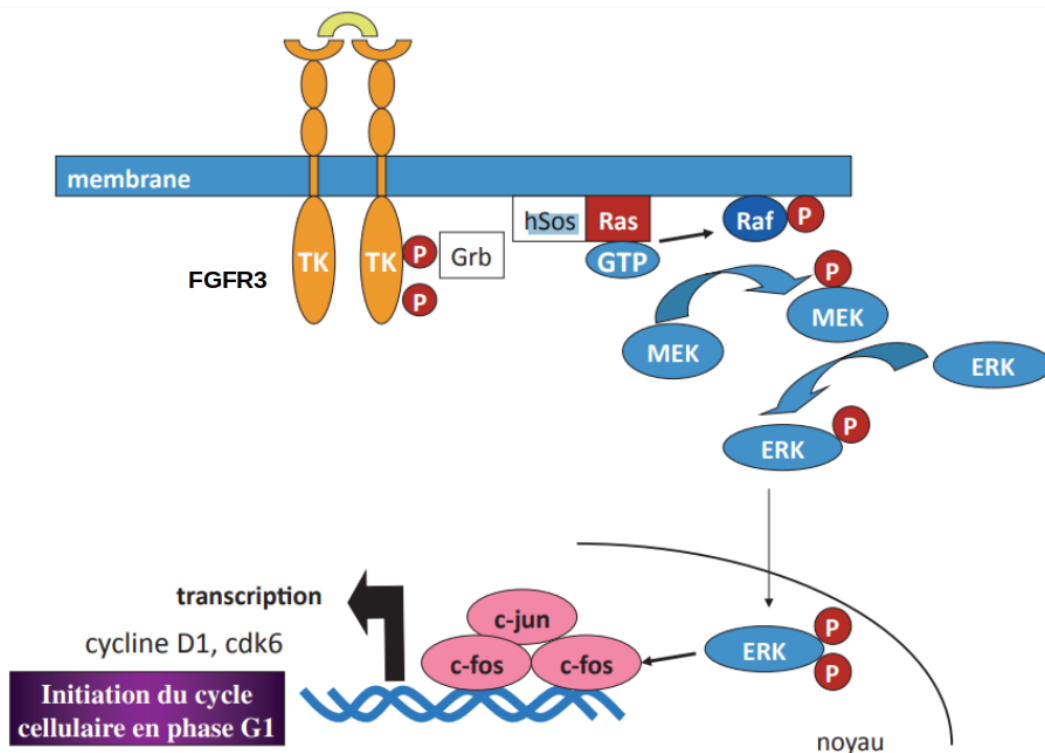


Figure 6: **La voie de signalisation RAS/MAPK** Figure adaptée d'après Lièvre and Laurent-Puig (2010).

2.3.3 La différenciation des cellules plasmocytaires

L'IRF4 est un facteur de transcription essentiel tout au long de la différenciation de la cellule B mais également pour la survie de la cellule de myélome.

L'expression d'IRF4 dans la lignée lymphocytaire B est plutôt biphasique. Dans les cellules B naïves, l'expression d'IRF4 est réprimée par l'intervention de MITF (Figure 7A). Lorsque la cellule B est activée, l'expression d'IRF4 va augmenter par l'intervention de deux voies de signalisation: les sous-unités NF- κ B et le facteur STAT6. IRF4 va alors lui même permettre l'expression de PRDM1 et de MYC. L'expression d'IRF4 est également régulée par un mécanisme de feed-back d'IRF4. L'expression de MYC va permettre un burst de division cellulaire tandis que PRDM1 va conduire à une différenciation d'une proportion des cellules en cellules à courte durée de vie sécrétrices d'Igs (Figure 7B).

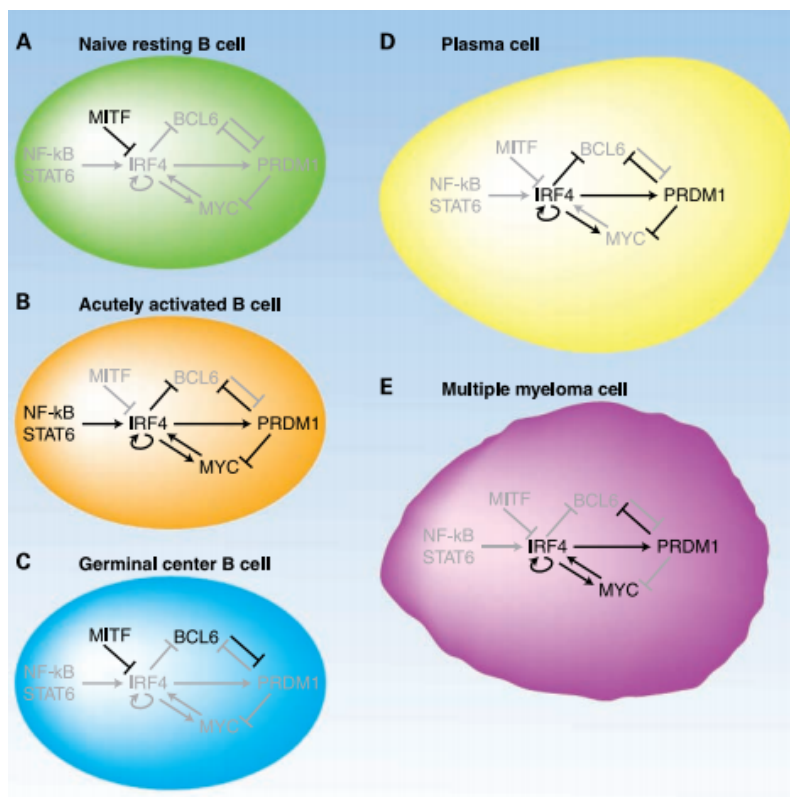


Figure 7: **Expression biphasique d'IRF4 selon le stade de différenciation de la cellule B.**, Figure adaptée de Shaffer et al. (2009).

IRF4 va de nouveau être réprimé dans les centres germinatifs. Les cellules expriment IRF8 et le niveau d'IRF4 diminue principalement du à l'absence de NF- κ B combinée à la répression par MITF. Ceci permet l'expression de BCL-6 qui va inhiber l'expression de PRDM1, ce qui bloque les cellules dans ce stade (Figure 7C).

Dans les plasmocytes, IRF4 est à nouveau exprimé grâce à l'intervention de différents stimuli. IRF4 réprime alors BCL-6 ce qui permet l'expression de PRDM1. PRDM1 réprime également MYC et BCL-6 amenant au phénotype de sécreteur d'Igs et termine la différenciation en cellule plasmocytaire (Figure 7D).

Dans la cellule de myélome, IRF4 est également exprimé ainsi que PRDM1 et MYC qui sont tous les deux sur-exprimés (Figure 7E). Les cellules de myélome sont dépendantes du niveau d'expression d'IRF4, une simple diminution entraîne une mort cellulaire. IRF4 contrôle un programme transcriptionnel aberrant qui contrôle l'expression de gènes importants pour le contrôle du cycle cellulaire, la régulation de la transcription et la différenciation du plasmocyte via PRDM1 (Figure 8). Cependant, les rares mutations oncogéniques sur le gène *IRF4* sont généralement associées à un bon pronostic.

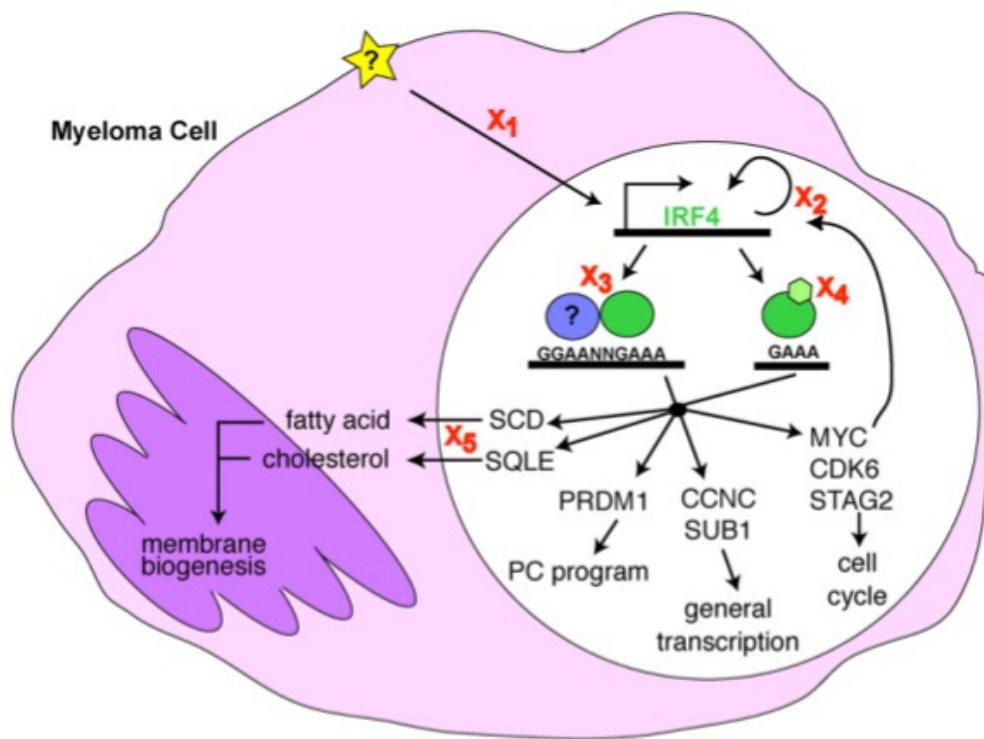


Figure 8: **Rôle d'IRF4 dans la survie et la prolifération de la cellule de myélome**, Figure adaptée de Shaffer et al. (2009); , les X rouges représentent les différentes cibles d'action de diverses drogues.

2.3.4 Autres voies perturbées dans le Myélome Multiple

Des mutations génétiques perturbant d'autres voies de signalisation ont également été répertoriées chez des patients atteints de MM. L'une de ces voies est celle de réponse aux dommages causés sur l'ADN: des mutations sur les gènes codants pour les kinases ATM et ATR et plus significativement sur le gène codant pour la TP53. Lors de cassures de l'ADN double brin, les kinases ATM et ATR vont activer la phosphorylation des protéines TP53, CHEK1 et de CHEK2 qui vont provoquer l'arrêt du cycle cellulaire, la réparation de l'ADN et l'apoptose de la cellule. Les mutations touchant la voie de réparation de l'ADN sont retrouvées chez environ 15% des patients. Ces mutations sont généralement associées à un mauvais pronostic chez les patients.

Une autre voie perturbée par des mutations sur les gènes est celle du cycle cellulaire. On retrouve en effet des mutations sur les gènes codant pour les Cyclines D1 et D3 (*CCND1* et *CCND3*), pour l'inhibiteur de la cycline D1 CDKN2C, la protéine RB1 et TP53. Comme indiqué précédemment, les translocations impactant *CCND1* sont retrouvées chez 15% des patients. Ces mutations impactent donc le cycle et peuvent à la fois provoquer l'arrêt du cycle et donc l'apoptose de la cellule ou au contraire la prolifération des cellules tumorales. Globalement, il y a toujours une cycline de perturbée dans le myélome: *CCND1* et *CCND3* par des translocations et des mutations et *CCND2* par les perturbations de c-MAF et MMSET.

2.4 Traitements du Myélome Multiple

La prise en charge du Myélome va dépendre de l'âge du patient mais également de son état général, des symptômes et du stade de la maladie. Si le patient ne présente pas de symptômes le traitement n'est pas systématique mais en revanche une surveillance accrue du patient et de l'évolution de sa maladie est mise en place.

L'un des principaux moyens de traitement du Myélome Multiple est l'autogreffe. Cependant, l'âge limite maximal pour une autogreffe est de 65 ans. Pour réaliser une autogreffe, un traitement dit d'induction de 4 cycles impliquant un inhibiteur du protéasome, un immunomodulateur et de la Dex est tout d'abord donné au patient dans le but de réduire de façon drastique la masse tumorale. Deux traitements d'induction sont en cours actuellement: le VRD (Bortezomib/Velcade, Lenalidomide/Revlimid, Dex) et le VTD (Bortezomib/Velcade, Thalidomide, Dex). Une fois les 4 cycles réalisés, les cellules souches hématopoïétiques périphériques du patient sont prélevées puis triées et congelées. Le patient va alors suivre une chimiothérapie additionnée de Melphalan qui aura pour effet de détruire les cellules de la moelle osseuse. Les cellules souches initialement prélevées et congelées sont ensuite réinjectées par transfusion 48 heures après pour initier une régénération de la moelle osseuse.

Suite à cette autogreffe, le patient recevra un traitement dit de consolidation pendant 2 mois après la greffe (correspondant à 2 cycles de traitement d'induction), puis sera sous immunomodulateur (lenalidomide) pendant 2 ans. L'autogreffe est également envisageable pour les patients de plus

de 65 ans ne présentant pas de problème cardiaque, rénal ou pulmonaire et ce jusqu'à 70 ans. Il faut également noter que des essais thérapeutiques qui ont pour objet les allogreffes, c'est-à-dire le prélèvement de cellules souches hématopoïétiques saines chez un donneur compatible, et l'injection après chimiothérapie chez le patient sont actuellement en cours.

Pour les patients plus âgés, le traitement consiste en une chimiothérapie combinée au Velcade (combinaison de Bortezomib/Velcade Melphalen et Prednisone = VMP) ou d'une combinaison de Revlimid/lenalidomide et Dex (RD). A noter que le RD est privilégié car son administration est plus facile pour le patient et nécessite moins de déplacements à l'hôpital.

Récemment, l'utilisation d'une nouvelle classe de médicaments, l'anticorps monoclonal daratumumab, en association avec les thalidomides (IMiD), le bortézomib (PI) et la Dex a amélioré la réponse et la survie sans rechute (Moreau et al., 2019). La Dex est également utilisée dans toutes les options de traitement pour les patients atteints de MM en rechute et/ou réfractaires (RRMM) en association avec les IMiD de deuxième génération, le lenalidomide et le pomalidomide et les IP carfilzomib et ixazomib (Gariani et al., 2018).

En parallèle de ces différents traitements, des approches pour traiter les symptômes du Myélome peuvent être utilisées telles que la radiothérapie pour soulager les douleurs osseuses ou des traitements médicamenteux pour lutter contre les infections, des dialyses pour les insuffisances rénales ou des compléments alimentaires pour compenser l'anémie.

2.5 Rechute du Myélome Multiple

Malgré l'émergence de ces nouveaux traitements, le Myélome Multiple reste une pathologie incurable et les patients présentent une durée de rechute allant de quelques mois à environ 10 ans. Cette rechute survient notamment par l'émergence d'une sous-population minoritaire qui est plus agressive et plus résistante aux traitements (Mateos and San Miguel, 2013).

Dans ces cas, des traitements dits de seconde ligne sont alors utilisés notamment deux types: l'utilisation d'anticorps anti-CD38 (Daratumumab, Isatuximab) ou des inhibiteurs du protéasome (Carfilzomib, Ixazomib). Malheureusement, ces différents traitements perdent drastiquement en efficacité au fur et à mesure des rechutes successives du patient.

3 Mécanismes de la régulation d'expression des gènes

Les nouvelles technologies de séquençage ont permis récemment une meilleure approche et compréhension du génome dans son ensemble. Ces approches ont pu ainsi mettre en évidence que le niveau de complexité d'un organisme n'est pas lié à son nombre de gènes ni à la taille de son génome. Par exemple, le génome d'un grain de riz est composé de 35 000 à 40 000 gènes quand seulement environ 20 000 gènes sont décrits pour l'être humain. La diversité et la complexité d'un organisme ne peuvent donc pas être uniquement attribuées à ces critères et d'autres paramètres tels que les modifications protéiques, épigénétiques ou bien encore les divers mécanismes d'épissage doivent être pris en compte pour appréhender la complexité d'un organisme. Ainsi, l'hypothèse de travail émise par Crick "un gène donne un acide ribonucléique (ARN) qui donne une protéine" a depuis longtemps évolué en fonction de l'organisme, du tissu et même de l'état de la cellule à un instant t . Aujourd'hui on peut affirmer qu'un gène sera transcrit en plusieurs ARNs qui pourront être traduits en plusieurs protéines qui selon les différentes modifications qu'elles subiront pourront jouer divers rôles pour la cellule.

De plus, les données de séquençage du génome entier ont révélé que sur l'ensemble du génome humain, la proportion de séquences dites "codantes" pour des protéines ne représentait qu'environ 1% du génome entier. Parmi le reste des régions dites "non codantes", on trouve des séquences permettant la transcription d'ARNs non codants, des séquences introniques mais aussi des séquences dites régulatrices plus ou moins proches des séquences codantes (Knight, 2003). Ces séquences sont appelées promoteurs lorsqu'elles se situent proche d'un transcription start site (TSS) ou bien enhancer (proximal ou distal) lorsque celles-ci sont situées à distance d'un TSS. Ces séquences régulatrices représenteraient à elles seules environ 40% du génome et leur utilisation en tant que régions régulatrices varierait entre deux tissus et même entre deux cellules d'un même tissu ce qui confère à la cellule une modulation de l'expression de ses gènes beaucoup plus fine. Ces différents mécanismes de régulation de l'expression des gènes à un niveau génomique sont donc un élément indispensable pour la compréhension de l'action moléculaire de la Dex dans le Myélocyte.

3.1 Transcription des gènes

Au cours des années, les connaissances sur la façon dont les gènes sont transcrits ont grandement évolué. Un gène est défini par une séquence d'ADN lue par un complexe protéique appelé machinerie transcriptionnelle. Cette machinerie lit la séquence d'ADN pour la retranscrire en ARN dit "messenger" (ARNm) qui sera ensuite exporté au niveau du cytoplasme pour être traduit en protéine. Cette transcription est effectuée par la RNA polymérase II. Une fois l'ADN transcrit en pré-ARNm, celui-ci va subir de nombreuses modifications dont la principale est l'épissage alternatif qui permet la production de différents transcrits à partir d'un seul et même gène (Lee and Rio, 2015).

L'épissage alternatif permet de définir deux types d'éléments présents dans la séquence du gène:

les introns et les exons. Les exons contiennent des parties codantes du gène et seront épissés pour composer l'ARN transcrit tandis que les introns sont des séquences non codantes qui séparent les exons. Lors de la maturation du pré-ARNm en ARNm, l'épissage permet d'éliminer les parties introniques et d'assembler des exons permettant la production de différents transcrits en fonction des besoins de la cellule. Une fois l'ARNm maturé, celui-ci est exporté dans le cytoplasme où il est traduit en protéine grâce à l'action des ribosomes. Ces protéines subissent ensuite toutes sortes de modifications post-traductionnelles (phosphorylation, ubiquitination, repliement...) afin de devenir fonctionnelles pour la cellule. Bien que ces mécanismes soient maintenant bien décrits, les séquences codantes de l'ADN ne représentent qu'une infime part de l'ensemble du génome. Parmi les régions non codantes, de nombreuses études ont démontré la présence d'éléments régulateurs de la transcription.

3.2 Régions régulatrices de la transcription et facteurs de transcription

Pour permettre l'initiation de la transcription des gènes, des éléments de régulation sont nécessaires. Ils sont majoritairement de deux types: les éléments de régulation très proches des gènes dits promoteurs et les éléments de régulation proximaux ou distaux dits enhancers. Ces éléments sont trouvés dans les régions de l'ADN dites non codantes.

Les promoteurs sont des séquences régulatrices généralement proches des TSS, ces séquences sont très souvent riches en bases GC (environ 70% des promoteurs humains). Les promoteurs sont indispensables pour la transcription du gène en ARNm, en effet, des études *in vitro* ont montré qu'une déplétion de ces éléments sur l'ADN inhibait l'expression du gène. Les promoteurs permettent le recrutement de la machinerie transcriptionnelle via des séquences de reconnaissance tels que la TATA box (5'-TATATA-3') qui est reconnue par la RNA polymérase II, ces séquences sont toutefois rares chez l'humain. Une fois la machinerie recrutée au niveau des promoteurs, la transcription peut être initiée et l'ADN transcrit. D'autres éléments de régulation peuvent également être trouvés dans les promoteurs comme la boîte CAT et boîte GC (5'-GGCCCATCCAT-3') (Govindan et al., 1991) qui permettent la fixation d'autres protéines servant au recrutement de la machinerie transcriptionnelle.

Les enhancers sont quant à eux des éléments de régulation de la transcription des gènes distaux. Ils peuvent être localisés à plusieurs centaines de kb du gène cible qu'ils vont réguler. Ces éléments contiennent des motifs de reconnaissance pour la liaison de protéines spécifiques à l'ADN. De nombreux enhancers peuvent réguler un même gène cible, de même un enhancer peut réguler différentes cibles selon l'environnement ou même le tissu. De façon générale, les enhancers agissent de façon synergique avec le promoteur du gène cible afin de booster ou au contraire inhiber la transcription du gène, on parle dans ce cas de silencers.

Outre la présence de la machinerie transcriptionnelle au niveau des promoteurs, il y a également de nombreuses séquences de reconnaissance protéique dans les promoteurs ainsi que dans les enhancers. Ces séquences permettent la liaison à l'ADN de protéines régulatrices que l'on appelle facteur de transcription (TF). Il existe de nombreuses familles de facteur de transcription comme

par exemple la famille IRF qui reconnaît la séquence ISRE (5'-AGTTTCNNTTTCNC/T-3'). L'activité et l'expression du facteur de transcription sont conditionnées par l'état de la cellule, le tissu ou le stade de différenciation de la cellule. Les TFs peuvent avoir une activité activatrice ou inhibitrice et dans certains cas les deux. Ainsi un facteur de transcription peut à la fois activer un gène A et en même temps inhiber un gène B.

3.3 Modifications post-traductionnelles des histones

L'ADN est une séquence nucléotidique non linéaire, cette chaîne est enroulée autour de complexes protéiques appelés octamère d'histones formant ainsi le nucléosome. Il existe plusieurs sortes d'histones: H3, H4, H2A et H2B. Chaque octamère d'histones est composé de deux exemplaires de chacun des quatre histones. Chaque histone contient deux domaines majeurs: un domaine central qui contient le "motif histone fold" composé de trois hélices et de deux boucles et une "queue" N-terminale dépourvue de structure secondaire et riche en résidus lysine et arginine (Figure 9A). Du fait de la composition de ces queues, les résidus qui la composent sont la cible de nombreuses

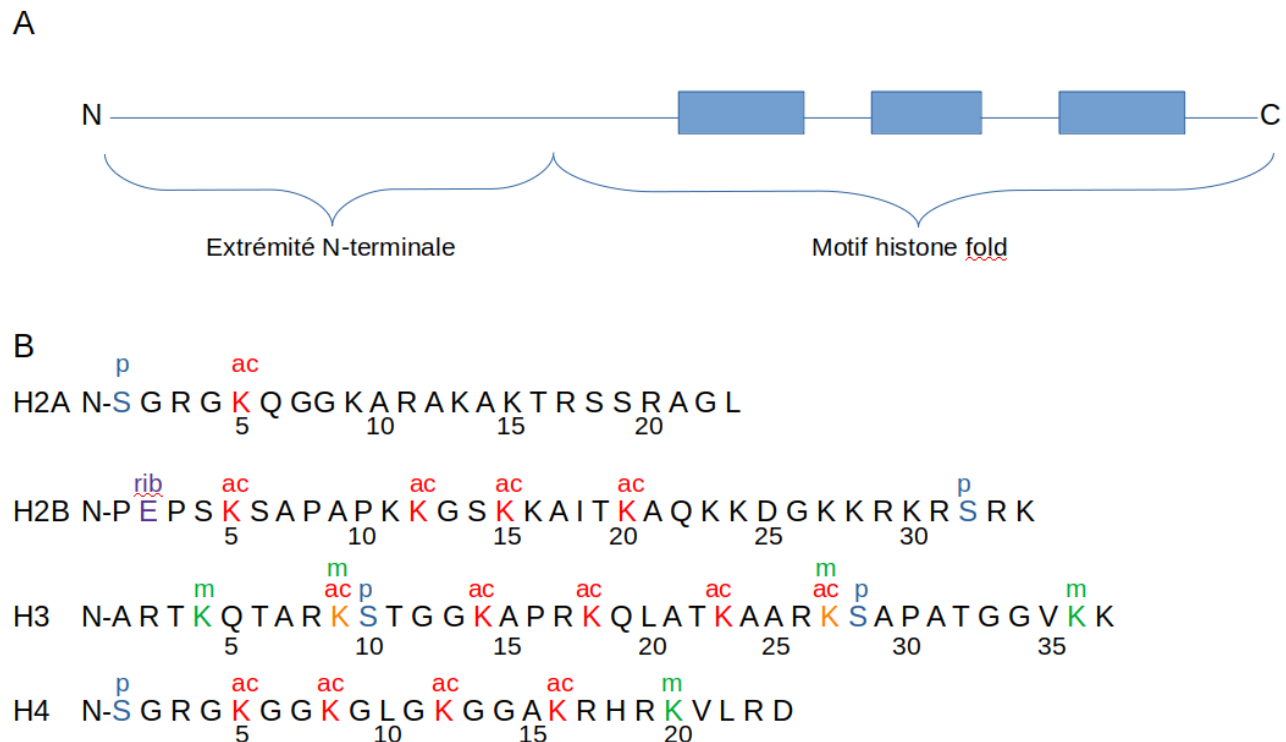


Figure 9: **Domaines et composition des sous-unités histones H2A et B, H3 et H4**
 (A) Domaines structuraux des sous-unités histones (B) Composition en acides aminés des queues N-terminales et positions des résidus pouvant subir des modifications post-traductionnelles; p: phosphorylation, ac: acétylation, rib: ribosylation, m: méthylation.

modifications post-traductionnelles (Hake et al., 2006). Ces modifications sont de plusieurs sortes: phosphorylation, acétylation, méthylation... Ces modifications peuvent donc affecter la charge des histones mais également l'accessibilité à l'ADN ainsi que les interactions protéines:protéines avec les nucléosomes (Figure 9B).

Les modifications post-traductionnelles au niveau de ces queues d'histones permettent de prédire l'activité transcriptionnelle d'une région régulatrice. Ainsi de nombreuses modifications appelées marques d'histones ont été identifiées. La combinaison de la présence de certaines marques permet ainsi de définir des états chromatinien dont va pouvoir dépendre la régulation de la transcription (Zhang et al., 2015).

Pour évaluer l'activité transcriptionnelle des régions régulatrices promoteurs et enhanceurs, ce sont les modifications sur les résidus lysine 4 et lysine 27 de l'histone H3 qui sont principalement utilisées (Ernst and Kellis, 2012). L'histone H3 peut être mono ou tri-méthylé sur la lysine 4 (H3K4me1 ou H3K4me3) et ainsi définir une région de type enhancer (H3K4me1) et une région de type promoteur (H3K4me3). Concernant l'activité transcriptionnelle de ces régions, c'est le résidu lysine 27 qui va être prépondérant (Figure 10). Sur ce résidu, une tri-méthylation (H3K27me3) est généralement associée à une répression ou une inhibition de la transcription (Figure 10A) alors qu'une acétylation (H3K27ac) est, à l'inverse, synonyme d'activation de la transcription (Figure 10).

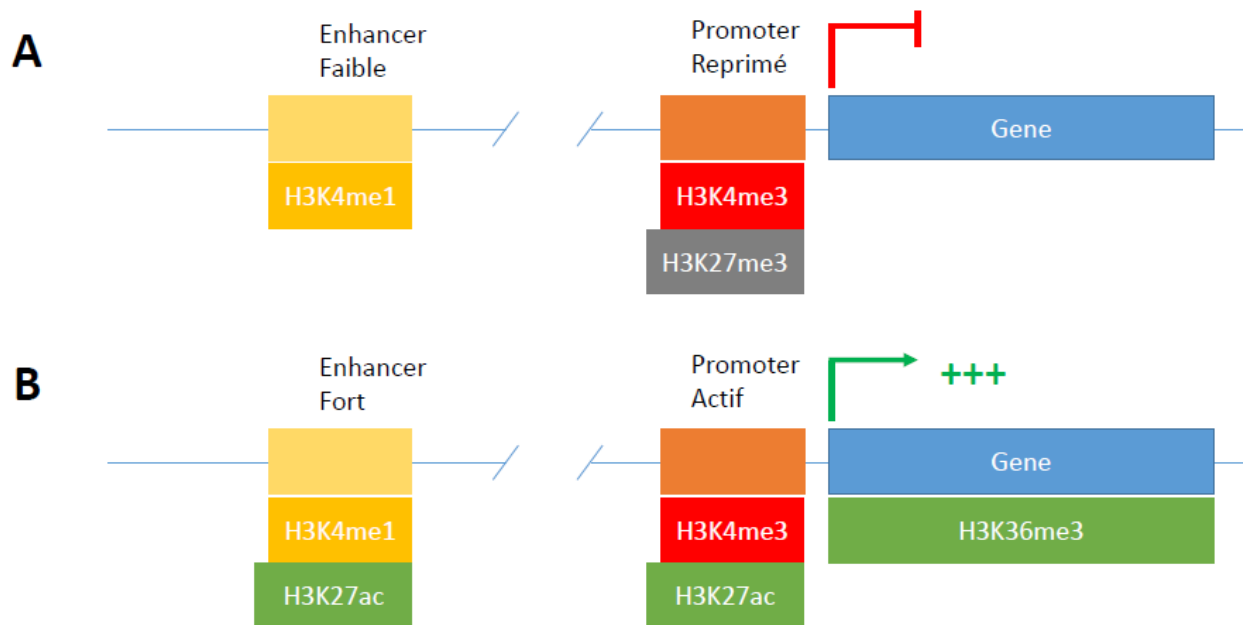


Figure 10: **Modulation de l'expression génique selon la présence ou l'absence de diverses marques d'histones** Marques d'histones fréquemment enrichies dans les régions régulatrices des gènes réprimés ou peu exprimés (H3K27me3) (A) et pour les gènes transcriptionnellement actifs (H3K27ac) (B).

3.4 Accessibilité de la chromatine

L'ADN n'existe pas sous forme libre dans la cellule, il est compacté à plusieurs niveaux, le niveau le plus élevé étant la super condensation donnant lieu aux chromosomes. A l'état enroulé l'ADN est appelé chromatine et peut être divisé en deux domaines: l'euchromatine et l'hétérochromatine (Figure 11B). L'euchromatine est le compartiment de l'ADN qui est décondensé pendant l'interphase du cycle cellulaire. L'hétérochromatine est, à l'inverse, un état de condensation qui ne change pas pendant le cycle cellulaire et est généralement localisé en périphérie du noyau alors que l'euchromatine occupe plutôt la partie centrale du nucléoplasme.

L'unité fondamentale de la chromatine est appelée nucléosome. Composé à la fois d'ADN et d'histones, le nucléosome constitue le premier niveau de compaction de l'ADN. Le nucléosome est composé d'une particule coeur (histone) qui est très conservée parmi les espèces et est constituée de 146pb d'ADN enroulé environ 1,7 tour autour d'un octamère d'histones (Figure 11A). Le nucléosome est également composé d'une région de liaison (région internucléosomale) qui relie les particules coeur dont la longueur varie selon les espèces. La distance internucléosomale varie généralement entre 160 et 240bp.

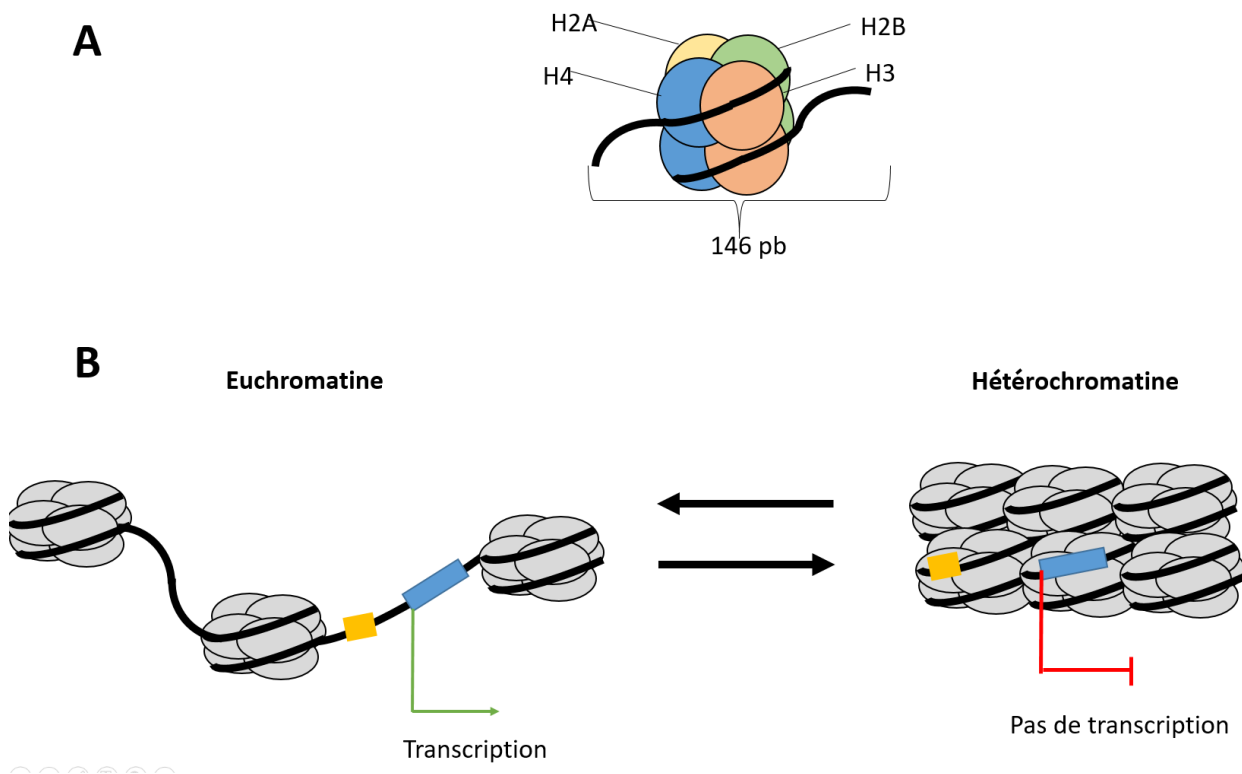


Figure 11: **Compaction et accessibilité de la chromatine.**(A) Composition du nucléosome avec ses différentes sous-unités. (B) L'accessibilité de la chromatine et le niveau de condensation de celle-ci permettent ou non la transcription des gènes cibles.

Selon l'état de la cellule à un instant t , les nucléosomes ne seront pas situés exactement à la même localisation dans le génome. Il existe des complexes de remodelage de la chromatine qui permettent de modifier la position des nucléosomes. Une des famille de facteurs de remodelage la plus connu est la famille SWI/SNF. Ces complexes comportent 8 à 15 sous-unités dont sept composent une sous-unité ATP-ase indispensable pour modifier la position de la particule coeur des nucléosomes. Ces complexes permettent d'augmenter l'accessibilité de l'ADN aux divers facteurs de transcription et aux nucléases. Le mécanisme de remodelage consiste généralement à faire "glisser" les histones sur l'ADN.

L'accessibilité de la chromatine est donc un élément essentiel dans les mécanismes de régulation de la transcription des gènes. Les gènes ainsi que les diverses régions régulatrices promoteurs et enhancers doivent en effet être accessibles pour permettre la fixation des facteurs de transcription et de la machinerie transcriptionnelle.

3.5 Conformation 3D de la chromatine

Le développement de nouvelles techniques de séquençage et particulièrement les techniques de HiC (Lieberman-Aiden et al., 2009) ont permis de mieux comprendre l'impact de l'organisation spatiale du génome sur la proximité physique des différents éléments de régulation de la transcription. Ces études ont montré que l'organisation du génome était un processus en plusieurs étapes impliquant la compaction de l'ADN en nucléosome, les fibres chromatinienues et la compartimentation des chromosomes. A des échelles inférieures à la mégabase, les fibres chromatinienues sont repliées en domaines globulaires appelés domaines topologiquement associés (TAD) (Figure 12). Ces domaines TAD se regroupent alors en deux compartiments principaux A et B caractérisés par un enrichissement de gènes transcriptionnellement actifs et de chromatine accessible pour le compartiment A et à l'inverse, de gènes éloignés les uns des autres et de chromatine inaccessible pour le compartiment B. Les TAD sont conservés et largement invariants à travers les types cellulaires, alors que les interactions chromatinienues à l'intérieur des TAD et les compartiments sont beaucoup plus tissus spécifiques. Une fonction importante des TAD est de restreindre les interactions d'éléments régulateurs d'un gène à l'intérieur d'un même TAD en évitant des interactions avec un TAD adjacent. Les frontières des TAD permettent également de restreindre la propagation de la chromatine répressive à l'intérieur des domaines actifs et inversement. Ces frontières sont enrichies en fixation de protéines structurales telles que le CTCF ou la cohésine. Les interactions médiées par la cohésine et par les boucles CTCF-CTCF sont essentielles pour l'organisation du génome en TAD. Cette organisation hiérarchique permet à la cellule d'avoir un profil d'expression tissu spécifique.

Enfin, au sein d'un même domaine de régulation, des boucles de régulation se forment pour permettre l'interaction physique d'un enhancer avec le promoteur d'un gène cible. Ainsi un enhancer peut réguler un gène à des centaines de kilobases de sa cible sur le même chromosome (interaction *cis*) ou sur un autre chromosome (interaction *trans*). La stabilisation de ces boucles dépend du facteur de transcription qui se fixe dans la région ainsi que de l'état cellulaire.

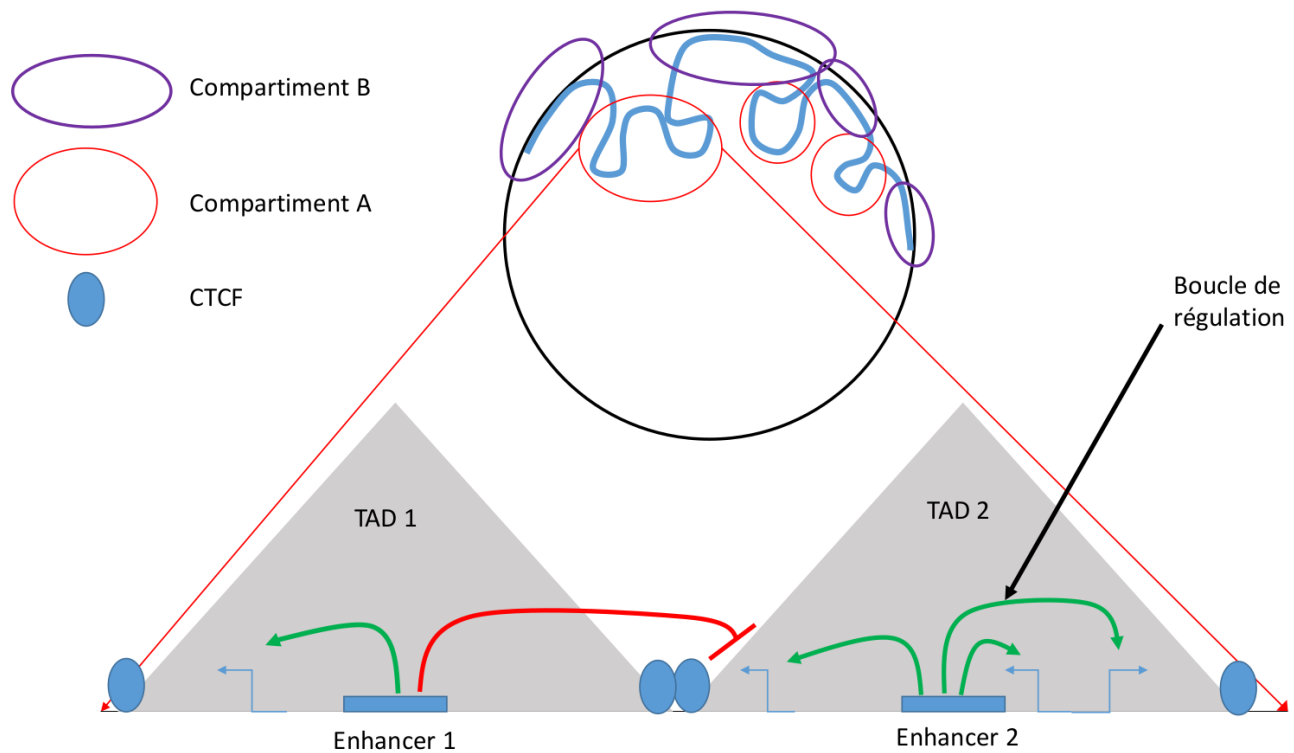


Figure 12: **Organisation structurale de la chromatine à différents niveaux**, Les deux compartiments représentent la chromatine libre et compactée, les domaines TAD sont formés par le CTCF. Les boucles de régulation se forment à l'intérieur de ces domaines.

4 Les Glucocorticoïdes

4.1 Rôle physiologique du Récepteur aux Glucocorticoïdes

A l'état physiologique, les GCs sont des hormones stéroïdes qui sont secrétées par les glandes adrénales principalement en réponse à un stress de l'organisme (Cruz-Topete and Cidlowski, 2015). Une fois secrété, le glucocorticoïde se fixe sur leur récepteur spécifique: le récepteur aux glucocorticoïdes (GR). Ainsi liés, ils agissent sur un grand nombre de métabolismes (métabolismes glucidique, lipidique et/ou protéique) mais également sur les éléments figurés du sang tels que les lymphocytes B, les mastocytes (activité anti-allergique) et les macrophages (diminution de l'activité anti-infectieuse) (Figure 13). Nous détaillerons d'abord le rôle des GCs dans divers métabolismes puis dans la médiation de l'inflammation avant d'évoquer son utilisation dans les pathologies.

4.1.1 Métabolisme

Les GCs interviennent dans de nombreux organes pour réguler des fonctions métaboliques importantes dans le corps humain (Cruz-Topete and Cidlowski (2015), Cohen and Steger (2017)):

- **Dans le foie:** Les GCs régulent divers processus à l'intérieur du foie. Des niveaux aberrants de GCs peuvent provoquer des hyperglycémies par induction de gènes impliqués dans la production d'enzyme gluconéogénique (G6Pase et PEPCK). Les GCs stimulent également la production de very low density protein (VLDL) conduisant à des troubles hypertriglycéridémiques. De plus, les GCs favorisent la sécrétion d'enzymes FAS et d'acetylCoA carboxylase (ACC) et à l'inverse, inhibent la β -oxydation d'acides gras libres (FFA) par interférence des GCs avec l'AcylCoA déshydrogénase ce qui peut conduire à des stéatoses hépatiques (maladie du foie gras). Finalement, les GCs en agissant sur les VLDL et les enzymes FAS et ACC contribuent à la lipogénèse dans le foie.
- **Dans les tissus adipeux:** Les GCs agissent dans les tissus adipeux mais de façon antagoniste selon la localisation (périphérique ou centrale/abdominale). Dans les tissus adipeux périphériques, les GCs provoquent la lipolyse et inhibent l'enzyme PEPCK et la lipoprotéine lipase (LPL). A l'inverse, dans les tissus centraux ou abdominaux, les GCs provoquent une hypertrophie des tissus par interaction avec les protéines C/EBP et PPAR γ . Des anomalies liées aux GCs dans les tissus adipeux peuvent provoquer de l'obésité et une résistance à l'insuline.
- **Dans les muscles squelettiques:** Les GCs permettent la stimulation de la myostatine qui inhibe la croissance des tissus musculaires. Les GCs régulent de plus l'utilisation du glucose et le renouvellement des protéines au travers de nombreuses voies: augmentation de la glutamine synthase qui favorise la mobilisation d'acides aminés; inhibition de la kinase AKT qui

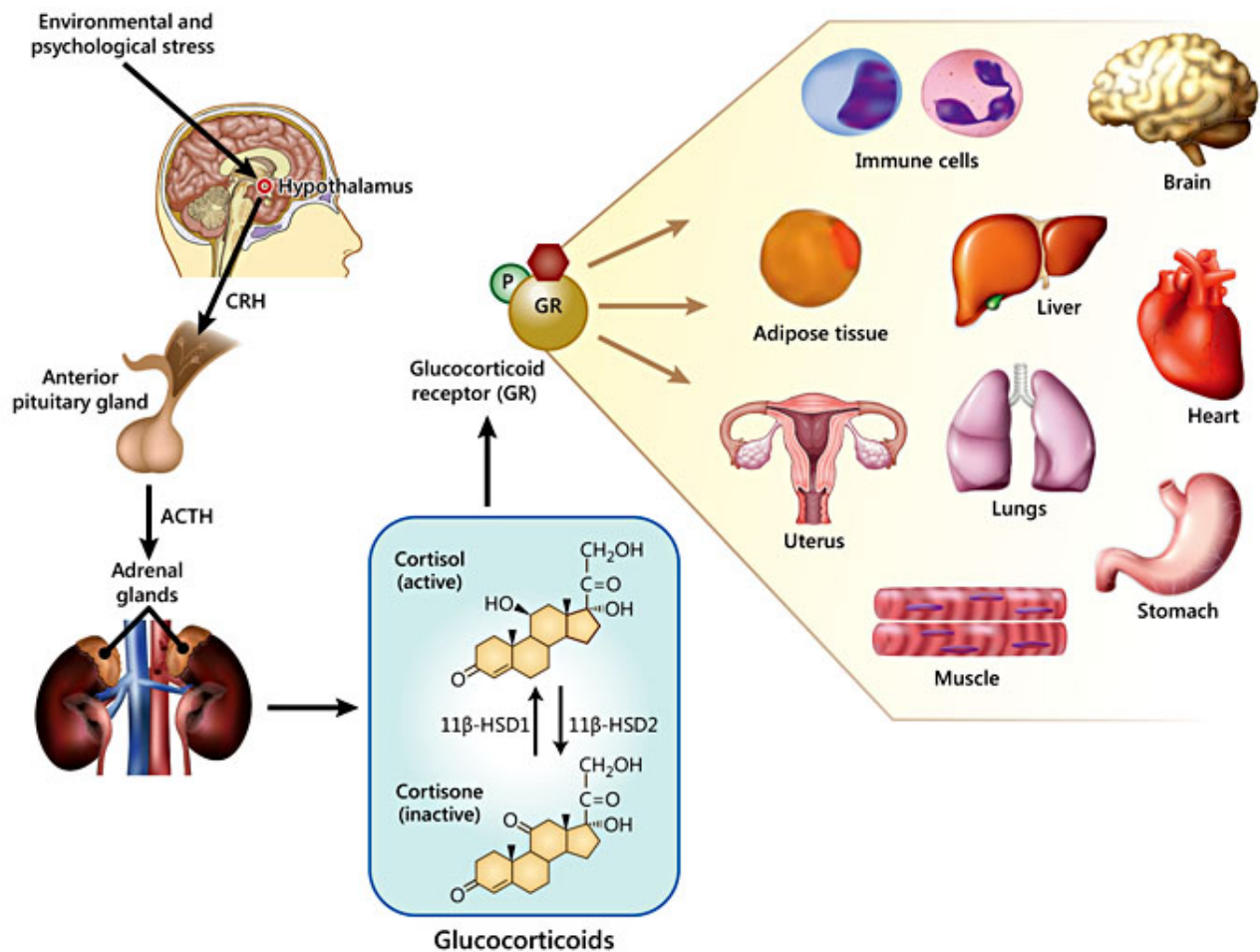


Figure 13: Les GCs régulent un très grand nombre de fonctions à l'état physiologique chez l'homme. L'hormone est synthétisée par les glandes adrénales pour agir via son récepteur. Figure tirée de Cruz-Topete and Cidlowski (2015).

inhibe l'assimilation du glucose (par inhibition de la translocation de la protéine GLUT4), la synthèse de glycogène (par inhibition de l'enzyme Glycogène synthase) et d'autres protéines. Les GCs stimulent les protéines FOXO1 et FOXO3 qui favorisent la dégradation de certaines protéines.

