



HAL
open science

Le Myélome Multiple et son environnement immunitaire à l'échelle de la cellule unique

Jonathan Cruard

► **To cite this version:**

Jonathan Cruard. Le Myélome Multiple et son environnement immunitaire à l'échelle de la cellule unique. Médecine humaine et pathologie. Nantes Université, 2023. Français. NNT : 2023NANU1033 . tel-04577631

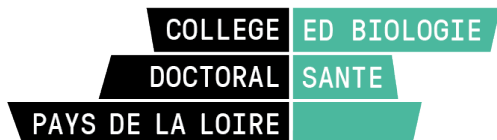
HAL Id: tel-04577631

<https://theses.hal.science/tel-04577631>

Submitted on 16 May 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT

NANTES UNIVERSITE

ÉCOLE DOCTORALE N° 605

Biologie, Santé

Spécialité : *Génétique, Génomique, Bioinformatique*

Par

Jonathan CRUARD

Le Myélome Multiple et son environnement immunitaire à l'échelle de la cellule unique

Thèse présentée et soutenue à Nantes, le 6 décembre 2023

Unité de recherche : CRCI²NA - Centre de Recherche en Cancérologie et Immunologie Intégrée Nantes Angers

Rapporteurs avant soutenance :

Mary CALLANAN Professeur des universités - Praticien hospitalier

Université de Bourgogne - INSERM U1231, Hôpital universitaire de Dijon

David GARRICK Chargé de recherche - INSERM U976, Institut de recherche Saint Louis, Université Paris Cité

Composition du Jury :

Présidente : Mary CALLANAN

Examineurs : Mary CALLANAN

David GARRICK

Dir. de thèse : Stéphane MINVIELLE

Co-dir. de thèse : Philippe MOREAU

Professeur des universités - Praticien hospitalier,
Université de Bourgogne - INSERM U1231, Hôpital universitaire de Dijon
Chargé de recherche - INSERM U976, Institut de recherche Saint Louis,
Université Paris Cité

Directeur de recherche CNRS, Nantes Université, CRCI²NA

Professeur des universités - Praticien hospitalier - Nantes Université, CRCI²NA

REMERCIEMENTS

Dans un premier temps, je remercie Stéphane Minvielle et Philippe Moreau, mes directeurs de thèse, qui m'ont offert l'opportunité de réaliser cette thèse. Ces quatre années passées au sein de l'équipe ont été très enrichissantes pour moi. Même si ce n'a pas été simple tous les jours de trouver la force de continuer à avancer malgré les difficultés.

J'adresse ensuite mes sincères remerciements aux membres du jury de ma thèse : Mary Callanan et David Garrick. Merci à eux de m'avoir fait l'honneur d'évaluer mon travail. Merci également pour l'intérêt qu'ils ont manifesté à l'égard de ces travaux.

Merci aussi aux membres de l'équipe, pour l'accueil chaleureux et les bons moments passés ensemble. Merci à Magali et Élise (et Nathalie) qui sont l'âme de cette équipe, à qui l'ont peu parler de tout, dans les bons comme les mauvais moments. Merci à Benjamin et Jean-Baptiste, mes premiers co-bureau, nous avons passé de bons moments ensemble. Merci à Jennifer et Victor mes seconds co-bureau, j'ai beaucoup apprécié travailler en votre compagnie, débriefer les réunions et tout simplement discuter. Merci aussi à vous deux d'avoir partagé avec moi votre expérience d'anciens doctorants, cela m'a beaucoup rassuré et cela m'a aidé à relativiser dans les moments difficiles. Merci aussi à tous les autres membres de l'équipe : Catherine, Anthony, Mia, Nils, Florence, Éric, Marie.

Merci aussi à Nathalie Labarrière, qui m'a ouvert les portes du CRCINA, en me faisant confiance pour un stage de master 1 puis de master 2, me permettant au passage de mettre un pied dans l'équipe de Stéphane.

Merci à mes bioloss préférés : Simon, Teddy, François, Florentin, Tanguy pour les bons moments de détente, autour d'une bière ou sur discord pour un peu de gaming.

Je remercie également ma famille qui m'a toujours soutenu dans mon parcours scolaire et universitaire, vous m'avez permis d'arriver là où j'en suis. Merci aussi à ma belle-famille pour le soutien, et l'intérêt que vous avez porté à mon travail.

Je sais que vous êtes tous fiers de moi, même si je ne suis pas toujours sûr de le mériter.

Enfin merci à Marie, ces années de thèse n'ont pas été facile pour nous deux. Elles nous ont apporté beaucoup de frustration sur le plan personnel. Merci pour ta patience et ton soutien tout au long du parcours. Merci pour tous ces moments qui m'ont permis de recharger mes batteries. Merci d'avoir pris le temps de relire et corriger cette thèse, qui sans toi ne serait pas d'aussi bonne qualité. T'avoir eu à mes côtés durant cette aventure donne tout son sens à l'accomplissement que représente pour moi cette thèse. Je t'aime.

Merci à vous tous qui faites partie de ma vie.

Jonathan Cruard

SOMMAIRE

Remerciements	3
Sommaire	4
Table des Figures	7
Liste des abréviations	10
1 Introduction	13
1.1 Le système immunitaire et l'hématopoïèse	14
1.1.1 Différenciation des cellules souches hématopoïétiques	14
1.1.2 La lignée myéloïde	14
1.1.3 Les lignées lymphoïdes T et NK	16
1.1.3.1 Les lymphocytes NK	16
1.1.3.2 Les lymphocytes T	17
1.1.4 La lignée lymphoïde B	19
1.1.4.1 Différenciation du B	19
1.1.4.2 Les plasmocytes	21
1.2 Le Myélome Multiple	24
1.2.1 Épidémiologie	24
1.2.2 Symptômes et Diagnostic	24
1.2.3 Genèse et stades précurseurs du Myélome Multiple (MM)	25
1.2.3.1 Les stades du MM	25
1.2.3.2 Origine génomique du MM	25
1.2.4 Prise en charge du MM	32
1.2.4.1 Le traitement du MM	32
1.2.4.2 Les grandes classes d'agents thérapeutiques utilisés dans le MM	34
1.2.4.3 La résistance du MM et ses mécanismes	39

