

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ DE NANTES

ÉCOLE DOCTORALE N° 605

BIOLOGIE SANTÉ

SPÉCIALITÉ : GÉNÉTIQUE, GÉNOMIQUE, BIO-INFORMATIQUE

Par

Jean-Baptiste Alberge

Étude des mécanismes de l'initiation et de la progression du myélome multiple par transcriptomique en cellule-unique et cartographie de l'hydroxyméthylome

Thèse présentée et soutenue le 15 février 2021 à Nantes

Unité de recherche : Centre de Recherche en Cancérologie et Immunologie Nantes-Angers U1232

Rapporteurs avant soutenance :

Jérôme Eeckhoute
Salomon Manier

Directeur de recherche, CNRS, Université de Lille
Maître de conférences des universités – Praticien hospitalier, Université de Lille

Composition du Jury :

Présidente : **Françoise Kraeber-Bodéré**

Examinatrices : **Anaïs Bardet**

Françoise Kraeber-Bodéré

Dir. de thèse : **Stéphane Minvielle**

Aurélien Sérandour

Chargé de recherche, CNRS, Université de Strasbourg

Professeur des universités – Praticien hospitalier, Université de Nantes

Directeur de recherche, CNRS, Université de Nantes

Maître de conférences des universités, Centrale Nantes

Invités

Nizar Bahlis
Philippe Moreau

Professeur associé, Université de Calgary, Canada

Professeur des universités – Praticien hospitalier, Université de Nantes

Titre : Étude des mécanismes de l'initiation et de la progression du myélome multiple par transcriptomique en cellule-unique et cartographie de l'hydroxyméthylome

Mots clés : 5-hydroxyméthylcytosine ; Génomique ; Myélome multiple ; Hétérogénéité clonale

Résumé : Le myélome multiple (MM) est une tumeur hématologique causée par la prolifération incontrôlée de plasmocytes à longue durée de vie. L'hétérogénéité génomique du MM est caractérisée par une diversité importante des altérations génétiques somatiques, des modifications épigénétiques, et des programmes transcriptionnels dont l'articulation reste mal comprise et que nous explorons en trois axes.

Nous montrons que (1) l'ADN du MM est globalement appauvri en 5-hydroxyméthylcytosine (5hmC) par rapport à l'ADN de plasmocytes non-tumoraux. Le faible niveau global de 5hmC dans l'ADN corrèle avec l'agressivité de la maladie. La marque est présente localement aux régions actives et transcrites de la chromatine, et reflète les sous-groupes moléculaires du MM (*CCND1*, *MMSET*, etc.) ainsi que leurs programmes d'expression génique associés.

Puis, (2) nous décrivons les anomalies génomiques associées à la progression clinique et à l'acquisition de la résistance au venetoclax, une thérapie ciblée anti-BCL2. Les altérations génomiques sélectionnées sont concentrées sur la famille du BCL2 de façon clonale ou sous-clonale. Ces résultats sont étayés par des tests fonctionnels et des analyses d'expression génique.

Enfin, (3) nous proposons d'associer la transcriptomique des cellules de myélome aux paramètres de rendu d'imagerie médicale pour définir de nouveaux indicateurs pronostiques qui reflètent l'expansion et l'agressivité de la maladie.

Ce travail contribue à notre compréhension de l'oncogenèse et de la progression du MM tout en fournissant des biomarqueurs susceptibles de trouver leur place en clinique.

Title: A Study of the Initiation and Progression of Multiple Myeloma with Single-Cell Transcriptomics and Genome-wide Mapping of Hydroxymethylation Technologies

Keywords: 5-hydroxymethylcytosine; Genomics; Multiple Myeloma; Clonal heterogeneity

Abstract: Multiple Myeloma (MM) is a hematological malignancy caused by the uncontrolled proliferation of long-lived plasma cells. Genomic heterogeneity of MM is characterized by a significant diversity of somatic genetic alterations, epigenetic modifications, and of transcriptional programs whose integration remains poorly understood, and that we discuss in three parts.

First, compared to normal plasma cells, we found global DNA 5-hydroxymethylcytosine (5hmC) levels to be lower in MM. Higher 5hmC levels correlated with a less severe disease. Local 5hmC remained at active and transcribed regions of the chromatin where it mirrored the MM molecular subgroups (*CCND1*, *MMSET*, etc.) as well as their gene expression programs.

Second, we described unique genomic abnormalities that were associated with clinical progression and acquired resistance to venetoclax, an anti-BCL2 targeted therapy. The BCL2 family of genes displayed numerous alterations that were clonally or sub-clonally selected. Functional tests and gene expression profiling underpinned these results.

Third, transcriptional programs of MM cells were associated with biomedical imaging parameters to define new prognostic markers that mirrored the expansion and severity of the disease.

Together, this work contributes to our understanding of MM oncogenesis and progression while it also unravels novel clinical biomarkers.

REMERCIEMENTS

Mes premiers remerciements s'adressent à mes encadrants : au Dr Aurélien Sérandour, pour son soutien depuis Centrale Nantes et ce projet de bio-informatique ambitieux qui s'est transformé en travail de thèse, pour toute l'expertise et le savoir partagés, pour sa patience avec moi. Merci aux Dr Stéphane Minvielle et Florence Magrangeas de m'avoir tant appris sur la recherche contre le myélome, de m'avoir toujours encouragé et de m'avoir aiguillé sur un site nantais chargé d'histoire.

C'est tout naturellement que je tiens à remercier la Pr Françoise Kraeber-Bodéré et les Dr Jérôme Eeckhoutte, Anaïs Bardet et Salomon Manier d'avoir accepté de discuter mon travail et de composer le jury de cette thèse.

Merci à Élise, Magali, Nathalie, Catherine, Loïc, Jonathan, Victor, Jennifer, Anthony, Bertrand, Aymeric, aux autres, pour ces moments partagés au quotidien, pour les cafés et déjeuners amicaux.

Merci aux personnels hospitaliers, Philippe, Cyrille, Clément, Yannick, et les autres pour leur confiance, leur énergie, leur enthousiasme, leur intelligence.

Merci aussi à l'autre « équipe myélome » et aux Dr Martine Amiot et Catherine Pellat-Deceunynck, pour leurs éclairages et leurs conseils sans commune mesure.

Merci à Nizar, merci à Paola pour leur accueil chaleureux au Canada, d'avoir pris le temps de partager leurs idées, leur recherche, de m'avoir fait découvrir le meilleur de Calgary et de l'Alberta. Merci aussi à James et Ranjan pour ces moments partagés loin de l'Europe.

Du fond du cœur, un grand merci à Pierre, Émilie, Taciën, Kathleen et les autres pour ces trois années à refaire le monde et à partager des corbeilles de frites au Coup du Lapin – Merci à Sylvain sans qui nous n'aurions peut-être jamais découvert ces plaisirs-là !

Merci à Cassandra, à ma famille, à mes amis pour leur présence tout au long de cette aventure.

TABLE DES MATIÈRES

Remerciements	1
Table des matières	3
Liste des abréviations	6
Liste des figures	8
Liste des tableaux.....	10
Partie 1. Introduction	11
1. La genèse du plasmocyte à longue durée de vie	11
1.1. Développement du lymphocyte B et activation du BCR.....	11
1.2. La différenciation terminale B	17
1.3. Le plasmocyte à longue durée de vie	17
1.4. La différenciation cellulaire B et les lymphomes B	20
2. Le myélome multiple : une hémopathie maligne du plasmocyte	21
2.1. Stades précurseurs de la maladie et progression du myélome.....	21
2.2. Diagnostic et épidémiologie du myélome.....	23
2.3. Interventions pharmaceutiques et essais cliniques.....	24
2.4. Intérêt de la TEP-FDG dans le myélome	27
2.5. Évaluation de la maladie résiduelle mesurable	28
2.6. Vie et survie du Myélome : le rôle de la famille du BCL-2.....	29
3. Le génome du myélome et ses anomalies	33
3.1. Variants génétiques hérités	35
3.2. Mutations acquises	38
3.3. Translocations	42
3.4. Gains et pertes de copies génomiques.....	43
3.5. Chromothripsies, chromoplexies, et événements complexes.....	45
3.6. Chromosomes minuscules doubles	47
4. Enjeux pronostiques et hétérogénéité de la transcription	49
4.1. Classification fonctionnelle des patients et hétérogénéité des myélomes.....	49
4.2. Prédiction d'événements et signature de haut-risque	52
4.3. Le transcriptome au-delà des gènes codants.....	54
4.4. Du séquençage de l'ARN « bulk » au « single-cell RNA-seq »	56
4.5. scRNA-seq des plasmocytes tumoraux	57

5. Cartographier l'hydroxyméthylation de l'ADN	59
5.1. Méthylation de l'ADN génomique	59
5.2. Dynamique de la méthylation de l'ADN dans le myélome multiple.....	60
5.3. Aspects fonctionnels de la méthylation de l'ADN.....	61
5.4. Hydroxyméthylation et déméthylation de l'ADN génomique.....	62
5.5. Localisation et niveaux de la 5hmC.....	64
5.6. Dépendance aux protéines TET et cancer.....	64
5.7. Vers un rôle fonctionnel de la 5hmC ?.....	65
5.8. Cartographie de l'hydroxyméthylation dans le myélome multiple	66
5.9. Séquençage de la méthylation et de l'hydroxyméthylation de l'ADN.....	66
<i>Problématisation et objets de la thèse</i>	68
<i>Évolution et sélection clonales : modèles et enjeux dans le myélome multiple.....</i>	71
<i>Partie 2. Résultats.....</i>	77
1. Cartographie et quantification de l'hydroxyméthylation dans le myélome multiple.....	77
1.1. Contexte et objectifs de l'étude	77
1.2. Méthodes bioinformatiques	79
1.3. Article et résumé de l'article : DNA hydroxymethylation is associated with disease severity and persists at enhancers of oncogenic regions in multiple myeloma	82
1.4. Méthodes supplémentaires	95
1.5. Figures supplémentaires	99
2. Résistance acquise au venetoclax et anomalies génomiques sélectionnées de la famille du BCL2.....	107
2.1. Résumé en français	107
2.2. Abstract.....	109
2.3. Introduction.....	110
2.4. Materials and methods	112
2.5. Results.....	119
2.6. Discussion.....	131
2.7. Supplementary figures	134
3. Étude d'association entre profils transcriptomiques et biomarqueurs d'imagerie par TEP-FDG	139
3.1. Introduction.....	140
3.2. Méthodes	141
3.3. Résultats	142

3.4. Discussion et conclusion.....	150
Discussion et conclusions.....	152
1. Découverte de nouveaux biomarqueurs et étude de la chromatine	152
1.1. Design expérimental pour quantifier la 5hmC	152
1.2. Nouveaux axes de description de la 5hmC.....	153
1.3. Bénéfices liés à l'utilisation de la 5hmC en clinique	154
1.4. Hétérogénéité génomique du myélome et hydroxyméthylation.....	154
1.5. Le myélome est-il un cancer de l'hématopoïèse ?.....	155
2. Diversité de la sélection clonale dans la résistance acquise au traitement .	156
2.1. Complémentarité de la génomique et des tests fonctionnels de sensibilité aux BH3-mimétiques	157
2.2. Intérêt des mutations dans le myélome.....	158
3. Intérêts pronostiques et scientifiques de la combinaison de la génomique à d'autres variables	159
3.1. Génomique et hétérogénéité spatiale de la maladie	159
3.2. Combinaison des marqueurs génomiques et d'imagerie	159
3.3. Retour au clone tumoral.....	160
Bibliographie	161

LISTE DES ABRÉVIATIONS

0-9

5caC : 5-carboxylcytosine
5fC : 5-formylcytosine
5hmC : 5-hydroxyméthylcytosine
5mC : 5-méthylcytosine

A

ASH : American Society of Hematology
ATAC : Assay for Transposase-Accessible Chromatin

B

B2M : Microglobuline-Bêta 2
BCR : Récepteur des cellules B
BER : Réparation par excision de base
bp : Paire de bases d'ADN

C

CAR-T : T à récepteur antigénique chimérique
cdDNA : ADN libre circulant
CpG : Dinucléotide Cytosine-phosphate-Guanine
CGI : Îlot CpG
ChIP : Immuno-précipitation de chromatine
chr : Chromosome (chr1q : bras long du chromosome 1)
CLL : Leucémie lymphoïde chronique
CLP : Progéniteur commun des lymphocytes
CMP : Progéniteur commun des cellules myéloïdes
CNA : Anomalie du nombre de copie

CNV : Variation du nombre de copies
CSR : Commutation isotypique
CTC : Cellule tumorale circulante

D

del : Déletion d'un locus (del(17p) : Déletion du bras court du chromosome 17)
DFCI : Dana Farber Cancer Institute
dmin : Chromosome minuscule double
DNMT : Méthyl-transférase à ADN

E

EMD : Maladie extra-médullaire

F

FDG : ¹⁸F-fluorodésoxyglucose
FISH : Hybridation in situ en fluorescence
FL : Lésion focale ostéolytique

G

gain : Gain d'un locus (gain(1q) : gain du bras long du chromosome 1)
GC : Centre germinatif
GEP70 : Signature de haut-risque à 70 gènes
GLUT : Transporteur de glucose
GRCh38 ou **hg38** : Génome humain de référence v. 38
GSEA : Gene Set Enrichment Analysis
GWAS : Genome-wide Association Study

H

HK : Hexokinase
HOVON : Fondation d'hémo-oncologie pour les adultes aux Pays-Bas
HRD : Hyperdiploïde

I

IFM : Intergroupe Francophone du Myélome
Ig : Immunoglobuline (IgH : chaîne lourde des Ig, IgK : chaîne *kappa* des Ig, ...)
IL6 : Interleukine 6
IMiD : Immunomodulateur
IMWG : International Myeloma Working Group
IRF4 : Interferon Regulatory Factor 4
IRM : Imagerie par résonance magnétique
ISS : International Staging System

K

kbp : Kilo-paire de bases (1 000 bp)

L

LDH : Lactate deshydrogénase
lncRNA : ARN long non-codant

M

Mbp : Méga-paire de bases (1 million de bp)
MCL : Lymphome à cellules du manteau
MGUS : Gammopathie monoclonale de signification indéterminée
miRNA : micro-ARN
MM : Myélome multiple
MRD : Maladie résiduelle mesurable

N

NDMM : Myélome multiple nouvellement diagnostiqué

NGF : Next Generation Flow cytometry

NGS : Next Generation Sequencing

O

OLS : Organe lymphoïde secondaire

OR : Odds ratio, rapport des chances

OS : Overall survival, survie globale

oxBS : Oxidative bisulfite

P

PCR : Réaction en chaîne par polymérase

PFS : Progression-free survival, survie sans progression

PI : Inhibiteur du protéasome

PMD : Maladie par-médullaire

R

R-ISS : ISS révisé

S

SCL-seq : Selective Chemical Labeling and sequencing

scRNA-seq : Single-cell RNA sequencing

SHM : Mutation hypersomatique

SMM : Myélome multiple indolent (smoldering)

SNP : Single-Nucleotide Polymorphism

SNV : Single-Nucleotide Variant

SUVmax : Consommation maximale standardisée de glucose

T

t(IgH, ...) : Translocation du locus (de l'IgH, ...)

TAD : Topologically Associated Domain

TEP : Tomographie par émission de positons

TEP-FDG : Tomographie par émission de positons au ¹⁸F-fluorodésoxyglucose

TET : Ten-Eleven Translocation

TPM : Transcrits par million

U

UAMS : University of Arkansas for Medical Sciences

UMI : Identifiant moléculaire unique

W

WES : Séquençage complet de l'exome (du génome codant)

WGBS : Séquençage complet au bisulfite

WGS : Séquençage complet du génome

LISTE DES FIGURES

Figure 1 - Différentiation précoce du lymphocyte B et différenciation en plasmocyte	12
Figure 2 - Recombinaison VDJ sur le locus de la chaîne lourde des immunoglobulines.....	12
Figure 3 - Logigramme de l'expression des chaînes légères des immunoglobulines	14
Figure 4 - Recombinaison VJ des loci IgK et IgL.....	14
Figure 5 - Activation du lymphocyte B et différenciation en plasmocyte	16
Figure 6 - Interactions du plasmocyte avec le microenvironnement de la moelle osseuse.	19
Figure 7 - Translocations majeures de l'IgH des lymphomes B et du myélome	19
Figure 8 - Modèle de progression du myélome multiple	22
Figure 9 - Hétérogénéité intra-patient inter-lésionnelle explorée à la TEP-FDG.....	26
Figure 10 - Transport du glucose et premières étapes du pathway de la glycolyse	26
Figure 11 - Induction de l'apoptose par molécules BH3-mimétiques.....	30
Figure 12 - Le génome du myélome multiple au diagnostic.....	34
Figure 13 - Chromothripsie des chromosomes 2, 4, 3, 8, 14, 16, 17, 18, 19 et 20	45
Figure 14 - Aspects pronostiques de la classification moléculaire fonctionnelle du myélome ..	50
Figure 15 - Aspects pronostiques de la classification à haut risque GEP70	52
Figure 16 - Hétérogénéité transcriptomique inter- et intra-patient.....	57
Figure 17 - Marques chimiques aux CpG et déméthylation de l'ADN.....	63
Figure 18 - Co-localisation de la 5hmC avec les marques de chromatine active.....	63
Figure 19 - Niveau de 5hmC exprimé en valeur relative de la 5hmC cérébrale.....	63
Figure 20 - Techniques de quantification de la méthylation au bisulfite	67
Figure 21 - Modèle d'évolution et de sélection clonales du myélome multiple	73
Figure 22 - Protocole SCL-seq pour l'enrichissement de la 5hmC génomique.....	78
Figure 23 - Déplétion de l'hydroxyméthylation dans le myélome	85
Figure 24 - Organisation des domaines de 5hmC dans le myélome.....	87
Figure 25 - Localisation spécifique de la 5hmC dans les sous-groupes de myélome	88
Figure 26 - Dynamique de la 5hmC à la rechute.....	89
Figure 27 - Profil Agilent BioAnalyzer d'une librairie de 5hmC-seq (SCL-seq).....	99
Figure 28 - MS quantification of 5mC in genomic plasma cell DNA.....	100
Figure 29 - Survival course depending on DNA methylation (5mC).....	101
Figure 30 - 5hmC association with expression and criteria to group 5hmC peaks.....	102
Figure 31 - 5hmC allows the identification of a putative CCND2 enhancer.....	103
Figure 32 - Myeloma 5hmC-enriched domains associate with H3K27ac super-enhancers.....	104
Figure 33 - 5hmC signal levels at WNT5B-associated domain are increased at relapse	105
Figure 34 - Molecular profiling and clinical characteristics of venetoclax-treated patients	120
Figure 35 - Dysregulation of BCL2 family members and acquired resistance to venetoclax ...	122
Figure 36 - BCL2 Asp111Ala mutation induces venetoclax resistance.....	126
Figure 37 - Acquired MCL1 dependency or co-dependency induces venetoclax resistance....	127

Figure 38 - Pseudotime analysis shows emergence of MCL1 dependent clones at resistance.	128
Figure 39 - BCL2L1 rearrangements with loss of BCL2 dependency.....	130
Figure 40 - D111A BCL2 mutation in primary MM cells.....	134
Figure 41 - BCL2 c.332T>G (p.Asp111Ala) digital PCR analysis.....	135
Figure 42 - Effect of D111A BCL2 mutation on the binding of venetoclax to BCL2.	136
Figure 43 - <i>BCLxL</i> / <i>BCL2L1</i> rearrangements confirmed by ATAC-Seq data.....	137
Figure 44 - Hétérogénéité tumorale intra-lésionnelle de fixation en TEP-FDG.....	140
Figure 45 - Répartition des variables de RNA-seq et d'imagerie au TEP-FDG	143
Figure 46 - Maladie para-médullaire et signature GEP70 définissent un groupe à très haut risque de progression.....	145
Figure 47 - Les TEP négatives sont associées à une dérégulation des programmes transcriptionnels incluant le métabolisme du glucose	147
Figure 48 - Analyse d'expression différentielle entre TEP anormales et normales.....	148
Figure 49 - Analyse d'ontologie des gènes différentiellement exprimés entre TEP normales et anormales.....	149

LISTE DES TABLEAUX

Tableau 1 - Critères de diagnostic du myélome multiple d'après l'IMWG 2014.....	22
Tableau 2 - Stratification de risque par ISS et R-ISS	22
Tableau 3 - Variables pronostiques en imagerie par TEP-FDG	26
Tableau 4 - Variants rares constitutifs associés au diagnostic de myélome multiple.....	36
Tableau 5 - Polymorphismes nucléotidiques communs constitutifs associés au diagnostic.....	36
Tableau 6 - Gènes affectés par des mutations somatiques récurrentes dans le myélome.....	39
Tableau 7 - Matrice de confusion des classifications de l'UAMS et de HOVON	50
Tableau 8 - Analyse de survie sans progression (PFS) des variables de rendu de TEP-FDG...	143

Partie 1. INTRODUCTION

1. LA GENÈSE DU PLASMOCYTE À LONGUE DURÉE DE VIE

Les acteurs cellulaires de l'immunité humorale sont responsables de nombreux cancers : le myélome multiple, auquel nous nous intéresserons particulièrement ici, est une hémopathie maligne dont la cellule d'origine est un plasmocyte à longue durée de vie et producteur d'anticorps. La connaissance de l'immunité humorale physiologique, capitale dans le monde moderne, nous aide à appréhender ce qui rend l'oncogénèse plasmocytaire si particulière. Ainsi la différenciation et l'activation du lymphocyte B comme l'oncogénèse B ont fait l'objet de nombreuses excellentes revues (Küppers, 2005; Kurosaki et al., 2015; Nutt et al., 2015; Ise and Kurosaki, 2019; Akkaya et al., 2020). Nous en reprenons ici les éléments qui nous semblent pertinents pour introduire l'étude de la génomique du myélome multiple.

L'acquisition de la mémoire B est communément divisée en trois phases : (1) la différenciation précoce du B dans la moelle osseuse, par ailleurs siège de l'hématopoïèse, (2) l'activation des lymphocytes B des centres germinatifs (GC) après stimulation antigénique via le récepteur des cellules B (BCR), et (3) la différenciation en lymphocyte B mémoire et plasmocytes à longue durée de vie. Ces phases de différenciation et d'activation sont le siège des recombinaisons VDJ, de l'hypermutation somatique (SHM) et de la commutation isotypique (CSR) : elles sont autant d'étapes susceptibles de générer, de manière accidentelle, une recombinaison anormale de l'ADN et d'être le siège d'une prolifération tumorale.

1.1. DÉVELOPPEMENT DU LYMPHOCYTE B ET ACTIVATION DU BCR

À partir de cellules souches hématopoïétiques sont produits des précurseurs myéloïdes (CMP) et lymphoïdes (CLP). Les CLP peuvent engendrer des lymphocytes Natural Killer (NK), T, et B. L'expression des facteurs de transcription E2A (*TCF3*)¹, EBF (*EBF1*) et PAX5 détermine la différenciation des CLP dans la lignée B précoce (pré-pro-B, Figure 1) (O'Riordan and Grosschedl, 1999).

¹ On note le nom d'une protéine humaine en majuscules, le symbole d'un gène humain en majuscules italiques

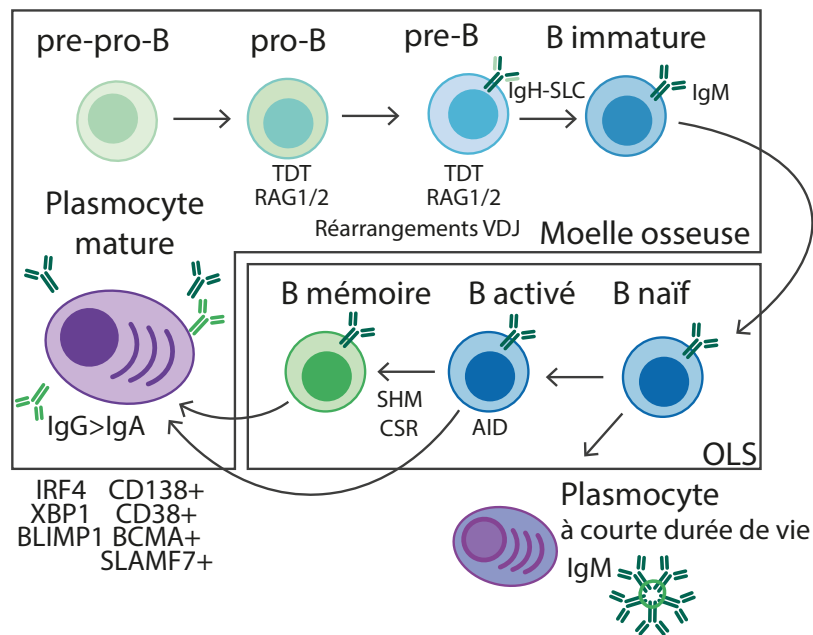


Figure 1 – Différentiation précoce du lymphocyte B et différenciation en plasmocyte. IgH: chaîne lourde des immunoglobulines ; SLC : chaînes légères de substitution ; TdT : nucléotidylexotransférase d'ADN ; RAG : gènes activant la recombinaison ; OLS : organe lymphoïde secondaire ; AID : Activation-induced cytidine deaminase ; SHM : Hypermutation somatique ; CSR : Commutation isotypique.

IgH chr14 : 105 586 437 – 106 879 844 (GRCh38)

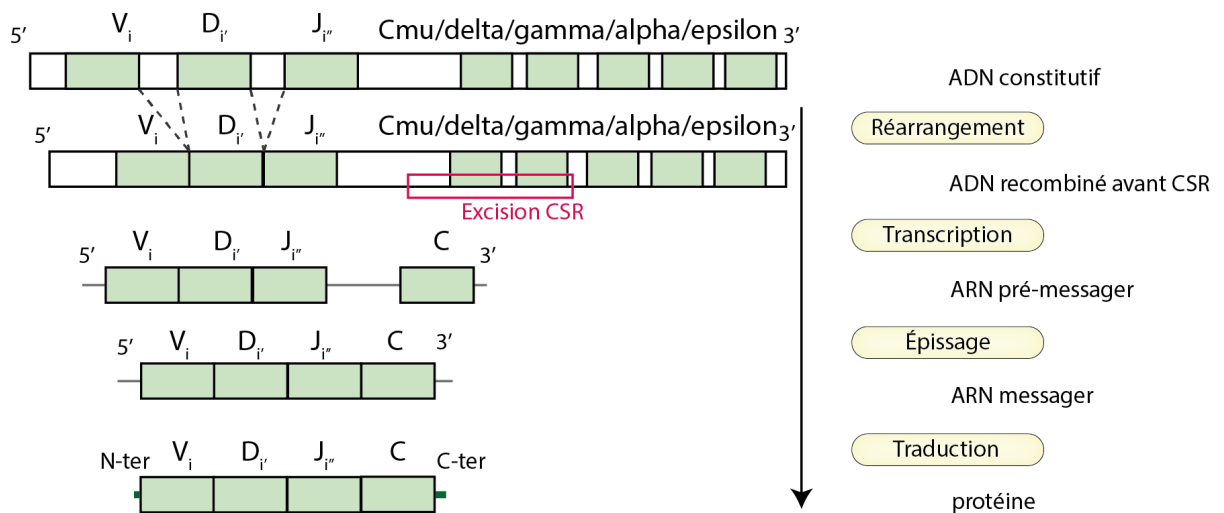


Figure 2 - Recombinaison VDJ sur le locus de la chaîne lourde des immunoglobulines. N-ter (resp. C-ter) : extrémité N-terminale (resp N-terminale). D'après les cours de l'IMGT (<http://atlasgeneticsoncology.org/Educ/PolyIgFr.html>).

En vue d'assembler le BCR, les précurseurs B subissent un premier réarrangement des fragments D-J du locus de la chaîne lourde des immunoglobulines (IgH) sur le chromosome 14 (chr14, Figure 2). Ensuite, le fragment V est recombiné avec le fragment D-J pour former le VDJ. Ce réarrangement est permis par l'activation d'enzymes recombinantes RAG1/RAG2 et TDT (*DNTT*). S'il n'est pas fonctionnel, le second allèle est réarrangé selon le même processus, sinon, le second allèle subit l'exclusion allélique. L'apposition du fragment constant au fragment VDJ se fait par épissage alternatif lors de la transcription.

Les lymphocytes pré-B expriment un pré-BCR composé de l'IgH réarrangé et de chaînes légères de substitution (SLC). Le pré-BCR est associé aux CD79A/B et sa survie est déterminée par sa capacité de signalisation.

Le réarrangement VJ de la chaîne légère *kappa* est initié sur le chromosome 2, et si aucun des deux allèles n'est fonctionnel, c'est le locus VJ *lambda* qui est arrangé sur le chromosome 22 suivant les mêmes règles (Figure 3). Le réarrangement *kappa* non-fonctionnel entraîne une délétion d'enhancer (une région régulatrice amplificatrice qui stimule la transcription d'un gène proximal) en aval d'IgK – cet enhancer est perdu dans environ 50% des LB *kappa* et dans les LB *lambda* (Figure 4). À l'inverse, le réarrangement *lambda* n'entraîne pas de délétion d'enhancer sur le locus IgL (Siminovitch et al., 1985). En pratique, deux tiers des lymphocytes B expriment la chaîne *kappa*.

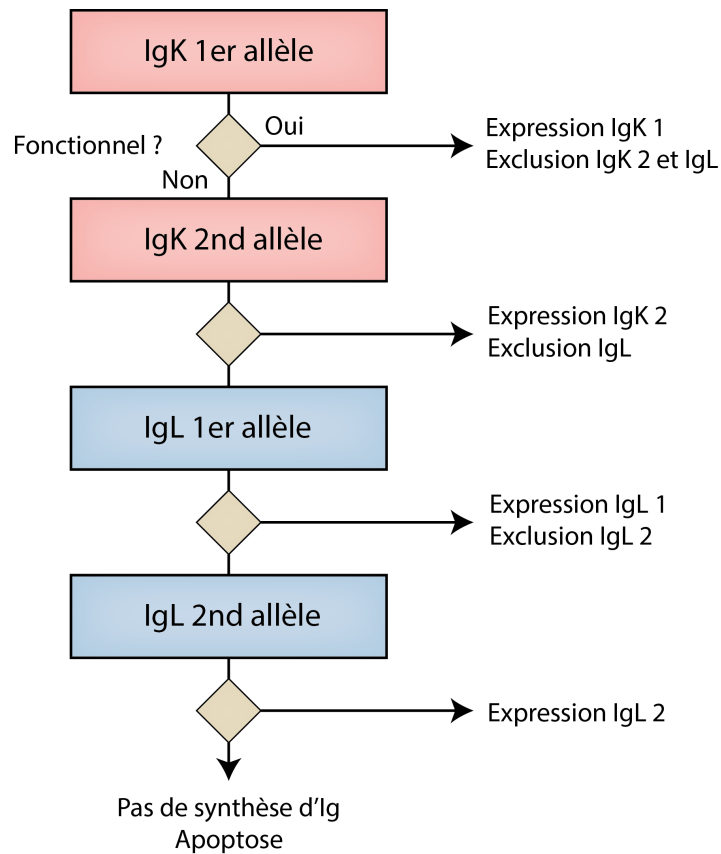


Figure 3 - Logigramme de l'expression des chaînes légères des immunoglobulines. IgK : chaîne légère *kappa* des immunoglobulines. IgL : chaîne légère *lambda* des immunoglobulines. D'après les cours de l'IMGT (<http://atlasgeneticsoncology.org/Educ/PolyIgFr.html>)

IgK chr2 : 88857361 – 90235368 (GRCh38)

IgL chr22 : 22 026 076 – 22 922 913 (GRCh38)

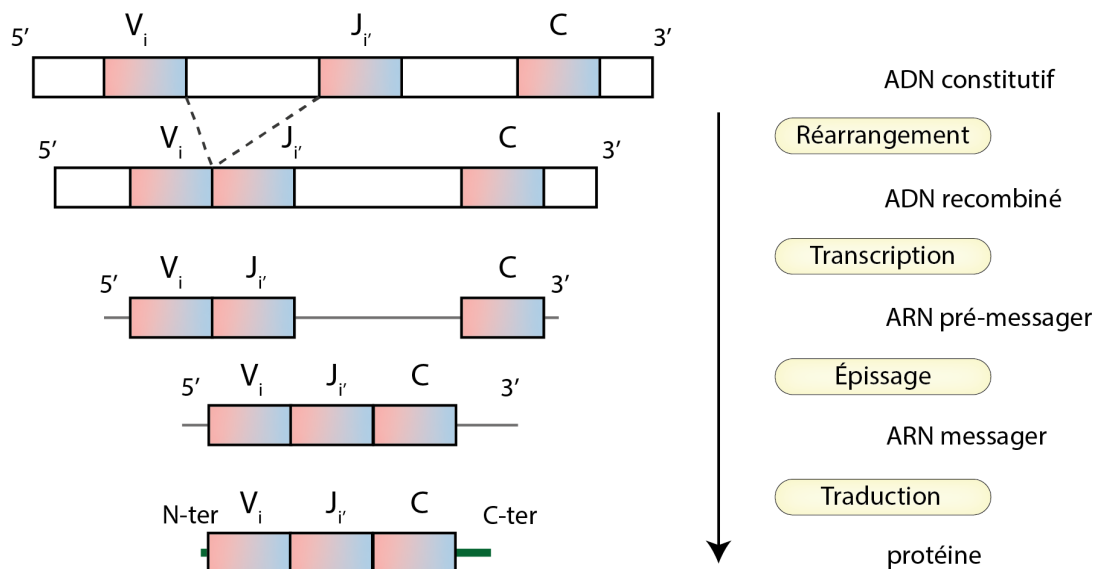


Figure 4 - Recombinaison VJ des loci IgK et IgL. N-ter (resp. C-ter) : extrémité N-terminale (resp N-terminale). D'après les cours de l'IMGT (<http://atlasgeneticsoncology.org/Educ/PolyIgFr.html>).

Une fois le BCR réarrangé, les lymphocytes B sont dits immatures et expriment un isotype IgM (Figure 1, Figure 5). Ils migrent vers les organes lymphoïdes secondaires (OLS) via la circulation sanguine. Les lymphocytes B naïfs activés par la première rencontre antigénique prolifèrent et suivent trois voies de différenciation possibles (Figure 5) :

- En plasmocytes à courte durée de vie, qui produisent des anticorps spécifiques de l'antigène présenté, d'isotype préférentiel IgM, mais parfois commuté en IgG ou IgA, et qui s'accumulent dans les OLS.
- En lymphocytes B des centres germinatifs, acteurs majeurs de la troisième phase de réponse immunitaire B, précurseurs des plasmocytes à longue durée de vie ou des B mémoire.
- En lymphocytes B indépendants du GC, qui peuvent à leur tour produire des plasmocytes à courte durée de vie, à longue durée de vie, ou des B mémoire indépendants des GC et de l'interaction avec les lymphocytes T.

Les paramètres qui déterminent la directionnalité de la différenciation des lymphocytes B activés en un des trois types cellulaires décrits ci-dessus sont encore mal connus. Il semble qu'un niveau important d'expression d'IRF4 (Interferon regulatory factor 4) joue en faveur d'une différenciation précoce en plasmocyte, et qu'au contraire, un niveau faible d'IRF4 entraîne une différenciation en lymphocyte B du centre germinatif. Le niveau d'IRF4 semble déterminé par l'affinité de l'interaction antigène-BCR. En somme, une interaction forte induit des niveaux élevés d'IRF4, plus susceptible d'entraîner une différenciation en plasmocyte, alors qu'une affinité faible de l'antigène pour le BCR entraînerait des niveaux faibles d'IRF4 et une inclinaison pour la différenciation en B des centres germinatifs (Ochiai et al., 2013; Taylor et al., 2015).

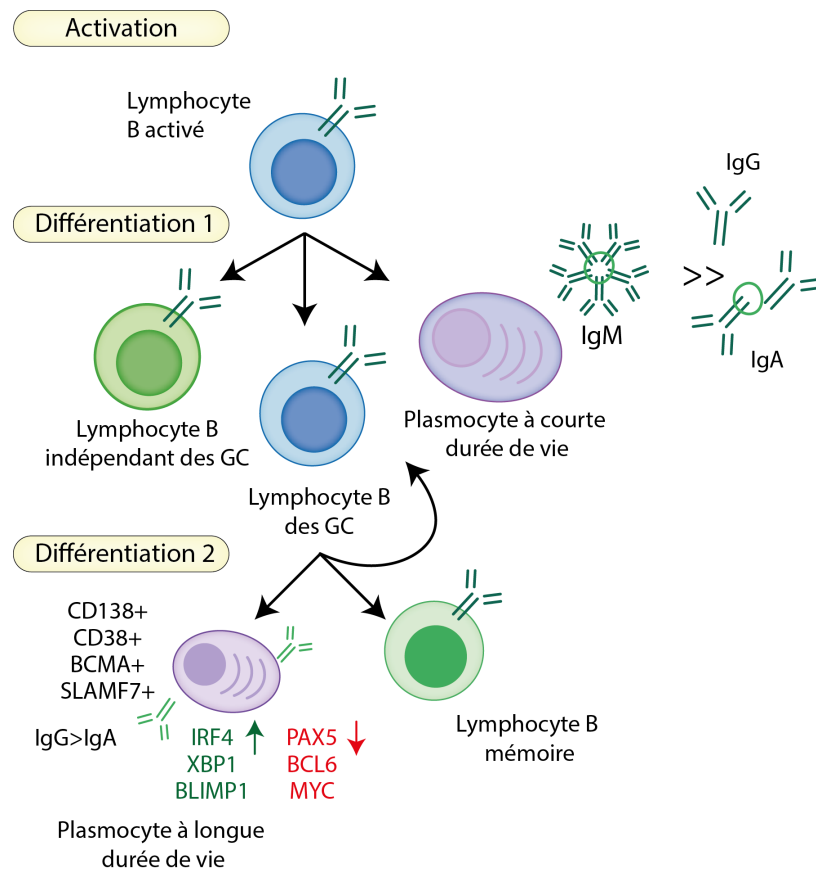


Figure 5 - Activation du lymphocyte B et différenciation en plasmocyte à longue durée de vie. BCL6 : B-cell lymphoma 6 ; BCMA : B-cell maturation antigen ; BLIMP1 : PRDM1 / PR domain zinc finger protein 1 ; CD138 : syndecan-1 / *SDC1* ; GC : centre germinatif ; IRF4 : Interferon regulatory factor 4 ; PAX5 : Paired box protein 5 ; XBP1 : X-box binding protein 1

1.2. LA DIFFÉRENTIATION TERMINALE B

Dans la troisième phase de la différenciation, les lymphocytes B du GC prolifèrent, subissent la SHM qui augmente l'affinité antigénique, puis empruntent une des voies de différenciation suivantes (Figure 5) :

- En plasmocyte à longue durée de vie présentant CD38, CD138, SLAMF7, BCMA, sécréteur d'anticorps, qui niche dans la moelle osseuse.
- En lymphocyte B mémoire à longue durée de vie, qui niche dans les OLS, caractérisé principalement par l'expression additionnelle du CD27.
- Un retour dans la zone de prolifération du GC, aussi appelée zone sombre, pour reprendre la maturation de l'affinité antigénique.

De manière intéressante, alors que la CSR est souvent associée à la SHM, notamment parce que l'activité enzymatique de la cytidine désaminase AID (*AICDA*) est nécessaire aux deux (Muramatsu et al., 2000), Roco et collègues ont montré récemment que le déclenchement de la CSR se fait avant l'entrée dans le GC et avant la SHM (Roco et al., 2019).

1.3. LE PLASMOCYTE À LONGUE DURÉE DE VIE

Les plasmocytes à longue durée de vie et producteurs d'anticorps sont généralement produits dans les GC en réponse à une stimulation antigénique dépendant de la réponse T, et présentent trois caractéristiques : ils sont reprogrammés par des facteurs de transcription, ils sécrètent des anticorps, et ils nichent dans la moelle osseuse où ils peuvent survivre de nombreuses années (Boise et al., 2014; Nutt et al., 2015; Ise and Kurosaki, 2019).

Les plasmocytes ont un profil transcriptomique distinct de la lignée B, et on retient communément que la différenciation est contrôlée par la baisse d'activité des facteurs de l'identité B PAX5 et BCL6 couplée à une augmentation des trois facteurs de transcription : IRF4, BLIMP1 (*PRDM1*), et XBP1 (Nutt et al., 2015). IRF4 est un facteur requis pour la différenciation et l'activation B, et à forte dose, il réprime BCL6 et active PRDM1. L'expression de PRDM1 nécessite IRF4 et réprime notamment l'expression de BCL6, PAX5, MYC et AID. XBP1 est activé avec la perte de PAX5 et est un acteur de la réponse aux protéines mal repliées (UPR), réponse induite par le stress du réticulum endoplasmique et associée à la forte production de protéines à sécréter (Nutt et al., 2015).

À propos, MYC, une cible de l'inhibiteur transcriptionnel BLIMP1, est probablement clé dans la différenciation terminale, car les B très affins expriment fortement MYC, ils sont plus susceptibles de se différencier en plasmocytes, et certains expriment aussi IRF4 (Ise and Kurosaki, 2019). Des travaux futurs pourraient élucider les mécanismes de différenciation terminale ; le rôle et la nécessité de la stimulation CD40 font notamment débat (Ise and Kurosaki, 2019).

La différenciation complète vers le plasmocyte est mal connue et fait l'objet de recherche active : il demeure que les plasmocytes humains normaux expriment fortement BLIMP1, n'expriment pas MYC, ne prolifèrent pas, et résident dans la moelle osseuse.

Le recrutement des plasmocytes dans la moelle osseuse est médié par l'axe CXCL12-CXCR4 (Hargreaves et al., 2001). CXCR4 est exprimé par les plasmocytes et CXCL12 est issu de cellules qui composent la niche de la moelle osseuse : les cellules stromales VCAM1+CXCL12+ (Figure 6). La niche est aussi composée de cellules assurant la survie des plasmocytes par la sécrétion de facteurs tels que l'interleukine-6 (IL6) et APRIL (A proliferation-inducing ligand, *TNFSF13*). Il n'y a pas de type cellulaire unique nécessaire à la sécrétion d'APRIL même si ce rôle a été attribué aux éosinophiles. Il est probable qu'un certain nombre agit de concert notamment dans la lignée hématopoïétique et que les niveaux d'IL6 et d'APRIL sont multifactoriels (Khodadadi et al., 2019). Enfin, l'axe APRIL-BCMA agit pour la survie des plasmocytes via MCL1, protéine anti-apoptotique de la famille du BCL2. Les plasmocytes ne survivent pas *in vivo* sans MCL1, et BCMA est requis pour une forte expression du MCL1 (Peperzak et al., 2013).

C'est également à ce stade terminal de différenciation que le CD138 (*SDC1*, Syndecan-1) est exprimé à la surface des plasmocytes. Le CD138 est utilisé comme marqueur ubiquitaire des cellules de myélome multiple comme des plasmocytes normaux. Le CD138 joue un rôle majeur dans la survie des plasmocytes en augmentant les signalisations d'IL6 et APRIL, qui confèrent un avantage à la survie des plasmocytes par rapport aux cellules n'expriment pas ou peu le CD138 (McCarron et al., 2017). Au-delà du CD138, on caractérise couramment les plasmocytes dans la moelle par les marqueurs de surface CD38, BCMA (*TNFRSF17*, CD269), et CS1 (*SLAMF7*, CD319).

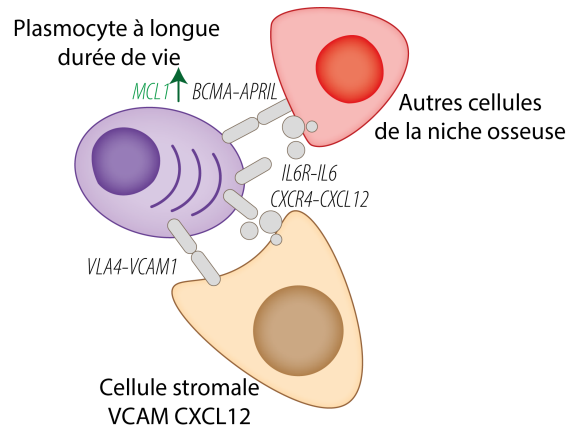


Figure 6 - Interactions du plasmocyte avec le microenvironnement de la moelle osseuse.

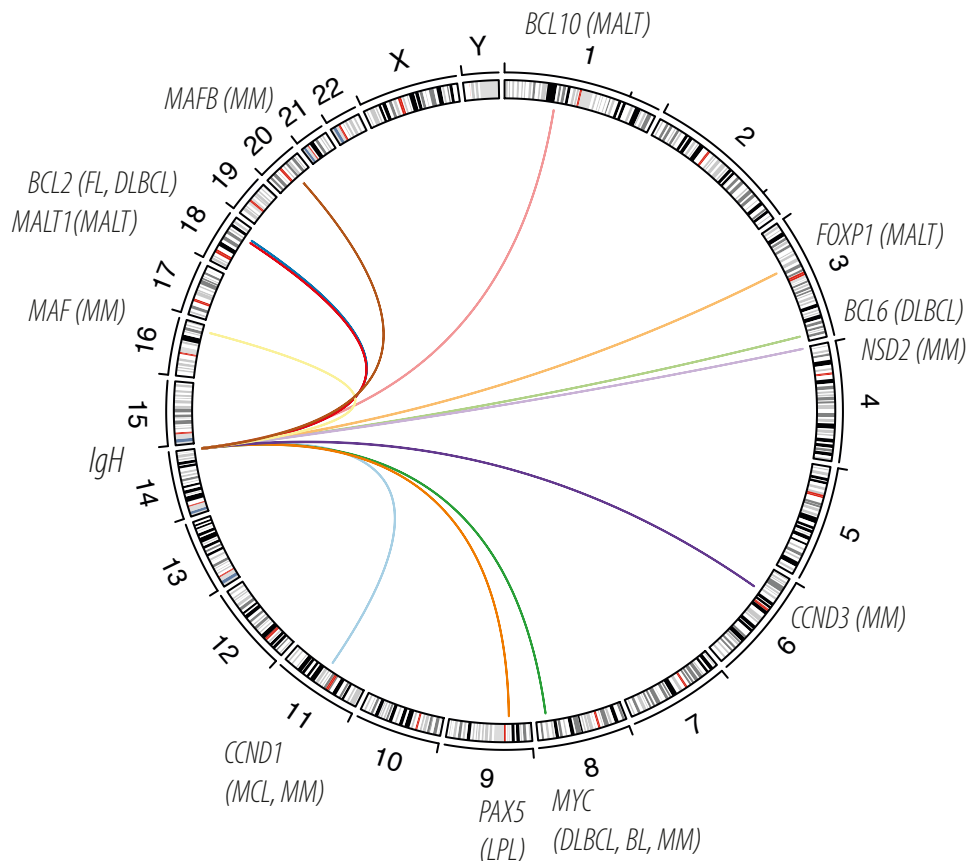


Figure 7 - Translocations majeures de l'IgH des lymphomes B et du myélome. BL : lymphome de Burkitt ; DLBCL : lymphomes diffus à grandes cellules B ; FL : lymphome folliculaire ; MALT : lymphome du tissu lymphoïde associé aux muqueuses ; LPL : lymphome plasmacytique ; MCL : lymphome à cellule du manteau ; MM : myélome multiple ; IgH : locus de la chaîne lourde des immunoglobulines. Représentation en « circos plot » réalisée avec R/circlize d'après les données de Küppers et al., 2005.

1.4. LA DIFFÉRENTIATION CELLULAIRE B ET LES LYMPHOMES B

Une caractéristique récurrente des lymphomes B est la présentation de translocations résumées en Figure 7 d'un locus des Ig (principalement IgH) avec un oncogène. Les translocations ne sont pas nécessairement équilibrées (c'est-à-dire qu'un bras de chromosome peut être perdu dans l'opération de réarrangement tout en donnant un clone tumoral viable), mais elles sont souvent clonales (c'est-à-dire présente dans toutes les cellules tumorales). On considère que la juxtaposition d'un enhancer des IgH à un oncogène est un marqueur de tumorigénicité caractéristique des lymphomes B et du myélome.

95% des lymphomes à cellules du manteau (MCL) présentent une translocation t(11;14) IgH-CCND1 qui recombine la cycline D1 au locus régulateur de l'IgH. La moitié des patients présentent également une délétion du locus chr13q14 et de la protéine du rétinoblastome (Rb). Aussi, 90% des lymphomes folliculaires présentent une translocation t(14;18) IgH-BCL2. Dans le lymphome de Burkitt, presque toujours lié à une infection au virus d'Epstein-Barr (EBV), MYC est systématiquement transloqué avec l'IgH en t(8;14) (Küppers, 2005). De manière intéressante, l'équipe de Vassetzky montre par FISH 3D dans une infection au virus de l'immunodéficience humaine (HIV) *in vitro* que la protéine virale Tat induit des cassures d'ADN RAG1-dépendantes, puis une co-localisation nucléaire des deux loci IgH et MYC, favorisant statistiquement leur recombinaison aberrante (Germini et al., 2017).

D'autre part, puisqu'il a été question de survie des progéniteurs B dans les paragraphes précédents, soulignons que la signalisation du BCR semble également nécessaire à la survie de la majorité des lymphomes à BCR (Küppers, 2005). La survie via le BCR est médiée par les protéines de la famille du BCL2 (Enders et al., 2003). La protéine anti-apoptotique BCL2 elle-même, qui est fortement exprimée dans les B matures, offre une vulnérabilité : BCL2 peut être ciblée par des molécules thérapeutiques, nous y reviendrons également.

2. LE MYÉLOME MULTIPLE : UNE HÉMOPATHIE MALIGNE DU PLASMOCYTE

2.1. STADES PRÉCURSEURS DE LA MALADIE ET PROGRESSION DU MYÉLOME

Le myélome multiple est une hémopathie maligne caractérisée par la prolifération anormale d'un clone plasmocytaire producteur d'anticorps, le plus souvent dans la moelle osseuse mais régulièrement présent aussi dans la circulation sanguine et d'autres tissus et organes extra-médullaires. Malgré une forte hétérogénéité clinique d'un patient à l'autre qui rend difficile la caractérisation unifiée du myélome multiple, les myélomes symptomatiques partagent des stades précurseurs : la gammopathie monoclonale de signification indéterminée (MGUS) et le myélome multiple indolent (SMM) (Rajkumar et al., 2014) (Figure 8).

Le MGUS concerne 3% des adultes de plus de 50 ans (Kyle et al., 2006). Certains patients présentent un MGUS stable et non-traité, durant des décennies, alors que d'autres progressent vers le stade symptomatique en quelques mois, si bien qu'on montre que la fréquence moyenne de progression clinique est d'1 à 2% par an pour un MGUS, et de 10% pour un SMM (Rajkumar et al., 2014; Kyle et al., 2018). Aujourd'hui, les SMM font l'objet d'une surveillance sans intervention thérapeutique (Rajkumar et al., 2015). La situation est susceptible d'évoluer dans les années à venir avec la reclassification des SMM (Rajkumar et al., 2014), les essais cliniques interventionnels chez les patients à risque (Mateos et al., 2013; Rajkumar et al., 2018; Landgren et al., 2020), et l'amélioration des modèles pronostiques de progression clinique (Lakshman et al., 2018; Bustoros et al., 2020).

Finalement, environ 20% des MGUS progressent vers un stade clinique incluant le myélome multiple mais pas seulement : on peut évoquer l'amylose de type AL et la macroglobulinémie de Waldenström (MW) (Kumar et al., 2017a), même si historiquement, notre équipe de recherche n'étudie pas ces cas cliniques. L'amylose de type AL est caractérisée par une accumulation d'IgK ou IgL clonale mal conformée sans prolifération oncogénique du clone plasmocytaire : ce n'est donc pas un cancer (Desport et al., 2012). La MW quant à elle est caractérisée par une prolifération clonale IgM très majoritairement associée à la mutation gain-de-fonction MYD88_L265P, un profil atypique par rapport aux myélomes multiples. MYD88_L265P est souvent complétée d'une mutation non-sens ou de décalage du cadre de lecture de CXCR4, qui tronque la queue cytosolique du récepteur de CXCL12 (Hunter et al., 2014; Bustoros et al., 2019; Yu et al., 2020).

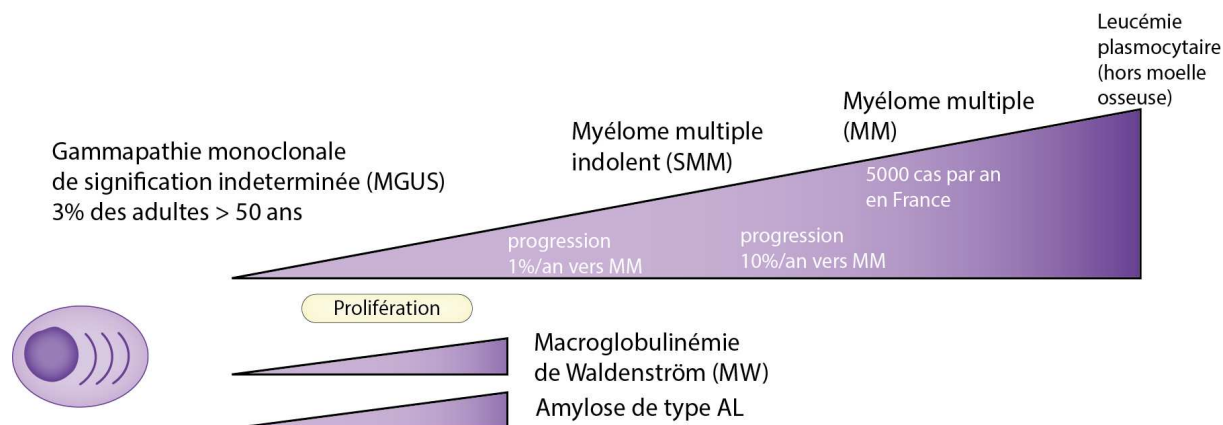


Figure 8 - Modèle de progression du myélome multiple depuis les stades indolents de la maladie. Le gradient de couleur représente l'avancement de la maladie.

Tableau 1 - Critères de diagnostic du myélome multiple d'après l'IMWG 2014. CRAB : hypercalcémie, insuffisance rénale, anémie, et lésions osseuses ; IRM : Imagerie par Résonance Magnétique ; TEP-FDG : Imagerie par tomographie à émissions de positons au 18-F-Fluorodésoxyglucose.

Taux de plasmocytes clonaux $\geq 10\%$ et un des critères suivants :	
- Critère CRAB avec atteinte de l'organe impliqué - Taux de plasmocytes clonaux $\geq 60\%$ - Ratio de chaîne légère libre sérique ≥ 100 - ≥ 2 lésions focales $\geq 5\text{mm}$ à l'IRM	
Critères CRAB	Hypercalcémie : calcium $>0.25\text{mg/dL}$ au-dessus de la limite normale ou $>11\text{mg/L}$ Insuffisance rénale : créatinine sérique $>2\text{mg/dL}$ ou clairance $<40\text{mL/min}$ Anémie : hémoglobine $>20\text{g/L}$ sous la limite de la normale ou $<100\text{g/L}$ Lésion osseuse : Lésion ostéolytique à l'IRM ou TEP-FDG.

Tableau 2 - Stratification de risque par ISS et R-ISS d'après Palumbo et al., 2015. B2M : microglobuline beta-2 ; LDH : lactate déshydrogénase. La FISH inspecte del(17p) sur TP53, t(4;14) et t(14;16).

International Staging System	Stade I : B2M $< 3.5\text{mg/L}$ et Albumine $\geq 3.5\text{g/dL}$
	Stade II : ni Stade I ni Stade III
	Stade III : B2M $\geq 5.5\text{mg/L}$
ISS révisé	Stade I : ISS-I, LDH normale et FISH normale
	Stade II : ni Stade I ni Stade III
	Stade III : ISS-3 et LDH ou FISH anormale

2.2. DIAGNOSTIC ET ÉPIDÉMIOLOGIE DU MYÉLOME

En France, il n'existe pas de centre organisé pour le dépistage des hémopathies malignes comme c'est le cas pour faciliter le diagnostic précoce du cancer colorectal ou du cancer du sein par exemple. La prévalence élevée des pics monoclonaux chez les personnes âgées par rapport à la rareté du cancer ne justifie pas cette approche. En revanche, les études observationnelles de l'Intergroupe Francophone du Myélome (IFM) GENOMGUS et CARRISMM définissent et valident les critères pronostiques de progression clinique. Aux États-Unis, le Dana Farber Cancer Institute propose la collecte de prélèvements biologiques de donneurs sains à visée de recherche en oncohématologie au sein des cohortes PCROWD et PROMISE. Un objectif de ces études est de montrer quels biomarqueurs (comme des mutations somatiques délétères connues) confèrent un risque élevé de progression vers un stade clinique de cancer hématologique et motiveraient une intervention thérapeutique précoce en amont du diagnostic de myélome.

Le myélome multiple établi d'après les critères de l'International Myeloma Working Group et reproduit en Tableau 1 représente environ 1-2% des cancers, 10% des cancers hématologiques (c'est le second plus courant après les lymphomes non-hodgkiniens) pour un âge médian au diagnostic de 70 ans chez l'homme et 74 ans chez la femme (INCa and AF3M, 2015; Kumar et al., 2017a). L'incidence est d'environ 5 000 cas annuels en France, en hausse pour toutes les classes d'âge depuis 1990 sans que le vieillissement de la population n'explique seul les niveaux d'augmentation. Les hommes sont légèrement plus touchés (54%), et s'il reste incurable, la durée médiane de survie a doublé avec l'introduction de nouveaux traitements dans les années 2000 pour dépasser les 6 ans (Kumar et al., 2008, 2014, 2017a).

La prolifération des cellules tumorales et la sécrétion importante d'anticorps et de cytokines est historiquement associée aux manifestations cliniques de la maladies connues sous l'acronyme CRAB pour hypercalcémie, insuffisance rénale, anémie, et lésions osseuses, ou *b*one disease with lytic lesions.

La ponction de moelle est requise pour prouver l'infiltration plasmocytaire, mais elle n'est pas dirigée sur les niches de prolifération détectées en imagerie médicale, on considère qu'on récupère des plasmocytes circulants dans la moelle. La ponction est invasive, généralement réalisée sur le sternum ou la crête iliaque, et le matériel biologique est précieux par rapport à du sang périphérique, en particulier dans les études longitudinales et de suivi de la réponse au traitement.

Étant donné la rareté des plasmocytes normaux au sein de la moelle osseuse, on considère souvent que la fraction de cellules CD138+ restreinte à un isotype *lambda* ou *kappa* correspond à l'envahissement tumoral – l'approximation ne tient pas dans l'évaluation fine de la maladie résiduelle mesurable (MRD) post-traitement par exemple. La stratification de risque des patients est systématiquement évaluée selon les critères de l'ISS (International staging system) et R-ISS (Revised-ISS) reproduits en Tableau 2 (Palumbo et al., 2015). La stratification de risque peut être complétée par d'autres facteurs pronostiques comme le nombre de cellules tumorales circulantes, la présence de maladie extra-médullaire, ou des marqueurs moléculaires qui font encore débat tels qu'une signature d'expression de haut risque GEP70, IFM-15, HOVON, ou la présence de mutation TP53 (Kumar et al., 2017a).

2.3. INTERVENTIONS PHARMACEUTIQUES ET ESSAIS CLINIQUES

Les traitements recommandés du myélome multiple en première ligne combinent un inhibiteur du protéasome (PI), un immunomodulateur (IMiD), et des corticoïdes. En pratique, cette association correspond souvent au bortezomib, lenalidomide et dexaméthasone (VRd) (Durie et al., 2017). L'essai clinique de phase 3 IFM/DFCI 2009 dont les données sont utilisées dans ce manuscrit montre l'intérêt pour la survie sans progression (PFS) de l'association du VRd avec une autogreffe de cellules souches de la moelle osseuse chez les patients qui y sont éligibles au diagnostic (Attal et al., 2017).

L'âge est une variable importante qui oriente la thérapie au diagnostic et délimite deux groupes de patients selon qu'ils sont diagnostiqués avant ou après 65 ans. Sauf pour les patients plus âgés dont l'état le permet, l'autogreffe est réalisée jusqu'à 65 ans. Chez les patients identifiés à haut-risque cytogénétique et éligibles, une double autogreffe de cellules souches est proposée.

Comme dans beaucoup de cancers et de disciplines médicales en général, la place de l'immunothérapie dans le traitement et la recherche contre le myélome augmente considérablement depuis les années 2010, en particulier chez les sujets en rechute pour lesquels les options thérapeutiques sont plus nombreuses. Les anticorps monoclonaux anti-CD38 (daratumumab ou isatuximab) et anti-SLAMF7 (elotuzumab) ont fait l'objet de développements spécifiques pour traiter le myélome. L'essai clinique de phase 3 CASSIOPEIA, dont une partie des données sont utilisées dans ce manuscrit, montre le bénéfice lié à l'ajout du daratumumab au triplet PI, IMiD et corticoïde dans la prise en charge des MM nouvellement diagnostiqués (NDMM) et éligibles à l'autogreffe (Moreau et al., 2019a).

Des protocoles cliniques dédiés sont mis en place pour les patients non-éligibles à l'autogreffe et montrent un bénéfice comparable (Facon et al., 2019).

Les immunothérapies de lymphocytes T à récepteur antigénique chimérique (CAR-T) font l'objet d'essais cliniques de phases précoces en ciblant BCMA, mais les premiers résultats indiquent une rechute rapide par perte de l'antigène décrite dès 2016 (Cohen et al., 2016; Simon and Riddell, 2020). La perte transmembranaire de BCMA peut être due au clivage de la protéine par le complexe de la gamma-sécrétase. La combinaison d'un inhibiteur de gamma-sécrétase et du CAR-T BCMA permettrait de restaurer l'expression membranaire de l'antigène et l'efficacité du traitement (Green et al., 2019; Pont et al., 2019). Un second mécanisme de perte de l'antigène est la délétion bi-allélique du gène *BCMA*, un phénomène présenté au congrès de l'American Society of Hematology (ASH) 2020 (Leblay et al., 2020; Samur et al., 2020) et qui motive l'exploration d'autres pistes de CAR-T ciblant par exemple GPRC5D (Smith et al., 2019; Simon and Riddell, 2020).

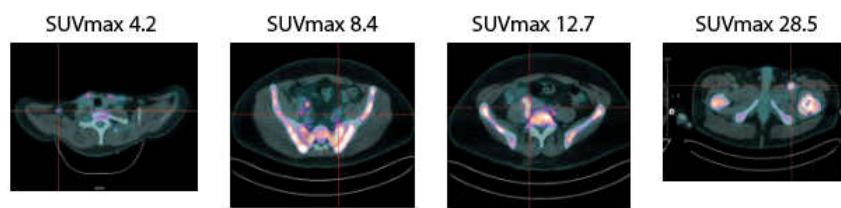


Figure 9 - Hétérogénéité intra-patient inter-lésionnelle explorée à la TEP-FDG. Illustration de l'équipe de recherche en oncologie nucléaire du CRCINA à Nantes. SUVmax : consommation maximale standardisée de glucose.