4.1.2 Inflammation

Les GCs jouent un rôle important dans la régulation de nombreuses voies immunitaires à travers l'interaction avec des voies de signalisation telles que la voie NF- κ B (Franchimont et al., 2002).

- **Rôle dans l'immunité innée:** Les GCs permettent la prolifération et la survie des

neutrophiles et à l'inverse induisent l'apoptose des éosinophiles et des basophiles. L'action des GCs sur les neutrophiles est permise par la down régulation de molécules d'adhésion des leucocytes (LFA-1 et L-sélectine) et de molécules d'adhésion endothéliale comme ICAM1. Les GCs inhibent également la sécrétion de chemokines. Concernant les éosinophiles, les GCs empêchent la migration de ces cellules en inhibant la molécule ICAM63 et en stimulant la sécrétion de chemokines telles que l'éotaxine.

- **Rôle dans l'immunité adaptative:** Les GCs bloquent la maturation des cellules dendritiques qui perdent ainsi leur fonction de présentation de l'antigène aux lymphocytes T, ce qui inhibe la capacité à générer une réponse immunitaire. Cette inhibition est due au blocage du complexe d'histocompatibilité MHC-II et des molécules co-stimulatrices telles que ICAM-1 et LFA-1 par les GCs. Les GCs bloquent également la sécrétion d'IL-12 par les monocytes, les macrophages et les cellules dendritiques ce qui inhibe le développement des cellules Th1 (diminution de l'immunité cellulaire) et à l'inverse, favorise le développement des cellules Th2 par activation de la sécrétion d'IL-10 par les macrophages. Les Th2 stimulent l'immunité humorale par sécrétion de diverses Interleukines (IL-4, IL-5, IL-10 et IL-13). Les GCs favoriseraient également le développement des cellules T régulatrices. Finalement, les GCs agissent en synergie avec l'IL-4 pour activer la différenciation des cellules B et permettre la sécrétion d'IgE et IgG4.

4.1.3 Utilisation dans diverses pathologies

De par leur capacité à induire l'apoptose des lymphocytes et leur activité anti-inflammatoire, les GCs, en particulier la dexaméthasone (Dex), une hormone de synthèse, sont utilisés dans le traitement de nombreuses hémopathies malignes et maladies infectieuses.

Leur utilisation a récemment été mise en exergue dans le traitement du SARS-COVID19. L'essai de grande ampleur RECOVERY (Group, 2020) a en effet révélé que la Dex permettait de réduire d'environ un tiers les décès chez les patients sous respiration artificielle et d'un cinquième les décès chez les patients recevant de l'oxygène ce qui en fait à l'heure actuelle l'un des traitements les plus prometteurs du SARS-COVID19.

Dans le cas des hémopathies malignes, la Dex est utilisée dans le traitement des leucémies et la résistance à cette molécule est de mauvais pronostic. En ce qui concerne le myélome, la Dex est utilisée en traitement de première ligne en combinaison avec d'autres molécules pour diminuer de façon drastique la masse tumorale puis est administrée à chaque étape du traitement.

Des études montrent que la réponse à la Dex est hétérogène en fonction des patients (Kervoëlen et al., 2015). Cette hétérogénéité peut en partie être expliquée par le niveau d'expression du GR. Les patients t(14,16) et t(4,14) surexprimant plus fortement le récepteur.

Cette étude démontre également une hétérogénéité inter-patient de réponse à la Dex car seulement 50% des patients atteints de myélome répondent à la Dex. De même, cette étude démontre que dans 31 lignées cellulaires de myélome représentant les différentes anomalies génétiques répertoriées, le niveau d'expression du GR semble influencer sur la mort de la cellule induite par cette molécule.

4.2 Mode d'action du récepteur aux glucocorticoïdes

4.2.1 Composition et structure du récepteur aux glucocorticoïdes

Pour exercer ses nombreuses fonctions au sein de l'organisme, le GC se fixe sur son récepteur: GR. Le gène codant pour ce récepteur, appelé *NR3C1* est localisé sur le chromosome 5. Il existe actuellement 5 isoformes du GR: le GR α qui est prédominant, le GR β (dominant négatif) et trois autres isoformes beaucoup moins caractérisés et abondants (GR γ , A et P).

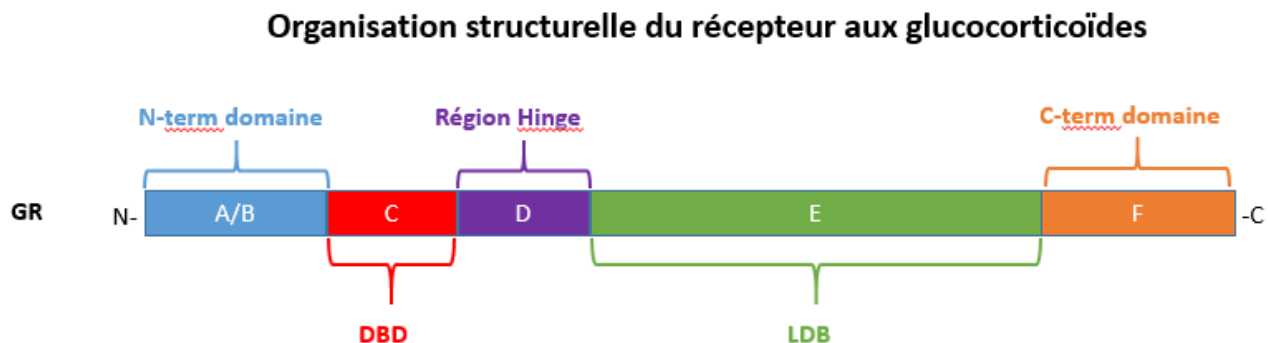


Figure 14: **Composition du récepteur nucléaire aux glucocorticoïdes.**

Le GR est composé de 6 domaines fonctionnels: les domaines A et B (N-terminaux) qui permettent l'initialisation et la régulation de la transcription des gènes cibles, le domaine de liaison à l'ADN (DBD) permettant la liaison à l'ADN via des structures en zinc finger et la dimérisation de la protéine, une région charnière qui relie les domaines fonctionnels et DBD aux deux domaines de liaison de l'hormone participant à la dimérisation du récepteur. A l'état basal, le GR est localisé dans le cytoplasme et est complexé avec des protéines chaperonnes telles que HSP90/70 et la p23 ou encore FKBP51 et 52. La sécrétion de GCs ou l'administration de la Dex va libérer le GR de son complexe pour permettre l'interaction avec son ligand.

4.2.2 Génomique appliquée au GR

A la suite de l'interaction avec son ligand, le GR s'homodimérise avant d'être internalisé dans le noyau de la cellule où il agit comme un facteur de transcription en se fixant sur l'ADN et régule ainsi la transcription de ses gènes cibles (Cohen and Steger (2017), Ratman et al. (2013), Reddy et al. (2009)) (Figure 15).

Il peut se fixer de façon directe à l'ADN grâce à la présence d'une séquence nucléotidique de reconnaissance (GRE) et grâce au domaine de liaison à l'ADN contenu dans la protéine (Figure 16A).

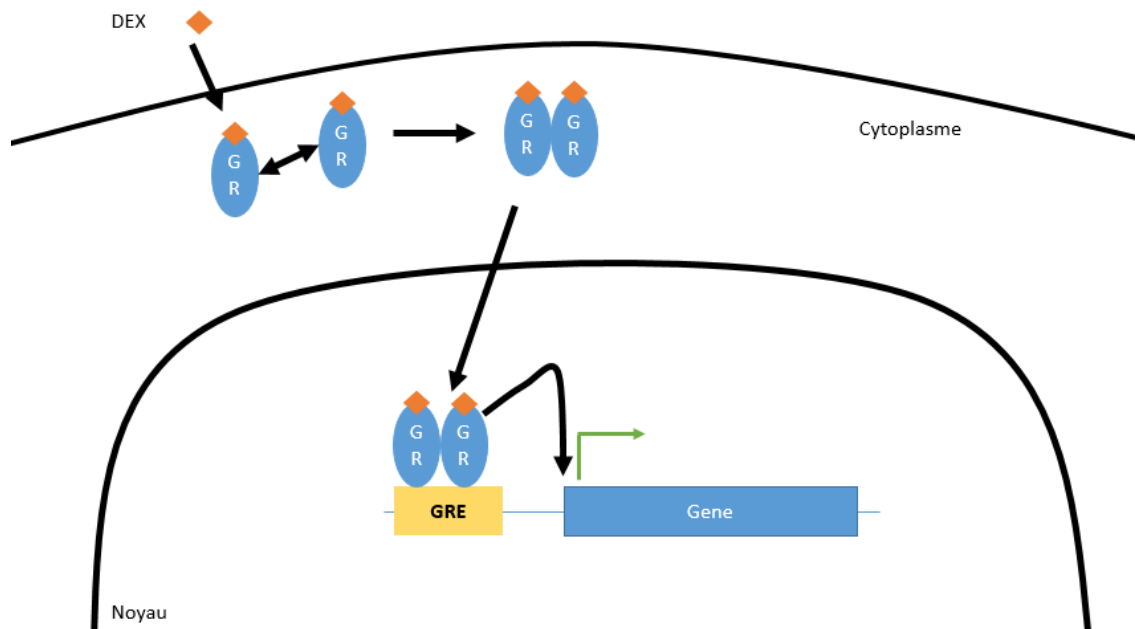


Figure 15: **Mécanismes d'action moléculaire du récepteur au glucocorticoïde sur la transcription des gènes.** L'hormone dans la cellule se lie au GR ce qui permet la dimérisation de celui-ci et son export dans le noyau. La fixation sur sa séquence consensus (GRE) permet l'activation de la transcription de ses gènes cibles.

La séquence GRE est une séquence de reconnaissance des facteurs de la famille des récepteurs nucléaires: 5'-CCTGCAnnnTGTTCT-3'. Certaines études démontrent que la fixation du GR sur cette séquence requiert la présence d'un homo-dimère GR mais que la fixation d'un monomère sur un demi-motif GRE est également possible même si les impacts sur la transcription de gènes cibles sont encore à ce jour méconnus (Figure 16B).

L'autre moyen de liaison à l'ADN est la fixation indirecte. En effet, le GR peut réguler la transcription de gènes en étant fixé à l'ADN via un facteur de transcription déjà présent et fixé à l'ADN (Figure 16C). Ce facteur est très souvent tissu voire cellule spécifique et maintient la région régulatrice ouverte et accessible pour la fixation du GR. Dans ce cas là, la présence de la séquence GRE n'est pas indispensable dans la région régulatrice. Il a été montré dans d'autres systèmes que le GR peut agir de cette façon sur les facteurs de transcription AP-1 ou NF- κ B. Les effets sur la transcription des gènes cibles selon le type de liaison à l'ADN ne sont pas à ce jour complètement élucidés. Certaines études décrivent qu'une liaison directe à l'ADN permettrait plutôt une activation de la transcription alors qu'une liaison indirecte jouerait un rôle antagoniste par rapport au facteur de transcription déjà présent. Enfin, de récentes études évoquent un autre mécanisme de régulation de la transcription via la fixation directe sur un motif dit GRE négatif (nGRE) (Figure 16D) qui favoriserait, à l'inverse du GRE, l'inhibition de la transcription du gène cible (Hudson et al. (2013), Surjit et al. (2011), Ratman et al. (2013)).

Des études démontrent que le motif n'est pas le seul élément déterminant la fixation du motif sur un site GRE mais que le landscape chromatinien est également prédéterminant pour cette fixation

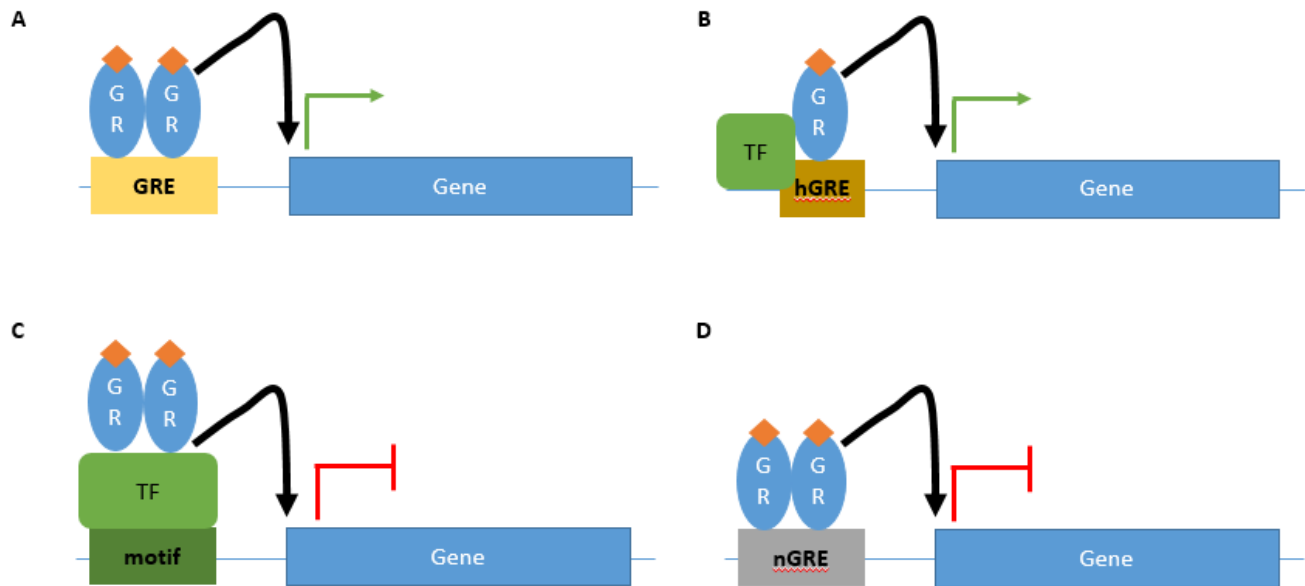


Figure 16: **Le GR régule la transcription de gènes cibles via différents mécanismes** (A) Fixation directe sur site consensus (GRE) et activation de la transcription du gène cible; (B) Fixation sur un demi-motif hexamerique (hGRE) et recrutement/interaction avec un co-facteur pour activer la transcription du gène; (C) Fixation indirecte via interaction avec un facteur de transcription déjà fixé pour exercer une activité transcriptionnelle antagoniste; (D) Fixation sur un site GRE négatif (nGRE) et inhibition de la transcription du gène cible.

(John et al. (2008); Vockley et al. (2016)). En effet, ces études démontrent que ce landscape prédéterminé au niveau de l'accessibilité et de l'activité des régions régulatrices tissus spécifiques est prédominant pour la fixation de la Dex.

L'action du GR sur la régulation de l'expression des gènes cibles est plus complexe que sa simple fixation sur l'ADN: celui-ci peut également permettre une ouverture de la chromatine par différents moyens:

- **Rôle dans l'ouverture de la chromatine:** Le GR interagit avec le complexe SWI/SNF. Ce complexe est composé d'une douzaine de protéines dont les différentes sous-unités de la protéine SMARCB1 et SMARCA4). L'interaction du GR avec ce complexe permet l'ouverture de nouvelles régions régulatrices accessibles au GR au sein de la cellule pour donner un nouveau niveau de régulation épigénétique de la transcription des gènes cibles (Grøntved et al., 2013).
- **Tetramerisation du GR:** Il a également été montré dans une étude récente (Paakinaho et al., 2010) que le GR pourrait non seulement s'homodimériser mais également lors de fortes doses de Dex former un homotetramère. Cette nouvelle structure permettrait non seulement

de se fixer sur les sites des régions régulatrices déjà actifs mais en plus de pénétrer des régions de la chromatine normalement inaccessibles pour s'y fixer et potentiellement réguler de nouvelles cibles (Figure 17).

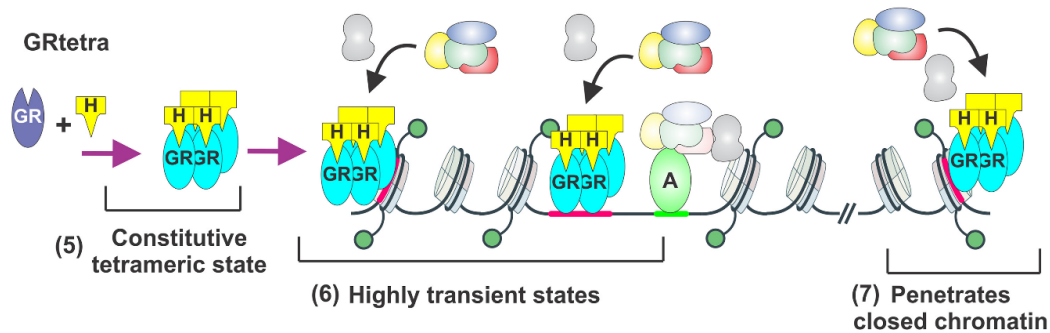


Figure 17: **La tetramérisation du GR permet de nouvelles fonctions moléculaires**, Figure tirée de Paakinaho et al. (2019).

- **Action sur la structure 3D:** Une étude récente réalisée sur une lignée cellulaire humaine de cancer du poumon (A549) (D'Ippolito et al., 2018) en utilisant des analyses de conformation 3D de l'ADN sur l'ensemble du génome (HiC) a démontré que le GR pouvait également jouer un rôle important au niveau de la structure 3D de l'ADN (Figure 18) en stabilisant les boucles de régulation. Cette étude démontre également que la majorité des interactions chromatinienne sont déjà pré-existantes et que l'addition de Dex altère la compartimentation des chromosomes mais pas les domaines TAD qui restent stables et non modifiés par le traitement à la Dex. En revanche, ils ont également montré que les interactions chromatinienne se répartissaient en réseaux et que les interactions d'un même réseau répondaient de façon coordonnée à la Dex et de façon générale, que les gènes répondant à la Dex étaient régulés par au moins une interaction chromatinienne dont la fréquence augmentait avec le traitement à la Dex. Ces études montrent également l'importance de la pré-activation des réseaux pour la fixation de la Dex.

Enfin, outre son rôle de fixation à l'ADN pour réguler la transcription des gènes cibles, le GR agit aussi sur la dégradation des ARN messagers. Certaines études (Park et al., 2016) ont rapporté que le GR se fixe aux ARNm pour réguler leur stabilité en formant des complexes avec les protéines PNRC2 et UPF1. La formation de ces complexes provoque ainsi le recrutement de la protéine DCPLA qui va déadenyler la queue polyA des ARNm et favoriser ainsi leur dégradation.

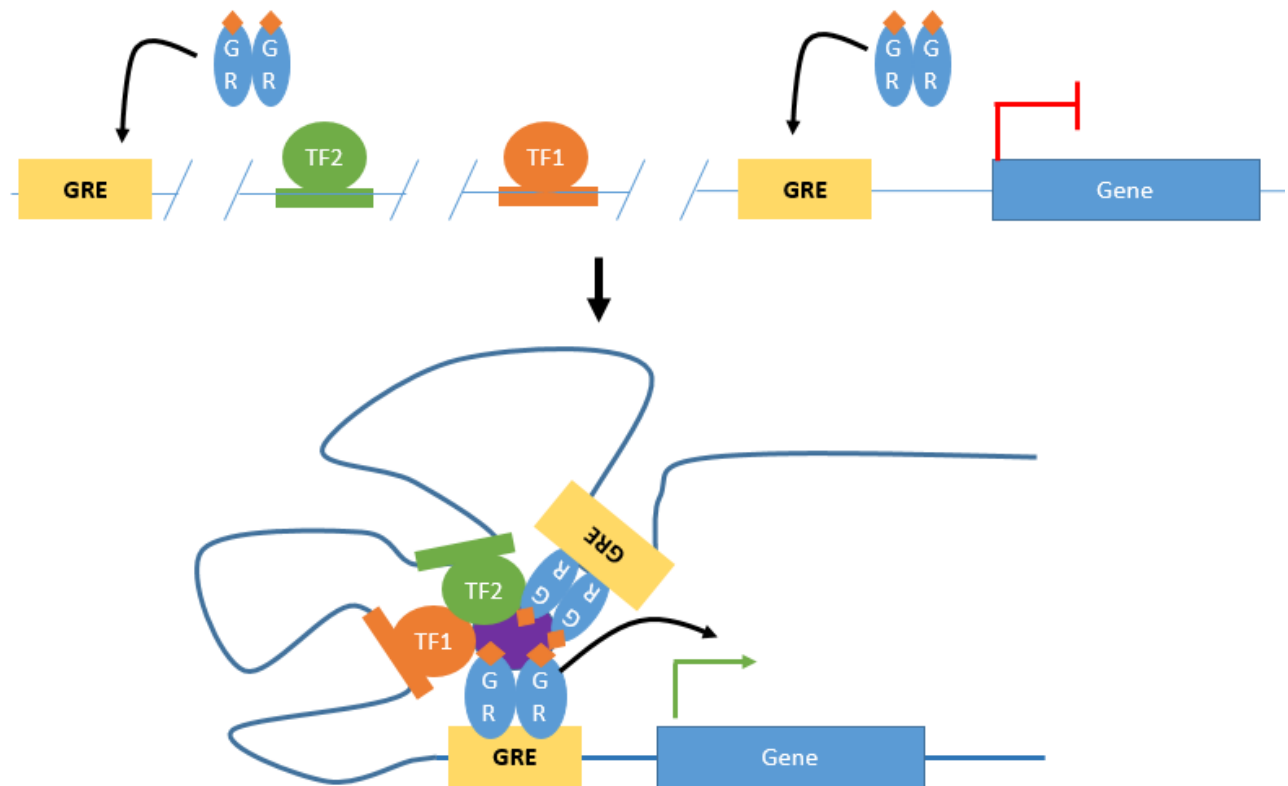


Figure 18: Le récepteur aux glucocorticoïdes agit sur la médiation des interactions enhancers-promoteurs à travers les boucles de régulation.

5 Sujet de thèse

Mon travail de thèse est à l'interface de la génomique, la cancérologie, l'informatique et la biologie moléculaire et a consisté à étudier les mécanismes de régulation épigénétique en réponse à un glucocorticoïde de synthèse, la Dexaméthasone, dans les cellules plasmocytaires malignes de patients atteints de myélome multiple (MM). La Dex exerce son effet en se fixant à son récepteur (GR) qui est alors transloqué dans le noyau où il agit comme un facteur de transcription et se fixe sur les séquences régulatrices du génome principalement des enhancers via sa séquence de reconnaissance (GRE : 5'-CCTGCAnnnTGTTCT-3') pour réprimer ou activer ses gènes cibles. La Dex est largement utilisée pour le traitement du MM en combinaison avec d'autres médicaments. Cependant, malgré un bénéfice certain, tous les patients ne répondent pas de la même façon. De plus, si l'efficacité de la Dex est attribuée à une apoptose spécifique des lymphocytes B matures, les mécanismes moléculaires mis en jeu après la fixation du GR sur l'ADN sont largement inconnus dans les plasmocytes.

Pour mieux comprendre le mode d'action moléculaire de la Dex et ses conséquences sur le transcriptome de l'ensemble de la population des cellules tumorales mais aussi sa variabilité cellule à cellule nous avons réalisé des études pan-génomiques qui intègrent des données d'accessibilité à la chromatine, de fixation de facteurs de transcription, de modification d'histone ainsi que des données de conformation de la chromatine couplées à des données de transcriptome en cellule unique (sc) obtenues dans un modèle de MM, la lignée MM.1S.

Mon travail durant cette thèse fut donc de répertorier et d'automatiser les différents outils nécessaires à l'analyse, de réaliser les analyses bioinformatiques de toutes les données biologiques que nous avons générées au laboratoire et de répondre aux différentes questions biologiques posées par le sujet de thèse. Ainsi, j'ai mis en place des pipelines d'analyses des données de ChIP-seq pour des marques d'histones (particulièrement H3K27ac) ainsi que pour des facteurs de transcription (GR et facteurs de la voie NF- κ B), pour des données d'ATAC-seq et enfin pour des données de HiChIP-seq (méthode de détection de boucles d'interaction chromatinienne associée à la présence d'une protéine spécifique). En parallèle, j'ai également analysé des données de RNA-seq et j'ai travaillé à l'intégration de l'ensemble de ces données avec des résultats de single-cell ATAC et RNAseq (réalisées par un autre doctorant).

6 Techniques et méthodes d'analyse

Les nouvelles avancées en matière de séquençage haut-débit ont permis ces dernières années de répondre à de nombreuses questions sur les mécanismes moléculaires de régulation de l'expression des gènes. Grâce à ces avancées, des analyses pan-génomiques de séquençage combinées à de nombreuses analyses bioinformatiques ont été réalisées afin de répondre à la problématique posée pour ce travail de thèse.

6.1 Matériel biologique utilisé

Dans cette étude, le modèle utilisé est une lignée cellulaire MM.1S. Cette lignée est issue d'un patient féminin de 42 ans atteint de myélome. Elle est obtenue à partir d'un clone de lymphocytes B du sang périphérique. Ce patient est devenu résistant aux thérapies basées sur les stéroïdes. La lignée MM.1S exprime le récepteur aux glucocorticoïdes ainsi que la chaîne légère λ des Igs. La lignée MM.1S est sensible à la Dex.

Cette lignée est classée dans les sous-populations de MM présentant une translocation 14;16. De ce fait, cette anomalie génétique conduit à une sur-expression du facteur de transcription c-MAF dans la cellule.

Cette lignée est cultivée dans un milieu RPMI additionné de L-Glutamine à 1% et de sérum fetal bovin à 10% (facteur de croissance). Les cellules sont cultivées à 37°C sous atmosphère à 95% d'air et 5% de CO₂ (Goldman-Leikin et al., 1989).

6.2 Techniques de biologie moléculaire

6.2.1 Immunoprécipitation de protéines

La technique de ChIP-seq permet d'identifier à l'échelle du génome entier l'ensemble des sites de fixation sur l'ADN d'un facteur de transcription ou d'un histone avec une modification post-traductionnelle spécifique (Park, 2009) (Figure 19).

La première étape de cette technique consiste à effectuer une fixation de l'ensemble des protéines liées à l'ADN dans la cellule (cross-link). Toutes les protéines (dont facteurs de transcription et histones) sont alors maintenues fixées à l'ADN. Le génome est ensuite fragmenté (par coupure enzymatique ou sonication), la taille de ces fragments dépend donc du type de cassure effectuée. Les fragments d'ADN d'intérêts vont ensuite être immuno-précipités grâce à un anticorps spécifique de la protéine d'intérêt (TF) ou de la marque d'histone (par exemple H3K27ac). L'ADN est ensuite purifié et amplifié par PCR après l'ajout d'adaptateurs spécifiques. L'ADN ainsi obtenu sera ensuite séquencé.

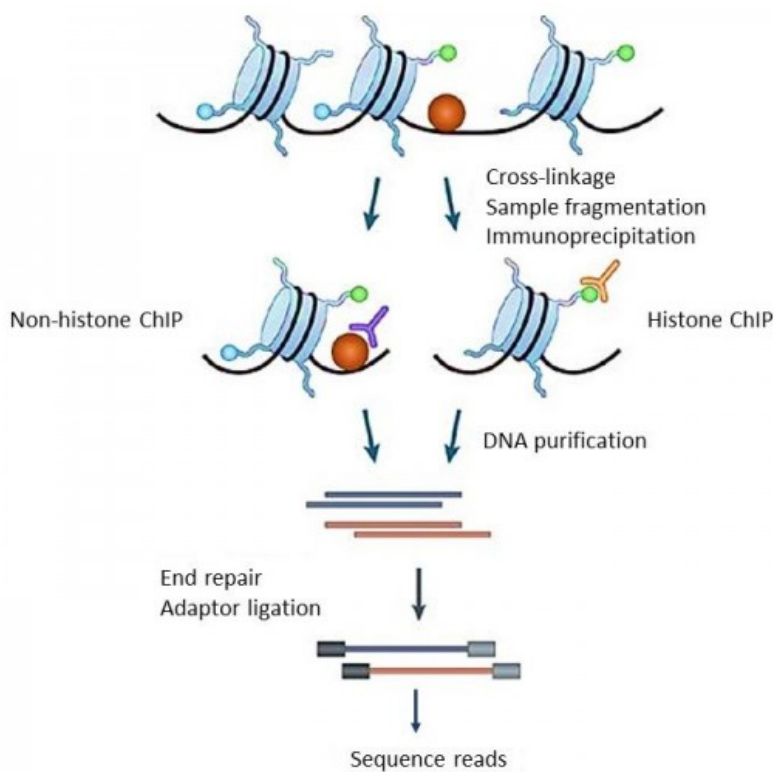


Figure 19: **Protocole de Chromatine ImmunoPrecipitation (ChIP) couplé au séquençage**, Figure issue du site: www.france-genomique.org.

6.2.2 Techniques de détection d'accessibilité de la chromatine

La fixation d'un facteur de transcription sur l'ADN est facilitée par l'ouverture de la chromatine. Il existe diverses techniques pour déterminer l'accessibilité de la chromatine tout au long du génome: le DNase-seq, le FAIRE-seq et l'ATAC-seq. Ces techniques reposent sur des méthodes différentes de fragmentation de l'ADN, par coupure enzymatique (DNase), sonication (FAIRE) ou par utilisation de l'enzyme transposase Tn5. Dans cette thèse, nous avons utilisé la méthode de séquençage ATAC-seq et son protocole proposé par l'équipe de Buenrostro (Buenrostro et al., 2015) avec une adaptation proposée dans une autre étude (Corces et al., 2016) qui est la technique la plus récente et qui a l'avantage de requérir peu de matériel biologique au départ (environ 50 000 cellules suffisent pour effectuer cette technique) (Figure 20).

Cette technique est de plus, assez rapide car elle repose juste sur l'incubation de la chromatine en présence d'une enzyme spéciale: la transposase Tn5 associée à deux adaptateurs. Pendant l'incubation, le complexe transposome va s'insérer dans l'ADN au niveau de ses sites de reconnaissance accessibles ce qui permet à la fois la fragmentation et l'indexation des séquences d'ADN. Cet ADN est ensuite purifié et amplifié par PCR avant d'être séquençé.

Cette technique permet d'identifier non seulement les régions accessibles de l'ADN mais également d'identifier les régions contenant un ou plusieurs nucléosomes en se basant sur la taille du fragment obtenu.

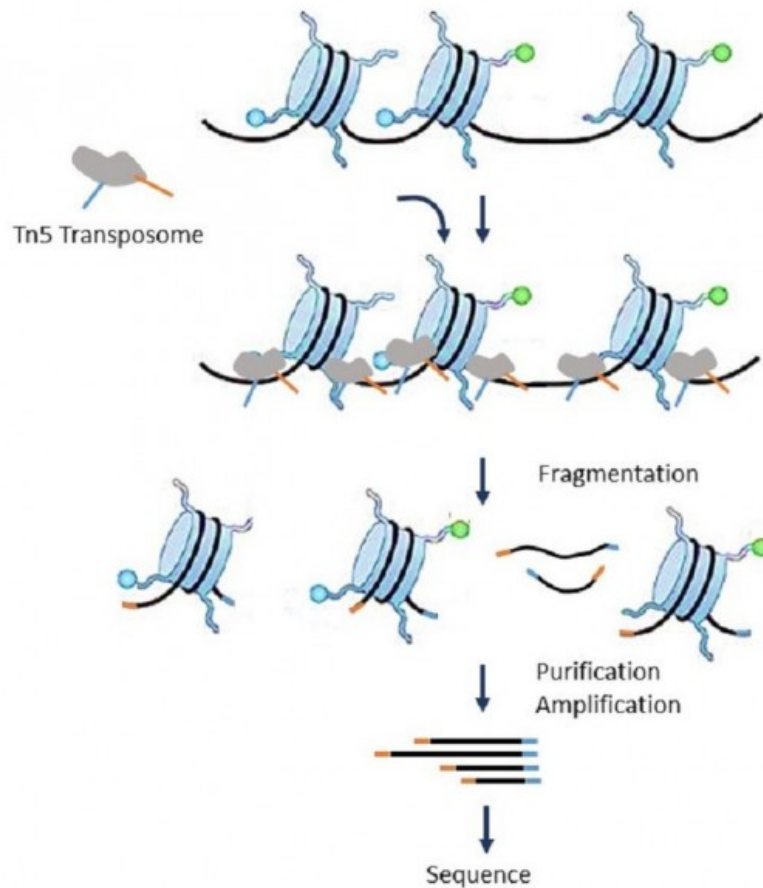


Figure 20: **Accessibilité de la chromatine liée à la transposase Tn5 (ATAC-seq)**, Figure issue du site: www.france-genomique.org.

6.2.3 Analyse des niveaux d'expression ARN

Pour connaître le niveau d'expression des transcrits au sein d'une population cellulaire, on utilise la technique du RNA-seq (Figure 21).

Les ARN messagers poly-adiénylés de la cellule sont capturés et fragmentés. On capture ces ARNm spécifiquement car la polyadiénylation leur offre une protection contre la dégradation et que la majorité des ARNs poly-adiénylés correspondent à des ARN messagers qui seront ensuite traduits en protéines. Des amorces s'hybrident sur ces fragments d'ARN avant synthèse de l'ADN complémentaire (ADNc). On obtient donc un ADN synthétique double brin auquel vont être liés

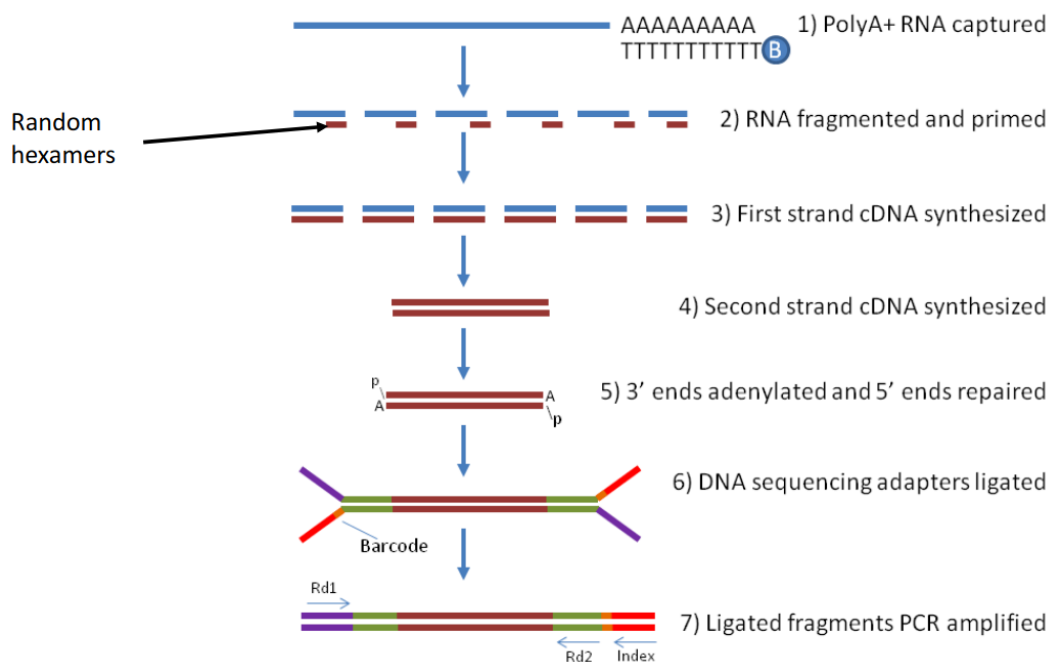


Figure 21: **Technique de séquençage des ARN messagers RNA-seq**, Figure issue du site: <https://www.labome.com/method/RNA-seq-Using-Next-Generation-Sequencing.html>.

des adaptateurs 5' et 3' avant amplification de ces fragments par PCR et suivi du séquençage haut débit.

Le résultat du séquençage va permettre d'évaluer le niveau de transcrit pour tous les gènes du génome mais permet également de travailler et de comprendre d'autres mécanismes de régulation de l'expression des gènes tels que l'épissage alternatif ou bien encore la détection de SNP (polymorphisme nucléotidique) sur les parties codantes des gènes correspondants. Dans les travaux présentés dans cette thèse, seuls les différentiels d'expression génique entre deux conditions seront étudiés.

6.2.4 Analyse des interactions ADN:ADN liées à la présence d'une protéine

Les trois premières techniques d'analyse présentées auparavant permettent donc de définir le paysage chromatinien et transcriptionnel d'une population cellulaire d'un point de vue linéaire. En revanche, de nombreuses études montrent que lors de l'analyse de régions régulatrices distales (enhancers), il est très souvent difficile d'identifier la bonne cible sur le génome car ces analyses ne prennent pas en considération la conformation 3D du génome.

Comme précisé dans la partie 3.5, les interactions chromatinienne peuvent être étudiées à différents niveaux: les compartiments A et B, les domaines TAD et enfin les boucles de régulation. Pour analyser ce phénomène la technique généralement employée est celle du HiC. Les limites de cette tech-

nique sont qu'elle nécessite une profondeur de séquençage conséquente pour obtenir une résolution suffisante pour l'étude des boucles de régulation et qu'elle n'est pas spécifique des interactions liées à un facteur ou une marque d'histone.

Une étude récente (Mumbach et al., 2016) propose un compromis entre la technique du HiC et celle du ChIP-seq appelée HiChIP (Figure 22). Cette technique permet donc d'étudier les interactions chromatiniennees liées à la présence d'un facteur de transcription spécifique ou encore d'une marque d'histone.

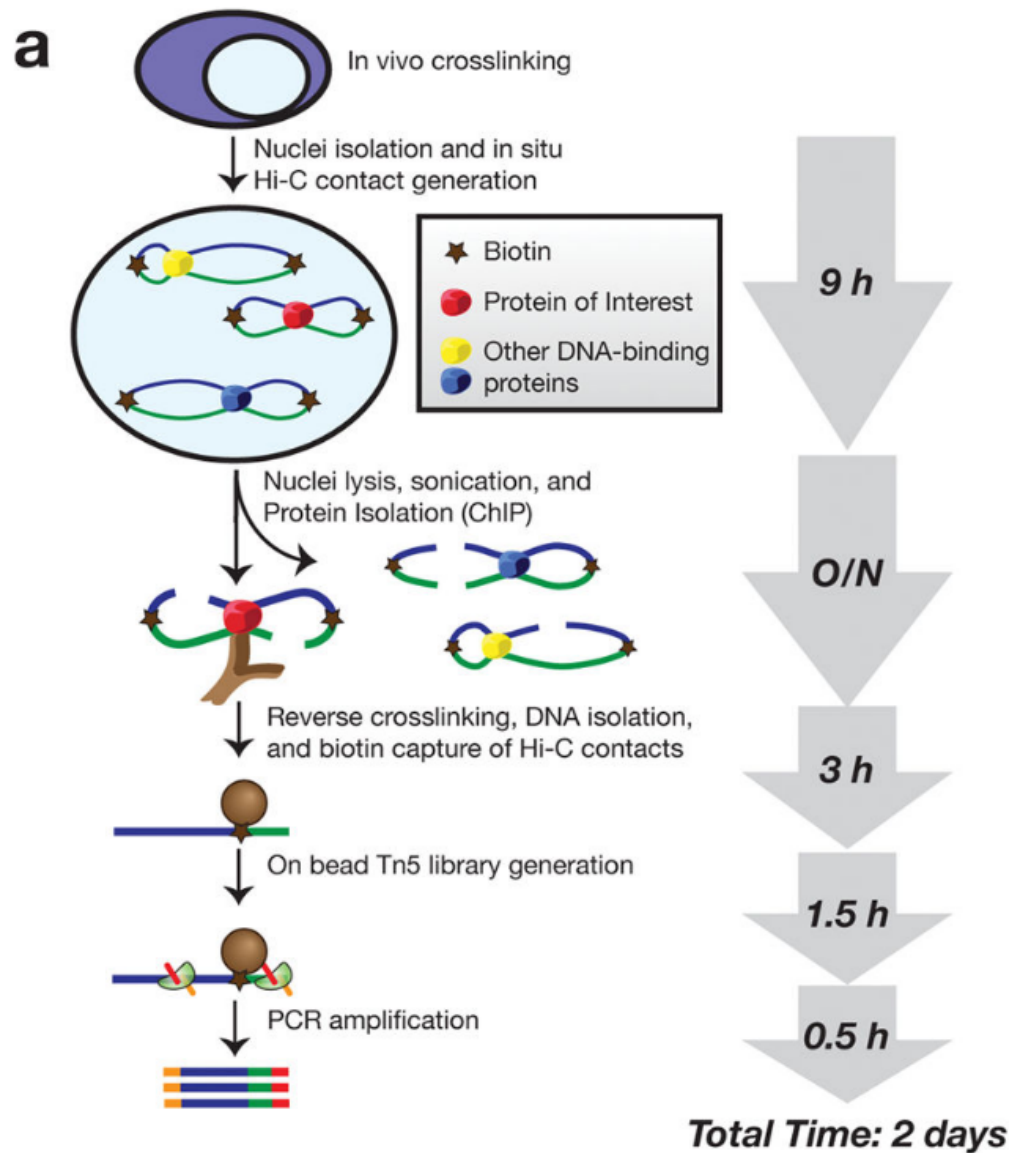


Figure 22: **Technique de séquençage HiChIP**; O/N: Overnight, Figure issue de Mumbach et al. (2016)

La chromatine est tout d'abord fixée avec les protéines de la même manière que pour le ChIP-

seq avant que l'ensemble des contacts chromatiniens soit capturé et circularisé par la biotine. Les noyaux cellulaires sont ensuite lysés pour récupérer ces ADN circulaires avant d'être coupés par sonication puis immunoprécipités avec l'anticorps spécifique de la protéine ou de la marque d'histone d'intérêt. Une fois les fragments voulus capturés, les protéines et l'ADN sont séparées par reverse cross-linking et l'ADN est purifié avant de capturer la biotine par billes magnétiques. Finalement la librairie est générée par utilisation de l'enzyme transposase Tn5 qui va fragmenter et indexer l'ADN qui est alors amplifié par PCR avant séquençage.

Cette technique permet donc de dresser une carte épigénétique de l'ensemble des interactions protéines-protéines ou marque d'histone spécifique au sein d'une population.

6.2.5 Techniques de séquençage nouvelle génération

Le séquençage de l'ADN permet ainsi d'identifier la composition nucléotidique de séquences d'intérêts importantes. Récemment, l'apparition des techniques de séquençage haut débit permet aisément de séquencer l'ensemble d'un génome en très peu de temps. Grâce à l'émergence de ces techniques, on peut ainsi séquencer l'ADN mais aussi l'ARN beaucoup plus rapidement et en nombre plus important que par les anciennes techniques tel que le séquençage de Sanger.

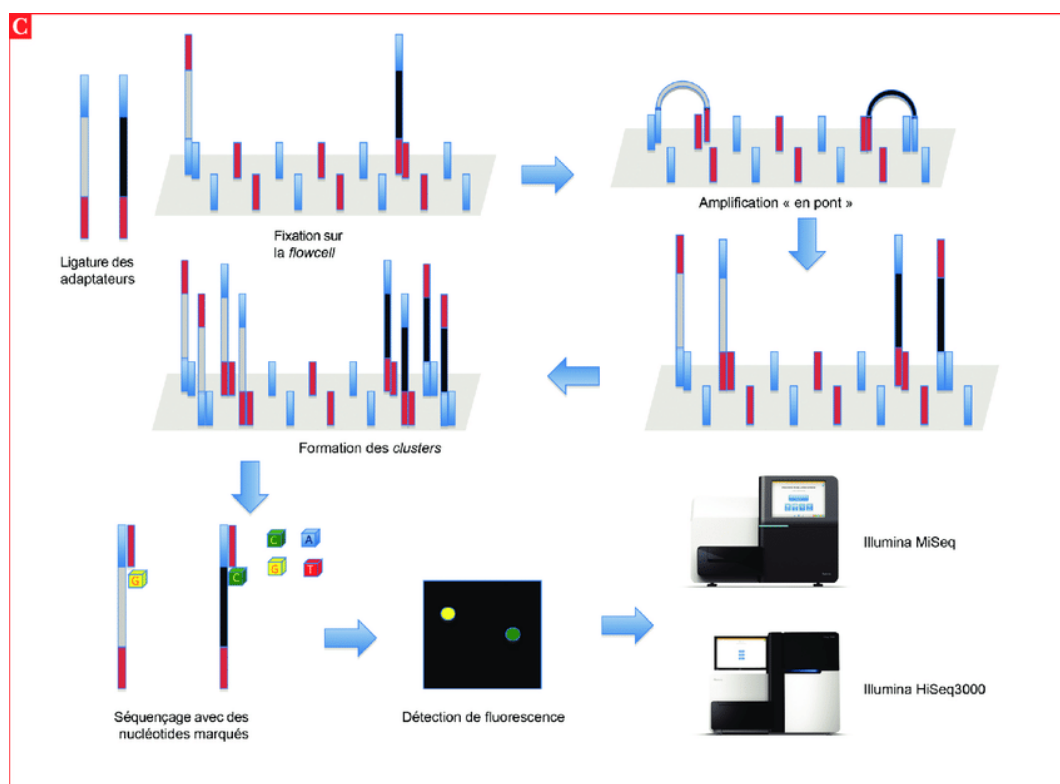


Figure 23: **Techniques de séquençage haut débit (NGS)** issue de Dortet et al. (2017).

Les technologies de séquençage haut débit (NGS) comportent trois étapes importantes quelque

soit la technique utilisée (Figure 23): la préparation de la banque qui suit une fragmentation du génome de manière aléatoire et la capture de ces fragments par liaison avec des petites séquences spécifiques ou immunoprécipitation, l'amplification de la banque grâce à une approche par PCR et enfin le séquençage de l'ADN ainsi obtenu. A ce jour, la technologie NGS la plus utilisée est la technologie Illumina qui permet l'identification simultanée des 4 bases d'ADN lorsqu'elles sont incorporées dans la chaîne. L'identification de ces bases se fait par lecture de la fluorescence émise à l'ajout de chaque base et qui permet de reconstituer la séquence d'ADN initiale.