2	Problématique et travaux de thèse	43
3	Méthodes	45
3.1	Traitement et préparation des échantillons	45
3.1.1	Prélèvement des échantillons	45
3.1.2	Au laboratoire	45
3.1.2.1	Traitement des échantillons et séquençage	45
3.2	Traitement des données de séquençage et contrôles qualité des profils d'ex- pression	46
3.2.1	Alignement des reads et création de la table de comptage	46
3.2.2	Contamination par les ARN ambiants	47
3.2.2.1	Principe de la contamination	47
3.2.2.2	SoupX et son fonctionnement	47
3.2.3	Présences de doublets dans les données	48
3.2.3.1	Origine des doublets	48
3.2.3.2	Les méthodes de détection des doublets	49
3.2.3.3	Méthodologie choisie pour notre pipeline	50
3.2.4	Variables de contrôle qualité	50
3.2.4.1	Nombre d'UMIs	51
3.2.4.2	Nombre de gènes détectés	53
3.2.4.3	Pourcentage de reads mitochondriaux	54
3.2.4.4	Pourcentage de reads ARN longs non codant (lncRNA)	54
3.2.5	Discussion des filtres	55
3.3	Normalisation	55
3.3.1	Objectif de la normalisation	55
4	Résultats	57
4.1	Étude du mode d'action de la dexaméthasone	57
4.1.1	Travail préliminaire, par Victor Bessonneau-Gaborit (GABORIT 2021)	57
4.1.2	Analyses en cellule unique de la lignée MM.1S exposée à la dexa- méthasone (dex)	59
4.1.3	Version finale de l'article présentant les résultats de notre étude du mode d'action de la dexaméthasone sur les cellules de myélome (BESSONNEAU-GABORIT et CRUARD 2023)	63

4.2	Étude du transcriptome en cellule unique du Myélome Multiple et son environnement immunitaire	81
4.2.1	Le jeu de données	81
4.2.2	L'environnement immunitaire	84
4.2.3	Composition de l'environnement immunitaire	86
4.2.4	Types cellulaires affinés	87
4.2.4.1	Changements de la composition des types cellulaires au cours de la genèse et de l'évolution de la maladie	89
4.2.5	Les interactions entre l'environnement et les plasmocytes	92
4.2.5.1	Variations quantitatives des interactions plasmocyte-environnement	92
4.2.5.2	Analyse de l'interaction (<i>CLEC2D</i>) LLT1-CD161 (<i>KLRB1</i>)	94
4.2.6	Analyse approfondie du compartiment T	98
4.2.6.1	Composition du compartiment T	98
4.2.6.2	État de la fonction T dans le MM	101
4.2.7	Les sous-populations de NK	104
4.2.8	Le compartiment myéloïde	108
4.2.9	Effet des BH3 mimétiques sur l'environnement immunitaire	111
4.2.9.1	Les BH3-mimétiques provoquent un remodelage de l'environnement immunitaire	111
4.2.9.2	La diversité de l'expression des gènes de la famille BCL2 .	115
5	Discussion et Conclusions	119
	Bibliographie	123

TABLE DES FIGURES

1.1	Schéma simplifié de l'hématopoïèse	15
1.2	La recombinaison VDJ	20
1.3	Les différents isotypes d'immunoglobulines (IGs)	20
1.4	L'acquisition de la mémoire B	22
1.5	Environnement de survie du plasmocyte	23
1.6	Le développement de la gammopathie monoclonale	26
1.7	Schéma des translocations du MM issues des événements génomiques impliqués dans la maturation B	28
1.8	Le génome du MM au diagnostic	29
1.9	Mécanismes d'action moléculaire du récepteur aux glucocorticoïdes sur la transcription des gènes	35
1.10	Schéma simplifié de la structure et du mode d'action des anticorps bispécifiques	38
1.11	Les membres de la famille BCL2	39
4.1	Schéma des techniques utilisées pour établir les effets de la dex sur le génome de la lignée MM.1S	58
4.2	Heatmap de corrélation des 51 gènes les plus sur-exprimés en réponse à la dex dans les données scRNA-seq	60
4.3	Schéma des différents mode d'action de la dexaméthasone sur la régulation de l'expression des gènes cibles	62
4.4	Frise chronologique des prélèvements	81
4.5	Tableau des échantillons issus de la cohorte MYRACLE	82
4.6	UMAP du dataset complet, environnement et tumeur	83
4.9	Tableau des types cellulaires et des marqueurs qui leur sont associés	85
4.10	Heatmap de l'abondance relative des principaux types cellulaires	86
4.11	Boxplots du pourcentage de cellules représenté par chaque type cellulaire	87
4.12	Dotplot des principaux marqueurs de chaque type cellulaire fin	88
4.13	La signature interféron dans le jeu de données	88