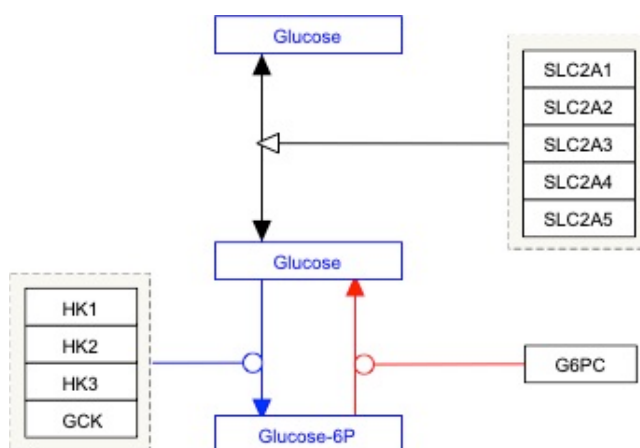


Figure 10 – Transport du glucose et premières étapes du pathway de la glycolyse. Image de WikiPathways Pathway:WP4361. La famille du SLC2A1-5 code pour les transporteurs de glucose (GLUT1-5). La famille des HK code pour les hexokinases. GCK : glucokinase ; G6PC : sous-unité catalytique de la glucose-6-phosphatase.

Tableau 3 - Variables pronostiques en imagerie par TEP-FDG d'après Michaud-Robert et al. 2020.

Lésion	Définition	Biomarqueur suggéré
Lésions focales (FL)	Consommation de FDG supérieure au bruit de fond sur deux sections successives, avec ou sans ostéolyse	Mauvais pronostic si > 3
Maladie extramédullaire (EMD)	Envahissement d'un tissu séparé de l'os	Mauvais pronostic si présent
Maladie paramédullaire (PMD)	Envahissement d'un tissu contigu à l'os	Mauvais pronostic si présent
SUVmax dans FL	Consommation de FDG maximale standardisée dans les lésions focales	Pourrait être de mauvais pronostic si > 4.2
TEP anormale	Présence de lésions focales ou d'EMD ou de PMD ou d'une consommation de FDG médullaire diffuse	L'absence d'anomalie est de bon pronostic

2.4. INTÉRÊT DE LA TEP-FDG DANS LE MYÉLOME

Chacune des deux études évoquées ci-dessus IFM/DFCI 2009 et CASSIOPEIA est accompagnée d'une étude secondaire visant à évaluer l'intérêt de l'imagerie par tomographie à émission de positron au ^{18}F -fluorodésoxyglucose (TEP-FDG) : respectivement IMAJEM, dont les résultats sont publiés (Moreau et al., 2017a), et CASSIOPET dont les résultats préliminaires ont été présentés au congrès de l'ASH 2019 (Moreau et al., 2019b).

L'intérêt de la TEP-FDG dans le MM a fait l'objet de revues de référence de l'équipe du Pr Kraeber-Bodéré (Bailly et al., 2019; Jamet et al., 2019a; Michaud-Robert et al., 2020). Brièvement, les lésions focales osseuses visibles en imagerie sont considérées depuis 2014 par l'IMWG comme un critère de définition du myélome encourageant une intervention thérapeutique, même si elles sont asymptomatiques (Rajkumar et al., 2014). Parmi les techniques d'imagerie, la TEP-FDG est utilisée pour évaluer et diagnostiquer le myélome tout en fournissant des variables pronostiques pour la survie et la rechute résumées en Tableau 3 : TEP anormale, nombre de lésions focales, présence de maladie extramédullaire et paramédullaire, consommation maximale standardisée de glucose (SUVmax) illustrée en Figure 9.

L'utilisation d'un traceur radioactif non-assimilé par le métabolisme du glucose (le FDG) permet de localiser les sites hyper-métaboliques dans le corps entier. Il est probable que le métabolisme du glucose joue un rôle majeur dans l'influence des variables d'imagerie (Figure 10), mais cette question fait encore débat (Rasche et al., 2017a), notamment pour les myélomes dépendant des voies métaboliques alternatives ou complémentaires à la consommation de glucose, par exemple glutaminique (Bolzoni et al., 2016).

Rasche et collègues ont montré que dans la recherche de myélome au diagnostic, les patients TEP-négatifs mais positifs à la séquence de diffusion (DWIBS), appelés « TEP-faux-négatifs » expriment pas ou peu d'*HK2* (Rasche et al., 2017a). Pour autant, la découverte n'est expliquée par aucun mécanisme biologique quand bien même elle serait exclusive à certains sous-groupes de myélome, et offre donc des opportunités de recherche.

Usmani et collègues ont montré que la prévalence de la maladie extra-médullaire est plus élevée statistiquement chez les patients aux signatures de haut-risque et aux groupes moléculaires MF (qui présentent une translocation MAF ou MAFB) et PR (qui présentent une signature de prolifération importante) et qui sont décrits en détail en paragraphe 4.1 page 49 (Usmani et al., 2012). Toutefois, aucun mécanisme biologique ne semble expliquer cette association et l'étude n'a pas été reproduite depuis.

Des questions restent en suspens dans l'association entre biologie moléculaire et variables d'imagerie : quels sont les myélomes diagnostiqués avec une TEP négative ? Est-ce que la biologie moléculaire confirme le meilleur pronostic clinique des TEP négatives ? Que représentent les variables métaboliques et de localisation spatiale de la tumeur par rapport à la biologie moléculaire de la tumeur ? Y a-t-il des différences moléculaires entre les différentes maladies, extra-médullaire (qui affecte un tissu sans os atteint de façon contiguë) et para-médullaire (qui affecte un tissu depuis un os attenant) ?

2.5. ÉVALUATION DE LA MALADIE RÉSIDUELLE MESURABLE

L'IMWG recommande d'effectuer une évaluation et un suivi de la réponse thérapeutique par TEP-FDG (Kumar et al., 2016; Cavo et al., 2017), qui permet la caractérisation des organes atteints et des clones résiduels localisés non-détectés lors de la ponction de moelle locale.

Dans le suivi de la réponse thérapeutique, l'imagerie vient en complément de l'évaluation de la maladie résiduelle mesurable (MRD) par une méthode standardisée nécessitant une ponction de moelle. La mesure du clone résiduel par séquençage ou cytométrie en flux de nouvelles générations (resp. NGS et NGF) atteint un seuil de détection de l'ordre de 1 pour 100 000 ou 1 pour 1 000 000 sur ce matériel biologique. À propos, la signification de l'acronyme MRD évolue depuis *Minimal Residual Disease* pour devenir *Measurable Residual Disease* (Paiva et al., 2019; Béné, 2020), mais les deux acronymes désignent la même technique.

Les deux techniques de quantification de la MRD NGS et NGF sont plus adaptées que les techniques biochimiques pour la surveillance des patients en réponse complète, dont la proportion augmente avec l'amélioration des traitements (Kumar et al., 2016; Moreau and Zamagni, 2017). Historiquement, le groupe espagnol et le consortium EuroFlow ont porté le NGF quand le groupe francophone a plutôt développé le NGS, mais les bénéfices des technologies sont comparables (Medina et al., 2020).

La MRD négative a une valeur pronostique pour la survie et pour la rechute, et semble effacer les effets de la classification de la sévérité de la maladie par R-ISS (Paiva et al., 2019).

Une nouvelle notion coïncide avec l'émergence du NGS/NGF : le pilotage par MRD, et l'essai européen PERSEUS visera à montrer si la MRD décisionnelle, c'est-à-dire qui conditionne l'inclusion dans des bras de traitements, a un intérêt dans le traitement du myélome multiple (Sonneveld et al., 2019).

L'étude des cellules tumorales circulantes (CTCs), de l'ADN libre-circulant (cfDNA) ou de contenu en exosomes sont trois outils supplémentaires de caractérisation de la tumeur et de suivi du traitement qui pourraient émerger dans la pratique clinique, mais leur utilisation fait débat et les limitations liées à la sensibilité sont importantes (Guo et al., 2018; Manier et al., 2018). L'étude des CTC et du cfDNA permettrait notamment de limiter l'acte invasif de ponction de moelle. Aussi, CTC et cfDNA pourraient améliorer la caractérisation de la diversité sous-clonale entre les niches de prolifération car les méthodes ne sont pas ciblées localement. Cette question fait débat puisqu'à l'inverse, les CTC sont circulantes par définition et pourraient donc biaiser la représentation des clones dont la survie dépend de l'interaction avec le microenvironnement de la moelle osseuse.

2.6. VIE ET SURVIE DU MYÉLOME : LE RÔLE DE LA FAMILLE DU BCL-2

La chaîne légère sérique (sFLC) des immunoglobulines est quantifiée lors du diagnostic clinique et peut être utilisée pour le suivi et l'évaluation de la réponse au traitement. De façon intéressante, la chaîne légère des immunoglobulines est presque toujours exprimée par les myélomes (>95%), contrairement à la chaîne lourde qui est détectée dans 80-85% des cas. Les myélomes à chaîne légère (LCMM) présentent majoritairement un réarrangement de l'IgH non-fonctionnel sur un allèle et un second allèle IgH en configuration germinale (Magrangeas et al., 2004), ce qui montre que l'expression de la chaîne lourde n'est pas nécessaire à la survie du plasmocyte tumoral.

Mais la survie anormale du plasmocyte malin proliférant peut aussi être considérée comme un échappement aux signaux de déclenchement de la mort cellulaire programmée : l'apoptose. La régulation intrinsèque de l'apoptose est médiée par les protéines de la famille de BCL2 résumée en Figure 11. On distingue couramment trois classes dans cette famille en fonction de leurs caractéristiques fonctionnelles et structurales :

(1) les protéines anti-apoptotiques (BCL2, MCL1, et BCLxL / *BCL2L1*), (2) les pro-apoptotiques multi-domaines BAX et BAK / *BAK1*, et les pro-apoptotiques « BH3-only » (entre autres BIM / *BCL2L11*, NOXA / *PMAIP1*, PUMA / *BBC3*).

Dans la voie physiologique de l'apoptose, un stress cellulaire induit ou active les BH3-only. La famille BH3-only est à son tour divisée en deux groupes : (A) les activateurs et (B) les sensibilisateurs. Les sensibilisateurs (par exemple NOXA) séquestrent les anti-apoptotiques et permettent la libération des activateurs. Les activateurs (PUMA et BIM) séquestrent les anti-apoptotiques et/ou activent directement BAX et BAK, qui oligomérisent à la surface de la membrane externe de la mitochondrie. La formation de pores et la perméabilisation qui en résultent permettent le relargage de protéases, d'activateurs de caspases et de cytochrome C. Ensemble, ces facteurs couplés à APAF1 forment l'apoptosome, qui active la cascade des caspases et le démantèlement de la cellule (Singh et al., 2019).

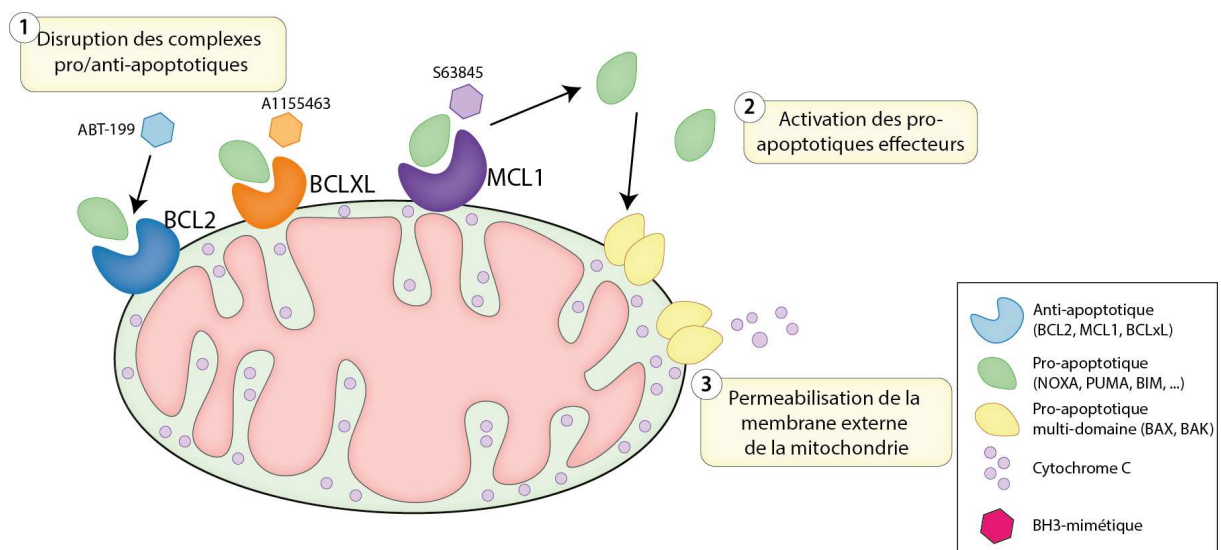


Figure 11 - Induction de l'apoptose par molécules BH3-mimétiques. Schéma de la mitochondrie inspiré de présentations du Dr Martine Amiot et du Dr Lawrence Boise. Chaque protéine anti-apoptotique est représentée avec une molécule spécifique à potentiel thérapeutique.

Le groupe d'Anthony Letai à Boston propose de caractériser un état cellulaire « primed for death », c'est-à-dire prêt à entrer en apoptose, par un essai fonctionnel : le BH3 profiling (Certo et al., 2006). L'essai consiste à exposer les cellules à des peptides mimant les domaines des BH3-only sensibilisateurs et à mesurer la mort ou les niveaux de cytochrome C induits par l'exposition au peptide. Connaissant a priori la co-dépendance des sensibilisateurs aux anti-apoptotiques, le niveau de mort induite par tel ou tel peptide donne la dépendance cellulaire à telle molécule ou couple de molécules anti-apoptotiques. L'objectif est double : d'abord, prédire les niveaux de

réponses aux traitements cytotoxiques non-ciblés par étude des stress inducteurs spécifiques, et ensuite, cibler spécifiquement le ou les anti-apoptotiques dont la cellule dépend pour sa survie (Figure 11).

Le venetoclax, première molécule spécifiquement anti-BCL2 disponible en clinique, est efficace dans de nombreux cancers hématologiques incluant leucémies lymphocytaires chroniques (CLL), MCL et FL (Roberts et al., 2016; Davids et al., 2017). Dans le MM, la sensibilité du venetoclax a d'abord été montrée dans les cellules présentant la translocation t(11;14) entre l'IgH et la Cycline D1 et qui représente environ 15% des patients (Touzeau et al., 2014; Kumar et al., 2017b). En association avec d'autres molécules, la sensibilité pourrait être élargie à d'autres groupes de patients.

Le ciblage spécifique du BCLxL affecte les mégacaryocytes et cause des thrombocytopénies – le développement de l'ABT-263/737 ciblant le BCL2 et le BCLxL a d'ailleurs été stoppé pour ces raisons (Vandenberg and Cory, 2013).

Finalement, d'autres molécules sont développées pour cibler le MCL1 (Kotschy et al., 2016). On a vu que l'expression du MCL1 était fortement liée à la survie des plasmocytes (Peperzak et al., 2013), son ciblage en clinique pourrait être à plus large spectre que BCL2. Néanmoins, les cardiomyocytes murins dépendent aussi du MCL1 et font craindre des toxicités cardiaques importantes chez l'homme (Thomas et al., 2013). Une alternative envisagée pour cibler le MCL1 est de le dégrader avec un PROTAC. Les PROTAC sont des molécules à deux domaines de liaison protéique reliés par un linker. L'un des deux domaines cible une protéine d'intérêt (MCL1) et l'autre domaine cible une E3-ubiquitine ligase qui enverrait la première cible pour dégradation au protéasome. Idéalement dans notre exemple, la ligase serait exprimée dans les plasmocytes mais pas dans les cardiomyocytes (Papatzimas et al., 2019).

Bien sûr, ce projet de médecine translationnelle ne représente pas le standard de traitement, d'ailleurs le venetoclax (ABT-199, anti-BCL2) n'a pas d'autorisation de mise sur le marché pour le MM et est associé à des décès dans l'essai BELLINI, où seuls des sous-groupes de patients comme les t(11;14) ou à forte expression de BCL2 ont montré un rapport risque-bénéfice positif (Kumar et al., 2020). En un sens, ce résultat favorise l'idée d'utiliser le répertoire des molécules BH3-mimétiques couplé à un essai fonctionnel personnalisé ou à une classification moléculaire pour prédire la sensibilité à un arsenal thérapeutique, de même qu'en infectiologie, l'antibiogramme prédit la sensibilité à un ou plusieurs antibiotiques disponibles en clinique.

3. LE GÉNOME DU MYÉLOME ET SES ANOMALIES

La génomique du myélome est un domaine d'étude qui a fait l'objet de nombreux travaux notamment sur le site nantais. Les analyses de FISH et de puces à ADN ont révélé certes une complexité et une hétérogénéité importantes d'un patient à l'autre, mais aussi des éléments récurrents susceptibles d'expliquer l'oncogenèse plasmocytaire ou même de dévoiler des vulnérabilités exploitables en clinique.

Environ 50% des myélomes présentent une translocation de leur chaîne lourde (IgH) avec un oncogène, parfois considérée comme événement primaire, c'est-à-dire à l'origine du myélome multiple – et donc clonale, ou présente dans toutes les cellules. De manière quasiment exclusive, l'autre moitié des patients présente une hyperdiploïdie (HRD), majoritairement de chromosomes impairs. D'autres anomalies des bras de chromosomes (gain1q, del13q et del17p) ont fait l'objet de nombreux travaux pour leur intérêt pronostique en termes de survie et de progression clinique.

Dans ce chapitre 3, nous complétons le dogme de l'antagonisme entre myélomes transloqués et myélomes hyperdiploïdes à la lumière des altérations du génome découvertes notamment grâce au NGS.

Les deux chapitres suivants, chapitre 4 à partir de la page 49 et chapitre 5 à partir de la page 59, développent respectivement les enjeux et connaissances spécifiques liés à l'étude des produits de la transcription et à la l'hydroxyméthylation de l'ADN.

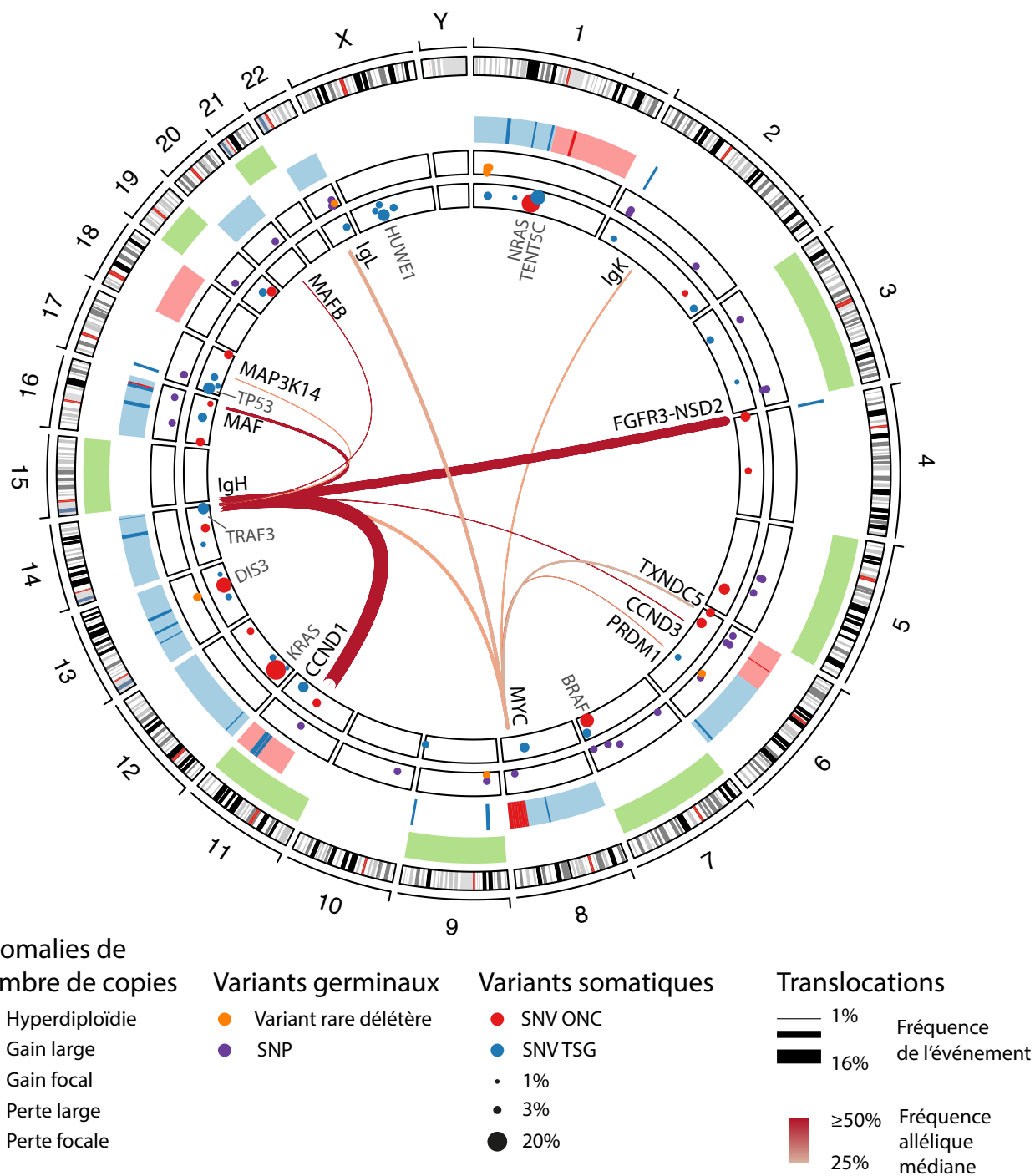


Figure 12 - Le génome du myélome multiple au diagnostic. Réalisé avec R/circlize. Gains et pertes de copies d'après Walker et al., 2018 et Manier et al., 2017. Variants germinaux d'après Pertesi et al, 2020. Variants somatiques d'après Walker et al., 2018 et la méthode 20/20. Translocations d'après Barwick et al., 2019. Les mutations somatiques et translocations sont reproduites lorsqu'elles affectent plus de 1% des patients des études respectives. Elles sont annotées si la prévalence est supérieure à 5%. SNP : Single nucleotide polymorphism ; SNV : Single nucleotide variation ; ONC : oncogène; TSG : gène suppresseur de tumeur.

3.1. VARIANTS GÉNÉTIQUES HÉRITÉS

La « part héritable » de la maladie est estimée par deux méthodes : l'une étudie les clusters de cas familiaux et leur pedigree à la recherche de variations génétiques rares associées à un OR élevé de développer la maladie, tandis que l'autre méthode cherche des variations génétiques communes à associer statistiquement à un risque (souvent plus faible) de contracter la maladie dans des cohortes plus grandes.

Parmi les regroupements de cas de cancers familiaux, les données de registres de cancers familiaux (notamment suédois) montrent que la famille rapprochée d'un patient a elle-même 2 à 4 fois plus de risque de développer un MGUS ou MM que le reste de la population, ce qui suggère d'abord que la prédisposition génétique à la maladie existe bien. L'association du risque d'incidence va dans le même sens pour les CLL et les lymphomes non-hodgkiniens (NHL) mais pas dans les autres leucémies, ni avec les lymphomes Hodgkiniens, ce résultat coïncide avec l'origine cellulaire B du myélome (Altieri et al., 2006; Pertesi et al., 2020).

On étudie les cas familiaux et héréditaires à la lumière de la biologie de la tumeur, car disons-le d'emblée, ces cas sont rares et font figure d'exception, néanmoins ils donnent des clés précieuses pour l'interprétation des cas pathologiques non-familiaux, et peut-être pour cibler les dysfonctionnements spécifiques à la lignée B ou au plasmocyte.

Les 6 gènes à variant rare et risque fort sont décrits dans le Tableau 4, reproduit d'après Pertesi et al. 2020 et ajoutés au résumé en Figure 12. Les mutations germinales délétères de *DIS3* présentent un intérêt particulier : *DIS3* est également muté somatiquement de façon récurrente dans le MM (11%) sous forme de mutation perte-de-fonction. *DIS3* code pour la partie catalytique du complexe de l'exosome, est situé sur le chromosome 13q, et la perte d'hétérozygotie de *DIS3* est courante puisque le chr13 est perdu dans la moitié des cas de MM. Les autres cibles identifiées incluent des gènes régulateurs du cycle cellulaire (notamment le suppresseur de tumeur *CDKN2A*), de la réparation de l'ADN (*USP45*), des marques d'histone et du remodelage de la chromatine (*KDM1A*, *ARID1A*). En somme, les exemples de myélomes familiaux peuvent guider l'interprétation du risque de développer la maladie et dévoilent trois mécanismes qui favoriseraient la tumeur : la dérégulation du cycle cellulaire, de la biologie de la chromatine, et du traitement des ARN.

Tableau 4 - Variants rares constitutifs associés au diagnostic de myélome multiple. Reproduit et commenté d'après Pertesi et al., 2020.

Gene	Variant	Commentaire et fonction du gène
<i>EP300</i>	c.2084G>A; p.Arg695His	22 variants-candidats, <i>EP300</i> est un exemple non-validé ; Acétyl-transférase d'histone
<i>CDKN2A</i>	24bp duplication	4 mélanomes dans la famille et 1 MM avec perte d'hétérozygotie de <i>CDKN2A</i> (cycle cellulaire)
	c.213C>A; p.Asn71Lys	Cancers multiples
<i>USP45</i>	c.1861C>G; p.Gln621Glu	Cohorte Utah, <i>USP45</i> régule la réparation de l'ADN ; Déubiquitination
	c.2071C>T; p.Gln691*	
<i>ARID1A</i>	c.268A>G; p.Ser90Gly	Cohorte Utah, <i>ARID1A</i> est membre du complexe SWI/SNF
	c.2668A>G; p.Met890Val	
<i>KDM1A</i>	c.805_806delAG; p.Arg269Aspfs*7	<i>KDM1A</i> est une H3K4 déméthylase, l'inhibition de <i>KDM1A</i> après stimulation antigénique chez la souris favorise la prolifération plasmocytaire
	c.707delA; p.Gln236Hisfs*3	
	c.1424T>C; p.Leu475Pro	
	c.2003G>C; p.Arg692Pro	
<i>DIS3</i>	c.1755+1G>T; p.Arg557Argfs*3	L'allèle mutant est systématiquement dégradé par NMD ; <i>DIS3</i> est une RNase du complexe de l'exosome
	c.2875T>C; p.*959Glnext*14	Variant commun à deux familles
	c.1883+1G>C	Un membre non-porteur présente une amylose AL

Tableau 5 - Polymorphismes nucléotidiques communs constitutifs associés au diagnostic de myélome multiple. Liste de variants issue de Pertesi et al., 2020. Ref : allèle de référence ; Alt : allèle alternatif ; OR : odds ratio ; kbp : 1 000 paires de bases

Chr	Pos (hg38)	rsID	Ref	Alt	OR	P value	Gènes (fonction ou pathway d'après UniProt)
2	25436375	rs6746082	A	C	1.29	1 × 10 ⁻⁷	<i>DTNB</i> (liaison au zinc) <200kbp (kilo-paire de base) de <i>DNMT3A</i>
2	25390277	rs7577599	T	C	1.23	1 × 10 ⁻¹⁸	
2	173944171	rs4325816	T	C	1.12	7 × 10 ⁻⁹	<i>SP3</i> (facteur de transcription)
3	41883906	rs1052501	C	T	1.32	7 × 10 ⁻⁹	<i>ULK4</i> (sérine/thréonine protéine kinase)
3	41950916	rs6599192	G	A	1.26	5 × 10 ⁻²⁰	
3	169774313	rs10936599	C	T	1.26	9 × 10 ⁻¹⁴	<i>MYNN</i> , <i>TERC</i> , <i>LRRC34</i>
3	169796797	rs10936600	A	T	1.2	1 × 10 ⁻¹⁶	
5	95907227	rs56219066	T	C	1.25	10 × 10 ⁻¹⁰	<i>ELL2</i> (Facteur d'élongation de la polymérase à ARN II)
5	95920020	rs1423269	A	G	1.17	8 × 10 ⁻¹²	
5	123407631	rs6595443	A	T	1.11	1 × 10 ⁻⁸	<i>CEP120</i> (liaison à l'extrémité C-terminale des protéines)
6	31139481	rs2285803	T	C	1.19	10 × 10 ⁻¹¹	Région du HLA
6	31148749	rs3132535	A	G	1.2	6 × 10 ⁻²¹	
6	15243787	rs34229995	C	G	1.37	6 × 10 ⁻⁸	<i>JARID2</i> (régulateur de méthyltransférase d'histone)
6	106219660	rs9372120	T	G	1.18	2 × 10 ⁻¹⁵	<i>ATG5</i> (autophagie)
7	21898622	rs4487645	C	A,T	1.24	3 × 10 ⁻²⁸	<i>DNAH11</i> , <i>CDCA7L</i>
7	106650672	rs17507636	C	T	1.12	9 × 10 ⁻⁹	<i>CCDC71L</i> (inconnue)
7	124943842	rs58618031	T	C	1.12	3 × 10 ⁻⁸	<i>POT1</i> (protection des télomères)
7	151253854	rs7781265	G	A	1.22	5 × 10 ⁻¹⁰	<i>SMARCD3</i> , <i>ABCF2</i> , <i>CHPF2</i>
8	127210176	rs1948915	T	C	1.15	3 × 10 ⁻¹²	<i>CCAT1</i> (inconnue)
9	21991924	rs2811710	C	T	1.15	4 × 10 ⁻¹¹	<i>CDKN2A</i> (cycle cellulaire)
10	28567890	rs2790457	G	A	1.12	3 × 10 ⁻⁶	<i>WAC</i> (liaison à la chromatine et à l'ARN Pol II)
11	69648142	rs9344	G	A	1.82	8 × 10 ⁻¹¹	<i>CCND1</i> (cycle cellulaire)
16	30689537	rs13338946	T	C	1.15	1 × 10 ⁻¹³	<i>FBR</i> , <i>SRCAP</i> , <i>PRR14</i> , <i>RNF40</i>
16	74630845	rs7193541	T	C	1.13	4 × 10 ⁻¹⁰	<i>RFWD3</i> (Ubiquitine ligase)
17	16945825	rs4273077	A	G	1.26	8 × 10 ⁻⁹	<i>TNFRSF13B</i> (Immunité et stimulation B)
17	16939677	rs34562254	G	A	1.3	1 × 10 ⁻¹⁹	
19	16327850	rs11086029	T	A	1.14	7 × 10 ⁻¹¹	<i>KLF2</i> (Facteur de transcription)
20	48738472	rs6066835	T	C	1.23	7 × 10 ⁻¹⁰	<i>PREX1</i> (Facteur d'échange GDP-GTP)
22	35303589	rs138740	C	T	1.18	6 × 10 ⁻⁸	<i>TOM1</i> (Transport protéique)
22	35304495	rs138747	A	T	1.21	3 × 10 ⁻⁸	
22	39146287	rs877529	G	A	1.23	8 × 10 ⁻¹⁶	<i>CBX7</i> (chromatine, composant du PcG PRC1-like) proximal d' <i>APOBEC3</i>
22	39150140	rs139402	T	C	1.23	4 × 10 ⁻²⁶	

Au contraire des anomalies rares à forte pénétrance, les études GWAS (*genome-wide association studies*) ont identifié 23 loci indépendants associés à un risque modéré de MM ($1 < OR < 1.4$) et un SNP (*single-nucleotide polymorphism*) supplémentaire est associé à un risque plus important de développer un MM t(11;14) ($OR = 1.8$) (Pertesi et al., 2020). Nous reproduisons ces données agrégées dans le Tableau 5.

Parmi les 23 autres loci associé à un risque de MM, on retrouve un SNP intronique de *CDKN2A*, suppresseur de tumeur déjà identifié parmi les cibles des mutations familiales rares, mais aussi des éléments de la biologie plasmocytaire comme *ELL2* ou *TNFRSF13B*. *ELL2* est un facteur d'élongation fortement exprimé dans les plasmocytes qui facilite la transcription d'un ARNm d'Ig sécrété contrairement aux Ig du lymphocyte B (Martincic et al., 2009), mais l'aspect causal du variant n'est pas établi. *TNFRSF13B*, plus connu sous le nom TACI, code pour un second récepteur très affin pour APRIL – on connaît mieux BCMA – et qui fait l'objet d'un CAR-T en développement (Lee et al., 2018). La fonction du variant faux-sens de *TNFRSF13B* n'est pas connue mais a aussi été associée à des niveaux élevés d'IgG chez des individus sains (Jonsson et al., 2017).

Les travaux GWAS montrent une incidence limitée des variants communs dans le risque de myélome multiple, qui nous poussent à travailler dans deux directions : (1) Faut-il étudier plutôt les effets secondaires des traitements ? On peut citer par exemple le travail du groupe nantais (Magrangeas et al., 2016) sur un SNP *PKNOX1* en 21q associé à la neuropathie induite par le bortezomib avec un OR égal à 2 ; et (2) le myélome multiple étant hétérogène, ne faut-il pas subdiviser la maladie en sous-groupes pour augmenter les chances de découvertes de SNP à impact modéré à moyen, comme l'ont fait Nils Weinhold *et al.* pour le SNP *CCND1* associé aux myélomes t(11;14) (Weinhold et al., 2013) ?

3.2. MUTATIONS ACQUISES

Le développement du NGS a permis de dresser un tableau des mutations somatiques dans le myélome multiple et d'identifier de nombreuses cibles thérapeutiques d'une part, mais aussi de définir plus généralement le caractère anormal du génome de myélome et l'évolution de la maladie d'autre part.

L'analyse de génome du cancer requiert des méthodes particulières : les outils développés pour l'analyse de génome constitutif doivent être adaptés à quatre caractéristiques présentées par les échantillons biologiques de tumeurs.

(1) Les échantillons biologiques ne sont pas purs en général. On parle de **cellularité**, ou de pureté cellulaire pour décrire la proportion de cellules tumorales dans l'échantillon. Certes, dans le myélome symptomatique, le marqueur CD138 nous permet dans la plupart des cas d'atteindre une pureté qui limite les erreurs d'analyse (>95% au laboratoire), mais les stades précurseurs faiblement envahis et les prélèvements sanguins sont particulièrement touchés par cette caractéristique.

(2) Réciproquement, il est critique de prêter attention à la non-contamination de l'échantillon d'ADN constitutif par la tumeur, en particulier dans les myélomes en phase leucémique, où la masse tumorale se retrouve dans la circulation sanguine. La **contamination inversée** (« Tumor in Normal » dans la littérature) limite fortement la détection de mutations somatiques dans l'échantillon tumoral, on cherche à l'éviter absolument.

(3) Le **nombre de copies** présentes pour un locus donné dans chaque noyau n'est en règle générale pas égal à 2. Au-delà des cas de tri- et tétra-ploïdies complètes, le nombre de copies d'ADN génomique varie aux échelles du gain complet de chromosome, de bras de chromosome ou de locus. Pour une copie supplémentaire d'un locus, on parle de gain du nombre de copies ; pour deux copies et plus, on parle d'amplification du nombre de copies.

(4) Les tumeurs ne sont pas homogènes en variants alléliques (ni en nombre de copies) : dans un modèle Darwinien d'évolution des sous-clones tumoraux en compétition pour proliférer, certaines mutations sont acquises plus ou moins tardivement dans l'évolution de la tumeur : on ne les trouve que dans une fraction des cellules, et donc une fraction des reads de séquençage. L'estimation de la **clonalité** des mutations et événements structuraux dépend de la cellularité et du nombre de copies au locus donné.

En résumé : **cellularité**, **contamination inversée**, **nombre de copies** et **clonalité** des mutations sont quatre aspects à considérer pour toute analyse de génome de cancer.

Les mutations ponctuelles (*single nucleotide variation*, SNV) « drivers », c'est-à-dire les mutations somatiques qui augmentent la prolifération et la survie des cellules tumorales (Martincorena et al., 2017), sont détectées dans les cohortes par des approches fréquentistes (avec MutSigCV ou dNdSCV) ou des approches fonctionnelles *in silico* (20/20 et SomInaClust). 95% des patients présentent des mutations drivers au diagnostic du myélome multiple selon ces critères, une proportion comparable aux autres cancers (Bolli et al., 2018; The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium, 2020).

Les mutations oncogéniques consensuelles du myélome ont rapidement émergé après la publication des premiers séquençages complets du génome (WGS) et de l'exome (WES) (Chapman et al., 2011). Nous reproduisons en Tableau 6 le tableau récapitulatif de la revue de Manier et collègues (Manier et al., 2017a), enrichi d'analyses et méta-analyses publiées après 2017.

Tableau 6 – Gènes affectés par des mutations somatiques récurrentes dans le myélome. Valeurs en pourcentage, d'après Manier et al., 2017, Bolli et al., 2018 et Walker et al., 2018. NDMM : myélome multiple nouvellement diagnostiqué. TSG : gène suppresseur de tumeur ; TF : facteur de transcription.

	MM N=38 (Chapman et al., 2011)	NDMM (N=463) (Walker et al., 2015a)	MM N=203 (Lohr et al., 2014)	MM N=67 (Bolli et al., 2014)	NDMM N=418 (Bolli et al., 2018)	NDMM N=1273 (Walker et al., 2018)	Fonction du gène ou voie de signalisation (Walker et al., 2018)
KRAS	26	21 (34 parmi t(11;14))	23	20	26	22	MAPK/ERK
NRAS	24	19 (26 parmi t(11;14))	20	20	22	17	MAPK/ERK
DIS3	11	9 (25 parmi t(4;14))	11	1	8	10	RNase (exosome)
FAM46C / TENT5C	13	6	11	10	8	9	Poly-A-polymérase
BRAF	-	7	6	12	9	8	MAPK/ERK
TP53	8	3	8	12	8	6	TSG
TRAF3	-	4	5	2	4	5	NFκB
HUWE1	-	-	-	-	-	5	Ubiquitine ligase
EGR1	-	4	4	6	3	4	TF
ATM	-	3	4	3	5	4	Sérine/thréonine protéine kinase
HIST1H1E / H1-4	-	3	-	-	2	4	Épigénétique
FGFR3	-	3 (17 parmi t(4;14))	-	-	-	4	MAPK/ERK
IRF4	-	3 (10 parmi t(11;14))	2	-	6	3	TF ; développement B
ACTG1	-	-	2	-	-	3	Cytosquelette
CYLD	-	2	2	1	3	3	NFκB
MAX	-	2	1	-	-	3	TF
PRDM1	-	-	5	-	3	2	TF ; développement B
SP140	-	-	4	6	5	2	Inconnue
CCND1	5	2	3	4	-	2	Cycle cellulaire
LTB	-	3	1	4	3	2	Réponse immunitaire
RB1	-	2	3	-	1	2	Cycle cellulaire

Le Tableau 6 indique des mutations codantes drivers déjà connues dans les cancers en général (*TP53*, *KRAS*, *BRAF*) qui ciblent les voies ERK/MAPK et de réparation de l'ADN. Des mutations *NRAS* et *KRAS* sont trouvées chez 43% des patients, presque tout le temps antagonistes l'une de l'autre et sans impact sur la survie, majoritairement sous-clonales donc associées à la progression mais pas à l'oncogenèse primaire (Bolli et al., 2014; Walker et al., 2015a). Sans tenir compte de la dissémination spatiale parmi les niches de prolifération, la diversité des gènes ciblés et des fractions alléliques des mutations rendent difficilement imaginable leur utilisation en clinique comme cela est fait dans d'autres cancers pour *BRAF_V600E* par exemple (*cf.* Fig. 3a *in* Rasche et al. 2017).

L'analyse WES/WGS met en lumière un ensemble de gènes mutés supplémentaires plus spécifiques à la biologie plasmocytaire. En particulier, les mutations *DIS3* (majoritairement non-sens) et *FAM46C/TENT5C* (non-sens et décalage du cadre ouvert de lecture, ORF) sont prédites pour avoir un rôle oncogénique dans une proportion importante de patients (jusqu'à 10% au diagnostic) et sont absentes ou très rares dans les autres types de cancers répertoriés dans le catalogue de mutations somatiques COSMIC v92. Notons que *DIS3* est déjà cité parmi les gènes mutés par des variants rares constitutifs délétères, et que la protéine normale participe à la dégradation des ARN via l'exosome. Les mutations de *FAM46C* sont associées aux myélomes hyperdiploïdes (Bolli et al., 2014; Walker et al., 2018). Le groupe de Dziembowski en Pologne a montré récemment que *FAM46C* possède une fonction de poly-A-polymérase non-canonique. *FAM46C* stabilise les ARNm et facilite la sécrétion d'Ig. L'extinction de l'expression de *FAM46C* dans des lignées cellulaires augmente leur prolifération, et la surexpression de *FAM46C* entraîne la mort cellulaire, ce qui indique bien un rôle de suppresseur de tumeur de *FAM46C*, potentiellement spécifique aux plasmocytes (Mroczek et al., 2017).

En résumé, une des particularités du myélome multiple dans le monde de la génomique du cancer est la mutation récurrente des mécanismes de traitement des ARN.

Les études WGS et WES ont également montré des mutations du génome non-codant ou dont la cible est inconnue. Une partie des mutations (si ce n'est la grande majorité) ne sont pas sélectionnées pour l'avantage sélectif qu'elles confèrent à la tumeur. Pour autant, ces mutations peuvent être informatives sur le processus d'oncogenèse.

Les signatures mutationnelles sont peut-être le plus grand succès d'analyse « métagénomique ». Elles utilisent les 6 mutations de nucléotides C·G → A·T, C·G → G·C, C·G→T·A, T·A→A·T, T·A→C·G, et T·A→G·C contextualisées en 5' et 3' pour obtenir $4 \times 6 \times 4 = 96$ trinuécléotides dont les fréquences d'apparition peuvent être associés à des mécanismes causaux. L'analyse des signatures mutationnelles se fait par factorisation de matrices non-négatives (NMF) d'après Alexandrov *et al.* ou par assignation aux signatures COSMIC (Alexandrov *et al.*, 2013, 2020). En plus des mutations signées par AID, APOBEC et le vieillissement au diagnostic (Walker *et al.*, 2015b; Maura *et al.*, 2019a, 2019b), le MM présente une nouvelle signature « melphalan » à la rechute, du nom de la chimiothérapie utilisée à haute dose pour l'autogreffe (Landau *et al.*, 2020). La présence de mutations clonales de signature « melphalan » à la progression soutient la théorie de sélection et d'expansion de clones minoritaires et résistants au traitement (Landau *et al.*, 2020; Rustad *et al.*, 2020a).

La détection de mutations sélectionnées sous pression thérapeutique est un domaine d'étude qui apparaît dans le MM. Les anomalies génomiques de *CRBN* acquises sous IMiD ont été décrites récemment (Gooding *et al.*, 2020). *CRBN* est ciblé par les IMiD (thalidomide, lenalidomide, pomalidomide et autres molécules de nouvelle génération). À l'acquisition de la résistance, un tiers des patients recevant un IMiD présente des nouvelles mutations, pertes de copies et variants structuraux qui affectent l'expression du *CRBN*. De façon intéressante, les anomalies sélectionnées de *CRBN* ne se limitent pas au gène mais touchent aussi la machinerie de l'épissage, et l'exon 10 qui code pour le domaine de liaison des IMiD tend à être moins exprimé chez ces patients (Gooding *et al.*, 2020).

L'analyse et la découverte du panorama des mutations ponctuelles codantes est riche mais peine encore à trouver des applications cliniques au quotidien. En fait, la pratique clinique est orientée par des découvertes à l'impact pronostique solidement démontré par plusieurs groupes de recherche et antérieur au séquençage haut-débit : les translocations IgH et les gains et pertes canoniques de bras de chromosomes.

3.3. TRANSLOCATIONS

La majorité des translocations du myélome impliquent le locus de la chaîne lourde des immunoglobulines (IgH) (González et al., 2007; Barwick et al., 2019) résumées en Figure 12, p. 34. Les translocations IgH dans le myélome apparaissent majoritairement dans les régions de commutation isotypique (CSR), laissant supposer que cet événement de recombinaison anormale intervient au moment de la CSR dans la maturation du lymphocyte B (Walker et al., 2013).

Les translocations IgH t(4;14) entre IgH et *MMSET* (hg38 : *NSD2*, hg19 : *WHSC1*) et t(14;16) entre IgH et *MAF* ont une valeur pronostique pour la survie des patients, et sont par conséquent systématiquement inspectées dans les essais de l'IFM (Avet-Loiseau et al., 2012). La t(4;14) peut être équilibrée, auquel cas la recombinaison entraîne une surexpression de deux oncogènes *FGFR3* et *NSD2* de part et d'autre du point de cassure en chr4. Chez un quart des patients, la translocation n'est pas équilibrée et seul *NSD2* est co-localisé avec les enhancers d'IgH (Keats et al., 2003). Des transcrits de fusion IgH-MMSET sont détectés dans la majorité des cas (Cleynen et al., 2017) et il semble que la protéine de fusion conserve la fonction de méthylation d'histone de *NSD2* (Marango et al., 2008). L'impact de l'expression de *NSD2* sur l'organisation tridimensionnelle de la chromatine n'est pas connu à ce jour.

À ce propos, l'implication du CSR dans la translocation IgH des myélomes n'est pas systématique ; en particulier, les t(11;14) et t(14;20) présentent 21 et 25% respectivement de points de cassures dans la zone de recombinaison D-J (Walker et al., 2013).

Plus récemment, l'équipe d'Emory University à Atlanta a décrit très précisément les translocations des myélomes de patients inclus dans la cohorte CoMMPASS. Barwick et collègues confirment ici que le locus IgH est transloqué dans 41% des patients nouvellement diagnostiqués, presque toujours à une fréquence allélique de 50% lorsque la recombinaison opère avec *CCND1*, *CCND3*, *MAF*, *MAFB* et *NSD2*, cela suggère une présence clonale de la translocation.

Les auteurs confirment également que 16% des patients présentent une t(11;14) IgH-*CCND1* au diagnostic, 11% une t(4 ;14) IgH-*NSD2*, 3.3% une t(14 ;16) IgH-*MAF*, 1.1% une t(6 ;14) IgH-*CCND3*, et 1.0% une t(14 ;20) IgH-*MAFB*, en accord avec la littérature (Manier et al., 2017a; Barwick et al., 2019) et reproduits en Figure 12. Une étude longitudinale du groupe francophone décrit des cas t(4;14) IgH-MMSET sous-clonales perdues au cours du traitement (Hébraud et al., 2013), qui restent minoritaires.

Étonnamment, les translocations IgH dans le myélome ne s'en tiennent pas aux figures quasi-clonales que l'on connaît et une autre classe de translocations IgH apparaît à des fréquences alléliques plus limitées (i.e. à un stade plus tardif de la pathogénèse) : les translocations t(8;14)/IgH-MYC. Ces translocations n'impliquent pas le CSR généralement, sont accompagnées d'amplifications focales génomiques, et sont réparties sur deux régions du chr8 : l'une centrée sur *MYC*, et la seconde autour de *PVT1*, télomérique de *MYC*. Les translocations de *MYC* s'accompagnent d'une surexpression de l'oncogène et d'amplifications focales aux points de cassure (Barwick et al., 2019).

MYC lui-même ne s'en tient pas à la translocation avec IgH. Les t(8;14) ne représentent que 16% des translocations de *MYC* dans le myélome et sont comparables en fréquence aux translocations IgL-MYC (Avet-Loiseau et al., 2001). L'étude récente de Barwick et collègues montre que *MYC* a de multiples partenaires de translocations, au-delà des classiques IgH et IgL, et fait appel plus généralement à un mécanisme de capture de région régulatrice (*enhancer hijacking*) dans des super-enhancers du plasmocyte.

En somme, la t(IgH) semble jouer généralement le rôle de « driver translocation » précoce de même qu'on parle couramment de « driver mutation » clonale, alors que les t(IgL) et/ou t(MYC) semblent offrir un enhancer supplémentaire pour la progression.

3.4. GAINS ET PERTES DE COPIES GÉNOMIQUES

Nous l'avons vu précédemment, les myélomes sont communément séparés en deux catégories : celle des t(IgH), pour environ 50% des cas au diagnostic, et celle des hyperdiploïdies (HRD) pour environ 50% des cas également. Les HRD sont caractérisés par de nombreux gains de chromosomes entiers pour un caryotype d'au moins 48 chromosomes représentés en Figure 12, p. 34. Les deux catégories sont quasiment exclusives l'une de l'autre.

Les chromosomes préférentiellement dupliqués sont impairs parmi 3, 5, 7, 9, 11, 15, 19, et 21. Les mécanismes menant à l'hyperdiploïdie sont inconnus, bien que l'hypothèse d'un événement catastrophique unique soit évoqué dans les leucémies aiguës lymphoblastiques (Onodera et al., 1992). En pratique, la clonalité des mutations acquises vient affaiblir la théorie de la mitose anormale unique. Par exemple, dans une cohorte de myélomes indolents du DFCI, les trisomies 3, 19 et 21 peuvent être sous-clonales, mais les trisomies 5, 11, 15 sont toujours clonales (Bustoros et al., 2020).

De même, par reconstruction chronologique des acquisitions des anomalies génétiques grâce à la fréquence allélique des mutations, Maura et collègues prédisent que le gain21 est plus tardif que gain11 ou gain5 (Maura et al., 2019b). Alors, deux questions demeurent : quelles sont les avantages conférés par chacun des chromosomes acquis ? Comment tirer parti de ces gains de chromosomes et exploiter des vulnérabilités avec des molécules thérapeutiques ?

La recherche dans ce domaine est plutôt influencée par le score prédictif de survie de chacune des anomalies (Avet-Loiseau et al., 2009). Trois anomalies majeures impliquant des bras de chromosomes sont couramment utilisées ou ont été utilisées à cet égard : la del17p, la del13 et le gain1q, qui sont donc supplémentaires à la t(IgH) et à l'HRD.

Le gain 1q est présent chez 22% des patients SMM, et entre 29% et 40% des NDMM, ce qui l'associe à la progression clinique (Walker et al., 2018; Bustoros et al., 2020). La valeur pronostique dans la survie est établie aussi par le groupe francophone (Avet-Loiseau et al., 2009). On mesure historiquement le gain1q par FISH sur *CKS1B*, dont le gain de copie est associé une hausse de l'expression (Fonseca et al., 2006). Néanmoins, la région communément gagnée est plus large, si bien que c'est plutôt le caractère pronostique que le rôle biologique joué par *CKS1B* qui justifie l'exploration de l'anomalie dans le myélome.

La del17p mesurée par FISH *TP53* est de mauvais pronostic pour la survie, elle est rare dans les MGUS, atteint 10% des patients MM, et augmente à la rechute. La délétion est souvent sous-clonale et le pourcentage pronostique de noyaux del(17p) varie selon les études (Avet-Loiseau et al., 2016). On associe la del17p à dérégulation de p53, acteur clé de la réparation de l'ADN, de l'apoptose et du cycle cellulaire. Del17p et mutations *TP53* sont fortement associées (Lodé et al., 2010). La désactivation biallélique de *TP53* est critique et définit (avec les patients chr1q≥4n+ISS3) un groupe de myélome dits « Double-Hit » à très haut risque avec les thérapies actuelles (Walker et al., 2019). Pour l'IFM, le double-hit complète mais ne remplace pas la valeur pronostique de la del17p seule en l'absence de mutation d'exon de *TP53* (Corre et al., 2020).

La monosomie du chromosome 13, ou del13, est présente dans environ 50% des patients au stade indolent comme au diagnostic (Walker et al., 2018; Bustoros et al., 2020). Le gène driver consensuel historique du chromosome 13 est le gène du rétinoblastome Rb (*RB1*) dont la délétion réactive le cycle cellulaire.

Notons que *BRCA2* et *DIS3* sont aussi présents sur le 13q. On associait historiquement la del13q à un mauvais pronostic, mais sa forte association avec d'autres facteurs l'a mise en retrait des diagnostics au profit de la t(4;14) plus couramment évaluée : 85% des patients t(4;14) présentent une del13q (Avet-Loiseau et al., 2007).

En somme, les anomalies des bras des chromosomes ont un impact pronostique fort sur la survie et partiellement indépendant des traitements. La recherche systématique des mutations n'a pas encore trouvé sa place dans la pratique clinique. Le rôle biologique des larges anomalies n'est que partiellement résolu, et se heurte à la fois à la taille des anomalies, et à la faible finesse de détection par FISH.

3.5. CHROMOTHRIPSIES, CHROMOPLEXIES, ET ÉVÉNEMENTS COMPLEXES

Les variants structuraux complexes ne sont pas antagonistes de larges événements « simples » capturés par FISH ou caryotypage. *MYC* présente par exemple de nombreux réarrangements complexes, et 75% des translocations de *MYC* résultent d'insertion template et entraînent un *hijacking* d'enhancer et/ou des amplifications focales (Hultcrantz et al., 2020).

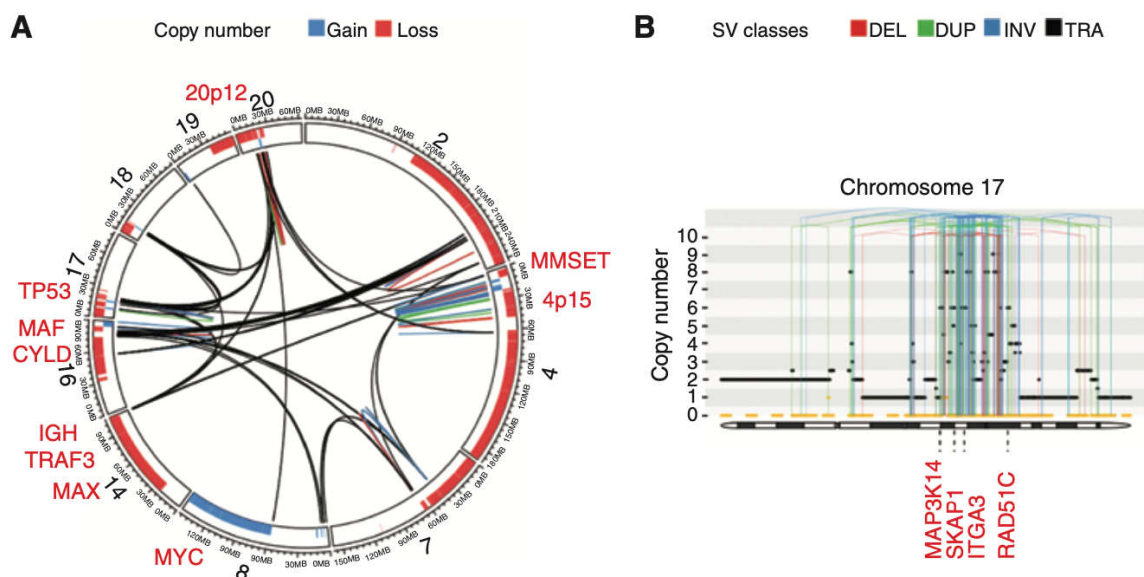


Figure 13 – Chromothripsie des chromosomes 2, 4, 3, 8, 14, 16, 17, 18, 19 et 20. Variant structural décrit par Rustad et al., 2020b. A : Circos plot décrivant les réarrangements structuraux, gains et pertes de copies. B : Chromothripsie vue depuis le chr17.

Les anomalies larges et complexes incluent la chromothripsie, un phénomène catastrophique de réarrangement de dizaines voire centaines de loci enchevêtrés a priori aléatoirement sur un ou quelques chromosomes. On la caractérise aussi par une oscillation rapprochée du nombre de copies (souvent 1 à 4) des loci impliqués dans le réarrangement, cette caractéristique est identifiable visuellement à partir de données de puces à ADN (Stephens et al., 2011; Korbel and Campbell, 2013).

La chromothripsie est décrite dans de nombreux cancers, et a été découverte en 2011 dans le myélome (Magrangeas et al., 2011). Les oscillations rapprochées de pertes et gains génomiques caractéristiques de la chromothripsie sont retrouvées dans 1.3% (n=10/764) des patients au diagnostic dans l'étude de l'IFM (Magrangeas et al., 2011). Il est probable que cet événement soit un marqueur fort d'agressivité : parmi les 10 cas identifiés dans cette première étude, 5 rechutent dans les 10 mois, parmi eux 4 décèdent, dont 3 en moins de 12 mois.

Plus récemment, l'étude de la cohorte CoMMPASS montre une prévalence plus élevée des chromothripsies au diagnostic (18 fois plus) illustrée par un exemple en Figure 13. 24% des patients présentent cet événement associé fortement (OR=6) avec une désactivation bi-allélique de *TP53*. La présence de chromothripsie est statistiquement associée avec l'OS et la PFS. La détection de chromoplexie, inversion, duplication, délétion, insertion ne sont pas associées à la survie mais ces événements sont décrits en détail par le groupe de Maura à New York (Rustad et al., 2020b). *MYC* et sa région régulatrice proximale forment le hotspot le plus courant du myélome qui subit des réarrangements complexes associés à des amplifications focales : insertion template, translocation, délétion menant à la juxtaposition de *MYC* avec un super-enhancer et à sa surexpression (Rustad et al., 2020b).

3.6. CHROMOSOMES MINUSCULES DOUBLES

MYC a fait l'objet d'une découverte supplémentaire dans le myélome : son implication dans un « chromosome minuscule double », ou « double-minute chromosome » abrégé dmin (Yamamoto et al., 2020). Les dmin sont des ADN circulaires dépourvus de centromère présents dans de nombreux cancers et découverts dans les années 1970 avec la systématisation du caryotypage de lignées cellulaires. Ils sont majoritairement porteurs d'oncogènes dont ils favorisent l'amplification et la surexpression (Hahn, 1993). Il est probable que des dmin aient été rapportés comme étant des amplifications intra-chromosomiques, auxquelles les dmin sont souvent associés. Pour des raisons techniques, car ils sont compliqués à assembler par séquençage, les dmin ont fait l'objet de peu de travaux à l'exclusion de MDM2, l'E3 ligase de p53, dont l'acronyme signifie Murine Double-Minute 2 (Fakharzadeh et al., 1991, 1993).

Dans le myélome multiple, en dehors de rares études de cas, on recense 2 publications qui annotent les dmin dans des cohortes cliniques : 1 dmin sur 24 patients (4%) dans l'étude du groupe de Cambridge (Chen et al., 1986) et 5 dmin sur 36 patients (14%) dans l'étude du groupe de Thiruvananthapuram dix ans plus tard (Ankathil et al., 1995). Les dmin identifiés par le groupe indien impliquent les chromosomes 7 (N=2), 8 (N=1) et 9 (N=2).

Nous avons vu que le myélome est caractérisé par des anomalies du génome très diverses, variables entre les patients. Les mutations et événements structuraux impactent l'oncogenèse à diverses échelles de l'ADN (linéaire et spatiales, du nucléotide au bras de chromosome) et à divers stades de la pathologie. Beaucoup d'événements structuraux ont un caractère prédictif de la sévérité de la maladie, à tel point qu'ils définissent des groupes de patients à risque. Pourtant, peu d'aspects fonctionnels liés aux anomalies génomiques ont été décrits. Dès lors, peut-on dire que la génomique du cancer dévoile des mécanismes suffisants à l'oncogenèse, en somme que le cancer est une maladie du génome ? Ou bien la génomique nécessite-t-elle une couche supplémentaire, fonctionnelle, liée à la transcription du génome anormal, pour mieux comprendre l'oncogenèse et la progression du myélome ?

4. ENJEUX PRONOSTIQUES ET HÉTÉROGÉNÉITÉ DE LA TRANSCRIPTION

4.1. CLASSIFICATION FONCTIONNELLE DES PATIENTS ET HÉTÉROGÉNÉITÉ DES MYÉLOMES

Contrairement à la génomique, la transcriptomique, c'est-à-dire l'étude des produits de la transcription du génome, décrit un événement majeur et unifiant de l'oncogenèse du myélome : la dérégulation précoce d'une cycline D (*CCND1*, *CCND2* ou *CCND3*). Logiquement, Bergsagel et Kuehl, qui ont découvert cette propriété avec les puces Affymetrix, proposent de l'associer aux HRD et t(IgH) pour définir arbitrairement la classification TC : Translocations et Cycline D (Bergsagel et al., 2005).

Puis, le groupe de l'UAMS propose d'utiliser les outils d'apprentissage statistique SAM et PAM développés par Tibshirani et Hastie à Stanford pour classer de façon non-supervisée les patients MM regroupés en cohortes d'entraînement et de test, et extraire les caractéristiques des 7 groupes résultants (Zhan et al., 2006). Classiquement, la normalisation des données préliminaires se fait par transformation logarithmique puis centrage avec ou sans réduction et conservation des gènes les plus variables dans la cohorte.

Ainsi à quelques exceptions près, le groupe **MS** correspond aux MMSET/t(4;14), avec (75%) ou sans (25%) expression de *FGFR3*.

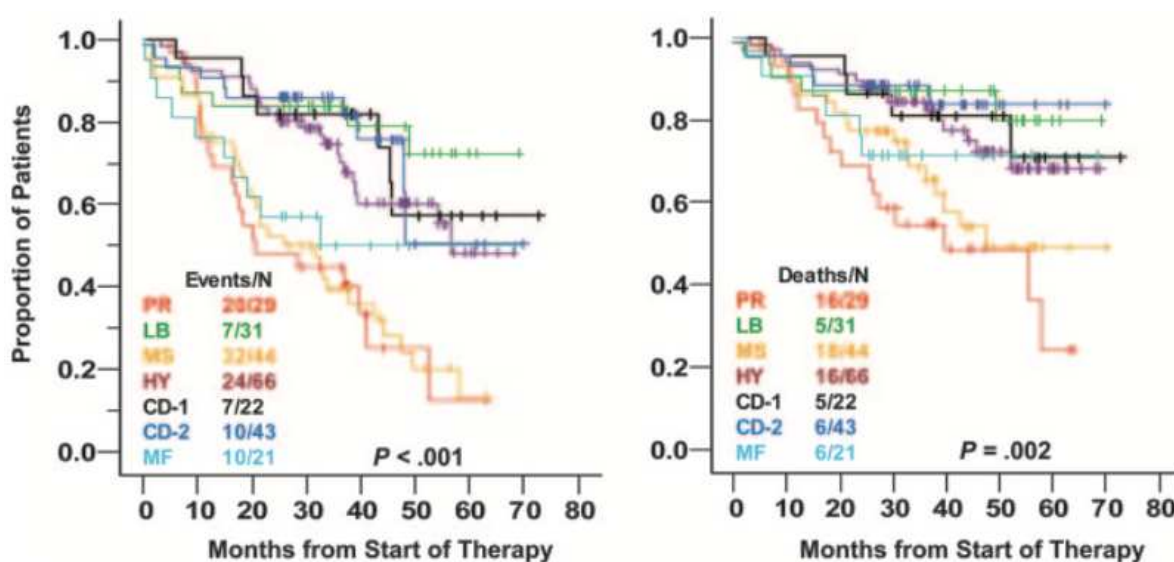
Le groupe **MF** regroupe les translocations de MAF et MAFB, ce qui suggère que le groupe MF permet de trouver les cibles des facteurs de transcription à l'expression forte. On y trouve aussi *CX3CR1*, *ITGB7*.