6.2.6 Immunoprécipitation de protéines suivie de spectrométrie de masse

Afin de déterminer l'ensemble des protéines interagissant physiquement avec une protéine d'intérêt, une nouvelle approche de spectrométrie de masse a été mise en place: la technique du RIME (Mohammed et al., 2016). Cette approche associe donc la technique classique d'immunoprécipitations

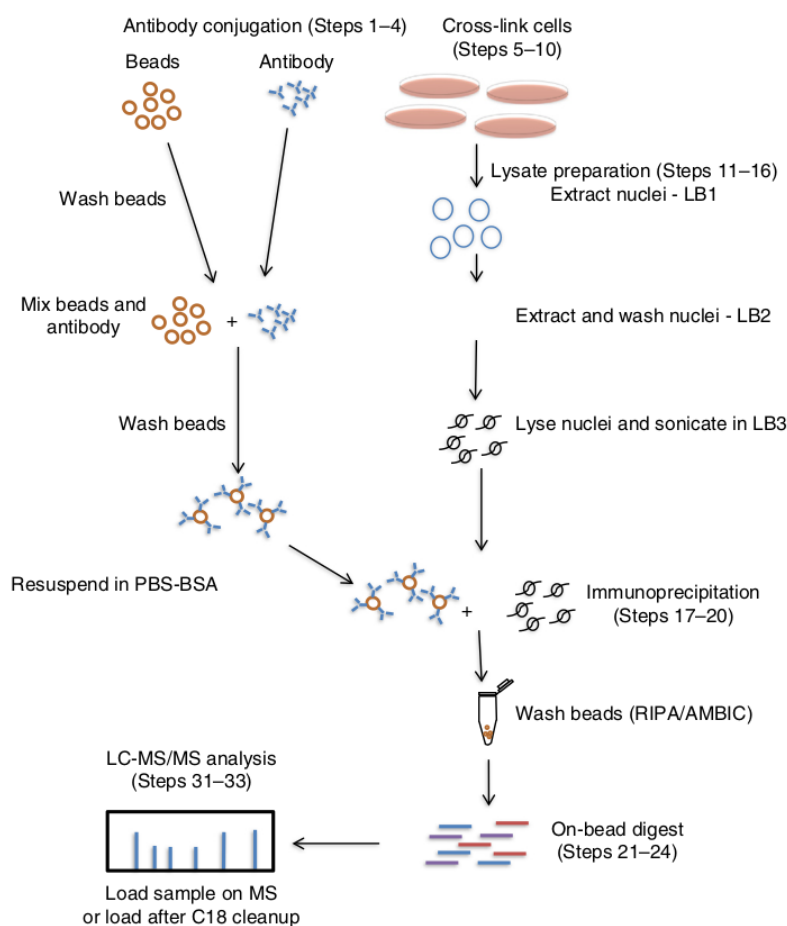


Figure 24: **Immunoprécipitation des protéines suivie d'une spectrométrie de masse (RIME)** issue de Mohammed et al. (2016).

de protéines. Les protéines sont fixées dans les cellules par cross-linking ensuite les noyaux sont ex-

traits puis lysés et fragmentés par sonication. En parallèle, les anticorps spécifiques de la protéine d'intérêt sont fixés sur des billes magnétiques. Ces billes sont mises en présence des fragments pour immunoprécipiter la protéine d'intérêt et capturer à la même occasion, l'ensemble des protéines interagissant physiquement avec la protéine d'intérêt. Après digestion enzymatique des protéines, celles ci sont alors identifiées par spectrométrie de masse (Figure 24).

6.2.7 Nouvelles approches en cellule unique

Ces différentes techniques décrites auparavant permettent de séquencer une population cellulaire entière et donc d'obtenir l'information moyennée sur cette population. Comme décrit précédemment, il existe au sein même d'une population cellulaire une grande hétérogénéité (état de la cellule, anomalies...) qui ne pourra pas être observée avec ces techniques de séquençage dites en bulk. Ces dernières années ont donc vu l'émergence de nouvelles approches en cellule unique (single cell) où chaque cellule sera identifiée individuellement pour observer l'information à l'échelle d'une cellule au lieu d'une population.

Single-Cell RNA-seq

Cette technique permet d'étudier le transcriptome d'une cellule unique et repose sur l'utilisation de séquences uniques (barcodes) pour identifier chaque cellule (Zheng et al., 2017).

Pour isoler les cellules individuellement, on utilise les principes de microfluidique. Les cellules sont d'abord isolées dans une micro gouttelette d'huile contenant les réactifs nécessaires à la réaction (oligonucléotides et retro-transcriptase) ainsi qu'avec une bille GEM (Gel bead in Emulsion) (Figure 25). Les GEM sont des billes présentant à leur surface des amorces spéciales. Ces amorces sont composées de 3 éléments: un barcode unique pour chaque GEM qui permettra d'identifier par la suite chaque cellule, un UMI (Unique Molecular Identifier) qui permet d'éviter les biais d'amplification (une séquence trop amplifiée sera détectée par sur-représentation de son UMI) et enfin une séquence poly(T) qui permet l'hybridation des ARNs messagers par complémentarité avec la queue poly(A).

Ensuite les cellules passent dans un microréacteur pour être lysées et les ARNm sont capturés puis l'ensemble de la séquence barcode, UMI et ARN est convertie en ADN complémentaire. Finalement, l'huile est retirée et l'ADNc est poolé et amplifié. Il ne reste alors que l'étape de la construction de la librairie (fragmentation, ligation des adaptateurs et amplification par PCR) et du séquençage haut-débit qui s'en suit.

10x Genomics Library Preparation

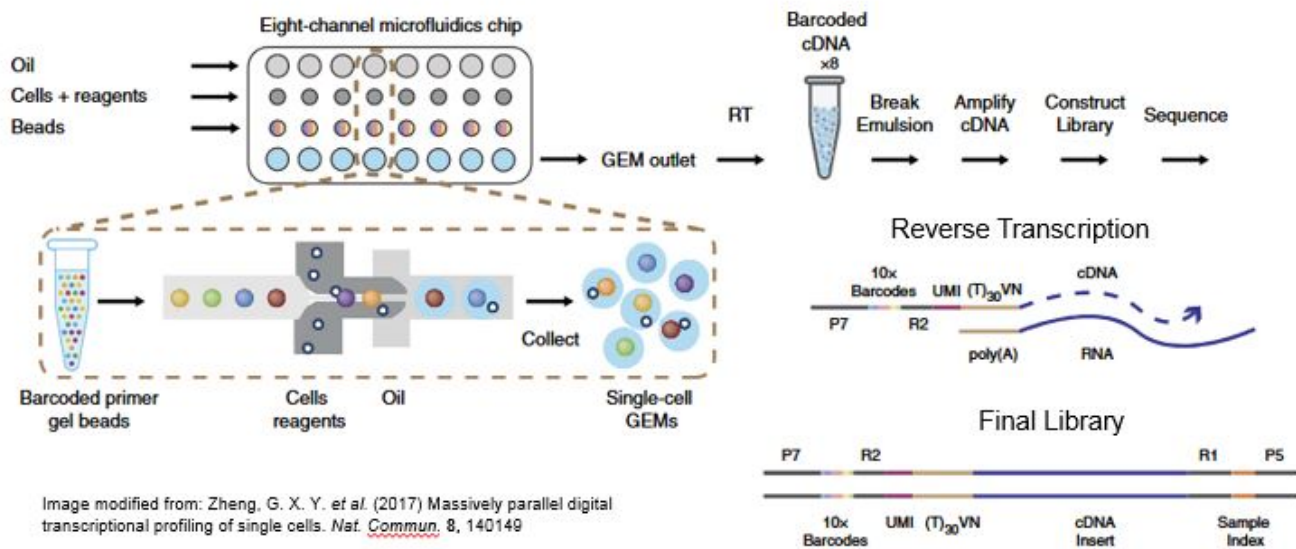


Figure 25: Préparation de librairie scRNA-seq figure issue de Zheng et al. (2017).

Couplage Single-Cell ATAC et RNA-seq

10x Genomics propose un nouveau protocole permettant à la fois de capturer les ARNs messagers nucléaires et les régions accessibles du génome au sein d'une même cellule: le single-cell multiome-seq.

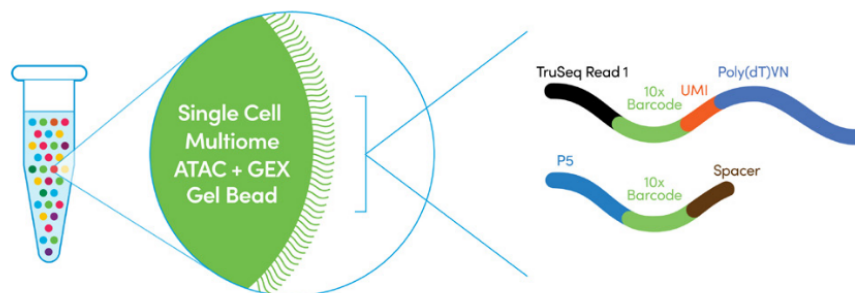


Figure 26: Composition des billes GEM pour le séquençage Multiome, par 10x Genomics.

Dans ce protocole, la lyse des cellules se fait avant capture par utilisation d'un détergent (digitonine) ainsi que la transposition. C'est donc uniquement les noyaux transposés qui sont capturés par les GEM. La transposition se fait dans une premier temps puis la rétro-transcription des

ARNs messagers nucléaires vient ensuite. Au niveau de la bille GEM, on va toujours retrouver les séquences nécessaires à la capture des ARNs messagers (barcode+UMI+queue poly(T)) mais on retrouve également des séquences contenant la séquence d'adaptateur P5, le barcode et un spacer qui va permettre la fixation des fragments d'ADN transposés pour la librairie ATAC (Figure 26).

L'utilisation de cette technique permet alors d'identifier les fragments de chromatine accessibles et le transcriptome associé au sein d'une même cellule.

6.2.8 Génération des données

Pour répondre à la problématique posée par le sujet de thèse, nous avons généré au sein de notre laboratoire une multitude de données de séquençage dites multi-omiques. L'ensemble de ces données est répertorié dans la Table 2.

Technique	Facteur ou marque ciblée	Condition	Type de séquençage
RNA-seq	Transcriptome	MM1S EtOH 4h	Paired-End (PE)
RNA-seq	Transcriptome	MM1S Dex 4h 0.1 μ M	PE
ChIP-seq	GR	MM1S EtOH	Single-End (SE)
ChIP-seq	GR	MM1S Dex 1h 0.1 μ M	SE
ChIP-seq	CTCF	MM1S EtOH	SE
ChIP-seq	CTCF	MM1S Dex 1h 0.1 μ M	SE
ChIP-seq	H3K27ac	MM1S EtOH	SE
ChIP-seq	H3K27ac	MM1S Dex 1h 0.1 μ M	SE
ATAC-seq	Chromatine	MM1S EtOH	PE
ATAC-seq	Chromatine	MM1S Dex 1h 0.1 μ M	PE
HiChIP-seq	H3K27ac	MM1S EtOH	PE
HiChIP-seq	H3K27ac	MM1S Dex 1h 0.1 μ M	PE
scRNA-seq	Transcriptome	MM1S EtOH 1h	SE
scRNA-seq	Transcriptome	MM1S EtOH 4h	SE
scRNA-seq	Transcriptome	MM1S Dex 1h 0.1 μ M	SE
scRNA-seq	Transcriptome	MM1S Dex 4h 0.1 μ M	SE
scMultiome-seq	Transcriptome et chromatine	MM1S EtOH 1h	SE
scMultiome-seq	Transcriptome et chromatine	MM1S EtOH 4h	SE
scMultiome-seq	Transcriptome et chromatine	MM1S Dex 1h 0.1 μ M	SE
scMultiome-seq	Transcriptome et chromatine	MM1S Dex 4h 0.1 μ M	SE

Table 2: Table récapitulative des données générées et analysées.

Les temps choisis sont de 1h pour l'ensemble des phénomènes épigénétiques étudiés et de 4h pour l'étude du transcriptome car les mécanismes de transcription répondant à une drogue ne sont pas visibles avant. Nous avons choisi une concentration en Dex de 0.1 μ M car c'est une dose suffisante pour induire la mort cellulaire dans la lignée MM.1S (Kervoëlen et al., 2015) (Figure 27)

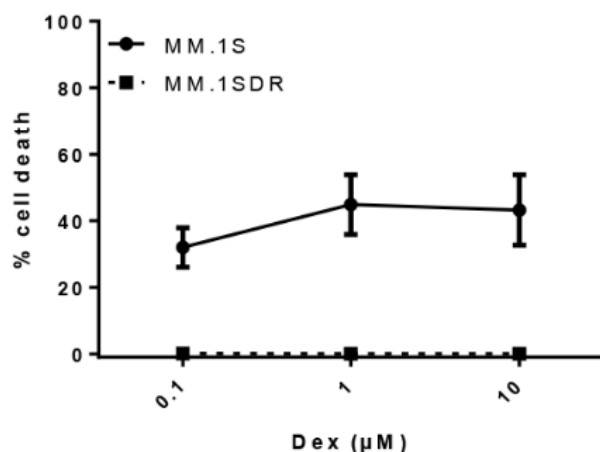


Figure 27: Dose de Dex induisant la mort cellulaire dans la lignée MM.1S, figure issue de Kervoëlen et al. (2015).

6.3 Analyses computationnelles

Le séquençage haut-débit permet d'explorer de nombreux mécanismes moléculaires importants pour la cellule mais au prix d'un volume de données très important (Un séquençage génère en effet de plusieurs dizaines à plusieurs centaines de millions de séquences ADN). Afin de pouvoir sortir de cette masse de données des informations relevantes biologiquement, il est nécessaire de passer par des approches de bioinformatique. Cette discipline est en plein essor depuis l'arrivée des microarrays puis du séquençage haut-débit et peut être très complexe selon la problématique étudiée. Je vais présenter ici les différentes approches (outils et pipelines) que j'ai été amené à utiliser durant cette thèse pour traiter et intégrer l'ensemble de ces données multi-omiques.

6.3.1 Préparation des données pour alignement

Le contrôle qualité post-séquençage des données est la première et non moindre importante des étapes de toute analyse bioinformatique. Peu importe le type de données NGS générées, il est essentiel d'évaluer la qualité de la librairie séquencée et du séquençage lui-même pour valider ces données.

Contrôle qualité post-séquençage

Lors du séquençage des données, la qualité de séquençage de chacune des bases composant le read est estimée. Ce score de Sanger, nommé Phred+33 et généralement codé en ASCII 33 donne

pour chacune des bases séquencée, sa probabilité d'erreur, c'est à dire la probabilité que la base soit fausse (soit par mauvaise inclusion d'un nucléotide lors de la phase d'incorporation soit par une mauvaise lecture de la fluorescence). Le score de Sanger est obtenu par la formule suivante:

$$Q_{sanger} = -10 \times \log_{10}(p),$$

De façon générale, on admet qu'une probabilité d'erreur p inférieure à 0.001 (équivalent à un score de Sanger supérieur à 30) est suffisante pour considérer que la base a été correctement séquencée. Les scores sont ensuite encodés en ASCII 33 ce qui permet d'attribuer dans le fichier fastq un caractère unique pour chaque base (exemple: '!' correspond à un score de 0 et 'I' correspond à un score de 40). Le score de Sanger varie de 0 à 40. Pour estimer la qualité de séquençage de nos données, j'utilise fastqc (Andrews et al., 2010). Fastqc est une suite d'outils permettant de réaliser le contrôle qualité des données il permet d'évaluer la qualité des bases grâce au score Phred+33 présent dans le fastq. Il permet d'estimer le score de qualité par séquence, le pourcentage de base GC par séquence ou encore la distribution de la taille des séquences ou le niveau de duplication des séquences.

Trimming des reads

Parfois lors de l'amplification des fragments par PCR, il est possible que les adaptateurs qui permettent le séquençage et l'identification des reads soient également amplifiés. Ces adaptateurs peuvent poser par la suite des problèmes d'alignement en créant de nombreux mismatches lors de l'alignement des reads. Il est donc essentiel de couper (trimming) les reads qui ont à leurs extrémités la présence de ces adaptateurs.

Dans cette thèse nous avons choisi d'utiliser l'outil bioinformatique trimmomatic (Bolger et al., 2014) pour réaliser cette tâche.

6.3.2 Transformée de Burrows-Wheeler pour l'alignement de reads

Une fois l'ensemble des séquences (reads) validées par les différents contrôles qualités (QC), il faut réaliser l'étape incontournable de l'analyse de données NGS: l'alignement des reads sur un génome de référence. Il existe plusieurs génomes de référence et plusieurs versions. Dans toute cette thèse le génome de référence utilisé est la version hg19 (Human Genome version 19) de refseq disponible sur UCSC (Kent et al., 2002).

L'alignement va permettre de dresser une carte génomique et de positionner l'ensemble des reads sur le génome pour identifier et localiser les régions ou transcrits obtenus par séquençage.

Alignement global

L'alignement des reads pour l'ensemble des données NGS exceptés le RNA-seq et le HiCHIP (cf

6.3.2 et 6.3.2), a été réalisé avec l'outil bioinformatique Bowtie2 (Langmead et al., 2009). Cet outil utilise un algorithme d'alignement bien connu appelé transformée de Burrows Wheeler (Figure 28 et Figure 29) qui permet une indexation du génome de référence. Le génome de référence est

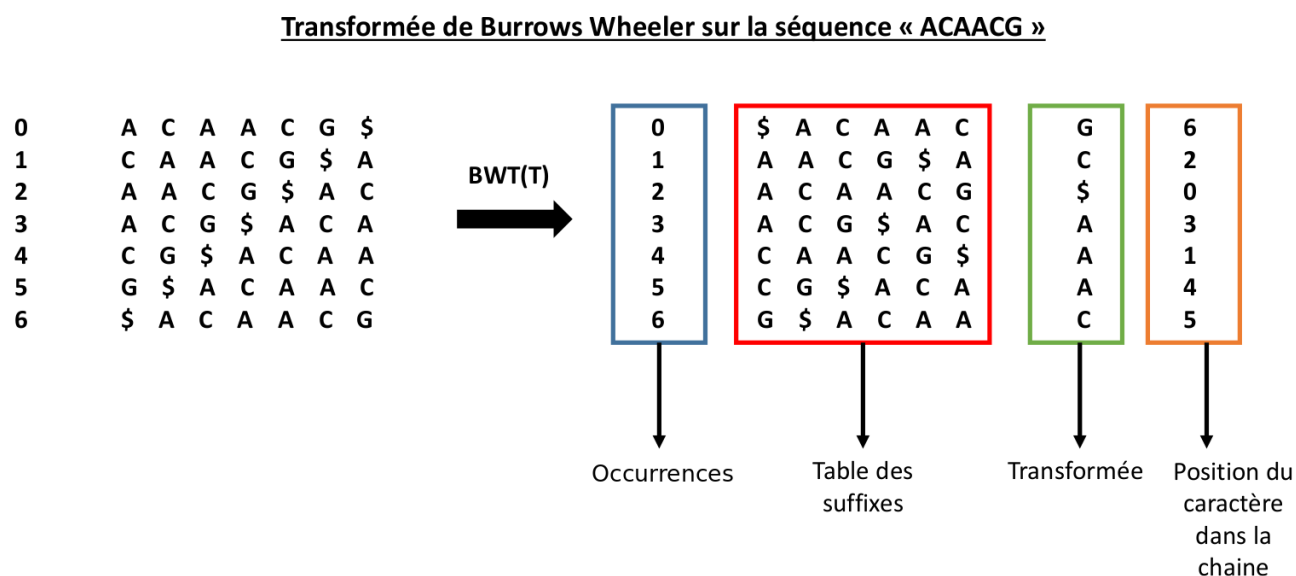


Figure 28: **Méthode de Burrows Wheeler pour indexation des séquences (ou génome) avant alignement.** L'algorithme itère sur la longueur de la chaîne de caractère avec un décalage d'une lettre à chaque occurrence pour générer une transformée à un caractère associé à sa position dans la chaîne.

donc transformé en une chaîne de caractères unique et l'index est créé via la transformation de Burrows-Wheeler (Figure 28). On réalise l'ensemble des rotations de cette chaîne avec à chaque fois un décalage de 1 pour récupérer à chaque rotation le dernier caractère de la chaîne et réaliser ensuite un tri alphabétique de sorte à ce que toutes les sous chaînes contenant un suffixe commun se retrouvent les unes à la suite des autres dans la table des suffixes. De plus la position du dernier caractère de chaque rotation est également enregistrée de manière à retrouver cette position au sein de la séquence initiale.

Concernant l'alignement des reads, chaque read est lu à l'envers en commençant donc par le dernier caractère et le read est ainsi reconstitué à partir de la transformée (Figure 29). Cette étape est déterminante pour identifier avec précision les endroits du génome matchant avec les reads. Il est possible d'admettre un ou plusieurs mismatch (équivalent à un changement, une délétion ou un ajout d'un nucléotide) dans l'alignement des reads (généralement 1 ou 2 maximum) pour éviter d'être biaisé du fait de la présence de nucléotides à fort polymorphisme dans le génome humain. L'alignement fournit généralement un fichier final (.sam ou binarisé .bam) qui répertorie pour chacun des reads son alignement, sa localisation génomique ainsi que diverses autres informations utiles pour la suite de l'analyse.

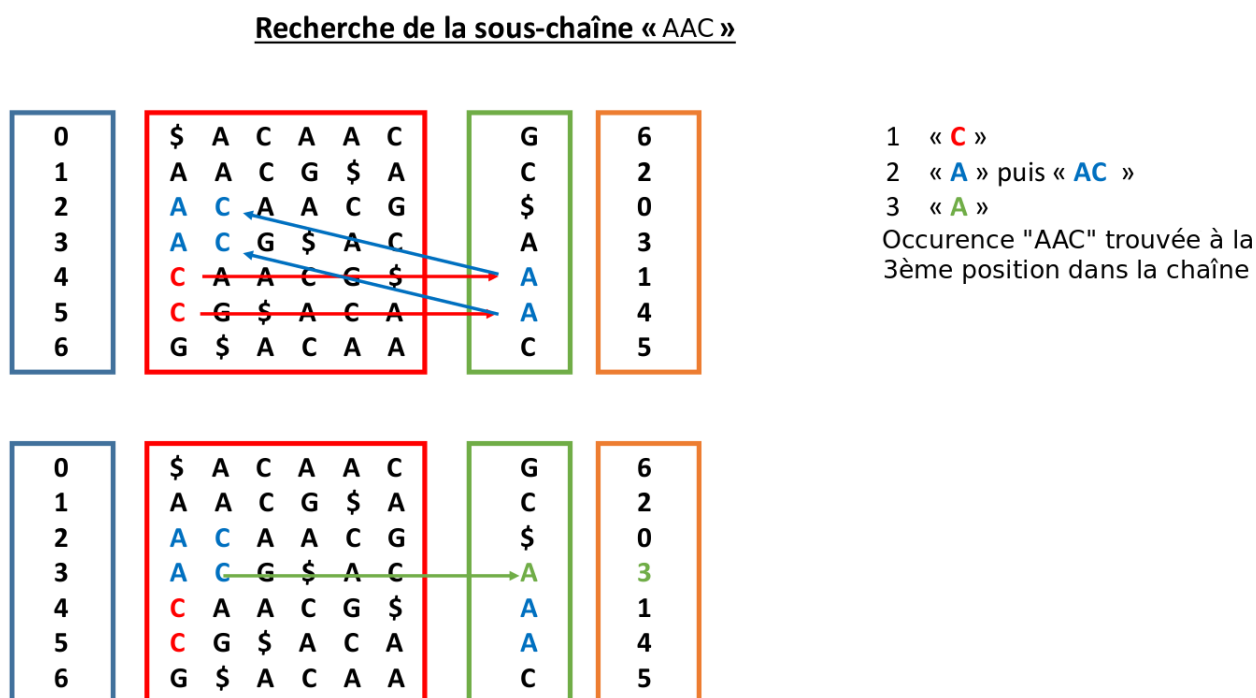


Figure 29: Utilisation de la transformée de Burrows-Wheeler pour l'alignement des reads issus du séquençage. La recherche d'une sous chaîne appelle les différents éléments générés par la transformée BW pour identifier à partir de la fin de la sous-chaîne, sa position dans la séquence initiale.

Spécificité liée au RNA-seq

L'alignement des reads issus de séquençage RNA-seq nécessite une prise en compte du phénomène d'épissage. En effet il est possible d'obtenir des reads mappant complètement dans un exon et des reads chevauchant deux exons. Sur le génome de référence la présence des introns va donc générer d'énormes mismatches pour ces reads. Nous avons donc utilisé l'outil d'alignement Tophat2 (Trapnell et al., 2009) qui utilise également l'outil Bowtie2 mais en prenant en compte ces reads chevauchant. Après une première étape d'alignement global, les reads qui ne se sont pas alignés complètement sont alors coupés en petites graines ("seed"). Ces graines sont de nouveau alignées sur le génome de référence en utilisant en plus un fichier d'annotation des exons pour créer une librairie exon-exon indexée sur laquelle les reads sont alignés, cela permet alors de reconstituer les différentes jonctions des exons. Ces petites graines sont alors réassemblées.

Spécificité liée au HiChIP

En ce qui concerne l'alignement de reads produits par technique du HiChIP-seq, deux principaux types de reads sont générés: les reads qui contiennent une séquence intégralement retrouvée au

sein d'un anchor (une interaction chromatinienne se fait entre deux régions distinctes du génome appelées anchors) et les reads hybrides contenant deux séquences partielles provenant des deux anchors de l'interaction. L'alignement se fait donc en deux étapes: un alignement global suivi

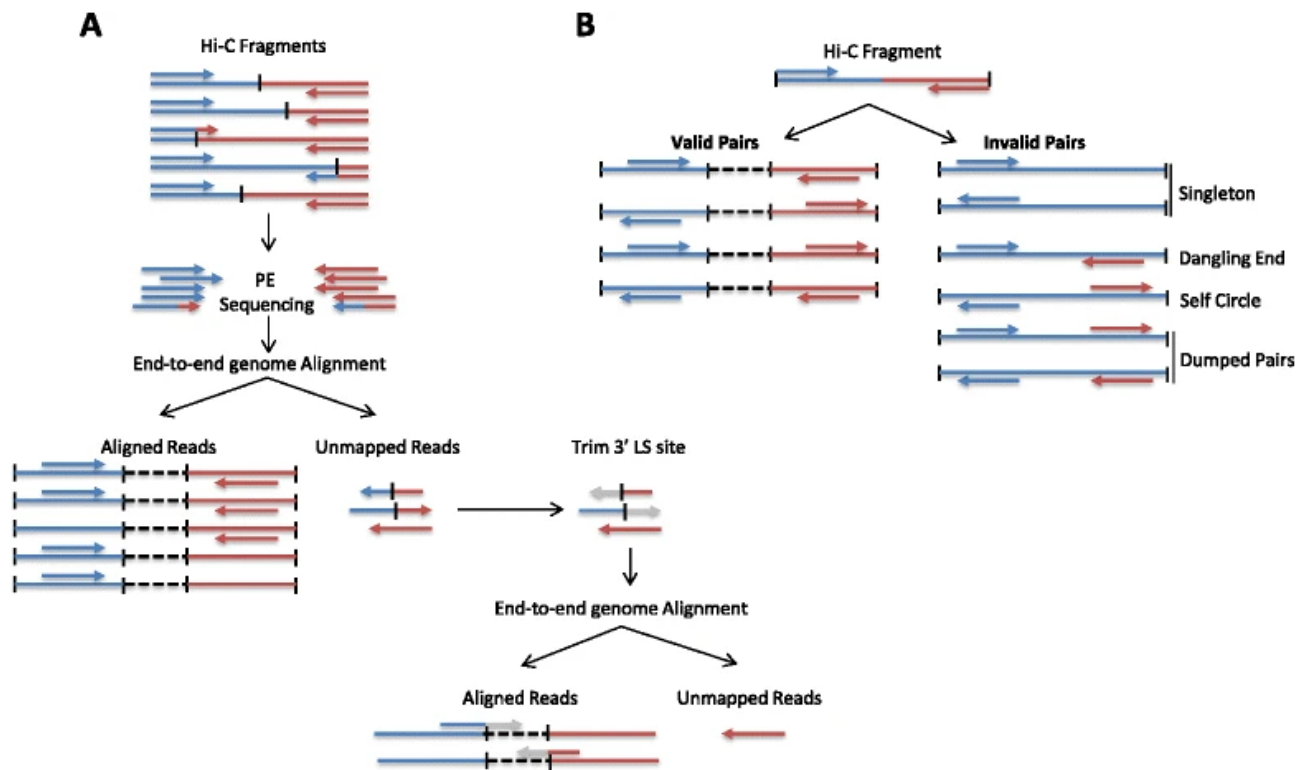


Figure 30: **Alignement de reads HiC et HiChIP par HiCpro.**, figure tirée de Servant et al. (2015). (A) Processus d'alignement d'HiC-Pro. (B) Identifications des paires valides et des paires invalides suite à l'alignement.

d'un alignement local. Ces alignements seront effectués par l'outil bioinformatique initialement implémenté pour aligner les reads de données HiC: HiCpro (Servant et al., 2015) qui utilise le logiciel Bowtie2 pour effectuer ces différents alignements. Les reads sont donc en premier lieu alignés sur le génome de référence de manière classique (type alignement Burrows-Wheeler Algorithm ou BWA). Les reads couvrant les jonctions de ligation des fragments qui ne sont pas alignés lors de la première étape sont donc coupés au niveau du site de ligation et l'extrémité 5' de ces reads est alors réalignée sur le génome de référence (alignement local) (Figure 30A). Une fois ces deux étapes réalisées, seuls les reads uniques et mappés sur un fragment de restriction (correspondant à l'enzyme de restriction utilisée lors de la coupure enzymatique) seront conservés. Seuls les paires considérées comme valides (R1 dans un anchor et R2 dans l'autre ou inversement) seront conservés. Ces paires valides sont identifiées grâce à la présence de deux sites de restriction (un dans chaque anchor). Les paires qui sont assignées à un seul et même enzyme de restriction sont classifiées en tant qu'auto circularisation du fragment ou en tant qu'extrémité non finie (dangling-end) et seront filtrées tout comme les fragments ou seuls un des deux reads à été mappé (singleton) (Figure 30B).

Filtration des reads

Une fois les données de séquençage alignées contre le génome de référence hg19, des étapes de filtrations sont alors nécessaires afin de ne conserver que les reads de qualité suffisante pour l'opération. Deux suites d'outils permettent d'effectuer diverses filtration: Samtools (Li et al., 2009) et GATK (McKenna et al., 2010)).

Ainsi une fois l'alignement achevé, les reads sont filtrés selon leurs score de mapping, à chaque alignement, un score est attribué qui représente la qualité du read mais également la qualité de son alignement. On considère que les reads mappés avec une qualité inférieure à 30 ne sont pas suffisants et ces reads seront supprimés de même que les reads non mappés sur le génome.

Ensuite, j'ai décidé de retirer les reads mappant dans les régions définies par le consortium ENCODE comme étant des régions blacklist, c'est à dire des régions télomériques, centromériques et autres séquences répétées tout au long du génome (Landt et al., 2012a).

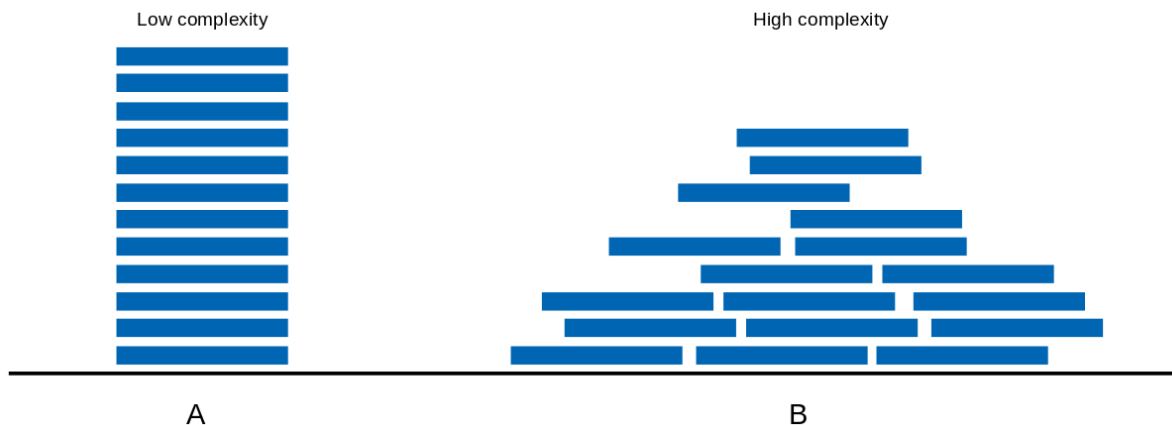


Figure 31: **La complexité de la librairie de séquençage est importante pour la qualité de l'analyse**, A: Les reads mappant exactement à la même position appauvrissent la complexité de la librairie et sont souvent des duplicats de PCR; B: Complexité attendue lors d'un séquençage normal.

Enfin, il peut y avoir dans les données des reads dits dupliqués qui sont dûs à des duplicats de PCR qui contribuent à appauvrir la complexité de la librairie. L'outil picard markDuplicates de GATK a donc été utilisé pour taguer ces reads et les retirer des fichiers (Figure 31 A et B).

une fois l'ensemble de ces filtrations effectuées (applicables essentiellement pour le ChIP-seq et l'ATAC-seq). On peut donc passer aux analyses plus ciblées dites de post-traitement des données.

6.3.3 Analyse post-alignement des données de ChIP-seq

Identification et sélection de pics

Ce traitement s'applique aux données de séquençage ATAC-seq et ChIP-seq. Les reads ainsi obtenus par ces techniques et alignés sur le génome de référence forment deux pics dont chacun correspond à un brin sur l'ADN. Ces deux pics entourent donc le vrai site de fixation de la protéine ou du nucléosome d'intérêt. Dans le cas du ChIP-seq, il est important d'avoir réalisé en même temps un contrôle généralement appelé Input procédé de la même manière que les échantillons test à l'exception de l'immuno-précipitation réalisée pour enrichir la protéine fixée. Cet Input va nous permettre de calculer la distribution du bruit de fond. Pour déterminer la localisation de nos protéines/nucléosomes d'intérêt, nous utilisons une Analyse basée sur un modèle (MACS) (Zhang et al., 2008). MACS introduit une méthode plus sophistiquée que d'autres algorithmes de sélection de pics pour modéliser la taille du fragment (obtenue après coupure enzymatique ou sonication). Le fragment ChIP-ADN doit être également séquençé des deux côtés ce qui permet de prédire que la densité de reads autour d'un vrai site de fixation doit montrer un motif d'enrichissement bimodal.

Il est donc possible de préciser en paramètre la taille de sonication utilisée (bandwidth) et un fold-enrichment de haute confiance. MACS va parcourir le génome avec des fenêtres de longueur $2 \times$ bandwidth pour trouver des régions avec une densité de reads supérieure au fold-enrichment relatif à une distribution génomique aléatoire des reads (Algorithme 1).

```

1 begin
2   Slide a window of  $2 \times$  bandwidth across genome;
3   Identify regions of moderate enrichment (mfold: 10-30 fold);
4   for each peak  $i$  of 1000 randomly chosen enriched regions do
5     Separate reads into + and - strand;
6     Calculate mode of + and - strand;
7      $d_i \leftarrow |\text{mode}_+ - \text{mode}_-|$ ;
8   end
9    $d \leftarrow \text{average}(d_i)$ ;
10 end
```

Algorithm 1: Estimation de la taille du fragment pour la détection des pics ChIP-seq et ATAC-seq.

Les reads de séquençage ChIP-seq sur les deux brins de l'ADN sont décalés en raison de la présence du facteur de transcription. Pour recentrer ces reads et identifier le bon sommet du pic, on décale les reads d'une distance équivalente à la moitié de la taille du fragment estimé vers leur extrémité 3' (Figure 32). Une fois les pics recentrés, un test statistique utilisant la loi de Poisson est utilisé pour éviter les biais. MACS utilise un paramètre dynamique, λ_{local} , défini pour chaque pic candidat comme:

$$\lambda_{local} = \max(\lambda_{BG}, \lambda_{1k}, \lambda_{5k}, \lambda_{10k})$$

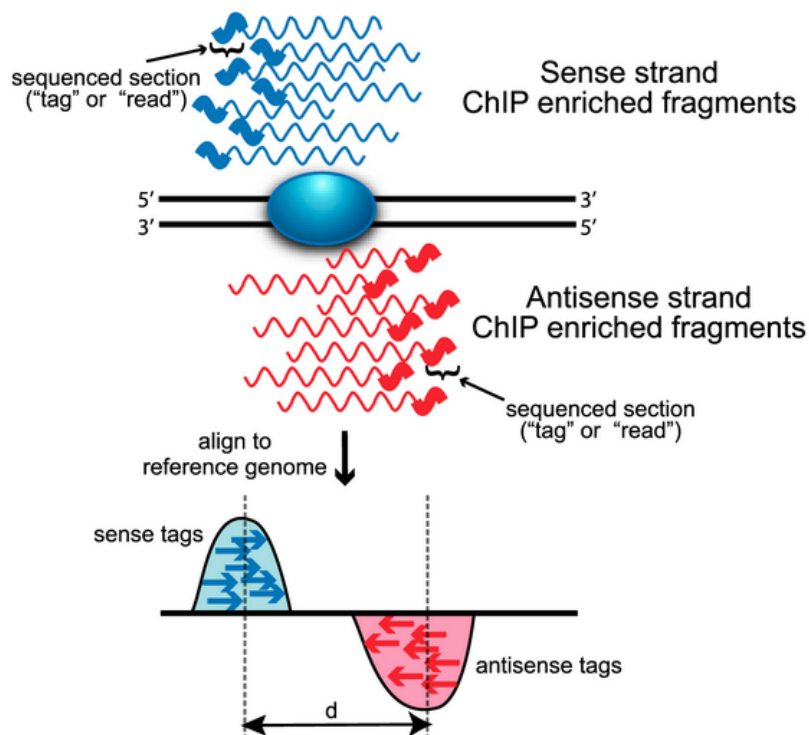


Figure 32: **Détection des pics ChIP-seq et ATAC-seq et shift réalisés pour combiner les données reads du brin + et du brin -, Figure issue de Zhang et al. (2008)**

Le paramètre λ_{BG} est calculé sur le génome entier et les autres sont calculés à partir de fenêtres de 1kb, 5kb et 10kb centrées sur la localisation du pic dans l'échantillon contrôle (Input). Le paramètre λ_{local} permet de réduire l'influence de biais locaux. MACS utilise ce paramètre pour estimer la p-value de chaque pic candidat. Les pics avec une p-value inférieure à un seuil défini de p-value (généralement 10^{-5}) ou de q-value ($FDR < 0.05$) sont conservés (Distribution de Poisson). Enfin, pour chacun des pics, le ratio entre le compte des reads ChIP-seq et le λ_{local} est reporté comme le fold enrichment.

Modèle de Markov caché: Description et applications

En biologie, l'analyse des séquences nucléotidiques et l'intégration de différentes données sont deux problématiques majeures dans notre compréhension du génome. Dans l'optique de la poursuite d'analyse post-alignement des données de ChIP-seq et ATAC-seq, nous allons détailler ici le modèle statistique des chaînes de Markov cachées (Eddy, 2004) qui seront utilisées pour l'annotation fonctionnelle d'une lignée. Le Hidden Markov Model (HMM) également appelé chaîne de Markov est caractérisé par deux composantes: les états et les transitions. Le modèle possède également un état initial ou début avec une certaine probabilité de passer dans chacun des états (la somme de

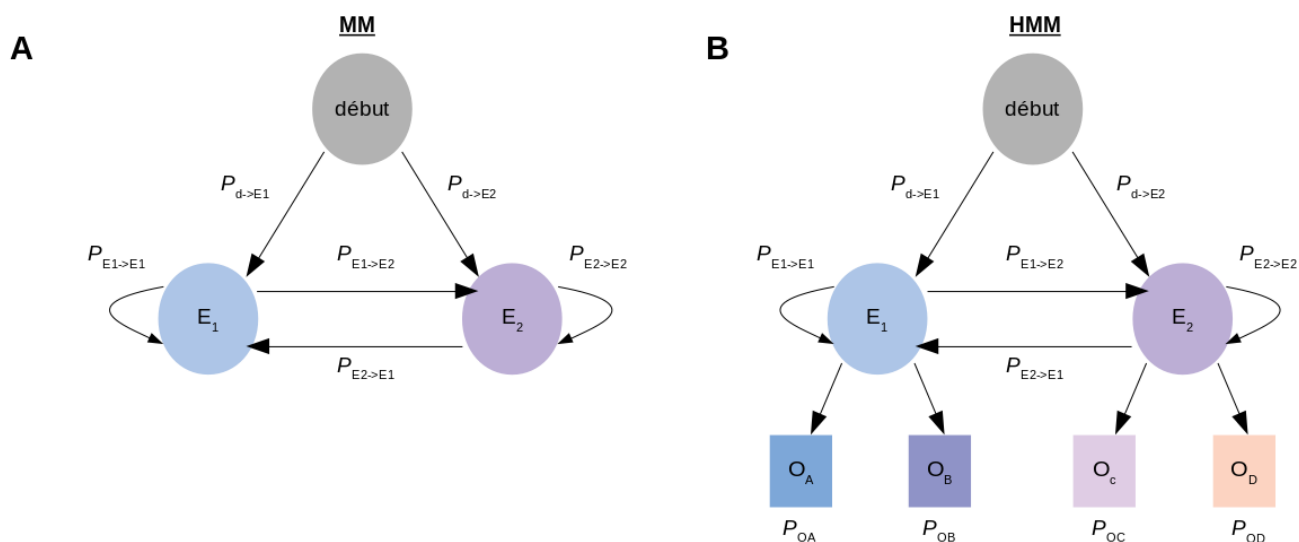


Figure 33: Schématisation d'un modèle de Markov (MM) (A) et d'un modèle de Markov caché (HMM) (B).

ces probabilités étant égale à 1). La transition représente la probabilité de passage de l'état 1 à l'état 2 (Figure 33A). Dans cet exemple de chaîne de Markov à deux états, la transition de l'état 1 ne peut se faire que vers l'état 2 ou vers lui-même. les probabilités de transitions sont donc toujours égales à 1 (exemple l'état E_1 à une probabilité de transition de 0.6 vers l'état E_2 et donc une probabilité de transition de 0.4 vers son propre état). Dans un modèle de Markov caché (Figure 33B), on considère que l'on ne connaît pas les caractéristiques des différents états, on dit qu'il sont donc cachés. A la place de ces états, le modèle repose sur l'émission d'observations pour chacun des états (dans la Figure 33B, il y a une certaine probabilité d'observer A et une autre d'observer B pour l'état E_1). Un modèle de Markov caché repose donc sur les transitions entre les états et les différentes émissions des observations de ces états pour établir le modèle. Je détaille ensuite comment ce modèle de Markov caché peut être appliqué dans une application concrète: chromHMM qui va permettre de définir des états de chromatine spécifique à une lignée cellulaire grâce à l'implémentation d'un modèle HMM.

Definition des états chromatinien

L'outil développé par Ernst et son équipe (Ernst and Kellis (2010) et Ernst and Kellis (2012)), implémente le modèle de Markov afin de définir dans un modèle cellulaire des états chromatinien fonctionnels (Figure 34).

Le génome va être intégralement découpé en plusieurs fenêtres de 200 paires de bases chacune et l'ensemble des marques d'histones étudiées va être "binarisée" (c'est à dire marquée comme 1 si elle est présente dans la fenêtre ou bien 0 si elle en est absente). Cette binarisation repose sur une distribution de Poisson. Une fois l'ensemble du génome binarisé, les paramètres d'émission et de transition du modèle sont calculés grâce à une phase d'apprentissage où le modèle sera affiné

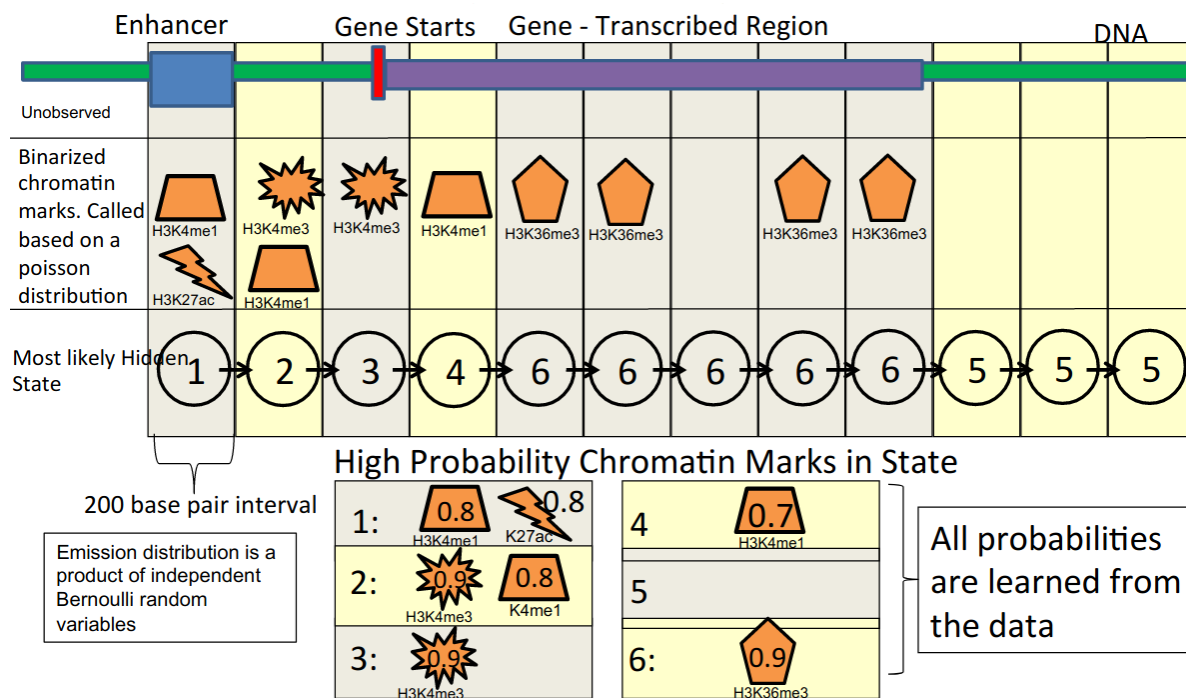


Figure 34: **Application des HMM pour l'annotation fonctionnelle du génome**, (Ernst and Kellis (2010) et Ernst and Kellis (2012)).

à chaque itération pour définir avec précision les états de chromatine attendus. Enfin, à chaque état sera associée la probabilité de présence de chacune des marques d'histones. L'interprétation de ces états nécessite par la suite une connaissance biologique du rôle fonctionnel des marques d'histones (exemple: dans la figure 34, l'état 1 qui a une très forte probabilité de présence des marques H3K4me1 et H3K27ac correspond à un état d'enhancers forts).

J'ai réalisé une annotation fonctionnelle du génome MM.1S grâce aux données de ChIP-seq MM.1S et à l'outil bioinformatique ChromHMM. Cinq marques d'histones disponibles sur le consortium ENCODE (Consortium et al., 2012) ont été utilisées: H3K4me1 qui définit une région régulatrice de type enhancer, H3K4me3 qui définit une région promotrice de type promoteur, H3K27ac qui indique une région régulatrice active, H3K27me3 qui définit une région régulatrice inactive et H3K36me3 qui correspond à un état de transcription active.

Recherche et découverte de motif

La recherche de motif est particulièrement utilisée pour des données de séquençage ChIP-seq de facteurs de transcription et éventuellement pour des données d'accessibilité de chromatine (ATAC-seq). Le but de cette recherche est d'identifier dans des séquences nucléotidiques d'intérêt des motifs (ou séquences) particulièrement enrichis. Cela permet par exemple d'identifier pour un facteur de transcription les sites où celui-ci se fixe de manière directe à l'ADN mais aussi d'éventuels motifs fixant traditionnellement un autre facteur de transcription et qui pourraient indiquer une

co-interaction et donc une fixation indirecte à l'ADN. Pour identifier ces motifs liés aux données de ChIP-seq, j'ai utilisé la suite MEME (Bailey et al., 2015) et particulièrement le pipeline MEME-ChIP (Machanick and Bailey, 2011) pour effectuer une recherche de motifs complète dans les données de ChIP-seq.