4.14	Heatmap de l'abondance relative des types cellulaires fins	89
4.15	Heatmap de corrélation des types cellulaires fins	91
4.16	Corrélations entre les proportions de certains types cellulaires	92
4.17	Diagramme de Venn récapitulant les interactions présentes au différents stades de la pathologie	93
4.18	Comparaison des interactions entre les différents stades de la maladie	94
4.19	Interaction <i>CLEC2D-KLRB1</i> entre l'environnement et les plasmocytes	95
4.20	L'expression des gènes impliqués dans l'interaction <i>CLEC2D-KLRB1</i>	96
4.21	Histogrammes de FACS des niveaux d'expression de la protéine LLT1 dans des lignées de myélome et dans des échantillons de patients	97
4.22	UMAP du compartiment T	99
4.23	Dotplots des principaux marqueurs des T	99
4.24	Boxplots des proportions de T au sein de chaque échantillon, par stade	100
4.25	Proportion de MAIT chez les patients avec des échantillons aux stades NDMM et ERRMM	101
4.26	Boxplot des ratios CD4/CD8 au sein des échantillons à chaque stade	101
4.27	Niveau d'expression des signatures d'exhaustion et de cytotoxicité au sein des populations T effectrices cytotoxiques	102
4.28	Expression des gènes associés à l'exhaustion et la dysfonction dans les lymphocytes T gamma delta ($T\gamma\delta$)	103
4.29	UMAPs du compartiment NK	105
4.30	Dotplots des principaux marqueurs des NK	106
4.31	Boxplots des proportions de NK au sein de chaque échantillons, par stade	106
4.32	Corrélations entre les proportions des types de NK enrichis au stade NDMM107	107
4.33	Dotplots de la signature de cytotoxicité dans les NK	107
4.34	UMAP du compartiment myéloïde	108
4.35	Dotplots des principaux marqueurs des types cellulaires de la lignée myéloïde	109
4.36	Boxplots des proportions des types cellulaires de la lignée myéloïde au sein de chaque échantillon, par stade	110
4.37	UMAP de l'environnement contrôle et BH3	111
4.38	Evolution des proportions des types cellulaires principaux après traitement au BH3 mimétiques	113
4.39	Evolution des proportions des types cellulaires fins après traitement au BH3 mimétiques	114

LISTE DES ABRÉVIATIONS

ACP analyse en composantes principales.

APC cellule présentatrice d'antigène.

ASCT auto-greffe de cellules souches hématopoïétiques (autologous stem cell transplantation en anglais).

ATAC-seq accessibilité de la chromatine (assay for transposable-accessible chromatin sequencing en anglais).

BCR récepteur des lymphocytes B.

ChIP-seq immunoprécipitation de chromatine.

CLP précurseurs lymphoïdes communs.

CMH II complexe majeur d'histocompatibilité de classe II.

CMP précurseurs myéloïdes communs.

DC cellules dendritiques.

del13 délétion du chromosome 13.

del17p délétion du 17p.

dex dexaméthasone.

ERRMM Early-RRMM.

FUMM Follow-up-MM.

GC centre germinatif.

GR récepteur aux glucocorticoïdes.

GRE GR-responsive element.

HSC cellules souches hématopoïétiques.

IFN γ interféron gamma.

IG immunoglobuline.

IL interleukine.

IMiD immunomodulateur.

IP inhibiteur du protéasome.

IRM imagerie par rayonnement magnétique.

lncRNA ARN longs non codant.

LRRMM Late-RRMM.

MAD déviations absolues de la médiane (median absolute deviation en anglais).

MAIT cellules T mucosales associées aux tissus (Mucosal-Associated Invariant T cells en anglais).

MGUS gammopathie monoclonale de signification indéterminée.

MM Myélome Multiple.

NBM moelle normale (Normal Bone Marrow en anglais).

ncMono Monocytes non conventionnel.

NDMM Myélome Multiple nouvellement diagnostiqué.

NF- κ B facteur nucléaire-kappa B.

NK natural killer.

OLS organes lymphoïdes secondaires.

pDC cellules dendritiques plasmacitoïdes.

pE enhancer prédominant.

RE réticulum endoplasmique.

RNA-seq séquençage de l'ARN.

RRG ratio of responding genes.

RRMM Myélome Multiple réfractaire et en rechute.

scRNA-seq séquençage de l'ARN en cellule unique.

SHM hypermutation somatique.

SI système immunitaire.

SMM myélome multiple indolent.

T $\gamma\delta$ lymphocytes T gamma delta.

TAM macrophages associés à la tumeur.

TCR récepteur des lymphocytes T.

TEP tomographie par émission de positons.

Th1 T helper 1.

Th2 T helper 2.

TNF β tumor necrosis factor beta.

Treg lymphocytes T régulateurs.

trNK tissue resident NK.

UMAP Uniform Manifold Approximation and Projection.

UPR réponse aux protéines mal repliées (Unfolded protein response en anglais).

INTRODUCTION

Le Myélome Multiple (MM) est une hémopathie maligne dont les cellules tumorales dérivent du plasmocyte à longue durée de vie. Au sein du système immunitaire (SI), les plasmocytes sont en charge de la production d'anticorps. Ils sont le stade de différenciation le plus avancé de la lignée lymphocytaire B et résident au sein de la moelle osseuse où ils perdurent jusqu'à plusieurs décennies. C'est donc également au sein de la moelle osseuse que se développe principalement le MM. La moelle osseuse est également le siège de l'hématopoïèse et donc en quelque sorte du SI. Le SI a pour rôle de protéger l'organisme contre les menaces telles que les infections, les parasites mais aussi les cellules cancéreuses. Dans divers cancers, il a été démontré que le SI joue un rôle majeur dans le contrôle et/ou la progression de la maladie. On observe notamment des changements de phénotypes induits par la tumeur et par l'exposition à l'environnement tumoral. Ces changements ont notamment été observés dans le contexte de tumeurs solides où l'environnement est dense et faiblement renouvelé et l'accès au site tumoral est donc difficile pour les cellules immunitaires. Dans le cas du MM, le milieu est "hybride". La moelle n'est pas à proprement parler un tissu solide car elle est fortement perfusée. Cependant certaines cellules y résident et ne sont, en temps normal pas observées dans le sang. Le MM est donc un cas particulier lorsque l'on s'intéresse aux interactions entre le SI et les tumeurs : les plasmocytes font partie du SI et la tumeur se développe au sein du tissu hématopoïétique primaire y attirant des cellules immunitaires matures.

1.1 Le système immunitaire et l'hématopoïèse

Le SI est chargé de protéger l'organisme contre diverses menaces telles que les virus, les bactéries ou encore les parasites mais aussi les cellules cancéreuses. Le SI est composé de nombreux acteurs cellulaires. Chacun de ces types cellulaires remplit un rôle spécifique au sein du système et interagit généralement avec le reste du SI via des mécanismes de régulation. Au sein du SI, on distinguera l'immunité innée de l'immunité adaptative, qui sont assurées par des types cellulaires distincts.