Les groupes **CD-1** et **CD-2** sont deux clusters de patients avec expression de *CCND1* ou *CCND3* majoritairement transloqués (Broyl et al., 2010). Les myélomes CD-2 expriment plus fortement *PAX5*, *MS4A1* (CD20), *CD79A*, et *VPREB3*, marqueurs de lymphocytes B y compris immatures. Le groupe CD-1 présente un équilibre entre patients t(11;14) et HRD, alors que le groupe CD-2 présente presque systématiquement la translocation de *CCND1* (Broyl et al. 2010, table S4) associée au phénotype B.

Le groupe **HY** est majoritairement composé de patients hyperdiploïdes (90%) et exprime fortement *TNFSF10* (TRAIL), associé à la mort des érythroblastes (Silvestris et al., 2002) et *DDIT3* associé à la maladie osseuse (Tian et al., 2003).

Tableau 7 - Matrice de confusion des classifications de l'UAMS et de HOVON reproduite depuis (Broyl et al., 2010). Le Taux d'erreur de classification (Taux Err. Cl.) est indiqué pour les groupes moléculaires identifiés par Zhan et collègues ainsi qu'une signature NFκB identifiée par Broyl et collègues.

Classe UAMS (Zhan et al., 2006) et signature NFκB										
Classe HOVON (Broyl et al., 2010)		CD-1	CD-2	MS	MF	HY	PR	LB	NFκB	Taux Err. Cl.
	CD-1	8	0	0	0	1	0	4	0	0.38
	CD-2	1	28	0	0	1	0	4	0	0.18
	MS	0	0	31	0	0	1	0	1	0.06
	MF	0	0	0	17	0	0	0	0	0.00
	HY	1	0	0	2	63	4	6	1	0.18
	PR	2	0	0	2	2	9	0	0	0.40
	LB	0	0	0	2	0	0	13	0	0.13
	NFκB	0	8	0	1	1	1	0	26	0.30
	CTA	0	5	0	0	5	8	0	4	
	PRL3	0	0	0	0	0	4	3	2	
	Myeloid	4	0	7	12	4	5	1	6	
	NA	3	0	0	0	1	0	5	0	



À l'inverse, le groupe **LB** (low bone disease) est retenu pour sous-représenter les patients à maladie osseuse à l'IRM (Imagerie par résonance magnétique) et sous-exprimer *DKK1*.

Le groupe **PR** quant à lui est caractérisé par une surexpression de gènes de la prolifération qu'on retrouve partiellement dans les classifications à haut risque clinique.

Cette classification fonctionnelle fait encore référence mais 27% des patients de l'UAMS étaient retirés de l'étude car présentant des caractéristiques myéloïdes contaminantes et/ou de plasmocytes normaux (Zhan et al., 2006). Plus tard, le groupe d'Erasmus MC propose une nouvelle classification qui réintègre les « Myéloïdes » et ajoute trois groupes : CTA, PRL3 et NFκB (Broyl et al., 2010) détaillée dans le Tableau 7. La considération du groupe myéloïde en tant que classe fonctionnelle ou comme contamination de l'environnement fait l'objet d'une controverse.

Le groupe **CTA** est un sous-groupe de PR pour Zhan et collègues qui exprime des antigènes de cancers du testicule : *MAGEA3/A6F/A12*, *PAGE1*, *GAGE12F*. Broyl montre qu'un groupe CTA existe indépendamment du groupe PR.

Le groupe **PRL3** surexprime *PRP4A3* (*PRL3*), mais n'est pas associé à des caractéristiques biologiques précises en dehors d'une sur-représentation des myélomes ISS 1 suggérant un niveau de sévérité moindre de la maladie.

Le groupe **NFκB** surexprime des gènes impliqués dans la voie de signalisation éponyme (*BCL10*, *TNFAIP3*, *CXCL8*, *GADD45B*) ainsi que *CD40*, impliqué dans l'activation de la voie de signalisation du même nom.

Ainsi les sous-groupes CD-1, CD-2, LB et HY représentent des groupes de faible risque, alors que les groupes MF, MS et PR sont plutôt des MM de haut risque (OS et PFS), et cette classification est indépendante d'autres prédicteurs clinico-biologiques : cytogénétique, B2M (Microglobuline-Bêta-2) et LDH (Lactate Deshydrogénase) (Figure 14).

En plus de l'aspect pronostique, l'intérêt majeur des groupes d'expression génique est d'associer une couche fonctionnelle aux translocations majeures d'IgH. La fonction cellulaire est déduite d'observations cliniques et des fonctions des gènes les plus différentiellement exprimés entre les classes, c'est-à-dire qui définissent le programme transcriptomique de la classe.

4.2. PRÉDICTION D'ÉVÉNEMENTS ET SIGNATURE DE HAUT-RISQUE

La nature des données collectées par puce ou séquençage de l'ARN permet également une approche supervisée des problèmes de classification. Les groupes de l'UAMS, d'Erasmus MC et l'IFM ont proposé trois signatures géniques pour identifier les myélomes à haut risque de rechute et/ou de décès (Shaughnessy et al., 2007; Decaux et al., 2008; Kuiper et al., 2012).

Le développement d'une signature d'expression génique se déroule en deux étapes : la première consiste à extraire une liste de gènes corrélés en expression avec la variable d'intérêt à apprendre (par exemple survie ou rechute avec un modèle de Cox univarié). La seconde étape consiste à combiner les gènes en une grandeur d'une seule dimension, par exemple le premier axe d'une analyse en composantes principales (ACP). Le test de la capacité du classifieur à séparer correctement les groupes vient valider la démarche mise en place.

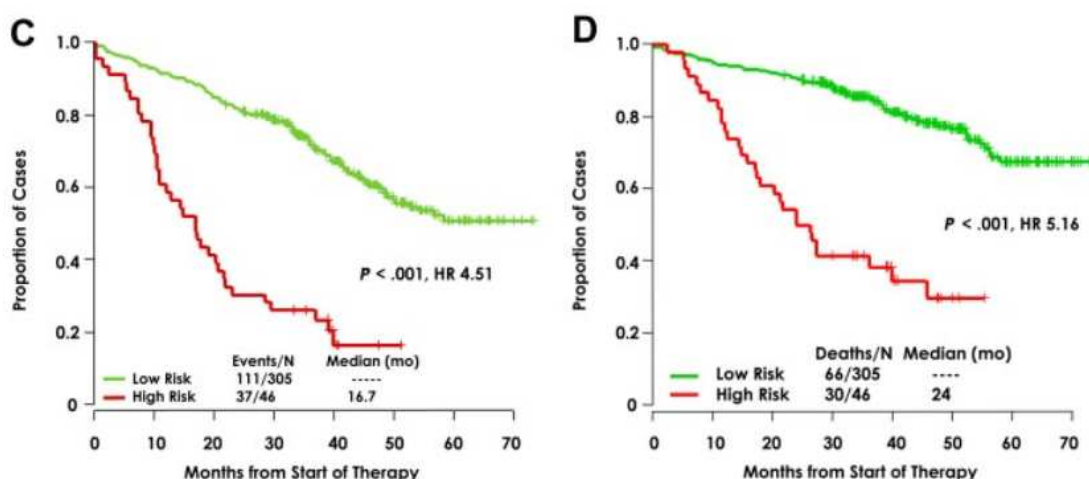


Figure 15 - Aspects pronostiques de la classification à haut risque GEP70. Survie sans progression (C) et survie globale (D) de la cohorte de l'UAMS avec signature GEP70. Figure issue de Shaughnessy et al., 2007.

Les efforts importants déployés à la définition de signatures de risque apportent des informations limitées sur la biologie du MM. Par exemple, le bras long du chromosome 1 concentre statistiquement plus de gènes liés au myélome à haut risque, en ligne avec les prédictions permises par la FISH et le nombre de copies, mais les gènes ne sont pas ou peu communs entre les signatures.

L'expression de la survivine (*BIRC5*) fait figure d'exception : *BIRC5* intervient dans les définitions de EMC-92, UAMS-17, UAMS-70 et GPI-50 et fait l'objet d'un intérêt particulier pour le développement de molécules antagonistes (Romagnoli et al., 2007; Altieri, 2013).

En pratique, les signatures moléculaires ou de haut-risque n'ont pas trouvé d'application en clinique malgré leur caractère pronostique indépendant statistiquement des autres variables clinico-biologiques mesurées au diagnostic. La puissance statistique des signatures est forte (Hazard Ratio compris entre 2 et 5 pour EMC-92 en fonction de la cohorte et du traitement, entre 4 et 5 pour GEP70 représentée en Figure 15). Cette puissance statistique est supérieure à celle de la del(17p) seule, de la beta-2-microglobuline et de l'albumine (Kuiper et al., 2012). Dans une étude postérieure, Kuiper et collègues confirment l'indépendance de l'EMC92 et de l'ISS ainsi que leur caractère additif pour prédire la survie (Kuiper et al., 2015).

L'introduction de ces signatures dans la pratique clinique fait encore débat aujourd'hui, en particulier l'indépendance entre signatures et R-ISS reste à déterminer, et plusieurs aspects techniques ne sont pas résolus : qualité de la transposition des signatures d'expression géniques entre laboratoires, technologies, et protocoles cliniques.

4.3. LE TRANSCRIPTOME AU-DELÀ DES GÈNES CODANTS

LES ARN LONGS NON-CODANTS

Parmi les ARN non-codants, l'étude des ARN longs non-codants (lncRNA) a bénéficié de l'essor du séquençage haut-débit de bibliothèques d'ARN poly-adenylés. On considère que les ARN d'au moins 200 nucléotides de longueur et à potentiel de traduction faible ou inexistant forment la classe des lncRNA. La majorité des gènes humains sont des lncRNA et certains ont des fonctions de régulation de la chromatine, de la transcription en *cis* et des modifications post-transcriptionnelles (Iyer et al., 2015). L'analyse des lncRNA est d'abord motivée par la présence « d'éléments ultra-conservés » pour environ 1% d'entre eux, et dans le cancer en particulier pour leur dérégulation courante (Iyer et al., 2015).

La valeur pronostique pour la survie et la rechute d'une signature d'expression composée seulement de lncRNA est prouvée (Samur et al., 2018). La signature proposée par Samur et collègues a un caractère pronostique indépendant de la MRD, de l'ISS et de la FISH. Ainsi l'expression des lncRNA a un caractère pronostic pour la survie indépendamment des ARN codant pour des protéines. En revanche, les effets de dosage liés aux gains et pertes de bras de chromosomes sur l'expression des lncRNA restent mal connus. Les technologies de séquençage couplé ADN et ARN dans la même cellule, ou la prédiction de nombre de copies à partir de données de transcription pourraient éclairer les conséquences du nombre de copies sur l'expression génique.

Plus spécifiquement, les lncRNA *PVT1*, *MALAT1* et *CRNDE* ont fait l'objet de travaux récents dans le myélome pour leur rôle dans la biologie de la tumeur.

L'expression du lncRNA *PVT1* proximal de *MYC* est requise pour sur-exprimer *MYC*, elle stabilise la protéine *MYC* et confère un avantage prolifératif dans les tumeurs *MYC*-amplifiées (Tseng et al., 2014). Les translocations de *MYC* dans le MM entraînent une surexpression à la fois de *MYC* et de *PVT1* susceptible d'engendrer une boucle de rétroaction positive entre les deux gènes (Mikulasova et al., 2020).

MALAT1 a aussi fait l'objet de nombreux travaux suite à la découverte du caractère pronostique de son expression dans le cancer du poumon et peut être considéré comme une cible thérapeutique potentielle dans le myélome (Amodio et al., 2018). Le ciblage spécifique de *MALAT1* entraîne une diminution de l'expression de *NRF1* et *NRF2*, deux facteurs de transcription qui contrôlent et activent les gènes du

protéasome. Indirectement, le ciblage de *MALAT1* par oligonucléotides anti-sens entraîne l'accumulation de protéines polyubiquitinées et la mort des cellules de myélome sur un mode d'action proche des inhibiteurs du protéasome (Amodio et al., 2018).

La fonction pro-croissance de *CRNDE* sur les cellules de myélome via l'augmentation de l'IL6R a été découverte par David et collègues (David et al., 2020). *CRNDE* fait partie des lncRNA surexprimés dans le myélome par rapport aux plasmocytes normaux, son expression forte est également associée à un mauvais pronostic clinique (David et al., 2020). La déplétion de *CRNDE* affecte de nombreux mécanismes tumoraux du myélome dont l'expression du récepteur à l'IL6 et conduit à une diminution de la survie, de la prolifération *in vitro* et de la maladie extramédullaire dans un modèle murin (David et al., 2020).

LES MICRO-ARN

Les microARN (miRNA) sont des ARN non-codants d'environ 21 bases qui régulent l'expression génique de façon post-transcriptionnelle. Les miRNA dérivent de gènes spécifiques (*host genes*) ou d'introns excisés d'ARNm et subissent une maturation médiée par Drosha et Dicer. Leur fonction biologique canonique est l'inhibition de la traduction d'ARNm par appariement partiel de leur séquence en 3' des régions non-traduites (3'-UTR) d'ARNm ciblés via le complexe miRNA-induced silencing complex (miRISC) (Krol et al., 2010).

L'axe LIN28/let-7 est peut-être le mécanisme de régulation des miRNA le plus étudié en laboratoire. Let-7 est une famille de miRNAs suppresseurs de tumeur qui ciblent *MYC* (Sampson et al., 2007) et *RAS* (Johnson et al., 2005). Les protéines Lin28 ciblent spécifiquement la famille Let-7, permettent la dégradation des miRNA Let-7 (Heo et al., 2008) et restaurent les mécanismes oncogéniques ciblés par les miRNA. Dans le myélome, Manier et collègues montrent que cet axe est fonctionnel et pertinent pour un ciblage thérapeutique, en particulier dans les tumeurs exprimant fortement *MYC* (Manier et al., 2017b). De plus, l'inactivation de *DIS3* dans le MM induit une augmentation directe de *LIN28B* et par la suite une augmentation de *MYC* et *RAS*, cibles de Let-7 (Segalla et al., 2015).

Les niveaux de miRNA Let-7b et miR-18a dans les exosomes de sérum de patient contribuent à la stratification du risque de rechute et de survie en analyse multivariée (Manier et al., 2017c), ce qui en fait un biomarqueur non-invasif pour évaluer la sévérité de la maladie et montre l'intérêt pronostique de l'étude des ARN non-codants.

4.4. DU SÉQUENÇAGE DE L'ARN « BULK » AU « SINGLE-CELL RNA-SEQ »

Dans l'étude des populations cellulaires qui composent le vivant, le RNA-seq a pris une place fondamentale ces dernières années, en ce qu'il permet notamment de quantifier le niveau d'expression des gènes à un niveau de résolution difficilement envisageable jusqu'au début des années 2000. Certes l'unité moléculaire fonctionnelle protéique n'est pas atteinte, mais le RNA-seq peut s'effectuer à l'échelle du génome, c'est-à-dire sans sélection a priori des gènes d'intérêt.

Le RNA-seq présente une limitation majeure : puisque le niveau de résolution est le tissu, il faut pouvoir isoler a priori la population d'intérêt, d'une part, et accepter que la mesure en moyenne sur l'échantillon masque l'hétérogénéité des sous-populations et des états cellulaires s'ils existent d'autre part. Or, le myélome multiple est une tumeur génétiquement hétérogène à l'échelle du patient et la mesure de l'expression en moyenne est susceptible de masquer la fonction moléculaire propre aux sous-clones.

Avec la publication de techniques de séquençage d'ARN en cellule-unique (scRNA-seq) (Tang et al., 2009; Macosko et al., 2015), et surtout la première solution commerciale à haut débit 10X Chromium (Zheng et al., 2017), l'accès à la technologie de séquençage unicellulaire se démocratise, et le scRNA-seq est élu « 2018 Breakthrough of the Year » par la revue Science pour son apport à la compréhension du développement embryonnaire.

Dans le MM, deux axes de recherche sont développés et ont fait l'objet de travaux majeurs : une description de l'hétérogénéité des cellules tumorales (Ledergor et al., 2018) que nous détaillons ci-après, et une cartographie de l'environnement immunitaire à divers stades de la maladie (Zavidij et al., 2020).

4.5. SCRNA-SEQ DES PLASMOCYTES TUMORAUX

Ledergor et collègues analysent l'hétérogénéité des plasmocytes tumoraux de 29 patients à divers stades de développement du MM ainsi que de 11 prélèvements de donneurs contrôles (Ledergor et al., 2018). La technologie utilisée (MARS-seq) ne fait pas appel à des émulsions microfluidiques mais isole les cellules les unes des autres dans des plaques 384 puits avec un cytomètre trieur sur la base de marqueurs phénotypiques (Jaitin et al., 2014). Cette méthode permet de quantifier des protéines membranaires en combinaison avec le transcriptome, ce qui ne sera permis plus tard en émulsion qu'avec le CITE-seq et le REAP-seq (Peterson et al., 2017; Stoeckius et al., 2017).

D'abord, Ledergor et collègues confirment l'établissement de programmes transcriptionnels malins propres aux patients atteints de myélome, alors que les plasmocytes normaux partagent leur programme transcriptionnel entre les patients (Figure 16). En d'autres termes, les cellules de chaque patient MM clusterisent entre elles et établissent une signature spécifique, alors que les plasmocytes normaux des différents patients clusterisent ensemble.

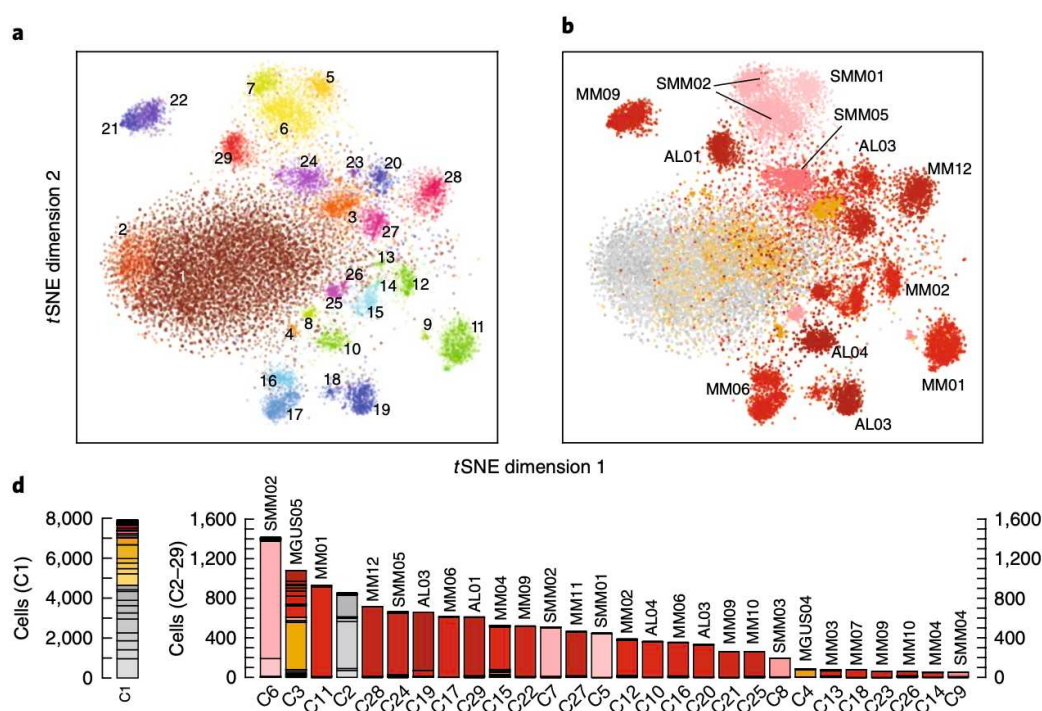


Figure 16 - Hétérogénéité transcriptomique inter- et intra-patient. Figure issue de Ledergor et al., 2018. Le panel A) montre un résultat de clustering non-supervisé de single-cell RNA sequencing qui recoupe les patients annotés en panels B) et D). À l'inverse, les donneurs sains (gris) sont regroupés ensemble par le clustering non-supervisé et par l'algorithme de réduction de dimension.

Ensuite, en utilisant un algorithme de prédiction de nombre de copies par moyennage de l'expression relative sur une fenêtre glissante de 100 gènes le long des bras de chromosomes, les auteurs sont capables de prédire l'état du nombre de copie par cellule (nombre de copies normal, diminué ou augmenté) et de l'associer au transcriptome scRNA. Un algorithme similaire est inclus dans le projet Trinity CTAT du Broad Institute (inferCNV). Le résultat principal de Ledergor et collègues est qu'il n'y a pas de lien entre les gènes dérégulés entre les sous-populations et le nombre de copies génomiques de ces gènes. D'une part, ce résultat pourrait suggérer une action en *trans* des anomalies de nombre de copies, et d'autre part, cela montre que localement, nombre de copies et programmes transcriptionnels peuvent être indépendants.

Enfin, les auteurs offrent des perspectives cliniques au single-cell RNA sequencing avec (1) la caractérisation des CTCs comparées au clone tumoral de la moelle osseuse ; (2) le suivi des cellules de la MRD sur le même principe. Ici, Ledergor et collègues montrent qu'il n'y a pas ou peu de différence transcriptomique entre les cellules des échantillons biologiques de CTC, de la MRD, et des clones d'origine de la ponction de moelle osseuse. Cette découverte ouvre la voie au suivi de précision de l'évolution tumorale en clinique avec le scRNA-seq dont des résultats préliminaires ont été présentés à l'ASH 2019 (Cohen et al., 2019).

5. CARTOGRAPHIER L'HYDROXYMÉTHYLATION DE L'ADN

5.1. MÉTHYLATION DE L'ADN GÉNOMIQUE

La méthylation de l'ADN est la marque épigénétique par excellence : elle modifie l'ADN sans muter le résidu cytosine qu'elle marque, peut modifier l'expression génique et est transmise automatiquement à la mitose. Ainsi une cytosine (C) devient 5-méthylcytosine (5mC) dans un contexte di-nucléotidique CpG chez les eucaryotes (Song et al., 2011b; Schübeler, 2015).

La fonction majeure de la méthylation de l'ADN est répressive, et se manifeste sous différentes formes stables : répression de l'expression génique, inactivation du X surnuméraire et de séquences répétées, gènes soumis à empreinte parentale (Jones, 2012).

Chez les vertébrés, l'ADN génomique est globalement méthylé sauf dans des régions riches en CpG appelés îlots CpG (CGI pour *CpG Island*) où la méthylation est quasi-absente. Les CGI sont caractérisés par une forte densité locale en CpG : longueur supérieure à 200bp, contenu en GC supérieur à 50% et ratio de CpG observé sur attendu supérieur à 0.6 (Gardiner-Garden and Frommer, 1987).

On associe les CGI aux gènes : les CGI sont typiquement situés à proximité du site d'initiation de la transcription (TSS) et sont associés à 70% des promoteurs chez l'homme. L'absence de méthylation d'un CGI est considérée comme une condition nécessaire mais pas suffisante à l'activité transcriptionnelle (Fan and Zhang, 2009).

Au cours du développement et de la différenciation cellulaire, la méthylation *de novo* est apposée par les méthyltransférases à ADN DNMT3A et DNMT3B (Okano et al., 1999) en combinaison avec leur partenaire non-enzymatique DNMT3L (Bourc'his et al., 2001).

Le maintien de la méthylation est assuré durant la réplication de l'ADN par DNMT1 associé à UHRF1 qui détecte les CpG hémi-méthylés (Bostick et al., 2007; Song et al., 2011b). DNMT1 est indispensable au renouvellement de tissus (Sen et al., 2010) y compris des cellules souches embryonnaires chez l'homme (Liao et al., 2015). Le niveau global de 5mC a tendance à diminuer avec le vieillissement ce qui en fait un prédicteur de l'âge (Goel et al., 2017).

5.2. DYNAMIQUE DE LA MÉTHYLATION DE L'ADN DANS LE MYÉLOME MULTIPLE

Logiquement, la notion de mémoire épigénétique motive l'étude de la méthylation de l'ADN dans les processus d'oncogenèse. Le cancer est globalement hypométhylé par rapport aux tissus sains, sans doute du fait du maintien erratique des profils de méthylation provoquant une déméthylation globale (Feinberg and Vogelstein, 1983).

L'hypométhylation aux CpG peut mener à la réactivation de gènes normalement éteints, en particulier de gènes favorisant la prolifération et le cycle cellulaire et de certains gènes d'expression testiculaire (Feinberg and Tycko, 2004) qui sont retrouvés dans les classifications moléculaires du myélome multiple (Zhan et al., 2006; Broyl et al., 2010).

La déméthylation des régions satellites centromériques et péri-centromériques, en particulier sur le chromosome 1, est associée à une instabilité chromosomique et pourrait favoriser les réarrangements catastrophiques oncogéniques y compris dans le MM (Sawyer et al., 1998, 2019).

À l'opposé, l'hyperméthylation ciblée de gènes suppresseurs de tumeurs est observée dans de nombreux cancers. Les gènes *RB1* et *CDKN2A* ont fait l'objet de nombreux travaux depuis les années 1980 y compris dans le myélome. Fonctionnellement, l'hyperméthylation serait associée à l'extinction de l'expression des gènes suppresseurs de tumeur (Sakai et al., 1991; Ng et al., 1997).

Les travaux sur la méthylation dans le myélome ont confirmé une hypométhylation globale du génome du myélome par rapport aux plasmocytes normaux (Salhia et al., 2010; Walker et al., 2011; Heuck et al., 2013). Pour autant, certains patients présentent une hyperméthylation globale importante (Agirre et al., 2015), caractéristique de l'hétérogénéité du myélome multiple. Surtout, les travaux d'Agirre et collègues ont montré que l'hyperméthylation locale dans le myélome est concentrée sur des enhancers introniques spécifiques du lymphocyte B et des cellules souches (Agirre et al., 2015), une caractéristique particulière du myélome en opposition aux tumeurs localement hyperméthylées sur les îlots CpG (Hansen et al., 2011; Berman et al., 2012).

Ces travaux sont à relire aujourd'hui à la lumière de la découverte de Weinberg et collègues, qui ont montré que l'apposition de l'H3K36me2 par NSD1 (paralogue de NSD2) est requise pour recruter DNMT3A et maintenir la méthylation de l'ADN intergénique (Weinberg et al., 2019).

Lorsque NSD2 est perdu, DNMT3A est redistribuée sur la marque H3K36me3 présente aux corps de gènes provoquant la baisse du niveau de méthylation de l'ADN intergénique (Weinberg et al., 2019). Le rôle de la re-distribution des marques H3K36me2/3 associé à la méthylation de l'ADN chez les patients t(4;14) sera probablement étudié à l'aune de ces nouveaux résultats et d'inhibiteurs spécifiques de l'interaction entre NSD2_PWWP1 et de l'H3K36me2 (Freitas et al., 2020), mais reste inconnu pour l'instant.

5.3. ASPECTS FONCTIONNELS DE LA MÉTHYLATION DE L'ADN

Les protéines de la famille MBD (methyl-CpG binding domain) MDB1, MDB2, MDB4 et MeCP2 ont une forte affinité pour les CpG méthylés et se lient préférentiellement à l'ADN dense en CpG. Elles pourraient ainsi contribuer à la fonction répressive de la méthylation de l'ADN (Defossez and Stancheva, 2011; Baubec et al., 2013).

In vitro, Yin et collègues ont montré dans un criblage SELEX de 542 DNA Binding Domains de facteurs de transcription que la majorité sont sensibles à l'état des méthylation des CpG (Yin et al., 2017). En particulier, les bHLH, bZIP et ETS sont inhibés par la méthylation, alors que les homéodomains, POU et NFAT sont plus affins pour les CpG méthylés (Yin et al., 2017).

Domcke, Bardet et collègues montrent néanmoins *in vivo* par enrichissement des sites hypersensibles à la Dnase I dans des cellules souches murines triple KO *Dnmt1/3a/3b* que la plupart des motifs de facteurs de transcription ne sont pas sensibles à la méthylation de l'ADN (Domcke et al., 2015). Le motif de NRF1 fait figure d'exception dans ce modèle, et en l'absence de méthylation de l'ADN, NRF1 se lie à de nombreux nouveaux loci où il favorise la transcription. L'induction de l'activité de méthyltransférase *de novo* rétablit la méthylation de l'ADN, NRF1 est redistribué sur le génome au bénéfice de la méthylation de l'ADN et fait apparaître un nouveau modèle : celui d'une compétition entre la méthylation de l'ADN et l'activité de facteurs de transcription (Domcke et al., 2015).

En somme, les travaux actuels sur la méthylation de l'ADN et les facteurs de transcription ou l'affinité de l'un pour l'autre sont prometteurs dans la recherche sur la différenciation cellulaire et le développement embryonnaire, et s'orientent vers un rôle déterminant de la régulation entre facteurs de transcription et les acteurs de la méthylation de l'ADN.

5.4. HYDROXYMÉTHYLATION ET DÉMÉTHYLATION DE L'ADN GÉNOMIQUE

On l'a vu, la déméthylation de l'ADN du génome est une marque du vieillissement physiologique comme de l'oncogenèse. Malgré leur caractère crucial dans la recherche médicale, les mécanismes responsables de la déméthylation globale de l'ADN ne sont pas complètement compris.

On distingue communément la déméthylation passive de la déméthylation active : la première est liée à une reproduction imparfaite des profils de méthylation par DNMT1 lors de la réplication de l'ADN, la seconde implique des enzymes dont nous détaillons l'état des connaissances ici.

En 2009, deux articles dans *Science* (re)découvrent simultanément l'existence d'une nouvelle modification de l'ADN dans le génome : la 5-hydroxyméthylcytosine (5hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Dans les années 1970, la 5hmC avait été découverte et réfutée (Penn et al., 1972; Kothari and Shankar, 1976), ce qui explique probablement qu'elle soit restée oubliée 30 ans durant.

Tahiliani et collègues identifient que la famille TET (Ten-Eleven Translocation) peut modifier les 5mC. Ils montrent successivement que TET1 catalyse la conversion de la 5mC en 5hmC dans les leucémies à fusion MLL-TET1, et que la répression de *TET1* par transfection de petit ARN interférent (siRNA) entraîne une réduction de la 5hmC (Tahiliani et al., 2009). TET1 et TET3 possèdent un domaine de liaison directe à l'ADN alors que TET2 n'en possède pas et lie l'ADN via son partenaire CXXC4/IDAX, ces deux gènes étant issus de la cassure d'un gène *TET* ancestral (Ko et al., 2013).

Le mécanisme majeur de la déméthylation de l'ADN (Figure 17) est l'oxydation successive par la famille TET de la 5mC en 5hmC, puis en 5-formylcytosine (5fC), enfin en 5-carboxylcytosine (5caC), elle-même clivée par la glycosylase TDG et prise en charge par le système Base-Excision Repair (BER) pour être remplacée par une cytosine complémentaire de la guanine située sur l'autre brin (Shen et al., 2013).

Kriaucionis et Heintz découvraient quant à eux que la 5hmC est présente de manière physiologique dans des cellules neuronales, mais ils ne la détectaient pas dans des lignées cellulaires cancéreuses (Kriaucionis and Heintz, 2009). Nous retrouvons ce résultat qui limite l'étude de la 5hmC *in vitro* sur la lignée cellulaire de myélome MM.1S par l'absence de signal en séquençage 5hmC-Seal.

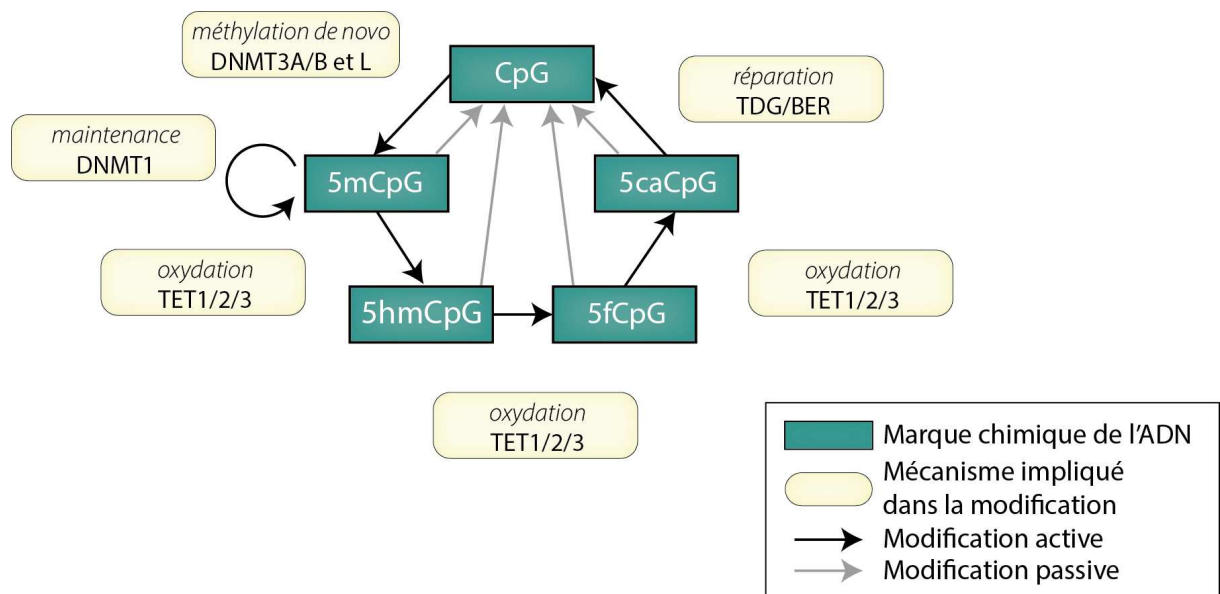


Figure 17 - Marques chimiques aux CpG et déméthylation de l'ADN. TDG : Thymine-ADN-glycosylase ; BER : Réparation par excision de base. DNMT : Méthyltransférase à ADN ; TET : Ten-eleven translocation méthylcytosine dioxygénase.

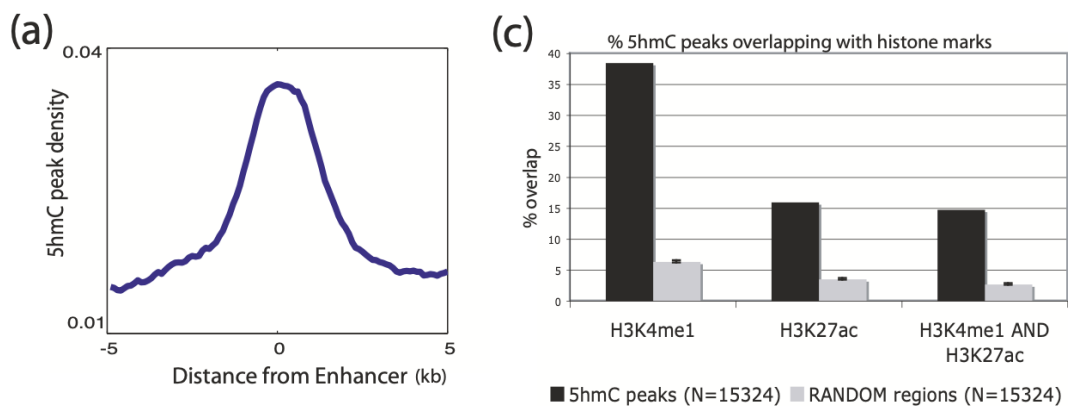


Figure 18 - Co-localisation de la 5hmC avec les marques de chromatine active. Figure issue de Stroud 2011. La 5hmC est colocalisée avec les enhancers (a) et avec les marques H3K4me1 H3K27ac (c) de cellules souches embryonnaires humaines.

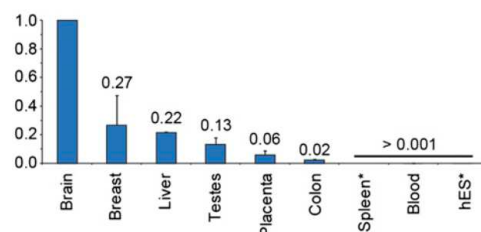


Figure 19 - Niveau de 5hmC exprimé en valeur relative par rapport à la 5hmC cérébrale. Figure issue de Nestor et al., 2012.

5.5. LOCALISATION ET NIVEAUX DE LA 5HMC

En plus de son rôle d'intermédiaire dans la déméthylation de l'ADN, l'hydroxyméthylation est également considérée comme une marque stable de l'ADN (Bachman et al., 2014). Elle est localisée préférentiellement dans les enhancers et dans les corps de gènes transcrits (Stroud et al., 2011; Sérandour et al., 2012; Mahé et al., 2017) (Figure 18).

Les niveaux globaux de 5hmC, contrairement aux niveaux de 5mC, sont très variables d'un tissu à l'autre, le cerveau en présente par exemple en moyenne 20 fois plus que le sang périphérique (Nestor et al., 2012) et (Figure 19).

La 5hmC a d'abord été étudiée dans les cancers hématologiques en lien avec les mutations perte-de-fonction affectant les protéines TET. En fait, la 5hmC est globalement perdue dans une grande variété de cancers solides et hématologiques à l'exception notable des leucémies réarrangées MLL-TET1 qui ont permis sa découverte (Tahiliani et al., 2009; Huang et al., 2013; Huang and Rao, 2014). La diminution des niveaux de 5hmC atteint un facteur 30 dans les tumeurs cérébrales (Jin et al., 2011).

Contrairement à la méthylation de l'ADN qui est globalement perdue dans des régions pauvres en gènes dans le cancer par rapport aux cellules normales, la 5hmC y est perdue dans les régions géniques, autour du promoteur, dans les exons, introns et les régions intergéniques (Lian et al., 2012; Jeschke et al., 2016).

5.6. DÉPENDANCE AUX PROTÉINES TET ET CANCER

Les protéines des familles DNMT et TET sont cruciales dans la différenciation cellulaire hématopoïétique.

Les souris KO pour *Tet1* sont viables et fertiles (Dawlaty et al., 2011) mais présentent des lymphomes B (Cimmino et al., 2015). Le promoteur de TET1 est hyperméthylé et le gène est transcriptionnellement éteint dans les lymphomes NHL, ce qui montre sa fonction de gène suppresseur de tumeur pour ces cancers hématologiques (Cimmino et al., 2015).

Les souris KO pour *Tet2* sont également viables et fertiles mais développent des leucémies chroniques (Li et al., 2011). *TET2* est régulièrement muté « perte-de-fonction » dans les leucémies myéloïdes aiguës où la présence d'une mutation de *TET2* est associée à l'hyperméthylation d'enhancers et l'extinction de gènes suppresseurs de tumeur (Rasmussen et al., 2015), et à une baisse globale des niveaux de 5hmC (Ko et al., 2015).

DNMT3A/B et *TET2* sont fréquemment mutés dans l'hématopoïèse clonale, les leucémies et les lymphomes (Challen and Goodell, 2020). L'inactivation de *TET2* chez l'homme et la souris induit un biais de différenciation hématopoïétique en faveur des monocytes, alors que l'inactivation de *DNMT3A* favorise la différenciation érythrocytaire. Le biais de différenciation serait dû à la sensibilité relative des facteurs de transcription impliqués dans l'hématopoïèse pour les CpG méthylés ou non-méthylés (Izzo et al., 2020; López-Moyado and Rao, 2020).

Tet3 n'est pas exprimé dans les cellules souches embryonnaires mais participe à la déméthylation du génome paternel dans le pronucleus, fortement hydroxyméthylé (Gu et al., 2011).

En résumé, *TET1* et *TET2* sont dispensables pour le développement embryonnaire mais leur suppression constitutive ou somatique cause de multiples tumeurs hématologiques.

5.7. VERS UN RÔLE FONCTIONNEL DE LA 5HMC ?

L'hydroxyméthylation est une marque (1) chimiquement stable, (2) dont les niveaux globaux dépendent du tissu, et (3) dont la localisation génomique est associée aux enhancers et régions activement transcrites. Son rôle fonctionnel propre fait débat.

Certaines protéines ont une affinité spécifique pour la 5hmC : UHRF2 est l'exemple le plus connu (Spruijt et al., 2013; Zhou et al., 2014). En outre, Khund-Sayeed et collègues montrent *in vitro* que sur certains motifs de facteurs de transcription, la 5hmC augmente l'affinité de TCF4 et USF1 pour l'ADN par rapport à la 5mC et à la cytosine nue. TCF4 et USF1 sont des membres de la famille bHLH, inhibés par la méthylation (Khund-Sayeed et al., 2016). Ces travaux suggèrent des rôles fonctionnels potentiels de la 5hmC, mais la rareté de la 5hmC et son contrôle difficile *in vitro* freinent l'évaluation de ces propriétés.

Au contraire, *in vivo*, les *DNMT3* et *TET* sont en compétition pour la méthylation de l'ADN : l'ADN reste méthylé malgré la présence conjointe des *DNMT* et *TET* dans un tiers des enhancers de cellules souches, où la combinaison joue en faveur du maintien de la méthylation (Charlton et al., 2020). La co-localisation des régulateurs de la méthylation et de la déméthylation permettrait un switch rapide entre les deux modifications de l'ADN, mais les travaux ne suggèrent pas de rôle fonctionnel à la marque 5hmC en elle-même (Charlton et al., 2020).

5.8. CARTOGRAPHIE DE L'HYDROXYMÉTHYLATION DANS LE MYÉLOME MULTIPLE

Les équipes de T. Fest et G. Salbert ont choisi d'aborder l'hydroxyméthylation dans le myélome multiple par l'axe de l'instabilité péricentromérique du chr1q12 et le gain de sa région proximale immédiate chr1q21. Ils ont montré par cartographie de l'hydroxyméthylome que la 5hmC est préférentiellement présente, chez les patients, dans les super-enhancers de la lignée cellulaire MM.1S (Chatonnet et al., 2020). En cherchant les CpG hydroxyméthylés dans le MM mais qui ne le sont pas dans les plasmablastes différenciés *in vitro*, les auteurs identifient une sur-représentation de CpG sur le chr1q21 associés à des gènes fortement exprimés dans le MM et de mauvais pronostic pour la survie. En particulier, l'analyse de la 5hmC a permis au groupe de Rennes de découvrir l'association entre la famille des 4 gènes du *FAM72* et la prolifération du MM, qui ouvre une voie supplémentaire dans l'étude du rôle biologique du gain du chr1q (Chatonnet et al., 2020).

Dans un travail concomitant et indépendant, nous proposons une étude de la cartographie de l'hydroxyméthylation dans la partie résultats de ce manuscrit (Alberge et al., 2020a).

5.9. SÉQUENÇAGE DE LA MÉTHYLATION ET DE L'HYDROXYMÉTHYLATION DE L'ADN

Les techniques de détection de la méthylation des CpG les plus courantes font appel à un traitement au bisulfite de sodium (par exemple le WGBS pour *whole-genome bisulfite sequencing*, ou encore les puces 450K chez Illumina). Le traitement en question modifie les cytosines en uraciles, lues comme des thymines par le séquençage Illumina (Figure 20). Les 5mC sont protégées du traitement au bisulfite de sodium, et sont bien lues comme des cytosines par séquençage Illumina. En comparant l'ADN traité à un échantillon non-traité ou à un génome de référence, on peut mesurer le taux de méthylation des CpG dans le génome.

En réalité, les marques 5mC et 5hmC sont toutes deux protégées du traitement au bisulfite de sodium (Huang et al., 2010 et Figure 20, panel A). Dit autrement, les données de WGBS ne permettent pas de faire la distinction entre 5mC et 5hmC. En général dans la littérature, la distinction est escamotée, probablement par soucis de simplicité car le taux de 5hmC est souvent négligeable devant la 5mC, par soucis de non-contradiction car les deux marques sont associées à des régions régulatrices

antagonistes, et pour des raisons historiques puisque la 5hmC est découverte tardivement.

L'*oxidative bisulfite* (oxBS), un protocole modifié du WGBS pour oxyder d'abord les 5hmC en 5fC (sensibles au bisulfite) permet de séquencer les 5mC spécifiquement (Figure 20, B et C). La comparaison de l'oxBS et du WGBS permet finalement d'identifier spécifiquement les 5hmC (Booth et al., 2012).

Une autre approche de séquençage des 5hmC au bisulfite, le TAB-seq, consiste à glycosyler spécifiquement les 5hmC pour les protéger de l'oxydation de l'ADN, avant oxydation et traitement au bisulfite (Yu et al., 2012).

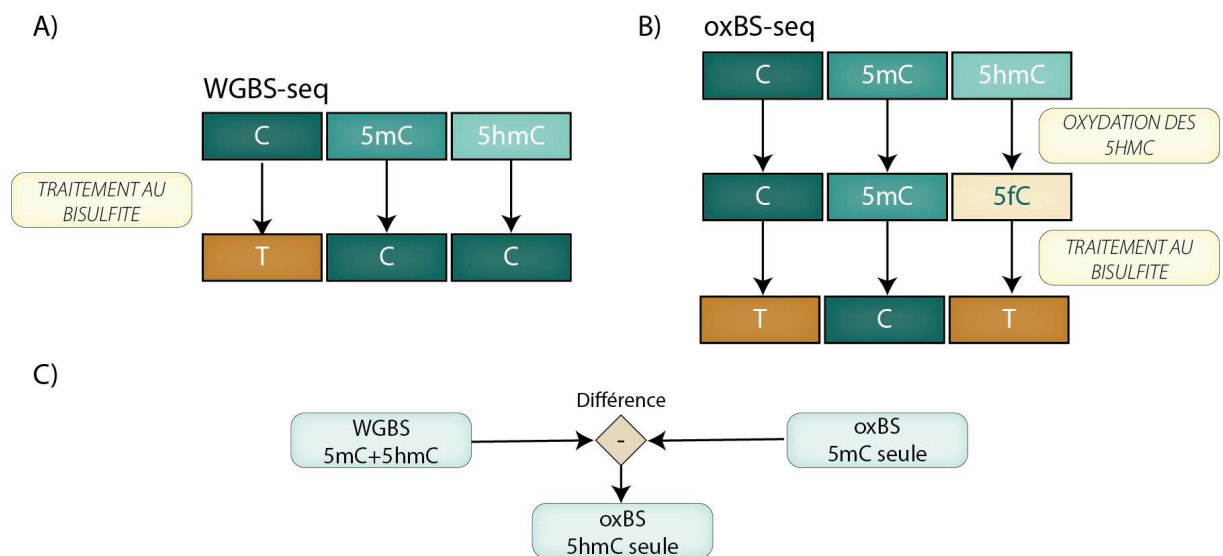


Figure 20 - Techniques de quantification de la méthylation et de l'hydroxyméthylation au bisulfite. Quantification des 5mC et 5hmC (A) ou 5mC seules (B) par séquençage avec traitement au bisulfite de sodium, avec ou sans oxydation des 5hmC en 5fC (formylcytosine). La comparaison entre WGBS et oxBS fournit un profil de la 5hmC (C). WGBS : Whole-genome bisulfite sequencing ; oxBS : oxidative bisulfite sequencing.

PROBLÉMATISATION ET OBJETS DE LA THÈSE

L'hétérogénéité moléculaire des plasmocytes tumoraux est caractérisée par une diversité importante des anomalies somatiques du génome inter- et intra-patient, aux niveaux génétiques, épigénétiques, et transcriptionnels. La diversité inter-patient se manifeste notamment par la multiplicité des événements oncogéniques « drivers » clonaux. Les mutations et anomalies de nombre de copies sont presque exclusivement sous-clonales. Elles reflètent l'hétérogénéité restreinte à l'endroit de la ponction de moelle, dont les techniques de séquençage unicellulaire commencent à révéler l'ampleur. L'imagerie apporte des informations sur la spatialisation de la tumeur et son étendue, qui pourraient ne pas être totalement indépendantes des programmes transcriptionnels malins. Enfin, l'étude de la diversité épigénétique reste à ce jour aussi confrontée à l'absence d'événement unificateur du myélome. Ces considérations nous invitent à nous demander : comment génome, transcriptome, et épigénome sont-ils reliés dans l'oncogenèse du myélome ? Les modifications épigénétiques reflètent-elles seulement le transcriptome, le génome, et la complexité de la maladie d'un patient à l'autre ? Ou bien peut-on en apprendre plus sur la pathologie elle-même ? En quoi les nouvelles techniques de séquençage peuvent-elles caractériser plus profondément l'hétérogénéité multi-niveaux ? Au-delà de la compréhension de la maladie, les interventions thérapeutiques peuvent être affectées par ces processus. Comment l'hétérogénéité est-elle perturbée par la pression de sélection des thérapies ciblées ? Connaissant cette hétérogénéité, quels nouveaux outils et nouveaux biomarqueurs peuvent être « transférés vers la clinique », comment, dans quelle mesure et pour quels usages ? Nous déclinons ces questions à travers trois axes.

Premièrement, nous savons que les mutations somatiques ne suffisent pas à expliquer l'oncogenèse, et que les modifications épigénétiques peuvent jouer un rôle important dans ce processus. Pourtant, peu de travaux ont caractérisé l'état de la chromatine, des modifications épigénétiques et des marques d'ADN dans le myélome. Nous décrivons la localisation sur le génome et la quantité de la marque de l'ADN récemment découverte 5-hydroxyméthylcytosine (5hmC), dans les sous-groupes moléculaires du myélome issus de l'IFM/DFCI 2009 et en comparaison de plasmocytes normaux. Partant des niveaux globaux de 5hmC, nous en déduisons des caractéristiques sur la maladie (survie, avancement, sévérité, etc.). Puis, connaissant les sous-groupes moléculaires de patients, nous tenterons d'en apprendre plus sur la répartition locale de la 5hmC restante. En particulier, nous proposerons de découvrir

de nouveaux enhanceurs sur la base de corrélations entre 5hmC locale et niveau d'expression des gènes proximaux.

Deuxièmement, les études de séquençage de l'ADN ont caractérisé en profondeur l'état du génome du myélome au diagnostic mais les anomalies acquises et sélectionnées par les traitements restent majoritairement inconnues. Or, dans le cadre de thérapies ciblées, les anomalies acquises peuvent dévoiler des biomarqueurs précoces de résistance et des nouvelles voies empruntées par la tumeur pour proliférer. Nous exploitons l'exemple du venetoclax, une thérapie ciblée anti-BCL2, pour caractériser de manière longitudinale l'hétérogénéité moléculaire des anomalies sélectionnées chez des patients réfractaires. Grâce aux méthodes de séquençage unicellulaire de nombre de copies, de transcriptome et d'accessibilité de la chromatine au sein d'un même prélèvement biologique, nous verrons comment la famille du BCL2 concentre les anomalies génomiques en question et quels essais *in vitro* viennent les corroborer.

Enfin, les signatures d'expression génique et l'imagerie médicale fournissent tous deux des biomarqueurs pronostiques pour la survie dans le myélome multiple. D'une part, les biomarqueurs d'imagerie par TEP-FDG reflètent l'avidité de la tumeur pour le glucose, laquelle est susceptible de dépendre des groupes fonctionnels et des anomalies génomiques. D'autre part, un groupe de patients encore mal caractérisé présente un profil normal au diagnostic en imagerie, et il est probable que le transcriptome puisse nous éclairer sur cette question pour affiner le caractère pronostique de la TEP. Pourtant, les associations statistiques entre biomarqueurs d'imagerie, d'expression génique et autres variables clinico-biologiques restent inconnues ou inexplicables. Ici, nous proposons d'utiliser les données de RNA-seq, de TEP-FDG, et de survie de la cohorte CASSIOPEIA pour aborder ces questions : quels myélomes restent « TEP-négatifs » malgré un diagnostic clinique ? Comment interagissent les variables d'imagerie et d'expression génique ? Les classifications à haut risque des deux techniques sont-elles indépendantes ? Le cas échéant, comment se complètent-elles ?

À travers ces trois axes : hydroxyméthylation de l'ADN, anomalies du BCL2 acquises par une thérapie ciblée, et intérêt pronostique de la génomique associé à l'imagerie, nous tentons d'éclairer notre compréhension du génome du myélome, de son hétérogénéité, de son évolution sous pression de sélection, et des implications cliniques potentielles de ces connaissances.

ÉVOLUTION ET SÉLECTION CLONALES : MODÈLES ET ENJEUX DANS LE MYÉLOME MULTIPLE

L'introduction de ce manuscrit et la problématisation de l'étude du génome du myélome multiple ont fait l'objet d'un résumé sous la forme d'une courte publication dans les *Correspondances en Onco-Hématologie* pour son dossier spécial « Myélome » (Alberge et Minvielle, 2020).

Évolution et sélection clonales : modèles et enjeux dans le myélome multiple

Clonal evolution and selection in multiple myeloma

J.B. Alberge*, S. Minvielle*

RÉSUMÉ

Dogme majeur de la biologie du cancer, l'évolution clonale de cellules tumorales en compétition pour proliférer est aujourd'hui révélée en détail par les techniques modernes de séquençage du génome. Dans l'étude du myélome multiple, la recherche de mutations et l'hétérogénéité clonale peinent à trouver leurs applications cliniques et la connaissance de la complexité de la maladie reste difficile à appréhender. Dans cet article, nous clarifions, à travers quelques exemples, comment la génomique des sous-clones tumoraux aide à quantifier le risque de progression dès les stades indolents de la maladie, d'une part, et nous chercherons à comprendre la sélection de clones résistants au traitement, d'autre part. L'étude approfondie du génome du cancer bouleverse nos connaissances concernant l'hétérogénéité intrapatient et fournit en même temps des modèles concrets d'évolution tumorale et de résistance acquise au traitement.

Mots-clés : Génomique – Myélome multiple – Évolution clonale – Sélection clonale.

SUMMARY

Clonal evolution of competitive tumor cell populations is a central dogma of cancer biology, on which high-throughput sequencing has shed light in recent years. In multiple myeloma, the discovery of mutations and clonal heterogeneity struggle to find a way towards clinical application, and the disease complexity remains hard to appreciate. Through examples, we clarify in this review how the analysis of tumor clones disclosed independent risks of progression at the indolent stages of the disease and acquired resistance to treatment. In short, studying the cancer genome provides insights in intrapatient heterogeneity and also unveils tangible examples of tumor evolution and treatment resistance.

Keywords : Genomics – Multiple myeloma – Clonal evolution – Clonal selection.

En 1976, Peter Nowell introduit dans un célèbre article, *The clonal evolution of tumor cell populations* [1], le dogme fondateur de l'étude de la biologie du cancer : les tumeurs ont pour origine une cellule unique, qui acquiert et accumule les mutations dans un processus de sélection des clones les plus compétitifs, c'est-à-dire les plus agressifs. En 2011, le premier séquençage complet d'un génome de myélome multiple (MM) vient compléter cette théorie et lui adjoindre une méthode systématique d'étude des anomalies génomiques sous-clonales, c'est-à-dire présentes seulement dans une fraction réduite de la tumeur [2]. À ce propos, la sémantique du clone génétique est commune à tous les cancers et sans lien avec la notion d'immunoglobuline monoclonale utilisée en hématologie. L'hétérogénéité intrapatient était certes constatée dans le MM, la perte partielle du bras court du chromosome 17 (del(17p)) en est un exemple à valeur pronostique. Désormais, le séquençage à haut débit

révèle la complexité génomique des cellules tumorales de façon globale et sans a priori, mais les applications cliniques peinent encore à émerger. La génomique du MM ouvre en fait 2 domaines d'étude, dont nous interrogeons ici les enjeux. D'une part, quelles sont les interactions entre anomalies génomiques et progression du MM à partir des stades asymptomatiques de la pathologie ? D'autre part, quelles sont les conséquences de l'hétérogénéité du MM sur la sélection induite par le traitement ?

Évolution clonale du myélome multiple

Altérations primaires

Les altérations chromosomiques du MM sont indissociables de la différenciation lymphocytaire B à l'origine des plasmocytes à longue durée de vie. Le réarrangement VDJ des gènes des immunoglobulines,

* Centre de recherche en cancérologie et immunologie Nantes-Angers (CRCINA), université de Nantes, Inserm, CNRS, CHU de Nantes.

DOSSIER

Tableau. Reproduction, d'après les données de B.A. Walker et al. [7], de 1 273 génomes séquencés ; 19 gènes présentent des mutations non synonymes chez plus de 3 % des patients. La clonalité est exprimée en moyenne de la fraction de cellules cancéreuses porteuses de la mutation.

Gène	Perte (%, n = 1 074)	Mutation (%, n = 1 273)	Clonalité moyenne (%)	Fonction
KRAS	–	22	58	MEK/ERK
NRAS	–	17	66	MEK/ERK
DIS3	40	10	63	Exonucléase
FAM46C	15	9	55	Nucléotidyltransférase
BRAF	–	8	66	MEK/ERK
TP53	9	6	68	Suppresseur de tumeur
HUWE1	–	5	59	Ubiquitine ligase
TRAF3	15	5	61	NF-κB
ATM	2	4	54	Sérine/thréonine kinase
EGR1	–	4	89	Régulateur transcriptionnel
HIST1H1E	–	4	82	Épigénétique
DUSP2	–	4	89	MEK/ERK
FGFR3	7	4	46	MEK/ERK
PRKD2	–	4	68	Sérine/thréonine kinase
UBR5	–	4	37	Ubiquitine ligase
CYLD	20	3	43	NF-κB
MAX	–	3	67	Facteur de transcription
IRF4	–	3	59	Développement B
ZNF292	–	3	49	Facteur de transcription

translocation du gène de leur chaîne lourde (*IGH*) avec un oncogène, considérée comme événement primaire. Cette translocation est ainsi à l'origine du MM, et donc clonale, présente dans toutes les cellules. De manière quasiment exclusive, l'autre moitié des patients présente une hyperdiploïdie de chromosomes impairs [4]. Cette première classification unifiée (figure, partie gauche) pose les bases d'une étude des sous-clones en compétition pour la prolifération au sein d'un même patient.

Altérations secondaires et progression du MM

On a vu que l'hyperdiploïdie et les translocations de *IGH* sont présentes dans la grande majorité des cellules tumorales, y compris au stade indolent (*smoldering multiple myeloma*, SMM) de la maladie. D'autres anomalies, telles que la *del(17p)* ou le gain du bras long du chromosome 1 (gain 1q), sont en revanche sous-clonales. Elles ne sont pas présentes dans toutes les cellules du MM et ne résultent donc probablement pas du même événement génétique que l'altération primaire [5]. De même, le réarrangement de l'oncogène *MYC* (translocation ou nombre de copies) est un événement secondaire qui touche moins de 5 % des patients au stade indolent, mais plus de 20 % au diagnostic. Il est désormais établi que la détection d'une telle altération est un facteur fort et indépendant de risque de progression vers un stade symptomatique de la maladie [5, 6]. Au diagnostic, les translocations de *MYC* sont décelées avec une fréquence de variant allélique moyenne de 26 %, ce qui pose la question de la capacité technique de détection précoce autant que de la répartition locale au sein des lésions.

De manière similaire, l'étude des mutations ponctuelles révèle un tropisme récurrent pour les voies de signalisation MAPK/ERK, NF-κB, la réparation de l'ADN, les modifications de la chromatine et la régulation du cycle cellulaire (tableau) [7], (figure, partie centrale). Par des approches en fréquence et fonctionnelles, on peut déduire de la récurrence de ces mutations leur caractère majeur pour la progression clinique, pour la prolifération (on parle de "driver mutation") et leur rôle biologique oncogénique ou suppresseur de tumeur. Ici non plus, il n'existe pas de mutation commune à tous les patients, et les mutations décelées au diagnostic restent majoritairement sous-clonales. Plus de 75 % des patients présentent une répartition complexe des mutations dans les lésions focales (on parle d'"hétérogénéité spatiale"). La dispersion des mutations observées par L. Rasche et al. [8] soutient le modèle d'une dissémination large du clone primaire, puis d'une évolution génétique des clones secondaires, restreinte localement et différenciée entre les lésions focales.

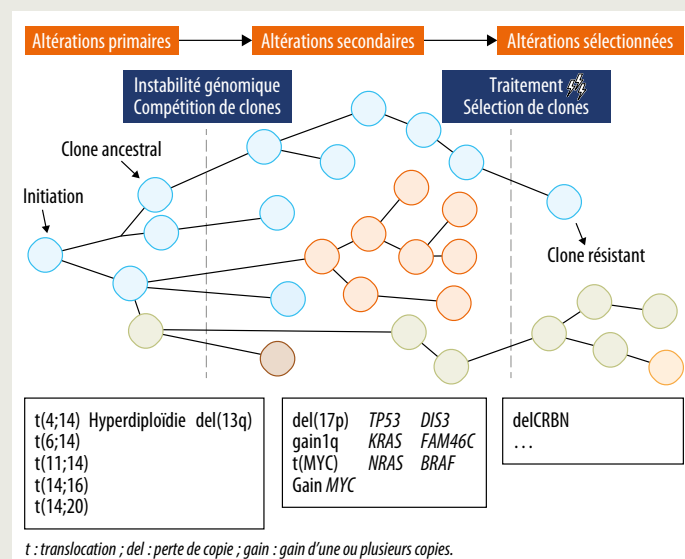


Figure. Modèle d'évolution et de sélection clonales du myélome multiple.

l'hypermutation somatique et la commutation de classe sont autant d'événements physiologiques d'édition du génome susceptibles de résulter en un réarrangement anormal [3]. Environ 50 % des MM présentent une

Sensibilité et résistance au traitement

Implications cliniques des mutations

Nous l'avons vu précédemment, le MM progresse avec des mutations et des réarrangements génomiques qui confèrent un avantage sélectif à un ou plusieurs sous-clones disséminés dans les lésions focales. La ponction de moelle osseuse hors des foyers de prolifération apporte donc une image partielle de la complexité clonale de la maladie. Si le séquençage à haut débit montre que les mutations du MM sont pour la plupart sous-clonales, rien n'empêche de penser que même des mutations clonales peuvent être des faux positifs, et que, en dehors de la crête iliaque, d'autres foyers n'en sont pas porteurs. La mutation actionnable en thérapie ciblée *BRAF V600E* en est peut-être le meilleur exemple récent [8, 9]. En ce sens, on préférera considérer de manière générale que l'analyse des mutations ponctuelles ainsi que des réarrangements structuraux reflète surtout l'instabilité de l'ADN, l'hétérogénéité des clones, leur succès d'adaptation et, en somme, la progression de la maladie.

Résistance acquise aux immunomodulateurs

Qu'en est-il de la sélection des clones avec le traitement ? Les immunomodulateurs (IMiD) tels que le lénalidomide ou le pomalidomide sont largement utilisés dans le traitement du MM et ciblent une protéine appelée céréblon (CRBN). Chez un tiers des patients réfractaires, la résistance acquise aux IMiD altère le gène *CRBN* (figure, partie droite) via de nombreuses voies : mutation ponctuelle, perte de copies, perte d'hétérozygotie, translocations et inversions. Au moins 1 voie non génétique est également empruntée : l'épissage alternatif de l'exon 10 du *CRBN*, qui code pour le domaine de liaison des 2 IMiD. Lorsque l'exon 10 du *CRBN* est épissé, la protéine traduite ne possède pas le domaine récepteur des IMiD, alors que la séquence génétique est bien présente dans l'ADN [10]. La diversité des altérations génomiques montre que l'instabilité du génome est une des clés dans la progression vers la résistance. L'épissage alternatif montre que les altérations non génétiques des cibles thérapeutiques peuvent aussi expliquer la résistance acquise. Une conséquence directe est que chez les patients réfractaires au lénalidomide, l'un et l'autre des mécanismes (altération génétique et épissage alternatif) sont des facteurs prédictifs d'un taux de réponse faible et d'une progression rapide sous pomalidomide [10]. Ici, la croissance d'un clone résistant au lénalidomide via une altération du *CRBN* est logiquement défavorable au pomalidomide. Le suivi et la détection précoce des anomalies

sélectionnées pourraient ainsi permettre d'ajuster les traitements (et donc leur efficacité) et d'améliorer la survie des patients.