Ce pipeline réalise divers types d'analyses sur les régions ChIP-seq données en entrée: découverte de motifs *de novo* dû à la fixation du facteur de transcription, enrichissements de motif déjà connus et répertoriés dans des bases de données (ici j'ai utilisé la base de donnée HOCOMOCO Kulakovskiy et al. (2018)), visualisation des motifs et localisation dans les régions données, enrichissement central des motifs *de novo* et déjà connus. MEME-ChIP utilise pour cela deux algorithmes : MEME qui utilise les modèles de maximisation des espérances (EM) pour découvrir des motifs de fixation de facteurs de transcription simples ou de complexes et DREME qui utilise un modèle non probabiliste pour décrire des motifs de fixation courts caractéristiques d'un seul facteur de transcription.



Figure 35: Exemple de résultats produit par le pipeline MEME-ChIP (Ma et al., 2014).

La comparaison de ces nouveaux motifs avec ceux de bases de données existantes est permise par l'algorithme TOMTOM qui va aligner ces motifs avec ceux des bases de données pour identifier avec précision l'identité des motifs.

Finalement, l'algorithme CentriMo permet de déterminer l'enrichissement central de l'ensemble de ces motifs (*de novo* et déjà connus) au sein des régions ChIP-seq données en entrée (Figure 35).

Identification des super-enhancers et circuits de régulations

Pour l'analyse du ChIP-seq de la marque H3K27ac correspondant aux régions régulatrices actives, je me suis intéressé à la détection de régions marquant l'identité cellulaire: les super-enhancers (Pott and Lieb, 2015). Pour ce faire, j'ai utilisé le logiciel ROSE (Whyte et al. (2013) Lovén et al. (2013)).

Ce logiciel permet de classer l'ensemble des régions ChIP-seq données en entrée selon la densité de signal ChIP-seq (Figure 36). Le point d'inflexion de la courbe obtenue, définie géométriquement, est utilisé pour définir le cut-off de détermination des super-enhancers. De plus les enhancers sont regroupés selon leurs distance mutuelles: si deux enhancers sont retrouvés à moins de 12.5kb l'un de l'autre, alors ils seront regroupés. Finalement, les enhancers localisés à moins de 2kb autour d'un TSS sont exclus pour éviter d'inclure les régions promotrices. Les super-enhancers sont donc des régions très enrichies en marque d'histone H3K27ac, regroupant un cluster d'enhancers individuel et éloigné suffisamment d'un site d'initiation de la transcription.

Afin de déterminer les circuits de régulations spécifiques de la lignée cellulaire étudiée, j'ai utilisé

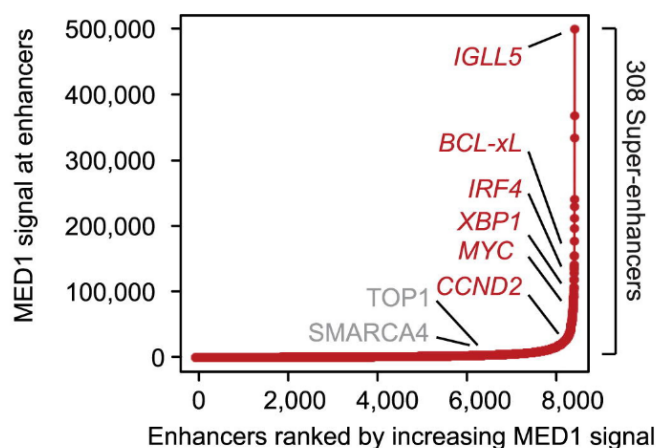


Figure 36: **Identification de super-enhancers à l'aide de données de ChIP-seq avec ROSE**, ChIP-seq réalisé sur le facteur de transcription MED1 (Love et al., 2017).

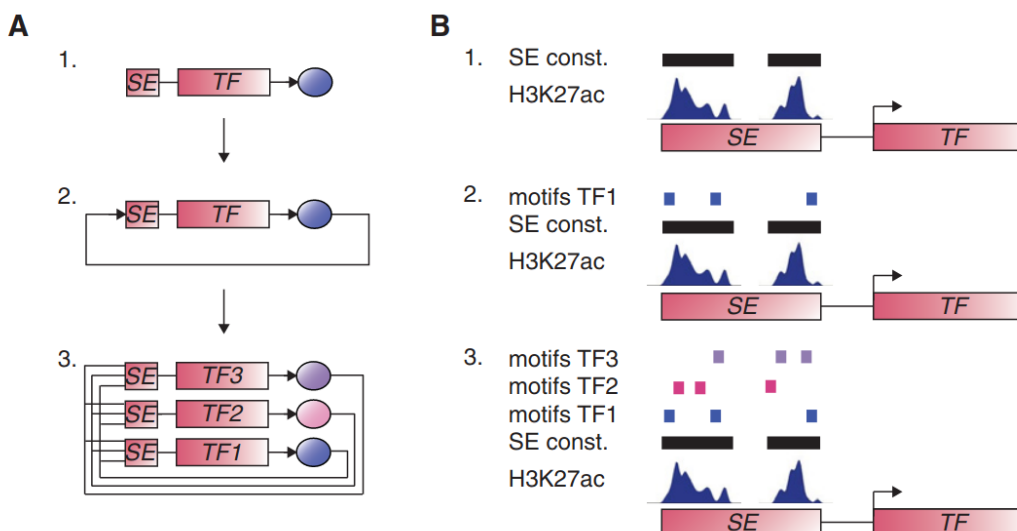


Figure 37: **Définition des circuits de régulation des facteurs de transcription**, figure issue de Saint-André et al. (2016).

un autre outil développé dans le laboratoire de Young: crcMapper (Saint-André et al. (2016) et Hnisz et al. (2013)). Cet outil est dépendant de l'identification des super-enhancers. La première étape consiste à associer les super-enhancers aux gènes codant pour un facteur de transcription proche (Figure 37A1 et B1). Une fois l'assignation effectuée, une recherche de motif de ce facteur de transcription dans le super-enhancer associé permet d'affirmer que le facteur de transcription auto-régule la transcription de son propre gène (Figure 37A2 et B2). Finalement, une liste de facteurs de transcription cible est établi et la présence des motifs dans les super-enhancers permettent de finaliser les circuits de régulations (Figure 37A3 et B3).

6.3.4 Analyse post-alignement des données d'ATAC-seq

Une subtilité dans le traitement des données ATAC-seq alignées doit impérativement être prise en compte. Du fait de l'intégration de l'enzyme transposase Tn5 dans le génome, lors de son intégration et de la réparation de l'ADN qui s'en suit, un décalage de 9 paires de bases se fait sur les sites d'insertions de l'enzyme (Viswanadham et al., 2019). Pour ce faire j'utilise samtools Li et al. (2009) pour corriger ce biais en déplaçant la position des reads mappés de +4 et -5 sur le brin positif et négatif respectivement.

Positionnement de nucléosomes et footprinting

Une fois la correction Tn5 effectuée, le séquençage paired-end ATAC-seq permet d'identifier plusieurs sortes de régions tout au long du génome en se basant sur la taille des fragments obtenus et la longueur d'une chaîne ADN contenant un nucléosome (146 paires de bases) (Buenrostro et al., 2013): des régions accessibles sans nucléosome d'une distance inférieure à 150 paires de bases, des régions contenant un seul nucléosome (mono) pour des fragments allant de 151 à 300 paires de bases et des régions contenant plus de deux nucléosomes pour des fragments de plus de 300 paires de bases (Figure 38). Pour les régions complètement libres en nucléosome, des approches

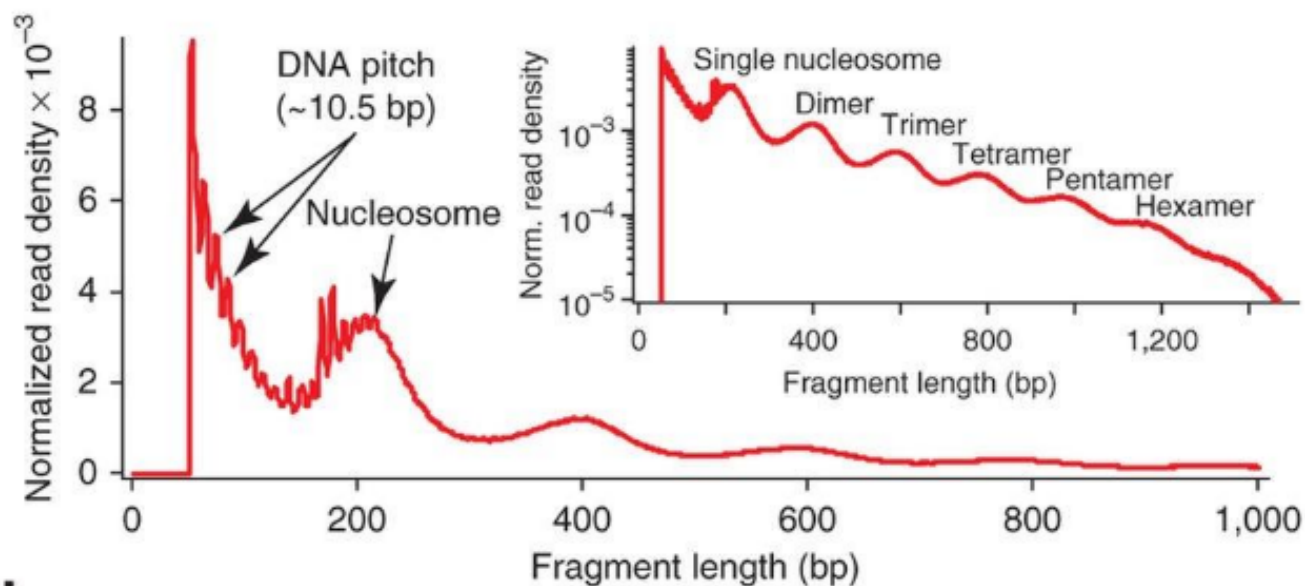


Figure 38: **Composition en nucléosome selon la taille des fragments ATAC-seq séquencés**, figure issue de Buenrostro et al. (2013).

de footprinting suggèrent que les données ATAC-seq permettent de déterminer la présence d'un facteur de transcription selon deux critères: la présence dans une région d'un motif de liaison à l'ADN et une chute de la densité de reads ATAC-seq au niveau de ce motif suggérant la présence de la protéine fixée à l'ADN. Pour réaliser cette analyse, j'ai utilisé l'outil PIQ-single (Thashim, 2017).

6.3.5 Analyse différentielle des données RNA-seq

Pour l'analyse des reads RNA-seq, je me suis uniquement intéressé aux différentiels d'expression entre la condition MM1S EtOH et la condition Dex.

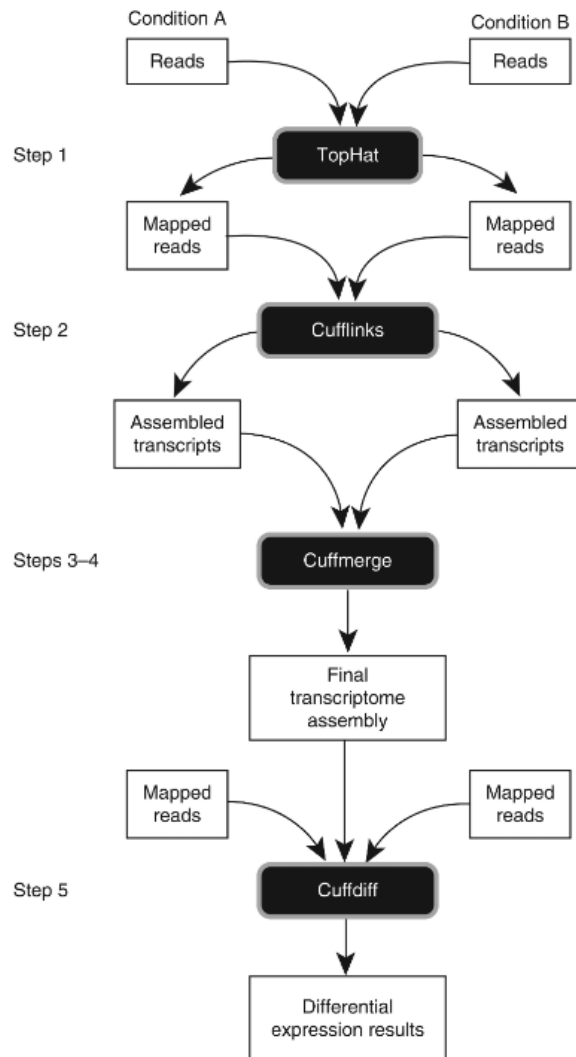


Figure 39: **Pipeline d'analyse différentielle de données RNA-seq**, figure issue de Trapnell et al. (2012).

Après alignement (cf. 6.3.2), j'utilise Cufflinks (Figure 39), un autre outil conçu par les équipes de Trapnell (Trapnell et al., 2012). Cufflinks utilise des données d'épissage connues pour les gènes dans le but de reconstituer les différents isoformes d'un même gène. Cufflinks va ainsi pouvoir quantifier avec parcimonie chaque fragment de transcription puis utilise un "meta-assembleur" nommé Cuffmerge pour créer un assemblage du transcriptome final en regroupant tous les transcrits générés. Finalement en réutilisant les reads mappés des différentes conditions et le transcriptome final généré, Cuffdiff va pouvoir calculer pour chaque échantillon l'expression des gènes et différents

isoformes. Les valeurs d'expression sont alors normalisées par calcul du fragment per kilo million (FPKM) calculé par l'équation suivante:

$$\begin{aligned} \text{Scaling Factor : } PM &= \frac{\text{Nombre de fragments Totaux}}{10^6} \\ FPM &= \frac{\text{Nombre de fragments mappés sur le gene}}{PM} \\ FPKM &= \frac{FPM}{\text{Taille du gene} * 10^3} \end{aligned}$$

Une fois les valeurs d'expressions normalisées obtenues, la variance de chaque transcrit est estimée et le log2 foldchange (FC) de chaque transcrit est estimé avec sa p-value et son false discovery rate (FDR) associé (correction par multi-test). On obtient donc finalement pour chacun des transcrits, sa valeur d'expression normalisée dans chaque échantillon (avec ses intervalles de confiance) ainsi que le différentiel calculé entre toutes les différentes conditions.

6.3.6 Analyse post-alignement des données de HiChIP-seq

Matrice de contacts

Le logiciel HiC-pro (Servant et al., 2015) utilisé pour effectuer l'alignement des reads HiChIP (cf. 6.3.2) permet également de construire les matrices de contact de l'ADN. Les paires de reads considérées comme valides sont utilisées pour construire ces matrices. Le génome va être découpé en différentes tailles pour construire différentes matrices. Ainsi on utilise des fenêtres génomiques d'une longueur de 20000pb, 40000pb, 150000pb, 500000pb et enfin 1000000pb. Ces différentes longueurs vont ainsi définir la résolution des contacts, ainsi, un découpage plus petit augmentera la précision des contacts détectés. Le nombre de contacts observés à l'intérieur d'une fenêtre est calculé.

Finalement, un algorithme de normalisation de ces matrices appelé normalisation ICE (Imakaev et al., 2012) est appliqué pour corriger les éventuels biais liés aux contacts de haute fréquence de certains loci par rapports à l'ensemble du génome.

Identifications des boucles chromatinienne

L'immunoprécipitation réalisée lors de la technique du HiChIP impose une différence dans l'identification des boucles chromatinienne sur le génome. J'utilise l'outil Hichipper (Lareau and Aryee (2018)) qui est spécifique de ce genre de données.

Celui-ci nécessite en entrée les matrices d'interactions générées par HiC-pro ainsi que la position des sites de l'enzyme de restriction utilisée (MboI dans notre cas) pour générer un contrôle qualité des données ainsi que l'identification des boucles et des anchors par un peak calling corrigé du biais provoqué par la position de pics au niveau de sites d'enzyme de restriction.

Pour identifier les différents anchors, Hichipper identifie donc les différents pics et étend l'anchor jusqu'aux sites d'enzyme de restriction les plus proches de chaque côté (Figure 40 haut). Lorsqu'un pic est localisé au niveau d'un site de restriction, hichipper étend l'anchor jusqu'aux sites les plus proches de chaque côté du site de restriction (Figure 40 bas).

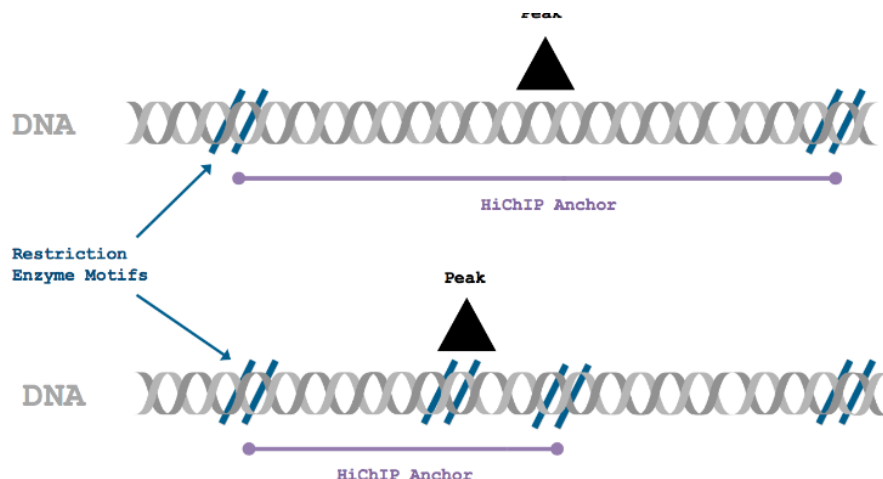


Figure 40: Définition des *anchors* liés aux pics HiChIP et aux sites de restrictions d'enzymes, selon Lareau and Aryee (2018).

6.3.7 Analyses bioinformatiques diverses

D'autres analyses bioinformatiques ont été réalisées au cours de cette thèse mais sont indépendantes du type de données initiales. c'est le cas des analyses ontologiques ou simplement du traitement de données qui seront détaillés ici.

Analyses ontologiques

L'analyse ontologique de régions génomiques permet d'associer à ces régions une fonctionnalité biologique. Pour ce faire j'ai utilisé l'outil en ligne GREAT (McLean et al., 2010) proposé par Stanford. GREAT définit par défaut un domaine de régulation basal à chaque gène (Figure 41). Ce domaine s'étend à 1kb en amont du site d'initiation de la transcription et à 5kb en aval du gène indépendamment des autres gènes. En plus, le domaine de régulation du gène est étendu jusqu'à 100kb vers le domaine de régulation le plus proche.

Chaque région est donc ensuite assignée aux différents domaines de régulation dans lesquels elle est localisée, et par des méthodes computationnelles, notamment un test utilisant une loi binomiale sur les régions et un autre une loi hypergéométrique sur les gènes, va identifier des voies de signalisations spécifiques du set de région donné. GREAT référence de nombreuses bases de données telles que GO (Genome Ontology), Panther ou encore MSigDB.

Manipulation des données

L'ensemble des données génomiques est manipulé grâce à une suite d'outils divers: Bedtools (Quinlan and Hall, 2010). Cette suite permet de nombreuses actions comme par exemple intersecter les positions génomiques de deux fichiers différents pour identifier leurs co-localisations ou à l'inverse les régions uniques. Pour la manipulation et la visualisation des données ChIP-seq et ATAC-seq



Figure 41: **Définition des domaines de régulation des gènes dans GREAT**, selon McLean et al. (2010).

essentiellement, j'ai utilisé la suite d'outils Deeptools (Ramírez et al., 2014). Cette suite permet de générer les fichiers aux formats BigWig pour faciliter leur visualisation. Les fichiers BigWig ont également été normalisés avec la méthode dite de la normalisation 1x (ou RPGC pour Read Per Genomic Content). Cette normalisation est effectuée sur des fenêtres génomiques d'une taille de 200 paires de bases. Les reads localisés dans ces fenêtres sont comptabilisés puis normalisés avec le facteur 1x.

Les fichiers générés vont servir ensuite à générer des heatmaps pour visualiser les signaux sur l'ensemble du génome ou sur des localisations spécifiques ou bien encore comparer deux conditions l'une versus l'autre. Ces fichiers sont également utilisés par deux génomes viewer: IGV et WashU epigenome Viewer (Robinson et al. (2011) et Li et al. (2019a)).

Finalement, Deeptools permet également de créer des tables de comptage pour effectuer des analyses différentielles entre plusieurs conditions.

Une autre suite d'outils a été fréquemment utilisée notamment pour réaliser l'annotation génomique des pics ATAC-seq et ChIP-seq mais également des interactions chromatinienne obtenues par HiChIP. La suite d'outils HOMER (Heinz et al., 2010) permet d'annoter ces régions en les catégorisant selon l'annotation de référence dans lesquelles elles sont localisées (régions intergéniques, promoteurs, introns, exons...). Il est également possible d'obtenir la distance de ces régions par rapport au site de transcription le plus proche et de connaître le gène le plus proche.

HOMER permet également d'effectuer une recherche de motifs de novo sur des régions de tailles variables et non centrées sur la présence d'un facteur de transcription ce qui permettra de faire des recherches de motifs rapides sur les données telles que les pics ATAC-seq ou encore les ancres des interactions chromatinienne. L'identification de toutes les positions d'un motif précis dans les régions génomiques est également possible avec cette suite d'outils.

6.3.8 Analyses spécifiques au séquençage en cellule unique

Les analyses single-cell RNA-seq et Multiome sur la lignée MM.1S ont été effectuées par un autre doctorant du laboratoire: Jonathan Cruard.

Traitement global des données

Les données de séquençage en cellule unique ont été démultiplexées et alignées sur le génome de référence GRCh38 (hg38, Schneider et al. (2017)) avec la suite d'outils cellranger pour le single-cell RNA-seq (Zheng et al., 2017) et cellranger-arc pour le séquençage scMultiome (Satpathy et al., 2019). Les matrices de comptage ainsi obtenues ont ensuite été chargées dans R avec le logiciel Seurat (Satija et al., 2015) et Signac (Stuart et al., 2020) pour la partie ATACseq. Les cellules détectées par Cellranger sont alors filtrées selon de multiples contrôles qualité.

Pour le RNAseq: le nombre d'UMI doit être compris entre 8,000 et 150, 000, le nombre de *Features* doit être supérieur à 2,000, la proportion de reads mitochondriaux doit être comprise entre 5 et 30% et enfin le pourcentage de long non-coding RNA doit être inférieur à 8%.

Pour l'ATACseq: le nombre d'UMI doit être compris entre 10,000 et 500,000, le nombre de *Features* doit être compris entre 5,000 et 60,000 l'enrichissement des TSS doit être supérieur à 3.5 et inférieur à 15, le signal correspondant au nucléosome ne doit pas être supérieur à 2 et la fraction de reads dans les pics doit être supérieure à 35%.

Estimation des ratios de gènes répondants à la Dex (RRG) et des ratio de régions d'ouvertures (ROR)

- **Ratio de gènes répondants à la Dex (RRG):** Pour chaque gène, nous avons calculé le 80ème percentile des valeurs d'expression dans les cellules non traitées. Nous avons ensuite calculé pour chaque cellule le RRG comme étant le pourcentage de gènes Dex-activés (scDAGs) avec un seuil supérieur à la valeur d'expression du gène à 4 heures et 24 heures après traitement à la Dex et en condition non traitées.
- **Ratio d'ouverture de région (ROR):** De la même façon, en se basant sur les données single-cell ATAC, nous avons calculé le ROR. Pour chaque cellule, nous avons estimé le pourcentage de régions avec une valeur supérieur au seuil contrôle (les seuils sont spécifiques à chaque région et définis comme le 80ème percentile des valeurs non nulles, suivant ce qui a été fait pour les données scRNA-seq).

6.3.9 Pipelines d'analyse bioinformatique

Au cours des ces années de thèse, j'ai eu la chance de manipuler toutes sortes de données de

séquençage haut débit, malgré l'existence de nombreux outils permettant l'analyse et la manipulation de ce genre de données, l'un des objectifs de ces travaux était de mettre en place au sein de l'équipe des pipelines d'analyses bioinformatiques permettant l'analyse automatisée de toutes sortes de données de séquençage. Ces pipelines ont tout d'abord été codés en Python avant d'être récemment implémentés en Snakemake et Conda. Ainsi je vais présenter ici les pipelines permettant le traitement des données de ChIP-seq, d'ATAC-seq et de HiChIP-seq. L'ensemble des pipelines est disponible sur mon gitlab :

<https://gitlab.univ-nantes.fr/crcina-eq11/bchippipeline>,

<https://gitlab.univ-nantes.fr/crcina-eq11/batacpipeline>,

<https://gitlab.univ-nantes.fr/crcina-eq11/bhichippipeline>

Pipelines d'analyse ChIP-seq et ATAC-seq

Le premier pipeline conçu est celui d'analyse de données de ChIP-seq (Figure 42). Celui-ci automatise le traitement des données brutes (.fastq). Du fait que la plupart de nos données ne sont pas en duplicats, nous ne pouvons pas vraiment utiliser les méthodes d'assertions des pics proposées par le consortium ENCODE (Landt et al., 2012b).

- **Contrôle qualité des reads (pre-processing):** La qualité du séquençage est examinée avant et après trimming par Fastqc, le trimming est réalisé par Trimmomatic et pour les données de séquençage paired-end, seuls les paires de reads entiers sont conservés.
- **Alignement des reads:** L'alignement global est réalisé par Bowtie2 avec des paramètres par défaut (`-sensitive -X 2000`, sans mismatch autorisés) avec le génome de référence indexé HG19. Les reads non mappés ou ayant une qualité de mapping inférieure à 30, dupliqués et mappant dans des régions répétées blacklist définies par ENCODE (Amemiya et al., 2019) seront filtrés pour l'analyse.
En ce qui concerne l'alignement des reads ATAC-seq une étape supplémentaire est ajoutée. Il s'agit du décalage Tn5 des reads pour corriger le biais dû à l'insertion de la transposase.
- **Peak Calling:** Il est réalisé par Macs2 avec ses paramètres par défaut. Seuls les pics ayant un FDR inférieur à 0.05 seront conservés pour la suite.

L'étape de post processing n'est à ce jour pas encore implémentée dans ces deux pipelines car elle dépend essentiellement des questions biologiques auxquelles souhaite répondre l'utilisateur. Il est néanmoins envisagé d'intégrer une recherche de motif lorsqu'il s'agit de ChIP-seq facteur de transcription, une annotation génomique des pics et également la génération des fichiers BigWig pour la visualisation.

Pipeline d'analyse HiChIP-seq

Le pipeline d'analyse HiChIP a directement été implémenté en Snakemake lors de mon séjour au DKFZ à Heidelberg (Allemagne) pour apprendre la technique du HiChIP. Il repose majoritairement sur les deux outils HiC-pro et Hichipper et est également divisé en trois grandes étapes (Figure 42):

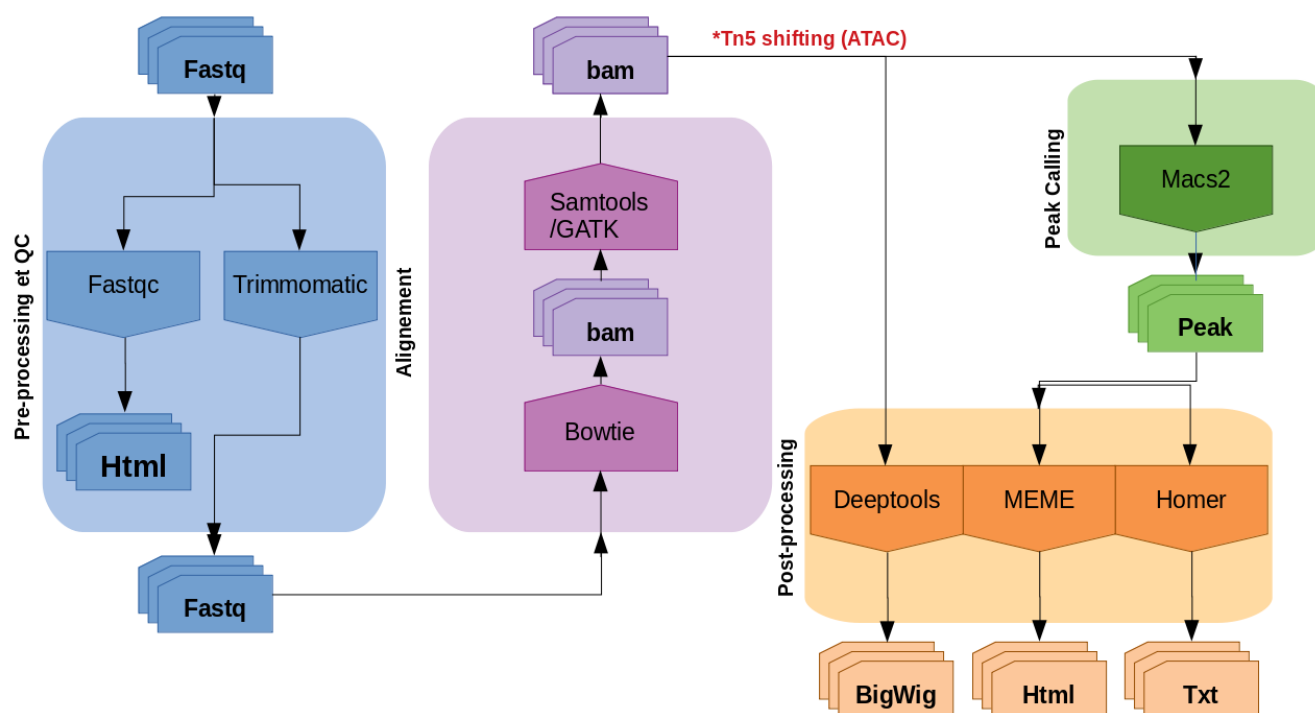


Figure 42: Pipeline d'analyse de données ChIP-seq et ATAC-seq mis en place.

- **Contrôle qualité des reads (pre-processing):** La qualité du séquençage est examinée avant et après trimming par Fastqc, le trimming est réalisé par Trimmomatic et pour les données de séquençage paired-end, seul les paires de reads entiers sont conservées.
- **Alignement et Construction des matrices de contacts:** HiC-pro est utilisé en mode séquentiel avec pour chaque étape des contrôles qualité qui permettent à l'utilisateur de vérifier les résultats de chaque étape: Tout d'abord une étape d'alignement, une de regroupement des échantillons et de construction des matrices de contacts et finalement une étape de normalisation ICE des matrices.
- **Identification des boucles:** Hichipper est utilisé avec trois fichiers en entrée: le fichier de pic ChIP-seq d'intérêt, la sortie d'HiC-pro et la position des sites d'enzyme de restriction. Cela va générer les fichiers des interactions et de leurs anchors ainsi que des fichiers permettant la visualisation

Une fois encore, ce pipeline va être amélioré pour permettre son utilisation sans avoir besoin d'un fichier de pics pour la partie identification des boucles et en réalisant une annotation génomique des interactions par défaut.

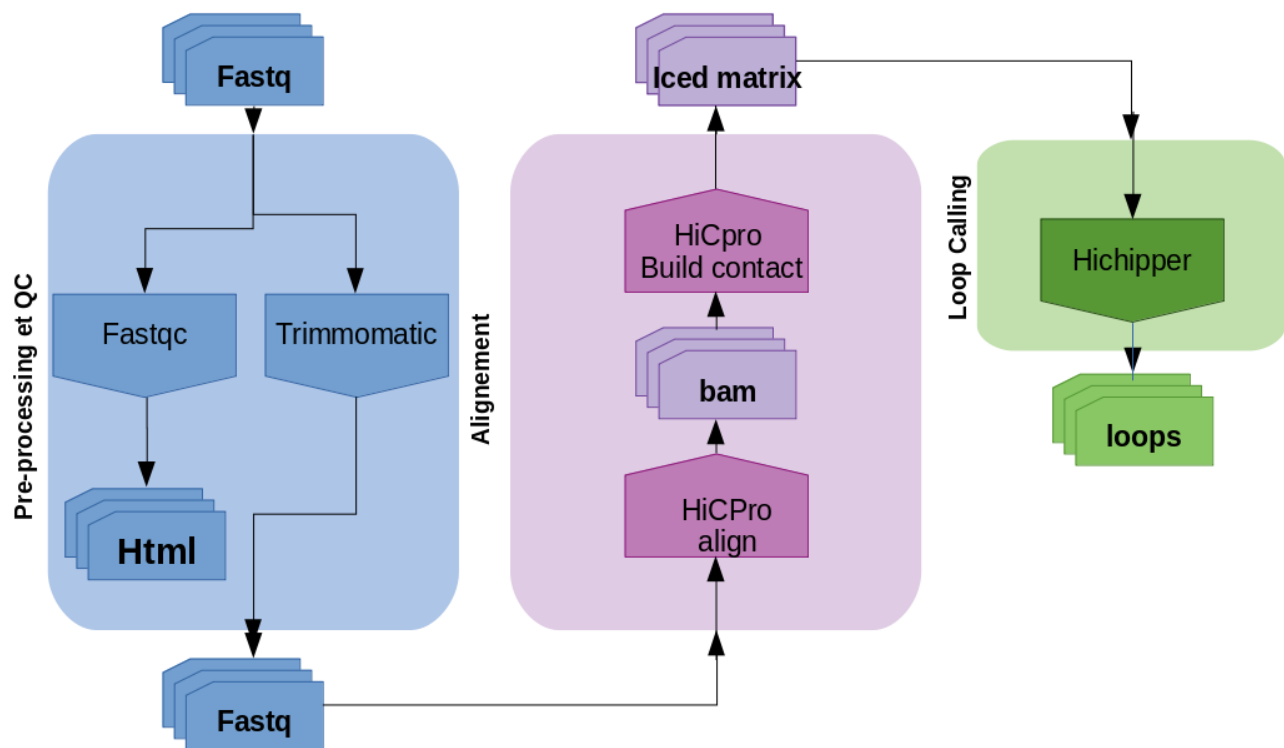


Figure 43: Pipeline d'analyse de données HiChIP mis en place.

7 Résultats

Les résultats obtenus pendant cette thèse peuvent être distingués en deux principales parties. Tout d'abord, nous avons effectué une analyse pan-génomique sur un bulk de cellules MM.1S pour déterminer la réponse à la Dex au sein d'une population cellulaire. Les résultats obtenus dans cette première partie ont été ensuite utilisés pour aller étudier les mécanismes de réponses à la Dex avec des approches de séquençage en cellule unique.

7.1 Le paysage chromatinien détermine la fixation du GR sur le génome

Cette première partie des résultats montre comment les différentes données pan-génomiques peuvent être intégrées afin de d'établir une carte de l'ensemble des sites de fixation du GR, de la présence de marques d'histones et de régions chromatinienes accessibles en combinaison avec des données transcriptomiques sur l'ensemble du génome (Figure 44). La combinaison de ces différentes données permet ainsi de mieux comprendre l'ensemble des mécanismes de régulation épigénétiques au sein d'une population cellulaire en réponse à la fixation du GR sur l'ADN.

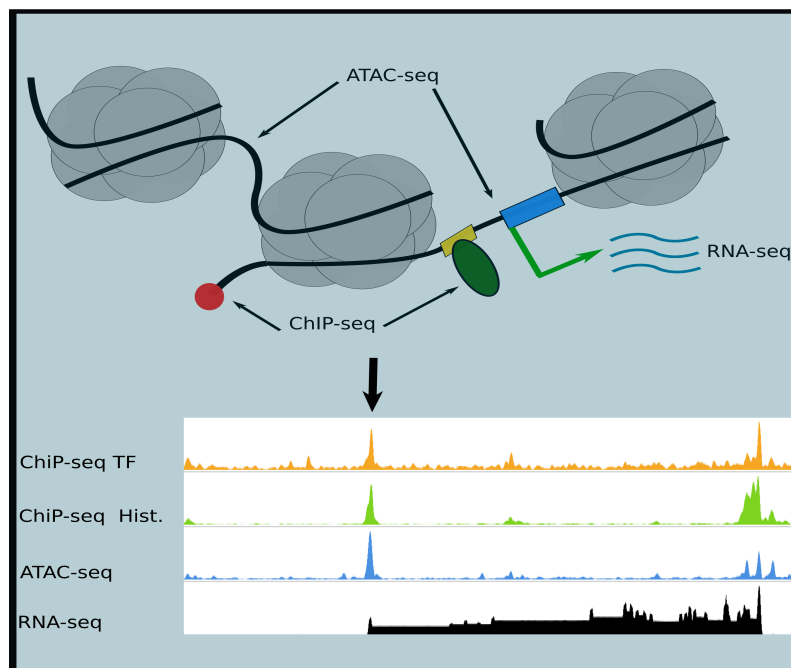


Figure 44: L'intégration des données de NGS ATAC-seq, ChIP-seq et RNA-seq déterminent le paysage chromatinien tout au long du génome.

7.1.1 Annotation fonctionnelle de la lignée cellulaire MM.1S

L'utilisation des données de ChIP-seq de marques d'histones permet de définir une carte fonctionnelle de l'ensemble du génome d'une lignée cellulaire. Pour cela, cinq marques d'histones ont été récupérées sur l'encyclopédie ENCODE (Zhang et al., 2020). Les données de ChIP-seq sont utilisées avec chromHMM (Ernst and Kellis, 2010) pour définir un modèle à 10 états chromatinien (Figure 45). Chaque marque définit ainsi une particularité génomique: H3K36me3 des régions transcriptionnellement actives, H3K27me3 des régions inactives ou réprimées, H3K4me1 des régions enhancers, H3K27ac des régions régulatrices actives et finalement H3K4me3 pour des régions promoteurs.

Les résultats obtenus sur des fenêtres génomiques de 200pb montrent un modèle à 10 états avec des probabilités de transitions plus élevées entre les différents états définissant des régions régulatrices 4 à 8 (Figure 45A).

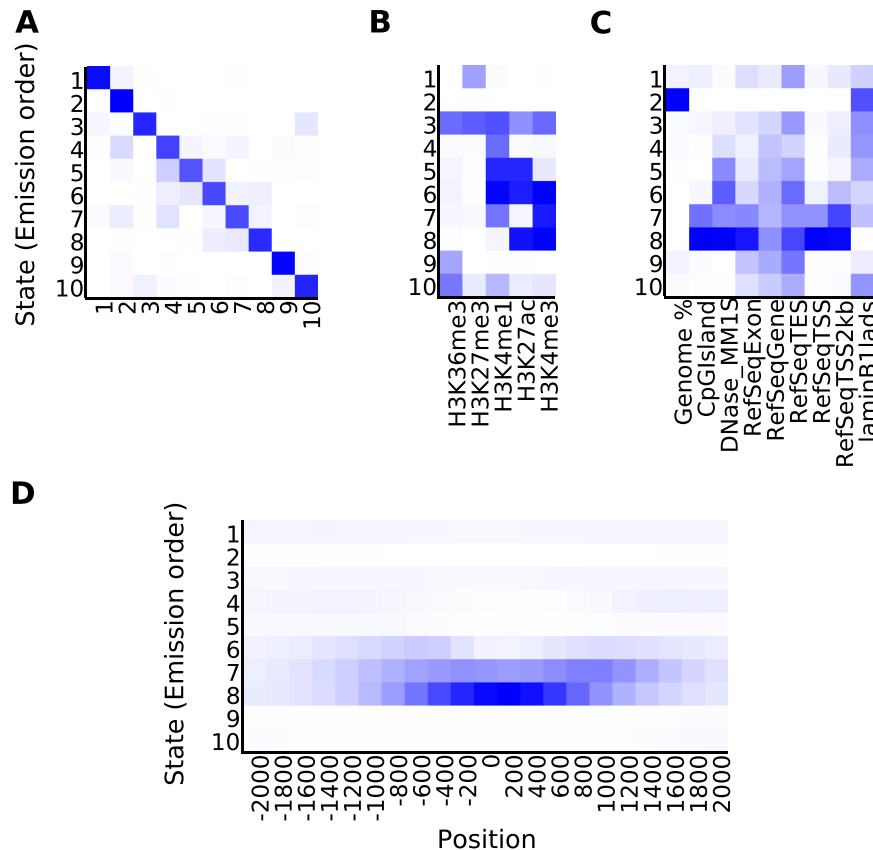


Figure 45: Utilisation des données de ChIP-seq de marques d'histones pour définir un modèle à 10 états chromatinien fonctionnels avec chromHMM (A) Heatmap de probabilité de transitions entre les états chromatinien. (B-D) Enrichissement des ChIP-seq de marques d'histones (B), de catégories génomiques (C) et position par rapport à la distance du TSS le plus proche (D) pour chacun des 10 états chromatinien du modèle.

Ces 10 états de chromatines montrent des enrichissements de marques chromatinienne spécifiques (Figure 45B). Les états chromatiniens peuvent être découpés en trois principaux groupes: des états de chromatine inactive identifiés par la présence de la marque H3K27me3 (état 1) de toutes ou d'aucune marque pour les états 3 et 2 respectivement, les régions transcriptionnellement actives marquées par la marque H3K36me3 (états 9 et 10) et finalement les régions régulatrices promoteurs et enhancers actifs ou non sont représentées par les états 4 à 8 avec un fort enrichissement des marques H3K4me1, H3K4me3 et H3K27ac.

Les différentes catégories d'annotations génomiques (Figure 45C) et la distance de ces états par rapport au TSS le plus proche (Figure 45D) sont deux données permettant de séparer de façon plus fine les trois grands groupes d'états identifiés. Le groupe correspondant à des états de chromatine inactive couvre une part importante du génome, particulièrement enrichis en Lamine B1, un marqueur d'hétérochromatine important (Camps et al., 2015) notamment pour l'état Hétérochromatine (état 2). Ces états sont généralement loin d'un site d'initiation de la transcription. De même, les états 9 et 10 associés à un état transcriptionnel montrent un enrichissement de catégories génomiques RefSeq Exon, Gene et TES soulignant le rôle de la marque H3K36me3 dans la régulation de la transcription (Kolasinska-Zwierz et al., 2009). Finalement on peut distinguer dans les états 4 à 8, plutôt associés à des régions régulatrices actives ou non, les promoteurs des enhancers grâce à ces différentes annotations. Les enhancers faibles (état 4) et forts (états 5 et 6) sont enrichis pour la technique d'accessibilité de la chromatine DNase-seq MM.1S et sont plutôt distants des TSS. Les promoteurs faibles et forts (respectivement états 7 et 8) sont enrichis en îlots CpG, des régions riches en bases C et G retrouvées généralement dans les promoteurs des gènes (Antequera, 2003), et proches des sites d'initiations de la transcription. L'ensemble de l'annotation de ces états et les marques d'histone principales correspondant à chaque état sont données dans la Table 3.

Etats	Marques	Annotation
E1	H3K27me3	Repressed Polycomb (Repr.)
E2	None	Heterochromatin (Heter.)
E3	All	Repeats/CNV (Rpts/CNV)
E4	H3K4me1	Weak Enhancers (WkEnh.)
E5	H3K4me1;H3K27ac	Strong Enhancers 1 (SgEnh1)
E6	H3K4me1;H3K4me3;H3K27ac	Strong Enhancers 2 (SgEnh2)
E7	H3K4me3;H3K4me1	Weak Promoters (WkProm.)
E8	H3K4me3;H3K27ac	Strong Promoters (SgProm.)
E9	H3K36me3	Weak Transcription (WkTx.)
E10	H3K36me3;H3K4me1	Transcription (Tx.)

Table 3: Annotation biologique et fonctionnelle du modèle à 10 états chromatiniens réalisée pour la lignée MM.1S.

Les 10 états chromatiniens ainsi obtenus ont été utilisés pour l'ensemble de l'étude afin d'annoter fonctionnellement les différents sites de fixation du GR sur le génome pour mieux caractériser son action.

7.1.2 La fixation du GR dépend du paysage chromatinien de la cellule

Les données de ChIP-seq réalisées en condition EtOH et Dex révèlent respectivement 2,878 et 20,587 sites de fixation du GR sur le génome (GRBS). Parmi l'ensemble de ces sites, j'ai conservé pour la suite de cette étude les sites GR uniquement spécifiques de la condition Dex ($n=18,844$ GRBS).

Tout d'abord, je me suis intéressé au paysage chromatinien sans Dex pour comprendre et définir les conditions dans lesquelles se fixe le GR. Le GR agissant comme un facteur de transcription, il se fixe préférentiellement dans des régions régulatrices du génome (Figure 46) (Vockley et al., 2016).

J'ai tout d'abord regardé à quelle distance du TSS le plus proche étaient localisés ces GRBS. En considérant qu'un pic localisé à moins de 5kb d'un TSS était considéré comme proximal et qu'un pic localisé à plus de 5kb d'un TSS était considéré comme distal, on remarque que 46% (8,476/18,444) des GRBS sont localisés à moins de 5kb du TSS le plus proche et que 54% (10,368/18,844) de ces GRBS sont identifiés dans des régions distales (Figure 46a gauche). En comparant ces GRBS avec l'annotation fonctionnelle chromHMM MM.1S (Figure 46a droite), on observe à nouveau la même répartition: la majorité des GRBS se fixant dans des régions annotées comme promoteurs ou enhancers (86.30%). On note cependant que pour l'ensemble de ces régions, celles-ci sont déjà particulièrement actives et que seule une faible proportion se fixe dans des régions inactives. Lorsque l'on s'intéresse à l'accessibilité de la chromatine pour les régions fixant le GR, on remarque que près des deux tiers (63.17%) de ces régions sont déjà accessibles dans la lignée MM.1S (Figure 46b). Ces résultats démontrent que l'état chromatinien de la cellule est capital pour la fixation du GR. En effet le GR se fixe dans des régions régulatrices déjà actives et accessibles pour effectuer son action.

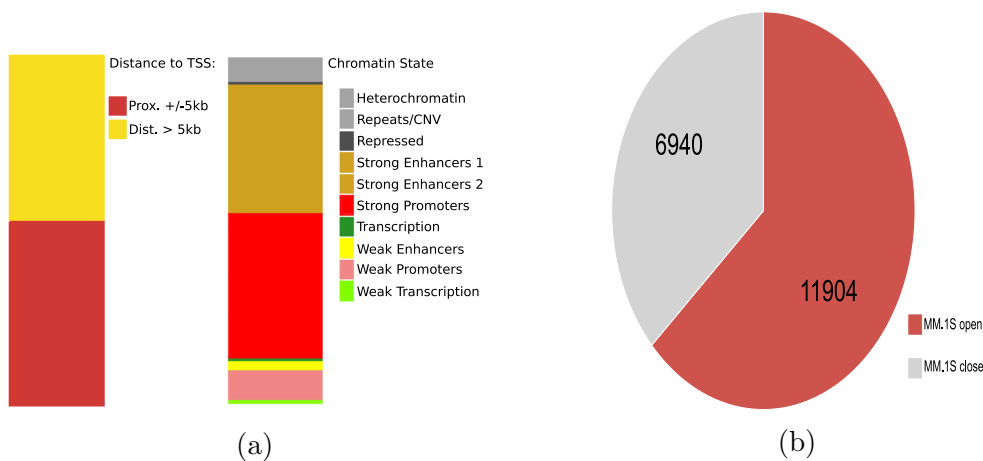


Figure 46: **Distribution Génomique des sites de fixation du GR dans la lignée MM.1S.** (a) Les histogrammes représentent la distribution des sites de fixation du GR en fonction de leur distance au TSS le plus proche (gauche; Prox. pour un site GR localisé à moins de 5kb d'un TSS et Dist. pour un site GR localisé à plus de 5kb d'un TSS) et en fonction de l'état chromatinien dans lequel le pic GR se trouve (droite). (b) Proportion de sites accessibles ATAC-seq fixant le GR dans la lignée MM.1S.