1.1.1 Différenciation des cellules souches hématopoïétiques

L'ensemble des cellules du SI sont produites au sein de la moelle osseuse à partir des cellules souches hématopoïétiques (HSC). Ces HSC résident dans la moelle et sont capables d'auto-renouvellement. L'hématopoïèse désigne le processus de production de l'ensemble des cellules sanguines à partir des HSC (figure 1.1). L'hématopoïèse est un processus complexe et très ramifié. L'engagement et la progression vers les types cellulaires les plus matures sont régulés par de nombreux signaux et notamment des interleukines (ILs). Dans un premier temps, les HSC engendrent deux types de progéniteurs : les précurseurs lymphoïdes communs (CLP) et les précurseurs myéloïdes communs (CMP). Les CMP donneront les différents types cellulaires de la lignée myéloïde : les érythrocytes, les mégacaryocytes, les cellules dendritiques et les monocytes et macrophages. De leur côté, les CLP sont à l'origine des lignées lymphocytaires T, B et natural killer (NK).

1.1.2 La lignée myéloïde

Une fois engagés dans la lignée myéloïde, les CMP vont maturer et se différencier pour engendrer divers types cellulaires, tels que les érythrocytes et les mégacaryocytes, respectivement à l'origine des hématies et des plaquettes. Les CMP se différencient également en myéloblastes, qui à leur tour deviendront les granulocytes et les monoblastes. Il existe trois sous-types de granulocytes : les basophiles, les neutrophiles et les éosinophiles. Les monoblastes quant à eux donneront les monocytes, qui eux même se différencieront en macrophage et en cellule dendritique. Les monocytes sont présents dans la moelle osseuse et circulent dans le sang. Le principal marqueur utilisé pour les identifier est le CD14. Ils se différencieront en macrophages lors de leur entrée dans les tissus. L'entrée dans les tissus est provoquée par divers signaux inflammatoires. Une fois sur site, les monocytes

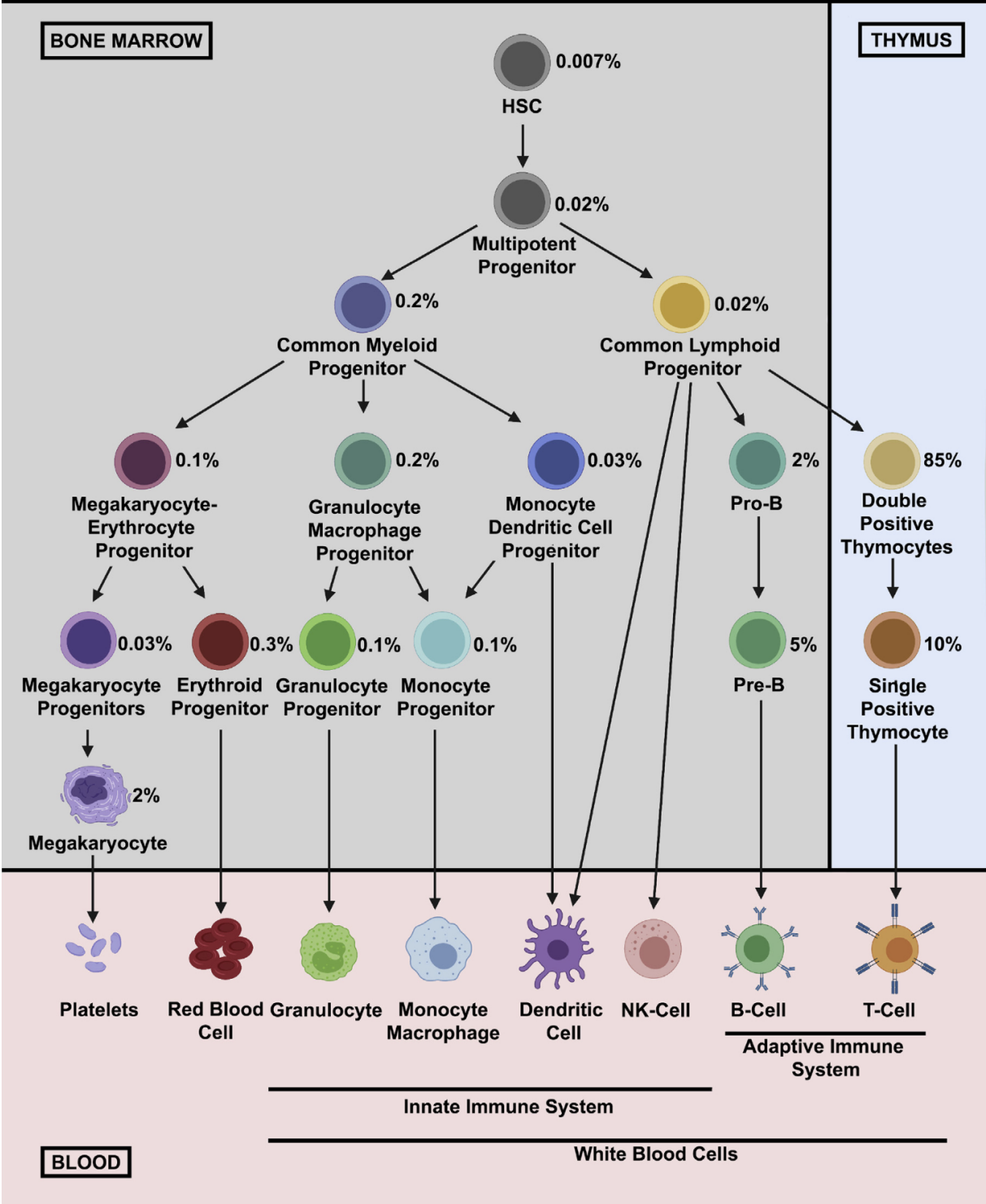


FIGURE 1.1 – Schéma simplifié de l'hématopoïèse, figure issue de COMAZZETTO, SHEN et MORRISON 2021