Conséquences thérapeutiques sur la sensibilité aux médicaments

Puisque les traitements interagissent avec les sous-clones, il semble utile de connaître la composition spécifique des clones de la maladie et leur évolution. Les mutations clonales des patients résistants sont encore mal connues, et la surveillance des sous-clones en compétition dans le MM peut paraître excessive dans la gestion et la prévention de la résistance au traitement. Pourtant, adapter le traitement en fonction des résistances acquises lors de processus darwiniens est courant dans d'autres domaines, comme le traitement du sida et des infections microbiennes [11]. Dans le MM en particulier, traquer les altérations génétiques par des techniques peu invasives, comme l'analyse des cellules tumorales circulantes (CTC) ou de l'ADN libre circulant (cfDNA), peut être prometteur, à condition que la sensibilité des méthodes choisies soit suffisante [12].

Hétérogénéité au-delà du génome

L'exemple de la résistance aux IMiD par une altération non génétique du *CRBN* nous interpelle sur la nature et l'utilité de la définition des clones : n'est-ce pas restreindre sa vision de l'hétérogénéité tumorale que de s'en tenir aux mutations du génome ? Si la résistance acquise au traitement n'est pas nécessairement génétique, alors il faut questionner plus en profondeur la nature de la clonalité du MM, de l'évolution et de la sélection des sous-clones. Les techniques de séquençage en cellule unique (*single-cell sequencing*) ciblant l'ARN, la méthylation de l'ADN ou les modifications d'histones peuvent nous éclairer sur les mécanismes biologiques de l'hétérogénéité (programmes transcriptionnels, entropie) et leur interaction avec le génome tumoral [13]. Jusqu'où la cellule tumorale peut-elle perdre son identité de plasmocyte à longue durée de vie tout en s'adaptant à son environnement ? Quelles sont les régions génomiques ou spatiales au sein du noyau de la cellule qui ne sont pas réparées, ou qui sont soumises à un contrôle d'intégrité moins fort de la part des systèmes de réparation de l'ADN ou du maintien des marques épigénétiques ? Peut-on relier l'hétérogénéité tumorale à la diversité phénotypique et à l'interaction avec l'environnement ? Au-delà du génome, de nombreux mécanismes biologiques, y compris épigénétiques, favorisent la diversité intra-

DOSSIER

patient et viennent compléter notre définition du clone tumoral. Ce changement modifie, certes, nos modèles de genèse et de progression du MM, mais peut fournir autant de pistes de recherche fécondes pour une meilleure compréhension de la maladie.

Conclusion

L'analyse du séquençage du génome est devenue essentielle à l'étude des cancers et soutient la théorie de l'évolution darwinienne de clones tumoraux en compétition pour proliférer. Le MM en particulier est caractérisé par une forte hétérogénéité des anomalies génétiques acquises et sélectionnées. Malgré cela, l'analyse des stades indolents révèle des biomarqueurs susceptibles de prédire le risque de progression clinique (par exemple, les altérations de *MYC* ou de réparation de

l'ADN). Au diagnostic, les mutations détectées reflètent surtout la progression du cancer et l'hétérogénéité clonale, car la dissémination des clones secondaires est une source importante de diversité des mutations actionnables par des thérapies ciblées. Enfin, la résistance acquise aux nouveaux traitements pourra être suivie et détectée de manière précoce par le séquençage précis d'anomalies sélectionnées connues ou restant à découvrir. Déjà, la recherche dans ce domaine révèle des mécanismes de résistance acquise qui nous invitent à étendre la notion de clone tumoral aux modifications épigénétiques, un domaine d'étude qui fait l'objet de nombreux travaux actuellement. En somme, l'étude approfondie de la génomique et de l'épigénomique des sous-clones de MM permet de mieux comprendre l'évolution et la sélection des cellules, tout en fournissant des modèles et des exemples concrets de progression clinique et de résistance au traitement. ■

J.B. Alberge déclare ne pas avoir de liens d'intérêts en relation avec cet article.
S. Minvielle n'a pas précisé ses éventuels liens d'intérêts.

RÉFÉRENCES

1. Nowell PC. The clonal evolution of tumor cell populations. *Science* 1976;194(4260):23-8.
2. Chapman MA et al. Initial genome sequencing and analysis of multiple myeloma. *Nature* 2011;471(7339):467-72.
3. Walker BA et al. Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells. *Blood* 2013;121(17):3413-9.
4. Manier S et al. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol* 2017;14(2):100-13.
5. Bustoros M et al. Genomic profiling of smoldering multiple myeloma identifies patients at a high risk of disease progression. *J Clin Oncol* 2020;38(21):2380-9.
6. Barwick BG et al. Multiple myeloma immunoglobulin lambda translocations portend poor prognosis. *Nat Commun* 2019;10(1):1911.
7. Walker BA et al. Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood* 2018;132(13):1461.
8. Rasche L et al. Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nat Commun* 2017;8(1):268.
9. Andrulis M et al. Targeting the BRAF V600E mutation in multiple myeloma. *Cancer Discov* 2013;3(8):862-9.
10. Gooding S et al. Multiple cereblon genetic changes associate with acquired resistance to lenalidomide or pomalidomide in multiple myeloma. *Blood* 2020. doi: 10.1182/blood.202007081.
11. Hughes D, Andersson DI. Evolutionary consequences of drug resistance: shared principles across diverse targets and organisms. *Nat Rev Genet* 2015;16(8):459-71.
12. Guo G et al. Genomic discovery and clonal tracking in multiple myeloma by cell-free DNA sequencing. *Leukemia* 2018;32(8):1838-41.
13. Nam AS et al. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat Rev Genet*. doi: 10.1038/s41576-020-0265-5.

Précautions et mode d'emploi...

DROIT À L'IMAGE DANS LA LITTÉRATURE MÉDICALE

Toute personne a droit au **respect de sa vie privée**, qui inclut également le droit à l'image (Code civil, article 9, CEDH, 23 juillet 2009).

Le médecin dont l'intention est d'utiliser les images ou les données d'un patient à des fins d'intérêt public, pour illustrer son propos dans le cadre d'un enseignement ou d'une publication scientifique, doit :

- obtenir l'autorisation explicite du patient, en d'autres termes un consentement écrit, celle des parents s'il s'agit d'un enfant ;
- avoir informé le patient au préalable, de manière claire et précise, de la finalité de l'utilisation, de la durée de cette dernière et de la nature des supports employés ;
- faire en sorte que l'identité du patient ne soit pas divulguée et qu'il ne soit pas reconnaissable.

Le respect de ces étapes permet de se prémunir de toute violation du secret médical tel que régi par le **Code de la santé publique** (art. R. 4127-73, al. 2 ; art. L. 1110-4), le **Code de déontologie médicale** (art. 73) et la **loi informatique et libertés** (art. 66).



Partie 2. RÉSULTATS

La partie *Résultats* de ce manuscrit est organisée en trois chapitres qui reprennent les axes déclinés par la problématisation : (1) la cartographie et la quantification de l'hydroxyméthylation dans le myélome multiple, (2) l'étude de la résistance acquise au venetoclax et des anomalies génomiques sélectionnées de la famille du BCL2 en chapitre 2, page 107, et (3) l'étude d'association entre profils transcriptomiques et biomarqueurs d'imagerie par TEP-FDG en chapitre 3, page 139.

1. CARTOGRAPHIE ET QUANTIFICATION DE L'HYDROXYMÉTHYLATION DANS LE MYÉLOME MULTIPLE

1.1. CONTEXTE ET OBJECTIFS DE L'ÉTUDE

Ce travail publié dans *Clinical Epigenetics* est initié en 2017 avec un financement de l'IFM et de la Ligue Contre le Cancer pour étudier l'hydroxyméthylome des cellules tumorales d'individus de l'IFM/DFCI 2009 (Attal et al., 2017) et sur des plasmocytes normaux issus de patients sans myélome. Le protocole finalement retenu pour la cartographie de la 5hmC est le séquençage par SCL-seq/hmC-Seal/5hmC-seq (Song et al., 2011a).

La méthode de capture de la 5hmC par SCL-seq consiste à enrichir des fragments d'ADN par (1) glycosylation des 5hmC, (2) biotinylation des 5hmC glycosylées et (3) enrichissement avec des billes de streptavidine (Song et al., 2011a). Finalement, l'amplification et le séquençage des fragments enrichis génèrent des données par chimie CLICK qui sont analysées sur le modèle du séquençage avec immunoprécipitation de la chromatine (ChIP-seq). Les données ont été générées par Aurélien Sérandour et Soline Denié, et analysées par Aurélien Sérandour et moi-même. Nous avons cartographié la 5hmC dans les plasmocytes purifiés de la moelle osseuse de 40 patients au diagnostic, dont 4 échantillons appariés à la rechute, et pour 5 moelles normales issues de prélèvements de chirurgie de hanche. Chaque librairie est également appariée avec un « Input » d'ADN non-glycosylé sur le modèle des Input IgG en ChIP-seq.

La distribution de la 5hmC caractérise le tissu cellulaire d'origine (Cui et al., 2020), et sa cartographie à partir de cfDNA est un biomarqueur précoce prometteur pour le diagnostic de nombreux cancers (Song et al., 2017).

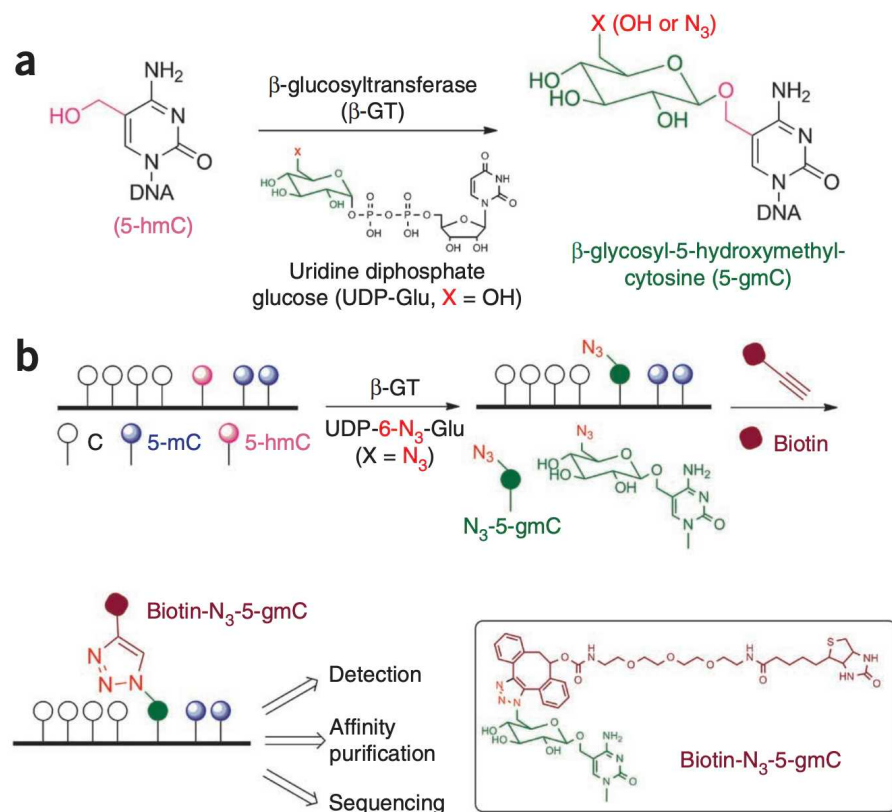


Figure 22 - Protocole SCL-seq pour l'enrichissement de la 5hmC génomique. Figure issue de Song et al., 2011. a) la beta-GT glucosyle spécifiquement les 5hmC avec l'UDP-Glu comme co-facteur. b) Le glucose de l'UDP-Glu est modifié pour contenir un groupe N₃, lequel est ensuite biotinylé. L'UDP-6-N₃-Glu biotinylée est ensuite enrichie avec des billes recouvertes de streptavidine à la forte affinité pour la biotine.

Dans notre étude, nous avons étudié de l'ADN de plasmocytes purifiés de la moelle osseuse, en gardant à l'esprit qu'une adaptation sur cfDNA serait envisagée pour la suite de ce travail. Sur cette base, ces travaux novateurs ont deux ambitions : l'une fondamentale sur la caractérisation de l'hydroxyméthylome d'une tumeur hématologique, le MM, l'autre appliquée en tirant parti des données clinico-biologiques de la cohorte IFM/DFCI 2009 pour découvrir des facteurs cliniques pronostiques et enrichir le diagnostic et le suivi des patients MM.

Dans le cadre de ce travail, nous avons eu l'occasion de collaborer avec le groupe du Pr Carell au Max Planck Institute à Munich pour quantifier de manière globale la 5mC et la 5hmC par spectrométrie de masse (Traube et al., 2019). La quantification du niveau global des bases modifiées dans les mêmes échantillons de l'IFM et de tissus sains pour lesquels nous cartographions la 5hmC a été faite par Mirko Wagner à Munich.

1.2. MÉTHODES BIOINFORMATIQUES

Pour chaque librairie, nous avons ciblé un séquençage single-end à 20M de reads par réplicat en suivant les recommandations du consortium ENCODE pour le ChIP-seq d'histone à « narrow peaks ».

DÉTECTION DE PICS DE 5HMC

Nous avons utilisé Cutadapt v1.18 (Martin, 2011) pour exciser les reliquats d'adaptateurs dans les données de séquençage avant d'aligner les fragments sur le génome de référence GRCh38 avec Bowtie2 v2.2.8 (Langmead and Salzberg, 2012) et les paramètres par défaut. Les reads dupliqués ou avec une faible qualité d'alignement ($Q < 30$) sont retirés.

En ChIP-seq, on utilise communément le « peak calling » pour définir des régions du génome statistiquement enrichies en fragments d'intérêt par rapport à un bruit de fond (*background*) et au séquençage de l'input. L'objectif est de définir des régions porteuses d'une marque d'histone, d'un facteur de transcription ou d'une marque sur l'ADN d'après l'hypothèse que la manipulation technique a sélectionné préférentiellement les fragments d'ADN d'intérêt. MACS est probablement l'algorithme de peak calling le plus utilisé par la communauté bio-informatique avec plus de 8000 citations (Zhang et al., 2008; Feng et al., 2012). MACS modélise d'abord la taille du fragment d pour déplacer les reads en 3' d'une distance $d/2$ vers les sites potentiels de fixation de la cible. Ensuite, MACS détecte les pics potentiels par un test de Poisson du nombre de reads dans une fenêtre de longueur $2d$ par rapport au nombre attendu de reads dans la même fenêtre sur le génome entier. Pour chaque pic potentiel, l'enrichissement observé dans le ChIP est enfin corrigé par rapport à l'input en rapportant le ratio $\text{ChIP} / \max(\text{background}, \text{control_1K}, \text{control_5K}, \text{control_10K})$ avec l'enrichissement contrôle respectivement sur le génome entier, dans l'input à 1 kilo-paire de bases (kbp), 5kbp, et 10kbp.

Cette procédure de peak calling fait consensus mais s'affranchit de deux aspects qui nous intéressent particulièrement en 5hmC dans le MM : Le premier est la normalisation pour le taux de GC, très variable localement et pas forcément capturé avec le même biais pour tous les fragments d'ADN. Le second aspect est que MACS ne fournit ni n'utilise de modélisation explicite du nombre de copies génomiques pour normaliser le signal alors que l'Input-seq peut estimer cette donnée. Ces deux considérations nous ont poussé à utiliser un deuxième algorithme en plus de MACS,

HMCan (Ashoor et al., 2013). HMCan adopte une philosophie proche de MACS en initialisant un réseau de markov caché par le résultat du même test de Poisson comparant nombre de read ChIP sur nombre de read attendu en background. Surtout, HMCan inclut la normalisation du signal pour le taux de GC et pour l'estimation du nombre de copies génomiques par Control-FREEC (Boeva et al., 2012).

Notons que malgré les nombreuses corrections de biais techniques, un ensemble de régions du génome est anormalement enrichi dans la plupart des protocoles NGS et affecte les analyses en aval. ENCODE fournit une liste de régions génomiques à exclure en amont du peak calling (Amemiya et al., 2019), que nous complétons de régions amplifiées spécifiquement dans nos input-seq de myélome ainsi que des chromosomes X et Y.

DÉFINITION DES « 5HMC DOMAINS »

De nombreux auteurs regroupent les pics de H3K27ac en domaines larges sur la base du modèle des super-enhancers (Lovén et al., 2013; Whyte et al., 2013). Les super-enhancers sont caractérisés en bioinformatique par réunion des pics de régions enhancers séparés d'une distance de moins de 12 500 bp avec l'algorithme ROSE, et le groupe de Boeva propose une méthode LILY sur le même modèle (Boeva et al., 2017).

Nous proposons d'étendre la caractérisation de « super-enhancers » aux domaines enrichis en 5hmC et donc de définir des « 5hmC domains ». En fait, lorsqu'on cherche à regrouper les pics de 5hmC en domaines enrichis larges, on retrouve le coude caractéristique des super-enhancers qui définit une distance critique à environ 12,5 kbp qui distingue entre des domaines denses et peu denses en pics. Nous détaillons ce concept dans notre article, et finalement, sur la base d'une liste de domaines de 5hmC communs à au moins deux patients dans notre cohorte, tous plus larges que 12,5 kbp, nous obtenons une matrice de score de niveau de 5hmC normalisée $M_{i,j}$ où i est un échantillon et j est un domaine de 5hmC.

ASSOCIATION D'UN DOMAINE 5HMC À UN GÈNE CIBLE

Une façon de rendre plus intelligible la localisation d'un domaine de 5hmC est de l'associer au nom d'un gène proximal.

Plusieurs possibilités s'offrent à nous : on peut annoter un domaine avec le nom du gène le plus proche. L'avantage majeur est que tous les domaines seront annotés. En échange, Il n'y a aucune certitude que le gène supposé cible soit exprimé.

De plus, l'annotation masque la fonction enhancer de régions codantes comme *PVT1* vis-à-vis de *MYC* ou *DUSP22* pour *IRF4*.

Une deuxième possibilité est d'utiliser le gène dont l'expression est la plus corrélée au signal de 5hmC dans la cohorte. Mais alors, où s'arrêter dans la recherche des gènes proximaux ? Nous proposons, en accord avec Boeva et collègues (Boeva et al., 2017), de chercher ce gène au sein des TAD (topologically associating domain) de la lignée B GM12878 identifiés par capture de la conformation de la chromatine par HiC. Dans le cas où le domaine 5hmC n'est pas associé à un TAD, nous associons le domaine avec le gène à l'expression la plus corrélée dans un périmètre d'1Mb, qui correspond à l'ordre de grandeur d'un TAD.

Pour des soucis d'intelligibilité, tous les domaines 5hmC portent le nom d'un gène même si le coefficient de corrélation est trop faible pour considérer qu'il y a effectivement une corrélation entre la 5hmC et l'expression. Réciproquement, on pourrait nommer un domaine 5hmC avec plusieurs gènes lorsque plusieurs corrélations linéaires passent un certain seuil. Nous avons fait le choix de conserver le gène le plus corrélé pour cet article.

SPÉCIFICITÉ D'UN DOMAINE 5HMC POUR UN GROUPE DE PATIENTS

Pour quantifier les différences entre deux groupes d'individus, on peut calculer la différence de niveau d'une variable entre ces deux groupes. La comparaison de classe en génomique utilise communément (1) un indicateur de différence moyenne associé à (2) une significativité de la différence corrigée pour (3) la réalisation de tests multiples.

Nous avons utilisé la mesure classique en génomique du Log2FoldChange moyen pour quantifier la différence de 5hmC entre deux groupes (exemple : groupe A versus non-groupe A), associée à un test non-paramétrique de Wilcoxon corrigé par la procédure de Benjamini-Hochberg.

Dans les tests pour lesquels des réplicats étaient disponibles, nous avons utilisé la méthode de DiffBind (Ross-Innes et al., 2012) avec DESeq2 (Love et al., 2014).

1.3. ARTICLE ET RÉSUMÉ DE L'ARTICLE : DNA HYDROXYMETHYLATION IS ASSOCIATED WITH DISEASE SEVERITY AND PERSISTS AT ENHANCERS OF ONCOGENIC REGIONS IN MULTIPLE MYELOMA

RÉSUMÉ DE L'ARTICLE EN FRANÇAIS

Dans cette étude, nous avons mesuré le niveau global et investigué la localisation de la marque chimique 5-hydroxyméthylcytosine (5hmC) sur l'ADN génomique de plasmocytes purifiés issus de ponctions de moelle osseuse. Les échantillons représentent 40 patients nouvellement diagnostiqués avec un myélome multiple (NDMM) et issus de la cohorte clinique IFM/DFCI 2009, dont 4 paires sont investiguées au diagnostic et à la rechute, ainsi que 5 plasmocytes purifiés de moelles contrôles.

Nous avons trouvé que les niveaux de 5hmC sont largement inférieurs dans les NDMM par rapport aux plasmocytes contrôles ($P < 0.0001$). Des niveaux élevés de 5hmC sont associés à un grade moins élevé de la maladie mesuré par International Staging System (ISS, $P < 0.05$), et tendent à être associés à une survie globale plus longue ($P < 0.1$).

Les données de séquençage et de cartographie de la 5hmC restante sur l'ADN ont montré que la 5hmC est organisée en large domaines qui chevauchent les marques de la chromatine active et ouverte. Nous avons trouvé que la 5hmC persiste à des niveaux élevés aux oncogènes majeurs du myélome dont *CCND1*, *CCND2* et *MMSET*. Ces domaines caractérisent également la 5hmC des sous-groupes de myélome. Des nouveaux domaines de 5hmC ont été découverts dans des régions non-codantes proximales de *CCND2* et *MYC* où la 5hmC est associée à la transcription et aux marques d'enhancers actifs.

En somme, le niveau global de 5hmC est associé avec des aspects cliniques de la maladie. La localisation génomique de la 5hmC fournit des informations majeures sur la chromatine active et la transcription, ce qui en fait une marque puissante pour étudier la biologie de la pathologie, découvrir des drivers oncogéniques et des régions enhancers sur de larges cohortes à partir d'ADN génomique brut.

RESEARCH

Open Access



DNA hydroxymethylation is associated with disease severity and persists at enhancers of oncogenic regions in multiple myeloma

Jean-Baptiste Alberge^{1,2}, Florence Magrangeas^{1,2,3}, Mirko Wagner⁴, Soline Denié¹, Catherine Guérin-Charbonnel^{1,2,5}, Loïc Campion^{1,2,5}, Michel Attal^{6,7}, Hervé Avet-Loiseau^{6,7}, Thomas Carell⁴, Philippe Moreau^{1,2,3}, Stéphane Minvielle^{1,2,3*} and Aurélien A. Sérandour^{1,8*} 

Abstract

Background: Multiple myeloma (MM) is a heterogeneous plasma cell malignancy that remains challenging to cure. Global hypomethylation correlates with an aggressive phenotype of the disease, while hypermethylation is observed at particular regions of myeloma such as B cell-specific enhancers. The recently discovered active epigenetic mark 5-hydroxymethylCytosine (5hmC) may also play a role in tumor biology; however, little is known about its level and distribution in myeloma. In this study, we investigated the global level and the genomic localization of 5hmC in myeloma cells from 40 newly diagnosed patients, including paired relapses, and of control individuals.

Results: Compared to normal plasma cells, we found global 5hmC levels to be lower in myeloma ($P < 0.001$). Higher levels of 5hmC were found in lower grades of the International Staging System prognostic index ($P < 0.05$) and tend to associate with a longer overall survival ($P < 0.1$). From the hydroxymethylome data, we observed that the remaining 5hmC is organized in large domains overlapping with active chromatin marks and chromatin opening. We discovered that 5hmC strongly persists at key oncogenic genes such as *CCND1*, *CCND2* and *MMSET* and characterized domains that are specifically hydroxymethylated in myeloma subgroups. Novel 5hmC-enriched domains were found at putative enhancers of *CCND2* and *MYC* in newly diagnosed patients.

Conclusions: 5hmC level is associated with clinical aspects of MM. Mapping 5hmC at a genome-wide level provides insights into the disease biology directly from genomic DNA, which makes it a potent mark to study epigenetics on large patient cohorts.

Keywords: Multiple myeloma, Epigenetics, DNA modifications, Hydroxymethylation

Introduction

Multiple myeloma (MM) is a plasma cell (PC) neoplasm with an incidence rate of 5/100,000 in Europe and it accounts for approximately 1% of all cancers. Median survival of patients has greatly improved in the last decade [1] with the use of novel strategies such as autologous

stem cell transplantation and new sets of drugs: immunomodulators, proteasome inhibitors, histone deacetylase (HDAC) inhibitors and monoclonal antibodies [2]. Yet, the treatment remains challenging as nearly all patients ultimately relapse with the emergence of a resistant subpopulation of malignant plasma cells. Malignant clones show a heterogeneous range of mutations and chromosomal abnormalities along with heterogeneous chromatin and epigenetic dysregulations at diagnosis and relapse that affects biological pathways such as MAPK, NF- κ B, or DNA-repair [3].

*Correspondence: stephane.minvielle@univ-nantes.fr; aurelien.serandour@ec-nantes.fr

¹ Université de Nantes, CNRS, Inserm, CRCINA, Nantes, France
Full list of author information is available at the end of the article



© The Author(s) 2020. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Genomic and transcriptomic studies have allowed a better understanding of the disease and identified key transcription factors (TFs) involved, such as IRF4, MYC, PRDM1 and XBP1 [3]. Recent epigenomics technologies can help to deepen our knowledge of the transcriptional programs shaping MM. However, epigenomics analysis through histone marks profiling with chromatin immunoprecipitation (ChIP-seq) or open chromatin mapping with assay for transposase-accessible chromatin (ATAC-seq) can be hard to set up on a cohort with limited material over a long-time frame and thus gives a limited insight into the disease's establishment and relapse. Studying DNA epigenetics marks is more adapted to this challenge.

Oxidative states of 5-methylated Cytosine (5mC) such as 5hmC, 5fC and 5caC were identified in genomic DNA a decade ago [4–6]. While DNMT1/3A/3B catalyzes DNA methylation at CpG, the TET proteins TET1/2/3 are responsible for the 5mC oxidation. Interestingly, TET1 and TET2 have been shown to have tumor suppressor roles in B cell lymphomas [7, 8]. 5fC and 5caC are almost undetectable in genomic DNA unless the glycosylase TDG gene is knocked-out [9, 10], whereas 5hmC can be found in all cell types at various levels [11]. 5hmC is believed to be a DNA demethylation intermediate in a process involving TET proteins, TDG and the Base Excision Repair system [12]. However, 5hmC has also shown to be in most cases a stable DNA modification and its abundance increases with DNA age [13, 14]. 5hmC is commonly accepted as a DNA mark associated with active chromatin [15–17] and is a powerful way to identify active genomic domains associated with a disease directly from genomic DNA or more recently from circulating DNA [18, 19].

In solid tumors, 5hmC is depleted compared to normal tissue [20] and some studies show that a lower 5hmC level is associated with poorer outcome [21–23]. When 5hmC is depleted, the mark is lost at genic regions, particularly around transcription start sites (TSS) and in gene bodies [24]. However, the putative tumor suppressor role of 5hmC is poorly characterized [20, 25].

Agirre et al. [26] described DNA methylation in an important number of MM samples. They identified very heterogeneous levels of methylation from one patient to another. They found that despite a global hypomethylation [27–29], local and extensive hypermethylation is present in MM at intronic enhancer regions that are associated with B cell-specific enhancers and stem cell development.

Chatonnet et al. [30] recently identified several hydroxymethylated CpGs in a limited panel of MM samples, yet the genome-wide mapping of 5hmC (hydroxymethylome) has never been studied in a well-established cohort.

In this study, we quantified the 5mC and 5hmC by MS and mapped the 5hmC genome-wide on plasma cell DNA from 40 patients newly diagnosed with MM between 2010 and 2012 [31] and representative of the main molecular sub-types, including 4 paired relapse samples, and of the plasma cells of 5 control individuals.

Results

5hmC negatively correlates with disease severity in newly diagnosed multiple myeloma

We studied a cohort of 40 patients newly diagnosed with multiple myeloma (NDMM) between 2010 and 2012 and 5 healthy bone marrow donors. NDMM were included in the “Intergroupe Francophone du Myélome/Dana Farber Cancer Institute 2009” cohort (IFM/DFCI) and have expression profile available [32]. Patients' characteristics are described in Additional file 1: Table S1.

We quantified by MS the global level of 5mC and 5hmC in 49 samples (40 diagnosis in 5mC, 39 in 5hmC with one failure, 4 relapses and 5 normal plasma cell samples). We found that both 5mC and 5hmC were significantly reduced in MM compared to normal plasma cells (NPC) (Fig. 1a and Additional file 2: Figure S2A). On average, 5mC is reduced ($P < 0.001$) by 34%, while 5hmC is reduced ($P < 0.001$) by 69% in MM with respect to NPC.

We found that 5hmC, but not 5mC, is reduced in MM stages II and III compared to Stage I (resp. by 25% and 31%, $P < 0.05$) regarding the International Staging System [34], a classification of patients based on beta2-microglobulin and albumin levels with a strong prognosis value (Fig. 1b right and Additional file 2: Figure S2B). 5mC and 5hmC global levels were not correlated to the sex or the age of the patients (Additional file 2: Figure S2C–F).

Survival analysis was performed on the global level of 5mC and 5hmC measured by MS. Higher 5hmC global level tends to be associated with good outcome (hazard ratio (HR) = 2.6, CI = [0.9, 7.8], $P = 0.066$) (Fig. 1c), whereas 5mC global level does not show significant association with survival ($P = 0.26$, Additional file 2: Figure S3).

5hmC persists at active and accessible chromatin of myeloma cells

These results encouraged us to characterize the hydroxymethylome by 5hmC-seq [35] in the 40 NDMM and 5 NPC to identify the genomic regions remaining marked by 5hmC.

The 5hmC signal predominantly marks strong enhancers, promoters, and to a lesser extent transcribed regions of NDMM (Fig. 1d, each left-side boxplot) when we used a ChromHMM functional annotation based on modified histone ChIP-seq data from the NCI-H929 cell line.

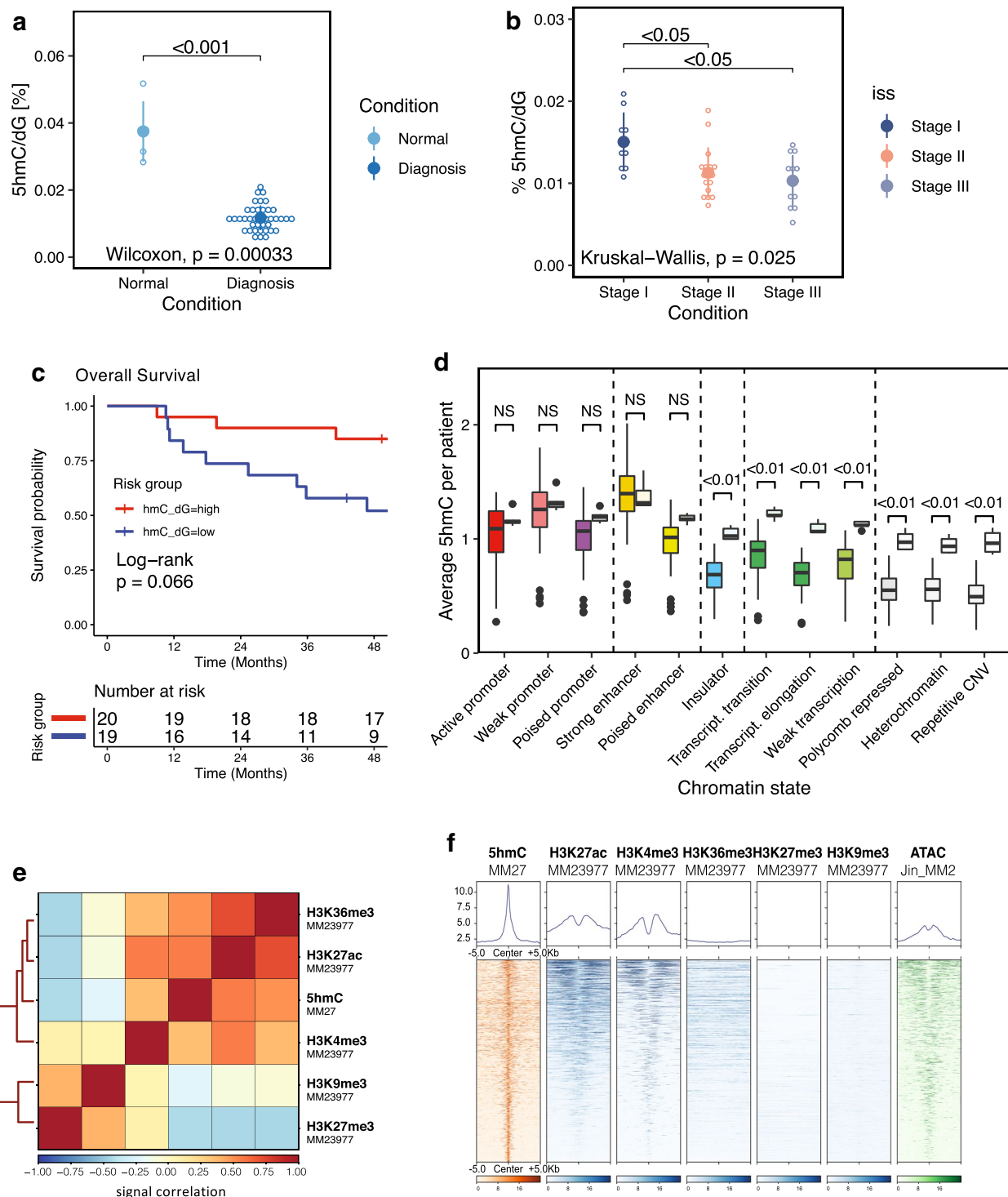


Fig. 1 DNA hydroxymethylation is depleted in NDMM and remains locally at active chromatin. **a** Dot plot of 5hmC global quantification by MS in normal plasma cells from healthy donors ($n = 5$), and of myeloma cells of patients at diagnosis ($n = 39$) and **b** at disease stage (ISS I $n = 9$; ISS II $n = 17$; ISS III $n = 12$; NA $n = 1$). **c** 5hmC level-based survival analysis with separation of two groups of NDMM ($n = 20$ and 19 , 1 measurement failure) **d** Distribution of the 5hmC signal at the different ChromHMM chromatin states in multiple myeloma ($n = 40$, each left-side boxplot) versus normal plasma cells ($n = 5$, right sides). **e** Correlation matrix between 5hmC signal (this study) and the histone marks H3K36me3, H3K27ac, H3K4me3, H3K9me3 and H3K27me3 (ChIP-seq data from the Blueprint Consortium) in MM patients. **f** Average signal and signal matrix at 5hmC peaks of 5hmC (this study), histone marks (Blueprint) and ATAC-seq [33] in MM patients. ISS, International Staging System, NDMM, Newly Diagnosed Multiple Myeloma

Insulator, Polycomb repressed, heterochromatin and repetitive regions are depleted in 5hmC in NDMM compared with the strong enhancers and promoters. In NPC however, 5hmC is relatively widespread (Fig. 1d, each right-side boxplot). When comparing NDMM and NPC, we show that 5hmC is globally depleted in the genome but persists in active promoters and enhancers (Fig. 1d).

Furthermore, we found 5hmC to be positively correlated with active chromatin (H3K27ac and H3K36me3, in a less extend with H3K4me3) but negatively correlated with inactive chromatin (H3K9me3 and H3K27me3) in independent MM patients (Fig. 1e). At local enriched regions, we found the 5hmC peaks to be associated with H3K27ac, H3K4me3 and broad H3K36me3 signals but not with H3K9me3 and H3K27me3 (Fig. 1f). The 5hmC signal is located in H3K27ac and H3K4me3 valleys proximal to histone mark peaks.

The 5hmC peaks are also associated with open chromatin as it is enriched in ATAC-seq signal (Fig. 1f). Finally, we observed that the level of 5hmC signal is enriched within gene bodies of highly transcribed genes, as expected from previous studies [16, 35, 36] (Additional file 2: Figure S4A).

All together, these results show that despite its global loss in NDMM compared to NPC, 5hmC is present at transcriptionally active chromatin. This makes 5hmC a useful tool to study active chromatin in DNA from patients' samples.

The 5hmC landscape in MM is organized in large 5hmC-enriched domains

Similar to H3K27ac super-enhancers [37], we found that 5hmC is also organized in large peak clusters that we called 5hmC-enriched domains (Fig. 2a, b; Additional file 2: Figure S4B, Additional file 1: Table S2).

When we ranked 5hmC-enriched domains of a t(4;14) patient from IFM/DFCI and H3K27ac super-enhancers of a t(4;14) patient [33], we found an important overlap between genomic domains (Fig. 2a). In particular, *FNDC3B*, *CREB3L2* and *NSD2* are among the strongest common active genomic domains (Fig. 2b). To go further into the characterization of 5hmC landscape in NDMM, we applied the same procedure to each of the NDMM samples and we kept all the domains ($n=1816$) overlapping in at least two samples. Interestingly, comparing with the literature, 41% of hypermethylated CpGs located in enhancer regions of B cells [26] were included within the 1816 5hmC domains, compared with only 11% ($\pm 0.5\%$) when we chose random CpGs (Additional file 2: Figure S4C). This suggests that 5hmC persists at hypermethylated loci of MM.

Similar to H3K27ac super-enhancers, 5hmC levels in enriched domains may correlate with a proximal

expressed gene ($R^2_{\max}=0.78$, $n=1816$, Additional file 1: Table S2). Briefly, 5hmC-enriched domains were associated with the most strongly correlated gene within its topologically associated domain (TAD) when we used TAD data from the GM12878 cell line (see Additional file 3: Methods). Remarkably, the strongest correlation between 5hmC and RNA expression is found at *NSD2* ($R^2=0.78$), and highly expressed genes such as *CCND2* ($R^2=0.64$), *GAS6* ($R^2=0.55$), *IL6R* ($R^2=0.49$) and *PRKCB* ($R^2=0.57$) display high correlation coefficients with their neighboring 5hmC domain (Fig. 2c and Additional file 1: Table S2).

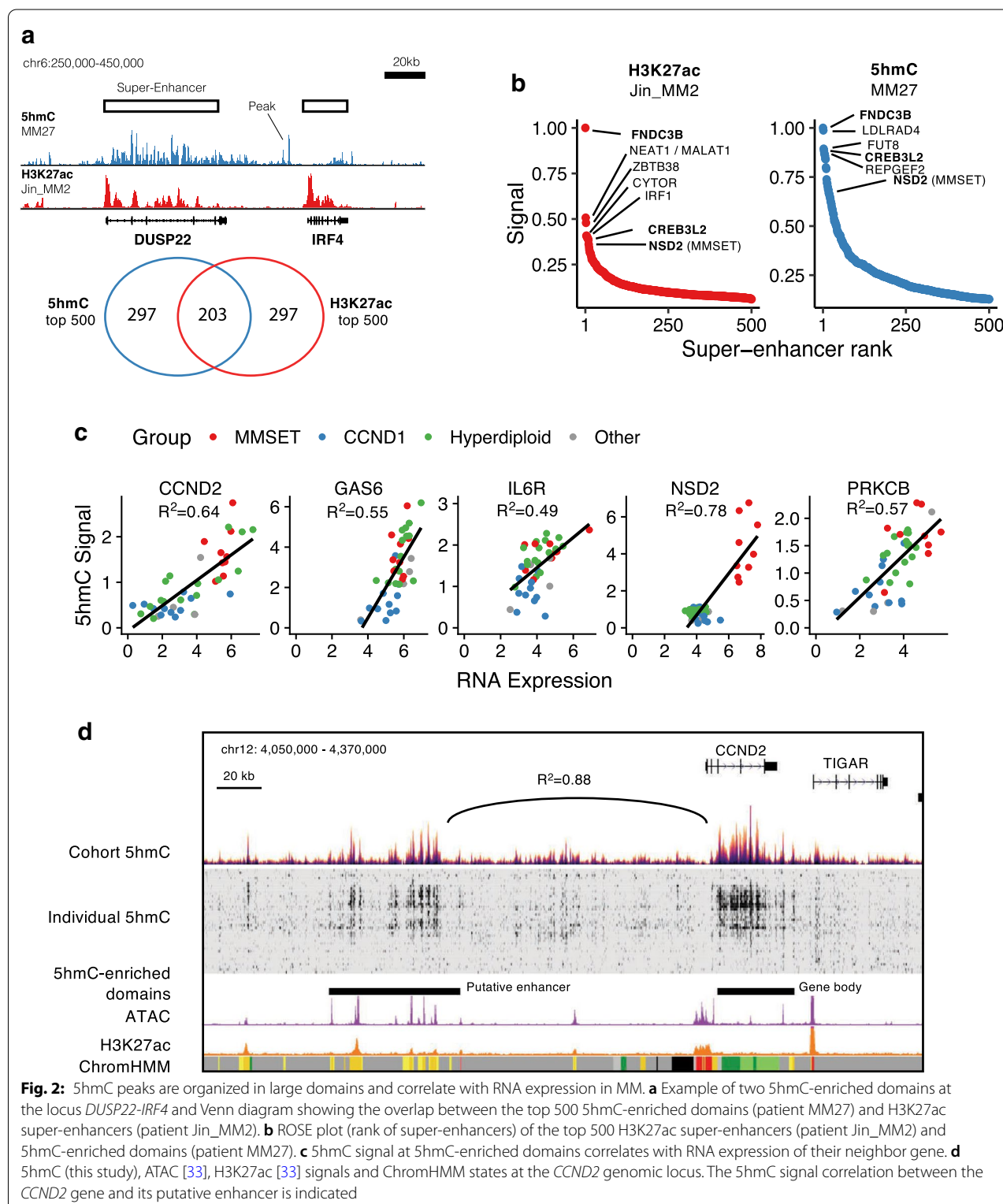
Several genes were found to have more than one neighboring 5hmC-enriched domain. In this case, two domains or more are found to correlate with the same proximal expressed gene (i.e., genes *SECTM1* and *IQSEC1* have four 5hmC-enriched domains each; *ALOX5*, *GRIK4* have 3 neighboring domains; *CCND2*, *FRZB*, 2 domains; see Additional file 1: Table S2 for complete records). Finally, 1468 5hmC-enriched domains are specific to their putative gene.

Regardless of proximity to an expressed gene, we found the strongest correlation between two 5hmC-enriched domains at the *CCND2* locus ($R^2=0.88$) with 5hmC signal in *CCND2* gene body (hg38, chr12:4,278,700–4,312,900) and 5hmC signal at an extragenic domain located 120 kb upstream (hg38, chr12:4,106,500–4,164,700; Fig. 2d and Additional file 2: Figure S5A). Both 5hmC signals correlate strongly with *CCND2* RNA expression and both domains are located in the same topological domain according to HiC data [38] (hg38, chr12:3,850,000–4,800,000; Additional file 2: Figure S5B). The extragenic domain is also marked by ATAC and H3K27ac signals [33] in independent MM patients suggesting that this genomic domain is functionally active. This strongly suggests that this upstream domain is the enhancer of *CCND2* gene in the MM context.

We took advantage of the 5hmC-enriched domains and our RNA-seq data to search for core regulatory circuits as Saint-André and colleagues [39]. The concept of core regulatory circuits is based on the fact that only a small subset of interconnected TFs is responsible for the control of the whole transcriptome. We identified motifs associated with 39 TFs that bind 5hmC-enriched domains and regulate transcription of genes in the vicinity (Additional file 2: Figure S5C). Top expressed TFs binding 5hmC-enriched domains include *XBPI*, *ATF4*, *KLF6*, *USF2*, *IRF4*, *PRDM1*, *IRF1*, *KLF13*, *USF1* and *TCF3* (Additional file 1: Table S3).

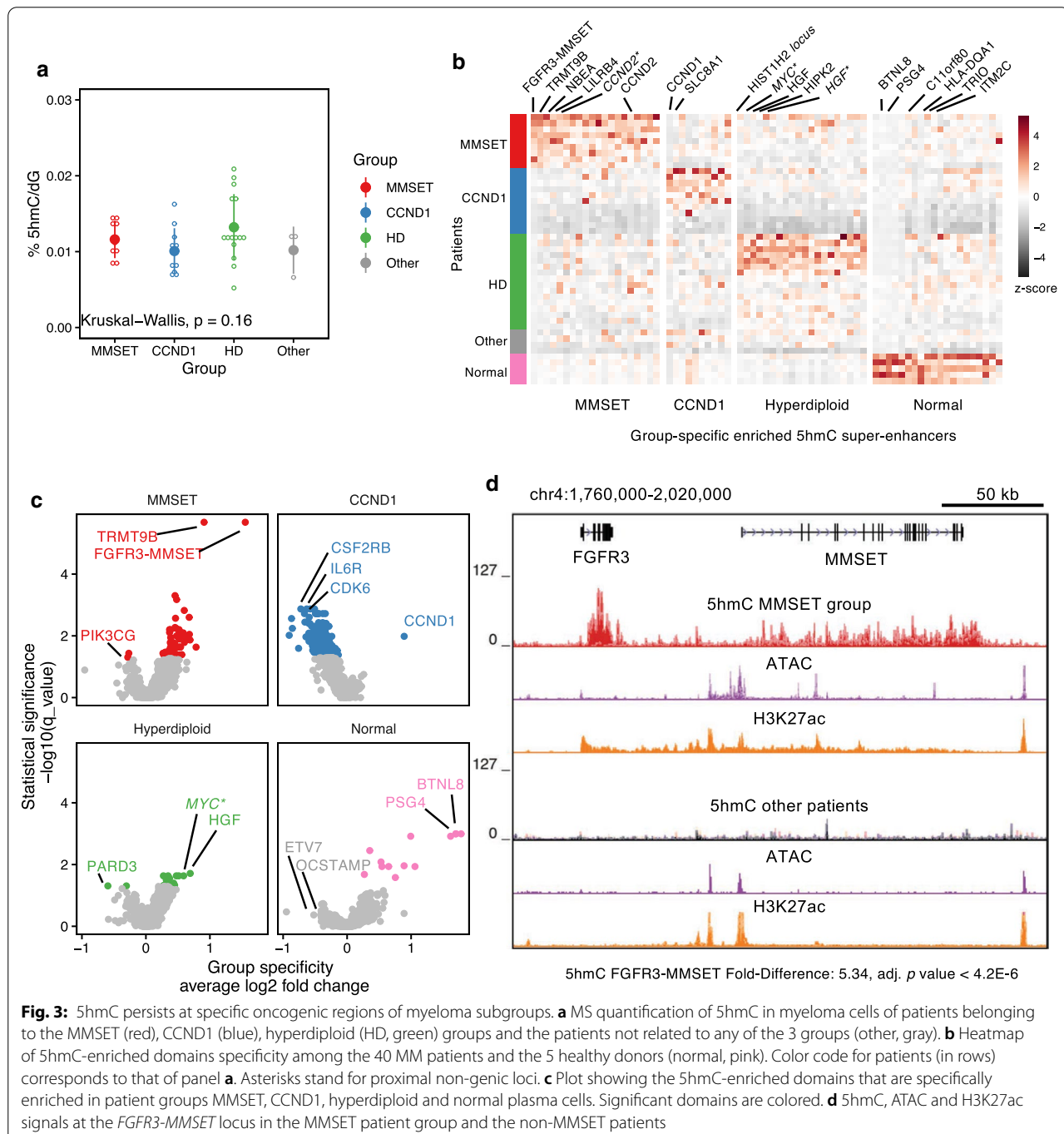
MM subgroups display specific 5hmC-enriched domains

Patients' samples were classified into 4 groups: MMSET [translocation t(4;14); 9 patients], *CCND1*



(RNA expression over 800 Transcripts per Million; 11 patients), hyperdiploid (16 patients; at least 2 odd chromosomal gains) and others (MM patients in none of

the aforementioned groups; 4 patients). Global level of 5hmC measured by MS shows no significant difference between MM groups (Fig. 3a). However, locally, some



5hmC-enriched domains are found to be group specific (Fig. 3b, c, Additional file 2: Figure S6A and Additional file 1: Table S4). Remarkably, the strongest specific 5hmC-enriched domain for the group MMSET is overlapping the *FGFR3-MMSET* locus ($P=1.6E-6$) followed by *CCND2*, *LILRB4*, *NBEA* and *TRMT9B* (Fig. 3d and Additional file 2: Figure S6B). This strong 5hmC enrichment in MMSET patients is also associated with

strong H3K27ac and ATAC-seq signals in independent MM patients [33]. Again, in patients from the CCND1 group, the strongest and most specific 5hmC signal was located in the *CCND1* gene itself (Fig. 3c and Additional file 2: Figure S6C, $P<0.05$). CCND1 group, which is determined by expression of CCND1 in absence of t(11;14) FISH, is marked by down-hydroxymethylation at *CSF2RB*, *IL6R* and *CDK6* loci ($P<0.05$). The three

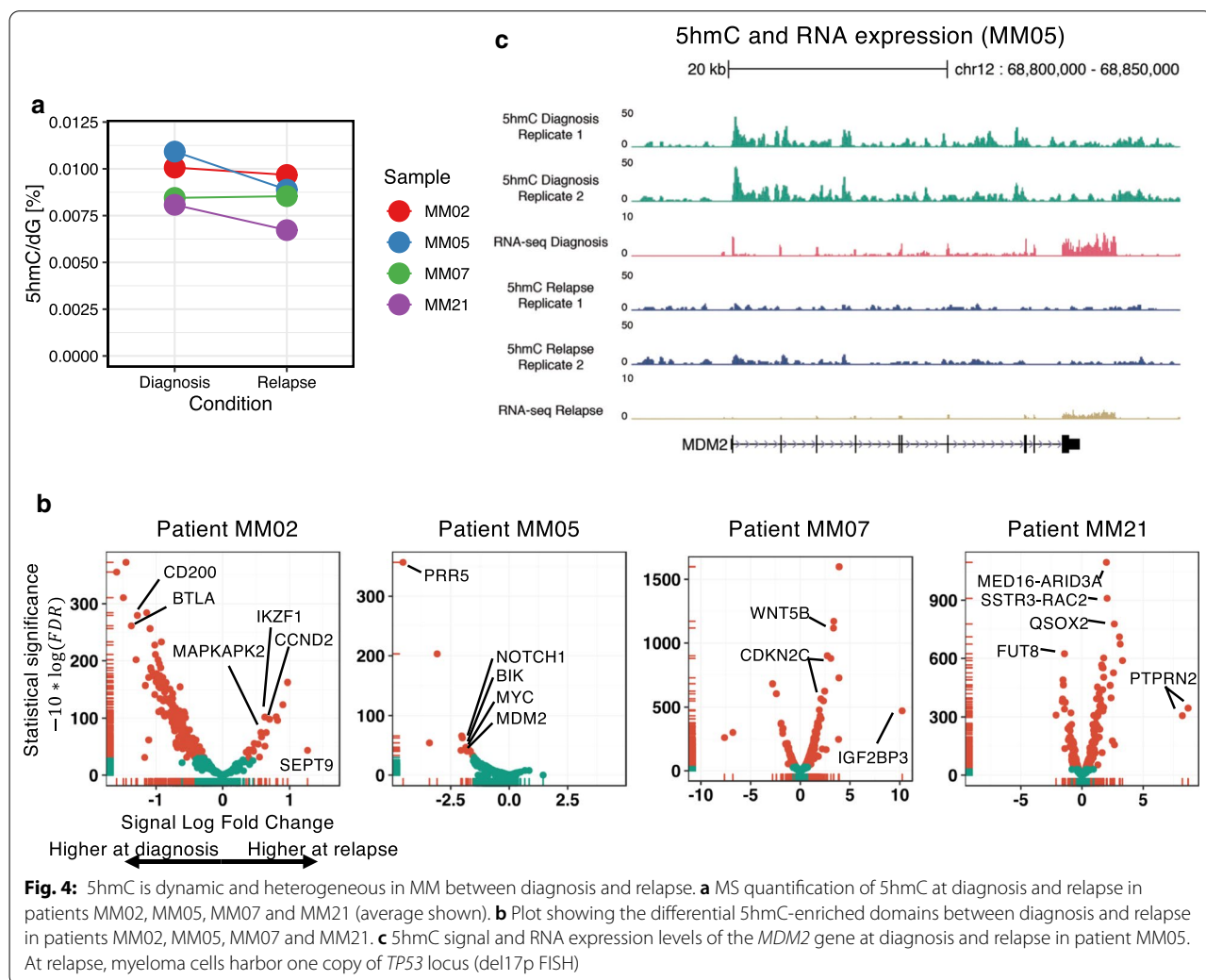
genes are also downregulated in RNA-seq from the same patients' samples ($P < 0.01$, data not shown).

The hyperdiploid group shows strongly specific 5hmC signal at *HGF* (hepatocyte growth factor, $P = 0.02$) and at the locus of *MYC* oncogene (Fig. 3b, c, Additional file 2: Figure S6D, $P = 0.02$). Normal plasma cells are enriched in 5hmC at *BTNL8*, *C11orf80*, *ITM2C*, *PSG4*, and *TRIO* genes (Fig. 3b, c). Of note, hematopoietic stem cell genes were not found over-represented when we performed a Gene Ontology term enrichment analysis. Main subgroups of NDMM thus show specific 5hmC at translocation breakpoints loci and at major oncogenes.

5hmC-enriched domains are dynamic between diagnosis and relapse in MM

To identify the genomic domains associated with MM progression, we quantified and sequenced 5hmC genome-wide in four MM pairs (diagnosis and first relapse) and identified differential 5hmC-enriched

domains using two replicates for each condition. CNV microarray data [40, 41] and RNA-seq were used at both time points to assess the progression of each patient. In three out of four paired samples, DNA hydroxymethylation slightly decreases at relapse (Fig. 4a). Although all patients' samples display significant changes of 5hmC localization at relapse, there was no significant change in 5hmC that overlapped the four cases when we used DiffBind for differential enrichment analysis [42] (Additional file 1: Table S5). Since 5hmC landscape at relapse seems heterogeneous and patient-dependent, we focused our analysis on each of the four cases. Myeloma cells of patient MM02 showed a translocation of the *MMSET* locus [t(4;14)] and deletion of chromosomes 1p, 13 and 17p at diagnosis. This subject progressed in 18 months, with his myeloma cells displaying an additional third copy of the chromosome arm 1q. Out of 560 consensus 5hmC-enriched domains, 269 (48%) were significantly reduced at relapse compared to diagnosis ($FDR < 0.05$), while 20



(3.5%) increased at relapse. Interestingly, among the most significant gain of 5hmC-enriched domains, we found the *CCND2* and *IKZF1* gene bodies at relapse (respectively, 1, 6 and 1, fivefold; Fig. 4b) associated with a higher transcriptional activity (respectively, 2, 4 and 1, fivefold). It is interesting to note that the drug lenalidomide is effective by inducing *IKZF1* proteasomal degradation via Cereblon [43, 44]. This suggests that the up-regulation of *IKZF1*, together with an increase in *CCND2* expression could favor disease progression. We also noticed a significant gain of 5hmC at *MAPKAPK2/MK2* at relapse. This gene is located on the chromosome arm 1 (gaining an extra-copy at relapse) and has been recently described as a poor prognosis factor in MM [45].

At diagnosis, patient MM05 displayed a classical hyperdiploid profile with amplification of chromosomes 3, 9, 11, 15, 19, in which we found 675 5hmC-enriched domains. At relapse 24 months later, myeloma cells displayed a 1-copy loss of the chromosome arm 17p and 22 5hmC-enriched domains (3.2%) significantly decreased. In particular, we found a lower expression and lower 5hmC signal at TP53 target gene *MDM2* (Fig. 4b, c).

Patient MM07 is another *MMSET* translocated patient [t(4;14)] with amplification of the chromosome arm 1q and deletions of chromosomes 13 and 17p (Fig. 4b) at diagnosis. At relapse, there are additional deletions of chromosomes arms 3p, 6q and 8q, and 5hmC-enriched domains increase at several genes including *CDKN2C*, *IGF2BP3* and *WNT5B* as well as their RNA expression ($\log_{FC} > 3$, Additional file 2: Figure S7).

Patient MM21 plasma cells were found to have a narrow *TET2* deletion at diagnosis, but surprisingly 2 heterozygous copies of *TET2* at relapse. *TET2* protein, which oxidizes 5mC in 5hmC, is frequently loss-of-function-mutated in cancer, especially in myeloid malignancies [46], but not in MM [32]. A total of 385 domains were found, out of which 86 (22%) decreased and 117 (30%) increased at relapse and global level of 5hmC decreased despite the reappearance of the missing *TET2* copy.

These data indicate that 5hmC signal varies between diagnosis and relapse in a patient-specific manner. However, several genes and enhancers, whose activities changed at relapse, could be identified as putative drivers of disease progression.

Discussion

We have shown that the epigenetic mark 5hmC is lower in MM than in normal plasma cells, and that it gradually decreases with the disease severity of patients while being independent of patient's age. Our observation corroborates the theory of DNA hydroxymethylation being linked to mitotic index [13, 14] as 5hmC decreases with the tumor severity. In that sense, global

levels of hydroxymethylation are consistent with the already-described global hypomethylation of MM. With a limited number of samples ($N=39$) and events ($N=16$), 5hmC level tends to correlate with longer survival ($P < 0.1$). This suggests that the 5hmC level is a prognosis biomarker for newly diagnosed MM. Independent cohort followed on a longer time will be needed to confirm the good-prognosis association that we observe. In any case, our study supports the potential use of 5hmC and 5mC quantification as a clinical biomarker in MM.

Our analysis shows that 5hmC localizes predominantly at transcriptionally active regions. Surprisingly, this is true despite a global and important loss of 5hmC. The mark seems to be maintained by TET proteins at highly active chromatin. This stays consistent with the fact that the DNA methylation in MM is reduced globally but maintained at intronic enhancers regions [26]. In a sense, there is an association between the remaining localization of 5mC and 5hmC in MM, although the marks are thought to flag different regions: inactive versus active chromatin.

Given that 5hmC efficiently marks active chromatin, we applied the H3K27ac super-enhancer discovery algorithms to the 5hmC signal and found similar regions. This analysis highlighted key components of the plasma cell biology (*CREB3L2* for example), and more importantly, of myeloma subgroups such as *MMSET* locus in t(4;14) patients. We show that defining 5hmC-enriched domains similar to H3K27ac super-enhancers, rather than studying single 5hmCpGs, is relevant and powerful to decipher major disease drivers on patient's genomic DNA.

We also show that major well-known translocation events produce massive oxidation at the recombinant loci, together with previously shown chromatin opening and high transcription level. In addition to translocation and myeloma subgroups, we found novel active domains, for instance, proximal to *CCND2* and *TRMT9B* in *MMSET* patients. De novo combined analysis of 5hmC and RNA expression revealed key transcription factors involved in the disease such as *IRF4*, *PRDM1* and *TCF3*. These TFs likely drive the malignant transcriptome as they are essential in most of the MM cell lines in CRISPR KO screening from DEPMap (<https://depmap.org/portal/>). Despite being difficult to drug, we believe that these TFs are therapeutic targets of interest.

Between diagnosis and relapse, we found a highly dynamic and patient-specific distribution of the 5hmC signal. This reflects that MM progression is highly heterogeneous, although we could find some consistency between 5hmC changes, RNA expression and copy number variations. In the future, it would probably be more meaningful to describe the dynamics of 5hmC on

more samples and at several time-points of the disease progression.

Profiling chromatin with histone marks (e.g., H3K27ac ChIP-seq) or chromatin accessibility (ATAC-seq) requires nuclei and thus limits the study of epigenetics on clinical cohorts with only genomic DNA stored for historical and practical reasons. The technical limitation of preserving frozen nuclei is overcome by mapping active domains of chromatin directly on genomic DNA through the 5hmC mark produced by the TET proteins. Our study shows indeed that the epigenetic mark 5hmC is valuable to discover active regulatory domains in genomic DNA from a cohort of patients without the need of chromatin extraction. Furthermore, it has been recently shown that it is also possible to map 5hmC on circulating DNA [18, 19]. This makes 5hmC not only biologically valuable, but it is also technically easier to map genome-wide or at key oncogenic loci. Taken together, these results show the value of epigenomics in retrospective studies and bring to light potential drug targets that drive the malignant transcriptome.

Conclusions

We show that DNA hydroxymethylation is an active chromatin mark that is globally depleted in malign plasma cells but remains at MM essential genes that drive the malignant transcriptional network. Furthermore, the global DNA hydroxymethylation level decreases with MM severity and tends to be associated with outcome. This observation needs to be further assessed. Remaining hydroxymethylation localizes at major translocation breakpoints and active chromatin. It is thus a potent mark to discover oncogenic drivers of the disease. We expect DNA hydroxymethylation, and DNA modifications in general, to become major clinical biomarkers in the future, especially when direct-DNA-sequencing will be efficient and cost-effective, allowing the genome-wide mapping of DNA modifications in a single run.

Methods

Genomic DNA extraction of normal plasma cells and multiple myeloma cells

Normal plasma cell and multiple myeloma cells collection and purification from human bone marrow are described in Additional file 3: Methods. Genomic DNA was extracted with Qiagen Allprep DNA/RNA Mini Kit (ref. 80204). DNA samples were dosed by DNA HS Qubit and the absence of contaminant RNA checked by RNA HS Qubit.

Patients selection for this study

No statistical method was used to predetermine sample size. In this study, we have selected 40 patient samples

from the IFM/DFCI cohort (NCT01191060) further detailed in Additional file 3: Methods. Samples were selected with: low level or absence of RNA in DNA samples, low level of rRNA in RNA-seq data, enough DNA material available and high percentage of CD138+ cells (98% in average in this study). Healthy individuals and myeloma patients are of comparable age (resp. 62 and 58 in median). Men were overrepresented in the healthy group ($N=4/5$). Neither control individuals nor myeloma patients were given vitamin C supplementation; however, nutritional aspects have not been checked before patient selection [47]

Digestion of genomic DNA and subsequent LC-MS analysis

The genomic levels of 5mC and 5hmC were quantified using a mass spectrometry-based stable isotope-dilution method [48]. For each LC-MS measurement, 70 ng of genomic DNA was digested to the nucleoside level. As heavy-atom-labeled internal standards, fixed quantities of D₃-5mC and D₂-¹⁵N₂-5hmC were added to the mixture. For each biological sample, two independent measurements (technical replicates) were taken. Quantitative LC-ESI-MS/MS analysis was performed using an Agilent 1290 UHPLC system coupled to an Agilent 6490 triple quadrupole mass spectrometer in conditions similar to Traube et al. [48]. Further details are described in Additional file 3: Methods.