Parmi ces différentes régions régulatrices fixant le GR, celles-ci peuvent être divisées en deux groupes: les régions déjà actives (A) et les régions inactives (I) (figure 47). Pour les régions déjà actives, 40.03% (7,543) des GRBS sont localisés dans des promoteurs forts, un quart des GRBS (25.55%) sont localisés dans les catégories enhancers forts 1 et 2 avec respectivement 17.52% (3,302) et 18.03% (3,398) sites GR. Ces résultats démontrent une forte affinité du GR pour la marque d'histone H3K27ac soulignant un tropisme évident du GR pour les régions déjà actives. Il est également important de remarquer qu'une proportion plus faible est localisée au niveau de régions inactives (13.40% soit 2,525 sites). Parmi ces sites, 2.68% (505) sont retrouvés dans l'hétérochromatine, 2.69% (507) localisés au niveau d'enhancers faibles et enfin 8.03% (1,513) localisés au niveau de promoteurs faibles.

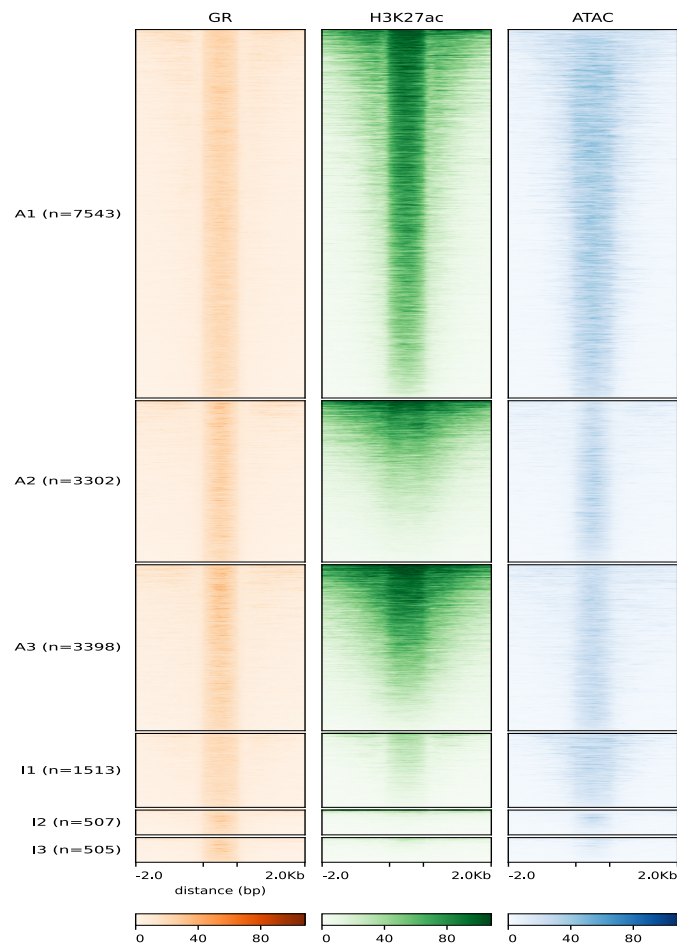


Figure 47: **La fixation de la Dex sur le génome est pré-déterminée par l'acétylation et l'accessibilité des régions**, Fixation de la Dex dans les régions actives définies par ChromHMM comme Strong Promoters (A1), Strong Enhancers 1 (A2), Strong Enhancers 2 (A3) et dans les régions inactives comme Weak Promoters (I1), Weak Enhancers (I2) ou Hétérochromatine (I3).

Ce paysage chromatinien peut être observé sur de nombreux gènes. Par exemple, le GR se fixe sur un enhancer et sur le promoteur de la forme courte du gène *TSC22D3* codant pour la protéine GILZ (Figure 48). Ce locus est identifié comme un gène de réponse aux glucocorticoïdes qui régule à la fois la réponse immunitaire innée et adaptative (Beaulieu and Morand, 2011) et qui pourrait inhiber la transcription de facteurs importants pour la voie NF- κ B et d'autres voies de signalisation. Le GR se fixe dans différentes régions autour de ce gène. Le GR se fixe sur deux enhancers, le premier en amont du gène et le deuxième à l'intérieur du premier exon du gène. Deux formes de GILZ ont déjà été décrites dans d'autres études (Bruscoli et al. (2010), Ayroldi et al. (2014)): la forme courte GILZ qui régule les effets anti-inflammatoires des GCs et son variant long-GILZ (L-GILZ) qui inhibe la croissance de la tumeur en activant la protéine p53 (Ayroldi et al., 2014). Dans notre système on remarque alors que seul le promoteur de la forme courte est déjà actif et fixe le GR.

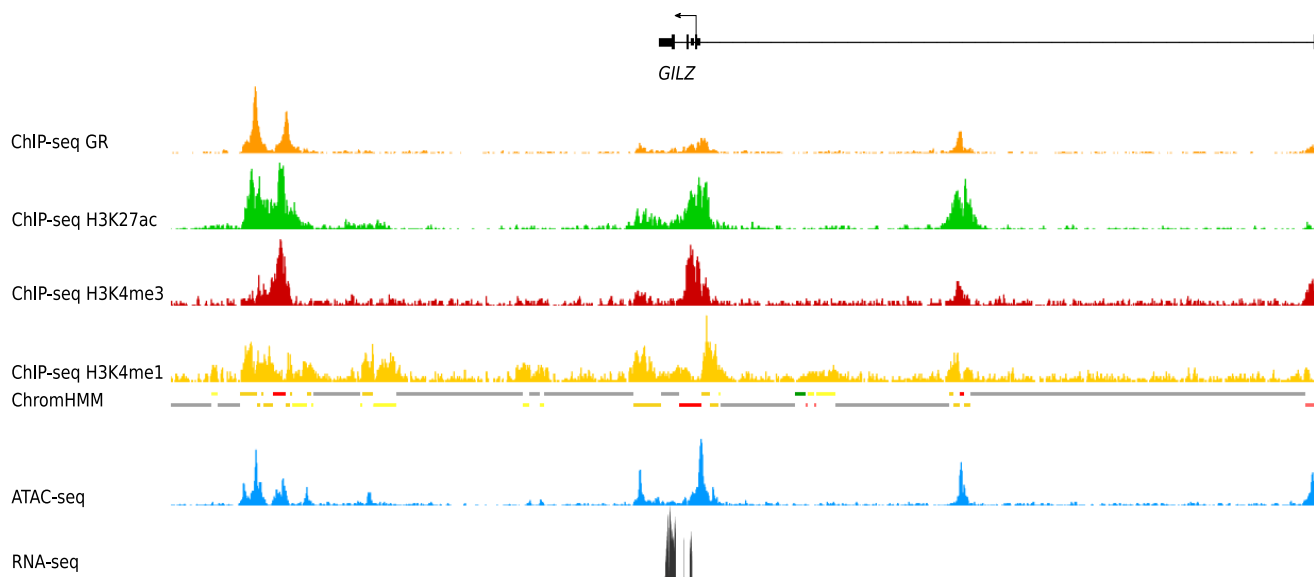


Figure 48: **Intégration des données pan-génomiques sur le locus du gène *TSC22D3* pour illustrer le rôle de la fixation du GR sur la régulation de la transcription.**

En parallèle, j'ai utilisé les données de ChIP-seq H3K27ac MM.1S pour déterminer les super-enhancers (SE) présents dans cette lignée. Les SE contrôlent la transcription de facteurs et de gènes clés pour l'identité de la cellule (Hnisz et al. (2013), Lovén et al. (2013)). Le ChIP-seq H3K27ac MM.1S a permis d'identifier 40,389 régions régulatrices actives dans notre modèle cellulaire. Ces régions sont classées selon l'intensité de leurs signaux et leurs distances au TSS ce qui nous permet d'identifier 815 SE (Figure 49). Le GR se fixe dans 99% (808/815) de ces SE. Parmi ces différents SE, différents gènes et facteurs de transcription ont pu être identifiés comme

étant sous le contrôle d'un de ces SE. Le premier SE régule la transcription du gène *IGLL5*, codant pour un polypeptide d'immunoglobuline lambda-like, un gène qui a récemment été identifié comme potentiellement transloqué avec le locus des chaînes lourdes des Igs (IGH) chez des patients atteints de myélome (White et al., 2018). D'autres facteurs de transcriptions tels que IRF4, XBP1 ou encore le gène *CCND2* sont également sous le contrôle d'un SE, ces différents facteurs ont déjà été décrits dans une autre étude (Lovén et al., 2013) comme étant importants pour la cellule de myélome. De plus, les facteurs de transcription importants du MM comme IKZF1, IKZF3, PRDM1, RUNX1 et MEF2C sont également sous contrôle d'un SE. Ces différents facteurs de transcription sont considérés comme de potentiels co-interacteur du GR.

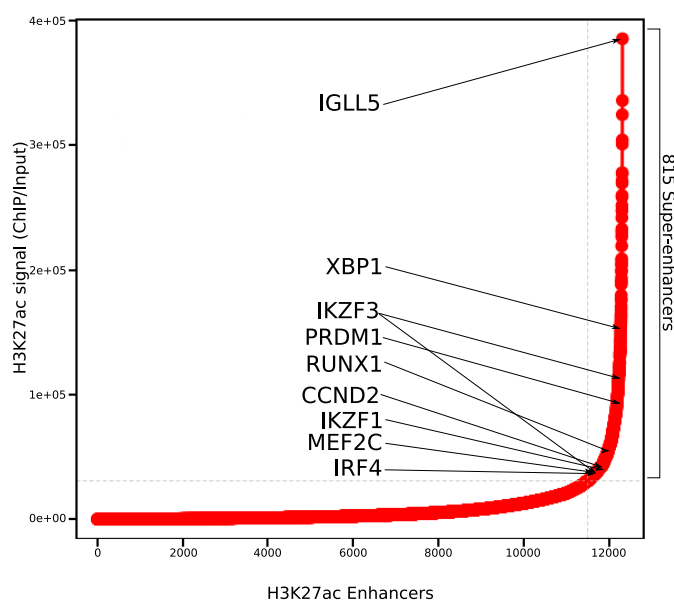


Figure 49: **Détermination des super-enhancers et facteurs de transcription cœurs de la lignée MM.1S.**

Afin de mieux caractériser ces sites de fixations, j'ai réalisé une découverte de motif *de novo* avec MEME (Bailey et al., 2015) pour les GRBS localisés dans les différentes catégories chromHMM (Figure 50). Sans surprise, un enrichissement important du motif de fixation du GR (GRE) est retrouvé dans les catégories Strong Enhancers 1 et 2, Weak Enhancers mais aussi dans la catégorie hétérochromatine. Ce résultat met en évidence une fixation directe à l'ADN du GR sur son motif dans ces régions et donc probablement une activation des différents gènes régulés par ces régions. Cependant, parmi les milliers de sites de fixation du GR sur le génome, la fixation sur un motif GRE ne représente qu'une faible proportion de l'ensemble des GRBS et ne suffit donc pas à expliquer l'ensemble des mécanismes moléculaires mis en place par le GR pour réguler la transcription des gènes cibles.

D'autres motifs importants sont également retrouvés, notamment le motif de type E-box, enrichi dans les strong enhancers 1, fixant de nombreux facteurs de transcription. Ce motif permet la

fixation de facteurs de transcription de type bHLH (beta Helix-Loop-Helix) certains facteurs de transcription de ce type tel HES1 ont déjà été décrits comme régulant la prolifération cellulaire dans les cellules de MM (Sachithanandan et al., 2017).

Finalement, j'ai pu identifier au niveau des promoteurs forts et des strong enhancers 2, un enrichissement du motif homo-dimère ISRE fixant les facteurs de transcription type IRF. Ce résultat suggère un rôle important du facteur de transcription spécifique du lignage B et fortement exprimé dans la lignée MM.1S : IRF4. IRF4 est un facteur essentiel lors de la différenciation du lignage B et est un facteur pionnier dans les cellules de MM (Shaffer et al., 2009). Ce résultat montre donc une possible fixation indirecte du GR via l'interaction avec d'autres facteurs de transcription importants pour la cellule.

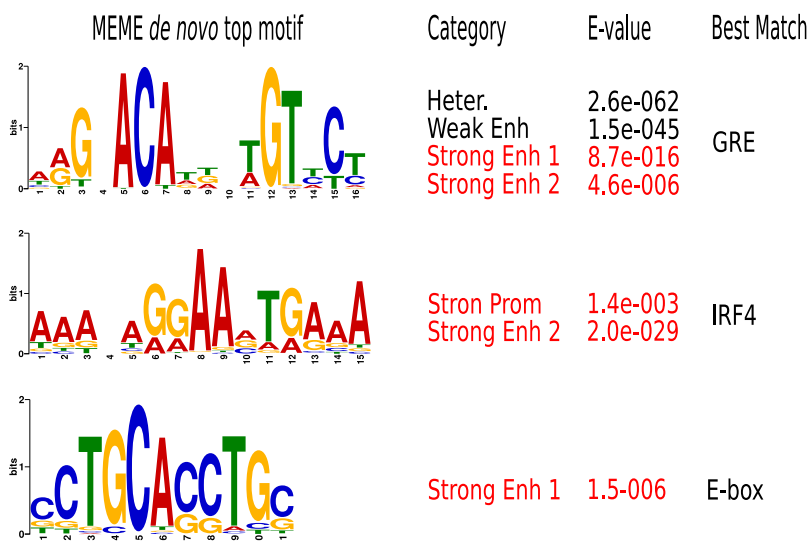


Figure 50: Recherche de motifs *de novo* pour les sites de fixation du GR selon les différents états chromatinien, mise en évidence de trois motifs principaux obtenus par MEME.

7.1.3 Le GR interagit avec un facteur pionnier du lignage B: IRF4

Ces premiers résultats suggèrent que l'état chromatinien de la cellule pourrait être maintenu via la présence de facteurs importants pour la cellule et notamment le facteur de transcription IRF4. Afin de vérifier cette hypothèse, nous avons récupéré les données de séquençage ChIP-seq IRF4 MM.1S (Lovén et al., 2013). Parmi l'ensemble des GRBS, une grande proportion (73.16% soit 13,786 sites) sont co-localisés avec un pic IRF4 (Figure 51a). Ce résultat démontre que le GR se fixe dans les régions où IRF4 est déjà présent soit pour agir en tant que co-interacteur soit pour agir en tant qu'antagoniste sur ces sites.

L'analyse ontologique par enrichissement de régions génomiques avec GREAT (McLean et al., 2010) des sites communs au GR et à IRF4 révèle alors un enrichissement de voies de signalisation importantes pour le lignage B (Shaffer et al., 2008) comme la réponse aux stimulations par les cytokines, l'activation des leucocytes ou encore la régulation positive de la prolifération des cellules B (Figure 52). Le GR interagit donc avec le facteur IRF4 sur des régions importantes pour le

lignage B probablement en agissant comme un antagoniste sur ces différents sites.

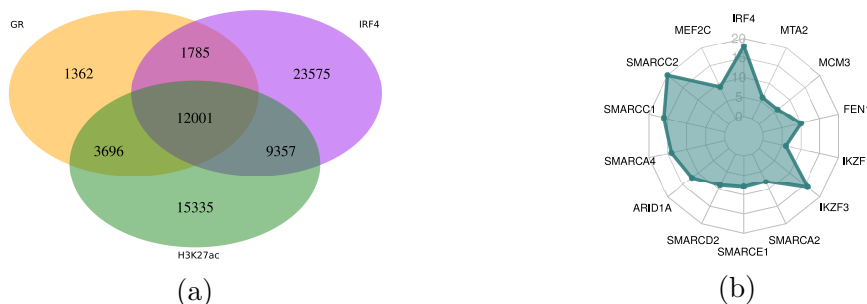


Figure 51: **Interaction du GR avec le facteur pionnier du lignage B IRF4.** (a) Diagramme de Venn montrant la col-localisation des sites de fixation du GR avec les sites de fixation IRF4 et les sites de localisation H3K27ac dans MM.1S. (b) Diagramme illustrant les principales protéines identifiées par la technique RIME et interagissant physiquement avec le GR.

Pour aller plus loin, j'ai séparé en trois catégories les sites fixant le GR et IRF4 selon la présence de trois motifs de fixation à l'ADN: le motif complet GRE, le demi-motif GRE fixant un monomère GR et enfin le motif de reconnaissance de l'interféron ISRE qui fixe le facteur IRF4 (Figure 53). Cette catégorisation repose sur une hypothèse de mécanismes moléculaires distincts: lorsque le GR se fixe sur son GRE via une liaison directe à l'ADN à côté d'un site IRF4, cette fixation entraîne une transcription des gènes de réponse à la Dex; de même lorsqu'il se fixe sur un demi-motif où le facteur IRF4 aurait un rôle de stabilisation de la fixation du GR dans ce cas là. Enfin la fixation du GR au niveau des sites ISRE se ferait de façon indirecte et le GR aurait alors un rôle antagoniste à IRF4 sur la transcription des gènes cibles.

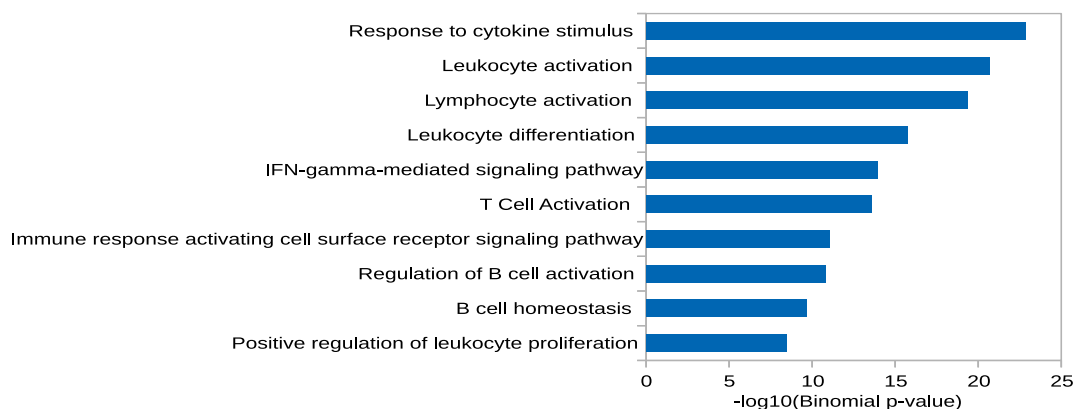


Figure 52: **GO biological process enrichis pour les sites co-localisant la fixation du GR et du facteur pionnier IRF4.**

L'analyse d'enrichissement de régions génomiques révèle différents GO biological process enrichis pour ces trois catégories. On observe que pour les régions contenant un motif GRE ou un demi-motif, le GR va intervenir dans des voies de signalisation déjà répertoriées en réponse à la Dex telles que la régulation des voies de signalisations apoptotiques intrinsèques et extrinsèques, l'inactivation

de l'activité des MAP kinases et la régulation de processus métaboliques (GRE) ou encore la voie de signalisation intrinsèque de l'apoptose et la régulation de la prolifération des cellules B (demi-motif GRE). En revanche, les voies de signalisations régulées par les régions avec un motif ISRE semblent être plus lignage-spécifiques avec des voies telles que la régulation de la production et de la réponse à l'interféron gamma ou encore la différenciation des cellules B matures. Ces résultats démontrent que la présence du motif de fixation à l'ADN influe pour déterminer le type de réponse engendrée par la Dex et le type de mécanisme mis en place par le GR pour moduler cette réponse.

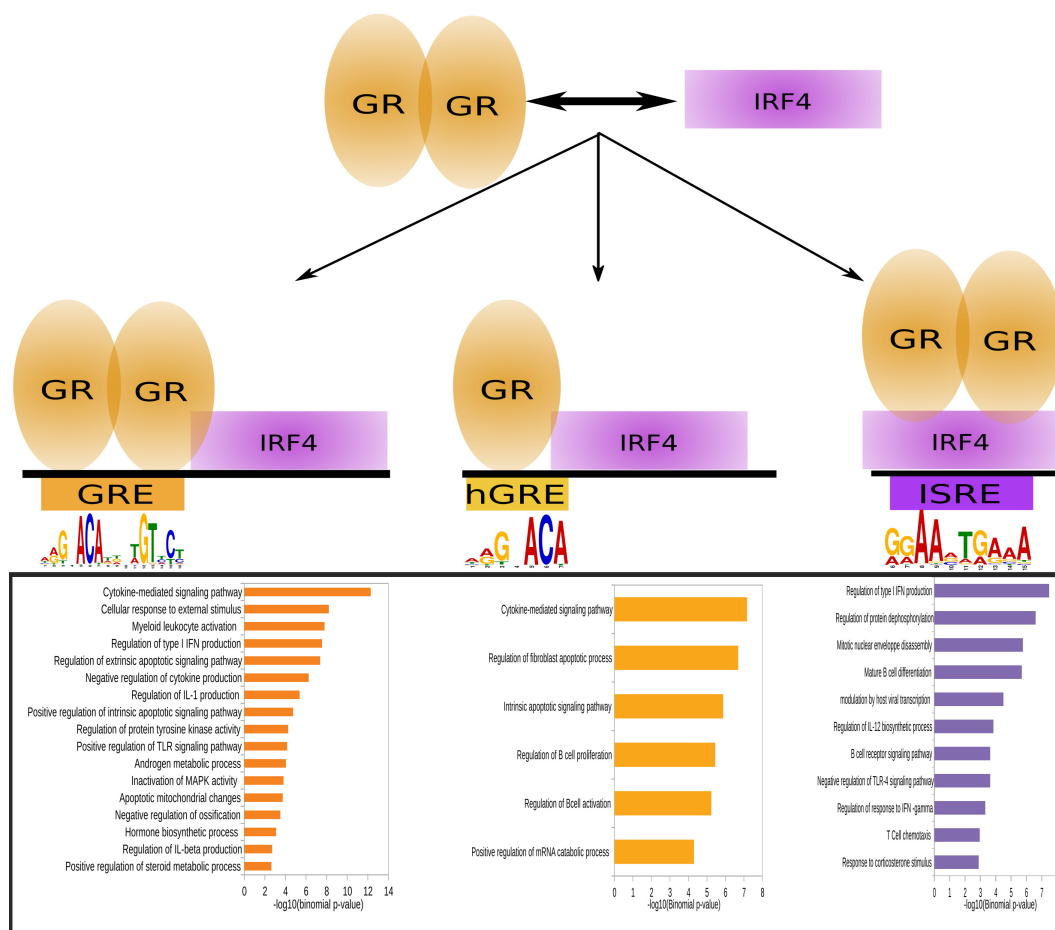


Figure 53: Schéma illustrant les différents termes de l'ontologie GO biological process enrichis pour l'association GR:IRF4 en fonction de la présence d'un motif de fixation sur l'ADN.

Finalement, les données de ChIP-seq IRF4 étudiées ayant été produites dans la lignée MM.1S sans traitement par la Dex, nous avons donc voulu vérifier si il existait bien une interaction physique

entre ces deux protéines. Nous avons donc réalisé une étude d'immuno-précipitation de protéines couplée à des techniques de spectrométrie de masse (RIME) afin de déterminer toutes les protéines interagissant physiquement avec le GR dans la lignée MM.1S avec Dex (Figure 51b). Les résultats de ces analyses confirment qu'il existe bel et bien une interaction physique entre le GR et le facteur de transcription IRF4. D'autres facteurs interagissant physiquement avec le GR sont également mis en évidences notamment IKZF1, IKZF3. Ces deux facteurs sont essentiels pour le MM (Krönke et al., 2014). Le GR interagit également avec MEF2C, un facteur important pour l'activité SE et l'expression de MYC dans les lymphomes primaires (Wang et al., 2020). Ces différents facteurs ont de plus été décrits précédemment, comme IRF4, sous le contrôle d'un super-enhancer dans la lignée MM.1S (Figure 49).

Hormis ces facteurs de transcription clés pour la cellule, le GR interagit avec de nombreuses protéines SMARCC2, SMARCC1, SMARCA4, SMARCD2, SMARCE1, SMARCA2 et ARID1A qui sont toutes des sous-unités du complexe protéique SWI/SNF. La mise en évidence de l'interaction du GR avec ce complexe protéique à déjà été montrée dans d'autres études (John et al. (2008); Johnson et al. (2018); Hoffman et al. (2018)) et suggère un autre mécanisme moléculaire d'action important du GR à travers l'ouverture et l'activation de nouvelles régions régulatrices.

Ces résultats confirment l'interaction du GR avec le facteur pionnier IRF4 pour réguler la transcription de gènes spécifiques du lignage B. En plus de cette interaction, les résultats obtenus avec le RIME suggèrent que le GR a également un effet sur l'ouverture et donc l'activation de régions régulatrices qui seront spécifiques de la réponse à la Dex.

7.1.4 La fixation du GR ouvre des nouveaux sites de fixation sur le génome

Nous avons donc identifié que le GR se liait à des protéines faisant partie du complexe de remodelage de la chromatine ce qui pouvait donc conduire à l'ouverture de nouveaux sites de fixation du GR. Nous avons donc voulu caractériser ces nouveaux sites.

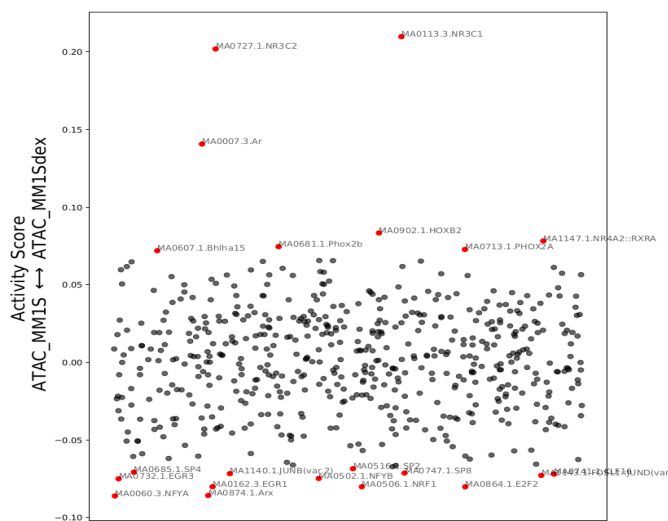


Figure 54: **Footprinting différentiel pour les données ATAC-seq en condition Dex versus EtOH.**

Dans le but d'identifier des motifs de fixation spécifiques à l'ouverture de la chromatine, des analyses différentielles de foot-printing sur les données ATAC-seq (Li et al., 2019b) en condition Dex versus EtOH ont été réalisées (Figure 54). Ces résultats montrent une ouverture spécifique au niveau de motifs de fixation pour la MM.1S en condition EtOH tels que les sites de fixations AP-1 (JUNB, JUND) ou encore le site de fixation du facteur de machinerie transcriptionnelle E2F2. En ce qui concerne la condition Dex, on observe une ouverture essentiellement pour le motif de reconnaissance du GR (NR3C1, NR3C2). En effet, lorsque l'on regarde les profils ATAC-seq des nouveaux sites ouverts ($n=334$), on remarque alors une empreinte plus marquée pour la condition Dex comparée à la condition EtOH (Figure 55).

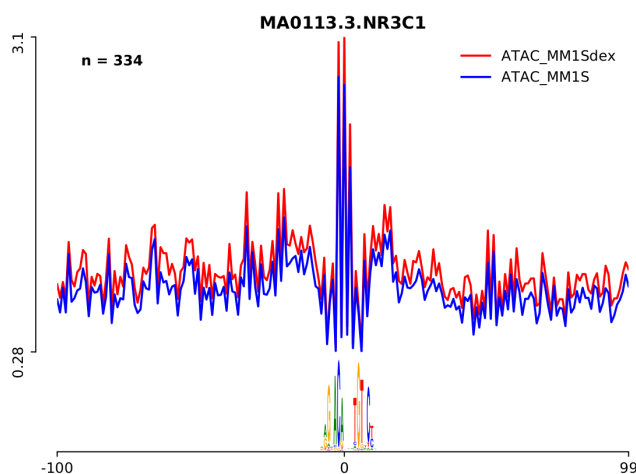


Figure 55: Le footprinting révèle l'ouverture de nouveaux sites GRE.

7.2 Le GR active des régions régulatrices Dex-spécifiques

Les résultats obtenus lors de cette première partie ont donc révélé que pour la majeure partie des sites de fixation du GR sur le génome, le paysage chromatinien déjà existant est prépondérant. En effet le GR possède un fort tropisme pour la marque H3K27ac et interagit avec de nombreux facteurs de transcription essentiels pour la cellule. Cependant, nous avons également montré que le GR se fixait sur une minorité de sites pour lesquels l'accessibilité de la chromatine augmentait. Nous avons donc émis l'hypothèse de l'activation d'une faible part de ces sites qui pourraient être signature de la réponse à la Dex.

Pour vérifier cette hypothèse, je me suis intéressé dans cette partie aux régions régulatrices qui s'activent en réponse à la Dex. Pour ce faire, j'ai effectué une analyse différentielle des 38,662 sites H3K27ac MM.1S Dex. Sur l'ensemble de ces sites, je n'ai conservé que les pics H3K27ac fixant le GR pour observer l'effet de la fixation de celui-ci sur l'acétylation des régions. Parmi les pics restants ($n=16,228$), j'ai estimé le Log FoldChange (LogFC) de chacun de ces pics pour distinguer ceux qui augmentent en réponse à la Dex. J'ai choisi de prendre un LogFC bas ($\text{LogFC} > 0.1$)

pour capturer 864 pics H3K27ac Dex-induits et 15,364 pics Dex-stables. Sur les pics H3K27ac Dex-induits, on observe une franche augmentation du signal H3K27ac avec Dex comparé au signal sans Dex alors qu'aucune différence n'est observée pour les pics H3K27ac Dex-stables (Figure 56). On remarque également que le signal ATAC-seq avec Dex est également plus important dans les régions H3K27ac Dex-induites comparé au signal ATAC-seq sans Dex. Ce résultat montre qu'à travers l'ensemble des sites GR sur le génome une faible proportion montre un changement dans l'acétylation et l'accessibilité ce qui indique que ces 864 régions H3K27ac Dex-induites dirigent probablement la réponse spécifique de la Dex.

J'ai ensuite effectué une annotation génomique de ces régions. J'ai donc regardé la position du site d'initiation de la transcription le plus proche de chaque région H3K27ac Dex-induite en considérant comme proximale une région H3K27ac Dex-induite localisée à moins de 5kb d'un TSS (Figure 57). On constate alors qu'une faible proportion de ces régions (132 régions, 15.28%) est localisée proche d'un TSS alors que la majeure partie de ces sites (732 régions, 84.72%) est retrouvée dans des régions distales que nous définissons comme enhancers. La distance au TSS le plus proche varie grandement pouvant aller jusqu'à plus de 5 mégabases. On remarque qu'un tiers de ces enhancers sont localisés entre 10 et 50kb d'un TSS (307 régions, 35.53%).

Ces régions se distribuent donc plutôt dans des régions distales du génome ce qui illustre que l'action principale du GR se passe dans les enhancers.

L'analyse différentielle des données RNA-seq en bulk produites pour la MM.1S 4h après traitement EtOH ou Dex montre que malgré les milliers de sites de fixation du GR sur le génome, seul un faible nombre de gènes est régulé par le GR. En effet, dans cette analyse différentielle, 171 gènes ont été identifiés comme étant inhibés ($\log FC < -1$; $FDR < 0.05$) et 432 gènes sont induits ($\log FC > 1$; $FDR < 0.05$). Ce résultat montre que le GR permet l'activation de la transcription plus que la répression de la transcription. Il est cependant important de noter que des gènes importants du myélome sont down régulés par la Dex comme *IRF4* ($\log FC = -1.30814$; $FDR < 0.0001$) ou bien encore *CXCL10* ($\log FC = -4.31126$; $FDR < 0.0001$), un gène exprimé normalement lors de la réponse à l'interféron gamma. Le rôle activateur de la Dex sur la transcription des gènes est donc important. Parmi les gènes induits ($FDR < 0.0001$) en réponse à la Dex, on retrouve des gènes clés tels que *BCL2L11* ($\log FC = 1.07856$), *TSC22D3* ($\log FC = 4.29259$), *CXCR4* ($\log FC = 2.9497$; FDR), *FKBP5* ($\log FC = 3.95052$), *CFLAR* ($\log FC = 1.38622$), *DDIT4* ($\log FC = 6.62855$), *IL6ST* ($\log FC = 1.31502$), et *CREB3L2* ($\log FC = 2.04835$; $FDR < 0.0001$). Beaucoup de ces gènes ont déjà été décrits comme étant régulés par la Dex dans d'autres études.

Lorsque l'on s'intéresse de nouveau aux régions H3K27ac Dex-induites et plus particulièrement à la distance par rapport à des éléments génomiques d'intérêt, on remarque que ces régions jouent un rôle important dans la régulation de la transcription (Figure 58). En effet, nous avons tout d'abord observé que ces régions H3K27ac Dex-induites étaient plus facilement retrouvées proche des gènes up-régulés en condition Dex par rapport aux gènes dont l'expression ne varie pas (Figure 58a). Ce résultat suggère que ces enhancers Dex-induits contrôlent directement la transcription des gènes cibles et que la fixation du GR dans ces enhancers conduit à une activation de la transcription.

De plus, nous nous sommes intéressés à la distance mutuelle de ces régions spécifiques et à la distance par rapport aux enhancers H3K27ac Dex-stables (Figure 58b). Il est alors intéressant de noter que les enhancers H3K27ac Dex-induits sont plus proches les uns des autres que des enhancers H3K27ac Dex-stables. Ce résultat nous conduit à penser que ces enhancers pourraient se

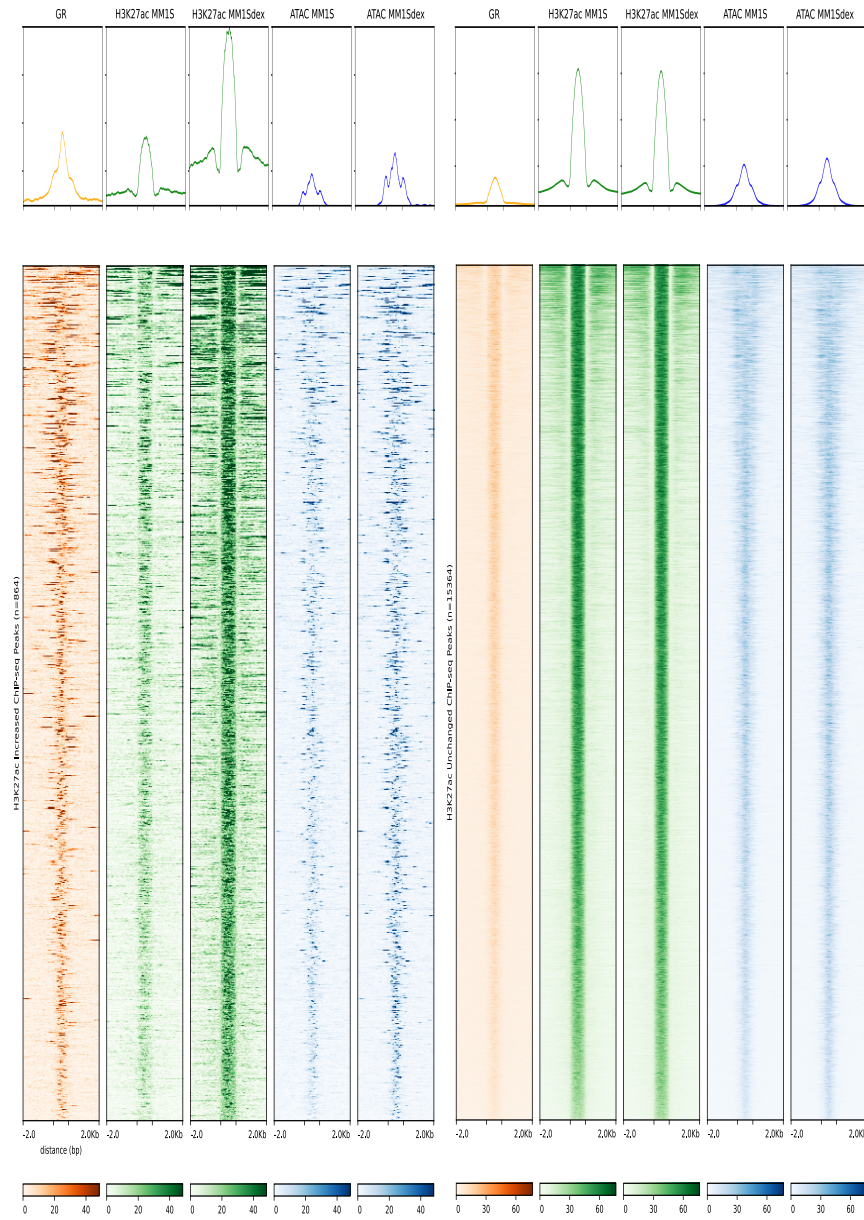


Figure 56: **Heatmap illustrant l'activation de régions H3K27ac liées à la fixation du GR**, les régions H3K27ac qui augmentent en réponse à la fixation de la Dex (gauche) ou qui ne bougent pas (droite) sont représentées.

regrouper entre eux pour agir en synergie sur la transcription des gènes cibles. Cette proximité entre enhancers H3K27ac Dex-induits permet également de penser que l'augmentation d'acétylation de ces régions peut amener à la formation de nouveaux super-enhancers contrôlant la transcription de gènes clés en réponse à la Dex. Une nouvelle analyse de détection de super-enhancers a donc été réalisée sur les pics H3K27ac MM.1S Dex (n=38,662). Les pics ont donc été classés selon leurs signal ChIP-seq ce qui a permis l'identification de 845 SE (Figure 59). Parmi ces différents

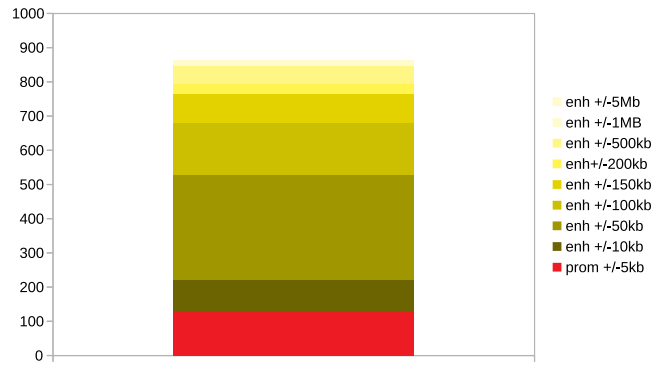


Figure 57: **Distribution génomique des régions H3K27ac Dex-induites**, annotation des régions par rapport à la distance au TSS le plus proche.

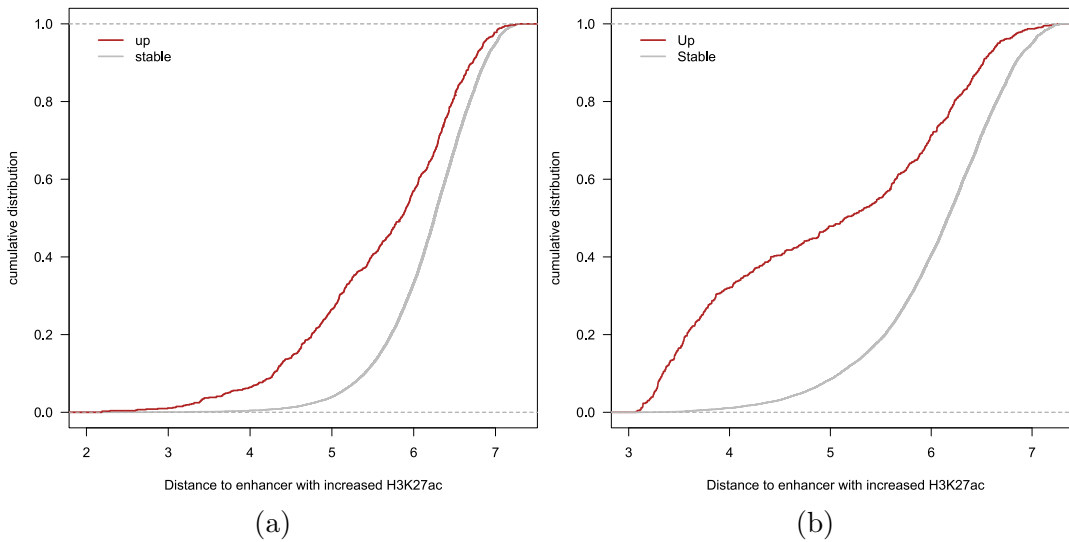


Figure 58: **Rôle des régions H3K27ac Dex-induites dans la régulation des gènes cibles.** (a) Distance des gènes up-régulés (rouge) ou stables (gris) par rapport aux régions H3K27ac Dex induites. (b) Distance des régions Dex-induites (rouge) ou stables (gris) par rapport aux régions H3K27ac Dex induites.

SE, 121 nouveaux SE spécifiques de la Dex ont été identifiés. De plus, du fait de l'augmentation d'acétylation dans ces régions, les SE présents avant Dex peuvent devenir plus grand ou même deux SE peuvent se regrouper en un seul plus grands. Ce résultat démontre que la fixation du GR sur un site particulier entraîne un rassemblement de différents enhancers pour activer la transcription de gènes cibles.

Parmi les cibles de ces SE, on remarque alors l'apparition de SE contrôlant directement l'expression de gènes importants dans la réponse à la Dex et déjà observés lors de l'analyse différentielle RNA-seq. C'est le cas notamment de 4 gènes particuliers: *DDIT4*, le facteur de transcription *CREB3L2*, *FKBP5* et enfin *TSC22D3*.

Les circuits de régulation pour les facteurs de transcription obtenus à partir des SE MM.1S avec

et sans Dex définissent les facteurs de transcription clés de la cellule. Parmi ces différents circuits, un grand nombre est retrouvé dans les deux conditions (IKZF1, XBP1, IRF4, HES1, PRDM1, KLF3, MEF2C, RUNX1, RUNX2, FOXO3...) alors que certains sont spécifiques d'une des deux conditions. Ainsi, on retrouve des circuits de régulations spécifiques de MM.1S sans Dex pour les facteurs JUNB et POU3F1. A l'inverse, les facteurs de transcriptions VDR, GLIS1, SOX2, TFAP2C, NR3C1 (GR), ATF4, ZBTB16, ETV6 et SREBF1 ont des circuits de régulation qui apparaissent en lien avec de nouveaux SE.

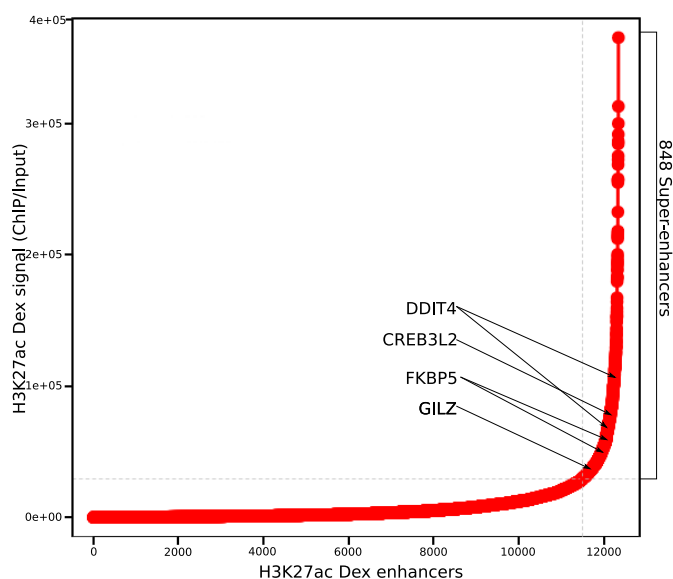


Figure 59: **Détermination des super-enhancers liés à la Dex et gène clés de réponse à la Dex.**

L'ensemble de ces résultats permet de mieux comprendre comment le GR va se fixer sur un site majeur et moduler tout l'état transcriptionnel d'une région. Un exemple illustrant parfaitement ce mécanisme est celui du locus du gène codant pour la protéine GILZ: *TSC22D3* (Figure 60). Cet exemple montre la fixation du GR dans des régions contenant un motif GRE à différents endroits (promoteurs et enhancers) autour du gène. Bien que ces régions soient déjà acétylées avant Dex, la fixation du GR conduit à l'augmentation de l'acétylation de ces régions ce qui amène également à l'apparition d'un super-enhancer contrôlant la transcription du gène.

L'augmentation d'acétylation sur les régions fixant le GR combinée à l'apparition du SE suggèrent une action synergique de ces régions conduisant à une augmentation de la transcription de ce gène. On note également pour certaines de ces régions, une augmentation en terme d'accessibilité de la chromatine.

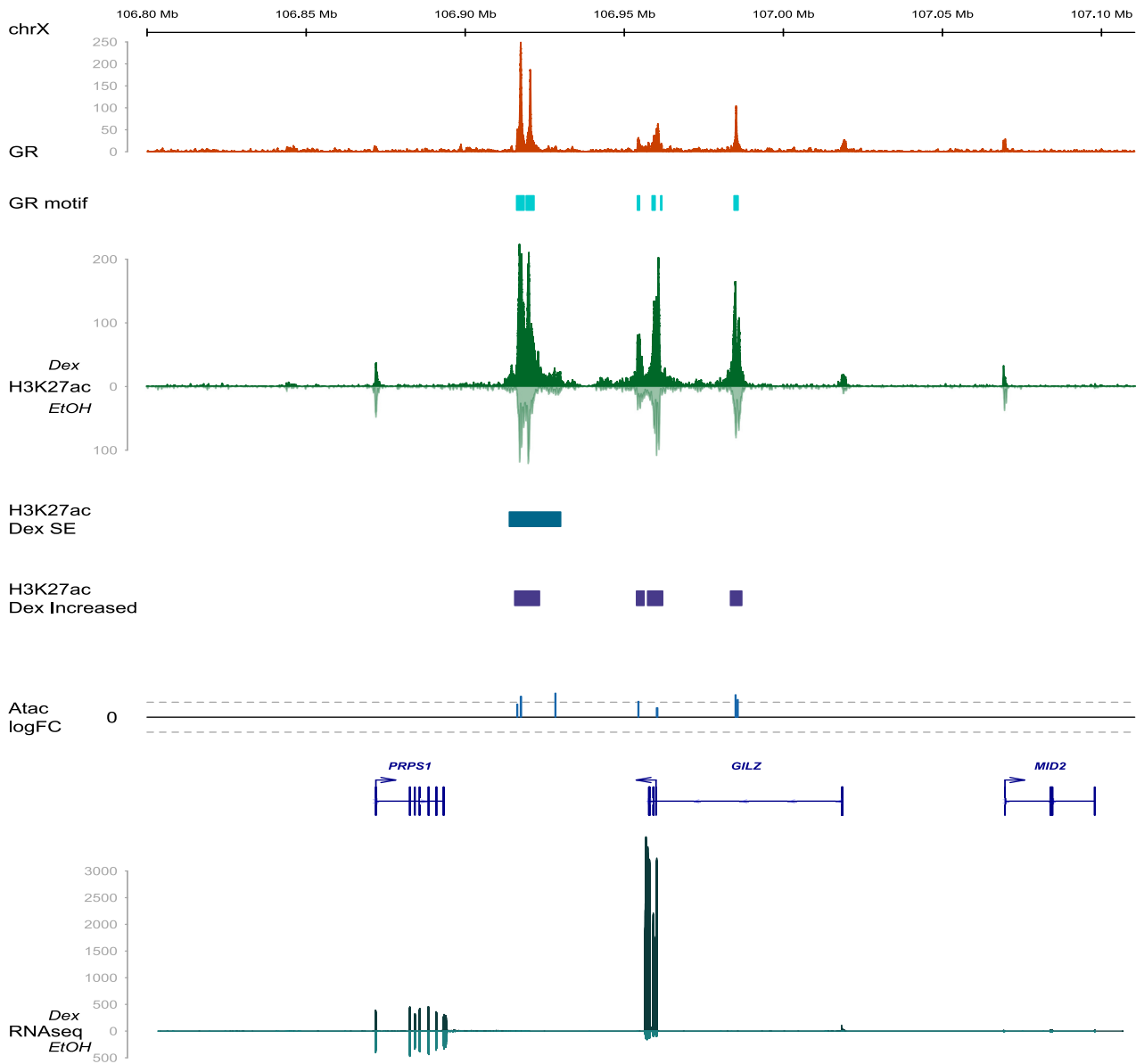


Figure 60: Exemple de fixation du GR sur le locus *TSC22D3* qui illustre le mécanisme de réponse à la Dex.