vont à leur tour renforcer le signal afin de recruter d'autres acteurs du SI. Le principal rôle des macrophages est la phagocytose, permettant d'éliminer les résidus de cellules ainsi que les pathogènes. Ils ont également pour rôle de présenter les antigènes issus de la phagocytose aux lymphocytes T, permettant leur activation. Parmi les fonctions des macrophages, on retrouve également la production de cytokines. Les monocytes et macrophages sont des types cellulaires extrêmement plastiques et de nombreux phénotypes existent. Les plus décrits et étudiés sont les sous-types dits M1 et M2 respectivement pro et anti-inflammatoires. Cependant, cette distinction est principalement issue d'expérimentations *in vitro* et ne semble pas aussi claire dans le contexte d'échantillons *in vivo*, où l'on peut observer des phénotypes hybrides ou intermédiaires.

1.1.3 Les lignées lymphoïdes T et NK

Il existe une grande diversité de types et de sous-types cellulaires au sein de la population de lymphocytes issue des CLP. Chacun de ces sous-types remplit une fonction ou une combinaison de fonctions qui lui est propre.

1.1.3.1 Les lymphocytes NK

L'un des destins possible des CLP est la différenciation en NK. Ces lymphocytes font partie du système immunitaire dit inné, puisqu'ils n'ont pas de récepteurs membranaires spécifiques d'un antigène contrairement aux lymphocytes T et B. Ces lymphocytes sont capables d'identifier et d'éliminer les cellules anormales telles que les cellules infectées par un virus ou les cellules tumorales. Ils remplissent ce rôle grâce à leurs fortes capacités cytotoxiques mais aussi par la production de cytokines pro-inflammatoires qui permettent le recrutement et la stimulation d'autres acteurs du SI. L'expression du récepteur CD56 (codé par le gène *NCAM1*) est l'un des principaux marqueurs utilisé pour identifier les lymphocytes NK. Les NK sont traditionnellement séparés en deux populations : les NK CD56_{dim} et les NK CD56_{bright}, présentant respectivement des niveaux intermédiaires et forts du CD56. Les NK CD56_{bright} sont considérés comme moins matures et représentent une minorité des NK en circulation dans le sang ($= < 15\%$). La proportion relative de ces deux sous-types varie en fonction du tissu considéré. Dans le cas de la moelle, la proportion est également en faveur des CD56_{dim}. Les NK CD56_{bright} ont des capacités cytotoxiques moindres que les NK_{dim}. À l'inverse, ils sont capables de produire des quantités plus importantes de cytokines. Au delà de cette classification binaire, il est possible de définir

d'autres sous-types de NK. La majorité d'entre eux est spécifique du tissu où ils résident : ce sont les tissu resident NK (trNK). Dans la moelle, les trNK sont caractérisés par l'expression de *CXCR6* et *CCR5* ou encore du *CD69*. Il existe également les NK adaptatifs caractérisés par l'expression de *KLRC2* et des gènes codants pour le complexe majeur d'histocompatibilité de classe II (CMH II). Ces NK sont principalement observés dans des contextes de cancers et d'infections virales et notamment au cytomégalovirus. Ils sont capables de persister dans le temps et de proliférer rapidement en cas de ré-infection, ce qui fait d'eux des acteurs du SI adaptatif (FREUD et al. 2017 ; CRINIER et al. 2021).

1.1.3.2 Les lymphocytes T

Les CLP engagés dans la lignée T vont migrer vers le thymus où ils vont poursuivre leur maturation. À ce stade, ils n'expriment pas les marqueurs de surface caractéristiques de la lignée T tels que le CD3, le récepteur des lymphocytes T (TCR) ou encore le CD4 et le CD8. Au sein du thymus, ces progéniteurs vont proliférer et passer par divers stades phénotypiques distingués par l'expression de marqueurs de surface. C'est au cours de cette différenciation que les lymphocytes T vont procéder au réarrangement VDJ (voir section 1.1.4.1 et figure 1.2) de la chaîne β puis α de leur TCR. L'expression du TCR ainsi obtenu couplée à l'expression du CD3 va permettre aux lymphocytes d'entrer dans un processus de sélection. À ce stade, ils sont doublement positifs en CD4 et CD8. La sélection se déroule en deux étapes. La première est une sélection positive, elle permet d'assurer la capacité du TCR réarrangé à se lier au CMH. La liaison au CMH va transmettre un signal de survie au lymphocyte, lui permettant d'échapper à l'apoptose. Les cellules subissent alors la sélection négative. Au cours de celle-ci, des macrophages et des cellules dendritiques présentent des CMH du soi aux lymphocytes. Ces CMH sont chargés ou non avec des antigènes du soi, en cas d'affinité de leur TCR pour ces antigènes, les lymphocytes reçoivent un signal de mort et entrent en apoptose. Cette étape de sélection est primordiale afin de prévenir l'apparition de lymphocytes T auto-réactifs qui s'attaqueraient alors à des cellules saines. À ce stade, les lymphocytes sont qualifiés de naïfs et perdent l'expression du CD4 ou du CD8. Ils sont par ailleurs incapables de proliférer. Ils quittent le thymus et entrent dans la circulation afin d'augmenter les chances de rencontre avec l'antigène dont ils sont spécifiques. La rencontre avec l'antigène et le signal de co-stimulation délivré par la cellule présentatrice d'antigène (APC) va permettre au lymphocyte T de s'activer, il récupère alors sa capacité de prolifération et se différencie en cellules T mémoire ou effectrice. L'expansion de ce clone de lymphocyte T donne