Selective chemical labeling of 5hmC coupled with sequencing (5hmC-seq)

For each sample, 550 ng of genomic DNA was sonicated with a Bioruptor Pico in Tris 10 mM pH 8 to obtain DNA fragments of 300 bp in average. 25 pg of 5hmC control spike-in was added to the sonicated DNA (control provided by the kit HydroxyMethyl Collector, Active Motif, ref. 55013). 50 ng of DNA was conserved at this stage to make the input library later. The remaining DNA was processed using the HydroxyMethyl Collector kit (method from Song et al. [35]) to glycosylate and biotinylate specifically the genomic 5hmC. After glycosylation and biotinylation, the DNA was purified with Ampure beads (Beckman Coulter, ref. A63881). The DNA fragments containing the biot-glu-5hmC were purified with Streptavidin beads (Active Motif, ref. 55013), eluted, purified with Ampure beads and finally eluted in 50 µL Tris pH 8. The 5hmC-seq libraries were prepared with the kit NEBNext Ultra II DNA library prep kit for Illumina (ref. E7645S) and indexed with NEBNext dual indexed primers (E7600S). The libraries were quality-checked by HS DNA Agilent BioAnalyzer (Additional file 2: Figure S1), dosed by DNA HS Qubit, pooled and submitted to the genome sequencing platform for Single-Read 50 bp Illumina HiSeq-2500 Rapid Run sequencing.

RNA-seq libraries

As the IFM-DFCI did not include RNA-seq data from patients at relapse, we produced additional RNA-seq data at diagnosis and at relapse for 4 MM patients (patient number MM02, MM05, MM07 and MM21). The RNA-seq libraries were prepared using the NEBNext Poly(A) mRNA Magnetic Isolation Module (NEB, ref. E74905) and the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (NEB, ref. E7760S) and sequenced by an Illumina Rapid Run HiSeq 2500 Single-Read 50 bp.

Bioinformatics (cf. Additional file 3: Methods).

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s13148-020-00953-y>.

Additional file 1. Figure S1: Agilent BioAnalyzer profile of a 5hmC-seq Illumina library. **Figure S2:** MS quantification of 5mC in genomic plasma cell DNA is independent of age and sex. **(A)** Dot plot of 5mC global quantification by MS in normal plasma cells from healthy donors (N=5), and of myeloma cells of patients at diagnosis (N=40). **(B)** Dot plot of 5mC global quantification by MS by disease stage (ISS I N=9; ISS II N=17; ISS III N=13; NA=1). 5mC **(C)** and 5hmC **(D)** dot plot of MS quantification depending on the sex of the patients. 5mC **(E)** and 5hmC **(F)** dot plot of MS quantification depending on the age of the patients. **Figure S3:** Survival course depending on DNA methylation (5mC) level-based separation of two risk groups of NDMM (n=20 and 20). **Figure S4:** 5hmC association with expression and criteria of 5hmC peaks to merge in 5hmC peak clusters. **(A)** Average level of 5hmC in all genes normalized to the same body length. Red line stands for average 5hmC in genes with high expression level (greater than 100 Reads per Kilobase Million (RPKM)). Orange line stands for medium expression level (between 10 and 100 RPKM). Green line represents lowly expressed genes (between 1 and 10 RPKM), and blue line stands for very lowly expressed genes (below 1 RPKM). **(B)** Stitching of 5hmC into 5hmC-enriched domains. The y-axis represents the number of peaks left after merging. The x-axis represents the distance between peaks to merge. Each 5hmC sample was analyzed (one color per patient). The distance 12.5 kb was chosen to stitch 5hmC peaks into the 5hmC-enriched domains that we describe in this study. **(C)** Fraction of overlap between 5hmC-enriched domains of this study and CpG from the Illumina 450K chip. The red bar represents overlap with hypermethylated CpGs in B cell-specific enhancers that were described by Agirre and colleagues (see Additional file 3: Methods). Blue bars represent random CpGs from the same chip. **Figure S5:** 5hmC allows the identification of a putative CCND2 enhancer. **(A)** Correlation between CCND2 expression, 5hmC at CCND2 gene body and 5hmC at the putative 5hmC enhancer across the 40 MM patients. **(B)** Hi-C contact map in lymphoblastoid cells (GM12878 cell line) at the CCND2 locus showing the spatial interaction between CCND2 gene and its putative enhancer. **(C)** Expression of core transcription factors predicted to orchestrate core regulatory circuitries with 5hmC and RNA expression genomic data. **Figure S6:** MM 5hmC-enriched domains associate with H3K27ac super-enhancers. Rank ordering of the 100 strongest 5hmC-enriched domains on average in the cohort **(A)**, in the MMSET group **(B)**, in the CCND1 group **(C)** and in the hyperdiploid group **(D)**. Color highlights domains present in only one of the ROSE plots by group. **Figure S7:** 5hmC signal levels at WNT5B-associated domain are increased at relapse in MM07. **(A)** Normalized 5hmC enrichment at WNT5B-associated domain. Point shapes match replicates. Fold change=1.3, p=0.003, FDR>0.1. **(B)** Gene expression levels in RPKM measured by RNA-seq at diagnosis and relapse for three genes surrounding the WNT5B-associated domain. **(C)** 5hmC genomic signal around WNT5B-associated domain. Colors match those of (A) and (B). 5hmC domain is depicted under signal tracks (hg38: chr12:1,517,750-1,621,200).

Additional file 2. Table S1: Patients characteristics, survival and mass spectrometry quantification of 5hmC and 5mC. **Table S2:** Scoring of 1816 5hmC-enriched domains across NDMM samples. **Table S3:** Motif analysis of core regulatory circuitries. **Table S4:** Scoring of groups-specific 5hmC-enriched domains. **Table S5:** Differential 5hmC-enriched domains between at diagnosis and relapse (DiffBind analysis).

Additional file 3. Methods: This file describes normal plasma cells purification, myeloma cells purification, 5mC and 5hmC dosage by mass spectrometry, bioinformatics methods and statistical analysis.

Abbreviations

5caC: 5-CarboxylCytosine; 5fC: 5-FormylCytosine; 5hmC: 5-Hydroxymethyl-Cytosine; 5hmCpG: 5'-HydroxymethylCytosine-phosphate-Guanine-3'; 5mC: 5-MethylCytosine; ATAC: Assay for transposase-accessible chromatin; ChIP: Chromatin immunoprecipitation; CI: Confidence interval; CNV: Copy number variation; CpG: 5'-Cytosine-phosphate-Guanine-3'; DNMT1 (3A/3B): DNA (cytosine-5)-methyltransferase 1 (3A/3B); FISH: Fluorescence in situ hybridization; HDAC: Histone deacetylase; HR: Hazard ratio; ISS: International Staging System; LC-MS: Liquid chromatography-mass spectrometry; MM: Multiple myeloma; MAPK: Mitogen-activated protein kinases; MS: Mass spectrometry; NDMM: Newly diagnosed multiple myeloma; NF-κB: Nuclear factor-kappa B; NPC: Normal plasma cell; PC: Plasma cell; ROSE: Rank of super-enhancers; RPKM: Reads per kilobase of transcript million mapped reads; rRNA: Ribosomal RNA; TAD: Topologically associated domain; TET1 (2/3): Ten-eleven translocation methylcytosine dioxygenase 1 (2/3); TFs: Transcription factors; TSS: Transcription start site.

Acknowledgements

We would like to thank M. Devic, E. Douillard, E. Ollivier and N. Roi for their technical support. We thank G. Salbert for the critical reading of this manuscript. We thank the Biogenouest sequencing platform GenoBird from Nantes for the Illumina sequencing and computing infrastructure. We thank the medical staff of the Nouvelles Cliniques Nantaises - Le Confluent that collected normal bone marrows during hip replacement surgery.

Authors' contributions

JBA, FM, SM and AS designed the project and analyzed the data. MA, HAL, PM and SM created the cohort and provided patient samples. AS and SD produced the 5hmC-seq data. MW and TC performed the quantitative mass spectrometry measurements and interpreted the data. JBA, CG and LC analyzed the survival data. AS and JBA wrote the manuscript. All authors read and corrected the manuscript.

Funding

This study was supported by the Chaire Mixte INSERM - Ecole Centrale de Nantes, the Intergroupe Francophone du Myelome, the Ligue Contre le Cancer, the I-SITE Next (ANR-16-IDEX-0007) and the SIRIC ILIAD (INCA-DGOS-379 Inserm-12558). JBA was supported by a Ph.D. Fellowship from INSERM and Région Pays de Loire. Research of MW and TC was supported by the Deutsche Forschungsgemeinschaft (SFB1309: TP A04, SFB1361: TP 02) and the European Research Council (ERC-AG) under the European Union's Horizon 2020 research and innovation program (Project ID: EpiR: 741912).

Availability of data and materials

Sequencing data are accessible at European Nucleotide Archive under accession number PRJEB32800. Mass spectrometry data are available in additional file table. Bioinformatics code is available on request.

Ethics approval and consent to participate Consent for publication

Multiple myeloma patients were included in the IFM/DFCI cohort (NCT03679351). All patients signed an informed consent form approved by the Toulouse Ethics Committee.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Université de Nantes, CNRS, Inserm, CRCINA, Nantes, France. ² Site de Recherche Intégrée Sur Le Cancer (SIRIC) ILIAD, INCA-DGOS-Inserm 12558,

Nantes, France. ³ Hematology Department, University Hospital, Nantes, France. ⁴ Ludwig Maximilian Universität München, Munich, Germany. ⁵ Institut de cancérologie de L'Ouest Site René-Gauducheau, Saint-Herblain, France. ⁶ Centre de Recherche en Cancérologie de Toulouse, Institut National de la Santé et de la Recherche Médicale U1037, Toulouse, France. ⁷ Unit for Genomics in Myeloma, Institut Universitaire du Cancer de Toulouse-Oncopole, University Hospital, Toulouse, France. ⁸ École Centrale de Nantes, Nantes, France.

Received: 23 June 2020 Accepted: 19 October 2020

Published online: 02 November 2020

References

- Kumar SK, Dispenzieri A, Lacy MQ, et al. Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia*. 2014;28(5):1122–8.
- Moreau P, Attal M, Hulin C, et al. Bortezomib, thalidomide, and dexamethasone 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*. 2019;394(10192):29–38.
- Manier S, Salem KZ, Park J, Landau DA, Getz G, Ghobrial IM. Genomic complexity of multiple myeloma and its clinical implications. *Nat Rev Clin Oncol*. 2017;14(2):100–13.
- Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009;324(5929):930–5.
- Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science*. 2009;324(5929):929–30.
- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. 2010;466(7310):1129–33.
- Cimmino L, Dawlaty MM, Ndiaye-Lobry D, et al. TET1 is a tumor suppressor of hematopoietic malignancy. *Nat Immunol*. 2015;16(6):653–62.
- Rasmussen KD, Jia G, Johansen JV, et al. Loss of *TET2* in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev*. 2015;29(9):910–22.
- He Y-F, Li B-Z, Li Z, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*. 2011;333(6047):1303–7.
- Shen L, Wu H, Diep D, et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*. 2013;153(3):692–706.
- Globisch D, Münzel M, Müller M, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE*. 2010;5(12):e15367.
- Kohli RM, Zhang Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*. 2013;502(7472):472–9.
- Bachman M, Uribe-Lewis S, Yang X, Williams M, Murrell A, Balasubramanian S. 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem*. 2014;6(12):1049–55.
- Bachman M, Uribe-Lewis S, Yang X, et al. 5-Formylcytosine can be a stable DNA modification in mammals. *Nat Chem Biol*. 2015;11(8):555–7.
- Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol*. 2011;12(6):R54.
- Sérandour AA, Avner S, Oger F, et al. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res*. 2012;40(17):8255–65.
- Mahé EA, Madigou T, Sérandour AA, et al. Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. *Genome Res*. 2017;27(6):947–58.
- Li W, Zhang X, Lu X, et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res*. 2017;27(10):1243–57.
- Song C-X, Yin S, Ma L, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res*. 2017;27(10):1231–42.
- Pfeifer GP, Xiong W, Hahn MA, Jin S-G. The role of 5-hydroxymethylcytosine in human cancer. *Cell Tissue Res*. 2014;356(3):631–41.
- Chen K, Zhang J, Guo Z, et al. Loss of 5-hydroxymethylcytosine is linked to gene body hypermethylation in kidney cancer. *Cell Res*. 2016;26(1):103–18.
- Johnson KC, Houseman EA, King JE, von Herrmann KM, Fadul CE, Christensen BC. 5-Hydroxymethylcytosine localizes to enhancer elements and is associated with survival in glioblastoma patients. *Nat Commun*. 2016;7(1):1–11.
- Misawa K, Yamada S, Mima M, et al. 5-Hydroxymethylcytosine and ten-eleven translocation dioxygenases in head and neck carcinoma. *J Cancer*. 2019;10(21):5306–14.
- Jeschke J, Collignon E, Fuks F. Portraits of TET-mediated DNA hydroxymethylation in cancer. *Curr Opin Genet Dev*. 2016;36:16–26.
- Ficz G, Gribben JG. Loss of 5-hydroxymethylcytosine in cancer: Cause or consequence? *Genomics*. 2014;104(5):352–7.
- Agirre X, Castellano G, Pascual M, et al. Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res*. 2015;25(4):478–87.
- Salhia B, Baker A, Ahmann G, Auclair D, Fonseca R, Carpten J. DNA methylation analysis determines the high frequency of genic hypomethylation and low frequency of hypermethylation events in plasma cell tumors. *Cancer Res*. 2010;70(17):6934–44.
- Walker BA, Wardell CP, Chiecchio L, et al. Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood*. 2011;117(2):553–62.
- Heuck CJ, Mehta J, Bhagat T, et al. Myeloma is characterized by stage-specific alterations in DNA methylation that occur early during myelomagenesis. *J Immunol*. 2013;190(6):2966–75.
- Chatonnet F, Pignarre A, Sérandour AA, et al. The hydroxymethylome of multiple myeloma identifies FAM72D as a 1q21 marker linked to proliferation. *Haematologica*. 2019. <https://doi.org/10.3324/haematol.2019.222133>.
- The International Myeloma Working Group. Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *Br J Haematol*. 2003;121(5):749–57.
- Morgan GJ, Walker BA, Davies FE. The genetic architecture of multiple myeloma. *Nat Rev Cancer*. 2012;12(5):335–48.
- Jin Y, Chen K, De Paepe A, et al. Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood*. 2018;131(19):2138–50.
- Greipp PR, Miguel JS, Durie BGM, et al. International staging system for multiple myeloma. *J Clin Oncol*. 2005;23(15):3412–20.
- Song C-X, Szulwach KE, Fu Y, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol*. 2011;29(1):68–72.
- Ficz G, Branco MR, Seisenberger S, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature*. 2011;473(7347):398–402.
- Whyte WA, Orlando DA, Hnisz D, et al. Master transcription factors and Mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153(2):307–19.
- Rao SSP, Huntley MH, Durand NC, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*. 2014;159(7):1665–80.
- Saint-André V, Federation AJ, Lin CY, et al. Models of human core transcriptional regulatory circuitries. *Genome Res*. 2016;26(3):385–96.
- Magrangeas F, Avet-Loiseau H, Gouraud W, et al. Minor clone provides a reservoir for relapse in multiple myeloma. *Leukemia*. 2013;27(2):473–81.
- Magrangeas F, Kuiper R, Avet-Loiseau H, et al. A genome-wide association study identifies a novel locus for bortezomib-induced peripheral neuropathy in European patients with multiple myeloma. *Clin Cancer Res*. 2016;22(17):4350–5.
- Ross-Innes CS, Stark R, Teschendorff AE, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481(7381):389–93.
- Krönke J, Udeshi ND, Narla A, et al. Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science*. 2014;343(6168):301–5.

44. Lu G, Middleton RE, Sun H, et al. The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science*. 2014;343(6168):305–9.
45. Gu C, Cheng H, Yang H, et al. MK2 is a therapeutic target for high-risk multiple myeloma. *Haematologica*. 2018. <https://doi.org/10.3324/haematol.2017.182121>.
46. Ko M, Huang Y, Jankowska AM, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*. 2010;468(7325):839–43.
47. Gillberg L, Ørskov AD, Nasif A, et al. Oral vitamin C supplementation to patients with myeloid cancer on azacitidine treatment: normalization of plasma vitamin C induces epigenetic changes. *Clin Epigenetics*. 2019;11(1):143.
48. Traube FR, Schiffrers S, Iwan K, et al. Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. *Nat Protoc*. 2019;14(1):283–312.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



1.4. MÉTHODES SUPPLÉMENTAIRES

NORMAL PLASMA CELLS PURIFICATION

After informed consent, the femoral canal of individuals with isolated hip osteoarthritis who were otherwise healthy was probed with a metal suction device following femoral neck removal. Bone marrow cells were suctioned into a tube that contained EDTA, placed on ice and immediately transported to our laboratory. BMMCs were purified by Ficoll. Normal plasma cells were FACS-sorted using a BD FACSAria III as CD38/CD138 positive and CD3/CD13/CD33 negative (antibodies from Becton Dickinson, ref. 345807, 555332, 555394, 555450 and 562935).

MYELOMA CELLS PURIFICATION

Bone marrow biopsies were realized on MM newly diagnosed patients from cohort IFM-DFCI 2009 (Intergroupe Francophone du Myélome - Dana Farber Cancer Institute) (Attal et al., 2017). All patients signed an informed consent form approved by the Toulouse Ethics Committee. All included patients were newly diagnosed with symptomatic MM based on International Myeloma Working Group 2003 Diagnostic Criteria (The International Myeloma Working Group*, 2003). All the samples were collected in France and processed at the University Hospital of Nantes. Bone Marrow Mononuclear Cells (BMMCs) were purified by Ficoll. Plasma cells were purified with anti-CD138 beads (Robosep platform, StemCell Technologies) and the CD138-positive percentage of cells was checked by immunofluorescence microscopy.

5mC AND 5hmC DOSAGE BY MASS SPECTROMETRY

The genomic levels of 5mC and 5hmC were quantified using a mass spectrometry-based stable isotope-dilution method (Traube et al., 2019). For each LC-MS-measurement (technical replicate), 70 ng of genomic DNA (gDNA) were digested to the nucleoside level using the Nucleoside Digestion Mix (ref. M0649S) from New England BioLabs. To this reason, a solution of 70 ng gDNA in 38 μ L of milliQ-water was prepared. As heavy-atom-labeled internal standards, 1.28 pmol D₅-5mC and 0.193 pmol D₅-¹⁵N₂-5hmC in 6 μ L of milliQ-water were added to the solution, followed by 5 μ L of the Nucleoside Digestion Mix Reaction Buffer (10x), and 1 μ L of the Nucleoside Digestion Mix. After incubation for 90 min at 37°C, the mixture was filtered using an AcroPrep Advance 96 filter plate 0.2 μ m Supor from Pall Life Sciences and subsequently analysed by LC-MS. For each biological sample, two independent

measurements (technical replicates) were performed. Quantitative LC-ESI-MS/MS analysis of the enzymatically digested DNA samples was performed using an Agilent 1290 UHPLC system coupled to an Agilent 6490 triple quadrupole mass spectrometer. The UHPLC-conditions used and the settings of the mass spectrometer were the same as previously published by Traube and colleagues (Traube et al., 2019).

BLUEPRINT DATA STATEMENT

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

BIOINFORMATICS

5hmC-SEQ ANALYSIS

5hmC-seq reads were trimmed with Cutadapt v1.18 (Martin, 2011) and mapped to human reference genome GRCh38 using Bowtie2 v2.2.8 (Langmead and Salzberg, 2012) with default parameters. Duplicates and reads of low mapping quality ($Q < 30$) were discarded. 5hmC peak clusters were defined similar to Boeva and colleagues (Ashoor et al., 2013; Boeva et al., 2017) with a stitching distance of 12.5kb. Blacklisted regions consist of chromosomes X, Y, ENCODE reference blacklist and peaks called on input-seq libraries. Peak calling was done with HMCAN (Ashoor et al., 2013) with default parameters. 5hmC peaks were then merged using LILY (Boeva et al., 2017) at a distance of 12.5 kb to produce what we call 5hmC-enriched domains. The value of 12.5 kb is the same as that of Boeva and colleagues (Figure 30) and global 5hmC signal was then re-normalized (Boeva et al., 2017). We defined a set of consensus peaks clusters by merging 5hmC-enriched domains and keeping arbitrary those occupied by at least two samples across the cohort. When specified, the process was applied to the cohort and normal plasma cells together. We excluded regions shorter than 12 kb. Average 5hmC signal depth over consensus peaks clusters was computed for each sample to obtain the 5hmC score matrix $M_{i,j}$ where i is a sample and j a consensus peak cluster.

Genes and 5hmC peaks clusters were associated to the Topologically Associated Domain (TAD) of the B-cell derived cell line GM12878 they are located in similar to Boeva and colleagues (Boeva et al., 2017).

Group specific 5hmC peaks clusters were identified by computing the average Log2 Fold-Change signal per region between groups and tested with a Wilcoxon rank-sum test adjusted with a Benjamini-Hochberg correction. For paired samples (two replicates of each diagnosis and relapse), we used DiffBind (Ross-Innes et al., 2012) with default parameters to find differentially enriched 5hmC regions.

Potential core regulatory circuitries were found with CRCMapper (Saint-André et al., 2016) using default parameters.

Overlap between 5hmC-enriched domains and previously published hypermethylated CpGs in MM (Agirre et al., 2015) was computed with R and compared to overlap with random CpGs obtained from the Illumina Methy450K table in UCSC (<https://genome.ucsc.edu/cgi-bin/hgTables>).

Gene Ontology enrichment analysis was performed with GSEA (Subramanian et al., 2005) and the 100 first gene sets with FDR<0.05 were inspected. The following Gene Ontology IDs were kept for hematopoietic stem cell gene enrichment analysis: GO:0060218, GO:0061484, GO:0035701, and GO:0071425.

CHROMATIN STATES ANNOTATION WITH CHROMHMM

Functional annotation was done with ChromHMM (Ernst and Kellis, 2012) based on modified histone ChIP-seq data from the cell lines MM.1S, NCI-H929 and GM12878. Individual track of NCI-H929 was kept for statistical analysis. We downloaded the ChIP-seq data from the ENCODE portal (ENCODE Project Consortium, 2012; Davis et al., 2018) (<https://www.encodeproject.org/>). The 105 libraries identifiers used can be communicated upon request.

RNA-SEQ ANALYSIS

IFM-DFCI RNA-seq data were obtained from Cleynen and colleagues (Cleynen et al., 2017) or generated by us. RNA-seq libraries were trimmed using Cutadapt v1.13 (Martin, 2011) with parameter $-m\ 20$ and aligned to hg38/GRCh38 using STAR v2.5.3a (Dobin et al., 2013) with default parameters. Genes were quantified using featureCounts (Liao et al., 2014) from the R package Rsubread and the Gencode v28 genome annotation.

CHIP-SEQ AND ATAC-SEQ ANALYSIS

ChIP-seq and ATAC-seq libraries were downloaded from the European Nucleotide Archive (project number PRJEB25605 (Jin et al., 2018)) or from Blueprint Epigenome project and were analyzed with the pyflow-ChIPseq (Terranova et al., 2018) and pyflow-ATACseq pipelines respectively with default parameters. Chromatin-states genome annotation was realized using ChromHMM(Ernst and Kellis, 2012) on ChIP-seq data from NCI-H929 cell line.

STATISTICAL ANALYSIS

Quantitative bio-clinical variables were described with median and interquartile range (IQR) or mean and standard deviation. Significance of mean difference between groups was assessed with Kruskal-Wallis method for multiple group testing, and/or Wilcoxon test for 2 groups. Qualitative variables were described using frequency of their respective modalities. Significance of heterogeneity of the distribution between groups was assessed with χ^2 Pearson's test (or Fisher's exact test if appropriate). DESeq2 FDR testing was used for diagnosis-relapse expression and 5hmC testing. Signal correlation was assessed with a Spearman correlation. Otherwise, Pearson correlation was used. For survival analysis, time-to-event was calculated from the randomization to the event date, *i.e.* death for OS, or to the last follow-up date. For MS quantitative variables, cohorts were then split at the median value of high and low 5hmc (and 5mC). Hazard ratio between groups were calculated using Univariate Cox model. Survival curves were calculated using the Kaplan-Meier method and groups were compared using a Log-rank test. P values were corrected for multiple testing with the Benjamini-Hochberg method. Adjusted p values under 0.05 were considered significant. All calculations were done using R 3.5.0 software.

1.5. FIGURES SUPPLÉMENTAIRES

Supplementary Figure 1

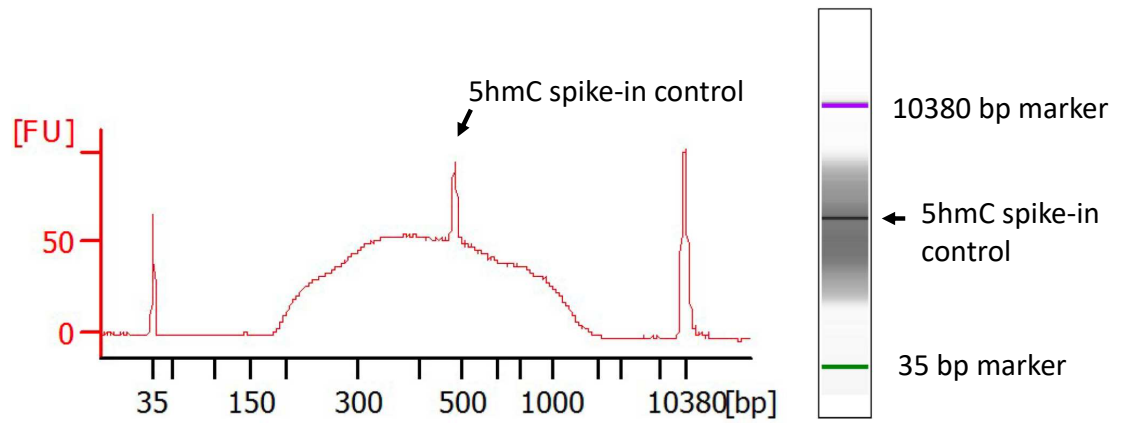


Figure 27 - Profil Agilent BioAnalyzer d'une librairie de 5hmC-seq (SCL-seq).

Supplementary Figure 2

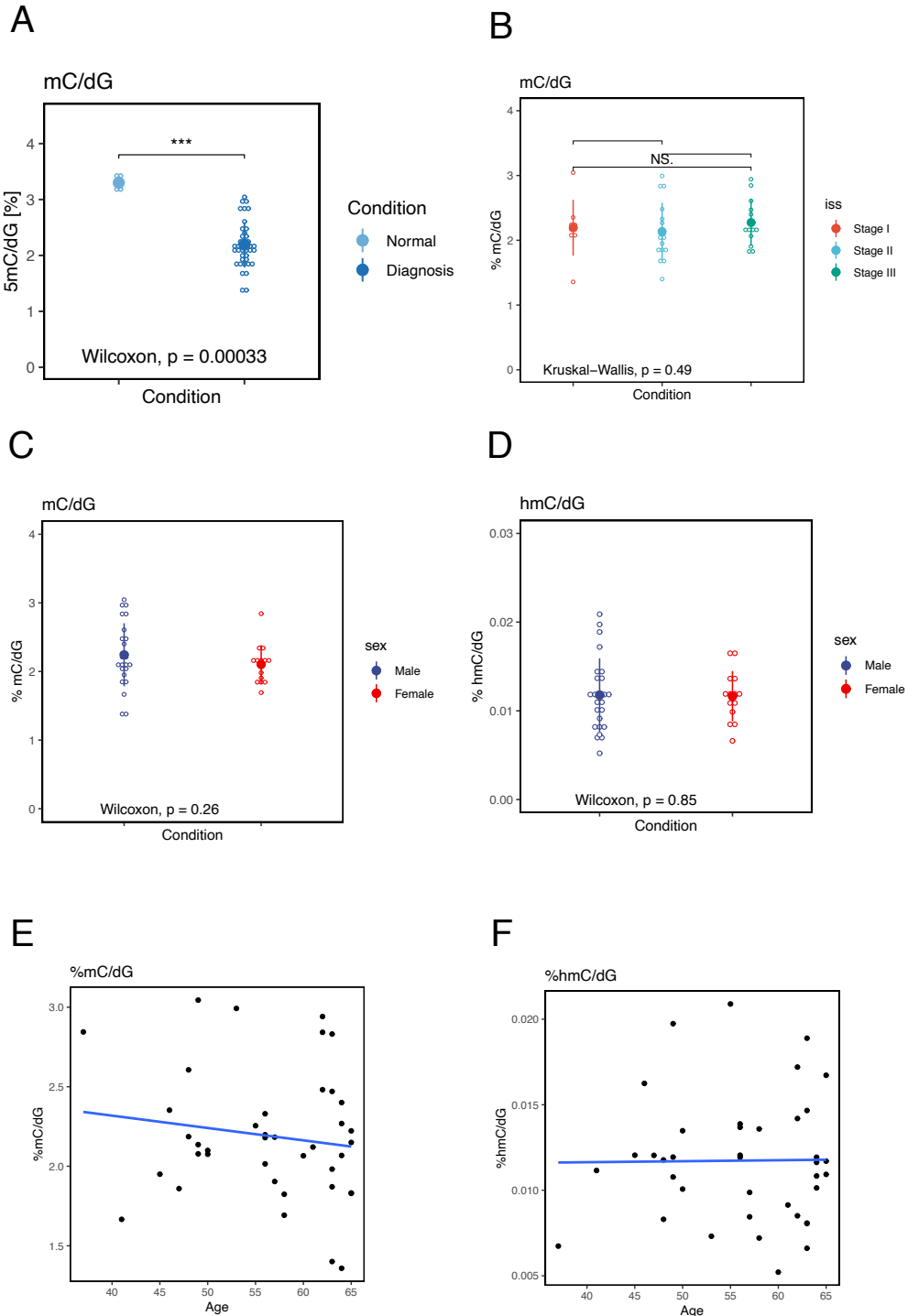


Figure 28 - MS quantification of 5mC in genomic plasma cell DNA is independent of age and sex. (A) Dot plot of 5mC global quantification by MS in normal plasma cells from healthy donors (N=5), and of myeloma cells of patients at diagnosis (N=40). (B) Dot plot of 5mC global quantification by MS by disease stage (ISS I N=9; ISS II N=17; ISS III N=13; NA=1). 5mC (C) and 5hmC (D) dot plot of MS quantification depending on the sex of the patients. 5mC (E) and 5hmC (F) dot plot of MS quantification depending on the age of the patients.

Supplementary Figure 3

A

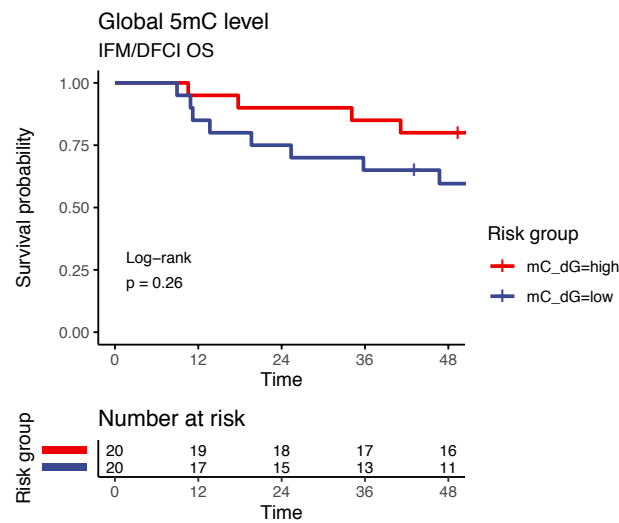
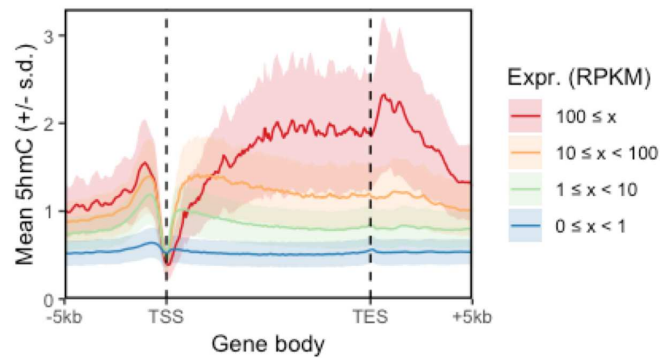


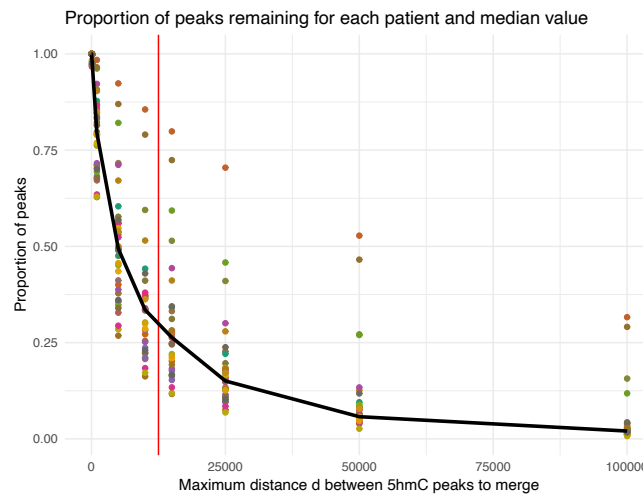
Figure 29 - Survival course depending on DNA methylation (5mC) level-based separation of two risks groups of NDMM (n=20 and 20)

Supplementary Figure 4:

A



B



C

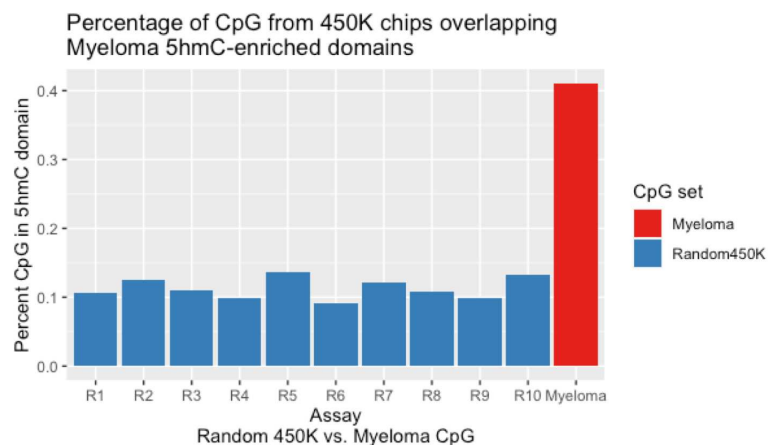


Figure 30 - 5hmC association with expression and criteria of 5hmC peaks to merge in 5hmC peak clusters. (A) Average level of 5hmC in all genes normalized to the same body length. RPKM: Reads per Kilobase Million. (B) Stitching of 5hmC into 5hmC-enriched domains. (C) Fraction of overlap between 5hmC-enriched domains of this study and CpG from the Illumina 450K chip by Agirre and colleagues (see Additional file 3: Methods).

Supplementary Figure 5

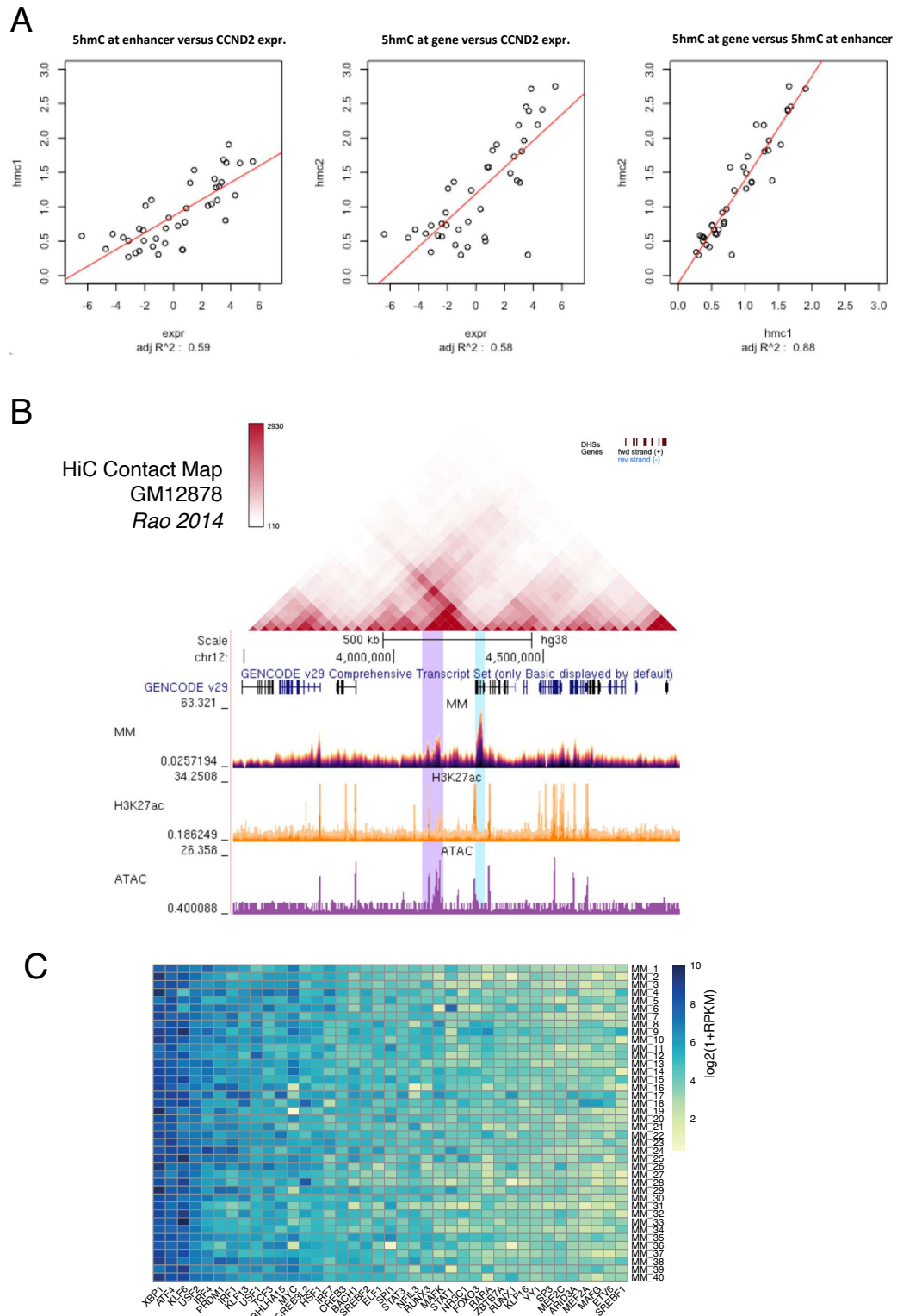


Figure 31 - 5hmC allows the identification of a putative CCND2 enhancer. (A) Correlation between CCND2 expression, 5hmC at CCND2 gene body and 5hmC at the putative 5hmC enhancer across the 40 MM patients. (B) Hi-C contact map in lymphoblastoid cells (GM12878 cell line) at the CCND2 locus showing the spatial interaction between CCND2 gene and its putative enhancer. (C) Expression of core transcription factors predicted to orchestrate core regulatory circuitries with 5hmC and RNA expression genomic data.

Supplementary Figure 6

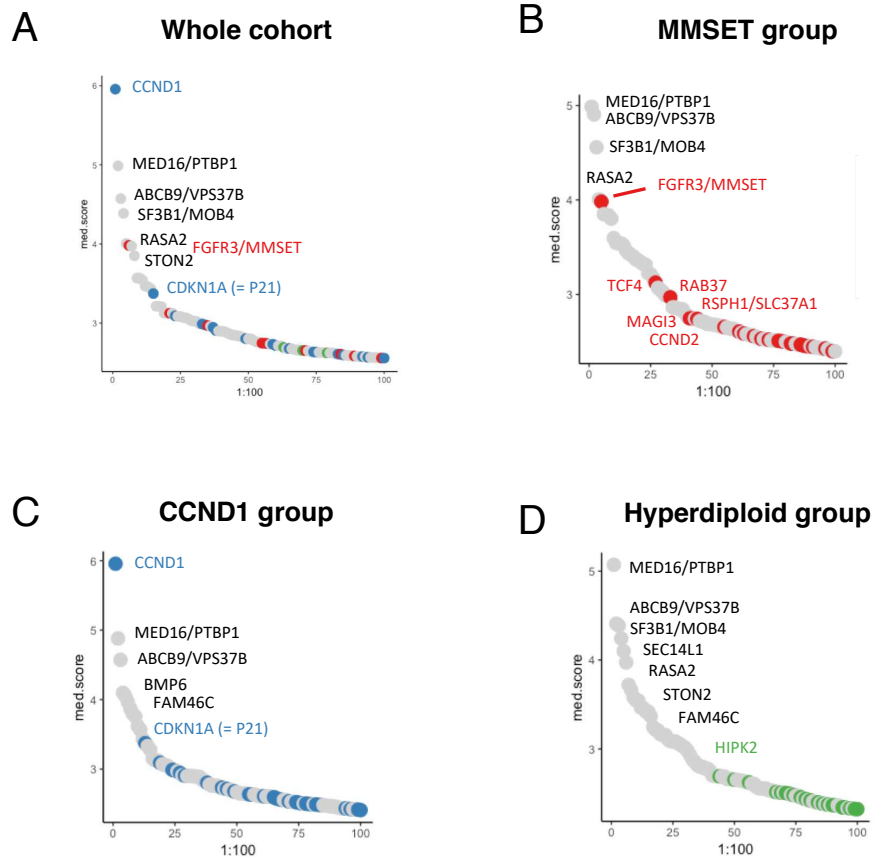


Figure 32 - Myeloma 5hmC-enriched domains associate with H3K27ac super-enhancers. Rank ordering of the 100 strongest 5hmC-enriched domains on average in the cohort (A), in the MMSET group (B), in the CCND1 group (C) and in the hyperdiploid group (D). Color highlights domains present in only one of the ROSE plots by group.

Supplementary Figure 7

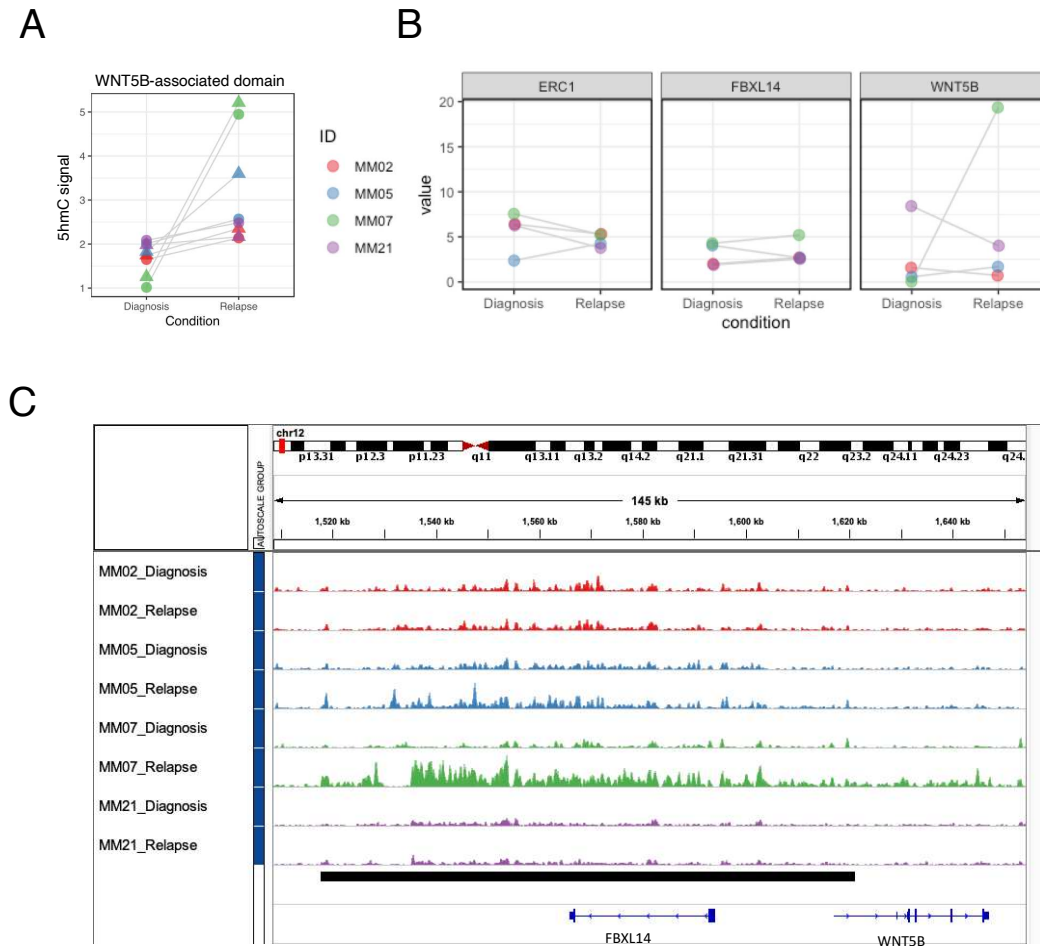


Figure 33 - 5hmC signal levels at WNT5B-associated domain are increased at relapse in MM07. (A) Normalized 5hmC enrichment at WNT5B-associated domain. Point shapes match replicates. Fold change=1.3, $p=0.003$, $FDR>0.1$. (B) Gene expression levels in RPKM measured by RNA-seq at diagnosis and relapse for three genes surrounding the WNT5B-associated domain. (C) 5hmC genomic signal around WNT5B-associated domain. Colors match those of (A) and (B). 5hmC domain is depicted under signal tracks (hg38: chr12:1,517,750-1,621,200).

2. RÉSISTANCE ACQUISE AU VENETOCLAX ET ANOMALIES GÉNOMIQUES SÉLECTIONNÉES DE LA FAMILLE DU BCL2

Ce travail a été réalisé à l'Université de Calgary au Canada en collaboration avec l'équipe de Dr Nizar Bahlis et Dr Paola Neri. Il a fait l'objet d'une présentation orale au congrès de l'ASH 2019 (Alberge et al., 2019). J'ai réalisé les analyses de données de séquençage single-cell avec l'équipe canadienne. Les manipulations expérimentales ont été faites par Dr Ranjan Maity et Dr Paola Neri. Le docking *in silico* du venetoclax sur le mutant de BCL2 a été réalisé par des collaborateurs de l'Université de Calgary. J'ai écrit l'article avec les Dr Bahlis et Neri. L'article va faire l'objet d'une soumission à un journal d'onco-hématologie.

2.1. RÉSUMÉ EN FRANÇAIS

L'inhibition spécifique de la protéine BCL2 avec le venetoclax / ABT-199 induit des réponses importantes chez des patients présentant un myélome multiple en rechute ou réfractaire (RRMM) et la translocation t(11;14) entre *CCND1* et la chaîne lourde du locus des immunoglobulines (*IGH*). La résistance acquise au venetoclax est courante, mais les mécanismes moléculaires qui y sont associés sont inconnus. Nous avons caractérisé les plasmocytes de 4 RRMM avant le début du traitement par le venetoclax et au moment de l'apparition de la résistance par trois technologies de séquençage en cellule-unique : le transcriptome (scRNA-seq), le nombre de copies (scCNV-seq) et l'accessibilité de la chromatine (scATAC-seq).

Nos collaborateurs à l'Université de Calgary ont découvert et caractérisé fonctionnellement un nouveau mutant de BCL2 (p.D111A) qui confère une résistance au venetoclax *in vitro*. Nous avons identifié une sélection récurrente clonale et sous-clonale des plasmocytes présentant des gains et amplifications de *MCL1* et associés à une augmentation de l'expression génique et de l'accessibilité de la chromatine à ce locus ; nous présentons des translocations originales de *BCLxL* avec les loci *IGH* et *IGL* également associées à un gain d'expression du gène et à une ouverture de la chromatine aux points de cassure.

Cette étude est la première à caractériser les mécanismes de l'acquisition de la résistance au venetoclax chez des patients MM et à l'échelle de la cellule-unique. Elle souligne l'importance de la détection précoce de clones dynamiques minoritaires et hautement résistants associés à un test fonctionnel de sensibilité au venetoclax pour améliorer l'efficacité du traitement et la survie globale des patients.

Mutations and Copy Number Gains of BCL2 Family Members Mediate Resistance to Venetoclax in Multiple Myeloma Patients

Jean-Baptiste Alberge^{1,2,*}, Ranjan Maity^{3,*}, Jonathan Cruard^{1,2}, Sarthak Sinha³, John McIntyre⁴, Madison Kong³, Fajer Hasan³, Florence Magrangeas^{1,2,5}, Stéphane Minvielle^{1,2,5}, Paola Neri^{3,4,‡}, Nizar J. Bahlis^{3,‡}

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

² Site de Recherche Intégrée sur le Cancer (SIRIC) « ILIAD », INCA-DGOS-Inserm_12558, Nantes, France

³ Arnie Charbonneau Cancer Research Institute, University of Calgary, Calgary, AB, Canada

⁴ Precision Oncology Hub, Tom Baker Cancer Centre, Calgary, AB, Canada

⁵ Hematology, University Hospital Hôtel-Dieu, Nantes, France

* Co-First authors

‡ Co-Last authors

2.2. ABSTRACT

Selective BCL2 inhibition with venetoclax induces deep responses in relapsed multiple myeloma (MM) patients. However, acquired resistance to venetoclax frequently occurs and the molecular mechanisms underlying this phenomenon remain unknown. Herein, we have profiled through single cell multi-omics approaches (scRNAseq, scCNVseq and scATAC-seq) BM aspirates collected from (n=4) MM patients treated with venetoclax prior starting therapy and at the time of acquired drug resistance.

We have discovered and functionally characterized a novel BCL2 mutation (D111A) that confers *in vitro* resistance to venetoclax. In addition, we have identified an enrichment of MM clones with *MCL1* locus copy number gain as well as a *BCL2L1* up-regulation and dependency as additional mediators of drug resistance.

Therefore, early detection and dynamic monitoring of these abnormalities (BCL2 mutant, BCL2L1 or 1q gains) with early therapeutic interventions targeting these branching clones may enhance venetoclax efficacy and improve survival in venetoclax treated MM patients.

STATEMENT OF SIGNIFICANCE

This is the first study to describe at single cell level the molecular mechanisms leading to venetoclax resistance in Myeloma. Our data emphasize that early detection of BCL2 abnormalities associated with the analysis of clonal dynamics and BH3 mimetic dependence are essential for the clinical management of venetoclax-treated patients.

2.3. INTRODUCTION

The members of the B cell leukaemia/lymphoma 2 (BCL2) family play a crucial role in the regulation of the intrinsic apoptotic pathway (Singh et al., 2019). They are classified into three groups according to structural and functional characteristics: anti-apoptotic members (i.e., MCL1, BCL2 and BCLxL), multidomain pro-apoptotic members (BAX and BAK), and BH3-only members (i.e., BIM, BID, BAD or NOXA) (Youle and Strasser, 2008). The overexpression of the anti-apoptotic proteins (BCL2, BCLxL, MCL1) ranks amongst one of the hallmarks of cancer that favors tumor cell growth and induces resistance to therapy (Touzeau et al., 2018).

Venetoclax is the first clinically available oral drug that targets specifically the anti-apoptotic BCL2 protein, displacing the pro-apoptotic proteins from their BH3 groove and leading to cell death (Souers et al., 2013). It has demonstrated significant anti-tumor activity in several BCL2-dependent hematological malignancies including chronic lymphocytic leukemia (CLL), mantle cell lymphoma (MCL) and follicular lymphoma (FL) (Roberts et al., 2016; Davids et al., 2017). In multiple myeloma (MM), sensitivity to venetoclax was first shown to be restricted to human cell lines and primary cells harboring the t(11;14) chromosomal rearrangement involving *CCND1* and the heavy locus of immunoglobulin (Touzeau et al., 2014, 2017, 2018; Kumar et al., 2017b). However, alone and in combination with proteasome inhibitors, venetoclax has also demonstrated ability to induce deep response in heavily pretreated MM patients, independent of t(11;14) status (Touzeau et al., 2014; Kumar et al., 2017b). The presence of the t(11;14) translocation is strongly associated with sensitivity to venetoclax and has been correlated with high *BCL2/BCL2L1* (encoding BCLxL) and *BCL2/MCL1* mRNA ratios (Kumar et al., 2017b). In addition with BCL2 members expression profile, functional tests as BH3 profiling or *in vitro* BH3 mimetic drug testing are also important to predict sensitivity to venetoclax (Morales et al., 2011; Matulis et al., 2019).

Despite significant clinical responses observed in relapsed/refractory multiple myeloma (RRMM) patients, sub-optimal responses and acquired resistance to venetoclax remain to be overcome. Recent reports on acquired resistance to venetoclax describe acquisition of Gly101Val mutation in *BCL2* (Blombery et al., 2019), as well as mutations in *BTG1* and homozygous deletions of *CDKN2A/B* (Herling et al., 2018) as causes of resistance in CLL. In FL, the Phe104Ile mutation in *BCL2* was also described as causes of venetoclax resistance together with copy number gain of *MCL1* or up-

regulation of *BCL2L1* (Blombery et al., 2019; Guièze et al., 2019). Of note, all *BCL2* mutations identified so far are described to occur at low allele frequencies and to be sub-clonal.

The mechanisms underlying the dependency of the t(11;14) subgroup of MM patients on *BCL2* remain to be elucidated and the mechanisms of resistance to *BCL2* inhibition with BH3 mimetic venetoclax are still unknown.

Given the extensive inter-patient genomic heterogeneity of MM patients (Manier et al., 2017a), and the fact that relapse is a result of a dynamic interplay of evolutionary processes, such as acquisition of new mutations, copy number changes, and selective pressure due to treatment used (Rasche et al., 2019), we have used a multi-omics single cell approach to investigate the acquired resistance to venetoclax in RRMM patients. As such, we have performed single cell mRNA expression (scRNAseq), copy number profiling (scCNVseq) as well as chromatin accessibility (scATAC-seq) on primary MM cells collected from patients treated with venetoclax prior to initiation of therapy and at disease progression. We identified and functionally characterized a novel *BCL2* mutation (Asp111Ala) that confers resistance to venetoclax. In addition, we reported an enrichment of MM clones with *MCL1* locus copy number gain as well as a *BCL2L1* up-regulation and dependency as additional mediators of drug resistance. Collectively, our data suggest that dynamic monitoring of emergent clones with acquired *BCL2* family members alterations together with characterizing the BH3 mimetic dependence of patient cells *ex vivo* could be used to design better treatment strategies to enhance venetoclax efficacy and improve survival in venetoclax treated MM patients.

2.4. MATERIALS AND METHODS

PATIENT SAMPLES

Experiments were conducted on primary patient samples provided by the Tom Baker Cancer Center, after informed consent, in accordance with the Declaration of Helsinki and following the approval by the Medical Center Institutional Review Board.

CELLS CULTURE AND CD138⁺ MM CELLS

KMS12PE cells were maintained in RPMI 1640 media with 10% FBS (LifeTechnology Canada, North York, ON), 100 U/ml penicillin, and 100 mg/ml streptomycin (LifeTechnology). Primary CD138⁺ MM cells were obtained from the BM aspirates of MM patients at the time of diagnostic procedure, using positive selection with CD138 microbeads (Miltenyi Biotech, Auburn, CA) and accordingly to manufacturer's instructions. CD138⁺ fractions were subsequently analyzed by flow cytometry in order to confirm the sample's purity. HEK293T cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS, 100 U/mL penicillin and 100 µg/mL streptomycin. All of the listed cells above were cultured at 37°C and 5% CO₂.

CELL VIABILITY

KMS12PE cells were seeded in 96-well plates at a total of 5×10^5 cells/mL and cells were treated with venetoclax alone, S63845 alone, and a combination of both venetoclax and S63845 with increasing concentrations ranging from 50 nM to 200 nM. The cellular proliferation was analyzed through 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) dye absorbance using CellTiter96TM AQueous Dye (Promega). The MTT assay data was analyzed using a SpectraMax iD3 Microplate Reader instrument (ThermoFisher Scientific).

ASSESSMENT OF APOPTOSIS

For Annexin/propidium iodide (PI) staining, an Annexin V apoptosis kit was used as per the manufacturer's instructions (BD Pharmingen). For patient samples, MM cells were subjected to 24-48 hours of drugs treatment and stained with CD138

antibody for flow cytometry by using BD FACS (BD Bioscience). MM cells were gated as CD138^{hi}.

SITE-DIRECTED MUTAGENESIS

pCDH-puro-BCL2 plasmid (#46971) was purchased from Addgene. BCL2-D111A point mutation was introduced by site directed mutagenesis by using forward primer 5'CGCCGCTACCGCCGCGCCTTCGCCGAGATGTCC3', reverse primer 5'GGACATCTCGGCGAAGGCGCGGCGGTAGCGGCG 3', and Quickchange II site directed mutagenesis kit from agilent (#200523).

LENTIVIRUS PRODUCTION AND VIRAL TRANSDUCTION

BCL2 lentiviral constructs were packaged into HEK293-LentiX. Cells were seeded 1 day prior to transfection and next day cells were transfected at 70%-80% confluency. For 10cm dish, 30 μ g of pCDH-puro-BCL2WT or BCL2D111A plasmid, 10 μ g of envelope plasmid pMD2.G and 20 μ g of packaging plasmid psPAX2 were transfected by using Calcium phosphate transfection reagent (Life Technologies). 48h and 72h post transfection supernatant was collected and filtered with 0.45 μ m filter (Millipore).

1 $\times$ 10⁵ KMS12PE cells were seeded in 1ml complete RPMI media in 24well plate and incubated with 6 μ g/ml of Polybrene (Sigma) followed by BCL2WT or BCL2D111A lentiviral infection. 12h after infected cells were washed and incubated with fresh media and 48h after infection, cells were cultured in medium containing 1 μ g/ml Puromycin for another 72h-96h. BCL2 expression was confirmed by Western Blot.

QUANTITATIVE WESTERN BLOT

Cells were grown in their respective media before total protein lysates were extracted in in-ice cold radioimmunoprecipitation assay (RIPA) lysis buffer [50mM Tris, 5M NaCl, 0.5mM EDTA, 0.5 mM EGTA and 1% Triton X-100, 1% NP40, 20% SDS containing CompleteTM EDTA-free protease inhibitor (Roche)]. The purified cellular lysates were loaded in sodium dodecyl sulfate (SDS)-gels (NuPage 4 – 12% bis – Tris gel, Invitrogen) followed by electroblotting into nitrocellulose 0.45 μ m membranes (BioRad, Canada). Western blot analysis was performed according to standard protocol with antibodies to MCL1, BCL2, BCLxL, and Actin. Membranes were imaged and band densitometry was performed using ImageJ. The ratio of band intensity of MCL1, BCL2, and BCLxL was calculated relative to loading control (Actin). Antibodies

were obtained from the following sources: MCL1 (Cell Signaling Technology), BCL2 (BD Bioscience, R&D Systems), BCL-xL (Cell Signaling Technology, R&D Systems), and Actin (Sigma-Aldrich).

CO-IMMUNOPRECIPITATION ASSAY

KMS12PE MCL1-V5 OVEREXPRESSION, BIM IP

pLX307-Mcl1-V5 (Addgene117726) lentiviral plasmid was purchased from Addgene and lentiviral particle was produced as previously described. KMS12PE cells were infected with lentiviral particles and stably expressed MCL1-V5. Cells were treated with either Venetoclax or S63845 or combination. 6hr following treatment cells were lysed with ONYX lysis buffer [20 mM Tris (pH 7.4), 135 mM NaCl, 1 mM EDTA, 1% Triton, 10% glycerol, 1.5 mM magnesium dichloride and CompleteTM EDTA-free protease inhibitor (Roche)] and immunoprecipitated with anti-BIM antibody followed by immunoblotted for BCL2 and MCL1.

KMS12PE BCL2 D111A OVEREXPRESSION, BIM IP

KMS12PE cells were infected with empty vector or pCDH-Puro-BCL2 or pCDH-Puro-BCL2D111A lentiviral particle and selected against antibiotic for stable expression. KMS12PE cells expressing BCL2WT or BCL2D111A were treated with venetoclax with 50nM or 100nM for 6hrs. Cells were subjected to lysis and immunoprecipitated for BIM and immunoblotted for BCL2, MCL1 and BIM respectively.

DROPLET DIGITAL PCR

A custom TaqMan® genotyping assay was designed to target *BCL2* NM_000633.2: c.332T>G (p.D111A) variant. The assay consisted of the following: forward primer 5'-GCAGCTGGCTGGACATCT-3', reverse primer 5'-GCCGGCGACGACTTCT-3', VIC labelled WT probe 5'-CCGCGACTTCGC-3' and FAM labelled mutant probe 5'-CCGCGCCTTCGC-3'. Each PCR reaction contained 7.5µL 2x QuantStudioTM 3D digital PCR master mix v2, 0.75µL 20x *BCL2* SNP assay, 4.75µL PCR grade water and 20ng of genomic DNA to a total volume of 15µL and loaded onto a QuantStudioTM 3D digital PCR 20K chip and sealed (ThermoFisher Scientific, Waltham MA). Digital PCR chips were run on a ProFlex 2x Flat PCR thermocycler using the following cycling conditions: 96°C for 10 minutes, then cycled at 60°C for 2 minutes and 98°C for 30

seconds x39 cycles, followed by a final 60°C elongation step for 2 minutes then 10°C hold. Chips were imaged on the QuantStudio™ 3D Digital PCR instrument and results analyzed using QuantStudio 3D Analysis Suite Software™ version 3.1.6 (ThermoFisher Scientific).

SINGLE-CELL PROTOCOLS

Unbiased mRNA and DNA profiling were conducted by using scRNA-seq, single cell ATAC-seq and sc-CNV-seq from the GemCode system (10x Genomics) accordingly to the manufacturer's protocols. In details, for single-cell library preparation for RNA-sequencing primary MM cells were processed according to 10X Genomics reagent Kits User Guide (CG00052 v2 Chemistry). Cells were partitioned into nanoliter-scale Gel Bead-In-Emulsions (GEMs) using 10X GemCode Technology. Primers containing (i) an Illumina R1 sequence, (ii) a 16 bp 10x barcode, (iii) a 10 bp Unique Molecular Identifier (UMI) and (iv) a poly-dT primer sequence were incubated with partitioned cells resulting in barcoded, full-length cDNA from poly-adenylated mRNA. Silane magnetic beads were used to remove leftover biochemical reagents/primers, then cDNA was amplified by PCR. Enzymatic Fragmentation and Size Selection was used to optimize cDNA amplicon size prior to library construction. R1 (read 1 primer sequence) were added during GEM incubation, whereas P5, P7, a sample index (i7), and R2 (read 2 primer sequence) were added during library construction via End Repair, A-tailing, Adaptor Ligation and PCR.