Les résultats démontrent que le GR joue un rôle non seulement dans l'activation et le regroupement d'enhancers (Petrovic et al. (2019), Mumbach et al. (2017)) mais également dans l'ouverture de régions. Malheureusement, ces analyses ne permettent pas de savoir avec certitude comment les enhancers interagissent entre eux ni de connaître avec précision le gène cible régulé par ces enhancers. Pour répondre à ces questions, nous avons effectué des analyses de conformation de la chromatine liée à la marque H3K27ac afin de comprendre le mécanisme du GR dans la régulation de ces interactions.

7.3 Action génomique du GR dans la régulation des boucles chromatinienne

Dans cette nouvelle partie, je vais présenter les résultats obtenus via l'analyse des interactions chromatinienne liées à la présence de la marque d'histone H3K27ac. Ces résultats ont été obtenus par technique du HiChIP, une méthode couplant immuno-précipitation de protéines et capture de conformation de la chromatine. Cette technique permet donc de détecter les régions sur le génome fixant la marque d'histone H3K27ac et permet d'étudier les interactions de type promoteurs/promoteurs, enhancers/enhancers et promoteurs/enhancers et ainsi de mieux comprendre les mécanismes moléculaires de régulation de la transcription des gènes régulés par le GR (Figure 61).

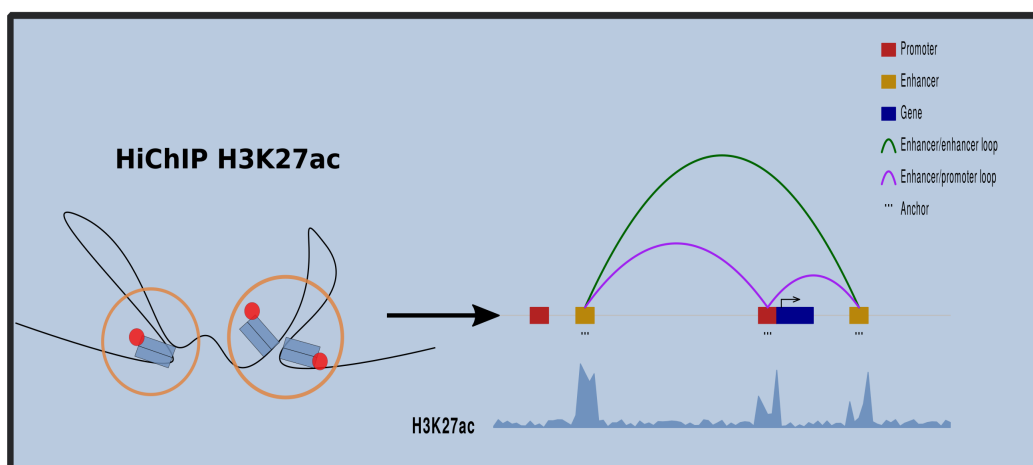


Figure 61: Schéma illustrant comment les boucles chromatinienne formées par la marque d'histone H3K27ac peuvent être étudiées par la technique du HiChIP.

7.3.1 Caractérisation des interactions chromatinienne

L'analyse des données HiChIP-seq pour la marque d'histone H3K27ac après addition d'EtOH ou de Dex et séquencés après une heure à permis d'identifier respectivement 21,248 (EtOH) et 23,277 (Dex) boucles d'interactions chromatinienne.

Afin de caractériser ces interactions d'un point de vue global, j'ai regardé tout d'abord les valeurs de signal ChIP-seq H3K27ac à l'intérieur ou hors des ancres dans les deux conditions (Figure 62). Les résultats montrent que peu importe la condition (EtOH ou Dex), on observe un signal ChIP-seq H3K27ac beaucoup plus intense ($p\text{-value} < 0.0001$ dans les deux conditions) à l'intérieur des ancres comparé au signal des pics H3K27ac localisés hors des ancres. On dénombre respectivement 50% et 55% des pics ChIP-seq H3K27ac localisés dans les ancres pour les interactions

HiChIP en EtOH et en Dex. Ceci souligne une corrélation forte entre les données de ChIP-seq et les données HiChIP ce qui vient valider la qualité de ces résultats. Les pics ChIP-seq localisés à l'extérieur d'anchors sont donc des pics de faible intensité qui peuvent être considérés comme des faux positifs ou secondaires.

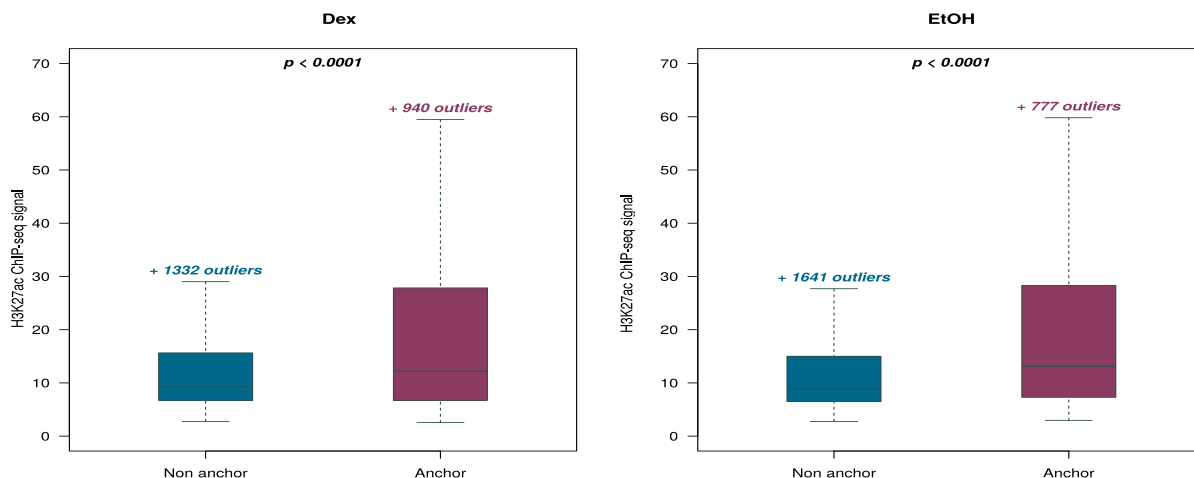


Figure 62: Signal ChIP-seq H3K27ac des pics localisés hors et à l'intérieur des *anchors* des interactions HiChIP H3K27ac.

Lorsque l'on compare les deux conditions EtOH et Dex, aucune différence notable n'est observée en terme de distribution des distances entre les deux anchors d'une même boucle (Figure 63 gauche), avec une distance médiane respective de 127kb et 125kb en EtOH et en Dex. De même, l'annotation génomique de ces interactions montre que la plus grande proportion de ces interactions concerne des boucles réunissant soit un promoteur et un enhancer soit deux enhancers (64% et 65% respectivement en EtOH et Dex) (Figure 64).

Cependant il est intéressant de noter que la distance entre deux anchors d'une même interaction est très variable et peut aller de quelques kilobases jusqu'à plusieurs mégabases. Aucune interaction trans, c'est-à-dire avec chacun des deux anchors localisés sur un chromosome différent, n'a été détectée pour les deux conditions. Ces résultats montrent également qu'un peu plus de la moitié des anchors ne sont concernés que par une seule interaction chromatinienne (Figure 63 droite). Pour le reste des anchors, ceux-ci contribuent à au moins deux interactions distinctes et ce nombre peut augmenter jusqu'à 28 pour la condition Dex. On note tout de même que les anchors des boucles chromatinienne en Dex contribuent à 3 interactions différentes (médiane) alors qu'elles ne contribuent qu'à deux interactions (médiane) en condition EtOH. Ce résultat sous-entend que les différentes régions régulatrices du génome interagissent ensemble pour former des groupes d'enhancers contribuant à moduler la transcription des gènes cibles.

De plus, j'ai annoté génomiquement ces interactions en fonction de la distance des anchors par rapport au TSS le plus proche (Figure 64). Les interactions sont annotées comme Prox-Prox si les deux anchors sont localisés à moins de 5kb de leurs TSS le plus proche, Prox-Dist lorsqu'un seul des deux anchors est localisé à moins de 5kb de son TSS le plus proche et Dist-Dist lorsque les deux anchors sont situés à plus de 5kb de leur site de transcription le plus proche. Si aucune différence n'est observée entre les deux conditions, on note qu'environ un cinquième des interactions sont de

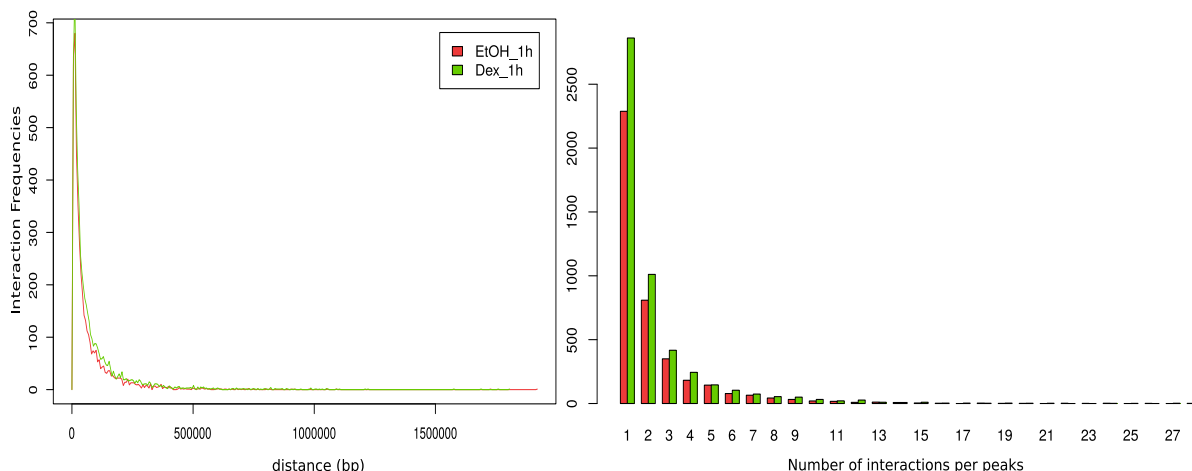


Figure 63: **Caractérisation des boucles d'interactions H3K27ac HiChIP**, Fréquence des interactions selon la distance entre les deux ancors de l'interaction (gauche) ou en fonction du nombre d'interactions associées à un ancor (droite).

la catégorie Dist-Dist et un peu plus de 40% des interactions sont de type Prox-Dist. Finalement environ 33% des interactions sont de type Prox-Prox. Ces résultats démontrent une nouvelle fois que les enhancers doivent se regrouper pour aller ensuite réguler le gène cible à travers une interaction de type Prox-Dist.

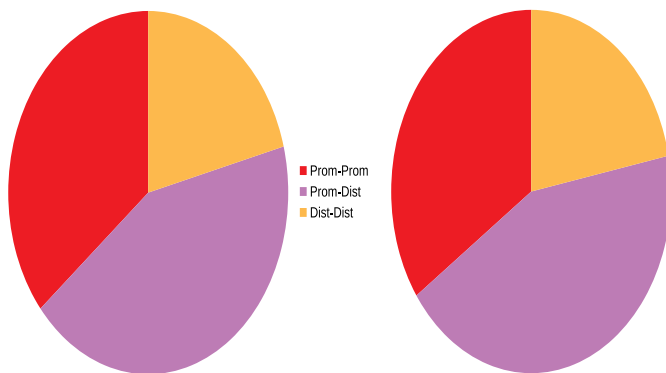


Figure 64: **Annotation génomique des boucles d'interactions H3K27ac HiChIP**, représentation des boucles obtenues en condition EtOH (Gauche) et Dex (Droite).

Pour revenir à la fixation du GR, on s'intéresse ensuite aux pics ChIP-seq GR à l'intérieur ou à l'extérieur des ancors des interactions HiChIP H3K27ac en condition Dex (Figure 65). On observe donc que les pics ChIP-seq GR localisés à l'intérieur des ancors sont de plus forte intensité comparé aux pics localisés à l'extérieur de ces ancors. Ceci souligne que les régions les plus importantes fixant le GR sont à l'intérieur d'interactions H3K27ac indiquant un rôle important du GR dans la régulation de ces boucles. En effet, on retrouve au moins un pic GR localisé dans les deux ancors pour 75% des boucles chromatinienes et 24% des ces boucles ont un pic GR dans un de leurs deux ancors.

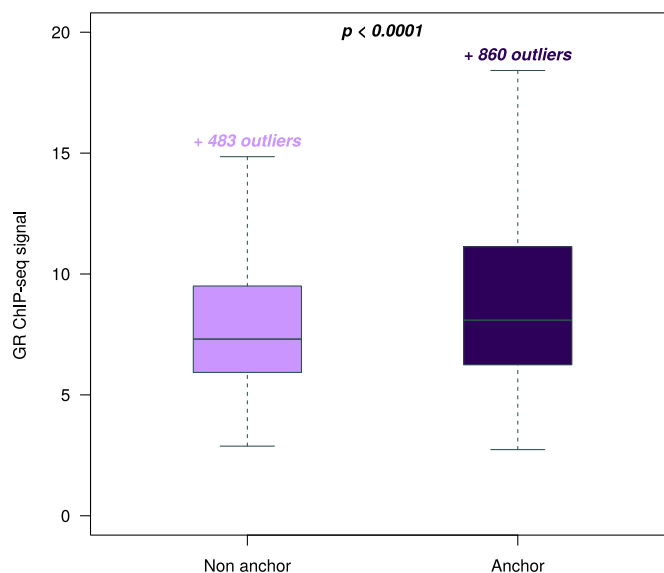


Figure 65: La fixation du GR se fait majoritairement dans les anchors des boucles d'interactions chromatinienne H3K27ac.

7.3.2 La fixation du GR dans les boucles pré-existante induit une régulation de ces boucles

Nous avons donc vu que le GR devait probablement jouer un rôle dans la régulation des boucles chromatinienne H3K27ac. Pour aller plus loin, nous avons réalisé une analyse différentielle de ces boucles en condition EtOH versus Dex (Figure 66) avec le package R edgeR (Robinson et al., 2010). Cette analyse a permis d'identifier 1,364 ($\text{LogFC} > 0.6$; $\text{FDR} < 0.05$) boucles chromatinienne qui augmentaient en réponse à la Dex (boucles Dex-induites) et 756 ($\text{LogFC} < -0.6$; $\text{FDR} < 0.05$) qui diminuaient en réponse à la Dex. Enfin, 25,765 boucles HiChIP H3K27ac ont été répertoriées comme ne bougeant pas en réponse à la Dex. Parmi les boucles Dex-induites, on retrouve des boucles de type Prox-Dist qui vont réguler des gènes précédemment identifiés en linéaire : *CXCR4*, *FKBP5*, *TSC22D3*, *BCL2L11* et *CFLAR*. Ceci implique que la régulation de l'expression de ces gènes par le GR se fait également au travers de la modulation des boucles chromatinienne. Pour aller plus loin, j'ai comparé les pics ChIP-seq GR localisés dans les anchors des boucles Dex-induites en comparaison avec les autres boucles (Figure 67). On observe une nouvelle fois que le signal ChIP-seq des pics localisés dans ces boucles est en moyenne supérieur ($p\text{-value} < 0.0001$) aux boucles qui ne varient pas. Ce résultat confirme le rôle du GR dans la régulation des boucles chromatinienne H3K27ac.

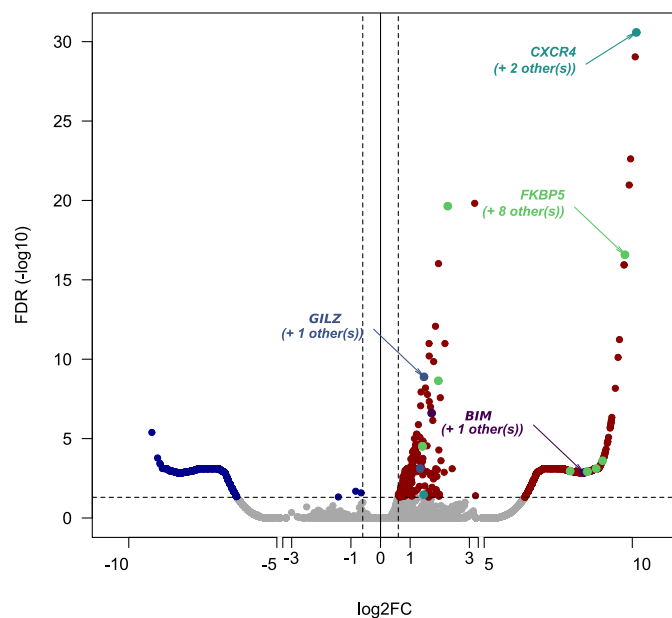


Figure 66: Analyse diff rentielle des boucles d'interactions H3K27ac en condition Dex versus EtOH.

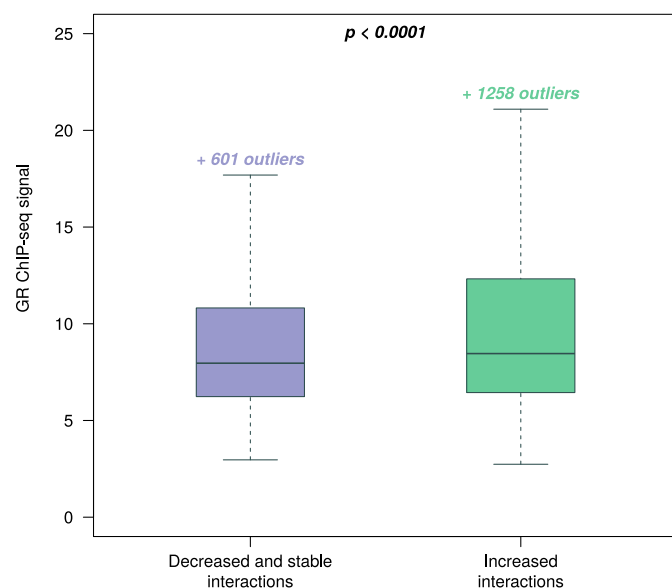


Figure 67: La fixation du GR est plus importante dans les boucles H3K27ac Dex-induites.

De plus, on constate un changement dans la r partition des cat gories d'annotation g n mique pour les boucles Dex-induites en comparaison avec les boucles qui ne bougent pas et les boucles qui diminuent (Figure 68). On observe que la proportion de boucles de type Dist-Dist est plus importante pour les boucles Dex-induites en comparaison avec les deux autres types de boucles. A l'inverse, la proportion de boucles de type Prox-Prox diminue. Ceci illustre une nouvelle fois que l'effet du GR est beaucoup plus visible au niveau des r gions r gulatrices de type enhancers plut t que dans les promoteurs.

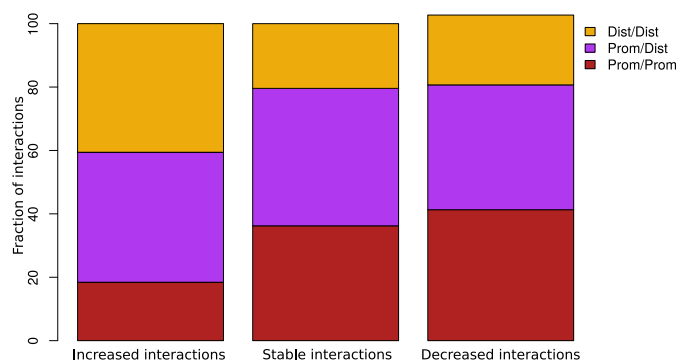


Figure 68: **Annotation g n mique des boucles chromatinienne H3K27ac dynamiques.**

Finalement, on remarque que le GR se fixe dans une forte proportion (1078/1834 soit 58.78%) des anchors des boucles Dex-induites et r gule la transcription de g nes d termin s comme  tant up-r gul s dont tous les g nes d termin s pr c demment (Figure 69). Ceci implique que la fixation du GR doit modifier la conformation de la chromatine ce qui va entra ner une r gulation des autres boucles chromatinienne.

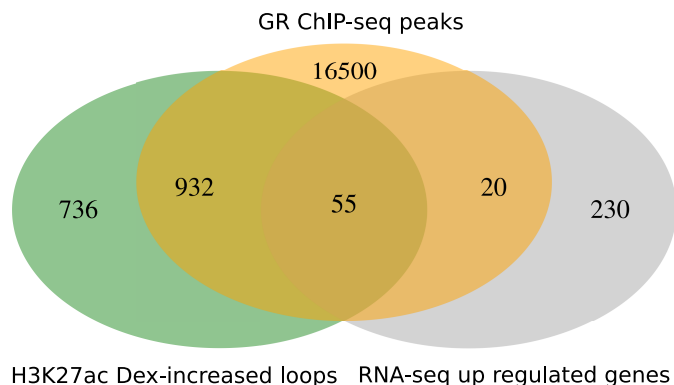


Figure 69: **Le GR se fixe dans les boucles H3K27ac Dex induites pour contr ler la transcription des g nes cl s de r ponse   la Dex.**

7.3.3 Rôle des régions H3K27ac Dex-induites au niveau des boucles de régulation

Nous nous sommes ensuite intéressés aux boucles Dex-induites en comparaison avec les régions H3K27ac Dex-induites déterminées précédemment (Figure 70). On remarque alors que plus de 40% des régions H3K27ac Dex-induites sont localisées à l'intérieur d'un anchor d'une boucle Dex-induite alors que cette proportion est de moins de 10% pour l'ensemble des enhancers et les régions H3K27ac Dex-stables. A l'inverse, quasiment aucune région H3K27ac Dex-induite n'est retrouvée dans des anchors de boucles diminuant en réponse à la Dex.

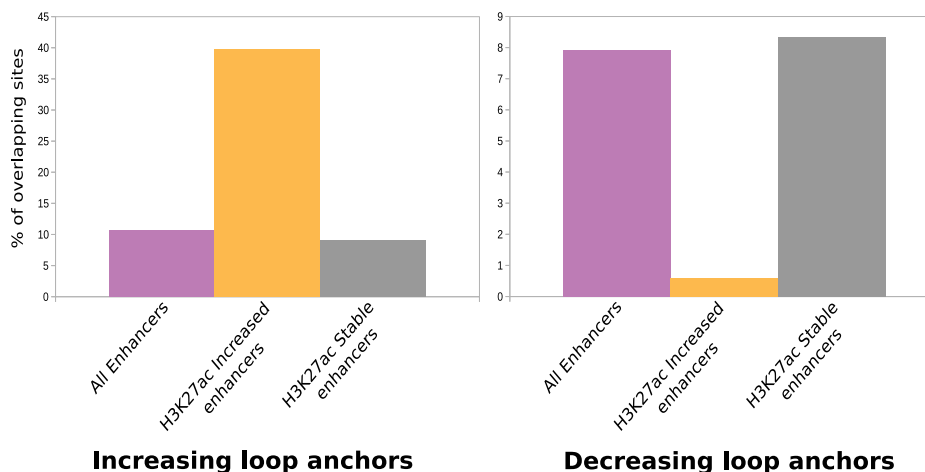


Figure 70: Les régions H3K27ac Dex-induites sont associées au boucles H3K27ac Dex-induites.

Parallèlement à cela, nous avons récupéré les anchors des boucles Dex-induites (n=1,383) et conservé parmi celles-ci uniquement les anchors contenant au moins un SE (n=420) (Figure 71). Nous nous sommes ensuite intéressés aux anchors contribuant aux interactions de type Prom-Dist (n=78) pour ainsi identifier 41 gènes up-régulés contrôlés par ce type d'interaction (Table4).

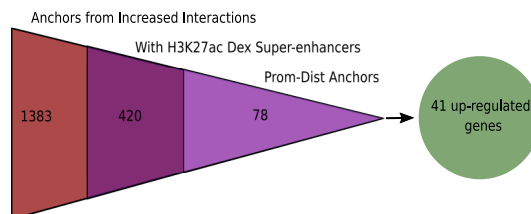


Figure 71: Les super-enhancers jouent également un rôle en contrôlant les boucles d'interactions chromatinienne et la transcription des gènes.

Gene Name	MM.1S FPKM	MM.1SDex FPKM	LogFC
<i>AANAT</i>	0.16	0.87	2.48
<i>ASCC1</i>	6.28	14.15	1.17
<i>ATP11B</i>	50.13	168.49	1.75
<i>BCL2L11</i> (BIM)	51.85	109.51	1.08
<i>C1QTNF6</i>	0.18	0.40	1.18
<i>CFLAR</i>	48.12	125.79	1.39
<i>CLEC3B</i>	1.99	6.15	1.63
<i>CREB3L2</i>	29.52	122.10	2.05
<i>F2RL3</i>	2.11	4.49	1.09
<i>FKBP5</i>	21.60	333.90	3.95
<i>FZD4</i>	0.24	4.79	4.31
<i>FZD7</i>	7.38	18.59	1.33
<i>INSR</i>	9.25	35.04	1.92
<i>ISG20</i>	62.53	176.47	1.50
<i>KCNJ15</i>	1.25	5.54	2.14
<i>KLF2</i>	23.60	92.31	1.97
<i>KLHL24</i>	9.21	20.54	1.16
<i>KLHL6</i>	0.06	0.83	3.77
<i>LRRK1</i>	5.85	18.75	1.68
<i>MAP3K5</i>	6.35	16.59	1.39
<i>MEI1</i>	17.84	63.73	1.84
<i>MTMR4</i>	17.05	41.54	1.28
<i>P4HA1</i>	15.53	108.70	2.81
<i>PFKFB2</i>	10.00	49.97	2.32
<i>PIGV</i>	5.70	14.06	1.30
<i>PIK3IP1</i>	15.96	71.01	2.15
<i>RGS1</i>	61.82	151.82	1.30
<i>RHOB</i>	43.80	103.39	1.24
<i>SCML4</i>	0.10	0.62	2.62
<i>SERTAD2</i>	13.42	35.64	1.41
<i>SMAP2</i>	52.43	281.99	2.43
<i>SOCS1</i>	4.41	8.91	1.01
<i>SPTBN1</i>	38.01	118.59	1.64
<i>ST6GAL1</i>	361.39	802.59	1.15
<i>TFAP2C</i>	3.26	8.26	1.34
<i>TP53INP1</i>	65.75	132.41	1.00
<i>TRIB1</i>	11.78	26.69	1.18
<i>TSC22D3</i> (GILZ)	20.25	396.80	4.29
<i>ZC3H12A</i>	8.81	84.57	3.26
<i>ZFP36L2</i>	98.66	314.90	1.67

Table 4: Gènes induits contrôlés par au moins une interaction chromatinienne qui augmente en réponse à la Dex et co-localisée avec un SE MM.1SDex.

Parmi les 41 gènes retrouvés régulés à la fois par un SE et par au moins une boucle Dex-induite, on retrouve les gènes *TSC22D3* (GILZ), *DDIT4*, *FKBP5*, *BCL2L11* (BIM) et *CREB3L2* alors que le gène *CXCR4* est seulement régulé par une boucle Dex-induite. Ceci renforce l'idée d'un regroupement d'enhancers lié à la fixation de la Dex qui va permettre l'activation de la transcription du gène cible.

Cette hypothèse est confirmée par l'exemple du locus *FKBP5* (Figure 72) où l'on peut constater que le GR se fixe sur des régions contenant un GRE. La fixation du GR dans ces sites induit une acétylation des régions et la formation de deux super-enhancers. L'ensemble de ces enhancers va interagir pour former un cluster qui va aboutir à l'augmentation forte de l'expression du gène *FKBP5*. D'autres exemples de ce mécanisme peuvent être observés sur les locus des gènes *TSC22D3*, *CXCR4*

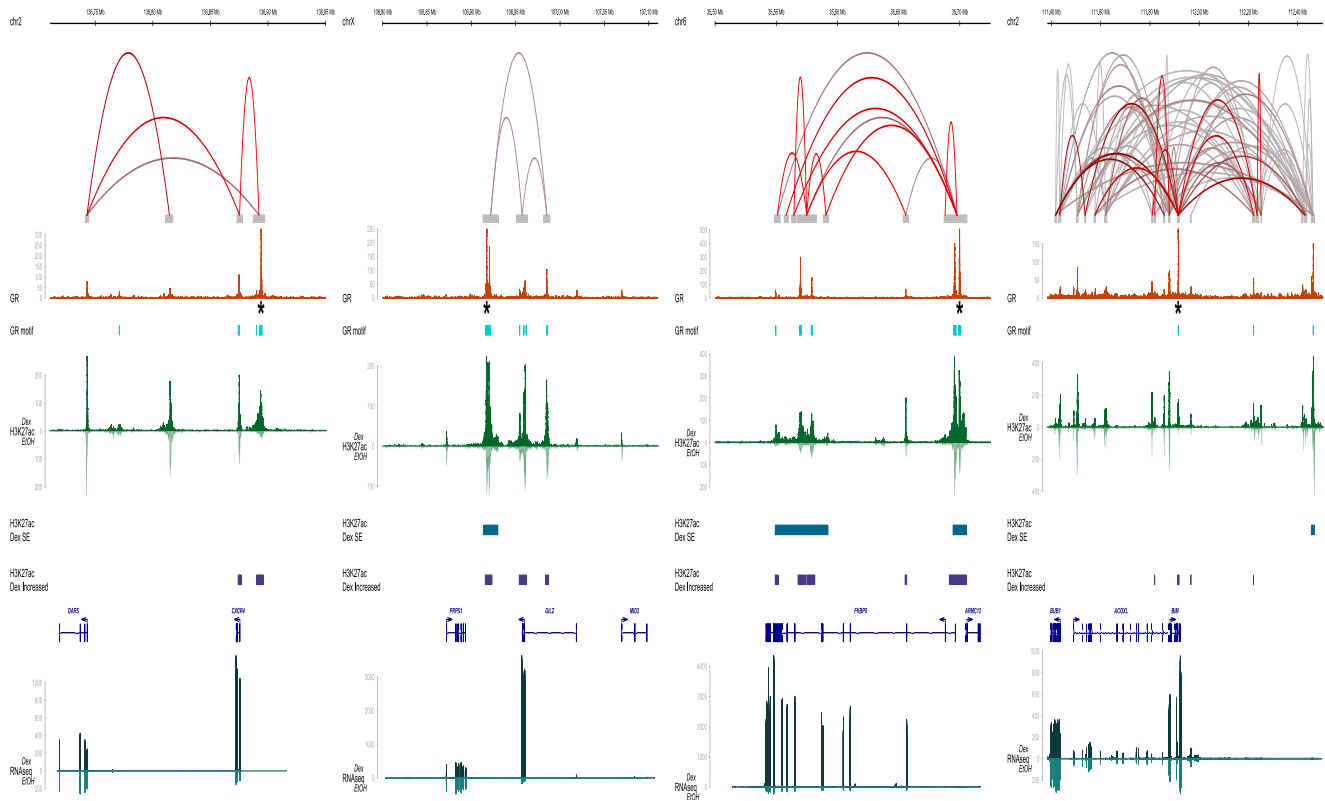


Figure 72: Les interactions H3K27ac permettent une organisation des enhancers en CLIQUES pour réguler la transcription des gènes, illustration de ce mécanisme sur les locus des gènes *CXCR4*, *TSC22D3* (GILZ), *FKBP5* et *BCL2L11* (BIM). (*) symbolise fixation principale du GR dans la région.

et *BCL2L11* (Figure 72) où l'on peut remarquer que pour chacun de ces loci, il y a une fixation principale du GR sur son motif consensus qui entraîne à la fois une augmentation de l'acétylation des régions et l'augmentation de la fréquence des boucles d'interaction chromatinienne. L'ensemble de ces phénomènes conduit alors pour ces trois locus à une augmentation de la transcription du

gène cible.

Parmi les réseaux de régulations identifiés, nous avons ensuite identifié le pic GR majoritaire. Ce pic possède le signal ChIP-seq le plus intense dans la région. Sur ce pic, le GR se fixe également sur son motif consensus. Nous avons émis l'hypothèse que cet enhancer principal est prédéterminant dans la régulation de la transcription du gène cible. Pour vérifier cette hypothèse, j'ai donc construit les réseaux de régulation sur les locus des gènes *BCL2L11* (Figure 73a) et *FKBP5* (Figure 73b). Pour ces deux exemples, j'ai modélisé l'absence d'interactions liées à l'enhancer principal (cercle vert). On observe dans les deux cas, lorsque cet enhancer est retiré, que le réseau de régulation est fortement perturbé. La majorité des boucles chromatinienes qui augmentent en réponse à la Dex disparaissent et le promoteur du gène cible est régulé par beaucoup moins d'enhancers Dex-induits.

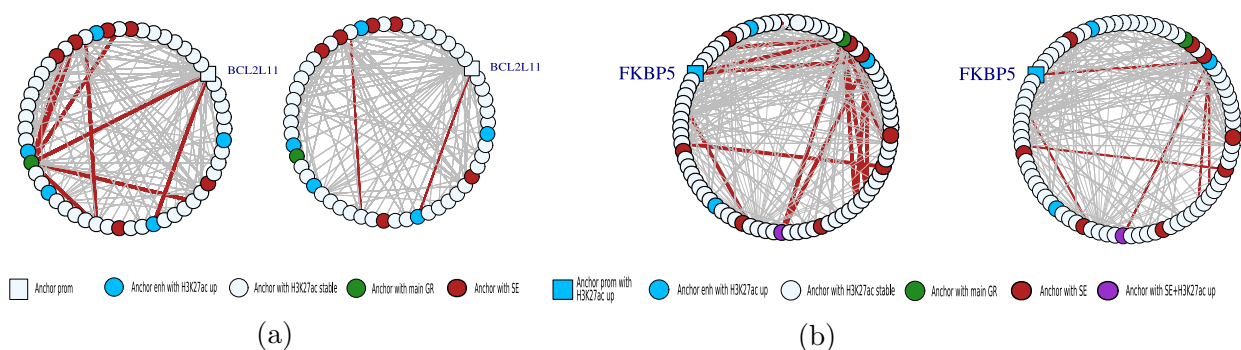


Figure 73: **Le rôle de l'enhancer majoritaire dans le recrutement d'enhancers secondaires pour contrôler la transcription des gènes**, exemple de régulation du gène *BCL2L11* (a) et *FKBP5* (b) par le réseau d'interaction HiChIP H3K27ac. Le pic principal (enhancer vert) représente le site principal de fixation du GR. Le réseau est montré avec les interactions liées à l'enhancer principal (gauche) ou sans (droite).

Ces résultats nous indiquent que les réseaux de régulation formés autour d'un gène en réponse à la Dex dépendent principalement d'un enhancer majoritaire dont l'activation permet le recrutement d'enhancers secondaires contribuant à augmenter la transcription du gène cible.

7.3.4 Le CTCF maintient les boucles chromatinienes stables

Des études récentes ont démontré le rôle de la protéine CTCF dans le maintien et la stabilisation des boucles d'interactions (Ren et al. (2017); Kubo et al. (2021); D'Ippolito et al. (2018)). Pour étudier ce phénomène, nous avons généré des données de ChIP-seq en condition EtOH et Dex afin de tenter d'expliquer comment sont modulées les boucles de régulation chromatinienne.

Tout d'abord, nous avons comparé les deux conditions (Figure 74). La heatmap illustrant le signal CTCF sur l'ensemble des pics en condition EtOH et Dex révèle que seulement un très faible nombre de sites de faible intensité varie en réponse à la Dex. Ceci indique que la fixation du CTCF sur le génome n'est en rien influencée par le traitement à la Dex.

Nous avons ensuite regardé pour tous les ancors des boucles chromatinienes si chacun de ces

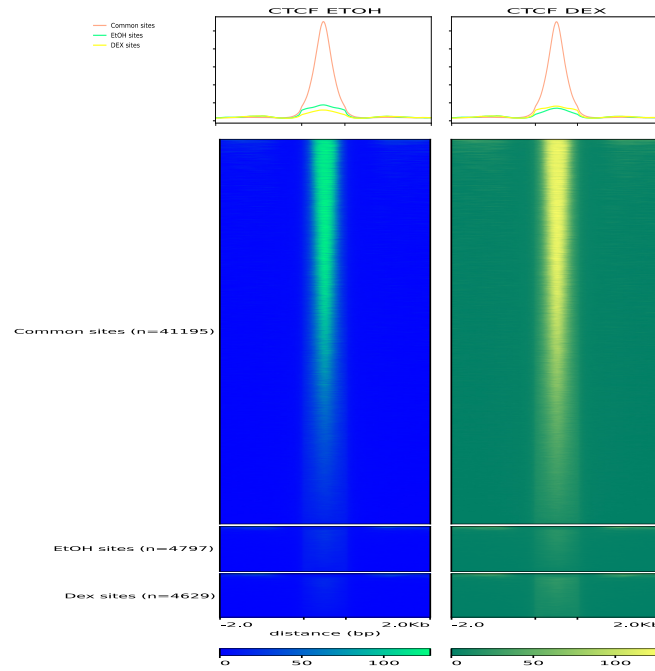


Figure 74: Heatmap illustrant la fixation du CTCF en condition EtOH et Dex.

anchors poss dait ou non un site CTCF (Figure 75). Lorsque l'on regarde l'ensemble des interactions ou les interactions stables, on remarque qu'une faible proportion de ces interactions n'a pas de pic CTCF dans aucun de leurs anchors et   l'inverse ont une grande proportion d'interactions stabilis es par un pic CTCF dans chacun des deux anchors. En ce qui concerne les boucles Dex-induites, la proportion de boucles qui ne contiennent aucun CTCF dans aucun des deux anchors est beaucoup plus importante et une proportion beaucoup plus faible de ces boucles ont un CTCF dans les deux anchors. Cela souligne que ces interactions sont donc beaucoup moins stables   travers une population cellulaire.

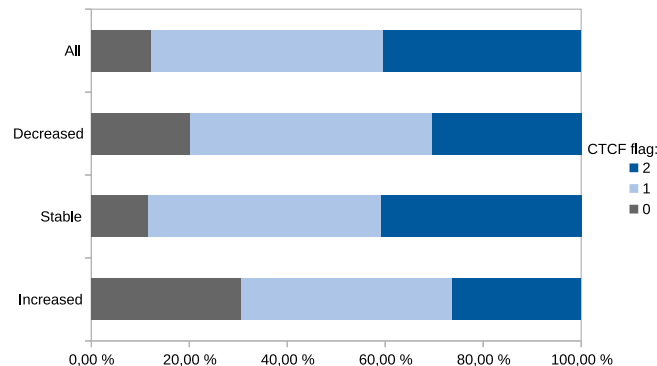


Figure 75: Le facteur CTCF permet de stabiliser les interactions stables et dynamiques des boucles H3K27ac Dex-induites.

7.3.5 La régulation des boucles chromatinienne est permise par l'ouverture de régions GR-spécifiques dans les anchors de ces interactions

En se focalisant sur les boucles Dex-induites, nous nous sommes ensuite intéressés à l'accessibilité de la chromatine au sein des anchors de ces interactions.

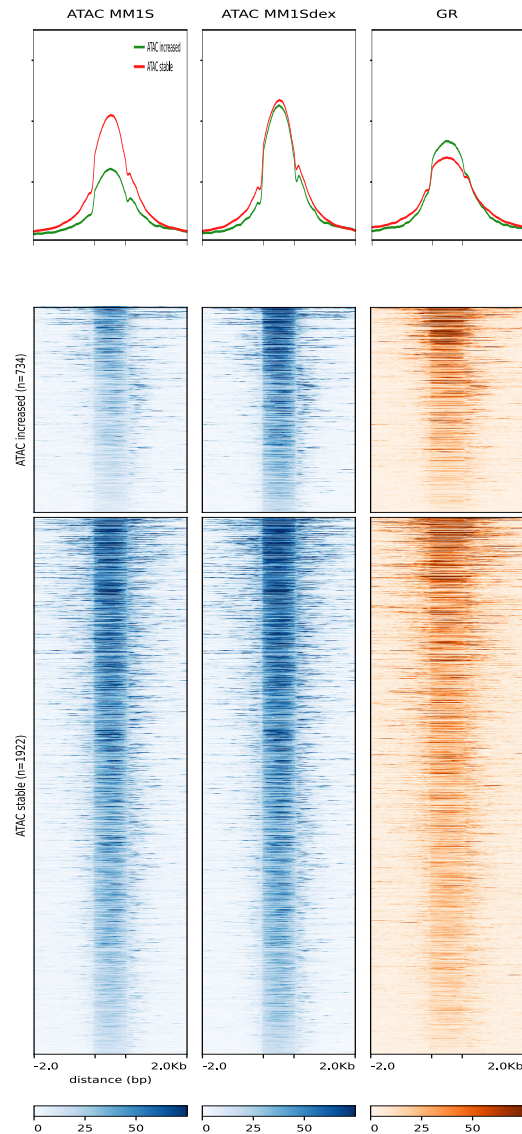


Figure 76: **Heatmap illustrant l'augmentation de l'accessibilité de la chromatine dans certaines régions ATAC-seq présentes dans les anchors des boucles H3K37ac Dex-induites.**

A l'aide d'une analyse différentielle des pics ATAC-seq en condition Dex versus EtOH, nous avons identifié 734 régions ATAC Dex-induites contre 1,922 régions ATAC Dex-stables (Figure 76). Pour

les r gions ATAC Dex-induites, nous observons une augmentation significative du signal ATAC-seq en Dex ainsi qu'un signal ChIP-seq GR l g rement plus fort que dans les ATAC Dex-stables, ce qui sugg re que le GR permet l'ouverture de r gions accessibles au sein des boucles Dex-induites. Une d couverte de motif *de novo*   ensuite  t  r alis e dans les r gions ATAC Dex-induites et Dex-stables (Figure 77). Certains motifs ont  t  retrouv s enrichis dans les deux cat gories comme par exemple le motif du facteur de transcription RUNX d j  mis en  vidence avec les facteurs de transcription coeurs de la cellule de my lome. Pour ce qui est des r gions ATAC Dex-induites, on remarque la pr sence de deux motifs importants correspondant   un demi-motif fixant IRF4 et   un motif GRE (Figure 77a). A l'inverse, nous ne voyons aucun enrichissement du motif GRE dans les r gions ATAC Dex-stables (Figure 77b). Et plus encore, un motif double ISRE est retrouv  enrichi dans ces r gions. D'autres motifs fixant des facteurs de transcription importants pour la cellule MM.1S comme AP-1 et NF- B sont  galement retrouv s. Ces r sultats sugg rent que l'ouverture de la chromatine dans les r gions ATAC-seq Dex-induites est permise par le GR en s'associant avec le complexe SWI/SNF pour ouvrir de nouvelles r gions de fixation directe. On peut  galement penser que la diff rence entre un demi-motif IRF4 et le motif long ISRE retrouv  respectivement dans les r gions ATAC Dex-induites et Dex-stables peut  tre expliqu  par le type d'interaction physique entre IRF4 et le GR. Dans les r gions ATAC Dex-induites, le GR doit pouvoir se fixer   c t  du monom re IRF4 alors que dans les ATAC qui ne varient pas, la r gion est d j  accessible et un homo-dim re IRF4 y est fix  et le GR doit venir se fixer de fa on indirecte.

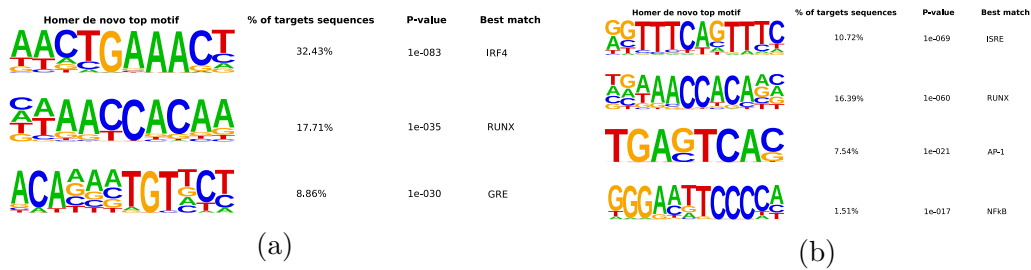


Figure 77: Recherche de motif *de novo* dans les r gions ATAC Dex-induites (a) et ATAC Dex-stables (b).

Ces r sultats peuvent  galement  tre observ s via le diff rentiel des interactions chromatininiennes (Figure 78). Cette figure nous montre que la plupart des boucles Dex-induites ont dans au moins un de leurs anchors un pic ATAC Dex-induit.

En comparant ces r gions ATAC Dex-induites avec les enhanceurs H3K27ac MM.1S en condition Dex (Figure 79), nous observons que 249 r gions (249/734 soit 33.92%) sont localis es   l'int rieur de SE et que le reste de ces r gions (442/734 soit 60.22%) est retrouv  au niveau d'un pic H3K27ac en condition Dex. Ce r sultat indique que l'ouverture des r gions est li    l'ac tylation et donc   la fixation du GR.

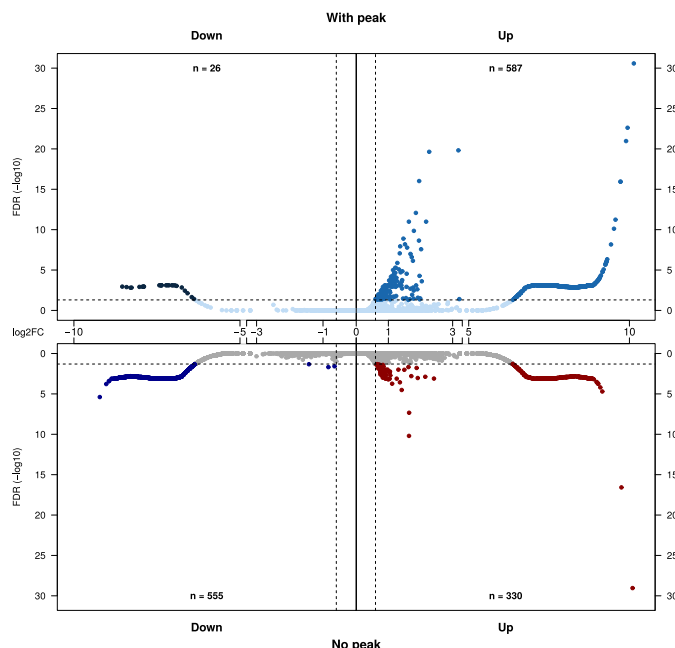


Figure 78: Volcano plot illustrant la répartition des régions ATAC Dex-induites sur le différentiel des boucles d'interactions H3K27ac.