naissance à diverses sous-populations spécialisées. Les cellules effectrices à durée de vie courte comprennent des sous-types spécialisés dans la sécrétion de cytokines ou avec une forte capacité cytotoxique. Dans le cas des T CD4, les cellules effectrices chargées de la production de cytokines sont appelées "helper", avec deux sous-populations principales aux profils cytokiniques distincts. La population T helper 1 (Th1) produit principalement de l'IL-2, de l'interféron gamma (IFN γ) et du tumor necrosis factor beta (TNF β). Cette population contribue à l'activation des lymphocytes cytotoxiques. La seconde population T helper 2 (Th2) produit de l'IL-4, IL-5, IL-6 et de l'IL-10 et participe à la mise en place de la réponse humorale en stimulant les lymphocytes B. Les lymphocytes mémoires quant à eux ont une durée de vie longue et assurent une réponse rapide en cas de contact avec l'antigène. L'efficacité accrue de cette réponse permet une seconde réaction plus efficace dans le cas où la primo-infection n'est pas encore jugulée mais aussi en cas de réinfection future. En parallèle, le thymus produit des lymphocytes T ayant pour rôle de limiter la prolifération des T effecteurs. Ils jouent un rôle essentiel dans la tolérance immunitaire, en limitant la reconnaissance d'antigènes du soi et d'antigènes non dangereux. Leur action immunosuppressive passe également par l'inhibition des APC. Cette population de lymphocytes T régulateurs (Treg) est généralement caractérisée par l'expression de FOXP3, qu'elle est la seule à exprimer parmi les lymphocytes T. Au sein de la population T, il existe également des lymphocytes aux caractéristiques particulières. Ils se distinguent par leurs TCR dont les fonctions diffèrent des lymphocytes conventionnels. Ces populations expriment des TCR à la variabilité limitée et qui ne se lient pas nécessairement au CMH. Les NK T notamment ont un TCR qui reconnaît les glycolipides présentés par le CD1D. La dénomination NK T dérive de leur expression simultanée de récepteurs associés aux lignées de lymphocytes T et NK. Parmi ces lymphocytes non conventionnels, on retrouve également les cellules T mucoales associées aux tissus (Mucosal-Associated Invariant T cells en anglais) (MAIT). Ils expriment un TCR semi-invariant reconnaissant le récepteur MR1 spécialisé dans la présentation d'antigène issus de métabolites bactériens. Enfin, les lymphocytes T gamma delta (T $\gamma\delta$) suivent un processus de recombinaison VDJ similaire aux lymphocytes T conventionnels cependant le répertoire T $\gamma\delta$ est plus limité et certaines combinaisons de gènes sont favorisées. Ces combinaisons sont généralement associées à des fonctions spécifiques. Une des caractéristiques intéressantes de cette population est sa capacité à reconnaître son antigène spécifique sans qu'il soit présenté par une molécule de la famille du CMH (LE BOURHIS, MBURU et LANTZ 2013).

1.1.4 La lignée lymphoïde B

Dans le cas du MM, nous nous intéressons plus particulièrement à la lignée B et sa différenciation dont le plasmocyte à longue durée de vie représente le stade final.

1.1.4.1 Différenciation du B

Les lymphocytes B sont issus des CLP. Ceux-ci s'engagent dans cette voie de différenciation par l'expression des facteurs de transcription codés par les gènes *TCF3*, *EBF1* et *PAX5*. Le pro-B obtenu va ensuite exprimer les gènes *RAG1*, *RAG2* et *DNTT*. Les enzymes issues de l'expression de ces gènes vont permettre la recombinaison VDJ (figure 1.2) qui conduira à la formation du récepteur des lymphocytes B (BCR). Dans un premier temps, c'est la chaîne lourde des immunoglobulines (IGs) située sur le chromosome 14 qui va subir la recombinaison. Celle-ci débute par l'association d'un gène D et d'un gène J par excision de l'ADN et par conséquent des gènes situés entre eux (figure 1.2). Ensuite, un gène V va être ajouté de façon similaire, par élimination de l'ADN le séparant du complexe DJ précédemment formé. Si le VDJ ainsi obtenu n'est pas fonctionnel, le second allèle subit alors le même processus. Dans un second temps, c'est la chaîne légère des IGs qui sera réarrangée, le processus est similaire bien que, dans ce cas il n'y a pas de gène D. La chaîne légère subit donc un unique réarrangement VJ. De même que pour la chaîne lourde, si le résultat obtenu n'est pas fonctionnel, le second allèle est réarrangé. Le réarrangement touche dans un premier temps la chaîne légère κ sur le chromosome 2. Si aucun de ses allèles n'est fonctionnel, c'est alors la chaîne λ sur le chromosome 22 qui est réarrangée.

Les lymphocytes B immatures ainsi formés subissent alors une phase de sélection similaire à celle décrite précédemment pour les lymphocytes T. Les lymphocytes B dont le BCR n'est pas capable de se lier au CMH seront éliminés ainsi que ceux spécifiques pour des antigènes du soi. Ce qui évite à nouveau la formation d'un répertoire immunitaire auto-immun. À ce stade, les lymphocytes expriment un isotype IgM des IGs (figure 1.3). Ces lymphocytes B entrent dans la circulation afin de rejoindre les organes lymphoïdes secondaires (OLS), à ce stade ils acquièrent l'expression de l'isotype IgD en plus de l'IgM (figure 1.3). Ces deux isotypes sont exprimés simultanément et reconnaissent le même antigène. Ces lymphocytes sont alors au stade B mature naïf.