For single-cell library preparation for ATAC-seq the desired number of nuclei were targeted and processed according to 10X Genomics Reagent Kits User Guide (CG000168). The Nuclei Isolation was performed as indicated in the nuclei Isolation protocol for Single Cell ATAC Sequencing (10x Genomics). Based on the starting number of cells and desired final nuclei concentration primary MM cells were washed, lysed and re-suspended in appropriate volume of chilled Diluted Nuclei Buffer. The resulting nuclei were then immediately used to generate scATAC-seq libraries. ScATAC-seq libraries were prepared according to 10X Genomics Reagent Kits User Guide. Briefly, the desired number of nuclei were combined with ATAC Buffer and master mix to form transposed Nuclei. Single-cell GEMs were then generated, amplified and subjected to bead clean-ups. Indexed sequencing libraries were constructed using Chromium i7 Sample Index and the barcode sequencing libraries subjected to a final bead clean-up prior to quantification.

For single-cell DNA library generation for CNV single-cell suspensions of primary MM cells were processed according to 10X Genomics Reagent Kits User Guide (CG000153). Single cells were partitioned in a hydrogel matrix by combining with a CB polymer to form Cell beads (CBs) using a microfluidic chip. Post a first encapsulation, CBs were treated to lyse the encapsulated cells and denature the genomic DNA (gDNA). The denatured gDNAs in the CB were then accessible to amplification and barcoding. A second microfluidic encapsulation step was then required to partition the CB with 10x barcode Gel Beads (GBs) to generate GEMs. Immediately after barcoding and amplification, 10x barcoded fragments were pooled and attached to standard Illumina adaptors.

For all the single cell methods, quality control and quantification was performed using a Kapa Library Quantification qPCR kit (Kapa Biosystems) on a BioRad qPCR instrument prior to preparing a single pool containing equal amounts of each library. This pool was then subjected to on-board cluster formation and sequencing on an Illumina NextSeq 500 sequencer with a high-output v2.5 150 sequencing kit for RNA-seq and ATAC-seq and 300 cycle sequencing kit for CNV-seq as per the standard Illumina protocols. After sequencing, the bcl data was converted to fastq data files using the Illumina BCL2FASTQ utility. Samples were processed with CellRanger suite v3.0 and downstream analyses were realized with Seurat, Monocle, Signac, and Cicero R packages.

STATISTICS AND BIOINFORMATICS

SINGLE-CELL DATA PROCESSING

Genomic sequence reads were treated with the CellRanger suite (cellranger v3.0.2, cellranger-atac v1.1.0, cellranger-dna v1.1.0 for scRNA-seq, scATAC-seq and scCNV-seq respectively) against the human reference genome GRCh38 with default parameters.

For gene expression data, samples were merged and cell barcodes with less than 1,000 UMIs were discarded. Plasma cells were selected based on immunoglobulin genes expression (UMI>100) as well as negative expression of *CD3D/G/E*, *CD14*, *FCGR3A* and *APOE*. SCRAN was used for normalization and the CellCycle Scoring function of Seurat (Butler et al., 2018; Stuart et al., 2019) was used to regress cell cycle genes at data standardization, during which immunoglobulin genes were discarded as described by Lederger and colleagues (Lederger et al., 2018). Differential gene

expression testing was run with MAST with default parameters (Finak et al., 2015). For paired analyses, matrices were randomly downsampled to achieve equal mean of UMI per cell in pre and post venetoclax samples.

For copy number variation data, samples were aggregated with custom R scripts available upon request and were further binned through a 2Mbp window. For readability purposes and when specified, copy number estimates greater than 4 were reduced to 4 and are marked “≥4”.

For chromatin accessibility data analysis of MM02 cells, samples were aggregated with the cellranger aggr function without normalization. Count matrices were filtered with Signac and custom R scripts. Single cells were kept if they matched all these requirements: at least 20 fragments in immunoglobulin regions, more than 1000 and less than 35000 fragments in peaks, more 30% of reads in peaks, less than 0.05% reads in ENCODE’s blacklist and a nucleosome signal <5%. Gene activity quantification was done with Signac and Ensembl v86 annotation for gene coordinates extended 2000bp of the gene body. Gene activity was then log-normalized. Pseudotime analysis was done with Cicero for Monocle3 package (Pliner et al., 2018) with LSI preprocessed method, UMAP was run with 100 neighbors, and clustering was run with Signac (Stuart et al., 2019) and a 0.2 resolution parameter for SNN clustering on peaks.

TRANSLOCATION DISCOVERY

Translocations were determined using manta v1.6.0 (Chen et al., 2016) in tumor only mode with default parameters. VCF output files were processed with custom R scripts available upon request. For CNV (read length=2x150), only translocations with split-reads involved were kept. Translocations are supported by at least 5 paired-read or 3 split-reads, with a minimum variant allele frequency of 10%, and a homology and insertion length of less than 5bp. Translocations were further filtered based on mappability: both breakpoints and 1kb around must strictly lay inside mappable regions according to CellRanger’s and DELLY (Rausch et al., 2012) mappable regions for HG38 genome. K-mer mappability was also checked and translocations with less than 50% mappability in 24mer UMAP track were filtered out (Karimzadeh et al., 2018). Finally, translocations were manually curated with IGV (Robinson et al., 2011) and a minimum mapping quality of 40. ATAC reads were also used to validate *CCND1* and *BCL2L1* translocations.

STATISTICAL ANALYSIS

Quantitative bio-clinical variables were described with median and interquartile range (IQR) or mean and standard deviation. Significance of mean difference between groups was assessed with Kruskal-Wallis method with Dunn post-hoc for multiple group testing, or Wilcoxon test for 2 groups. Qualitative variables were described using frequency of their respective modalities. Significance of heterogeneity of the distribution between groups was assessed with χ^2 Pearson's test (or Fisher's exact test if appropriate). Difference of gene expression level between groups of cells is expressed in average log fold change (avgLogFC) and significance of the difference is computed with MAST package and R 3.6.0 and adjusted with the Bonferroni procedure.

2.5. RESULTS

MOLECULAR PROFILING AND CLINICAL CHARACTERISTICS OF VENETOCLAX-TREATED PATIENTS

As shown in Figure 34,A, bone marrow (BM) aspirates (n=8) were collected from four RRMM patients treated with venetoclax prior to initiation of therapy and at disease progression. Following magnetic sorting, a total of 78,000 CD138⁺ cells from paired samples were subjected to unbiased mRNA and DNA profiling by performing single-cell RNA, ATAC, and copy number sequencing using the GemCode system (10x Genomics). The clinical characteristics of the patients enrolled on this study are shown in Table 1. All patients received venetoclax and dexamethasone at time of MM relapse and achieved PR and VGPR. The clinical responses and duration of response data of each patient are presented in Figure 34,B. Of note, two patients had a duration of response of 12 months. All patients harbor the t(11;14) translocation between CCND1 and the immunoglobulin heavy chain (IgH) locus. The molecular profile of one the patients (MM04) is shown in Figure 34,C. Of note, the hematoxylin and eosin stain (H&E) for CD138 confirmed the presence of MM cells and the over-expression of BCL2. The FISH using probes for the CCND1 gene (red) and IGH locus (green), and red and green fusion demonstrated the t(11;14) corresponding to IGH-CCND1 rearrangement.

Shown in Figure 34,D is the UMAP clustering of CD138^{pos} myeloma cells based on the cells mRNA transcript levels in one of our patients with t(11,14) disease (MM04). The clonality of the plasma cells is shown by the restriction of *lambda* light chain expression vs *kappa* (Figure 34,E). Of note, high expression of CCND1 and BCL2 mRNA levels and not BCL2L1 (encoding BCLxL) were observed in this particular patient before venetoclax therapy.

Figure 1

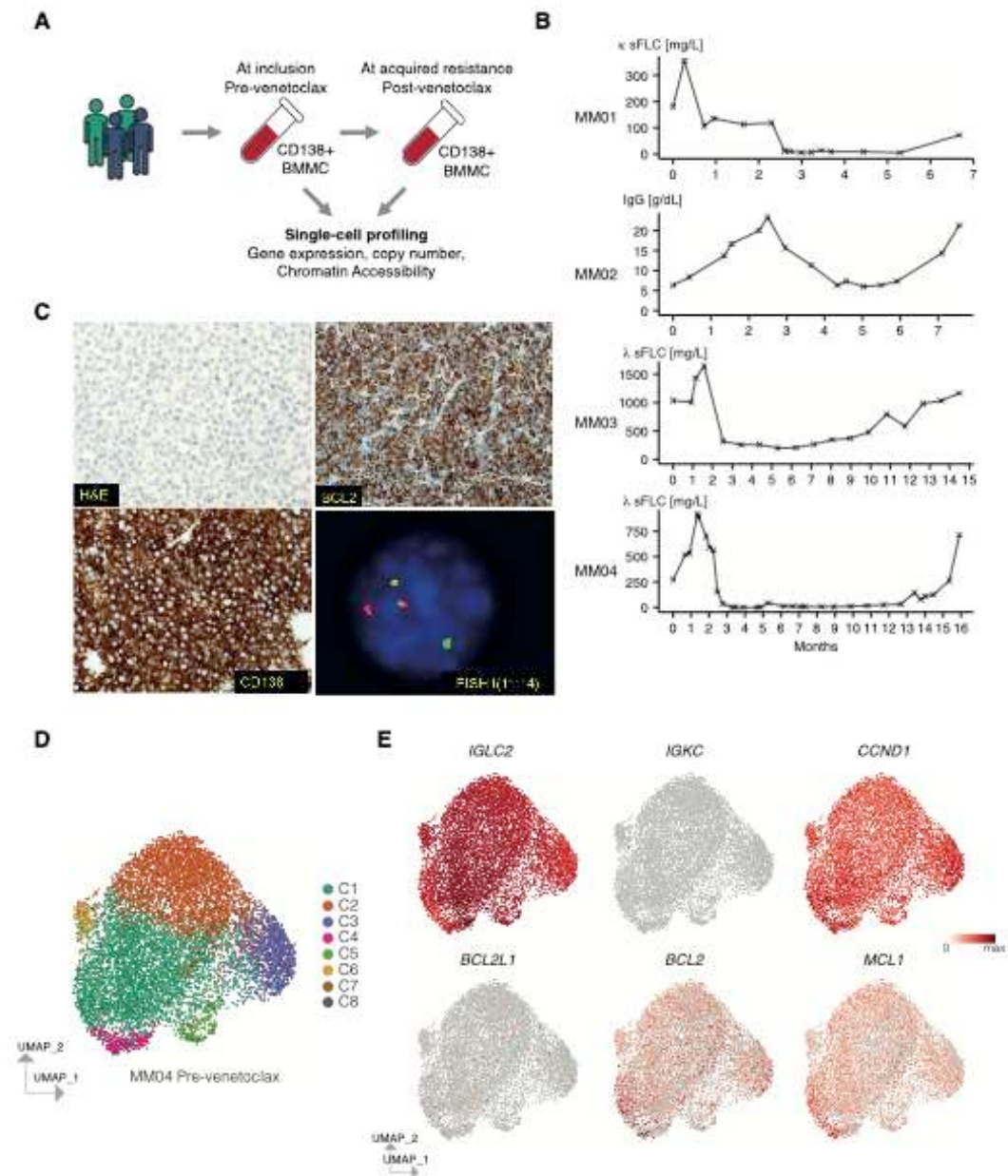


Figure 34 - Molecular profiling and clinical characteristics of venetoclax-treated patients

A, Overview of the experimental design. Paired BM samples were collected from 4 patients treated with venetoclax prior starting therapy and at the time of progression., A total of 78,000 single-cells were profiled through gene expression, copy number, and chromatin accessibility technologies. B, The clinical responses obtained by measuring serum *kappa* or *lambda* Free Light Chain (sFLC) or monoclonal protein over time during venetoclax therapy and duration of response data of each patient are presented here. C, Hematoxylin and eosin stain (H&E 400x) for CD138 and Bcl2. The FISH results using probes for the CCND1 gene (red) and IGH gene (green), and red and green fusion confirm the presence of the t(11;14). D-E, Uniform manifold approximation and projection (UMAP) to visualize single-cell gene expression data for sample MM04 pre-treatment. 8,598 cells were profiled for gene expression. The color represents unsupervised clustering (panel D) or expression level (panel E). Scale is linear from 0 (grey) to the maximum expression value (dark red) of each gene. The clonality of the plasma cells is shown by the restriction of *lambda* light chain expression vs *kappa*. High expression of CCND1 and BCL2 and not BCL2L1 (encoding BCLXL) was observed in this particular patient.

DYSREGULATION OF BCL2 FAMILY MEMBERS AND ACQUIRED RESISTANCE TO VENETOCLAX

Since members of the BCL2 family are critical regulators of apoptosis, and interactions between pro-survival and pro-apoptotic members determine cell fate, we next evaluated at single cell level the mRNA changes of BCL2 members observed in our t(11;14) patients pre-venetoclax and at the time of venetoclax resistance. As shown in the violin plots presented in Figure 35,A, we identified numerous changes in both pro-apoptotic and anti-apoptotic family members, including *BCL2*, *MCL1*, *BCL2L1*, *BCL2L11* (encoding BIM), *PMAIP1* (encoding NOXA), and *BBC3* (encoding PUMA) at the time of progression and venetoclax resistance. *MCL1* increased expression was seen in all patients, and in some (n=2) an increased expression of the other anti-apoptotic gene *BCL2L1* was observed. Interestingly, this *BCL2L1* upregulation was seen in the two patients with a low increase of *MCL1*. Minor changes were observed in other pro-apoptotic BCL2 family members such as *PMAIP1*, *BBC3* and *BCL2L11*. However, these changes were neither mutually exclusive nor universal across the samples tested.

In addition, single-cell genomic copy number profiling of the same genes reveals also a heterogeneous and subclonal evolution at time of acquired resistance (Figure 35,B). Clones with more than 4 copies of *BCL2* and the proximal *PMAIP1* on chromosome 18q21 were lost during treatment in patient MM01, while copies of 18q21 remained at constant frequencies in the other ones. Genomic copies of *MCL1* were gained in two samples at time of acquired resistance and associated with an upregulation of the anti-apoptotic gene. *BCL2L1* was gained in another patient (MM03) and associated with its upregulation, while two other patients partially lost additional copies of *BCL2L1*. Additional recurrent copy number variations included the loss of sub-clones with three copies of *BBC3* in two patients with no changes in gene expression. Taken together, these data highlight the complexity and heterogeneity of the potential mechanisms of venetoclax resistance identified in individual MM patient.

Figure 2

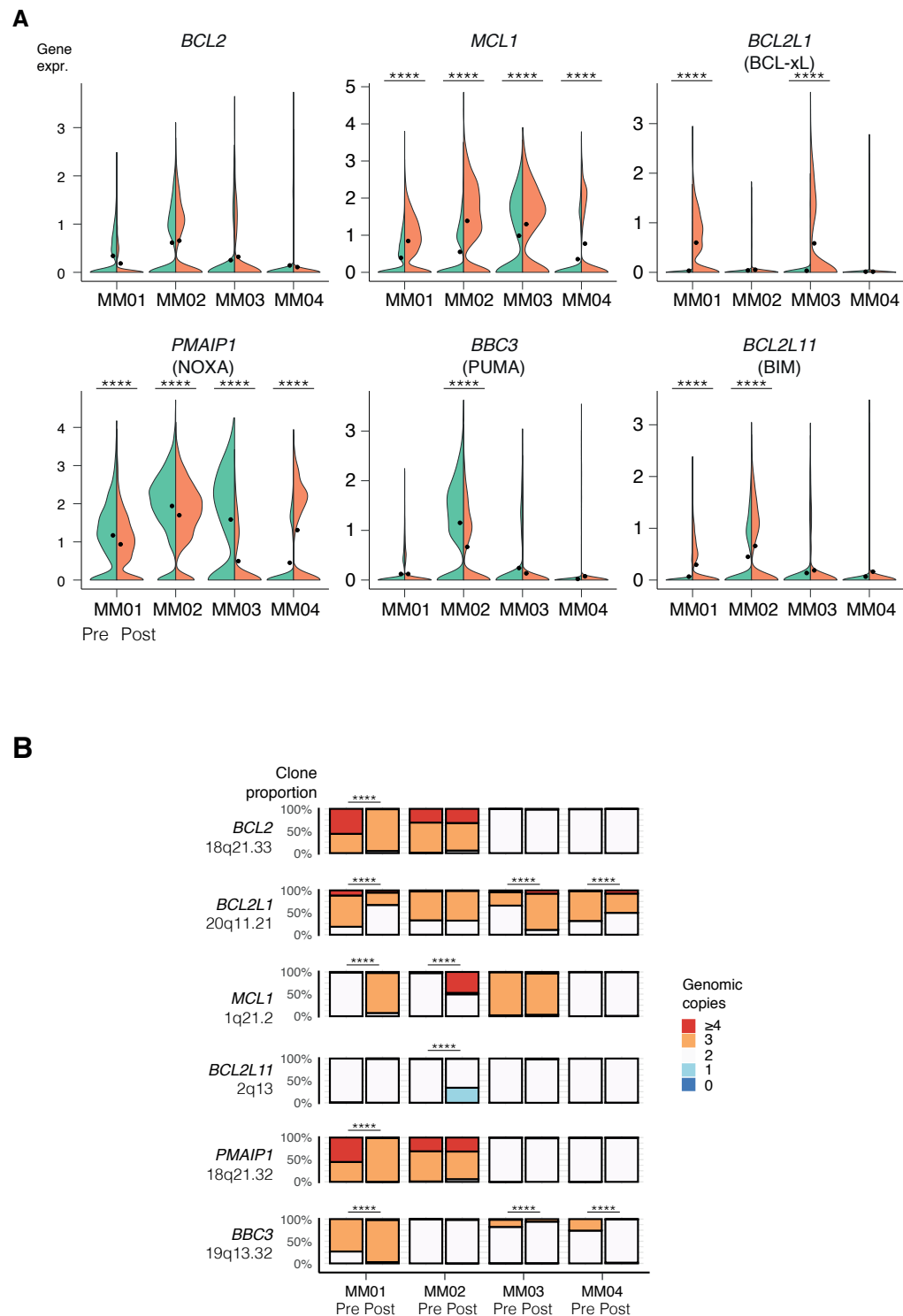


Figure 35 - Dysregulation of BCL2 family members and acquired resistance to venetoclax. A, These violin plots summarize at single MM cell level the mRNA expression of six anti- and pro-apoptotic proteins in individual t(11;14) patients pre-venetoclax (in green) and at the time of disease progression (in red). B, Stacked bar plots showing the clonal proportion and number of genomic copies for the six loci encompassing BCL2 family. The color matches the number of genomic copies. (****: adjusted p value < 1E-4)

BCL2 ASP111ALA MUTATION INDUCES VENETOCLAX RESISTANCE

Of interest, in one patient MM04 at the time of progression we observed only minor decrease of *BCL2* and small increases in *MCL1* and *PMAIP1* expression (Figure 35,A). However, despite the lack of significant changes in the expression levels of other pro- or anti-apoptotic proteins, the *ex vivo* profiling of this patient MM cells did show a nearly 1000-fold increase in venetoclax IC₅₀ at the time of clinical progression confirming acquired venetoclax resistance (Supplementary Fig. S1, Figure 40). In addition, as shown in Figure 36,A, single-cell copy number analysis performed prior starting venetoclax and at disease progression did not demonstrate any new copy number gains or losses. In particular there was no 1q21 amplification or gain. Of note, we discovered that this patient acquired a *de novo* *BCL2* mutation on exon 2[NM_000633.2:c.332A>C, p.(Asp111 to Ala), proximal to the hydrophobic groove that binds venetoclax. To determine whether the mutation was sufficient to confer resistance to venetoclax, the mutation was overexpressed in a MM cell-line harboring t(11;14) such as KMS12PE cells, highly sensitive to venetoclax at low nM concentrations. As shown in Figure 36,B, Sanger sequencing was used to determine the nucleotide sequences of our region of interest and confirm the DNA variation on exon 2. As shown in Figure 36,C, the expression of the mutant D111A *BCL2* in KMS12PE cells resulted in nearly 10 folds increase in the IC₅₀ of venetoclax compared to cells transfected with empty vector or a vector expressing WT *BCL2*. Importantly, in *BCL2* IP experiments, while venetoclax displaced BIM from endogenous or WT *BCL2*, it fails to do so in cells expressing the mutant D111A (Figure 36,D). The mutation was further confirmed with ddPCR at time of relapse (P1404) and not prior starting therapy (P1217). As shown in Figure 36,E, the *BCL2* Asp111Ala mutant was present at an allele frequency of 1.124% (CI:0.857-1.472%) post-venetoclax. KMS12PE expressing WT *BCL2* and D111A were used as negative and positive controls respectively (Supplementary Fig. S2A,B, Figure 41).

The crystal structure of venetoclax bound to *BCL2* has been recently reported and demonstrated that the chlorophenyl moiety of venetoclax binds to the p2 hydrophobic groove of *BCL2* while its side chain binds into the p4 hydrophobic groove (Birkinshaw et al., 2019). By performing computational analyses and molecular dynamic simulations on WT- and mutant D111A *BCL2* we observed that that the D111A mutation increased the flexibility of the $\alpha 3$ helix with loss of intra-helical hydrogen bonding. Along with the increased flexibility of *BCL2*, we also observed that A111

interacts with the chlorophenyl moiety of venetoclax leading to increased dynamics of the rest of the ligand with excursions to a solvent-exposed conformation which is unfavorable for venetoclax binding (Figure 36,F).

Using a streptavidin pull-down assay, we have also examined the impact of this D111A BCL2 mutation on the binding of biotin-tagged venetoclax molecules to BCL2. As shown in Supplementary Fig. S3 (Figure 42), in KMS12PE cells stably expressing FLAG-tagged mutant BCL2, venetoclax co-pulled down with WT but not with D111A mutant BCL2 confirming the impaired binding of venetoclax to BCL2 predicted by the dynamic simulation studies.

ACQUIRED MCL1 DEPENDENCY OR CO-DEPENDENCY INDUCES VENETOCLAX RESISTANCE

In almost all patients we observed a significant gain in the *MCL1* transcript levels with no change in *BCL2* at the time of resistance (Figure 35,A). Of note, in patient MM02, single-cell CNV analysis prior treatment showed a focal copy number gain of 1q in a small clonal fraction of 2.2% cells. However, at the time of acquired resistance we observed a focal copy number gain of 2 to 10 Mbp surrounding the *MCL1* locus in 50% of the cells with nearly 12 copies in some clones suggesting a MCL1 dependency at disease progression. In addition, in a subclonal population of cells no *MCL1* gain was observed but 1 copy loss of *BCL2L11* locus was noted (Figure 37,A-B).

Ex vivo venetoclax profiling performed on this patient confirmed a reduction in the sensitivity to venetoclax at time of relapse together with a MCL1 dependency as shown in Figure 37,C-D. While prior starting venetoclax therapy, low dose of drug (100 nM) was able to induce after 12h of exposure cell death in nearly 90% of primary MM cells (Figure 37,C), at progression, only 30% of MM cells were dead after 48 hours exposure to venetoclax (Figure 37,D, third panel). Of note, the MCL1 inhibitor S63845 was able to induce cell death in 30% of the cells but the combination of the two showed an additive effect and ability to induce cell death of 70% of primary cells (Figure 37,D, second and fourth panels).

Next, we wanted to investigate whether MCL1 over-expression was sufficient to trigger venetoclax resistance in MM cells. MCL1 was overexpressed in the KMS12PE BCL2-dependent cell line with a V5 tag (Figure 37,E). As shown in Figure 37,F, MCL1 overexpression is sufficient to induce resistance to venetoclax. Of note, BIM co-IP confirmed that the overexpressed free MCL1 acts as sink for the venetoclax released

BIM from BCL2 and this induced BIM priming of MCL1 making the cells sensitive to the combination of BCL2 and MCL1 inhibitors (Figure 37,G).

Consistent with the increased *MCL1* mRNA expression at progression and the co-existence of BCL2 and MCL1 dependent clones scATAC profiling of this patient (MM02) also demonstrated increased chromatin accessibility at the *MCL1* locus at progression in most but not all cells with the persistence of a BCL2 dependent clone (Figure 38,A-C). Consistent with a branching evolutionary model, cell trajectory analysis and pseudotime ordering of cells using Cicero (Pliner et al., 2018) revealed the emergence over time of an MCL1 dependent clone as the disease evolved from its original BCL2 dependent cluster at pseudotime T0 (Figure 38,D,E).

Légende Figure 36 - BCL2 Asp111Ala mutation induces venetoclax resistance

A, Single cell copy number analysis pre- (top) and post venetoclax at disease progression (bottom) for patient MM04. The chromosomes are linearly represented on the x-axis at a 2Mb resolution. The rows indicate CNV events per cells and cell clusters. The color code indicates number of copies from blue to red and is explained on the right. B, Sanger sequencing results indicating the nucleotide sequences of our region of interest (exon2 of BCL2) and confirming the DNA variation in the KMS12PE transfected with the D111A mutant vs BCL2 WT. C, KMS12PE cells were transfected with empty vector (KMS12PE_EV), BCL2 WT (KMS12PE_BCL2_WT) or BCL2 D111A mutant (KMS12PE_BCL2_D111A) followed by 24hours of venetoclax (0-200 nM). Cell viability was assessed by MTT assay. Results represent the mean $\pm$ SD of three independent experiments. D, Immunoblot of BIM and BCL2 in KMS12PE_EV, KMS12PE_BCL2_WT and KMS12PE_D111A_mutant after 6h of venetoclax treatment (0-100 nM) and following BCL2 overnight immunoprecipitation. Of note, venetoclax is able to displace BIM from BCL2 only in the WT KMS12PE but not in the D111A mutant. E, BCL2 c.332T>G (p.Asp111Ala) digital PCR analysis on MM04. As shown on the right the BCL2 D111A mutant was detected only at the time of progression (1.1% VAF). F, Computational analyses and molecular dynamic simulations models on venetoclax binding to WT_BCL2 (green) and mutant D111A_BCL2 (purple). Of note, the D111A mutation increased the flexibility of the BCL2 α 3 helix with loss of intra-helical hydrogen bonding and increased excursions to a solvent-exposed conformation which is unfavorable for venetoclax binding.

Légende Figure 37 - Acquired MCL1 dependency or co-dependency induces venetoclax resistance

A, Single cell copy number analysis pre- (top) and post venetoclax at disease progression (bottom) for patient MM02. The chromosomes are linearly represented on the x-axis at a 2Mb resolution. The rows indicate CNV events per cells and cell clusters. The color code indicates number of copies from blue to red and is explained on the right. B, Focal copy number gain of ~ 100 kbp surrounding the MCL1 locus on 1q. Of note, a gain was present in a small clonal fraction of cells (2.2%) pre-treatment (on top) while the amplification is found in 50% of clones at time of progression (on bottom). C and D, Ex vivo venetoclax profiling of MM02 primary cells pre- (panel C) and post-venetoclax (D). Cells were treated with venetoclax (100nM) for 12-48hours and/or the MCL1 inhibitor S63845 (50nM) for 48hours. Cell death was assessed in CD138 positive cells by Annexin V staining. Percentage value represents fraction of live cells after each treatment. E, Immunoblot of MCL1 and V5 tag in KMS12PE cells transfected with_EV (KMS12PE_EV) or MCL-1-V5 (KMS12PE_MCL1). F, Cell viability results assessed by MTT assay in KMS12PE_EV and KMS12PE_MCL1 over-expressed cells treated with venetoclax (0-1000 nM) for 24hours. Results represent the mean $\pm$ SD of three independent experiments. G, Immunoblot of MCL1 and BCL2 in KMS12PE cells transfected with_EV (KMS12PE_EV) or MCL-1-V5 (KMS12PE_MCL1) after 6h of treatment with venetoclax (0-200 nM) and/or the MCL1 inhibitor S63845 (0-200 nM) following BIM overnight immunoprecipitation.

Figure 3

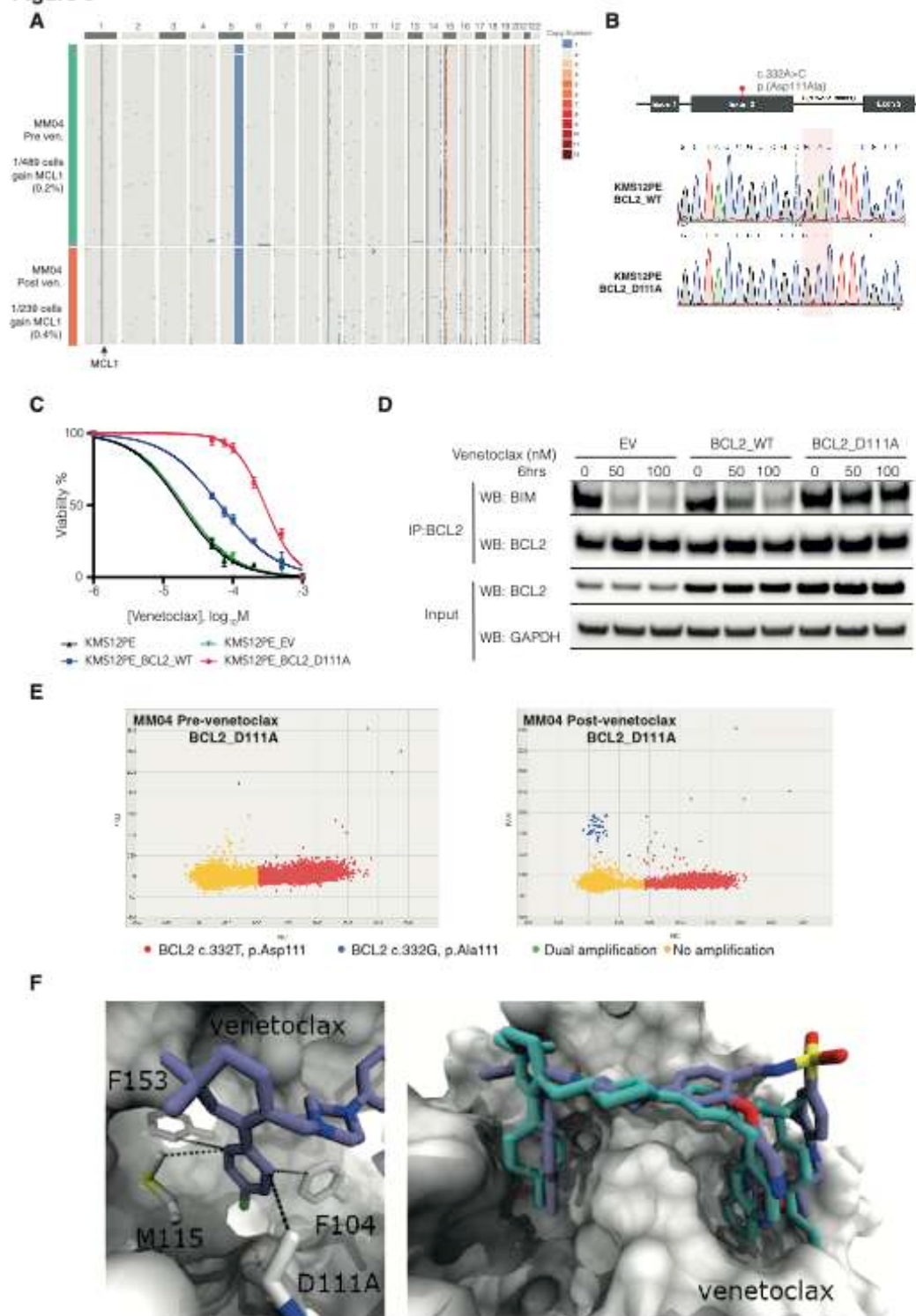


Figure 36 - BCL2 Asp111Ala mutation induces venetoclax resistance (légende page précédente).

Figure 4

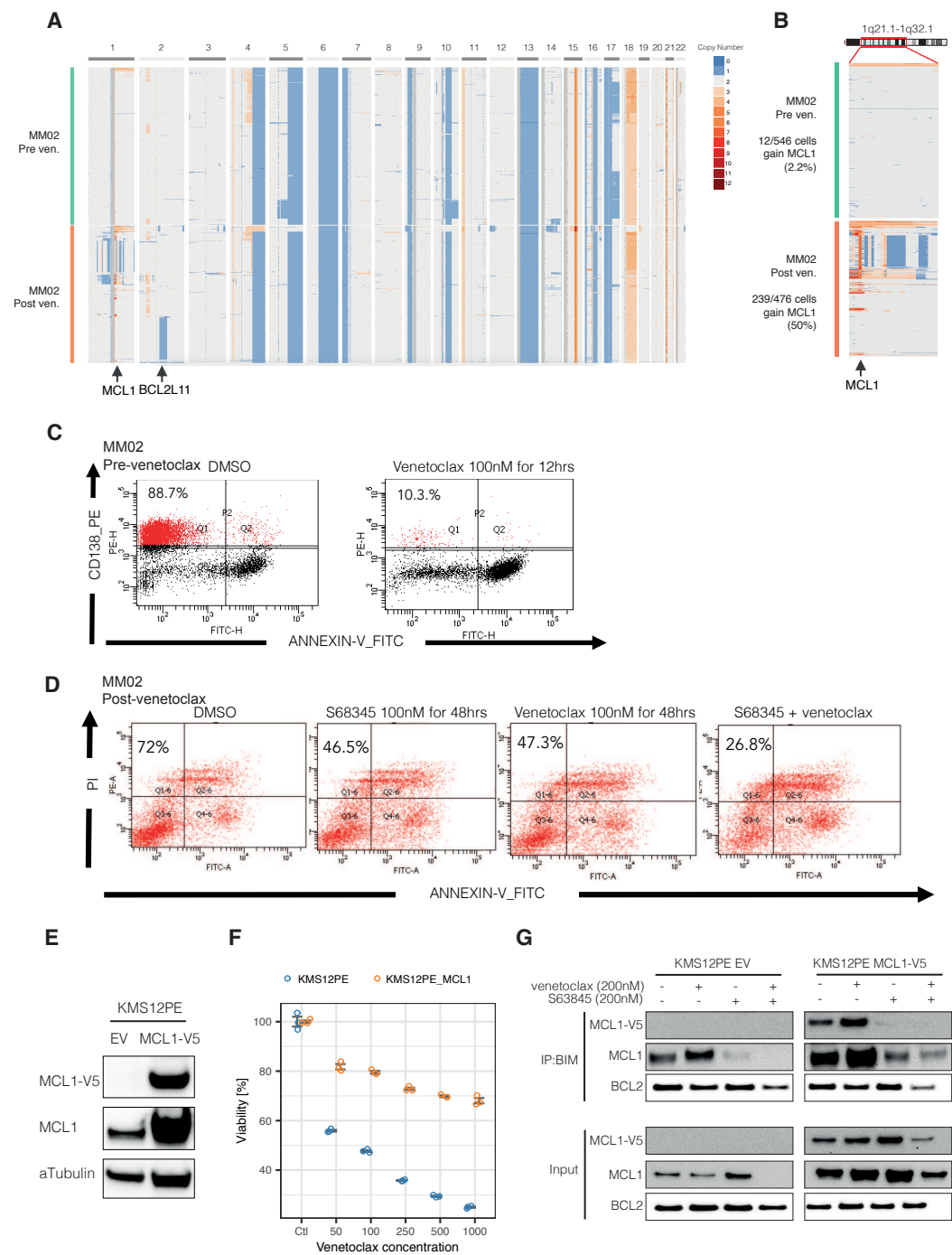


Figure 37 - Acquired MCL1 dependency or co-dependency induces venetoclax resistance (légende pages précédentes)

Figure 5

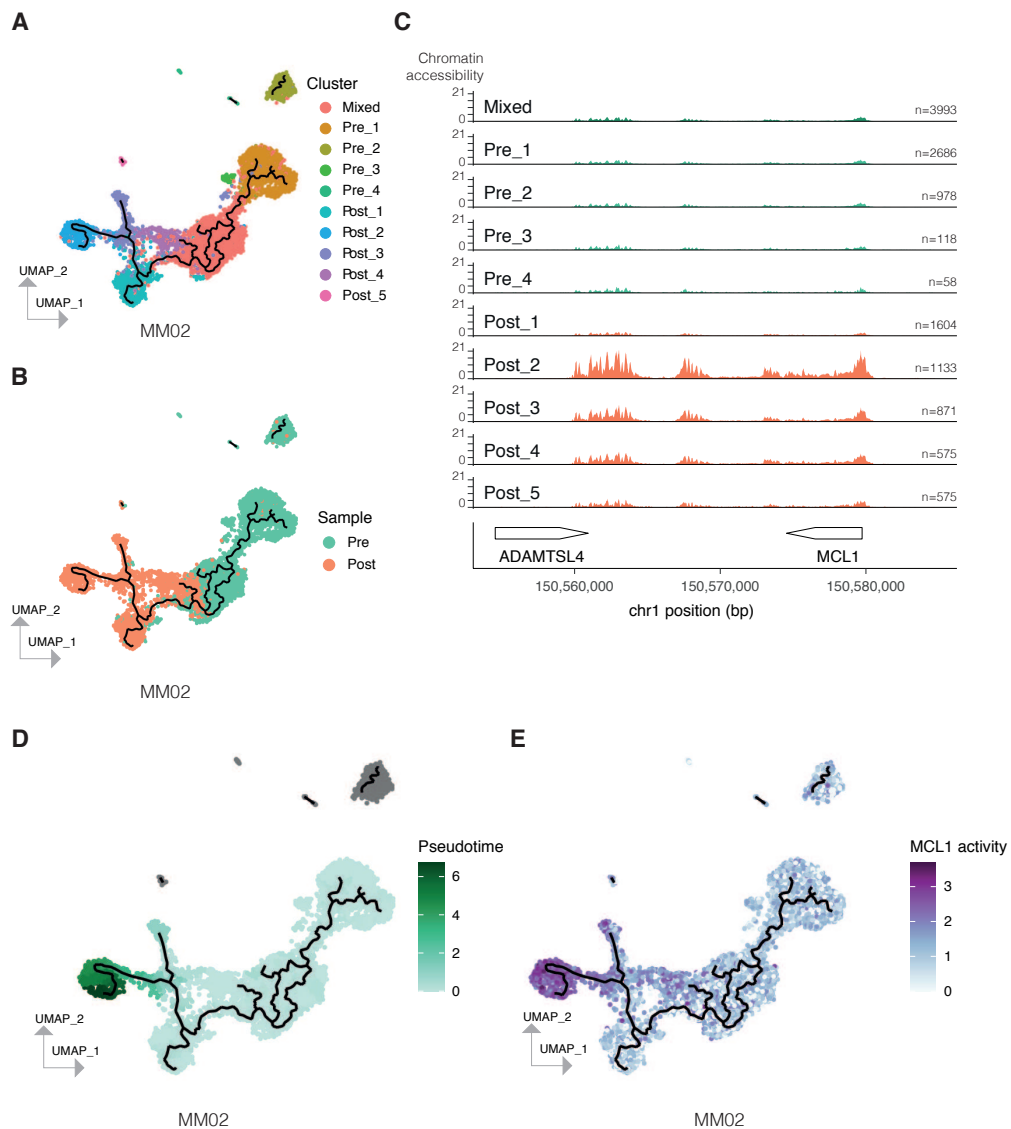


Figure 38 - Pseudotime analysis shows emergence of MCL1 dependent clones at resistance. A-B, UMAP projection to visualize single-cell ATAC-seq data for patient MM02 before (n=7112) and after (n=4966) treatment. The color represents unsupervised clustering in panel A and sample type (pre- or post-venetoclax) in panel B. C, Normalized coverage plot of scATAC-seq data matching the clones and colors presented in panel. Of note, MCL1 accessibility was increased in all cell clusters at progression. D, Cell trajectory analysis and pseudotime ordering of cells using Cicero R package. Pre-venetoclax sample is used as root for pseudotime ordering of cells. E, UMAP to visualize single-cell ATAC-seq data for MM02 based on MCL1 activity.

BCL2L1 REARRANGEMENTS WITH LOSS OF *BCL2* DEPENDENCY LEADS TO VENETOCLAX RESISTANCE

As gains and losses only partially recapitulate the observed resistance to venetoclax, whole genome sequencing analysis was also used to search for structural variants. Of interest, we found *BCL2L1* rearrangements leading to upregulation of *BCL2L1* in the post-treatment samples of two patients (MM01 and MM03). As shown in Figure 39,A in patient MM01, characterized by high *BCL2L1* expression (AvgLogFC=0.7, $p < 1E-4$) but low copy number, we found a novel translocation between IgL enhancer on chromosome 22 (chr22:22,960,001-22,980,000, hg38) and the *BCL2L1* locus on chromosome 20. The same translocation was also found in independent ATAC sequencing reads (Supplementary Fig. S4A, Figure 43). Of note, the *BCL2L1*-IgL translocation was accompanied by focal gains in copy number on chromosome 22 (Figure 39,B) as well as a newly opened chromatin region in the resistant sample (Supplementary Fig. S4B, Figure 43). Furthermore, in patient MM03, we found a novel rearrangement juxtaposing *BCL2L1*, and *CCND1* on chromosome 11. As shown in Figure 39,C, *CCND1* locus on chromosome 11 was translocated with IgH on chromosome 14 as a result of t(11;14). This rearrangement was accompanied by genomic gains of *BCL2L1*, 3' to the rearrangement in 67% of the cells (297/440). Of note, in the same cells, we found an additional copy (n=4) of chr11 matching the two breakpoints: 5' of MYEOV for the t(11;14), and 3' of MYEOV for the t(11;20) (Figure 39,D). The same rearrangement involving *BCL2L1* and *CCND1* was also found in independent ATAC sequencing reads (Supplementary Fig. S4C, Figure 43).

Ex vivo functional assay performed on these samples further confirmed the loss of *BCL2* dependency and venetoclax resistance of these patients. Data for MM01 are shown in Figure 39,E,F.

Taken together, our data suggest that complex genomic rearrangements involving *BCL2L1* can drive aberrant malignant circuitry endowing resistance to anti-*BCL2* agents in MM cells.

Figure 6

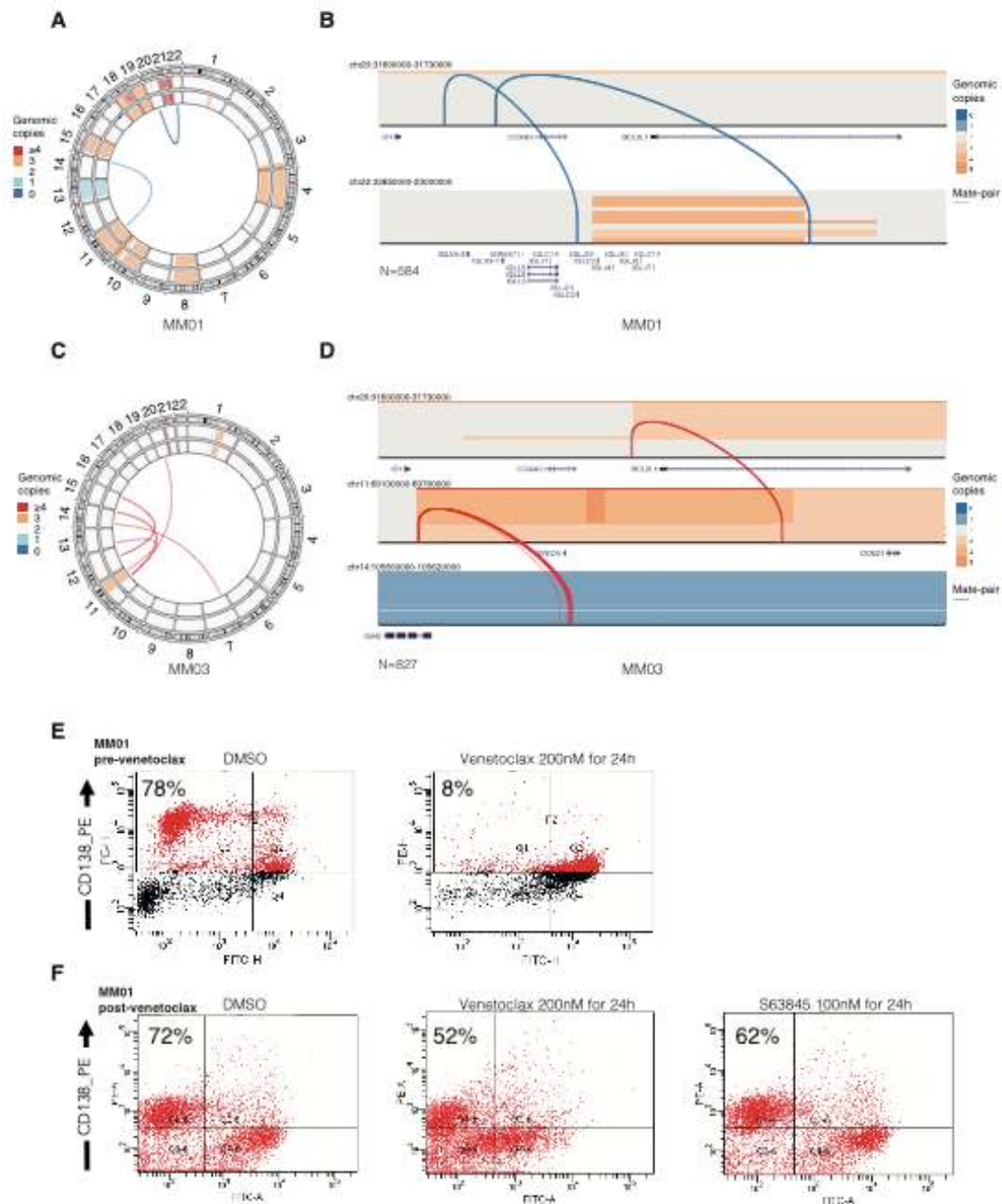


Figure 39 - BCL2L1 rearrangements with loss of BCL2 dependency leads to venetoclax resistance. A and C, these circos plots show the key translocations and copy number abnormalities found in patient MM01 and MM03, respectively. The chromosomes are arranged around the circle starting from the top at chromosome 1 and continuing clockwise to chromosome 22. The outer circle indicates the pre-venetoclax sample, the inner circle the post-venetoclax. B and D, Paired-end sequencing representation of scCNV data at breakpoints locations proximal to BCL2L1 and IgL loci for MM01 and MM03 respectively. The lines represent DNA fragments mapping to both chromosomes. The color codes from blue to red in all panels indicates the number of copies identified in each cell clusters and has been generated by using the CellRanger DNA Loupe software. E, F Ex vivo venetoclax profiling of MM01 primary cells pre- (panel E) and post-venetoclax (F). Cells were treated with venetoclax (200nM) and/or the MCL1 inhibitor S63845 (100nM) for 24hours. Cell death was assessed in CD138 positive cells by Annexin V staining. Percentage value represents fraction of live cells after each treatment.

2.6. DISCUSSION

Treatment of MM patients in the relapsing setting remains very challenging, primarily due to the heterogeneity of disease at relapse and the absence of clear biological-based recommendations regarding the choice of salvage therapies. As such, there is an unmet need for targeted therapy as a way of overcoming resistance to current therapies and improving patient outcomes.

Venetoclax is a potent and selective inhibitor of BCL2 that has showed significant activity, as both a single agent and in combination with other therapies in RRMM (Kumar et al., 2017b; Moreau et al., 2017b), particularly in those harboring t(11;14), which express high levels of BCL2 relative to BCLxL and MCL1 (Kumar et al., 2017b). However, despite the clinical efficacy treatment resistance still occurs and is an important cause of treatment failure. One of the possible explanations for this phenomenon is the fact that the expression of BCL2 family members in MM cells is heterogeneous, and upregulation of alternate pro-survival proteins such as MCL1 or BCLxL results in a relative resistance of malignant plasma cells to venetoclax killing (Punnoose et al., 2016; Seiller et al., 2020). Furthermore, the development of venetoclax binding site mutations (Gly101Val, Asp103Tyr, and Phe104Ile) have recently emerged as important contributors to resistance in CLL and FL patients. Therefore, understanding the underlying cellular and genomic mechanisms that contribute to these modes of resistance is central to optimizing the use of this effective therapy.

Intra-tumor heterogeneity is a common feature across diverse cancer types including MM (Melchor et al., 2014). Dynamic genetic and epigenetic changes among intra-tumoral sub-populations over time and following therapy present also a key challenge to cancer treatment (Gerlinger et al., 2012). Therefore, understanding this variability within and across patients is critical for the identification of mechanism of drug resistance.

In this study we have integrated multiple single-cell approaches in paired samples of RRMM patients treated with venetoclax to unmask the genetic and epigenetic modifications contributing to venetoclax resistance. As such, scRNAseq, scCNVseq as well as single cell ATAC-seq were performed on primary MM cells collected from patients treated with venetoclax prior to initiation of therapy and at disease progression. By performing this comprehensive multi-omics approach we identified numerous changes in both pro-apoptotic and anti-apoptotic family members, including *BCL2*, *MCL1*, *BCL2L11*, *PMAIP1* and *BBC3* at the time of venetoclax

resistance. However, these changes were neither mutually exclusive nor universal across the samples tested, confirming that MM is particularly heterogeneous with respect to anti-apoptotic protein dependencies and the expression patterns of these molecules by itself cannot be used as biomarkers to predict response to venetoclax.

Of note, we report for the first time in MM a new mutation (D111A) on exon 2 of *BCL2* proximal to the hydrophobic groove where venetoclax binds. As previously reported for the other venetoclax binding site mutations (Gly101Val, Asp103Tyr, and Phe104Ile) reported in CLL and FL patients (Blombery et al., 2019), the mutation was sub-clonal and present at low allele frequency (AF) at time of disease progression. Given the observed sub-clonality of the *BCL2* D111A mutation we have functionally characterized the effect of the mutation on impairing the binding of venetoclax to *BCL2* and extensively validated *in vitro* its role in the development of venetoclax resistance. VAF is the result of cell-intrinsic fitness advantages or cell-extrinsic factors such as immune or environmental (Rozhok and DeGregori, 2015). It depends on the sample's tumor content, the hetero- or homozygous status, the copy number variation of the considered locus, and the clonal or sub-clonal nature of the mutation. Although tumor cell content may also account for the low VAF observed in the D111A mutation, our results indicate that most *BCL2* mutations accumulate in sub-clones, reflecting ongoing genomic instability in MM cells.

Notably, we also observed an *MCL1* locus copy number gain associated with increased gene expression in several patients at the time of disease progression and venetoclax resistance. *MCL1* dependency has been shown to increase at relapse in MM (Gomez-Bougie et al., 2018) and *MCL1* over-expression is frequently due to either gene amplification (1q amplification) or oncogenic pathways (Beroukhim et al., 2010). In our study, we observed that in one patient, *MCL1* was locally amplified but the amplification didn't encompass *CKS1B*. This finding could partially explain why some patients are sensitive to *MCL1* inhibition irrespective of their 1q21 status assessed by FISH (Slomp et al., 2019). However, the exact molecular mechanism by which the significant and focal *MCL1* amplification occurred (e.g. double-minute chromosome, breakage-fusion-bridge mechanism) remains to be elucidated. Furthermore, since cells that are not bearing the *MCL1* amplification have lost one copy of *BCL2L11* encoding BIM, it appears that clonal selection during the course of the disease could have favored *MCL1* dependency. Therefore, our data provide a strong rationale for combination therapies using *BCL2* and *MCL1* inhibitors. Two *MCL1* inhibitors (AMG176 and AZD5991) have recently demonstrated anti-MM activity in a *MCL1*-

dependent manner in preclinical studies, and are currently being evaluated in early phase trials (NCT02675452, NCT03465540 and NCT03218683) (Caenepeel et al., 2017; Hird et al., 2017). However, caution should be taken regarding the potential toxicity of such approaches, especially after the recent cardiac toxicity signal observed with AMG176.

Lastly, we report that two complex rearrangements involving *BCL2L1* can significantly affect cell death induced by venetoclax. In one patient, *BCL2L1* was found translocated with the IgL enhancer – t(20;22). In the other one, we describe a novel *CCND1-BCL2L1* translocation in a t(11;14) patient. It is not clear whether the translocation is juxtaposing *BCL2L1* to der(14) and the immunoglobulin heavy chain enhancer. The genomic gain of 1Mbp at *BCL2L1* in the same sample is also a potential driver of *BCL2L1* expression. To the best of our knowledge, this is the first report of translocations of *BCL2L1* as a mechanism occurring during MM progression under venetoclax. *BCL2L1* was found translocated with IgH in one patient from the CoMMPASS cohort (Barwick et al., 2019), but was never reported with IgL. Genomic instability with the identification of a complex mutational landscape and a branching pattern of clonal evolution represents a hallmark of MM cells (Neri and Bahlis, 2013). Here we reported on how genetic changes involving *BCL2L1* provide growth advantage of the fittest clone and promote the acquisition of venetoclax resistance leading to disease progression.

To our knowledge, this is the first study to describe *BCL2* mutation, *MCL1* gain and *BCL2L1* rearrangements as major causes of resistance to BCL2 inhibition in MM. Altogether, our data suggest that venetoclax resistance in MM is a heterogeneous phenomenon, both between and within individual patients. Constitutive overexpression of these pro-survival proteins in MM results from a range of microenvironmental signals and different genetic lesions. This complex regulation of *MCL1*, *BCL2*, and *BCLxL* offers multiple direct and indirect targets for therapeutic intervention. Therefore, early therapeutic interventions targeting these branching clones may enhance venetoclax efficacy and improve survival in venetoclax treated MM patients. In addition, our study emphasizes that early detection and dynamic monitoring of these abnormalities is essential for the clinical management of these patients and should be considered prior to initiate therapy since the sequence of these therapies may influence patterns of resistance.

2.7. SUPPLEMENTARY FIGURES

Supplementary Figure 1

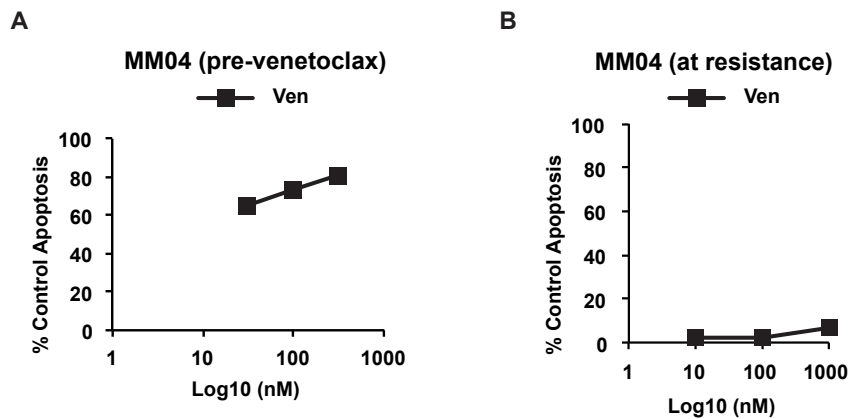


Figure 40 - D111A BCL2 mutation in primary MM cells contributes to resistance to venetoclax.

A,B Ex vivo venetoclax profiling of MM04 primary cells pre- (panel A) and post-venetoclax (B). Cells were treated with venetoclax (0-1000nM) for 24hours. Cell death was assessed in CD138 positive cells by Annexin V staining. Percentage value represents fraction of live cells after treatment.

Supplementary Figure 2

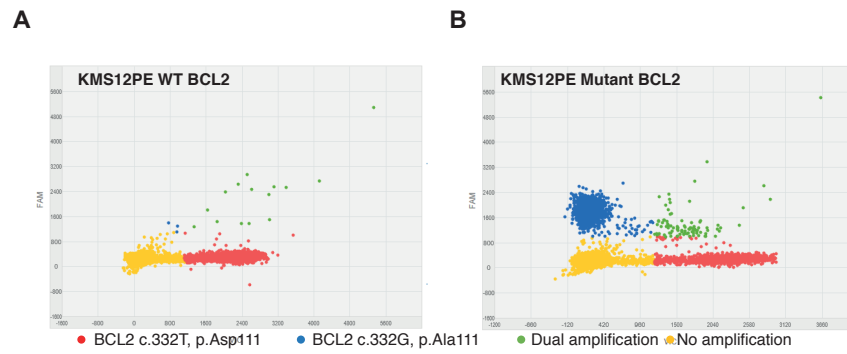


Figure 41 - BCL2 c.332T>G (p.Asp111Ala) digital PCR analysis on KMS12PE control MM cell line.

KM12PE cell line were transfected with empty vector (A) or the D111A BCL2 mutant (B). As expected, the BCL2 D111A mutant was detected in the KMS12PE_BCL2_D111A with a VAF of 58%.

Supplementary Figure 3

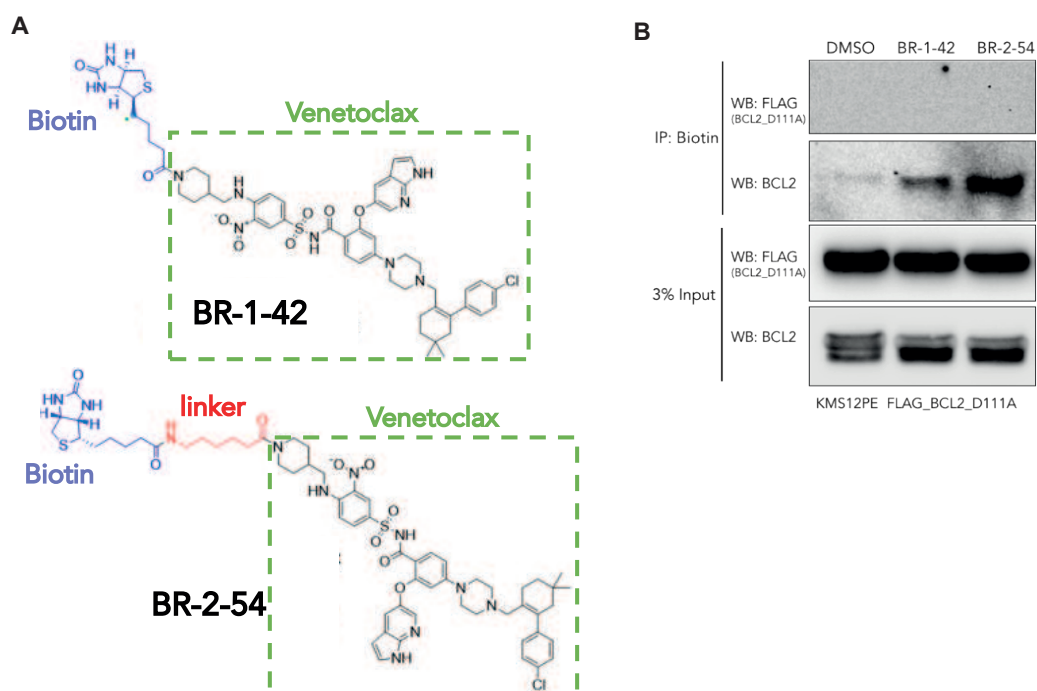


Figure 42 - Effect of D111A BCL2 mutation on the binding of biotin-tagged venetoclax molecules to BCL2.

A, Two venetoclax molecules with biotin tags were generated by using a short (BR-1-42) and long linker (BR-2-54) to allow biotin immunoprecipitation experiment. B, Streptavidin pull-down assay was performed and immunoblot of BCL2 and FLAG measured to examine the impact of this D111A BCL2 mutation on the binding of biotin-tagged venetoclax molecules to BCL2. Of note venetoclax co-pulled down with WT but not with D111A mutant BCL2 confirming the impaired binding of venetoclax to BCL2 predicted by the dynamic simulation studies.

Supplementary Figure 4

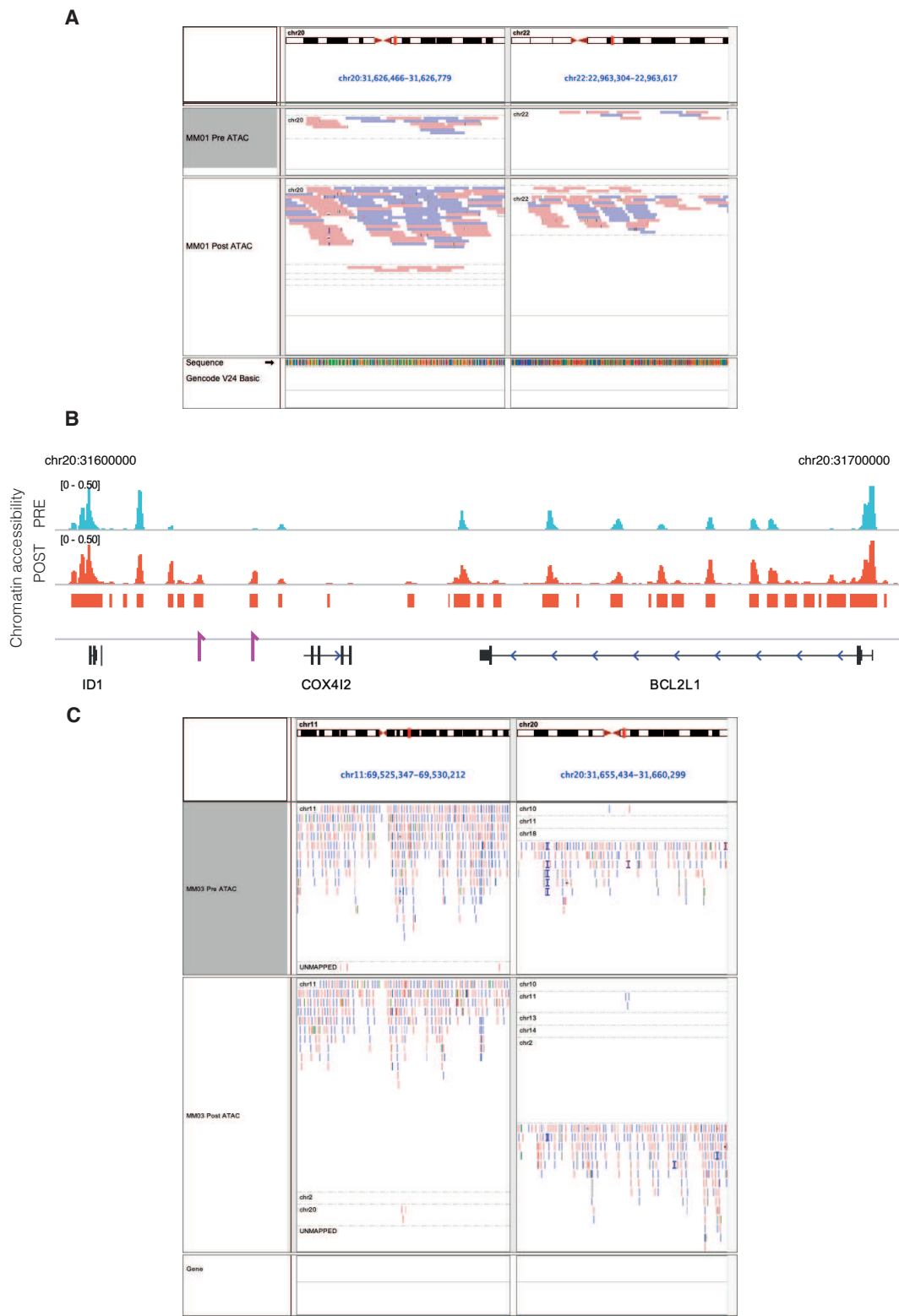


Figure 43 – *BCLxL* / *BCL2L1* rearrangements confirmed by ATAC-Seq data (légende page suivante).