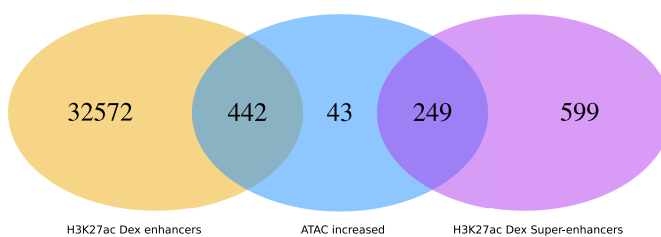


Figure 79: Co-localisation des régions ATAC Dex-induites avec les super-enhancers et enhancers H3K27ac en condition Dex.

L'ensemble de ces résultats montre donc que la réponse à la Dex est contrôlée par de nombreux mécanismes épigénétiques au sein d'une population cellulaire. Ces résultats posent la question de savoir si tous ces sites réagissent de façon homogène lorsque l'on regarde les cellules individuellement. En effet, il a été démontré qu'il existait une importante hétérogénéité intra-tumorale dans la pathologie du Myélome Multiple. On sait également que tous les patients ne répondent pas à la Dex de la même manière.

Pour répondre à ces différentes questions, nous avons donc effectué une étude en cellule unique pour essayer de mieux comprendre la réponse à la Dex dans cette lignée cellulaire.

7.4 Hétérogénéité de l'accessibilité de la chromatine en réponse à la Dex

La première partie de cette étude a donc révélé le rôle pré-déterminant d'un enhancer particulier au sein d'un réseau de régulation. Dans cette seconde partie, nous nous sommes donc intéressés à l'étude des régions régulatrices en cellule unique afin de voir si ces clusters d'enhancers identifiés en bulk s'ouvrent de façon simultanée au sein des mêmes cellules en réponse à la Dex.

Afin de répondre à cette question, nous avons réalisé un séquençage en cellule unique ATAC-seq pour les cellules MM.1S avec Dex ($0.1\mu\text{M}$) ou EtOH après 1 heure et 4 heures. Les données ont été normalisées et soumises à une approximation uniforme et projetées afin de réduire les dimensions (UMAP; Figure 80).

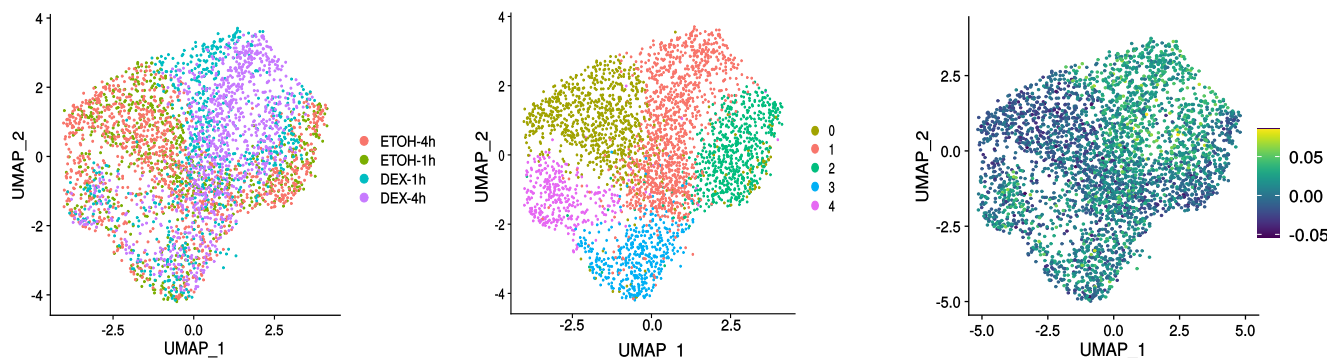


Figure 80: **Clustering des cellules en réponse à la Dex ou non selon l'accessibilité de la chromatine**, clustering selon la condition (gauche), groupes identifiés (milieu) et selon les régions ouvertes Dex-induites détectées en bulk (droite).

On peut remarquer en premier lieu qu'au sein de notre population cellulaire, une grande proportion des cellules se distinguent selon le traitement qu'elles ont reçu (Figure 80 gauche). Il est intéressant de noter que la distinction entre cellules traitées à 1 heure et à 4 heures est moins visible même si on observe plus de différences dans la condition Dex.

Les clusters de cellules ont ensuite été calculés et coloriés pour définir 5 clusters distincts, le cluster 0 nous permet d'identifier presque uniquement des cellules exclusivement traitées avec l'EtOH alors qu'au contraire, le cluster 1 est composé exclusivement de cellules traitées par la Dex (Figure 80 milieu) ce qui indique que le paysage chromatinien détecté en cellule unique différencie particulièrement les cellules traitées par la Dex des cellules traitées par l'EtOH. Parmi ces 5 clusters, seuls ces deux clusters permettent de distinguer efficacement les deux conditions (Dex et EtOH). On note cependant que le cluster 4 semble regrouper des cellules en phase G2M du cycle cellulaire (Figure 81).

Nous nous sommes ensuite intéressés à l'accessibilité de la chromatine sur les 734 régions régulatrices Dex-induites détectées avec les données d'ATAC-seq en bulk (Figure 76; Figure 80 droite). Nous avons observé que ce set de régions constituait une signature des cellules ayant eu un traitement par la Dex comparé aux cellules traitées avec l'EtOH.

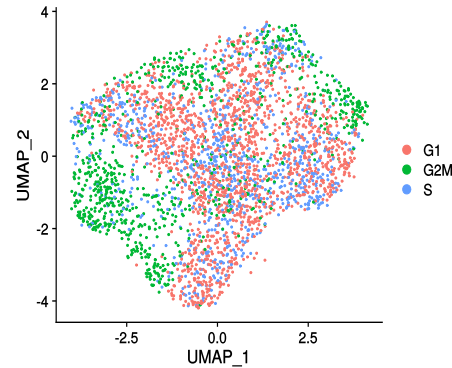


Figure 81: **Clustering des cellules en fonction de la phase du cycle cellulaire.**

Nous avons ensuite estimé le ratio d'accessibilité de ces régions génomiques dans chaque cellule après 1 heure de traitement par la Dex ou l'EtOH (ROR). En se basant sur ce calcul dérivé de la méthode proposé par l'équipe d'Hoffman (Hoffman et al., 2020), nous avons trouvé que les régions étaient significativement plus accessibles dans chaque cellule traitée par la Dex en comparaison des cellules traitées par l'EtOH (Figure 82; $p\text{-val} < 0.0001$).

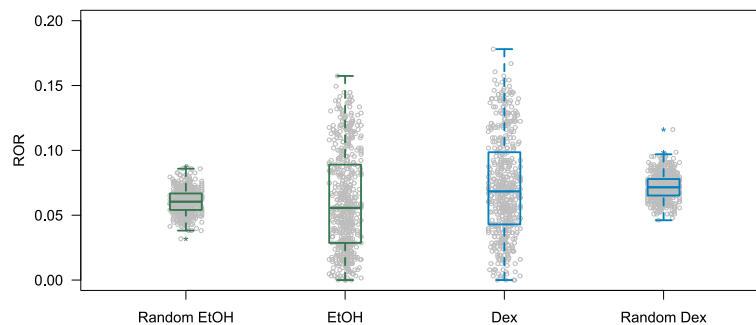


Figure 82: **Estimation du ratio de régions accessibles (ROR) sur les régions ATAC Dex-induites**

Ensuite, nous avons examiné les différences d'accessibilité de la chromatine entre les cellules traitées par la Dex ou l'EtOH dans les régions régulatrices autour des gènes *GILZ*, *FKBP5*, *BIM* et *CXCR4* (Figure 83 à 86). Sur ces régions, nous avons pris comme référence les pics ChIP-seq GR détectés en bulk au sein d'ancres d'interactions chromatiniques dynamiques et pour lesquels un motif GRE était retrouvé. Ainsi, plusieurs sites ont été identifiés sur les régions régulatrices des gènes *GILZ* et *FKBP5* (Figure 83 et 84; G1 à G4) et un seul site dans les régions régulatrices des gènes *BIM* et *CXCR4* (Figure 85 et 86).

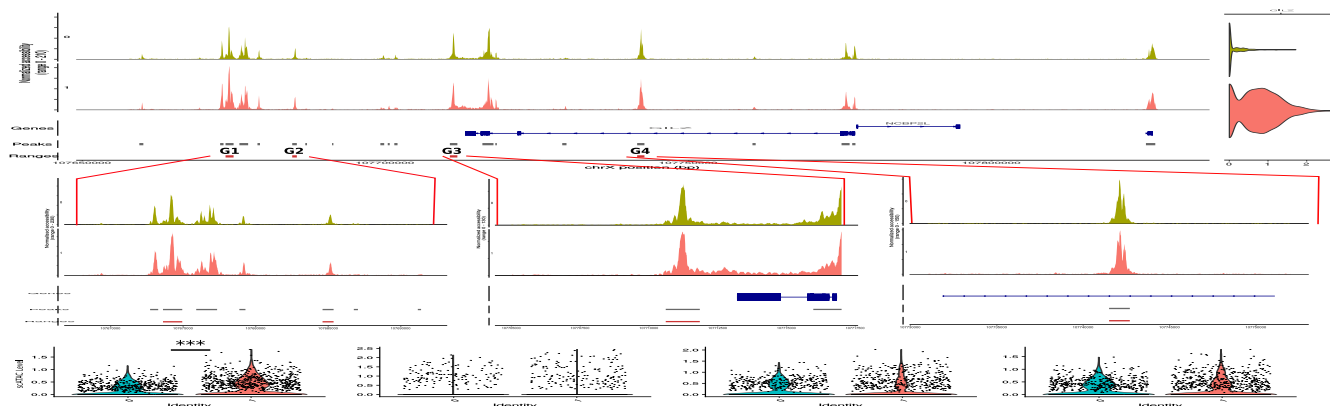


Figure 83: **Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique.** Visualisation sur le locus du gène *TSC22D3* (*GILZ*). Les résultats sont montrés pour le cluster 0 (Dex) et le cluster 1 (EtOH). Les blocs rouges représentent les régions fixant le GR détectées en bulk.

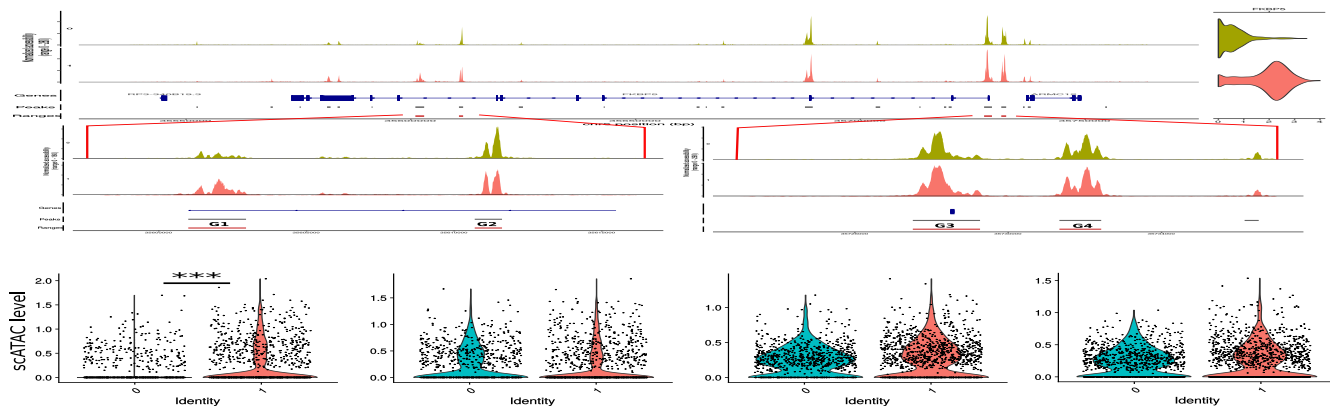


Figure 84: **Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique.** Visualisation sur le locus du gène *FKBP5*. Les résultats sont montrés pour le cluster 0 (Dex) et le cluster 1 (EtOH). Les blocs rouges représentent les régions fixant le GR détectées en bulk (G1, G2, G3 et G4).

Sur ces différentes régions, nous notons que seulement un site augmentait en accessibilité pour le cluster correspondant au cellules traitées par la Dex (cluster 1). Nous avons remarqué également que pour la majorité des cas (sauf pour *FKBP5*), ce changement d'accessibilité était observé dans l'enhancer pré-dominant de la région.

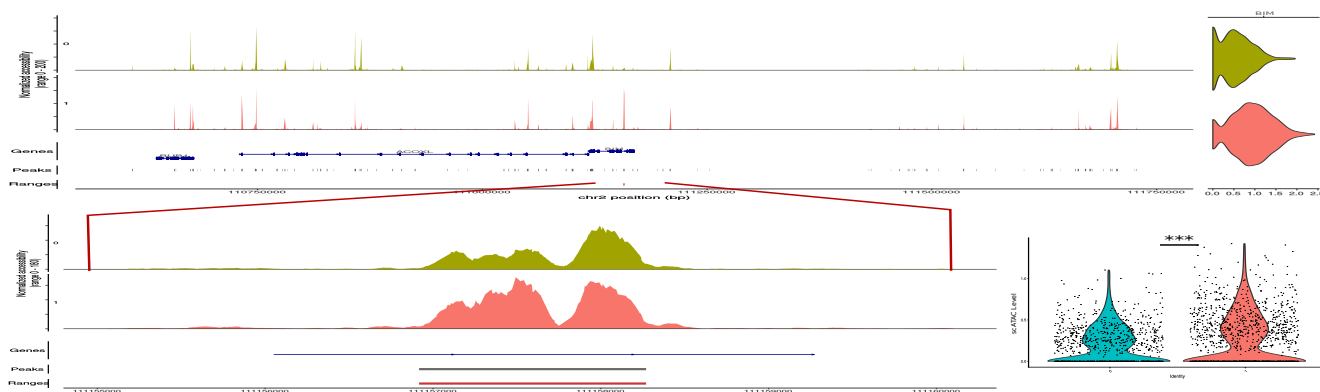


Figure 85: **Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique.** Visualisation sur le locus du gène *BCL2L11* (BIM). Les résultats sont montrés pour le cluster 0 (Dex) et le cluster 1 (EtOH). Les blocs rouges représentent les régions fixant le GR détectées en bulk.

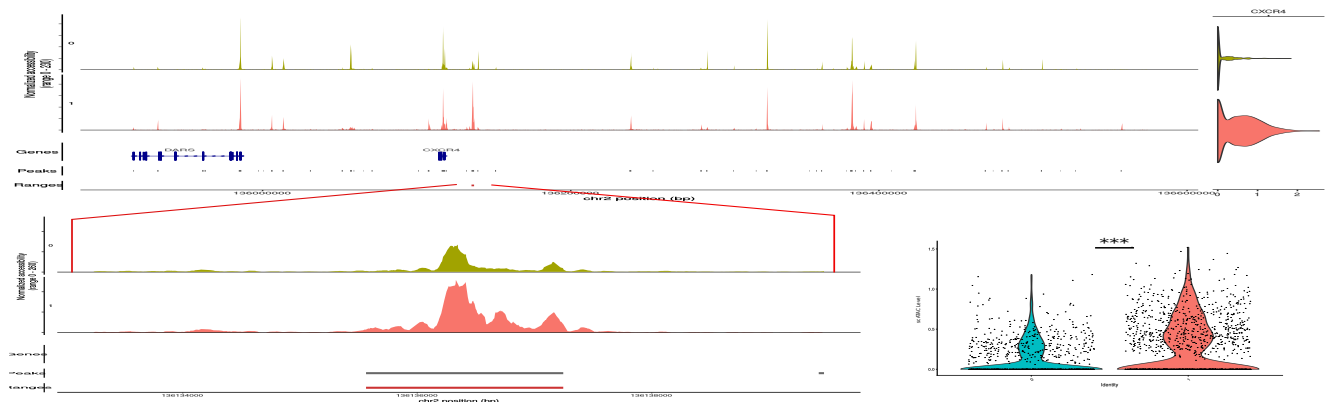


Figure 86: **Visualisation de l'ouverture de régions en lien avec la transcription du gène en cellule unique.** Visualisation sur le locus du gène *CXCR4*. Les résultats sont montrés pour le cluster 0 (Dex) et le cluster 1 (EtOH). Les blocs rouges représentent les régions fixant le GR détectées en bulk.

Finalement, dans l'objectif de montrer que ce gain d'accessibilité de la chromatine était variable dans les cellules MM.1S, les régions enhancers régulant les 41 gènes identifiés en bulk (Figure 71 et Table 4) avec une augmentation d'accessibilité de la chromatine significative ont été sélectionnées

et le ratio de cellules accessibles (RAC) à été calculé. Comparé à la médiane de base (basée sur les cellules traitées à l'EtOH), la médiane de cellules accessibles est significativement plus élevée après traitement par la Dex ($p\text{-value}=0.0003$) ce qui souligne plutôt une augmentation d'accessibilité de la chromatine pour ces régions enhancers dans chaque cellule (Figure 87). De plus, les valeurs de RAC sont beaucoup plus étendues dans les cellules traitées avec la Dex ce qui suggère une importante variabilité dans le nombre de cellules ayant leurs enhancers régulant les gènes de réponse à la Dex accessibles.

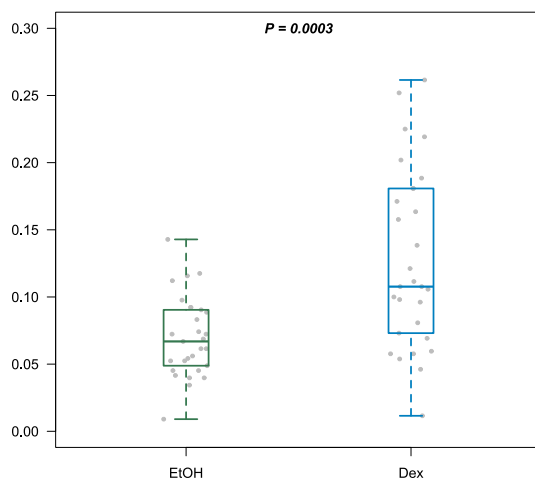


Figure 87: Estimation du ratio de cellules accessibles (RAC) basé sur les régions enhancers de 26 des 41 gènes Dex-up-régulés.

7.5 Hétérogénéité transcriptomique en réponse à la Dex

L'intégration des données pan-génomiques de ChIP-seq GR, d'HiChIP H3K27ac et de RNA-seq en bulk nous a permis d'identifier un set de gènes de réponse à la Dex. Ces gènes sont imbriqués dans une dynamique d'interaction chromatinienne similaire, alors que les analyses d'ATAC-seq en cellule unique ont révélé une variabilité de l'accessibilité de la chromatine au sein des enhancers prédominants dans la régulation des gènes de réponse à la Dex. Par conséquent, une hétérogénéité transcriptionnelle cellule à cellule peut être impliquée parmi ces gènes.

Pour évaluer cette hétérogénéité, nous avons réalisé des analyses de RNA-seq en cellule unique pour des cellules MM.1S récupérées après 4 heures et 24 heures en présence de Dex ($0.1\mu\text{M}$) ou d'EtOH. Nous nous sommes focalisés sur les gènes les plus fortement induits par la Dex. Nous avons alors analysé la distribution des logFC supérieurs à 0 et obtenu une distribution bimodale comprenant une petite partie des logFC positifs étant éloignés de la part principale, nous avons ensuite utilisé un modèle gaussien pour identifier une petite sous-population de 51 gènes avec un logFC élevé appelée single-cell Dex-activated genes (scDAGs) (Figure 88).

Tous les scDAGs exceptés 2 gènes (*PTMA* et *TXNIP*) ont été identifiés dans les trois jeux de données (ChIP-seq GR, HiChIP H3K27ac et RNA-seq en bulk; Figure 71).

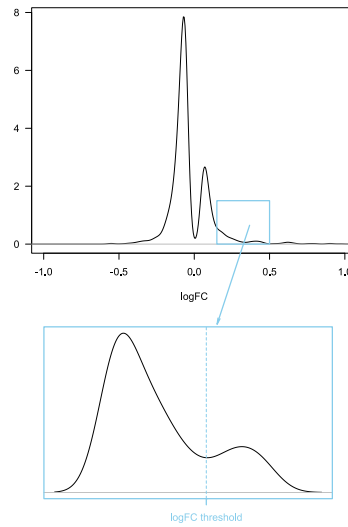


Figure 88: **Identification des logFC élevés après application d'un modèle Gaussien.**

Dans le but de mesurer les niveaux de variabilité transcriptionnelle cellule à cellule, nous avons calculé les informations d'entropie d'expression des scDAGs (Pastore et al. (2019); Landau et al. (2014)) et trouvé que parmi les cellules exposées à la Dex, l'expression médiane augmentait conduisant à une augmentation de la fraction de cellule exprimant le gène (fpc) (Figure 89).

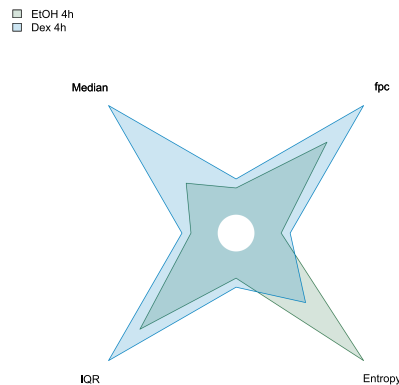


Figure 89: **Diagramme illustrant la médiane, le rang inter-quartile, la fraction par cellule (fpc) et l'entropie pour les cellules traitées en EtOH ou en Dex, pour chaque paramètre, les valeurs maximales sont utilisées en référence.**

Par conséquent, nous avons observé une diminution de l'entropie, c'est-à-dire moins d'hétérogénéité entre les cellules. Cependant, l'expression des gènes entre les cellules reste hétérogène, avec un rang inter-quartile (IQR) encore élevé.

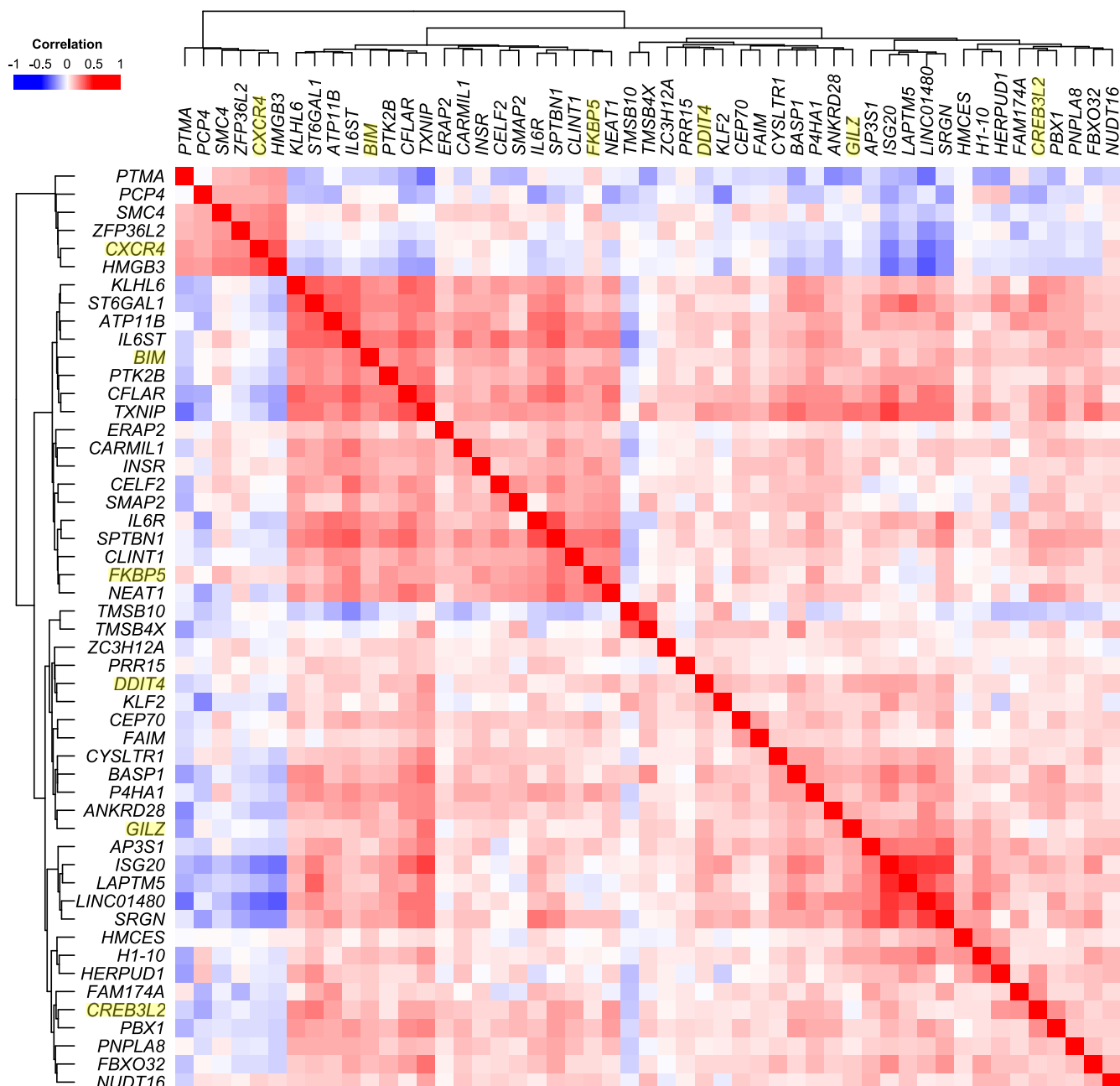


Figure 90: Heatmap montrant la co-expression des 51 gènes les plus différentiels entre EtOH et Dex.

Afin d'aller plus loin dans l'analyse de la variabilité transcriptionnelle qui persiste après exposition à la Dex, nous avons étudié les relations entre les 51 scDAGs (Figure 90B; Figure 91 (b) et (c)). Bien que ces corrélations ne soient pas très fortes, nous avons observé une meilleure corrélation entre les scDAGs par rapport à des gènes pris aléatoirement (Figure 92).

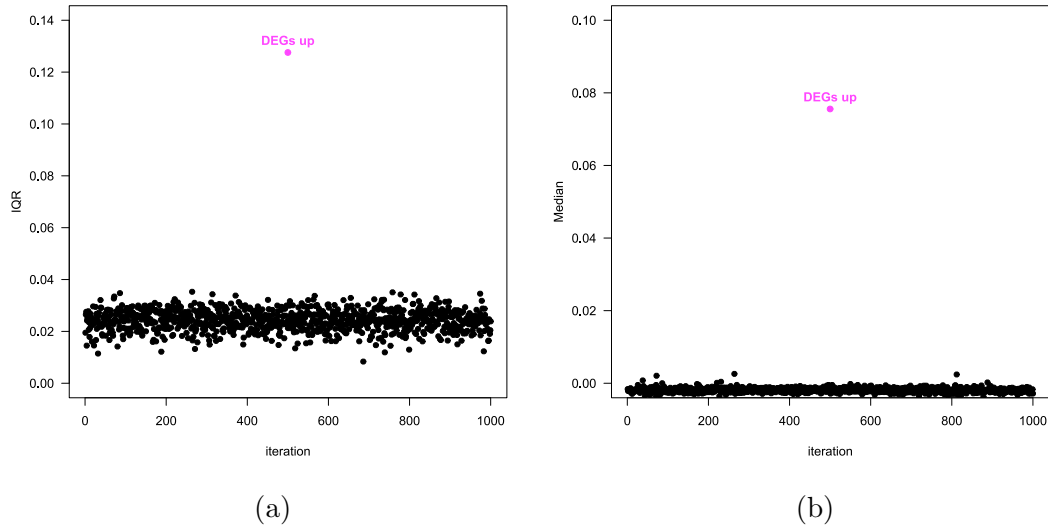


Figure 91: **Comparaison des IQR et des médianes des scDAGs par rapport à des gènes sélectionnés aléatoirement.** (a) rang inter-quartiles (IQR). (b) médianes.

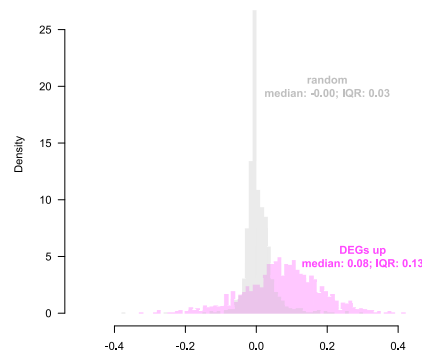


Figure 92: **Distribution de la densité des coefficients de corrélation pour des gènes sélectionnés aléatoirement et les scDAGs.**

Nous avons remarqué que parmi l'ensemble des 51 scDAGs, ceux-ci se regroupaient en deux groupes principaux, un large cluster (cluster 1) incluant le gène pro-apoptotique *BIM* et des gènes de réponse à la Dex ubiquitaires comme *GILZ*, *FKBP5* et *DDIT4* et un second cluster 2 regroupant 6 gènes dont *CXCR4* (Figure 90). Ces résultats suggèrent qu'au sein des cellules traitées par la Dex, il existe une population de cellules répondant fortement à ce traitement et qui expriment une large majorité des scDAGs (cluster 1) et qu'une autre population va moins répondre à ce traitement en exprimant un nombre réduit de scDAGs. Pour vérifier cette hypothèse, nous avons tout d'abord utilisé une méthode récemment proposée par l'équipe d'Hoffman (Hoffman et al., 2020), afin d'estimer le nombre de scDAGs exprimés dans chaque cellule après 4h et 24h d'exposition à la Dex ou à l'EtOH (Figure 95; Figure 94).

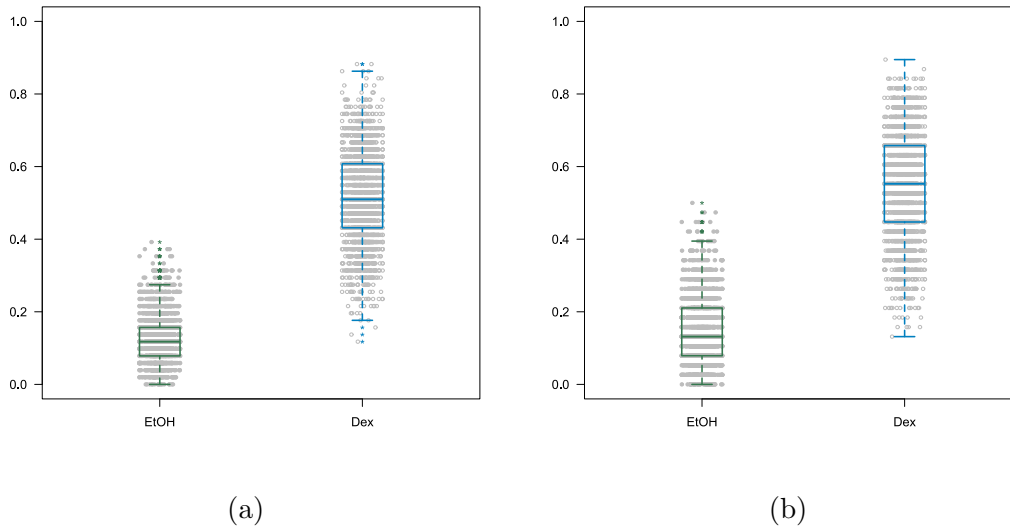


Figure 93: **Estimation du ratio de gènes répondant à la Dex (RRG)**, ratios calculés à 4h (a) et 24h (b).

Nous avons trouvé que les cellules traitées par l'EtOH avaient un ratio de gènes répondant à la Dex (RRG) médian de 11.8% (6/51 scDAGs) alors que les cellules traitées par la Dex ont un RRG médian de 51.0% (26/51 scDAGs). Cette valeur de RRG est également similaire pour les cellules traitées par la Dex à 24 heures (41.2%; 21/51 scDAGs).

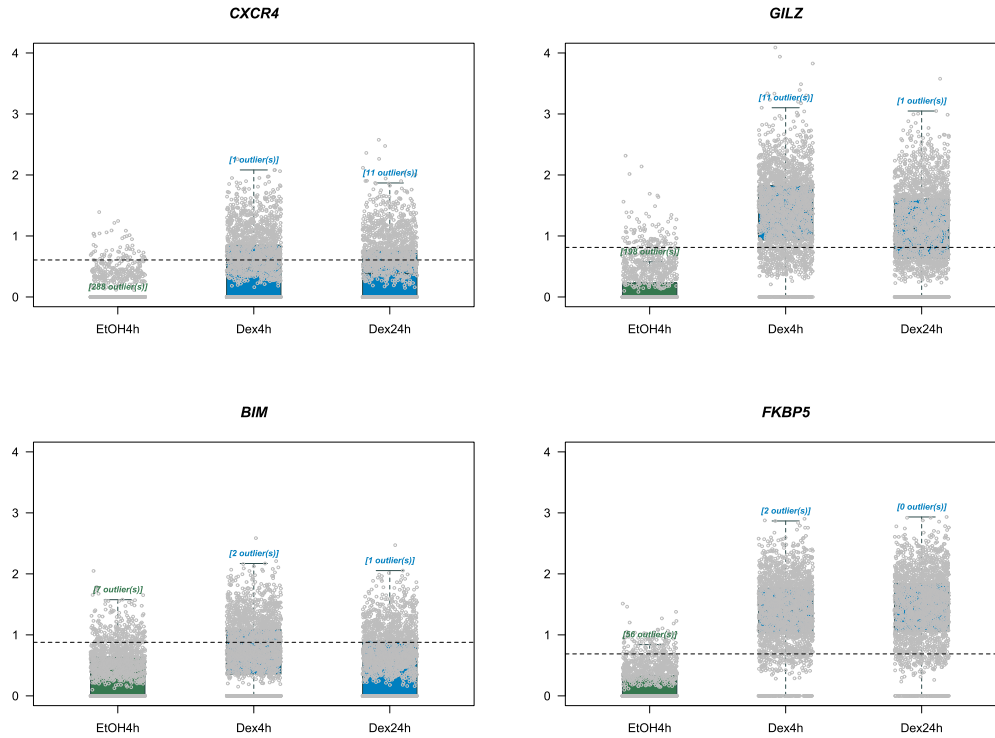


Figure 94: **Boxplots illustrant la distribution des valeurs scRNA-seq pour des gènes d'intérêt**, les valeurs sont montrées pour les gènes après 4h d'exposition à l'EtOH et 4 heures et 24 heures d'exposition à la Dex. Les valeurs de chaque cellule sont représentées. La ligne pointillée représente le seuil spécifique du gène.

Nous avons ensuite visualisé la variabilité transcriptionnelle des scDAGs en réalisant une réduction de dimension grâce aux méthodes de d'approximation et de projection uniforme des données de single-cell (UMAP) (Becht et al., 2019), comme nous pouvions le penser, la réduction UMAP sépare parfaitement les cellules en fonction de la condition (Figure 95 gauche).

Finalement, nous avons coloré cette UMAP en fonction du RRG de chaque cellule (Figure 95 droite). Il est intéressant de noter que bien qu'il existe une importante hétérogénéité cellule à cellule, les cellules qui répondent fortement à la Dex ont tendance à se regrouper alors que les cellules répondant peu sont localisées autour. Les UMAP fusionnées et colorées selon l'expression des gènes anti-corrélés *BIM* et *CXCR4* ($R^2 = -0.02$) montre clairement que dans la plupart des cellules, l'expression de ces deux gènes est mutuellement exclusive (Figure 96).

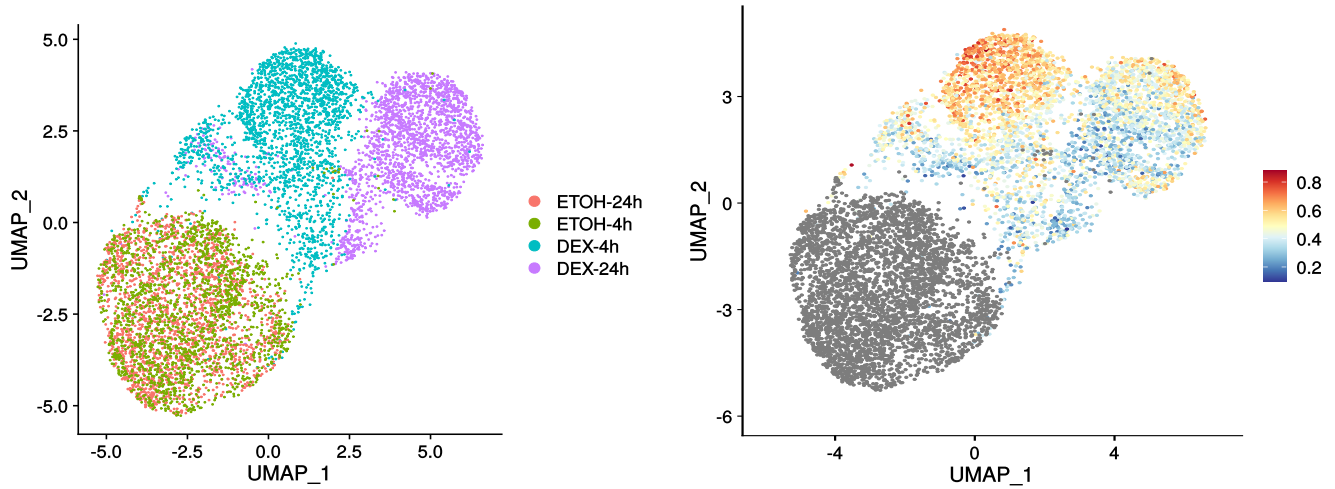


Figure 95: **UMAP single cell RNAseq**, les cellules sont colorées selon le traitement (gauche) ou selon le ration de gènes réponses (RRG) de chaque cellules (droite).

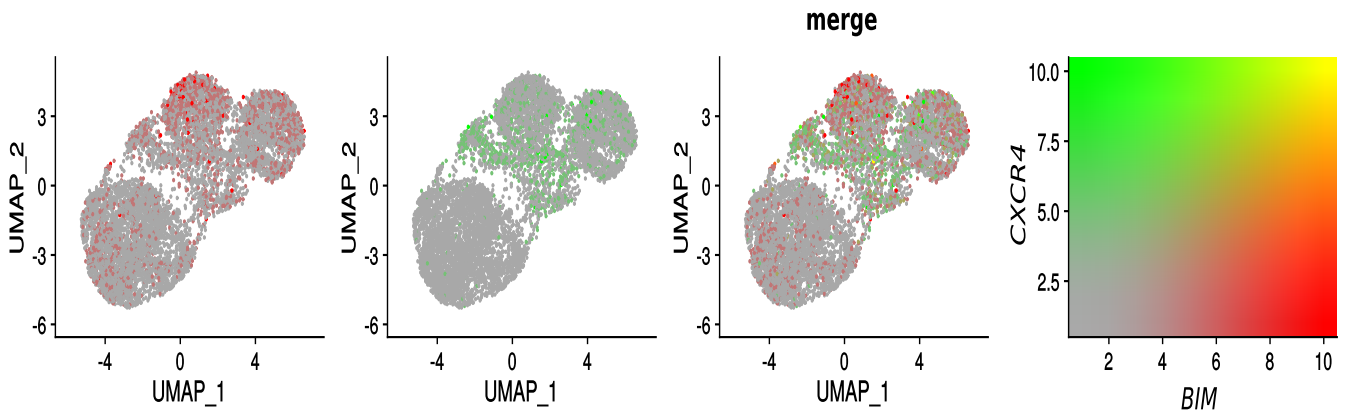


Figure 96: **UMAP** révélant l'expression anti-corrélée de *BIM* et *CXCR4* dans les cellules, les UMAP sont colorées selon l'expression de chaque cellule du gène *BIM* (rouge) et *CXCR4* (vert).

Les analyses scRNA-seq révèlent qu'en moyenne, les cellules MM.1S expriment seulement la moitié de l'ensemble des gènes de réponses à la Dex. De plus, pour la majeure partie des cellules à faible réponse (i.e. $RR < 50\%$) le gène induisant la mort cellulaire induit par les GCs le plus important (*BIM*) n'est pas exprimé.

8 Discussion et conclusion

8.1 Discussion

Bien que l'efficacité de la Dex dans le MM puisse largement être attribuée à l'apoptose induite par le GR, la réponse génomique au traitement à la Dex dans les cellules plasmocytaires de myélome reste méconnue. Étant donné que la Dex est utilisée à chaque étape du traitement, il était urgent d'étudier le mode d'action moléculaire de la Dex au moyen de nouveaux outils génomiques afin de mieux comprendre la résistance au traitement et de donner de nouvelles perspectives d'options de combinaisons de thérapies.

Dans cette étude, nous avons pu confirmer que le paysage épigénétique spécifique des cellules plasmocytaires était maintenu par le facteur pionnier lignage spécifique IRF4 et que le GR se fixait de façon omniprésente sur de nombreux sites d'enhancers actifs pré-programmés. Cependant, malgré une forte association de la marque H3K27ac avec la fixation du GR dans des enhancers distaux impliqués dans des interactions de longue portée pour former des réseaux lignage spécifique, le changement d'activité enhancer provoqué par le traitement à la Dex est limité à un faible nombre d'enhancers distaux. Ces éléments contribuent à la régulation de gènes de réponse à la Dex d'une manière combinatoire. En effet, notre étude a montré que la fixation du GR modifie la marque H3K27ac à travers des clusters d'enhancers recouvrant plusieurs kilo-bases d'ADN génomique appelés SE. Ces enhancers dynamiques sont associés avec des promoteurs de gènes cibles de la Dex via des interactions chromatiniennes et assurent une forte expression du gène. De plus, nos cartes de connections régulatrices à haute résolution ont révélé que la fixation du GR active non seulement un large éventail d'enhancers mais aussi permet l'augmentation des fréquences d'interactions des boucles d'ADN régulatrices reliant les multiples enhancers distaux à leurs gènes cibles précédemment évoqués en tant que cliques d'enhancers (Mumbach et al., 2017). Ce mode d'action est similaire à ce qui a déjà été décrit pour la voie NOTCH (Wang et al. (2014); Petrovic et al. (2019)). Globalement, nos résultats ont démontré que cette combinaison spécifique d'enhancers collaborent entre eux pour produire une réponse propre et sélective des GCs. Parmi ces enhancers, nous avons montré l'importance du motif GRE et de la distance au TSS pour la fixation du GR en accord avec d'anciennes études (Vockley et al. (2016); McDowell et al. (2018)). Ensemble, ces données suggèrent l'existence d'une hiérarchie au sein des clusters d'enhancers avec un enhancer prédominant et des enhancers secondaires qui collaborent suivant la fixation du GR. Ce résultat a déjà été décrit dans d'autres lignages (Shin et al. (2016); Carleton et al. (2017)). L'équipe de Jing et al. a montré qu'un enhancer spécifique est nécessaire pour réguler l'interaction avec le promoteur pour activer la transcription de *BCL2L1* (BIM) (Jing et al., 2018). Dans cette étude, nous montrons également que cet enhancer est impliqué dans la plupart des interactions dynamiques en réponse à la Dex. Nous pouvons donc en déduire que cet enhancer est l'élément régulateur central qui collabore d'une façon additive avec d'autres enhancers afin de permettre l'expression pleine de BIM. Avec ces résultats, l'ingénierie génomique ciblée pourrait être utilisée pour tester directement cette hypothèse et déterminer l'importance des autres enhancers (Thomas et al., 2021).

Comme nous avons trouvé que l'accessibilité de la chromatine était en corrélation avec la fréquence d'interaction des enhancers entre eux ou avec des gènes cibles proches, les données de single-cell

ATAC-seq et RNA-seq ont été intégrées avec nos jeux de données HiChIP et ChIP-seq afin d'étudier l'accessibilité de la chromatine de ces enhancers dans chaque cellule individuellement. Nos résultats montrent que le traitement Dex provoque et augmente le nombre de régions accessibles dans une même cellule en particulier pour les enhancers prédominants. Malgré ces résultats, l'accessibilité de la chromatine reste variable entre les cellules. Bien sûr, l'augmentation de l'accessibilité de la chromatine n'est pas toujours détectée lors des analyses ATAC-seq en bulk. Ces résultats mis ensemble suggèrent qu'une cellule, après traitement par la Dex, est plus sujette à montrer toutes les bonnes connectivités pour l'expression du gène cible.

Pour certains gènes, l'exposition à la Dex ne change pas l'accessibilité de la chromatine. Dans ces cas-là, le GR dans ces régions Dex-induites peut agir en tant que stabilisateur des boucles d'interactions chromatiniques pré-déterminées comme suggéré par les travaux de D'ippolito (D'Ippolito et al., 2018). Même si d'autres études seront requises pour analyser l'hétérogénéité chromatinique parmi les cellules individuellement en couplant des études de conformation de chromatine en cellule unique (ChIA-Drop, Zheng et al. (2019)) avec des études épigénomiques en cellule unique (protocoles CUT&Run et CUT&Tag; Hainer et al. (2019); Kaya-Okur et al. (2019)), notre étude démontre l'importance de prendre en considération l'ensemble des enhancers impliqués dans la régulation spatiale d'un gène ainsi que le niveau d'analyse en cellule unique afin de mieux comprendre le mode d'action des facteurs de transcription.

Il était intéressant pour nous de montrer que la variabilité d'accessibilité de chromatine était associée à une réponse hétérogène à la Dex dans les cellules de Myélome Multiple. Ce résultat a été montré avant pour le cancer du poumon (Hoffman et al., 2020). En effet, nous avons trouvé une variabilité intra-cellulaire en regardant l'expression de gènes de réponse à la Dex, premièrement chaque cellule exprimait en moyenne approximativement la moitié des réseaux transcriptionnels induits par la Dex et dans un second temps, un sous-ensemble des cellules n'expriment pas BIM, un médiateur majeur de l'apoptose par la Dex dans les cellules de MM (Kervoëlen et al., 2015). Nos résultats mettent en évidence la réciprocity complexe entre la modulation de l'accessibilité cellule à cellule et les boucles d'interactions E-P et E-E qui augmentent avec la fixation du GR, particulièrement illustré par les expressions mutuellement exclusives de *BIM* et *CXCR4* et nous offre de nouvelles connaissances sur les mécanismes de résistance aux drogues en considérant que le niveau de GR peut être un événement limitant pour la réponse à la Dex (Kervoëlen et al. (2015); Heuck et al. (2012)).

Finalement, nos résultats montrent que les facteurs de transcription Ikaros (IKZF1) et Aiolos (IKZF3) sont parmi les quelques co-interacteurs du GR suggérant que ces facteurs peuvent jouer un rôle dans la réponse à la Dex. Une étude récente montre que ces TFs sont dégradés par les IMiDs (Sievers et al., 2018). Depuis que ces deux drogues sont combinées pour traiter les patients atteints de MM, nous ne pouvons pas exclure un rôle antagoniste de ces molécules, des études approfondies sont nécessaires pour identifier les gènes cibles, s'il y en a, du complexe GR-IKZF1/3.

Le fait que *CXCR4*, connu pour être important pour le myélome et sa subsistance (Alsayed et al., 2007), est up-régulé avec l'exposition à la Dex et que cette expression est restreinte aux cellules qui n'expriment pas le gène pro-apoptotique *BIM* soulève la question de la résistance à cette drogue. Un essai de phase Ib/II utilisant une combinaison d'anticorps IgG4 anti-CXCR4 humains en combinaison avec des lenalidomides et de la Dex ou des bortezomib et de la Dex montrent un fort taux de réponse (Ghobrial et al., 2019). Il serait intéressant de poursuivre dans cette voie afin de

déterminer si, dans cette combinaison, la Dex pourrait être plus efficace.