Lorsque le BCR des lymphocytes B naïfs se lie à son antigène, un processus d'internalisation du récepteur se déclenche. L'antigène est par la suite lié à un CMH II puis présenté

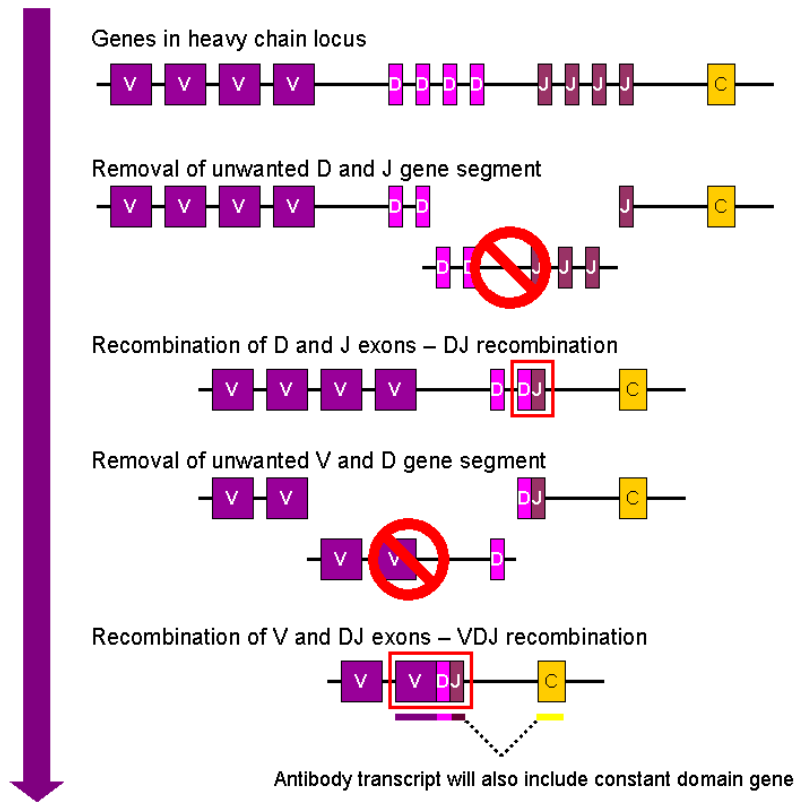


FIGURE 1.2 – Schéma simplifié de la recombinaison VDJ

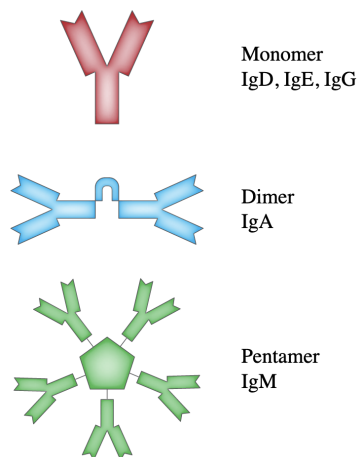


FIGURE 1.3 – Les différents isotypes d'IGs

à la surface de la cellule. L'interaction de ce CMH II chargé de l'antigène avec le TCR d'un lymphocyte Th2, va permettre au lymphocyte B de proliférer et de se différencier. Trois destins sont possibles pour ces lymphocytes : plasmocyte à courte durée de vie, lymphocyte B du centre germinatif (GC) et lymphocyte B indépendant du GC (figure 1.4 : phase 1). Les plasmocytes à courte durée de vie permettent la mise en place rapide d'une production d'anticorps reconnaissant l'antigène et donc de juguler la progression du pathogène. Les anticorps de cette première réponse ont cependant une relativement faible affinité pour l'antigène. Les lymphocytes B du GC entrent dans le GC pour proliférer et subir la maturation d'affinité (figure 1.4, phase 2). L'augmentation de l'affinité est permise par le mécanisme d'hypermutation somatique (SHM). La SHM consiste en une modification aléatoire de la séquence de l'IG. Elle est suivie d'une sélection des lymphocytes les plus réactifs ainsi obtenus par interaction avec les lymphocytes T helper folliculaires. Les lymphocytes B sélectionnés se différencient alors en B mémoires, en plasmocytes à longue durée de vie ou bien retournent dans le GC pour un nouveau cycle de maturation d'affinité. Les lymphocytes B mémoires ont une durée de vie longue, ils permettent de pérenniser les capacités de défenses de l'organisme face aux pathogènes déjà rencontrés. Il est à noter que les B mémoires possèdent une diversité d'anticorps plus élevée et donc une affinité plus faible pour le pathogène primaire que les plasmocytes à longue durée de vie. Cette diversité couplée à leur capacité à entrer dans un nouveau cycle de maturation d'affinité permet une réponse efficace dans le cas d'infection par un pathogène homologue ou par un variant (AKKAYA, KWAK et PIERCE 2019).

1.1.4.2 Les plasmocytes

Comme évoqué dans la section précédente, il existe deux types de plasmocytes respectivement à courte et longue durée de vie. Le plasmocyte à longue durée de vie est le stade de différenciation final de la lignée lymphocytaire B. Il est en charge de la production d'anticorps solubles hautement spécifiques d'un unique antigène du pathogène ayant déclenché la réponse immunitaire. La différenciation en plasmocyte est régie par l'expression des facteurs de transcription suivants : IRF4, PRDM1 et XBP1. IRF4 est exprimé avant et après la différenciation en plasmocyte. Ce facteur de transcription a un effet dose dépendant : une faible expression permet le maintien de l'expression des gènes de l'identité B et l'entrée dans le GC. Fortement exprimé, il promeut le passage au stade plasmocytaire, notamment en permettant l'expression de *PRDM1*. PRDM1 réprime alors l'expression des gènes de l'identité B. XBP1 est quant à lui associé à la réponse aux pro-