Légende de **Figure 43** – BCLxL / BCL2L1 rearrangements confirmed by ATAC-Seq data.

A, Representation of ATAC paired-end sequencing reads of MM01 mapping to *BCL2L1* (left) and IgL (right) translocation breakpoints on chr20 and chr 22, pre- (top) and post-venetoclax (bottom panel). Genomic reads are colored by strand and grouped by chromosome of mate.

B, Chromatin accessibility track of MM01 pre- (top) and post-venetoclax (bottom panel) in MM01. Of note, two novel opened regions map to the breakpoint location.

C, Representation of ATAC paired-end sequencing reads of MM03 mapping to *CCND1* (left) and *BCL2L1* (right) translocation breakpoints on chr11 and chr 20, pre- (top) and post-venetoclax (bottom panel). Genomic reads are colored by strand and grouped by chromosome of mate.

3. ÉTUDE D'ASSOCIATION ENTRE PROFILS TRANSCRIPTOMIQUES ET BIOMARQUEURS D'IMAGERIE PAR TEP-FDG

Ce travail a été réalisé en collaboration avec l'équipe de recherche en oncologie nucléaire des Pr Françoise Kraeber-Bodéré et Michel Chérel. Il fait l'objet d'une présentation « poster » au congrès de l'ASH 2020 (Alberge et al., 2020b). L'étude est portée par l'équipe d'imagerie médicale, par l'équipe de recherche du Dr Stéphane Minvielle et par le service d'hématologie du CHU de Nantes avec Dr Cyrille Touzeau et Pr Philippe Moreau. J'ai réalisé les analyses de données RNA-seq et les analyses statistiques. Les données d'imagerie sont fournies et interprétées par l'IFM, Dr Bastien Jamet, Dr Clément Bailly, Dr Thomas Carlier, Dr Caroline Bodet-Milin.

Genome-Wide Transcriptome Analysis Identifies Molecular Patterns of FDG-PET/CT Biomarkers in MM Patients from the Cassiopet Study

Jean-Baptiste Alberge^{1,2}, Bastien Jamet^{1,3}, Clément Bailly^{1,2,3}, Cyrille Touzeau^{1,2,4}, Jonathan Cruard^{1,2}, Thomas Carlier^{1,2,3}, Caroline Bodet-Milin^{1,2,3}, Michel Chérel^{1,2,5}, Philippe Moreau^{1,2,4}, Stéphane Minvielle^{1,2,4} and Françoise Kraeber-Bodéré^{1,2,3,5}

1 Site de Recherche Intégrée sur le Cancer (SIRIC), Imaging and Longitudinal Investigations to Ameliorate Decision-making (ILIAD), INCA-DGOS-Inserm 12558, Nantes, France

2 Université de Nantes, CHU Nantes, CNRS, Inserm, CRCINA, Nantes, France

3 Nuclear Medicine Unit, University Hospital, Nantes, France

4 Haematology Department, University Hospital, Nantes, France

5 Nuclear Medicine Unit, ICO Cancer Center Gauducheau, Saint-Herblain, France

3.1. INTRODUCTION

INTÉRÊT DE LA TEP-FDG DANS LE MYÉLOME MULTIPLE

L'imagerie par tomographie à émission de positron (TEP) au ^{18}F -fluorodésoxyglucose (FDG) est recommandée par l'IMWG depuis 2017 au diagnostic et au cours du traitement du myélome multiple (Cavo et al., 2017). Cette technique d'imagerie évalue à la fois les lésions osseuses actives et les sites de prolifération extra-médullaire. Plusieurs biomarqueurs pronostiques sont extraits de la TEP-FDG : nombre de lésions focales, présence ou non de maladie extra-médullaire (EMD), para-médullaire (PMD), absorption maximale standardisée de glucose (SUVmax) (Moreau et al., 2017a; Abe et al., 2019; Jamet et al., 2019b). Pour autant, 10 à 20% des scanners sont négatifs au diagnostic et l'étendue des connaissances sur les liens entre aspects moléculaires de la maladie et biomarqueurs d'imagerie reste limitée.

ASSOCIATION DE L'IMAGERIE ET DES PROGRAMMES TRANSCRIPTOMIQUES

Les approches transcriptomiques et d'imagerie sont opposées par nature : les biomarqueurs d'imagerie sont calculés aux endroits de prolifération tumorale ou de forte avidité pour le FDG, alors que la ponction de moelle n'est pas dirigée sur les lésions ostéolytiques, en général dans le sternum ou la crête iliaque. On considère que les cellules étudiées en RNA-seq sont des cellules « circulantes » de la moelle osseuse, alors que l'imagerie montre une hétérogénéité spatiale y compris intra-lésionnelle (Figure 44). Enfin, certaines TEP sur lesquelles aucune anomalie significative de fixation du FDG n'est mise en évidence sont dites « normales » (ou négatives) malgré le diagnostic clinique établi notamment par la présence de plasmocytes tumoraux dans la ponction de moelle osseuse.

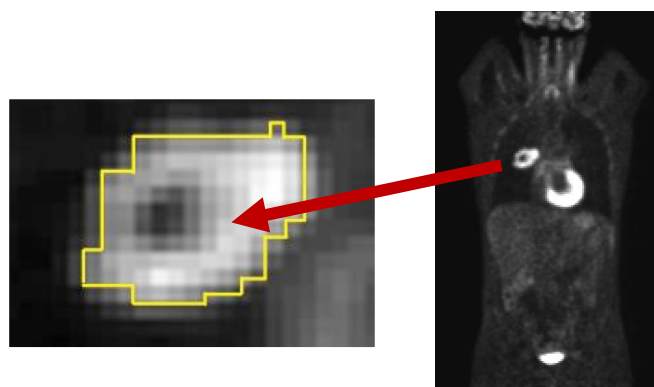


Figure 44 - Hétérogénéité tumorale intra-lésionnelle de fixation en TEP-FDG. Image de l'équipe d'oncologie nucléaire du CRCINA, Nantes.

Puisque les deux techniques (transcriptomique et imagerie) identifient des groupes de patient à haut risque et qu'elles reflètent au moins un clone ancestral commun, il est possible que les caractéristiques des deux se recoupent partiellement et aient un intérêt pronostique complémentaire. Dans ce domaine, peu de choses sont connues. Le groupe de Little Rock montre que l'hexokinase-2 *HK2* (voir Figure 10 page 26) est fortement sous-exprimée chez les patients négatifs à la TEP-FDG mais positifs à la séquence de diffusion, une forme d'IRM corps entier (Rasche et al., 2017a). Cela suppose que les « faux-négatifs » ont un métabolisme d'assimilation du glucose non-canonique, ou une prolifération moins dépendante du glucose. Toutefois, le mécanisme de répression de l'*HK2* et la voie de prolifération empruntée ne sont pas connues. De plus, cette étude ne traite pas des « faux-négatifs » pour lesquels ni TEP ni IRM ne détectent de maladie au diagnostic. Quant aux sept sous-groupes moléculaires (CD-1, CD-2, HY, LB, MF, MS, et PR) et à la signature de haut-risque GEP70, le même groupe de Little Rock montre que les sous-groupes PR, MF, et GEP70 présentent plus de maladie extramédullaire (EMD) (Usmani et al., 2012). Le groupe MS / t(4;14) est aussi de mauvais pronostic pour la survie mais n'est pas associé à l'EMD, ce qui montre que les deux biomarqueurs n'identifient pas les mêmes patients. Ces études n'ont pas été reproduites par d'autres groupes de recherche : qu'en est-il dans la cohorte CASSIOPEIA / CASSIOPET ? Aussi, la distinction entre maladie paramédullaire et extramédullaire n'était pas faite par l'UAMS. Quelles sont les caractéristiques des deux groupes de maladies ? Enfin, comment caractériser les TEP normales au-delà de la diminution d'*HK2* ?

3.2. MÉTHODES

Un groupe de 136 patients de CASSIOPET, une étude parallèle (*companion study*) à la cohorte CASSIOPEIA (ClinicalTrials.gov, NCT02541383) a été soumis à une quantification du transcriptome (RNA-seq) sur cellules de myélome purifiées de la moelle osseuse au diagnostic en plus de l'imagerie par TEP-FDG. Le séquençage se fait sur NovaSeq 6000 à 100 millions de reads par individu sur le site nantais. Les reads sont alignés avec STAR v2.7.3 sur hg38 et sont annotés sur ENSEMBL v99. Les signatures moléculaires (Zhan et al., 2006; Shaughnessy et al., 2007) sont calculées d'après données normalisées et standardisées (log2 du nombre de transcrits par million (TPM), centré réduit). Le facteur TPM est calculé sans tenir compte des gènes des immunoglobulines définis par la base de données d'anticorps et d'immunoglobulines IMGT (Barwick et al., 2019). L'argument maximum de la moyenne pondérée est utilisé pour assigner une classe moléculaire à un patient.

Pour la classification à haut risque, on utilise le seuil de 0,5 sur la moyenne signée qui correspond au même pourcentage de patients que dans la publication de l'UAMS (13%). On utilise la normalisation vst (variable stabilizing transformation) du paquet R DESeq2 pour la représentation graphique des données normalisées. Le taux de pureté en plasmocytes et le centre collecteur (fondation néerlandaise HOVON ou IFM) sont systématiquement inclus dans la modélisation linéaire de DESeq2. Sur les données d'imagerie, on retient comme variables : le rendu (TEP normale/anormale, variable binaire), la présence d'EMD ou de PMD (variables binaires), le nombre de lésions focales (variable de classe ordonnée : 1, 2, 3, 4, 5 à 10, >10), et la SUVmax (variable continue positive supérieure à 1). On utilise un test de wilcoxon pour une variable continue entre deux classes, un test de Kruskal-Wallis avec test post-hoc de Dunn entre plusieurs classes. On teste l'association entre deux variables de classe par un test du χ^2 ou un test exact de Fisher selon le nombre d'observations. On considère un résidu du test du $\chi^2 \geq 2$ en valeur absolue comme post-hoc. On utilise le test de tendance de Cochran-Armitage pour les variables de classes ordonnées, et le test de Wald du paquet R/DESeq2 pour identifier les gènes dont l'expression varie entre les conditions du test ($H_0: \log_2FC=0$). Pour les analyses de survie, la durée sans événement est calculée de la date de randomisation à la date de l'événement ou à la dernière date de suivi. Le hazard ratio est calculé selon un modèle de Cox univarié. Les courbes de survie sont calculées selon la méthode de Kaplan-Meier. Les valeurs p ajustées selon la méthode de Benjamini-Hochberg et inférieures à 0,05 sont considérées comme significatives.

3.3. RÉSULTATS

La classification moléculaire fonctionnelle de l'UAMS est présentée en Figure 45, panel A et montre une proportion de patients comparable aux études précédentes (Zhan et al., 2006; Broyl et al., 2010). La proportion de patients à profil d'expression à haut risque est représentée en Figure 45, panel B. Les variables d'imagerie utilisées pour l'analyse sont présentées sur les panels C à H de la même figure et sont fournies par l'IFM. La TEP-FDG est positive pour 109 patients sur 139 (79%), dont 20 (14%) et 16 (12%) présentent également une PMD ou une EMD respectivement.

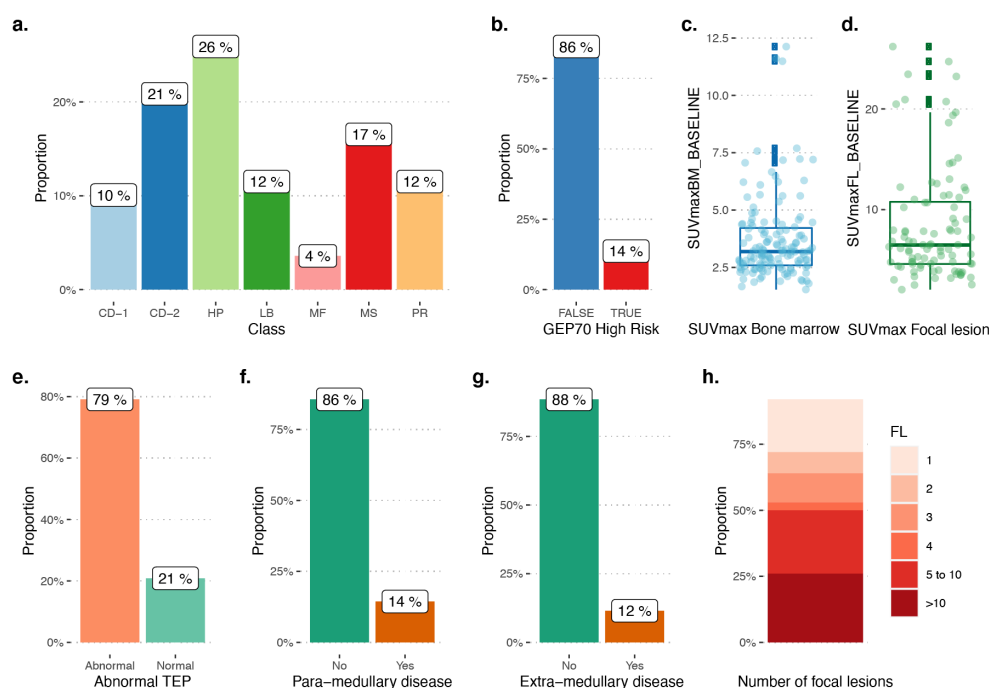


Figure 45 – Répartition des variables de RNA-seq et d'imagerie au TEP-FDG. Chaque échantillon (N=139) est attribué à une des catégories des classifications suivantes : moléculaire d'après (Zhan, 2006) (a.), CEP70 de haut risque d'après (Shaughnessy, 2007) (b.). Les consommations maximales de glucose (SUVmax) sont indiquées dans la moelle osseuse (c.) et dans les lésions ostéolytiques (d., N=92). Les TEP sont classées en catégories normales ou anormales (e.), avec maladie para-médullaire (f.), extra-médullaire (g.) et selon le nombre de lésions ostéolytiques visibles à l'examen (h.).

Tableau 8 - Analyse de survie sans progression (PFS) univariée d'après un modèle de cox. HR : Hazard ratio ; conf : intervalle de confiance ; LRT : likelihood ratio test. La significativité liée aux variables d'imagerie est corrigée pour les tests multiples.

	Beta	HR (95% conf)	LRT	LRT p-value	Adj.p-value	Signif.
PET (Normal or abnormal)	-2	0,14 (0,019-1)	7	0,008	0.01	*
EMD (Extra-medullary disease)	1,2	3,3 (1,4-7,8)	5,7	0,02	0.03	*
PMD (Para-medullary disease)	1,5	4,6 (2-11)	11	0,0009	0.005	**
SUVMax$\geq$4 (Bone marrow SUVmax)	0,38	1,5 (0,65-3,3)	0,8	0,4	0.4	
FL$>$3 (Number of focal lesions)	1,1	3,1 (1,4-7,1)	7,9	0,005	0.0125	*
Molecular Class			7,6	0,3		
CD-1	-1	0,4 (0-3,1)				
CD-2	-0,3	0,7 (0,2-2,5)				
LB	-1,2	0,3 (0-2,5)				
MF	0,1	1,1 (0,1-9,2)				
MS	0,2	1,2 (0,4-3,7)				
PR	0,9	2,4 (0,8-7,3)				
CEP70 (High risk or normal)	1,4	4,1 (1,7-9,5)	8,6	0,003		**

INTÉRÊT PRONOSTIQUE DE L'ASSOCIATION DES VARIABLES D'IMAGERIE ET D'EXPRESSION

La médiane de suivi est de 26 mois (95% : 21 - 33 mois), durant lesquels 25 patients (18%) progressent et 9 décèdent (6%). Les analyses de survie sont limitées à la PFS, l'OS a été écartée en raison du nombre restreint d'événements. L'analyse univariée de Cox (Tableau 8) montre que l'imagerie et l'expression génique ont un caractère pronostique pour la progression. En particulier, la PMD et la signature de haut risque GEP70 sont fortement associées à la rechute (respectivement $HR_{PMD}=4,6$, $CI : [2-11]$, et $HR_{GEP70}=4,1$ (1,7-9,5)).

La combinaison de la PMD et de la GEP70 en un modèle multivarié de Cox ($p=0.0004$) montre l'indépendance des deux variables pour prédire la rechute (Figure 46, panel A, PMD : $HR=3,5$, $CI=[1,5-8,5]$, $p=0,005$, GEP70 : $HR=2,9$, $CI=[1,2-7]$, $p=0,02$).

Pour analyser le caractère additif de la PMD et la GEP70, les patients ont été stratifiés sur la base de la PMD et de la GEP70 pour définir quatre classes : la classe double-positive PMD et GEP70 ($n=8$), la classe double-négative ($n=108$), et les classes PMD seule ($n=12$) et GEP70 seule ($n=11$). Les courbes de Kaplan-Meier en Figure 46, panel B, montrent que la stratification fait émerger un groupe à très haut risque de progression : les patients « doubles positifs » (PFS médiane inférieure à 18 mois, $p=9E-7$). Le mauvais pronostic du groupe PMD et GEP70 double-positif est indépendant de l'ISS et du R-ISS (resp. $p=5E-4$ et $p=5E-5$). Par ailleurs, les 8 doubles-positifs sont répartis sur les trois classes de la R-ISS : R-ISS I : 1 patient, R-ISS II : 4 patients et R-ISS III : 3 patients.

Parmi les patients à TEP anormale, les patients positifs GEP70 (GEP70pos) présentent une PMD plus fréquente que les GEP70-négatifs (GEP70neg, $p=0,003$, Figure 46, panel C) mais la majorité des patients à PMD sont GEP70neg. Le groupe proliférant (PR) est également associé à la PMD. Par ailleurs, aucun des patients du groupe à faible maladie osseuse (LB) ne présente de PMD malgré une TEP anormale (Figure 46, panel D, non-significatif). Une analyse parallèle comparable pour l'EMD montre une absence d'association statistique entre EMD et GEP70 ainsi qu'entre EMD et classification moléculaire (Figure 46, panels E et F).

Pris ensemble ces résultats montrent que le caractère pronostique de la PMD s'additionne à celui de la GEP70 pour définir un groupe de patients à très haut risque de progression.

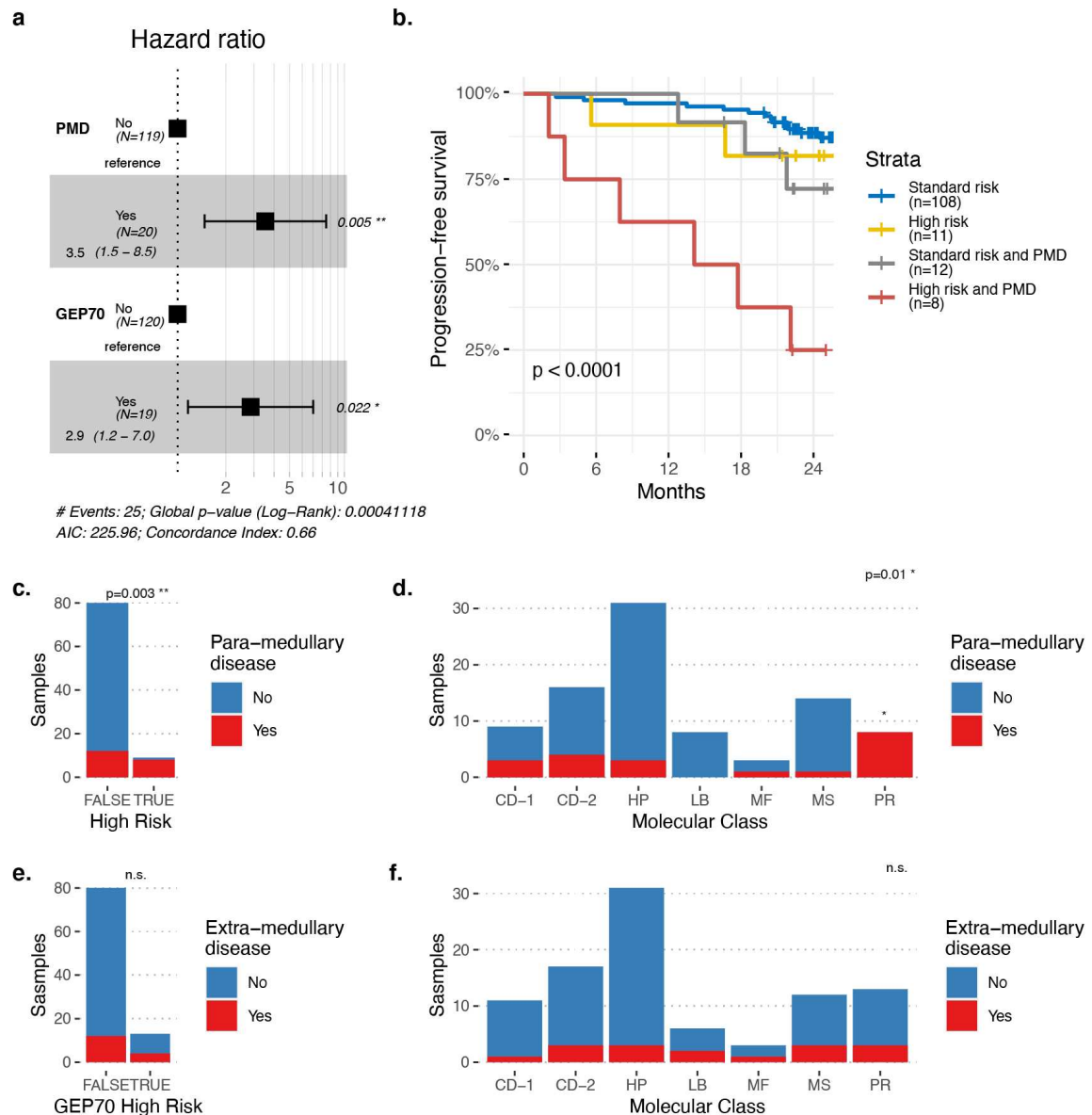


Figure 46 - Maladie para-médullaire et signature GEP70 définissent un groupe à très haut risque de progression. **a.** Graphe de forêts multivariées sur la survie sans progression. **b.** Courbe de Kaplan-Meier de la survie sans progression stratifiée par risque GEP70 et la PMD. **c, d.** Répartition des patients présentant une maladie para-médullaire à l'inclusion en fonction du risque GEP70 (**c**) et de la classification moléculaire (**d**). La couleur représente la présence (rouge) ou l'absence (bleu) de PMD. **e, f.** Représentation de la maladie extra-médullaire (EMD) selon les codes couleurs de **C, B**.

TEP NORMALE ET EXPRESSION DES GÈNES DU MÉTABOLISME DU GLUCOSE

Par opposition à la PMD, la TEP normale (négative) est de bon pronostic. Pour chercher à comprendre quels patients sont susceptibles de présenter une TEP normale au diagnostic, et pour affiner le bon pronostic conféré par la faible avidité pour le FDG, nous avons étudié l'expression des transporteurs de glucose (*GLUTs*, famille *SLC2A*) et des hexokinases (*HKs*) dans la cohorte. *SLC2A2* et *SLC2A4* sont peu exprimés (<0.02 TPM en moyenne) et ont été retirés de l'étude. *HK2*, première enzyme de la glycolyse, est plus faiblement exprimée chez les patients à TEP normale (Fold Change = 2.0, $p=0.04$, Figure 47 panel A), tout comme *GLUT5* (*SLC2A5*, FC=4.2, $p=7E-4$) qui code pour un récepteur canonique du fructose mais pas du glucose. À l'inverse, *SLC2A3* (*GLUT3*) est sur-exprimé chez les patients à TEP normale (FC=2.0, $p=0.003$). L'expression des *GLUTs* et *HKs* est dérégulée entre TEP normales et anormales, ce qui suggère que les caractéristiques moléculaires des patients pourraient différer entre ces deux groupes.

Pour comprendre quels sous-groupes de myélomes sont plus susceptibles de présenter une TEP normale, nous avons exploré les deux classifications GEP70 et UAMS. La classification GEP70 s'avère indépendante de l'expression des *GLUTs* et *HKs* (Figure 47, panel B), ainsi que de la TEP normale (Figure 47, panel C). TEP normale, de bon pronostic, et GEP70, de mauvais pronostic, semblent se compléter en analyse de Kaplan-Meier et dans un modèle multivarié de Cox (Figure 47, panels D et E) et séparent deux groupes presque disjoints de patients (2 patients sur 139 ont une TEP normale et une GEP70 de haut risque).

La combinaison linéaire des *GLUTs* et *HKs* significatives permet de séparer TEP normales et anormales mais ne gagne pas en puissance statistique par rapport à *SLC2A5* seule (Figure 47, panel F), et la somme de l'expression des *GLUTs* et *HKs* est indépendante de la TEP normale (Figure 47, panel G).

La classification moléculaire de l'UAMS, au contraire, est statistiquement associée au rendu de l'imagerie TEP (Figure 47, panels H et I). En particulier, le groupe LB présente plutôt des TEP normales ($p<0.05$), et une sous-expression de l'*HK2* ($p=9E-4$) et de *GLUT5* ($p=1.6E-5$) par rapport au groupe de référence HP. En résumé, la classification moléculaire du myélome est associée à la TEP normale et à la dérégulation des gènes du métabolisme du glucose.

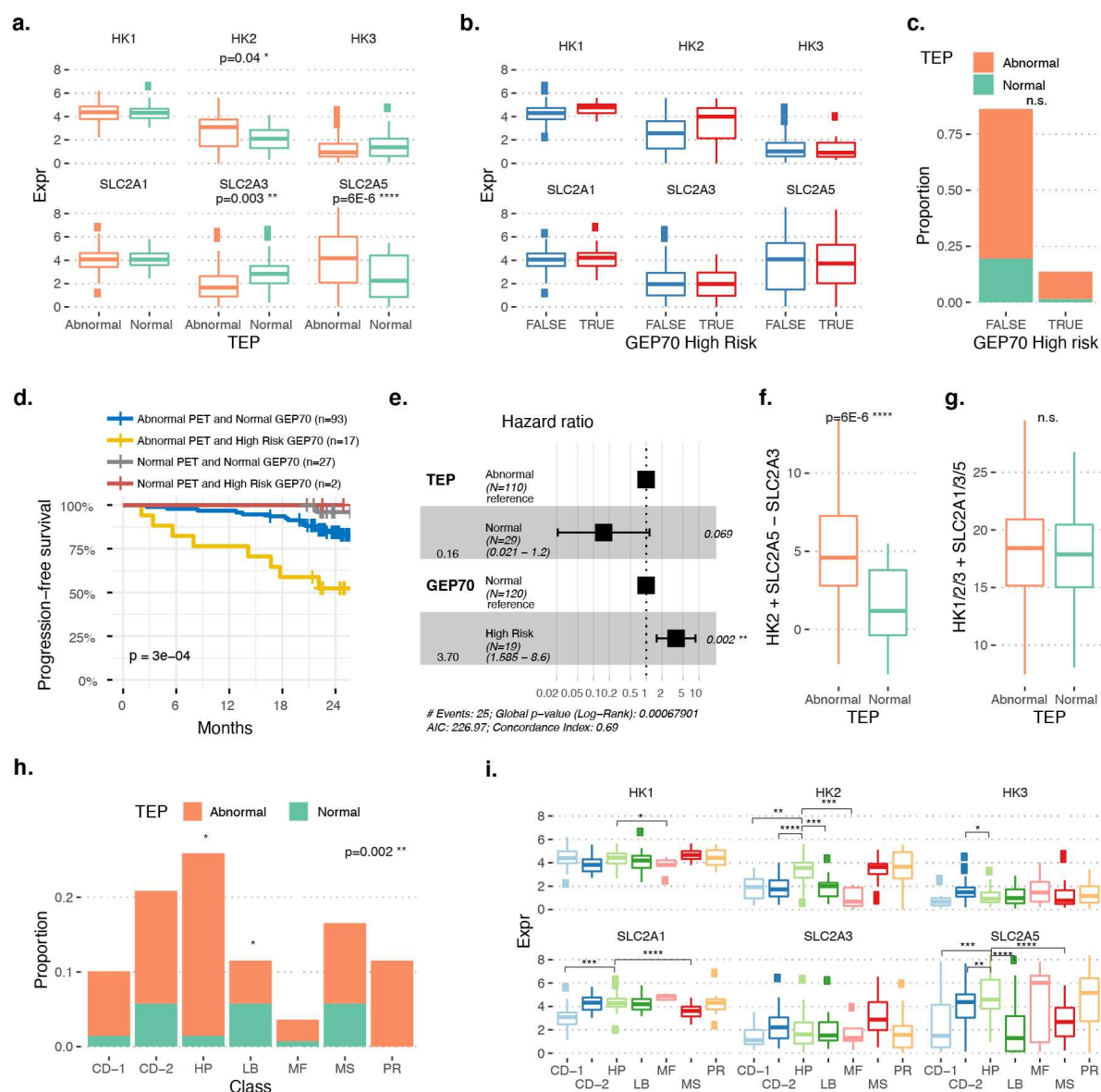


Figure 47 - Les TEP négatives sont associées à une dérégulation des programmes transcriptionnels incluant le métabolisme du glucose. **a.** Comparaison de l'expression des hexokinases (HKs) et des transporteurs de glucose (GLUTs) dans les myélomes à TEP normales ou anormales. **b.** Comparaison de l'expression des HKs et GLUTs entre myélomes à haut risque et à risque normal d'après la signature GEP70. **c.** Répartition des TEP normales entre myélomes à haut risque et myélomes à risque normales d'après la signature GEP70. **d.** Courbe de Kaplan-Meier de la survie sans progression (PFS) combinée entre TEP normales et score de haut risque. **e.** Graphique en forêt multivariée sur les TEP et GEP70 sur la survie sans progression. **f.** Moyenne pondérée des HKs et GLUTs dérégulées entre TEP normales et anormales. **g.** Somme comparée des HKs et GLUTs entre TEP normales et anormales. **h.** Répartition des TEP normales parmi les sous-groupes UAMS. **i.** Niveaux d'expression des HKs et GLUTs parmi les groupes de la classification moléculaire UAMS.

La classification moléculaire des patients à TEP normale est concentrée sur les patients LB, mais elle ne suffit pas à caractériser spécifiquement la normalité de la TEP au diagnostic. Une analyse différentielle de l'expression génique a donc été réalisée avec DEseq2 et montre que 1 202 gènes sont dérégulés entre les deux conditions. Les gènes moyennement à fortement exprimés (≥ 500 ARNm en moyenne) et sur lesquels la condition a un effet important ($\text{Log2FC} \geq 1$ en valeur absolue) sont présentés en Figure 48. Le criblage de banques de gènes par Gene Set Enrichment Analysis (GSEA) avec MSigDB est présenté en Figure 49. L'analyse d'ontologie montre qu'une partie des patients à TEP anormale (Figure 48, gauche et Figure 49, A) présente une signature de prolifération importante, alors que l'autre partie surexprime des antigènes de lymphocytes (*CD19*, *TNFSF8/CD30L*, *TNFSF10/TRAIL*) et *SLC2A5/GLUT5*. À l'inverse, les TEP normales (Figure 48, droite et Figure 49, B) présentent une ontologie de machinerie cellulaire (sécrétion, membrane, exocytose), et une expression régulière de *SLC2A3/GLUT3* en accord avec la Figure 47.

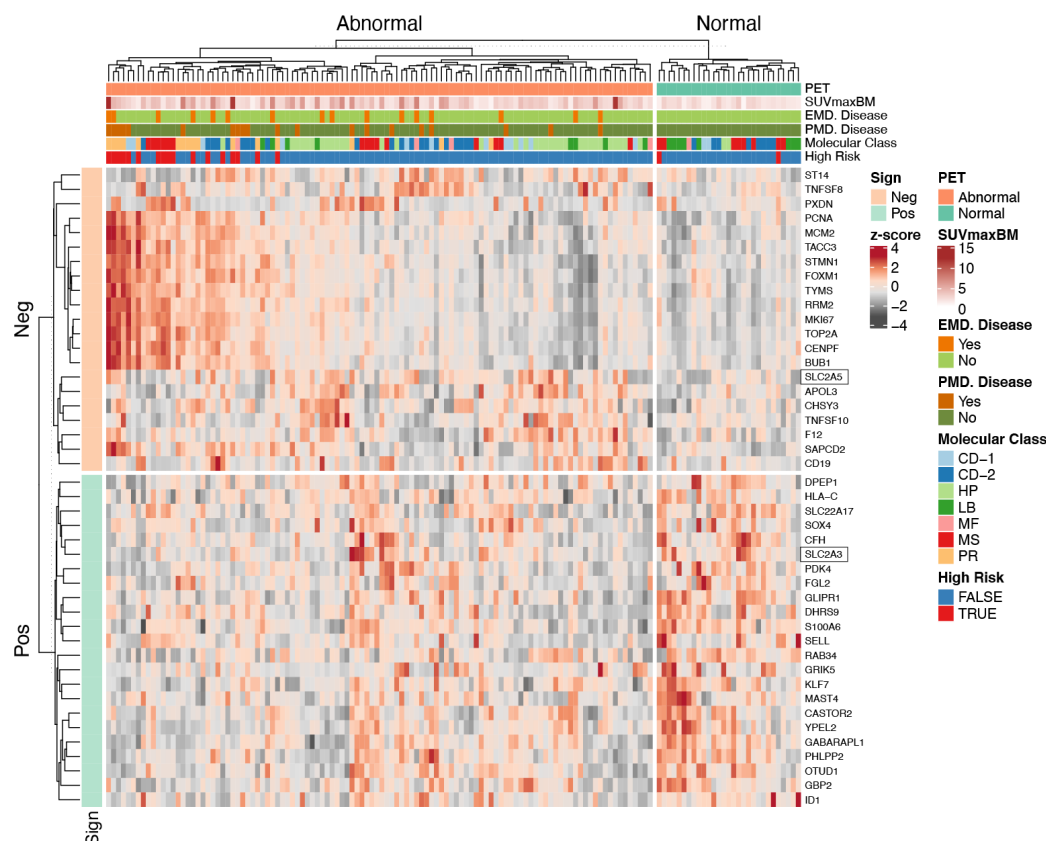


Figure 48 - Analyse d'expression différentielle entre TEP anormales (colonnes de gauche) et normales (droite). Les gènes représentés sont moyennement à fortement exprimés (≥ 500 comptes en moyenne), significativement sur- ou sous-exprimés (resp. lignes oranges et vertes) avec un $\text{log2foldchange} \geq 1$ en valeur absolue et une p -value ajustée < 0.05 . Les gènes de la famille des transporteurs de glucoses SLC2A sont encadrés.

a.

Entrez Gene Id	Gene Symbol	GO_MITOTIC_CELL_CYCLE	GO_CELL_CYCLE_PROCESS	GO_CELL_CYCLE	GO_CELL_CYCLE_PHASE_TRANSITION	GO_REGULATION_OF_CHROMOSOME_ORGANIZATION	GO_REGULATION_OF_CELL_CYCLE_PROCESS	GO_REGULATION_OF_CELL_CYCLE	GO_CHROMOSOME_SEPARATION	GO_REGULATION_OF_ORGANELLE_ORGANIZATION	GO_REGULATION_OF_CHROMOSOME_SEGREGATION	Entrez	Gene Description
1063	CENPF												centromere protein F [Source:HGNC Symbol;Acc:HGNC:1857]
699	BUB1												BUB1 mitotic checkpoint serine/threonine kinase [Source:HGNC Symbol;Acc:HGNC:1148]
10460	TACC3												transforming acidic coiled-coil containing protein 3 [Source:HGNC Symbol;Acc:HGNC:11524]
4171	MCM2												minichromosome maintenance complex component 2 [Source:HGNC Symbol;Acc:HGNC:6944]
5111	PCNA												proliferating cell nuclear antigen [Source:HGNC Symbol;Acc:HGNC:8729]
6241	RRM2												ribonucleotide reductase regulatory subunit M2 [Source:HGNC Symbol;Acc:HGNC:10452]
2305	FOXM1												forkhead box M1 [Source:HGNC Symbol;Acc:HGNC:3818]
7298	TYMS												thymidylate synthetase [Source:HGNC Symbol;Acc:HGNC:12441]
4288	MKI67												marker of proliferation Ki-67 [Source:HGNC Symbol;Acc:HGNC:7107]
7153	TOP2A												DNA topoisomerase II alpha [Source:HGNC Symbol;Acc:HGNC:11989]
3925	STMN1												stathmin 1 [Source:HGNC Symbol;Acc:HGNC:6510]
89958	SAPCD2												suppressor APC domain containing 2 [Source:HGNC Symbol;Acc:HGNC:28055]
8743	TNFSF10												TNF superfamily member 10 [Source:HGNC Symbol;Acc:HGNC:11925]
944	TNFSF8												TNF superfamily member 8 [Source:HGNC Symbol;Acc:HGNC:11938]
7837	PXDN												peroxidasin [Source:HGNC Symbol;Acc:HGNC:14966]
6768	ST14												suppression of tumorigenicity 14 [Source:HGNC Symbol;Acc:HGNC:11344]
930	CD19												CD19 molecule [Source:HGNC Symbol;Acc:HGNC:1633]
6518	SLC2A5												solute carrier family 2 member 5 [Source:HGNC Symbol;Acc:HGNC:11010]
2161	F12												coagulation factor XII [Source:HGNC Symbol;Acc:HGNC:3530]
80833	APOL3												apolipoprotein L3 [Source:HGNC Symbol;Acc:HGNC:14868]
337876	CHSY3												chondroitin sulfate synthase 3 [Source:HGNC Symbol;Acc:HGNC:24293]

b.

Entrez Gene Id	Gene Symbol	GO_SECRETION	GO_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND	GO_CELLULAR_RESPONSE_TO_OXYGEN_CONTAINING_COMPOUND	GO_VESICLE_MEMBRANE	GO_EXOCYTOSIS	GO_CARBOHYDRATE_HOMEOSTASIS	GO_MYELOID_LEUKOCYTE_MEDIATED_IMMUNITY	GO_SECRETORY_GRANULE_MEMBRANE	GO_MYELOID_LEUKOCYTE_ACTIVATION	GO_WHOLE_MEMBRANE	Entrez	Gene Description
2901	GRIK5												glutamate ionotropic receptor kainate type subunit 5 [Source:HGNC Symbol;Acc:HGNC:4583]
6659	SOX4												SRY-box transcription factor 4 [Source:HGNC Symbol;Acc:HGNC:11200]
8609	KLF7												Kruppel like factor 7 [Source:HGNC Symbol;Acc:HGNC:6350]
6402	SELL												selectin L [Source:HGNC Symbol;Acc:HGNC:10720]
3107	HLA-C												major histocompatibility complex, class I, C [Source:HGNC Symbol;Acc:HGNC:4933]
11010	GLIPR1												GLI pathogenesis related 1 [Source:HGNC Symbol;Acc:HGNC:17001]
6515	SLC2A3												solute carrier family 2 member 3 [Source:HGNC Symbol;Acc:HGNC:11007]
10875	FGL2												fibrinogen like 2 [Source:HGNC Symbol;Acc:HGNC:3696]
5166	PK4												pyruvate dehydrogenase kinase 4 [Source:HGNC Symbol;Acc:HGNC:8812]
1800	DPEP1												dipeptidase 1 [Source:HGNC Symbol;Acc:HGNC:3002]
3397	ID1												inhibitor of DNA binding 1, HLH protein [Source:HGNC Symbol;Acc:HGNC:5360]
729438	CASTOR2												cytosolic arginine sensor for mTORC1 subunit 2 [Source:HGNC Symbol;Acc:HGNC:37073]
83871	RAB34												RAB34, member RAS oncogene family [Source:HGNC Symbol;Acc:HGNC:16519]
23710	GABARAPL1												GABA type A receptor associated protein like 1 [Source:HGNC Symbol;Acc:HGNC:4068]
51310	SLC22A17												solute carrier family 22 member 17 [Source:HGNC Symbol;Acc:HGNC:23095]
3075	CFH												complement factor H [Source:HGNC Symbol;Acc:HGNC:4883]
2634	GBP2												guanylate binding protein 2 [Source:HGNC Symbol;Acc:HGNC:4183]
6277	S100A6												S100 calcium binding protein A6 [Source:HGNC Symbol;Acc:HGNC:10496]
23035	PHLPP2												PH domain and leucine rich repeat protein phosphatase 2 [Source:HGNC Symbol;Acc:HGNC:29149]
10170	DHRS9												dehydrogenase/reductase 9 [Source:HGNC Symbol;Acc:HGNC:16888]
220213	OTUD1												OTU deubiquitinase 1 [Source:HGNC Symbol;Acc:HGNC:27346]
375449	MAST4												microtubule associated serine/threonine kinase family member 4 [Source:HGNC Symbol;Acc:HGNC:19037]
388403	YPEL2												yippee like 2 [Source:HGNC Symbol;Acc:HGNC:18326]

Figure 49 - Analyse d'ontologie des gènes différentiellement exprimés entre TEP normales et anormales. MSIGDB v7.2 a été utilisé séparément sur les gènes fortement surexprimés dans les TEP anormales (A, haut) et normales (B, bas). Les ensembles de gènes de Gene Ontology ont été utilisés et seuls les 10 plus significatifs avec $FDR \leq 0.05$ sont montrés.

3.4. DISCUSSION ET CONCLUSION

Nous montrons dans ce chapitre des associations originales entre rendus d'imagerie et expression génique dans le myélome, ainsi qu'un détail des profils moléculaires de TEP normales par rapport à l'état de la connaissance (Rasche et al., 2017a).

Maladie para-médullaire (PMD) et signature d'expression de haut risque (GEP70) sont corrélées, contrairement à EMD et GEP70, ce qui suggère que ces deux types de lésions reflètent deux mécanismes liés à des mécanismes tumoraux distincts. La combinaison de la PMD et de la GEP70 en groupe « double-positifs » PMD et GEP70 en fait un marqueur intéressant de patients à très haut risque (6%) indépendant du R-ISS dans cette cohorte. De manière intéressante, le seul patient R-ISS de stade 1 mais double positif pour la PMD et la GEP70 a progressé en quatre mois et est décédé dans les deux ans. Des études supplémentaires devront évaluer ces modèles pour la survie globale, dans des cohortes supplémentaires et recevant d'autres traitements. L'association PMD et GEP70 n'est pas la première classification de très haut risque, on pourra également l'associer aux patients « Double-Hit » qui présentent une anomalie de *TP53* bi-allélique ou une amplification du chr1q avec un R-ISS élevé (Walker et al., 2019).

La deuxième partie des résultats s'est focalisée sur la caractérisation du groupe de patients avec une TEP normale, ou négative, au diagnostic. Les patients à TEP normale sont plus susceptibles d'appartenir au sous-groupe LB (*low bone disease*) de la classification de l'UAMS, un résultat cohérent puisque le groupe LB est caractérisé par un faible nombre de lésions détectées à l'IRM, et les deux groupes sont de bon pronostic clinique. Cette observation n'avait pas été faite par les études précédentes et est probablement due à la différence entre la définition de « TEP normales » dans notre étude et celle des « TEP fausses-négatives » de l'étude de l'UAMS (Rasche et al., 2017a).

Nous confirmons que les TEP normales sous-expriment *HK2*, et nous montrons que *GLUT3* et surtout *GLUT5* sont dérégulés dans une plus grande mesure que l'*HK2* entre TEP normales et anormales. De façon plus générale, lorsqu'on étend l'analyse à l'ensemble des gènes, deux signatures transcriptomiques ressortent pour les MM à TEP positive : la première implique des gènes de prolifération (*MKI67*, *PCNA*, *TOP2A*, *STMN1*) et les groupes proliférant (PR et MS). La seconde implique *GLUT5* et les antigènes de lymphocytes *CD19*, *CD30L* et *TRAIL*, ce qui suggère qu'un phénotype particulier est associé à l'avidité pour le glucose indépendamment de la prolifération. Cette observation pourra être validée au niveau protéique en cytométrie.

En résumé, les TEP normales semblent liées à la fois à une prolifération moins importante et à un phénotype de différenciation B/plasmocytaire distinct, bien que leur caractérisation ne soit pas encore totalement élucidée.

Les MM à TEP normale et faible SUVmax pourraient emprunter une voie métabolique alternative au glucose pour proliférer, comme le métabolisme glutaminique (Bolzoni et al., 2016), mais l'expression génique ne soutient cette hypothèse en général, sauf pour de rares patients qui expriment fortement les transporteurs glutaminiques.

La forte expression de *GLUT5*/*SLC2A5* est associée à la TEP positive. Ce résultat est inattendu puisque que *GLUT5* ne transporte pas le glucose mais le fructose (Burant et al., 1992; Nomura et al., 2015). *SLC2A5* est de manière générale le gène du métabolisme du glucose le plus discriminant entre TEP normales et anormales. La forte expression du *GLUT5* dans les tissus avides pour le FDG est évoquée dans la littérature où elle est associée à la TEP fausse-positive (Bono et al., 2017) dans les tumeurs bénignes ovariennes ou « en désaccord » (*discrepancy*) avec une faible expression du *GLUT1* dans le cancer du sein (Hamann et al., 2018). Des validations supplémentaires seront nécessaires pour expliquer le caractère artefactuel, non-habituel ou compensatoire de la forte expression de *GLUT5* chez les patients avides pour le FDG.

Nous avons donné ici un premier aperçu des associations statistiques de l'expression génique et biomarqueurs d'imagerie médicale. Nous confirmons d'abord que la sévérité de la pathologie se retrouve à la fois dans ses aspects moléculaires (signature de haut risque) et d'imagerie (présence de maladie para-médullaire), et que ces classes sont en fait liées voire s'additionnent. La TEP normale reflète l'hétérogénéité transcriptomique de la maladie via l'expression des *GLUTs*, *HKs* et par classification moléculaire. Finalement, des traceurs autres que le FDG pourront également aider à tisser plus de liens entre programmes d'expression génique et rendus d'imagerie médicale (Jamet et al., 2019b).

DISCUSSION ET CONCLUSIONS

Ma thèse contribue à la description de l'hétérogénéité génomique inter- et intra-patient dans le myélome multiple, et fournit des exemples concrets de l'impact de la diversité génomique de cette tumeur sur les aspects cliniques de la maladie.

1. DÉCOUVERTE DE NOUVEAUX BIOMARQUEURS ET ÉTUDE DE LA CHROMATINE

Les travaux réalisés autour de la cartographie de l'hydroxyméthylation des cytosines (5hmC) sur l'ADN des cellules de myélome multiple sont novateurs. La distribution et le niveau de 5hmC étaient inconnus dans la plupart des tumeurs en 2017 à l'initiation de ce projet. L'état général de la chromatine dans le myélome était aussi globalement inconnu et nos résultats viennent s'ajouter aux travaux d'autres groupes qui décrivent l'ouverture et l'activation de la chromatine (Jin et al., 2018; Ordoñez et al., 2020). Nous avons réussi à montrer que la 5hmC est bien présente dans le myélome multiple et qu'un niveau faible de 5hmC est associé à une sévérité plus élevée de la maladie, contrairement au niveau global de méthylation (5mC). Pour aller au-delà de la recherche fondamentale et faire du niveau global de 5hmC un biomarqueur, il faudra (1) tester les résultats préliminaires sur plus de patients, (2) s'assurer de la reproductibilité de la manipulation technique, et (3) l'intégrer dans une problématique clinique.

1.1. DESIGN EXPÉRIMENTAL POUR QUANTIFIER LA 5HMC

Notre étude a bénéficié de l'environnement structurant précieux de l'IFM pour la recherche. Ce réseau s'est assuré de l'homogénéité du traitement des échantillons biologiques de la cohorte clinique IFM/DFCI 2009, de la purification des plasmocytes, de l'extraction et la conservation de l'ADN ainsi que du suivi clinique d'un nombre important de patients. Dans notre étude en particulier, le nombre d'échantillons analysés a été fixé par des contraintes techniques et de coût et était sans doute insuffisant. Maintenant que nous connaissons mieux la distribution des niveaux de 5hmC dans le myélome, nous pourrions adopter une approche similaire à celles des essais cliniques (Lachin, 1981) pour estimer le nombre d'échantillons suffisants pour tester nos hypothèses.

En plus du nombre important d'échantillons nécessaires, nous devons nous assurer de la reproductibilité de la technique et des analyses. La reproductibilité de la

quantification de la 5hmC est facilitée par la publication de protocoles (Song et al., 2011a; Traube et al., 2019), mais la technique que nous avons choisie rend compliquée la comparaison d'échantillons de patients différents. L'enrichissement nécessite en effet deux contrôles : un référentiel d'hydroxyméthylation normale, sur des régions « contrôles » du génome par exemple, et un échantillon « contrôle » pour normaliser les biais d'amplification, de nombre de copies et de séquençage (input-sequencing). La quantification par OxBS/BS donnerait un pourcentage de 5hmC sur un locus donné qui facilite la comparaison d'échantillons. La définition d'un pourcentage « plancher » à partir duquel le marqueur est positif rend le travail plus intelligible et me semble plus approprié. En revanche, l'approche OxBS/BS sur une cohorte de patients ne pourra être envisagée qu'avec une baisse importante du coût du séquençage. L'approche par puces ADN nous aurait permis de gagner en puissance statistique et de comparer plus de patients, mais elle aurait limité notre étude à 5% des CpG du génome humain et ne permettait pas une approche *genome-wide*, ce qui aurait limité les chances de découvrir de nouvelles régions régulatrices.

1.2. NOUVEAUX AXES DE DESCRIPTION DE LA 5HMC

L'approche que j'ai choisie, de montrer la corrélation entre niveau global de 5hmC et sévérité de la maladie (diagnostic contre contrôle, ISS, OS), a joué un rôle dual auprès des relecteurs. D'abord, il a semblé convaincre de la puissance pronostique potentielle de cette marque chimique stable déposée sur l'ADN. Cette étape est importante puisque la 5hmC a été découverte récemment et qu'elle n'a pas de rôle fonctionnel clairement établi dans le cancer. Mais cette approche a aussi joué en notre défaveur pour intéresser le lectorat : pourquoi s'intéresser à un nouveau biomarqueur s'il est déjà corrélé aux biomarqueurs existants ?

Cette critique, que je partage, nous invite à développer de nouveaux axes autour de la quantification de la 5hmC : est-ce un prédicteur susceptible d'affiner les variables utilisées en routine ? Peut-on quantifier la 5hmC de manière non-invasive via l'ADN libre circulant (cfDNA, Li et al. 2017) ou la purification des cellules tumorales circulantes (CTC, Manier et al., 2018) ? On imagine par exemple que l'étude d'une signature « Myélome » dans la 5hmC sanguine pourrait aider à la détection précoce du cancer ou au suivi clinique de la maladie, une approche existante avec la méthylation dans d'autres cancers (Kang et al., 2017). Aussi, la 5hmC pourrait caractériser des marqueurs prédictifs de la progression du myélome depuis des stades indolents.

1.3. BÉNÉFICES LIÉS À L'UTILISATION DE LA 5hmC EN CLINIQUE

Contrairement aux signatures d'expression génique qui nécessitent l'extraction d'ARN, et contrairement à l'étude de la chromatine, qui nécessite des techniques de conservation spécifiques, la 5hmC se mesure directement sur l'ADN génomique. En clinique, l'utilisation d'ADN génomique pour caractériser l'activité de la chromatine a un intérêt évident : éviter de stocker de l'ARN et des noyaux. Le même échantillon d'ADN pourrait ainsi alimenter la pratique de routine (translocations, nombre de copies, mutations ciblées), la recherche en génomique (mutations non-codantes, signatures mutationnelles, sélection clonale, ...), et la recherche en épigénomique (régions régulatrices, chromatine active, ...). Puisque l'ADN a été un support de stockage de nombreux essais cliniques, il est finalement plus aisé de construire des études rétrospectives avec la 5mC et 5hmC qu'avec par exemple les marques d'histone ou l'ouverture de la chromatine.

1.4. HÉTÉROGÉNÉITÉ GÉNOMIQUE DU MYÉLOME ET HYDROXYMÉTHYLATION

Localement, à l'échelle d'un gène par exemple, les niveaux variables de 5hmC permettent de découvrir des associations originales. Nous avons montré que la corrélation linéaire du niveau de 5hmC entre deux loci permet de découvrir des régions régulatrices potentielles des gènes cyclines D. Ce « continuum biologique » de l'expression génique en cohorte n'existe pas avec les lignées cellulaires, et fait l'objet de débats. Par exemple, peut-on prédire une t(11;14) sur la base de l'expression de *CCND1* ? L'étude des lignées cellulaires suggère que oui, l'étude des cohortes n'est pas si catégorique. En conséquence, les modèles linéaires des études cliniques nous semblent plus robustes qu'avec des lignées cellulaires, car la continuité biologique est plus susceptible de protéger des fausses corrélations dues aux données aberrantes ou extrêmes résumées dans le fameux quartet d'Anscombe.

Pour étudier des mécanismes causaux et non des corrélations, nous avons besoin d'un modèle à perturber *in vitro*. Or, nous n'avons pas de modèle cellulaire pour étudier la 5hmC dans le MM. Malheureusement, aucune lignée de myélome séquencée par le Translational Genomics Research Institute (TGen) n'exprime fortement les gènes *TET1/2/3* responsables de l'oxydation de la méthylation en hydroxyméthylation des cytosines. Nous avons trouvé que le niveau de 5hmC par dot blot et par séquençage dans la lignée MM.1S était trop faible pour faire office de modèle cellulaire.

À l'avenir, peut-être pourrions-nous cribler une banque de lignées cellulaires de myélome à la recherche de 5hmC par spectrométrie de masse (Traube et al., 2019), ou bien induire l'expression des protéines TET dans une lignée d'intérêt.

1.5. LE MYÉLOME EST-IL UN CANCER DE L'HÉMATOPOÏÈSE ?

Les régulateurs épigénétiques *DNMT3A* et *TET2* concentrent la majorité des mutations somatiques chez les individus sains âgés qui présentent une hématopoïèse clonale (Challen and Goodell, 2020). Les mutations de *DNMT3A* et *TET2* sont récurrentes dans les cancers hématologiques et font l'objet d'articles dédiés dans le *New England Journal of Medicine* (Delhommeau et al., 2009; Ley et al., 2010). Les deux protéines orchestrent la différenciation cellulaire hématopoïétique et la méthylation du génome : il est naturel d'étudier leur rôle, ainsi que la distribution de la méthylation, dans le myélome multiple.

Pourtant, *DNMT3A* et *TET2* sont rarement mutées dans le myélome, qui est un cancer de lymphocyte B pleinement différencié en plasmocyte. D'autre part, l'événement oncogénique primaire et de mauvais pronostic à la prévalence la plus importante est la t(4;14). La t(4;14) juxtapose *NSD2*, méthyltransférase d'histone, au locus des immunoglobulines et cause sa forte sur-expression. *NSD2* catalyse la diméthylation d'H3K36, qui suffit à activer l'expression génique, et la t(4;14) cause une augmentation globale et intergénique de l'H3K36me2 associée à une perte globale d'H3K27me3 (Li et al., 2009; Kuo et al., 2011; Popovic et al., 2014). La découverte récente que la marque H3K36me2 recrute *DNMT3A* à la chromatine (Weinberg et al., 2019) nous invite à faire deux hypothèses : (1) que *DNMT3A* méthyle *de novo* les régions marquées de l'H3K36me2 dans les myélomes t(4;14), et (2) que les myélomes t(4;14) sont associés à une chromatine désorganisée de manière anarchique.

Les techniques de séquençage unicellulaire (scRNA-seq, scATAC-seq, scCNV-seq, etc.) pourraient montrer par exemple une plus grande diversité des sous-clones chez les patients t(4;14) associée au mauvais pronostic clinique. Cet axe de recherche serait bien distinct de l'étude de la différenciation hématopoïétique linéaire défectueuse déséquilibrée par les anomalies ciblant *DNMT3A* et *TET2*.

De plus, ce nouvel axe de recherche peut être exploré dans d'autres tumeurs hématologiques : 10% des leucémies lymphoblastiques aiguës pédiatriques (B-ALL) présentent la mutation activatrice *NSD2_E1099K* (Pierro et al., 2020). Cette mutation somatique est associée à une augmentation et à une redistribution de l'H3K36me2, ainsi qu'à un transcriptome malin et à une prolifération accrue (Jaffe et al., 2013;

Swaroop et al., 2019; Pierro et al., 2020), mais son incidence sur la méthylation de l'ADN et l'hétérogénéité transcriptomique reste à élucider.

Au-delà du clone génétique et de la différenciation cellulaire, la modification non-génétique du cancer nous encourage en tout cas à étendre notre définition du clone tumoral aux clones épigénétiques.

2. DIVERSITÉ DE LA SÉLECTION CLONALE DANS LA RÉSISTANCE ACQUISE AU TRAITEMENT

L'évolution clonale des cellules tumorales en compétition pour proliférer est un dogme majeur de la cancérologie (Nowell, 1976), mais peu de travaux observent finement ce phénomène à l'œuvre *in vivo*. Notre description des anomalies génomiques de la famille du BCL2 et sélectionnés sous pression de la molécule anti-BCL2 venetoclax, fournissent un exemple concret d'émergence de clones résistants au traitement. Si cette étude a été entièrement conçue par nos collaborateurs de l'Université de Calgary au Canada, les analyses bio-informatiques que j'ai mises en place ont permis de quantifier les résultats, de les mettre en forme, et de les compléter par des événements qui n'avaient pas été vus « à la main ».

Le venetoclax a montré son efficacité dans les myélomes de type t(11;14) réfractaires ou en rechute (Kumar et al., 2017b, 2020). Pourtant, la résistance des cellules de myélome au venetoclax est courante et est une cause d'échec du traitement. Puisqu'une des hypothèses avancées est que l'hétérogénéité génomique des cellules de myélomes induit une expression variable des gènes de la famille du BCL2, nous nous sommes intéressés au profil de cette famille dans des données de séquençage unicellulaire de nombre de copies (single-cell CNV-seq), d'ouverture de la chromatine (single-cell ATAC-seq) et d'expression (single-cell RNA-seq).

Nous observons deux événements structuraux récurrents : (1) le gain ou l'amplification de l'anti-apoptotique *MCL1* associée à une ouverture de la chromatine et une augmentation de son expression, et (2) la translocation de l'anti-apoptotique *BCLxL* avec IgL ou IgH associée à une ouverture de la chromatine aux points de cassure qui pourrait être due à une amplification focale, et à une augmentation de l'expression. Pour autant, aucune des anomalies récurrentes acquises n'est systématique ni exclusive, c'est-à-dire qu'aucune anomalie n'est commune à tous les patients et que les patients présentent plusieurs de ces anomalies en même temps.

La surexpression de l'anti-apoptotique *MCL1* est un phénomène attendu à la rechute, puisque le gain du chr1q est associé à la progression clinique indépendamment du traitement. Aussi, la plus forte dépendance fonctionnelle au *MCL1* à la rechute est décrite en détails par l'équipe du Dr Martine Amiot (Gomez-Bougie et al., 2018). Mais notre étude montre pour la première fois que le gain du gène *MCL1* peut se faire sans le gain proximal du gène *CKS1B*, qui est le marqueur du gain1q utilisé en routine. Ce résultat suggère que ce sont bien les clones « *MCL1* » qui émergent par pression de sélection. D'un point de vue pratique, cela signifie aussi que la FISH de *CKS1B* n'est pas suffisante pour caractériser le gain du chr1q sous venetoclax. De plus, dans un clone minoritaire qui ne présente pas de gain1q, on trouve une délétion mono-allélique du locus du pro-apoptotique *BIM*, ce qui renforce encore l'hypothèse que la pression du traitement favorise l'émergence de clones résistant spécifiquement à l'anti-BCL2. Plus généralement, l'utilisation de traitements ciblés comme le venetoclax invite à utiliser des nouveaux marqueurs génétiques dans la pratique clinique pour prédire la sensibilité et surveiller la réponse.

2.1. COMPLÉMENTARITÉ DE LA GÉNOMIQUE ET DES TESTS FONCTIONNELS DE SENSIBILITÉ AUX BH3-MIMÉTIQUES

Si une partie des résultats liés à l'étude de la sélection clonale sous pression du venetoclax a été validée sur un modèle *in vitro* par nos collègues de l'Université de Calgary, il demeure que pour un patient donné, le lien de causalité entre anomalie acquise et résistance observée n'est pas montré.

La sensibilité au traitement dépend aussi de facteurs extrinsèques à la cellule tumorale (immunitaire, environnemental, ...), et de facteur intrinsèques non-génétiques (interactions protéiques, signalisation, ...). Le système de culture de moelle osseuse *ex vivo* et de test de sensibilité aux BH3-mimétiques mis en place par les équipes du Dr Martine Amiot à Nantes et Dr Lawrence Boise à Atlanta est complémentaire à la génomique en ce sens : il fournit un résultat fonctionnel qui se veut prédictif de la sensibilité clinique. En revanche, l'approche fonctionnelle ne détecte pas la présence d'un clone minoritaire en cours de sélection, ni des facteurs comme la délétion bi-allélique d'un gène suppresseur de tumeur ou l'amplification d'un oncogène. À mon sens, on peut résumer ainsi la complémentarité des approches dans le futur : le test fonctionnel pourrait prédire la sensibilité globale au traitement, et la génomique fournirait des raisons suffisantes d'exclure le traitement, comme par exemple la présence d'un clone complètement résistant.

2.2. INTÉRÊT DES MUTATIONS DANS LE MYÉLOME

Dans le myélome multiple en général, les mutations ponctuelles sont sous-clonales et ne sont pas communes aux patients. On peut donc considérer qu'elles reflètent surtout l'hétérogénéité de la maladie et le succès d'adaptation des clones. Nos travaux sur le venetoclax et les travaux d'autres équipes sur la résistance acquise aux immunomodulateurs IMiD (Gooding et al., 2020) suggèrent aussi que la génomique peut trouver sa place en clinique dans la prédiction de la réponse au traitement et dans la surveillance au cours du traitement. Encore une fois, cette piste de recherche médicale demande à ce que (1) les résultats soient montrés sur un nombre important de patients, (2) que la manipulation technique soit reproductible, et (3) qu'elle s'intègre dans une problématique clinique.