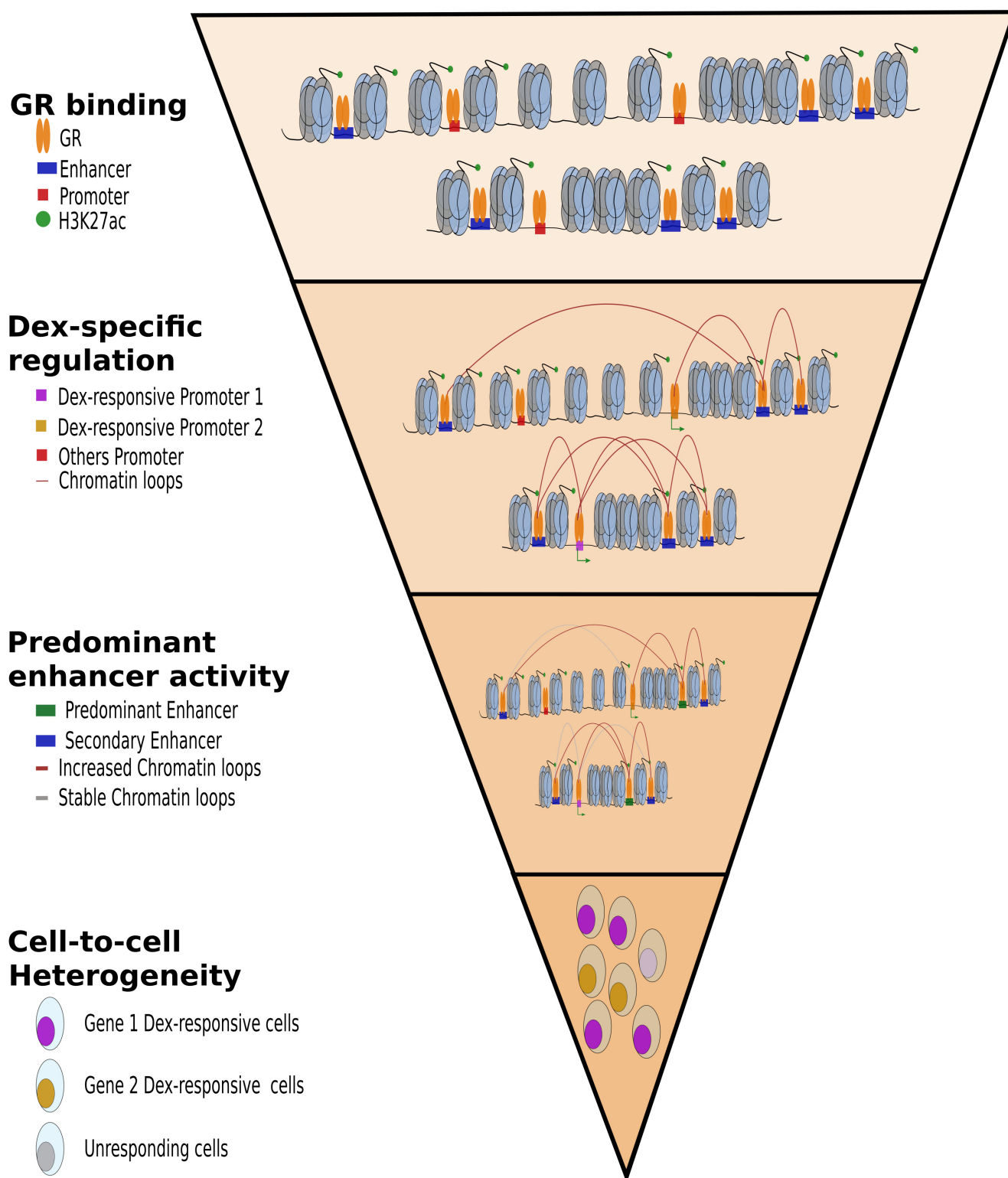


Figure 97: Schéma récapitulant l'effet de la fixation du GR sur la régulation des gènes de réponse à la Dex.

8.2 Conclusion

Cette thèse m'a donc permis d'appréhender une question biologique à l'aide de moyens bioinformatiques. Les analyses pan-génomiques réalisées ont permis de mieux comprendre le mécanisme de régulation de la transcription du GR au sein d'une cellule. Malgré ces résultats intéressants, de nombreux points restent encore à élucider pour totalement comprendre les mécanismes de régulation de la Dex. Au travers de cette étude j'ai eu la chance de pouvoir manipuler toutes sortes de données multi-omiques en bulk et en cellule unique et de me poser les questions sur les différentes méthodes d'analyse et sur les interprétations pour répondre à une problématique biologique.

Dans la continuité de cette étude nous envisageons de mettre en place une étude de séquençage en cellule unique de chromosome conformation capture (HiC) dans notre modèle cellulaire mais également chez les patients atteints de Myélome à différents stades de la maladie. Cette étude permettrait notamment de mieux comprendre l'hétérogénéité conformationnelle des patients en réponse à la Dex.

D'un autre point de vue, une étude plus fonctionnelle sera également menée grâce aux progrès des biotechnologies afin de décrire le rôle fonctionnel des enhancers en réponse à la Dex. Pour cela nous envisageons de réaliser des techniques CRISPR/CAS9 sur des enhancers identifiés dans cette étude afin d'étudier le rôle de ces enhancers dans la formation de clusters (cliques) et donc le recrutement d'enhancers secondaires pour contrôler la transcription du gène.

Finalement, comme décrit lors de la présentation de la pathologie, les mécanismes de résistance à la Dex sont à ce jour encore méconnus du fait, une nouvelle fois d'une importante hétérogénéité. Les résultats ont montré que tous les patients ne répondaient pas de la même manière à la Dex et donc leurs tumeurs ne résistent pas de la même façon. Pour cela, une troisième étude pourra être mise en place incluant un autre modèle de lignée de myélome: la MM.1R issue de la MM.1S qui est résistante aux glucocorticoïdes par délétion d'un allèle du gène *NR3C1* et mutation sur l'autre allèle. Dans cette étude, nous comptons analyser les différents mécanismes présentés dans cette thèse à travers la lignée MM.1R mais aussi au cours du temps en réalisant une cinétique des données de HiChIP-seq.

References

- Yazan Alsayed, Hai Ngo, Judith Runnels, Xavier Leleu, Ujjal K Singha, Costas M Pitsillides, Joel A Spencer, Teresa Kimlinger, Joanna M Ghobrial, Xiaoying Jia, et al. Mechanisms of regulation of cxcr4/sdf-1 (cxcl12)-dependent migration and homing in multiple myeloma. *Blood*, 109(7):2708–2717, 2007.
- SF Altekruze, CL Kosary, M Krapcho, N Neyman, R Aminou, W Waldron, J Ruhl, N Howlader, Z Tatalovich, H Cho, et al. Seer cancer statistics review, 1975-2007, national cancer institute. *Bethesda (MD), based on November*, 2009.
- Haley M Amemiya, Anshul Kundaje, and Alan P Boyle. The encode blacklist: identification of problematic regions of the genome. *Scientific reports*, 9(1):1–5, 2019.
- Simon Andrews et al. Fastqc: a quality control tool for high throughput sequence data, 2010.
- Francisco Antequera. Structure, function and evolution of cpg island promoters. *Cellular and Molecular Life Sciences CMLS*, 60(8):1647–1658, 2003.
- Emira Ayroldi, Antonio Macchiarulo, and Carlo Riccardi. Targeting glucocorticoid side effects: selective glucocorticoid receptor modulator or glucocorticoid-induced leucine zipper? a perspective. *The FASEB Journal*, 28(12):5055–5070, 2014.
- Timothy L Bailey, James Johnson, Charles E Grant, and William S Noble. The meme suite. *Nucleic acids research*, 43(W1):W39–W49, 2015.
- Elaine Beaulieu and Eric F Morand. Role of gilz in immune regulation, glucocorticoid actions and rheumatoid arthritis. *Nature Reviews Rheumatology*, 7(6):340, 2011.
- Etienne Becht, Leland McInnes, John Healy, Charles-Antoine Dutertre, Immanuel WH Kwok, Lai Guan Ng, Florent Gehroux, and Evan W Newell. Dimensionality reduction for visualizing single-cell data using umap. *Nature biotechnology*, 37(1):38–44, 2019.
- Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics*, 30(15):2114–2120, 2014.
- Stefano Bruscoli, Valerio Donato, Enrico Velardi, Moises Di Sante, Graziella Migliorati, Rosario Donato, and Carlo Riccardi. Glucocorticoid-induced leucine zipper (gilz) and long gilz inhibit myogenic differentiation and mediate anti-myogenic effects of glucocorticoids. *Journal of Biological Chemistry*, 285(14):10385–10396, 2010.
- Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang, and William J Greenleaf. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dna-binding proteins and nucleosome position. *Nature methods*, 10(12):1213, 2013.
- Jason D. Buenrostro, Beijing Wu, Howard Y. Chang, and William J. Greenleaf. Atac-seq: A method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology*, 109(1):21.29.1–21.29.9, 2015. doi: 10.1002/0471142727.mb2129s109. URL <https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/0471142727.mb2129s109>.

REFERENCES

- Jordi Camps, Michael R Erdos, and Thomas Ried. The role of lamin b1 for the maintenance of nuclear structure and function. *Nucleus*, 6(1):8–14, 2015.
- Julia B Carleton, Kristofer C Berrett, and Jason Gertz. Multiplex enhancer interference reveals collaborative control of gene regulation by estrogen receptor α -bound enhancers. *Cell systems*, 5(4):333–344, 2017.
- Michael A Chapman, Michael S Lawrence, Jonathan J Keats, Kristian Cibulskis, Carrie Sougnez, Anna C Schinzel, Christina L Harview, Jean-Philippe Brunet, Gregory J Ahmann, Mazhar Adli, et al. Initial genome sequencing and analysis of multiple myeloma. *Nature*, 471(7339):467–472, 2011.
- Daniel M Cohen and David J Steger. Nuclear receptor function through genomics: lessons from the glucocorticoid receptor. *Trends in Endocrinology & Metabolism*, 28(7):531–540, 2017.
- Davi Collares. Post-translational modifications of relb nf-kb subunit and associated functions. *Cells*, 5: 22, 05 2016. doi: 10.3390/cells5020022.
- ENCODE Project Consortium et al. An integrated encyclopedia of dna elements in the human genome. *Nature*, 489(7414):57, 2012.
- M Ryan Corces, Jason D Buenrostro, Beijing Wu, Peyton G Greenside, Steven M Chan, Julie L Koenig, Michael P Snyder, Jonathan K Pritchard, Anshul Kundaje, William J Greenleaf, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nature genetics*, 48(10):1193–1203, 2016.
- Diana Cruz-Topete and John A Cidlowski. One hormone, two actions: anti-and pro-inflammatory effects of glucocorticoids. *Neuroimmunomodulation*, 22(1-2):20–32, 2015.
- Yulia N Demchenko, Oleg K Glebov, Adriana Zingone, Jonathan J Keats, P Leif Bergsagel, and W Michael Kuehl. Classical and/or alternative nf- κ b pathway activation in multiple myeloma. *Blood, The Journal of the American Society of Hematology*, 115(17):3541–3552, 2010.
- MA Dimopoulos, E Kastritis, L Rosinol, J Bladé, and H. Ludwig. Pathogenesis and treatment of renal failure in multiple myeloma. *Leukemia*, 22(8):1485–1493, 2008.
- Anthony M D’Ippolito, Ian C McDowell, Alejandro Barrera, Linda K Hong, Sarah M Leichter, Luke C Bartelt, Christopher M Vockley, William H Majoros, Alexias Safi, Lingyun Song, et al. Pre-established chromatin interactions mediate the genomic response to glucocorticoids. *Cell systems*, 7(2):146–160, 2018.
- Laurent Dortet, RÃ©my Bonnin, and Thierry Naas. Impact du s  quen  age d’adn    haut d  bit sur la surveillance des   pid  mies de bact  ries multi-r  sistantes aux antibiotiques. *Feuilles de biologie*, 354, 01 2017.
- Sean R Eddy. What is a hidden markov model? *Nature biotechnology*, 22(10):1315–1316, 2004.
- Jan B Egan, Chang-Xin Shi, Waibhav Tembe, Alexis Christoforides, Ahmet Kurdoglu, Shripad Sinari, Sumit Middha, Yan Asmann, Jessica Schmidt, Esteban Braggio, et al. Whole-genome sequencing of multiple myeloma from diagnosis to plasma cell leukemia reveals genomic initiating events, evolution, and clonal tides. *Blood, The Journal of the American Society of Hematology*, 120(5):1060–1066, 2012.

REFERENCES

- Jason Ernst and Manolis Kellis. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology*, 28(8):817–825, 2010.
- Jason Ernst and Manolis Kellis. Chromhmm: automating chromatin-state discovery and characterization. *Nature methods*, 9(3):215–216, 2012.
- Denis Franchimont, Tomoshige Kino, Jerome Galon, Gianfranco Umberto Meduri, and George Chrousos. Glucocorticoids and inflammation revisited: the state of the art. *Neuroimmunomodulation*, 10(5):247–260, 2002.
- Joanna Gariani, Olwen Westerland, Sarah Natas, Hema Verma, Gary Cook, and Vicky Goh. Comparison of whole body magnetic resonance imaging (wbmri) to whole body computed tomography (wbct) or 18f-fluorodeoxyglucose positron emission tomography/ct (18f-fdg pet/ct) in patients with myeloma: systematic review of diagnostic performance. *Critical reviews in oncology/hematology*, 124:66–72, 2018.
- Irene M Ghobrial, Ravi Vij, David Siegel, Ashraf Badros, Jonathan Kaufman, Noopur Raje, Andrzej Jakubowiak, Michael R Savona, Mihaela Obreja, and Jesus G Berdeja. A phase ib/ii study of oprozomib in patients with advanced multiple myeloma and waldenström macroglobulinemia. *Clinical Cancer Research*, 25(16):4907–4916, 2019.
- Robin E Goldman-Leikin, Helen R Salwen, CV Herst, Daina Variakojis, Mei Lu Bian, Michelle M Le Beau, Peter Selvanayagan, Robert Marder, Rhonda Anderson, Sigmund Weitzman, et al. Characterization of a novel myeloma cell line, mm. 1. *The Journal of laboratory and clinical medicine*, 113(3):335–345, 1989.
- MV Govindan, F Pothier, S Leclerc, R Palaniswami, and B Xie. Human glucocorticoid receptor gene promotor—homologous down regulation. *The Journal of steroid biochemistry and molecular biology*, 40(1-3):317–323, 1991.
- Lars Grøntved, Sam John, Songjoon Baek, Ying Liu, John R Buckley, Charles Vinson, Greti Aguilera, and Gordon L Hager. C/ebp maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. *The EMBO journal*, 32(11):1568–1583, 2013.
- RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with covid-19 preliminary report. *New England Journal of Medicine*, 2020.
- Sarah J Hainer, Ana Bošković, Kurtis N McCannell, Oliver J Rando, and Thomas G Fazzio. Profiling of pluripotency factors in single cells and early embryos. *Cell*, 177(5):1319–1329, 2019.
- Sandra B Hake, Benjamin A Garcia, Elizabeth M Duncan, Monika Kauer, Graham Dellaire, Jeffrey Shabanowitz, David P Bazett-Jones, C David Allis, and Donald F Hunt. Expression patterns and post-translational modifications associated with mammalian histone h3 variants. *Journal of Biological Chemistry*, 281(1):559–568, 2006.
- Sven Heinz, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C Lin, Peter Laslo, Jason X Cheng, Cornelis Murre, Harinder Singh, and Christopher K Glass. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Molecular cell*, 38(4):576–589, 2010.

REFERENCES

- Christoph J Heuck, Jackie Szymonifka, Emily Hansen, John D Shaughnessy, Saad Z Usmani, Frits Van Rhee, Elias Anaissie, Bijay Nair, Sarah Waheed, Yazan Alsayed, et al. Thalidomide in total therapy 2 overcomes inferior prognosis of myeloma with low expression of the glucocorticoid receptor gene nr3c1. *Clinical Cancer Research*, 18(19):5499–5506, 2012.
- Denes Hnisz, Brian J Abraham, Tong Ihn Lee, Ashley Lau, Violaine Saint-André, Alla A Sigova, Heather A Hoke, and Richard A Young. Super-enhancers in the control of cell identity and disease. *Cell*, 155(4): 934–947, 2013.
- Jackson A Hoffman, Kevin W Trotter, James M Ward, and Trevor K Archer. Brg1 governs glucocorticoid receptor interactions with chromatin and pioneer factors across the genome. *Elife*, 7:e35073, 2018.
- Jackson A Hoffman, Brian N Papas, Kevin W Trotter, and Trevor K Archer. Single-cell rna sequencing reveals a heterogeneous response to glucocorticoids in breast cancer cells. *Communications biology*, 3(1):1–11, 2020.
- William H Hudson, Christine Youn, and Eric A Ortlund. The structural basis of direct glucocorticoid-mediated transrepression. *Nature structural & molecular biology*, 20(1):53, 2013.
- Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R Lajoie, Job Dekker, and Leonid A Mirny. Iterative correction of hi-c data reveals hallmarks of chromosome organization. *Nature methods*, 9(10):999–1003, 2012.
- Duohui Jing, Yizhou Huang, Xiaoyun Liu, Keith CS Sia, Julia C Zhang, Xiaolu Tai, Meng Wang, Cara E Toscan, Hannah McCalmont, Kathryn Evans, et al. Lymphocyte-specific chromatin accessibility pre-determines glucocorticoid resistance in acute lymphoblastic leukemia. *Cancer cell*, 34(6):906–921, 2018.
- Sam John, Peter J Sabo, Thomas A Johnson, Myong-Hee Sung, Simon C Biddie, Stafford L Lightman, Ty C Voss, Sean R Davis, Paul S Meltzer, John A Stamatoyannopoulos, et al. Interaction of the glucocorticoid receptor with the chromatin landscape. *Molecular cell*, 29(5):611–624, 2008.
- Thomas A Johnson, Răzvan V Chereji, Diana A Stavreva, Stephanie A Morris, Gordon L Hager, and David J Clark. Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. *Nucleic acids research*, 46(1):203–214, 2018.
- Hatice S Kaya-Okur, Steven J Wu, Christine A Codomo, Erica S Pledger, Terri D Bryson, Jorja G Henikoff, Kami Ahmad, and Steven Henikoff. Cut&tag for efficient epigenomic profiling of small samples and single cells. *Nature communications*, 10(1):1–10, 2019.
- Jonathan J Keats, Marta Chesi, Jan B Egan, Victoria M Garbitt, Stephen E Palmer, Esteban Braggio, Scott Van Wier, Patrick R Blackburn, Angela S Baker, Angela Dispenzieri, et al. Clonal competition with alternating dominance in multiple myeloma. *Blood*, 120(5):1067–1076, 2012.
- W James Kent, Charles W Sugnet, Terrence S Furey, Krishna M Roskin, Tom H Pringle, Alan M Zahler, and David Haussler. The human genome browser at ucsc. *Genome research*, 12(6):996–1006, 2002.
- Charlotte Kervoëlen, Emmanuelle Ménoret, Patricia Gomez-Bougie, Régis Bataille, Catherine Godon, Séverine Marionneau-Lambot, Philippe Moreau, Catherine Pellat-Deceunynck, and Martine Amiot. Dexamethasone-induced cell death is restricted to specific molecular subgroups of multiple myeloma. *Oncotarget*, 6(29):26922, 2015.

REFERENCES

- Julian C Knight. Functional implications of genetic variation in non-coding dna for disease susceptibility and gene regulation. *Clinical Science*, 104(5):493–501, 2003.
- Paulina Kolasinska-Zwierz, Thomas Down, Isabel Latorre, Tao Liu, X Shirley Liu, and Julie Ahringer. Differential chromatin marking of introns and expressed exons by h3k36me3. *Nature genetics*, 41(3):376, 2009.
- Jan Krönke, Namrata D Udeshi, Anupama Narla, Peter Grauman, Slater N Hurst, Marie McConkey, Tanya Svinkina, Dirk Heckl, Eamon Comer, Xiaoyu Li, et al. Lenalidomide causes selective degradation of ikzf1 and ikzf3 in multiple myeloma cells. *Science*, 343(6168):301–305, 2014.
- Naoki Kubo, Haruhiko Ishii, Xiong Xiong, Simona Bianco, Franz Meitinger, Rong Hu, James D Hocker, Mattia Conte, David Gorkin, Miao Yu, et al. Promoter-proximal ctcf binding promotes distal enhancer-dependent gene activation. *Nature Structural & Molecular Biology*, 28(2):152–161, 2021.
- W Michael Kuehl and P Leif Bergsagel. Multiple myeloma: evolving genetic events and host interactions. *Nature Reviews Cancer*, 2(3):175–187, 2002.
- Ivan V Kulakovskiy, Ilya E Vorontsov, Ivan S Yevshin, Ruslan N Sharipov, Alla D Fedorova, Eugene I Rumynskiy, Yulia A Medvedeva, Arturo Magana-Mora, Vladimir B Bajic, Dmitry A Papatsenko, et al. Hocomoco: towards a complete collection of transcription factor binding models for human and mouse via large-scale chip-seq analysis. *Nucleic acids research*, 46(D1):D252–D259, 2018.
- Shaji K. Kumar, Vincent Rajkumar, Robert A. Kyle, Mark Van Duin, Pieter Sonneveld, María Victoria Mateos, Francesca Gay, and Kenneth C. Anderson. Multiple myeloma. *Nature Reviews Disease Primers*, 3, July 2017. ISSN 2056-676X. doi: 10.1038/nrdp.2017.46.
- Dan A Landau, Kendell Clement, Michael J Ziller, Patrick Boyle, Jean Fan, Hongcang Gu, Kristen Stevenson, Carrie Sougnez, Lili Wang, Shuqiang Li, et al. Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer cell*, 26(6):813–825, 2014.
- Stephen G Landt, Georgi K Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, Serafim Batzoglou, Bradley E Bernstein, Peter Bickel, James B Brown, Philip Cayting, et al. Chip-seq guidelines and practices of the encode and modencode consortia. *Genome research*, 22(9):1813–1831, 2012a.
- Stephen G Landt, Georgi K Marinov, Anshul Kundaje, Pouya Kheradpour, Florencia Pauli, Serafim Batzoglou, Bradley E Bernstein, Peter Bickel, James B Brown, Philip Cayting, et al. Chip-seq guidelines and practices of the encode and modencode consortia. *Genome research*, 22(9):1813–1831, 2012b.
- Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L Salzberg. Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome biology*, 10(3):1, 2009.
- Caleb A Lareau and Martin J Aryee. hichipper: a preprocessing pipeline for calling dna loops from hichip data. *Nature methods*, 15(3):155–156, 2018.
- Yeon Lee and Donald C Rio. Mechanisms and regulation of alternative pre-mrna splicing. *Annual review of biochemistry*, 84:291–323, 2015.

REFERENCES

- Daofeng Li, Silas Hsu, Deepak Purushotham, Renee L Sears, and Ting Wang. Washu epigenome browser update 2019. *Nucleic acids research*, 47(W1):W158–W165, 2019a.
- Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et al. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–2079, 2009.
- Yuying Li, Zhonghua Du, Xu Wang, Guanjuan Wang, and Wei Li. Association of il-6 promoter and receptor polymorphisms with multiple myeloma risk: A systematic review and meta-analysis. *Genetic testing and molecular biomarkers*, 20(10):587–596, 2016.
- Zhijian Li, Marcel H Schulz, Thomas Look, Matthias Begemann, Martin Zenke, and Ivan G Costa. Identification of transcription factor binding sites using atac-seq. *Genome biology*, 20(1):1–21, 2019b.
- Erez Lieberman-Aiden, Nynke L Van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *science*, 326(5950):289–293, 2009.
- Astrid Lièvre and Pierre Laurent-Puig. La voie de signalisation ras/mapk. *Cancéro digest*, 2010.
- Jens G Lohr, Petar Stojanov, Scott L Carter, Peter Cruz-Gordillo, Michael S Lawrence, Daniel Auclair, Carrie Sougnez, Birgit Knoechel, Joshua Gould, Gordon Saksena, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. *Cancer cell*, 25(1):91–101, 2014.
- Michael I Love, Matthew R Huska, Marcel Jurk, Robert Schöpflin, Stephan R Starick, Kevin Schwahn, Samantha B Cooper, Keith R Yamamoto, Morgane Thomas-Chollier, Martin Vingron, et al. Role of the chromatin landscape and sequence in determining cell type-specific genomic glucocorticoid receptor binding and gene regulation. *Nucleic acids research*, 45(4):1805–1819, 2017.
- Jakob Lovén, Heather A Hoke, Charles Y Lin, Ashley Lau, David A Orlando, Christopher R Vakoc, James E Bradner, Tong Ihn Lee, and Richard A Young. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, 153(2):320–334, 2013.
- Wenxiu Ma, William S Noble, and Timothy L Bailey. Motif-based analysis of large nucleotide data sets using meme-chip. *Nature protocols*, 9(6):1428–1450, 2014.
- Philip Machanick and Timothy L Bailey. Meme-chip: motif analysis of large dna datasets. *Bioinformatics*, 27(12):1696–1697, 2011.
- Florence Magrangeas, Hervé Avet-Loiseau, Wilfried Gouraud, Laurence Lodé, Olivier Decaux, Pascal Godmer, Laurent Garderet, Laurent Voillat, Thierry Facon, Anne Marie Stoppa, et al. Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia*, 27(2):473–481, 2013.
- Salomon Manier, Karma Z Salem, Jihye Park, Dan A Landau, Gad Getz, and Irene M Ghobrial. Genomic complexity of multiple myeloma and its clinical implications. *Nature reviews Clinical oncology*, 14(2):100, 2017.

REFERENCES

- María-Victoria Mateos and Jesús F San Miguel. How should we treat newly diagnosed multiple myeloma patients? *Hematology 2013, the American Society of Hematology Education Program Book*, 2013(1): 488–495, 2013.
- Ian C McDowell, Alejandro Barrera, Anthony M D’Ippolito, Christopher M Vockley, Linda K Hong, Sarah M Leichter, Luke C Bartelt, William H Majoros, Lingyun Song, Alexias Safi, et al. Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. *Genome research*, 28(9):1272–1284, 2018.
- Aaron McKenna, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernyt-sky, Kiran Garimella, David Altshuler, Stacey Gabriel, Mark Daly, et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation dna sequencing data. *Genome research*, 20(9): 1297–1303, 2010.
- Cory Y McLean, Dave Bristor, Michael Hiller, Shoa L Clarke, Bruce T Schaar, Craig B Lowe, Aaron M Wenger, and Gill Bejerano. Great improves functional interpretation of cis-regulatory regions. *Nature biotechnology*, 28(5):495–501, 2010.
- Jonathan S Mitchell, Ni Li, Niels Weinhold, Asta Försti, Mina Ali, Mark Van Duin, Gudmar Thorleifs-son, David C Johnson, Bowang Chen, Britt-Marie Halvarsson, et al. Genome-wide association study identifies multiple susceptibility loci for multiple myeloma. *Nature communications*, 7(1):1–9, 2016.
- Moshe Mittelman. The implications of anemia in multiple myeloma. *Clinical lymphoma*, 4:S23–S29, 2003.
- Hisham Mohammed, Christopher Taylor, Gordon D Brown, Evaggelia K Papachristou, Jason S Carroll, and Clive S D’santos. Rapid immunoprecipitation mass spectrometry of endogenous proteins (rime) for analysis of chromatin complexes. *Nature protocols*, 11(2):316–326, 2016.
- Philippe Moreau, Michel Attal, Cyrille Hulin, Bertrand Arnulf, Karim Belhadj, Lotfi Benboubker, Marie C Béné, Annemiek Broijl, Hélène Caillon, Denis Caillot, et al. Bortezomib, thalidomide, and dexametha-sone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (cassiopeia): a randomised, open-label, phase 3 study. *The Lancet*, 394(10192):29–38, 2019.
- Gareth J Morgan, Brian A Walker, and Faith E Davies. The genetic architecture of multiple myeloma. *Nature Reviews Cancer*, 12(5):335–348, 2012.
- Maxwell R Mumbach, Adam J Rubin, Ryan A Flynn, Chao Dai, Paul A Khavari, William J Greenleaf, and Howard Y Chang. Hichip: efficient and sensitive analysis of protein-directed genome architecture. *Nature methods*, 13(11):919–922, 2016.
- Maxwell R Mumbach, Ansuman T Satpathy, Evan A Boyle, Chao Dai, Benjamin G Gowen, Seung Woo Cho, Michelle L Nguyen, Adam J Rubin, Jeffrey M Granja, Katelynn R Kazane, et al. Enhancer connectome in primary human cells identifies target genes of disease-associated dna elements. *Nature genetics*, 49(11):1602, 2017.
- Ville Paakinaho, Harri Makkonen, Tiina Jaaskelainen, and Jorma J Palvimo. Glucocorticoid receptor activates poised fkbp51 locus through long-distance interactions. *Molecular endocrinology*, 24(3):511–525, 2010.

REFERENCES

- Ville Paakinaho, Thomas A Johnson, Diego M Presman, and Gordon L Hager. Glucocorticoid receptor quaternary structure drives chromatin occupancy and transcriptional outcome. *Genome research*, 29(8):1223–1234, 2019.
- Antonio Palumbo, Sara Bringhen, Heinz Ludwig, Meletios A Dimopoulos, Joan Bladé, Maria V Mateos, Laura Rosiñol, Mario Boccadoro, Michele Cavo, Henk Lokhorst, et al. Personalized therapy in multiple myeloma according to patient age and vulnerability: a report of the european myeloma network (emn). *Blood, The Journal of the American Society of Hematology*, 118(17):4519–4529, 2011.
- Ok Hyun Park, Joori Park, Mira Yu, Hyoung-Tae An, Jesang Ko, and Yoon Ki Kim. Identification and molecular characterization of cellular factors required for glucocorticoid receptor-mediated mrna decay. *Genes & development*, 30(18):2093–2105, 2016.
- Peter J Park. Chip-seq: advantages and challenges of a maturing technology. *Nature reviews genetics*, 10(10):669–680, 2009.
- Alessandro Pastore, Federico Gaiti, Sydney X Lu, Ryan M Brand, Scott Kulm, Ronan Chaligne, Hongcang Gu, Kevin Y Huang, Elena K Stamenova, Wendy Béguelin, et al. Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in cll. *Nature communications*, 10(1):1–11, 2019.
- Jelena Petrovic, Yeqiao Zhou, Maria Fasolino, Naomi Goldman, Gregory W Schwartz, Maxwell R Mumbach, Son C Nguyen, Kelly S Rome, Yogev Sela, Zachary Zapataro, et al. Oncogenic notch promotes long-range regulatory interactions within hyperconnected 3d cliques. *Molecular cell*, 73(6):1174–1190, 2019.
- Relja Popovic, Eva Martinez-Garcia, Eugenia G Giannopoulou, Quanwei Zhang, Qingyang Zhang, Teresa Ezponda, Mrinal Y Shah, Yupeng Zheng, Christine M Will, Eliza C Small, et al. Histone methyltransferase mmset/nsd2 alters ezh2 binding and reprograms the myeloma epigenome through global and focal changes in h3k36 and h3k27 methylation. *PLoS Genet*, 10(9):e1004566, 2014.
- Sebastian Pott and Jason D Lieb. What are super-enhancers? *Nature genetics*, 47(1):8–12, 2015.
- Aaron R Quinlan and Ira M Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–842, 2010.
- S Vincent Rajkumar and Robert A Kyle. Multiple myeloma: diagnosis and treatment. In *Mayo Clinic Proceedings*, volume 80, pages 1371–1382. Elsevier, 2005.
- Fidel Ramírez, Friederike Dündar, Sarah Diehl, Björn A Grüning, and Thomas Manke. deeptools: a flexible platform for exploring deep-sequencing data. *Nucleic acids research*, 42(W1):W187–W191, 2014.
- Dariusz Ratman, Wim Vanden Berghe, Lien Dejager, Claude Libert, Jan Tavernier, Ilse M Beck, and Karolien De Bosscher. How glucocorticoid receptors modulate the activity of other transcription factors: a scope beyond tethering. *Molecular and cellular endocrinology*, 380(1-2):41–54, 2013.
- Timothy E Reddy, Florencia Pauli, Rebekka O Sprouse, Norma F Neff, Kimberly M Newberry, Michael J Garabedian, and Richard M Myers. Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome research*, 19(12):2163–2171, 2009.

REFERENCES

- Gang Ren, Wenfei Jin, Kairong Cui, Joseph Rodriguez, Gangqing Hu, Zhiying Zhang, Daniel R Larson, and Keji Zhao. Ctf-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Molecular cell*, 67(6):1049–1058, 2017.
- James T Robinson, Helga Thorvaldsdóttir, Wendy Winckler, Mitchell Guttman, Eric S Lander, Gad Getz, and Jill P Mesirov. Integrative genomics viewer. *Nature biotechnology*, 29(1):24–26, 2011.
- Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2010.
- Sasikala P Sachithanandan, Deepak Arya, Shuling Zhang, Sudhir Krishna, and Beverly Mock. Hes1, a bhlh transcription factor regulate cell proliferation and promotes apoptosis in multiple myeloma cells. *Clinical Lymphoma, Myeloma and Leukemia*, 17(1):e48, 2017.
- Violaine Saint-André, Alexander J Federation, Charles Y Lin, Brian J Abraham, Jessica Reddy, Tong Ihn Lee, James E Bradner, and Richard A Young. Models of human core transcriptional regulatory circuitries. *Genome research*, 26(3):385–396, 2016.
- Rahul Satija, Jeffrey A Farrell, David Gennert, Alexander F Schier, and Aviv Regev. Spatial reconstruction of single-cell gene expression data. *Nature biotechnology*, 33(5):495–502, 2015.
- Ansuman T Satpathy, Jeffrey M Granja, Kathryn E Yost, Yanyan Qi, Francesca Meschi, Geoffrey P McDermott, Brett N Olsen, Maxwell R Mumbach, Sarah E Pierce, M Ryan Corces, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral t cell exhaustion. *Nature biotechnology*, 37(8):925–936, 2019.
- Valerie A Schneider, Tina Graves-Lindsay, Kerstin Howe, Nathan Bouk, Hsiu-Chuan Chen, Paul A Kitts, Terence D Murphy, Kim D Pruitt, Françoise Thibaud-Nissen, Derek Albracht, et al. Evaluation of grch38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome research*, 27(5):849–864, 2017.
- Nicolas Servant, Nelle Varoquaux, Bryan R Lajoie, Eric Viara, Chong-Jian Chen, Jean-Philippe Vert, Edith Heard, Job Dekker, and Emmanuel Barillot. Hic-pro: an optimized and flexible pipeline for hi-c data processing. *Genome biology*, 16(1):1–11, 2015.
- Arthur L Shaffer, NC Tolga Emre, Laurence Lamy, Vu N Ngo, George Wright, Wenming Xiao, John Powell, Sandeep Dave, Xin Yu, Hong Zhao, et al. Irf4 addiction in multiple myeloma. *Nature*, 454(7201):226, 2008.
- Arthur L Shaffer, NC Tolga Emre, Paul B Romesser, and Louis M Staudt. Irf4: immunity. malignancy! therapy? *Clinical Cancer Research*, 15(9):2954–2961, 2009.
- Ha Youn Shin, Michaela Willi, Kyung Hyun Yoo, Xianke Zeng, Chaochen Wang, Gil Metser, and Lothar Hennighausen. Hierarchy within the mammary stat5-driven wap super-enhancer. *Nature genetics*, 48(8):904–911, 2016.
- Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2016. *CA: a cancer journal for clinicians*, 66(1):7–30, 2016.

REFERENCES

- Rebecca L Siegel, Kimberly D Miller, and Ahmedin Jemal. Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1):7–34, 2019.
- Quinlan L Sievers, Georg Petzold, Richard D Bunker, Aline Renneville, Mikołaj Ślabicki, Brian J Liddicoat, Wassim Abdulrahman, Tarjei Mikkelsen, Benjamin L Ebert, and Nicolas H Thomä. Defining the human c2h2 zinc finger degrome targeted by thalidomide analogs through crbn. *Science*, 362(6414), 2018.
- Tim Stuart, Avi Srivastava, Caleb Lareau, and Rahul Satija. Multimodal single-cell chromatin analysis with signac. *bioRxiv*, 2020.
- Milan Surjit, Krishna Priya Ganti, Atish Mukherji, Tao Ye, Guoqiang Hua, Daniel Metzger, Mei Li, and Pierre Chambon. Widespread negative response elements mediate direct repression by agonist-liganded glucocorticoid receptor. *Cell*, 145(2):224–241, 2011.
- Thashim. Piq dna footprinting, 2017. URL <https://bitbucket.org/thashim/piq-single>.
- Henry F Thomas, Elena Kotova, Swathi Jayaram, Axel Pilz, Merrit Romeike, Andreas Lackner, Thomas Penz, Christoph Bock, Martin Leeb, Florian Halbritter, et al. Temporal dissection of an enhancer cluster reveals distinct temporal and functional contributions of individual elements. *Molecular Cell*, 81(5):969–982, 2021.
- Cole Trapnell, Lior Pachter, and Steven L Salzberg. Tophat: discovering splice junctions with rna-seq. *Bioinformatics*, 25(9):1105–1111, 2009.
- Cole Trapnell, Adam Roberts, Loyal Goff, Geo Pertea, Daehwan Kim, David R Kelley, Harold Pimentel, Steven L Salzberg, John L Rinn, and Lior Pachter. Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nature protocols*, 7(3):562, 2012.
- Vinayak V Viswanadham, Vinay S Mahajan, and Shiv Pillai. A bayesian approach for correcting tn5 transposition bias in atac-seq footprinting. *bioRxiv*, page 525808, 2019.
- Christopher M Vockley, Anthony M D’Ippolito, Ian C McDowell, William H Majoros, Alexias Safi, Lingyun Song, Gregory E Crawford, and Timothy E Reddy. Direct gr binding sites potentiate clusters of tf binding across the human genome. *Cell*, 166(5):1269–1281, 2016.
- Brian A Walker, Eileen M Boyle, Christopher P Wardell, Alex Murison, Dil B Begum, Nasrin B Dahir, Paula Z Proszek, David C Johnson, Martin F Kaiser, Lorenzo Melchor, et al. Mutational spectrum, copy number changes, and outcome: results of a sequencing study of patients with newly diagnosed myeloma. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 33(33):3911, 2015.
- Chong Wang, Luyao Zhang, Liangru Ke, Weiyue Ding, Sizun Jiang, Difei Li, Yohei Narita, Isabella Hou, Jun Liang, Shijun Li, et al. Primary effusion lymphoma enhancer connectome links super-enhancers to dependency factors. *Nature communications*, 11(1):1–13, 2020.
- Hongfang Wang, Chongzhi Zang, Len Taing, Kelly L Arnett, Yinling Joey Wong, Warren S Pear, Stephen C Blacklow, X Shirley Liu, and Jon C Aster. Notch1-rbpj complexes drive target gene expression through dynamic interactions with superenhancers. *Proceedings of the National Academy of Sciences*, 111(2):705–710, 2014.

REFERENCES

- Brian S White, Irena Lanc, Julie O’Neal, Harshath Gupta, Robert S Fulton, Heather Schmidt, Catrina Fronick, Edward A Belter, Mark Fiala, Justin King, et al. A multiple myeloma-specific capture sequencing platform discovers novel translocations and frequent, risk-associated point mutations in *igll5*. *Blood cancer journal*, 8(3):1–10, 2018.
- Warren A Whyte, David A Orlando, Denes Hnisz, Brian J Abraham, Charles Y Lin, Michael H Kagey, Peter B Rahl, Tong Ihn Lee, and Richard A Young. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*, 153(2):307–319, 2013.
- Zhigang Xie and Wee Joo Chng. Mmset: role and therapeutic opportunities in multiple myeloma. *BioMed research international*, 2014, 2014.
- Jing Zhang, Donghoon Lee, Vineet Dhiman, Peng Jiang, Jie Xu, Patrick McGillivray, Hongbo Yang, Jason Liu, William Meyerson, Declan Clarke, et al. An integrative encode resource for cancer genomics. *Nature communications*, 11(1):1–11, 2020.
- Tianyi Zhang, Sarah Cooper, and Neil Brockdorff. The interplay of histone modifications–writers that read. *EMBO reports*, 16(11):1467–1481, 2015.
- Yong Zhang, Tao Liu, Clifford A Meyer, Jérôme Eeckhoutte, David S Johnson, Bradley E Bernstein, Chad Nusbaum, Richard M Myers, Myles Brown, Wei Li, et al. Model-based analysis of chip-seq (macs). *Genome biology*, 9(9):R137, 2008.
- Bo Zhao, Luis A Barrera, Ina Ersing, Bradford Willox, Stefanie CS Schmidt, Hannah Greenfeld, Hufeng Zhou, Sarah B Mollo, Tommy T Shi, Kaoru Takasaki, et al. The NF- κ B genomic landscape in lymphoblastoid b cells. *Cell reports*, 8(5):1595–1606, 2014.
- Grace XY Zheng, Jessica M Terry, Phillip Belgrader, Paul Ryvkin, Zachary W Bent, Ryan Wilson, Solongo B Ziraldo, Tobias D Wheeler, Geoff P McDermott, Junjie Zhu, et al. Massively parallel digital transcriptional profiling of single cells. *Nature communications*, 8(1):1–12, 2017.
- Meizhen Zheng, Simon Zhongyuan Tian, Daniel Capurso, Minji Kim, Rahul Maurya, Byoungkoo Lee, Emaly Piecuch, Liang Gong, Jacqueline Jufen Zhu, Zhihui Li, et al. Multiplex chromatin interactions with single-molecule precision. *Nature*, 566(7745):558–562, 2019.

Annexes

Article publié-Genome Medecine.

RESEARCH

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

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

*Correspondence:
stephane.minvielle@univ-nantes.fr

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

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

†Co-senior

Abstract

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

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

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

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

Article soumis-Genome Biology.

Gaborit et al.

RESEARCH

Chromatin accessibility combined with enhancer clusters activation mediates heterogeneous response to dexamethasone in myeloma cells

Victor Gaborit^{1,2†}, Jonathan Cruard^{1†}, Catherine Guérin-Charbonnel^{1,3}, Jennifer Derrien¹, Jean-Baptiste Alberge¹, Elise Douillard^{1,4}, Nathalie Roi^{1,4}, Magali Devic^{1,4}, Loïc Campion^{1,3}, Frank Westermann⁵, Phillipe Moreau^{1,4}, Carl Herrmann⁶, Jérémie Bourdon², Florence Magrangeas^{1,4*^} and Stéphane Minvielle^{1,4*^}

*Correspondence:

florence.magrangeas@chu-nantes.fr;

stephane.minvielle@univ-nantes.fr

²Université de Nantes, LS2N,

CNRS, Nantes, France

³Université de Nantes, CNRS,

INSERM, CRCINA F-44000,

Nantes, France

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

[†] Co-author Co-senior

Abstract

Background: Glucocorticoids (GC) effects occur through binding to the GC receptor (GR) which, once translocated to the nucleus, binds to GC response elements (GREs) to activate or repress target genes. Among GCs, dexamethasone (Dex) is widely used in treatment of multiple myeloma (MM), mainly in combination regimens. However, despite a definite benefit, all patients do not respond in the same way. Moreover, while GC efficacy can be largely attributed to lymphocyte-specific apoptosis, its molecular basis remains elusive.

Results: To determine the functional role of GR binding in MM.1S myeloma cells, we generated bulk and single-cell multi-omic data and high-resolution contact maps of active enhancers and target genes. We show that a minority (6%) of GR binding sites are associated with enhancer activity gains and increased interaction loops. We find that enhancers contribute to regulate gene activity through combinatorial assembly of large stretches of enhancers and/or enhancer cliques. Furthermore, one enhancer, proximal to GR-responsive genes, is predominantly associated with increased chromatin accessibility and higher H3K27ac occupancy. Finally, we show that Dex exposure leads to co-accessibility changes between predominant enhancer and other regulatory regions of the interaction network. Notably, this epigenomic variability is associated with cell-to-cell transcriptional heterogeneity. As consequences, *BIM* critical for GR-induced apoptosis and *CXCR4* protective from chemotherapy-induced apoptosis are rather upregulated in different cells.

Conclusion: In summary, our work provides new insights into the molecular mechanisms involved in Dex escape.

Keywords: multiple myeloma; enhancers cluster; chromatin co-accessibility and gene expression cell-to-cell heterogeneity; drug escape

Titre : Mode d'action moléculaire de la dexaméthasone dans le myélome multiple

Mots clés : Myélome Multiple, Récepteurs au glucocorticoïde, hétérogénéité, chromatine, single-cell

Résumé : Le Myélome Multiple est une hémopathie maligne caractérisée par une prolifération incontrôlée des plasmocytes dans la moelle osseuse. Lors du développement de cette pathologie, la cellule fait l'acquisition de nombreuses anomalies génétiques. Ces anomalies ne sont pas retrouvées chez tous les patients ni même dans toutes les cellules d'une même tumeur. Cette maladie est donc caractérisée par une grande hétérogénéité inter et intra-patient mais également intra-clonale. Parmi les divers traitements utilisés, la Dexaméthasone (Dex), une hormone glucocorticoïde est largement utilisée en combinaison avec d'autres médicaments. La Dex exerce son effet en se fixant à son récepteur (GR) pour agir comme un facteur de transcription en se fixant sur les séquences

régulatrices du génome, principalement des enhancers, via des séquences de reconnaissance (GRE). Malgré un bénéfice certain de cette molécule en tant que traitement, tous les patients ne répondent pas de la même façon. De plus, si l'efficacité de la Dex est attribuée à une apoptose spécifique des lymphocytes B matures, les mécanismes moléculaires mis en jeu sont encore largement inconnus dans les plasmocytes.

Pour mieux comprendre ce mode d'action, des analyses pangénomiques intégrant des données d'accessibilité à la chromatine, de fixation de facteur de transcription, de modification d'histones ainsi que des données de conformation de la chromatine couplées à des analyses en cellule unique ont été réalisées.

Title : Molecular action of dexamethasone into Multiple Myeloma pathology

Keywords : Multiple Myeloma, Glucocorticoid Receptor, Heterogeneity, chromatin, single-cell

Abstract : Multiple Myeloma is a malignant hemopathy characterized by an aberrant proliferation of plasma cell in bone marrow. During development of this pathology, the cell acquires many genetic abnormalities. Not all the patients share those abnormalities and not even all the cells form one tumor. This pathology is characterized by a strong heterogeneity inter- and intra-patient but also intra-clonal. Among all treatments used, Dexamethasone (Dex), a synthetic glucocorticoid is widely used in combination with other drugs. Dex acts as a transcription factor by interacting with its receptor (GR) and binds to regulatory regions of genome, mostly enhancers, through recognizing DNA sequences (GRE). Despite a certain benefit of this molecule as treatment, all the patients not

respond the same way. In addition, if B lymphocytes apoptosis is attributed to Dex efficiency, the molecular mechanisms involved are largely unknown in plasma cells.

In order to better understand this mode of action, pan-genomics analysis integrating chromatin accessibility data, transcription factor binding, histones modifications and conformation data combined with single cell analysis have been performed.