Le couplage de la génomique à un test fonctionnel pourrait demander une centralisation des échantillons et une capacité à travailler sur matériel congelé à l'avenir, car les techniques peuvent être compliquées à mettre en place dans tous les centres hospitaliers. La reproductibilité technique de l'analyse de séquençage est permise par les outils bio-informatiques standardisés développés par exemple par le Broad Institute. Enfin, la question posée en recherche pourrait être liée directement à l'utilisation des BH3-mimétiques : quels sont les patients susceptibles de bénéficier de telle ou telle molécule, seule ou en combinaison ? Cette question fait l'objet d'un axe du Site de Recherche Intégrée sur le Cancer (SIRIC) ILIAD à Nantes sur le myélome et le cancer du sein (Gomez-Bougie et al., 2018; Lohard et al., 2020). L'étude longitudinale pourrait permettre une détection précoce de la résistance acquise par identification de variants dans les cellules tumorales circulantes et l'ADN libre-circulant (Manier et al., 2018), notamment dans le cas de thérapies ciblées où l'intégrité de la cible biologique peut être contrôlée. Enfin, les anomalies clonales sélectionnées et les tests fonctionnels pourraient indiquer dans quelles directions opère la pression de sélection sur les sous-clones de myélome. Le biais de sélection clonale induit par une molécule pourrait révéler des vulnérabilités nouvelles à d'autres traitements. En fait, la résistance au traitement des clones résiduels peut jouer en notre faveur et s'établir au profit d'une « sensibilité collatérale » (ou « vulnérabilité acquise ») à d'autres options thérapeutiques (Zhao et al., 2016; Wang et al., 2018; Acar et al., 2020), un modèle démontré dans d'autres cancers. À notre connaissance, cet axe de recherche reste inexploité dans le myélome.

3. INTÉRÊTS PRONOSTIQUES ET SCIENTIFIQUES DE LA COMBINAISON DE LA GÉNOMIQUE À D'AUTRES VARIABLES

Nous avons présenté des relations originales entre les biomarqueurs de la TEP-FDG et le RNA-seq au diagnostic et qui sont discutées dans le paragraphe 3.4 (p. 150). Ces résultats s'inscrivent dans une problématique plus large : comment intégrer différentes approches technologiques pour en apprendre plus sur la maladie ?

3.1. GÉNOMIQUE ET HÉTÉROGÉNÉITÉ SPATIALE DE LA MALADIE

Nous avons montré qu'à l'échelle moléculaire, les deux types de lésions para-médullaire (PMD) et extra-médullaire (EMD) ont des caractéristiques distinctes. La présence de PMD est statistiquement associée à l'appartenance aux sous-groupes proliférant et de haut risque des classifications de l'UAMS. À l'inverse, nous avons montré que l'EMD n'est associée à aucune des deux classifications.

Les programmes transcriptionnels malins que nous mesurons sont issus des plasmocytes purifiés de la moelle osseuse mais pas des lésions extra-médullaires. Il se pourrait que le transcriptome tumoral au sein des lésions extra-médullaires soit distinct de ceux de la moelle osseuse et des lésions para-médullaires contiguës à l'os. L'observation d'un phénotype propre à la maladie extramédullaire dans un modèle murin appuie cette hypothèse (Roccaro et al., 2015). Les études spatiales des mutations disséminées dans les lésions (de Haart et al., 2016; Rasche et al., 2017b) et des spécificités phénotypiques de l'EMD (Bladé et al., 2011) apportent un regard supplémentaire sur l'hétérogénéité du myélome intra-patient et sur lequel nous pourrions également nous appuyer.

3.2. COMBINAISON DES MARQUEURS GÉNOMIQUES ET D'IMAGERIE

PMD et signature d'expression de haut risque peuvent être associées dans une stratification clinique discutée également dans le paragraphe 3.4, page 150, qui concerne 6% des patients pour une PFS médiane inférieure à 18 mois. Le caractère additif de l'information pronostique apportée par chacun des marqueurs pourrait fournir une classification de risque affinée par rapport à chacune des deux variables considérées indépendamment. Cette nouvelle classification nécessite d'être testée sur une cohorte de taille plus importante et d'être comparée avec d'autres classifications de très haut risque (Walker et al., 2019). La validation en survie globale requiert une durée de suivi étendue de plusieurs années par rapport à notre étude.

L'augmentation du nombre de patients et de cohortes testées permettra de resserrer l'intervalle de confiance autour d'une valeur significative de rapport des risques et de décliner notre modèle en fonction des interventions thérapeutiques et des stades de la maladie. De nouvelles signatures d'expression pourraient permettre de tester génétiquement le bénéfice lié à l'ajout de traitements comme cela est fait dans le cancer du sein (Kwa et al., 2017; Sparano et al., 2018), et de prédire des variables originales comme les effets indésirables spécifiques liés à l'apparition des nouvelles classes de médicaments dans le MM (Ludwig et al., 2018). Il demeure que, contrairement à ce que pouvait suggérer notre première étude portant sur la 5hmC, la corrélation de plusieurs biomarqueurs n'est pas un frein à leur intérêt pronostique combiné.

3.3. RETOUR AU CLONE TUMORAL

En somme, l'hétérogénéité inter- et intra-patient dans le myélome est caractérisée par une complexité de la maladie aux niveaux génétique, épigénétique et transcriptionnel. Cette complexité reflète l'avancement de la maladie, l'hétérogénéité fonctionnelle des clones tumoraux, et leur sélection sous pression thérapeutique.

Nous l'avons vu avec le Chapitre 2 notamment, les analyses *single-cell* permettent de pallier des biais des technologies *bulk* et de gagner en puissance descriptive à l'échelle du patient. À l'échelle du sous-clone désormais, c'est le séquençage en cellule-unique qui devient la première brique de description de l'hétérogénéité tumorale. Il est fort à parier qu'à ce niveau, les technologies existantes (RNA-seq, ATAC-seq, ...), émergentes (HiC, ChIP-seq, HiChIP ...), voire multi-omiques (NMT-seq, G&T-seq, RNA+ATAC-seq, ...) vont nous inviter à :

- (1) étendre la notion de clone génétique au clone épigénétique,
- (2) étudier l'évolution temporelle et spatiale de la tumeur à la lumière des avantages relatifs des sous-clones à se multiplier,
- (3) intégrer les interactions entre les cellules tumorales et le microenvironnement médullaire dans l'étude de la compétition clonale pour survivre et proliférer.

BIBLIOGRAPHIE

Abe, Y., Narita, K., Kobayashi, H., et al. (2019). Pretreatment 18F-FDG PET/CT combined with quantification of clonal circulating plasma cells as a potential risk model in patients with newly diagnosed multiple myeloma. *Eur. J. Nucl. Med. Mol. Imaging* 46, 1325–1333.

Acar, A., Nichol, D., Fernandez-Mateos, J., et al. (2020). Exploiting evolutionary steering to induce collateral drug sensitivity in cancer. *Nat. Commun.* 11, 1–14.

Agirre, X., Castellano, G., Pascual, M., et al. (2015). Whole-epigenome analysis in multiple myeloma reveals DNA hypermethylation of B cell-specific enhancers. *Genome Res.* 25, 478–487.

Akkaya, M., Kwak, K., and Pierce, S.K. (2020). B cell memory: building two walls of protection against pathogens. *Nat. Rev. Immunol.* 20, 229–238.

Alberge, J.B., et Minvielle, S. (2020). Évolution et sélection clonales : modèles et enjeux dans le myélome multiple. *Corresp. En Onco-Hématologie* XV, 293–296.

Alberge, J.-B., Sinha, S., Maity, R., et al. (2019). IGLL5-BCL2L1 Rearrangement with Loss of BCL2 Dependency As Mechanism of Venetoclax Resistance in Multiple Myeloma (MM). *Blood* 134, 686–686 (Supplements ASH conference).

Alberge, J.-B., Magrangeas, F., Wagner, M., et al. (2020a). DNA hydroxymethylation is associated with disease severity and persists at enhancers of oncogenic regions in multiple myeloma. *Clin. Epigenetics* 12, 163.

Alberge, J.-B., Jamet, B., Bailly, C., et al. (2020b). Genome-Wide Transcriptome Analysis Identifies Molecular Patterns of FDG-PET/CT Biomarkers in MM Patients from the Cassiopet Study. *Blood* 136, 26–26 (Supplements ASH conference).

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., et al. (2013). Signatures of mutational processes in human cancer. *Nature* 500, 415–421.

Alexandrov, L.B., Kim, J., Haradhvala, N.J., et al. (2020). The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101.

Altieri, D.C. (2013). Targeting survivin in cancer. *Cancer Lett.* 332, 225–228.

Altieri, A., Chen, B., Bermejo, J.L., et al. (2006). Familial risks and temporal incidence trends of multiple myeloma. *Eur. J. Cancer* 42, 1661–1670.

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* 9, 9354.

Amodio, N., Stamato, M.A., Juli, G., et al. (2018). Drugging the lncRNA MALAT1 via LNA gapmeR ASO inhibits gene expression of proteasome subunits and triggers anti-multiple myeloma activity. *Leukemia* 32, 1948–1957.

Ankathil, R., Madhavan, J., Gangadharan, V.P., et al. (1995). Nonrandom karyotype abnormalities in 36 multiple myeloma patients. *Cancer Genet. Cytogenet.* 83, 71–74.

Ashoor, H., Hérault, A., Kamoun, A., et al. (2013). HMCAN: a method for detecting chromatin modifications in cancer samples using ChIP-seq data. *Bioinformatics* 29, 2979–2986.

Attal, M., Lauwers-Cances, V., Hulin, C., et al. (2017). Lenalidomide, Bortezomib, and Dexamethasone with Transplantation for Myeloma. *N. Engl. J. Med.* 376, 1311–1320.

Avet-Loiseau, H., Gerson, F., Magrangeas, F., et al. (2001). Rearrangements of the c-myc oncogene are present in 15% of primary human multiple myeloma tumors. *Blood* 98, 3082–3086.

Avet-Loiseau, H., Attal, M., Moreau, P., et al. (2007). Genetic abnormalities and survival in multiple myeloma: the experience of the Intergroupe Francophone du Myélome. *Blood* 109, 3489–3495.

Avet-Loiseau, H., Li, C., Magrangeas, F., et al. (2009). Prognostic Significance of Copy-Number Alterations in Multiple Myeloma. *J. Clin. Oncol.* 27, 4585–4590.

Avet-Loiseau, H., Attal, M., Campion, L., et al. (2012). Long-Term Analysis of the IFM 99 Trials for Myeloma. *J. Clin. Oncol.* 30, 1949–1952.

Avet-Loiseau, H., Fonseca, R., Siegel, D., et al. (2016). Carfilzomib significantly improves the progression-free survival of high-risk patients in multiple myeloma. *Blood* 128, 1174–1180.

Bachman, M., Uribe-Lewis, S., Yang, X., et al. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* 6, 1049–1055.

Bailly, C., Bodet-Milin, C., Bourgeois, M., et al. (2019). Exploring Tumor Heterogeneity Using PET Imaging: The Big Picture. *Cancers* 11, 1282.

Barwick, B.G., Neri, P., Bahlis, N.J., et al. (2019). Multiple myeloma immunoglobulin lambda translocations portend poor prognosis. *Nat. Commun.* 10, 1–13.

Baubec, T., Ivánek, R., Lienert, F., et al. (2013). Methylation-Dependent and -Independent Genomic Targeting Principles of the MBD Protein Family. *Cell* 153, 480–492.

Béné, M.-C. (2020). MRD: minimale ou mesurable ? *Corresp. En Onco-Hématologie* XV, 219.

Bergsagel, P.L., Kuehl, W.M., Zhan, F., et al. (2005). Cyclin D dysregulation: an early and unifying pathogenic event in multiple myeloma. *Blood* 106, 296–303.

Berman, B.P., Weisenberger, D.J., Aman, J.F., et al. (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* 44, 40–46.

Beroukhi, R., Mermel, C.H., Porter, D., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature* 463, 899–905.

Birkinshaw, R.W., Gong, J., Luo, C.S., et al. (2019). Structures of BCL-2 in complex with venetoclax reveal the molecular basis of resistance mutations. *Nat. Commun.* 10, 1–10.

Bladé, J., Fernández de Larrea, C., Rosiñol, L., et al. (2011). Soft-Tissue Plasmacytomas in Multiple Myeloma: Incidence, Mechanisms of Extramedullary Spread, and Treatment Approach. *J. Clin. Oncol.* 29, 3805–3812.

Blombery, P., Anderson, M.A., Gong, J., et al. (2019). Acquisition of the Recurrent Gly101Val Mutation in BCL2 Confers Resistance to Venetoclax in Patients with Progressive Chronic Lymphocytic Leukemia. *Cancer Discov.* 9, 342–353.

Boeva, V., Popova, T., Bleakley, K., et al. (2012). Control-FREEC: a tool for assessing copy number and allelic content using next-generation sequencing data. *Bioinformatics* 28, 423–425.

Boeva, V., Louis-Brennetot, C., Peltier, A., et al. (2017). Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nat. Genet.* 49, 1408–1413.

Boise, L.H., Kaufman, J.L., Bahlis, N.J., et al. (2014). The Tao of myeloma. *Blood* 124, 1873–1879.

Bolli, N., Avet-Loiseau, H., Wedge, D.C., et al. (2014). Heterogeneity of genomic evolution and mutational profiles in multiple myeloma. *Nat. Commun.* 5, 2997.

Bolli, N., Biancon, G., Moarii, M., et al. (2018). Analysis of the genomic landscape of multiple myeloma highlights novel prognostic markers and disease subgroups. *Leukemia* 32, 2604–2616.

Bolzoni, M., Chiu, M., Accardi, F., et al. (2016). Dependence on glutamine uptake and glutamine addiction characterize myeloma cells: a new attractive target. *Blood* 128, 667–679.

Bono, Y., Mizumoto, Y., Nakamura, M., et al. (2017). FDG-PET-positive ovarian thecoma with GLUT5 expression: Five cases. *J. Obstet. Gynaecol. Res.* 43, 599–603.

Booth, M.J., Branco, M.R., Ficz, G., et al. (2012). Quantitative Sequencing of 5-Methylcytosine and 5-Hydroxymethylcytosine at Single-Base Resolution. *Science* 336, 934–937.

Bostick, M., Kim, J.K., Estève, P.-O., et al. (2007). UHRF1 Plays a Role in Maintaining DNA Methylation in Mammalian Cells. *Science* 317, 1760–1764.

Bourc’his, D., Xu, G.L., Lin, C.S., et al. (2001). Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536–2539.

Broyl, A., Hose, D., Lokhorst, H., et al. (2010). Gene expression profiling for molecular classification of multiple myeloma in newly diagnosed patients. *Blood* 116, 2543–2553.

Burant, C.F., Takeda, J., Brot-Laroche, E., et al. (1992). Fructose transporter in human spermatozoa and small intestine is GLUT5. *J. Biol. Chem.* 267, 14523–14526.

Bustoros, M., Sklavenitis-Pistofidis, R., Kapoor, P., et al. (2019). Progression Risk Stratification of Asymptomatic Waldenström Macroglobulinemia. *J. Clin. Oncol.* 37, 10.

Bustoros, M., Sklavenitis-Pistofidis, R., Park, J., et al. (2020). Genomic Profiling of Smoldering Multiple Myeloma Identifies Patients at a High Risk of Disease Progression. *J. Clin. Oncol.* 38, 2380–2390.

Butler, A., Hoffman, P., Smibert, P., et al. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.

Caenepeel, S.R., Belmontes, B., Sun, J., et al. (2017). Abstract 2027: Preclinical evaluation of AMG 176, a novel, potent and selective Mcl-1 inhibitor with robust anti-tumor activity in Mcl-1 dependent cancer models. *Cancer Res.* 77, 2027–2027.

Cavo, M., Terpos, E., Nanni, C., et al. (2017). Role of 18F-FDG PET/CT in the diagnosis and management of multiple myeloma and other plasma cell disorders: a consensus statement by the International Myeloma Working Group. *Lancet Oncol.* 18, e206–e217.

Certo, M., Moore, V.D.G., Nishino, M., et al. (2006). Mitochondria primed by death signals determine cellular addiction to antiapoptotic BCL-2 family members. *Cancer Cell* 9, 351–365.

Challen, G.A., and Goodell, M.A. (2020). Clonal hematopoiesis: mechanisms driving dominance of stem cell clones. *Blood* 136, 1590–1598.

Chapman, M.A., Lawrence, M.S., Keats, J.J., et al. (2011). Initial genome sequencing and analysis of multiple myeloma. *Nature* 471, 467–472.

Charlton, J., Jung, E.J., Mattei, A.L., et al. (2020). TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet.* 52, 819–827.

Chatonnet, F., Pignarre, A., Sérandour, A.A., et al. (2020). The hydroxymethylome of multiple myeloma identifies FAM72D as a 1q21 marker linked to proliferation. *Haematologica* 105, 774–783.

Chen, K.C., Bevan, P.C., and Matthews, J.G. (1986). Analysis of G banded karyotypes in myeloma cells. *J. Clin. Pathol.* 39, 260–266.

Chen, X., Schulz-Trieglaff, O., Shaw, R., et al. (2016). Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* 32, 1220–1222.

Cimmino, L., Dawlaty, M.M., Ndiaye-Lobry, D., et al. (2015). TET1 is a tumor suppressor of hematopoietic malignancy. *Nat. Immunol.* 16, 653–662.

Cleynen, A., Szalat, R., Kemal Samur, M., et al. (2017). Expressed fusion gene landscape and its impact in multiple myeloma. *Nat. Commun.* 8, 1893.

Cohen, A.D., Garfall, A.L., Stadtmauer, E.A., et al. (2016). B-Cell Maturation Antigen (BCMA)-Specific Chimeric Antigen Receptor T Cells (CART-BCMA) for Multiple Myeloma (MM): Initial Safety and Efficacy from a Phase I Study. *Blood* 128, 1147–1147 .

Cohen, Y.C., Zada, M., Gatt, M.E., et al. (2019). Kydar Multicenter Trial of Quadruple Regimen for Induction Resistant Myeloma Combined with Translational Single-Cell Analysis Identifies Potential Drivers of Advanced Resistance, Including Novel Immune Checkpoints. *Blood* 134, 982–982.

Corre, J., Perrot, A., Caillot, D., et al. (2020). Del17p without TP53 mutation confers poor prognosis in intensively treated newly diagnosed multiple myeloma patients. *Blood* blood.2020008346 (en cours d'édition).

Cui, X.-L., Nie, J., Ku, J., et al. (2020). A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun.* *11*, 6161.

David, A., Zocchi, S., Talbot, A., et al. (2020). The long non-coding RNA CRNDE regulates growth of multiple myeloma cells via an effect on IL6 signalling. *Leukemia* 1–12.

Davids, M.S., Roberts, A.W., Seymour, J.F., et al. (2017). Phase I First-in-Human Study of Venetoclax in Patients With Relapsed or Refractory Non-Hodgkin Lymphoma. *J. Clin. Oncol.* *35*, 826–833.

Davis, C.A., Hitz, B.C., Sloan, C.A., et al. (2018). The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* *46*, D794–D801.

Dawlaty, M.M., Ganz, K., Powell, B.E., et al. (2011). Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development. *Cell Stem Cell* *9*, 166–175.

Decaux, O., Lodé, L., Magrangeas, F., et al. (2008). Prediction of Survival in Multiple Myeloma Based on Gene Expression Profiles Reveals Cell Cycle and Chromosomal Instability Signatures in High-Risk Patients and Hyperdiploid Signatures in Low-Risk Patients: A Study of the Intergroupe Francophone du Myélome. *J. Clin. Oncol.* *26*, 4798–4805.

Defossez, P.-A., and Stancheva, I. (2011). Biological Functions of Methyl-CpG-Binding Proteins. In *Progress in Molecular Biology and Translational Science*, (Elsevier), pp. 377–398.

Delhommeau, F., Dupont, S., Valle, V.D., et al. (2009). Mutation in TET2 in Myeloid Cancers. *N. Engl. J. Med.* *360*, 2289–2301.

Desport, E., Bridoux, F., Sirac, C., et al. (2012). AL Amyloidosis. *Orphanet J. Rare Dis.* *7*, 54.

Dobin, A., Davis, C.A., Schlesinger, F., et al. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* *29*, 15–21.

Domcke, S., Bardet, A.F., Adrian Ginno, P., et al. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* *528*, 575–579.

Durie, B.G.M., Hoering, A., Abidi, M.H., et al. (2017). Bortezomib with lenalidomide and dexamethasone versus lenalidomide and dexamethasone alone in patients with newly diagnosed myeloma without intent for immediate autologous stem-cell transplant (SWOG S0777): a randomised, open-label, phase 3 trial. *The Lancet* *389*, 519–527.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* *489*, 57–74.

Enders, A., Bouillet, P., Puthalakath, H., et al. (2003). Loss of the Pro-Apoptotic BH3-only Bcl-2 Family Member Bim Inhibits BCR Stimulation-induced Apoptosis and Deletion of Autoreactive B Cells. *J. Exp. Med.* *198*, 1119–1126.

Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* *9*, 215–216.

Facon, T., Kumar, S., Plesner, T., et al. (2019). Daratumumab plus Lenalidomide and Dexamethasone for Untreated Myeloma. *N. Engl. J. Med.* *380*, 2104–2115.

Fakharzadeh, S.S., Trusko, S.P., and George, D.L. (1991). Tumorigenic potential associated with enhanced expression of a gene that is amplified in a mouse tumor cell line. *EMBO J.* *10*, 1565–1569.

Fakharzadeh, S.S., Rosenblum-Vos, L., Murphy, M., et al. (1993). Structure and Organization of Amplified DNA on Double Minutes Containing the *mdm2* Oncogene. *Genomics* *15*, 283–290.

Fan, S., and Zhang, X. (2009). CpG island methylation pattern in different human tissues and its correlation with gene expression. *Biochem. Biophys. Res. Commun.* *383*, 421–425.

Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. *Nat. Rev. Cancer* *4*, 143–153.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* *301*, 89–92.

Feng, J., Liu, T., Qin, B., et al. (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* *7*, 1728–1740.

Finak, G., McDavid, A., Yajima, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* *16*.

Fonseca, R., Van Wier, S.A., Chng, W.J., et al. (2006). Prognostic value of chromosome 1q21 gain by fluorescent in situ hybridization and increase CKS1B expression in myeloma. *Leukemia* *20*, 2034–2040.

Freitas, R.F. de, Liu, Y., Szewczyk, M.M., et al. (2020). Discovery of Small-Molecule Antagonists of the PWWP Domain of NSD2. *BioRxiv* 2020.11.25.398586.

Gardiner-Garden, M., and Frommer, M. (1987). CpG Islands in vertebrate genomes. *J. Mol. Biol.* *196*, 261–282.

Gerlinger, M., Rowan, A.J., Horswell, S., et al. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* *366*, 883–892.

Germini, D., Tsfasman, T., Klibi, M., et al. (2017). HIV Tat induces a prolonged MYC relocalization next to IGH in circulating B-cells. *Leukemia* *31*, 2515–2522.

Goel, N., Karir, P., and Garg, V.K. (2017). Role of DNA methylation in human age prediction. *Mech. Ageing Dev.* *166*, 33–41.

Gomez-Bougie, P., Maiga, S., Tessoulin, B., et al. (2018). BH3-mimetic toolkit guides the respective use of BCL2 and MCL1 BH3-mimetics in myeloma treatment. *Blood* *132*, 2656–2669.

González, D., van der Burg, M., García-Sanz, R., et al. (2007). Immunoglobulin gene rearrangements and the pathogenesis of multiple myeloma. *Blood* *110*, 3112–3121.

Gooding, S., Ansari-Pour, N., Towfic, F., et al. (2020). Multiple Cereblon genetic changes associate with acquired resistance to Lenalidomide or Pomalidomide in Multiple Myeloma. *Blood* 2020007081 (en cours d'édition).

Green, D.J., Pont, M., Cowan, A.J., et al. (2019). Response to Bcma CAR-T Cells Correlates with Pretreatment Target Antigen Density and Is Improved By Small Molecule Inhibition of Gamma Secretase. *Blood* 134, 1856–1856.

Gu, T.-P., Guo, F., Yang, H., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610.

Guièze, R., Liu, V.M., Rosebrock, D., et al. (2019). Mitochondrial Reprogramming Underlies Resistance to BCL-2 Inhibition in Lymphoid Malignancies. *Cancer Cell* 36, 369–384.e13.

Guo, G., Raje, N.S., Seifer, C., et al. (2018). Genomic discovery and clonal tracking in multiple myeloma by cell-free DNA sequencing. *Leukemia* 32, 1838–1841.

de Haart, S.J., Willems, S.M., Mutis, T., et al. (2016). Comparison of intramedullary myeloma and corresponding extramedullary soft tissue plasmacytomas using genetic mutational panel analyses. *Blood Cancer J.* 6, e426.

Hahn, P.J. (1993). Molecular biology of double-minute chromosomes. *BioEssays* 15, 477–484.

Hamann, I., Krysz, D., Glubrecht, D., et al. (2018). Expression and function of hexose transporters GLUT1, GLUT2, and GLUT5 in breast cancer—effects of hypoxia. *FASEB J.* 32, 5104–5118.

Hansen, K.D., Timp, W., Bravo, H.C., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat. Genet.* 43, 768–775.

Hargreaves, D.C., Hyman, P.L., Lu, T.T., et al. (2001). A Coordinated Change in Chemokine Responsiveness Guides Plasma Cell Movements. *J. Exp. Med.* 194, 45–56.

Hébraud, B., Caillot, D., Corre, J., et al. (2013). The translocation t(4;14) can be present only in minor subclones in multiple myeloma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* 19, 4634–4637.

Heo, I., Joo, C., Cho, J., et al. (2008). Lin28 Mediates the Terminal Uridylation of let-7 Precursor MicroRNA. *Mol. Cell* 32, 276–284.

Herling, C.D., Abedpour, N., Weiss, J., et al. (2018). Clonal dynamics towards the development of venetoclax resistance in chronic lymphocytic leukemia. *Nat. Commun.* 9, 727.

Heuck, C.J., Mehta, J., Bhagat, T., et al. (2013). Myeloma Is Characterized by Stage-Specific Alterations in DNA Methylation That Occur Early during Myelomagenesis. *J. Immunol.* 190, 2966–2975.

Hird, A.W., Secrist, J.P., Adam, A., et al. (2017). Abstract DDT01-02: AZD5991: A potent and selective macrocyclic inhibitor of Mcl-1 for treatment of hematologic cancers. *Cancer Res.* 77, DDT01-02-DDT01-02.

- Huang, Y., and Rao, A. (2014). Connections between TET proteins and aberrant DNA modification in cancer. *Trends Genet.* *30*, 464–474.
- Huang, H., Jiang, X., Li, Z., et al. (2013). TET1 plays an essential oncogenic role in MLL-rearranged leukemia. *Proc. Natl. Acad. Sci.* *110*, 11994–11999.
- Huang, Y., Pastor, W.A., Shen, Y., et al. (2010). The Behaviour of 5-Hydroxymethylcytosine in Bisulfite Sequencing. *PLoS ONE* *5*, e8888.
- Hultcrantz, M., Yellapantula, V., and Rustad, E.H. (2020). Genomic profiling of multiple myeloma: New insights and modern technologies. *Best Pract. Res. Clin. Haematol.* *33*, 101153.
- Hunter, Z.R., Xu, L., Yang, G., et al. (2014). The genomic landscape of Waldenstrom macroglobulinemia is characterized by highly recurring MYD88 and WHIM-like CXCR4 mutations, and small somatic deletions associated with B-cell lymphomagenesis. *Blood* *123*, 1637–1646.
- INCa, and AF3M (2015). Comprendre le Myélome Multiple (INCa et IF3M).
- Ise, W., and Kurosaki, T. (2019). Plasma cell differentiation during the germinal center reaction. *Immunol. Rev.* *288*, 64–74.
- Iyer, M.K., Niknafs, Y.S., Malik, R., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* *47*, 199–208.
- Izzo, F., Lee, S.C., Poran, A., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. *Nat. Genet.* *52*, 378–387.
- Jaffe, J.D., Wang, Y., Chan, H.M., et al. (2013). Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. *Nat. Genet.* *45*, 1386–1391.
- Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., et al. (2014). Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science* *343*, 776–779.
- Jamet, B., Bailly, C., Carlier, T., et al. (2019a). FDG PET in Multiple Myeloma. In *Molecular Imaging in Multiple Myeloma*, C. Nanni, S. Fanti, and L. Zanoni, eds. (Cham: Springer International Publishing), pp. 27–38.
- Jamet, B., Bailly, C., Carlier, T., et al. (2019b). Interest of Pet Imaging in Multiple Myeloma. *Front. Med.* *6*.
- Jeschke, J., Collignon, E., and Fuks, F. (2016). Portraits of TET-mediated DNA hydroxymethylation in cancer. *Curr. Opin. Genet. Dev.* *36*, 16–26.
- Jin, S.-G., Jiang, Y., Qiu, R., et al. (2011). 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res.* *71*, 7360–7365.
- Jin, Y., Chen, K., De Paepe, A., et al. (2018). Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood* *131*, 2138–2150.
- Johnson, S.M., Grosshans, H., Shingara, J., et al. (2005). RAS Is Regulated by the let-7 MicroRNA Family. *Cell* *120*, 635–647.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* *13*, 484–492.

Jonsson, S., Sveinbjornsson, G., de Lapuente Portilla, A.L., et al. (2017). Identification of sequence variants influencing immunoglobulin levels. *Nat. Genet.* *49*, 1182–1191.

Kang, S., Li, Q., Chen, Q., et al. (2017). CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. *Genome Biol.* *18*, 1–12.

Karimzadeh, M., Ernst, C., Kundaje, A., et al. (2018). Umap and Bismap: quantifying genome and methylome mappability. *Nucleic Acids Res.* *46*, e120–e120.

Keats, J.J., Reiman, T., Maxwell, C.A., et al. (2003). In multiple myeloma, t(4;14)(p16;q32) is an adverse prognostic factor irrespective of FGFR3 expression. *Blood* *101*, 1520–1529.

Khodadadi, L., Cheng, Q., Radbruch, A., et al. (2019). The Maintenance of Memory Plasma Cells. *Front. Immunol.* *10*.

Khund-Sayeed, S., He, X., Holzberg, T., et al. (2016). 5-Hydroxymethylcytosine in E-box motifs ACAT|GTG and ACAC|GTG increases DNA-binding of the B-HLH transcription factor TCF4. *Integr. Biol.* *8*, 936–945.

Ko, M., An, J., Bandukwala, H.S., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* *497*, 122–126.

Ko, M., An, J., Pastor, W.A., et al. (2015). TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol. Rev.* *263*, 6–21.

Korbel, J.O., and Campbell, P.J. (2013). Criteria for Inference of Chromothripsis in Cancer Genomes. *Cell* *152*, 1226–1236.

Kothari, R.M., and Shankar, V. (1976). 5-methylcytosine content in the vertebrate deoxyribonucleic acids: Species specificity. *J. Mol. Evol.* *7*, 325–329.

Kotschy, A., Szlavik, Z., Murray, J., et al. (2016). The MCL1 inhibitor S63845 is tolerable and effective in diverse cancer models. *Nature* *538*, 477–482.

Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* *324*, 929–930.

Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* *11*, 597–610.

Kuiper, R., Broyl, A., de Knecht, Y., et al. (2012). A gene expression signature for high-risk multiple myeloma. *Leukemia* *26*, 2406–2413.

Kuiper, R., van Duin, M., van Vliet, M.H., et al. (2015). Prediction of high- and low-risk multiple myeloma based on gene expression and the International Staging System. *Blood* *126*, 1996–2004.

Kumar, S.K., Rajkumar, S.V., Dispenzieri, A., et al. (2008). Improved survival in multiple myeloma and the impact of novel therapies. *Blood* *111*, 2516–2520.

Kumar, S.K., Dispenzieri, A., Lacy, M.Q., et al. (2014). Continued improvement in survival in multiple myeloma: changes in early mortality and outcomes in older patients. *Leukemia* 28, 1122–1128.

Kumar, S.K., Paiva, B., Anderson, K.C., et al. (2016). International Myeloma Working Group consensus criteria for response and minimal residual disease assessment in multiple myeloma. *Lancet Oncol.* 17, e328–e346.

Kumar, S.K., Rajkumar, V., Kyle, R.A., et al. (2017a). Multiple myeloma. *Nat. Rev. Dis. Primer* 3, 17046.

Kumar, S.K., Kaufman, J.L., Gasparetto, C., et al. (2017b). Efficacy of venetoclax as targeted therapy for relapsed/refractory t(11;14) multiple myeloma. *Blood* 130, 2401–2409.

Kumar, S.K., Harrison, S.J., Cavo, M., et al. (2020). Venetoclax or placebo in combination with bortezomib and dexamethasone in patients with relapsed or refractory multiple myeloma (BELLINI): a randomised, double-blind, multicentre, phase 3 trial. *Lancet Oncol.* 0.

Kuo, A.J., Cheung, P., Chen, K., et al. (2011). NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. *Mol. Cell* 44, 609–620.

Küppers, R. (2005). Mechanisms of B-cell lymphoma pathogenesis. *Nat. Rev. Cancer* 5, 251–262.

Kurosaki, T., Kometani, K., and Ise, W. (2015). Memory B cells. *Nat. Rev. Immunol.* 15, 149–159.

Kwa, M., Makris, A., and Esteva, F.J. (2017). Clinical utility of gene-expression signatures in early stage breast cancer. *Nat. Rev. Clin. Oncol.* 14, 595–610.

Kyle, R.A., Therneau, T.M., Rajkumar, S.V., et al. (2006). Prevalence of Monoclonal Gammopathy of Undetermined Significance. *N. Engl. J. Med.* 354, 1362–1369.

Kyle, R.A., Larson, D.R., Therneau, T.M., et al. (2018). Long-Term Follow-up of Monoclonal Gammopathy of Undetermined Significance. *N. Engl. J. Med.* 378, 241–249.

Lachin, J.M. (1981). Introduction to sample size determination and power analysis for clinical trials. *Control. Clin. Trials* 2, 93–113.

Lakshman, A., Rajkumar, S.V., Buadi, F.K., et al. (2018). Risk stratification of smoldering multiple myeloma incorporating revised IMWG diagnostic criteria. *Blood Cancer J.* 8, 1–10.

Landau, H.J., Yellapantula, V., Diamond, B.T., et al. (2020). Accelerated single cell seeding in relapsed multiple myeloma. *Nat. Commun.* 11.

Landgren, C.O., Chari, A., Cohen, Y.C., et al. (2020). Daratumumab monotherapy for patients with intermediate-risk or high-risk smoldering multiple myeloma: a randomized, open-label, multicenter, phase 2 study (CENTAURUS). *Leukemia* 34, 1840–1852.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

- Leblay, N., Maity, R., Barakat, E., et al. (2020). Cite-Seq Profiling of T Cells in Multiple Myeloma Patients Undergoing BCMA Targeting CAR-T or Bites Immunotherapy. *Blood* *136*, 11–12 (Supplements ASH conference).
- Ledergor, G., Weiner, A., Zada, M., et al. (2018). Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat. Med.* *24*, 1867–1876.
- Lee, L., Draper, B., Chaplin, N., et al. (2018). An APRIL-based chimeric antigen receptor for dual targeting of BCMA and TACI in multiple myeloma. *Blood* *131*, 746–758.
- Ley, T.J., Ding, L., Walter, M.J., et al. (2010). DNMT3A Mutations in Acute Myeloid Leukemia. *N. Engl. J. Med.* *363*, 2424–2433.
- Li, W., Zhang, X., Lu, X., et al. (2017). 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res.* *27*, 1243–1257.
- Li, Y., Trojer, P., Xu, C.-F., et al. (2009). The target of the NSD family of histone lysine methyltransferases depends on the nature of the substrate. *J. Biol. Chem.* *284*, 34283–34295.
- Li, Z., Cai, X., Cai, C.-L., et al. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* *118*, 4509–4518.
- Lian, C.G., Xu, Y., Ceol, C., et al. (2012). Loss of 5-Hydroxymethylcytosine Is an Epigenetic Hallmark of Melanoma. *Cell* *150*, 1135–1146.
- Liao, J., Karnik, R., Gu, H., et al. (2015). Targeted disruption of DNMT1 , DNMT3A and DNMT3B in human embryonic stem cells. *Nat. Genet.* *47*, 469–478.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* *30*, 923–930.
- Lodé, L., Eveillard, M., Trichet, V., et al. (2010). Mutations in TP53 are exclusively associated with del(17p) in multiple myeloma. *Haematologica* *95*, 1973–1976.
- Lohard, S., Juin, P.P., and Barillé-Nion, S. (2020). Mitotic stress-induced secretome primes cancer cells to apoptosis and maximizes paclitaxel response in breast tumors when combined with BCL-xL-targeting BH3 mimetics. *Mol. Cell. Oncol.* *7*, 1735912.
- Lohr, J.G., Stojanov, P., Carter, S.L., et al. (2014). Widespread Genetic Heterogeneity in Multiple Myeloma: Implications for Targeted Therapy. *Cancer Cell* *25*, 91–101.
- López-Moyado, I.F., and Rao, A. (2020). DNMT3A and TET2 mutations reshape hematopoiesis in opposing ways. *Nat. Genet.* *52*, 554–556.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* *15*, 550.
- Lovén, J., Hoke, H.A., Lin, C.Y., et al. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. *Cell* *153*, 320–334.
- Ludwig, H., Delforge, M., Facon, T., et al. (2018). Prevention and management of adverse events of novel agents in multiple myeloma: a consensus of the European Myeloma Network. *Leukemia* *32*, 1542–1560.

Macosko, E.Z., Basu, A., Satija, R., et al. (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161, 1202–1214.

Magrangeas, F., Cormier, M.-L., Descamps, G., et al. (2004). Light-chain only multiple myeloma is due to the absence of functional (productive) rearrangement of the IgH gene at the DNA level. *Blood* 103, 3869–3875.

Magrangeas, F., Avet-Loiseau, H., Munshi, N.C., et al. (2011). Chromothripsis identifies a rare and aggressive entity among newly diagnosed multiple myeloma patients. *Blood* 118, 675–678.

Magrangeas, F., Kuiper, R., Avet-Loiseau, H., et al. (2016). A genome-wide association study identifies a novel locus for bortezomib-induced peripheral neuropathy in European patients with multiple myeloma. *Clin. Cancer Res.* 22, 4350–4355.

Mahé, E.A., Madigou, T., Sérandour, A.A., et al. (2017). Cytosine modifications modulate the chromatin architecture of transcriptional enhancers. *Genome Res.* 27, 947–958.

Manier, S., Salem, K.Z., Park, J., et al. (2017a). Genomic complexity of multiple myeloma and its clinical implications. *Nat. Rev. Clin. Oncol.* 14, 100–113.

Manier, S., Powers, J.T., Sacco, A., et al. (2017b). The LIN28B/let-7 axis is a novel therapeutic pathway in multiple myeloma. *Leukemia* 31, 853–860.

Manier, S., Liu, C.-J., Avet-Loiseau, H., et al. (2017c). Prognostic role of circulating exosomal miRNAs in multiple myeloma. *Blood* 129, 2429–2436.

Manier, S., Park, J., Capelletti, M., et al. (2018). Whole-exome sequencing of cell-free DNA and circulating tumor cells in multiple myeloma. *Nat. Commun.* 9, 1691.

Marango, J., Shimoyama, M., Nishio, H., et al. (2008). The MMSET protein is a histone methyltransferase with characteristics of a transcriptional corepressor. *Blood* 111, 3145–3154.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* 17, 10.

Martincic, K., Alkan, S.A., Cheadle, A., et al. (2009). Transcription elongation factor ELL2 directs immunoglobulin secretion in plasma cells by stimulating altered RNA processing. *Nat. Immunol.* 10, 1102–1109.

Martincorena, I., Raine, K.M., Gerstung, M., et al. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. *Cell* 171, 1029–1041.e21.

Mateos, M.-V., Hernández, M.-T., Giraldo, P., et al. (2013). Lenalidomide plus Dexamethasone for High-Risk Smoldering Multiple Myeloma. *N. Engl. J. Med.* 369, 438–447.

Matulis, S.M., Gupta, V.A., Neri, P., et al. (2019). Functional profiling of venetoclax sensitivity can predict clinical response in multiple myeloma. *Leukemia* 33, 1291–1296.

Maura, F., Degasperi, A., Nadeu, F., et al. (2019a). A practical guide for mutational signature analysis in hematological malignancies. *Nat. Commun.* 10, 2969.

Maura, F., Bolli, N., Angelopoulos, N., et al. (2019b). Genomic landscape and chronological reconstruction of driver events in multiple myeloma. *Nat. Commun.* 10, 3835.

McCarron, M.J., Park, P.W., and Fooksman, D.R. (2017). CD138 mediates selection of mature plasma cells by regulating their survival. *Blood* 129, 2749–2759.

Medina, A., Puig, N., Flores-Montero, J., et al. (2020). Comparison of next-generation sequencing (NGS) and next-generation flow (NGF) for minimal residual disease (MRD) assessment in multiple myeloma. *Blood Cancer J.* 10, 1–10.

Melchor, L., Brioli, A., Wardell, C.P., et al. (2014). Single-cell genetic analysis reveals the composition of initiating clones and phylogenetic patterns of branching and parallel evolution in myeloma. *Leukemia* 28, 1705–1715.

Michaud-Robert, A.-V., Jamet, B., Bailly, C., et al. (2020). FDG-PET/CT, a Promising Exam for Detecting High-Risk Myeloma Patients? *Cancers* 12, 1384.

Mikulasova, A., Ashby, C., Tytarenko, R.G., et al. (2020). Microhomology-mediated end joining drives complex rearrangements and overexpression of MYC and PVT1 in multiple myeloma. *Haematologica* 105, 1055–1066.

Morales, A.A., Kurtoglu, M., Matulis, S.M., et al. (2011). Distribution of Bim determines Mcl-1 dependence or codependence with Bcl-xL/Bcl-2 in Mcl-1-expressing myeloma cells. *Blood* 118, 1329–1339.

Moreau, P., and Zamagni, E. (2017). MRD in multiple myeloma: more questions than answers? *Blood Cancer J.* 7, 1–4.

Moreau, P., Attal, M., Caillot, D., et al. (2017a). Prospective Evaluation of Magnetic Resonance Imaging and [18F]Fluorodeoxyglucose Positron Emission Tomography-Computed Tomography at Diagnosis and Before Maintenance Therapy in Symptomatic Patients With Multiple Myeloma Included in the IFM/DFCI 2009 Trial: Results of the IMAJEM Study. *J. Clin. Oncol.* 35, 2911–2918.

Moreau, P., Chanan-Khan, A., Roberts, A.W., et al. (2017b). Promising efficacy and acceptable safety of venetoclax plus bortezomib and dexamethasone in relapsed/refractory MM. *Blood* 130, 2392–2400.

Moreau, P., Attal, M., Hulin, C., et al. (2019a). Bortezomib, thalidomide, and dexamethasone 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, 29–38.

Moreau, P., Zweegman, S., Perrot, A., et al. (2019b). Evaluation of the Prognostic Value of Positron Emission Tomography-Computed Tomography (PET-CT) at Diagnosis and Follow-up in Transplant-Eligible Newly Diagnosed Multiple Myeloma (TE NDMM) Patients Treated in the Phase 3 Cassiopeia Study: Results of the Cassiopet Companion Study. *Blood* 134, 692–692.

Mroczek, S., Chlebowska, J., Kuliński, T.M., et al. (2017). The non-canonical poly(A) polymerase FAM46C acts as an onco-suppressor in multiple myeloma. *Nat. Commun.* 8, 619.

Muramatsu, M., Kinoshita, K., Fagarasan, S., et al. (2000). Class Switch Recombination and Hypermutation Require Activation-Induced Cytidine Deaminase (AID), a Potential RNA Editing Enzyme. *Cell* 102, 553–563.

Neri, P., and Bahlis, N.J. (2013). Genomic instability in multiple myeloma: mechanisms and therapeutic implications. *Expert Opin. Biol. Ther.* 13, S69–S82.

Nestor, C.E., Ottaviano, R., Reddington, J., et al. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* 22, 467–477.

Ng, M.H.L., Chung, Y.F., Lo, K.W., et al. (1997). Frequent Hypermethylation of p16 and p15 Genes in Multiple Myeloma. *Blood* 89, 2500–2506.

Nomura, N., Verdon, G., Kang, H.J., et al. (2015). Structure and mechanism of the mammalian fructose transporter GLUT5. *Nature* 526, 397–401.

Nowell, P. (1976). The clonal evolution of tumor cell populations. *Science* 194, 23–28.

Nutt, S.L., Hodgkin, P.D., Tarlinton, D.M., et al. (2015). The generation of antibody-secreting plasma cells. *Nat. Rev. Immunol.* 15, 160–171.

Ochiai, K., Maienschein-Cline, M., Simonetti, G., et al. (2013). Transcriptional Regulation of Germinal Center B and Plasma Cell Fates by Dynamical Control of IRF4. *Immunity* 38, 918–929.

Okano, M., Bell, D.W., Haber, D.A., et al. (1999). DNA Methyltransferases Dnmt3a and Dnmt3b Are Essential for De Novo Methylation and Mammalian Development. *Cell* 99, 247–257.

Onodera, N., McCabe, N.R., and Rubin, C.M. (1992). Formation of a hyperdiploid karyotype in childhood acute lymphoblastic leukemia. *Blood* 80, 203–208.

Ordoñez, R., Kulis, M., Russiñol, N., et al. (2020). Chromatin activation as a unifying principle underlying pathogenic mechanisms in multiple myeloma. *Genome Res.* 30, 1217–1227.

O’Riordan, M., and Grosschedl, R. (1999). Coordinate Regulation of B Cell Differentiation by the Transcription Factors EBF and E2A. *Immunity* 11, 21–31.

Paiva, B., Puig, N., Cedena, M.-T., et al. (2019). Measurable Residual Disease by Next-Generation Flow Cytometry in Multiple Myeloma. *J. Clin. Oncol.* 38, 784–792.

Palumbo, A., Avet-Loiseau, H., Oliva, S., et al. (2015). Revised International Staging System for Multiple Myeloma: A Report From International Myeloma Working Group. *J. Clin. Oncol.* 33, 2863–2869.

Papatzimas, J.W., Gorobets, E., Maity, R., et al. (2019). From Inhibition to Degradation: Targeting the Antiapoptotic Protein Myeloid Cell Leukemia 1 (MCL1). *J. Med. Chem.* 62, 5522–5540.

Penn, N.W., Suwalski, R., O’Riley, C., et al. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* 126, 781–790.

Peperzak, V., Vikström, I., Walker, J., et al. (2013). Mcl-1 is essential for the survival of plasma cells. *Nat. Immunol.* *14*, 290–297.

Pertesi, M., Went, M., Hansson, M., et al. (2020). Genetic predisposition for multiple myeloma. *Leukemia* *34*, 697–708.

Peterson, V.M., Zhang, K.X., Kumar, N., et al. (2017). Multiplexed quantification of proteins and transcripts in single cells. *Nat. Biotechnol.* *35*, 936–939.

Pierro, J., Saliba, J., Narang, S., et al. (2020). The NSD2 p.E1099K Mutation Is Enriched at Relapse and Confers Drug Resistance in a Cell Context–Dependent Manner in Pediatric Acute Lymphoblastic Leukemia. *Mol. Cancer Res.* *18*, 1153–1165.

Pliner, H.A., Packer, J.S., McFaline-Figueroa, J.L., et al. (2018). Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. *Mol. Cell* *71*, 858–871.e8.

Pont, M.J., Hill, T., Cole, G.O., et al. (2019). γ -Secretase inhibition increases efficacy of BCMA-specific chimeric antigen receptor T cells in multiple myeloma. *Blood J. Am. Soc. Hematol.* *134*, 1585–1597.

Popovic, R., Martinez-Garcia, E., Giannopoulou, E.G., et al. (2014). 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*, e1004566.

Punnoose, E.A., Levenson, J.D., Peale, F., et al. (2016). Expression Profile of BCL-2, BCL-XL, and MCL-1 Predicts Pharmacological Response to the BCL-2 Selective Antagonist Venetoclax in Multiple Myeloma Models. *Mol. Cancer Ther.* *15*, 1132–1144.

Rajkumar, S.V., Dimopoulos, M.A., Palumbo, A., et al. (2014). International Myeloma Working Group updated criteria for the diagnosis of multiple myeloma. *Lancet Oncol.* *15*, e538–e548.

Rajkumar, S.V., Landgren, O., and Mateos, M.-V. (2015). Smoldering multiple myeloma. *Blood* *125*, 3069–3075.

Rajkumar, S.V., Voorhees, P.M., Goldschmidt, H., et al. (2018). Randomized, open-label, phase 3 study of subcutaneous daratumumab (DARA SC) versus active monitoring in patients (Pts) with high-risk smoldering multiple myeloma (SMM): AQUILA. *J. Clin. Oncol.* *36*, TPS8062–TPS8062.

Rasche, L., Angtuaco, E., McDonald, J.E., et al. (2017a). Low expression of hexokinase-2 is associated with false-negative FDG–positron emission tomography in multiple myeloma. *Blood* *130*, 30–34.

Rasche, L., Chavan, S.S., Stephens, O.W., et al. (2017b). Spatial genomic heterogeneity in multiple myeloma revealed by multi-region sequencing. *Nat. Commun.* *8*, 268.

Rasche, L., Kortüm, K.M., Raab, M.S., et al. (2019). The Impact of Tumor Heterogeneity on Diagnostics and Novel Therapeutic Strategies in Multiple Myeloma. *Int. J. Mol. Sci.* *20*, 1248.

Rasmussen, K.D., Jia, G., Johansen, J.V., et al. (2015). Loss of *TET2* in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis. *Genes Dev.* 29, 910–922.

Rausch, T., Zichner, T., Schlattl, A., et al. (2012). DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339.

Roberts, A.W., Davids, M.S., Pagel, J.M., et al. (2016). Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia. *N. Engl. J. Med.* 374, 311–322.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., et al. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.

Roccaro, A.M., Mishima, Y., Sacco, A., et al. (2015). CXCR4 Regulates Extra-Medullary Myeloma through Epithelial-Mesenchymal-Transition-like Transcriptional Activation. *Cell Rep.* 12, 622–635.

Roco, J.A., Mesin, L., Binder, S.C., et al. (2019). Class-Switch Recombination Occurs Infrequently in Germinal Centers. *Immunity* 51, 337–350.e7.

Romagnoli, M., Trichet, V., David, C., et al. (2007). Significant impact of survivin on myeloma cell growth. *Leukemia* 21, 1070–1078.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, 389–393.

Rozhok, A.I., and DeGregori, J. (2015). Toward an evolutionary model of cancer: Considering the mechanisms that govern the fate of somatic mutations. *Proc. Natl. Acad. Sci.* 112, 8914–8921.

Rustad, E.H., Yellapantula, V., Leongamornlert, D., et al. (2020a). Timing the initiation of multiple myeloma. *Nat. Commun.* 11, 1917.

Rustad, E.H., Yellapantula, V.D., Glodzik, D., et al. (2020b). Revealing the impact of structural variants in multiple myeloma. *Blood Cancer Discov.*

Saint-André, V., Federation, A.J., Lin, C.Y., et al. (2016). Models of human core transcriptional regulatory circuitries. *Genome Res.* 26, 385–396.

Sakai, T., Toguchida, J., Ohtani, N., et al. (1991). Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am. J. Hum. Genet.* 48, 880–888.

Salhia, B., Baker, A., Ahmann, G., et al. (2010). DNA Methylation Analysis Determines the High Frequency of Genic Hypomethylation and Low Frequency of Hypermethylation Events in Plasma Cell Tumors. *Cancer Res.* 70, 6934–6944.

Sampson, V.B., Rong, N.H., Han, J., et al. (2007). MicroRNA Let-7a Down-regulates MYC and Reverts MYC-Induced Growth in Burkitt Lymphoma Cells. *Cancer Res.* 10.

Samur, M.K., Minvielle, S., Gulla, A., et al. (2018). Long intergenic non-coding RNAs have an independent impact on survival in multiple myeloma. *Leukemia* 32, 2626–2635.

Samur, M.K., Fulciniti, M., Aktas-Samur, A., et al. (2020). Biallelic Loss of BCMA Triggers Resistance to Anti-BCMA CAR T Cell Therapy in Multiple Myeloma. *Blood* 136, 14–14 (Supplements ASH conference).

Sawyer, J.R., Tricot, G., Mattox, S., et al. (1998). Jumping Translocations of Chromosome 1q in Multiple Myeloma: Evidence for a Mechanism Involving Decondensation of Pericentromeric Heterochromatin. *Blood* 91, 1732–1741.

Sawyer, J.R., Tian, E., Walker, B.A., et al. (2019). An acquired high-risk chromosome instability phenotype in multiple myeloma: Jumping 1q Syndrome. *Blood Cancer J.* 9, 1–12.

Schübeler, D. (2015). Function and information content of DNA methylation. *Nature* 517, 321–326.

Segalla, S., Pivetti, S., Todoerti, K., et al. (2015). The ribonuclease DIS3 promotes let-7 miRNA maturation by degrading the pluripotency factor LIN28B mRNA. *Nucleic Acids Res.* 43, 5182–5193.

Seiller, C., Maiga, S., Touzeau, C., et al. (2020). Dual targeting of BCL2 and MCL1 rescues myeloma cells resistant to BCL2 and MCL1 inhibitors associated with the formation of BAX/BAK hetero-complexes. *Cell Death Dis.* 11, 1–14.

Sen, G.L., Reuter, J.A., Webster, D.E., et al. (2010). DNMT1 maintains progenitor function in self-renewing somatic tissue. *Nature* 463, 563–567.

Sérandour, A.A., Avner, S., Oger, F., et al. (2012). Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. *Nucleic Acids Res.* 40, 8255–8265.

Shaughnessy, J.D., Zhan, F., Burington, B.E., et al. (2007). A validated gene expression model of high-risk multiple myeloma is defined by deregulated expression of genes mapping to chromosome 1. *Blood* 109, 2276–2284.

Shen, L., Wu, H., Diep, D., et al. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* 153, 692–706.

Silvestris, F., Cafforio, P., Tucci, M., et al. (2002). Negative regulation of erythroblast maturation by Fas-L+/TRAIL+ highly malignant plasma cells: a major pathogenetic mechanism of anemia in multiple myeloma. *Blood* 99, 1305–1313.

Siminovitch, K.A., Bakhshi, A., Goldman, P., et al. (1985). A uniform deleting element mediates the loss of κ genes in human B cells. *Nature* 316, 260–262.

Simon, S., and Riddell, S.R. (2020). Dual Targeting with CAR T Cells to Limit Antigen Escape in Multiple Myeloma. *Blood Cancer Discov.* 1, 130–133.

Singh, R., Letai, A., and Sarosiek, K. (2019). Regulation of apoptosis in health and disease: the balancing act of BCL-2 family proteins. *Nat. Rev. Mol. Cell Biol.* 20, 175–193.

Slomp, A., Moesbergen, L.M., Gong, J., et al. (2019). Multiple myeloma with 1q21 amplification is highly sensitive to MCL-1 targeting. *Blood Adv.* 3, 4202–4214.

Smith, E.L., Harrington, K., Staehr, M., et al. (2019). GPRC5D is a target for the immunotherapy of multiple myeloma with rationally designed CAR T cells. *Sci. Transl. Med.* *11*.

Song, C.-X., Szulwach, K.E., Fu, Y., et al. (2011a). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* *29*, 68–72.

Song, C.-X., Yin, S., Ma, L., et al. (2017). 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res.* *27*, 1231–1242.

Song, J., Rechko, O., Bestor, T.H., et al. (2011b). Structure of DNMT1-DNA Complex Reveals a Role for Autoinhibition in Maintenance DNA Methylation. *Science* *331*, 1036–1040.

Sonneveld, P., Broijl, A., Gay, F., et al. (2019). Bortezomib, lenalidomide, and dexamethasone (VRd) ± daratumumab (DARA) in patients (pts) with transplant-eligible (TE) newly diagnosed multiple myeloma (NDMM): A multicenter, randomized, phase III study (PERSEUS). *J. Clin. Oncol.* *37*, TPS8055–TPS8055.

Souers, A.J., Levenson, J.D., Boghaert, E.R., et al. (2013). ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* *19*, 202–208.

Sparano, J.A., Gray, R.J., Makower, D.F., et al. (2018). Adjuvant Chemotherapy Guided by a 21-Gene Expression Assay in Breast Cancer. *N. Engl. J. Med.* *379*, 111–121.

Spruijt, C.G., Gnerlich, F., Smits, A.H., et al. (2013). Dynamic Readers for 5-(Hydroxy)methylcytosine and Its Oxidized Derivatives. *Cell* *152*, 1146–1159.

Stephens, P.J., Greenman, C.D., Fu, B., et al. (2011). Massive Genomic Rearrangement Acquired in a Single Catastrophic Event during Cancer Development. *Cell* *144*, 27–40.

Stoeckius, M., Hafemeister, C., Stephenson, W., et al. (2017). Simultaneous epitope and transcriptome measurement in single cells. *Nat. Methods* *14*, 865–868.

Stroud, H., Feng, S., Morey Kinney, S., et al. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* *12*, R54.

Stuart, T., Butler, A., Hoffman, P., et al. (2019). Comprehensive Integration of Single-Cell Data. *Cell* *177*, 1888–1902.e21.

Subramanian, A., Tamayo, P., Mootha, V.K., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* *102*, 15545–15550.

Swaroop, A., Oyer, J.A., Will, C.M., et al. (2019). An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene* *38*, 671–686.

Tahiliani, M., Koh, K.P., Shen, Y., et al. (2009). Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* *324*, 930–935.

Tang, F., Barbacioru, C., Wang, Y., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* *6*, 377–382.

Taylor, J.J., Pape, K.A., Steach, H.R., et al. (2015). Apoptosis and antigen affinity limit effector cell differentiation of a single naïve B cell. *Science* 347, 784–787.

Terranova, C., Tang, M., Orouji, E., et al. (2018). An integrated platform for genome-wide mapping of chromatin states Using High-throughput ChIP-sequencing in Tumor Tissues. *J. Vis. Exp.*

The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium (2020). Pan-cancer analysis of whole genomes. *Nature* 578, 82–93.

The International Myeloma Working Group* (2003). Criteria for the classification of monoclonal gammopathies, multiple myeloma and related disorders: a report of the International Myeloma Working Group. *Br. J. Haematol.* 121, 749–757.

Thomas, R.L., Roberts, D.J., Kubli, D.A., et al. (2013). Loss of MCL-1 leads to impaired autophagy and rapid development of heart failure. *Genes Dev.* 27, 1365–1377.

Tian, E., Zhan, F., Walker, R., et al. (2003). The Role of the Wnt-Signaling Antagonist DKK1 in the Development of Osteolytic Lesions in Multiple Myeloma. *N. Engl. J. Med.* 349, 2483–2494.

Touzeau, C., Dousset, C., Le Gouill, S., et al. (2014). The Bcl-2 specific BH3 mimetic ABT-199: a promising targeted therapy for t(11;14) multiple myeloma. *Leukemia* 28, 210–212.

Touzeau, C., Le Gouill, S., Mahé, B., et al. (2017). Deep and sustained response after venetoclax therapy in a patient with very advanced refractory myeloma with translocation t(11;14). *Haematologica* 102, e112–e114.

Touzeau, C., Maciag, P., Amiot, M., et al. (2018). Targeting Bcl-2 for the treatment of multiple myeloma. *Leukemia* 32, 1899–1907.

Traube, F.R., Schiffers, S., Iwan, K., et al. (2019). Isotope-dilution mass spectrometry for exact quantification of noncanonical DNA nucleosides. *Nat. Protoc.* 14, 283–312.

Tseng, Y.-Y., Moriarity, B.S., Gong, W., et al. (2014). PVT1 dependence in cancer with MYC copy-number increase. *Nature* 512, 82–86.

Usmani, S.Z., Heuck, C., Mitchell, A., et al. (2012). Extramedullary disease portends poor prognosis in multiple myeloma and is over-represented in high-risk disease even in the era of novel agents. *Haematologica* 97, 1761–1767.

Vandenberg, C.J., and Cory, S. (2013). ABT-199, a new Bcl-2-specific BH3 mimetic, has in vivo efficacy against aggressive Myc-driven mouse lymphomas without provoking thrombocytopenia. *Blood* 121, 2285–2288.

Walker, B.A., Wardell, C.P., Chiecchio, L., et al. (2011). Aberrant global methylation patterns affect the molecular pathogenesis and prognosis of multiple myeloma. *Blood* 117, 553–562.

Walker, B.A., Wardell, C.P., Johnson, D.C., et al. (2013). Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells. *Blood* 121, 3413–3419.

Walker, B.A., Boyle, E.M., Wardell, C.P., et al. (2015a). Mutational Spectrum, Copy Number Changes, and Outcome: Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma. *J. Clin. Oncol.* *33*, 3911–3920.

Walker, B.A., Wardell, C.P., Murison, A., et al. (2015b). APOBEC family mutational signatures are associated with poor prognosis translocations in multiple myeloma. *Nat. Commun.* *6*, 1–11.

Walker, B.A., Mavrommatis, K., Wardell, C.P., et al. (2018). Identification of novel mutational drivers reveals oncogene dependencies in multiple myeloma. *Blood* *132*, 587–597.

Walker, B.A., Mavrommatis, K., Wardell, C.P., et al. (2019). A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis. *Leukemia* *33*, 159–170.

Wang, L., Leite de Oliveira, R., Huijberts, S., et al. (2018). An Acquired Vulnerability of Drug-Resistant Melanoma with Therapeutic Potential. *Cell* *173*, 1413–1425.e14.

Weinberg, D.N., Papillon-Cavanagh, S., Chen, H., et al. (2019). The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* *573*, 281–286.

Weinhold, N., Johnson, D.C., Chubb, D., et al. (2013). The CCND1 c. 870G> A polymorphism is a risk factor for t (11; 14)(q13; q32) multiple myeloma. *Nat. Genet.* *45*, 522–525.

Whyte, W.A., Orlando, D.A., Hnisz, D., et al. (2013). Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* *153*, 307–319.

Yamamoto, K., Yakushijin, K., Ito, M., et al. (2020). MYC amplification on double minute chromosomes in plasma cell leukemia with double IGH/CCND1 fusion genes. *Cancer Genet.* *242*, 35–40.

Yin, Y., Morgunova, E., Jolma, A., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* *356*.

Youle, R.J., and Strasser, A. (2008). The BCL-2 protein family: opposing activities that mediate cell death. *Nat. Rev. Mol. Cell Biol.* *9*, 47–59.

Yu, M., Hon, G.C., Szulwach, K.E., et al. (2012). Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* *7*, 2159–2170.

Yu, X., Li, W., Deng, Q., et al. (2020). MYD88 L265P Elicits Mutation-specific Ubiquitination to Drive NF-κB Activation and Lymphomagenesis. *Blood* *blood.2020004918* (en cours d'édition).

Zavidij, O., Haradhvala, N.J., Mouhieddine, T.H., et al. (2020). Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma. *Nat. Cancer* *1*, 1–14.

Zhan, F., Huang, Y., Colla, S., et al. (2006). The molecular classification of multiple myeloma. *Blood* *108*, 2020–2028.

Zhang, Y., Liu, T., Meyer, C.A., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol.* 9, 1–9.

Zhao, B., Sedlak, J.C., Srinivas, R., et al. (2016). Exploiting Temporal Collateral Sensitivity in Tumor Clonal Evolution. *Cell* 165, 234–246.

Zheng, G.X.Y., Terry, J.M., Belgrader, P., et al. (2017). Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* 8, 14049.

Zhou, T., Xiong, J., Wang, M., et al. (2014). Structural Basis for Hydroxymethylcytosine Recognition by the SRA Domain of UHRF2. *Mol. Cell* 54, 879–886.