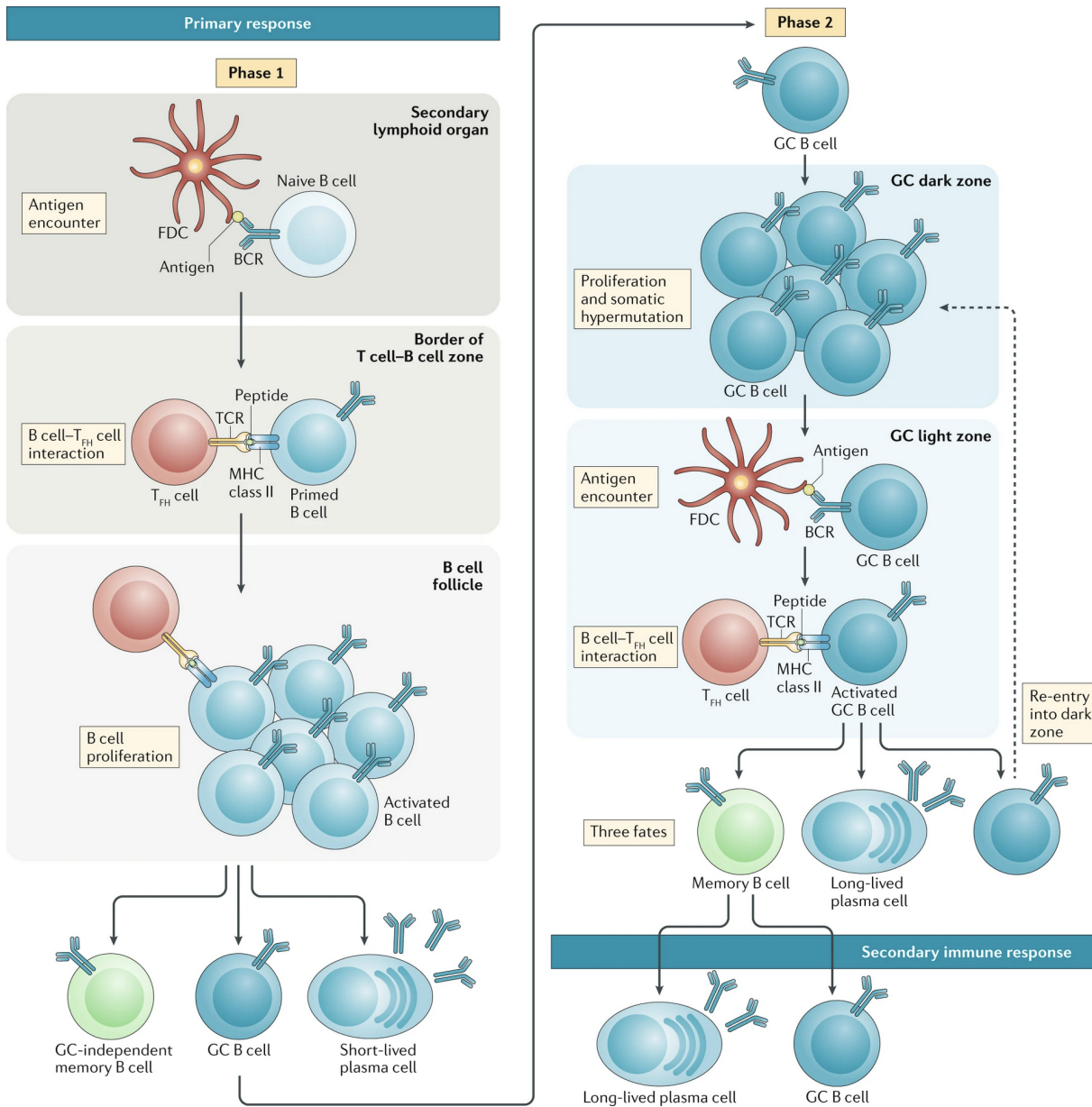


FIGURE 1.4 – L’acquisition de la mémoire B, figure issue de AKKAYA, KWAK et PIERCE 2019

téines mal repliées (Unfolded protein response en anglais) (UPR). L'UPR est déclenchée par le stress du réticulum endoplasmique (RE). La production intensive d'anticorps par les plasmocytes les expose particulièrement à ce stress.

Les plasmocytes résident au sein de la moelle osseuse. Ils sont recrutés sur site par les cellules stromales qui expriment et produisent du CXCL12. Les plasmocytes sont attirés par chimiotactisme lors de la liaison du CXCL12 sur le récepteur CXCR4 présent à leur surface. La rétention des plasmocytes au sein de la moelle implique quant à elle l'interaction de divers récepteurs présents sur le plasmocyte tels que VLA4, CD44, CD28 et CD93. Les plasmocytes persistent dans la moelle pendant plusieurs années. Dans un contexte non pathologique, ils sont incapables de proliférer. Leur capacité de persistance n'est pas intrinsèque, lorsqu'ils sont sortis de l'environnement de la moelle cela conduit rapidement à leur mort. Il semble que les principaux facteurs permettant la survie des plasmocytes soient IL-6 et le TNFSF13 (aussi appelé APRIL) produits par des cellules présentes dans l'environnement de la moelle. La liaison du TNFSF13 avec son récepteur TNFRSF17 (aussi appelé BCMA) provoque l'expression de MCL1, une protéine anti-apoptotique de la famille BCL2. La demi-vie de la protéine MCL1 est très courte ce qui explique la nécessité d'une exposition continue à l'environnement de la moelle pour la survie des plasmocytes. Les plasmocytes sont généralement caractérisés par l'expression de TNFRSF17, de SLAMF7, du CD38 et enfin de SDC1 (CD138).

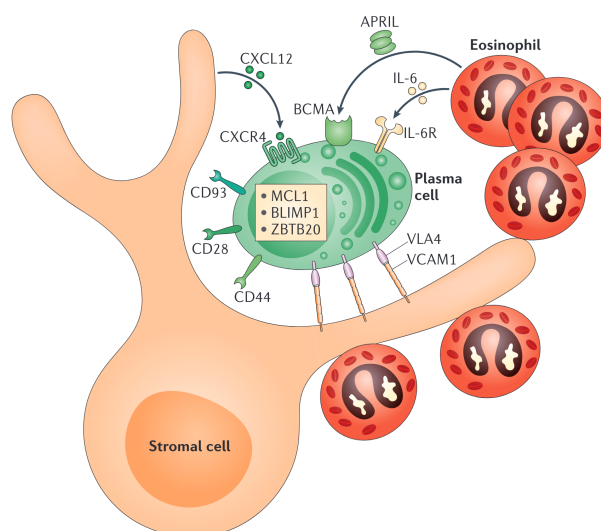


FIGURE 1.5 – Environnement de survie du plasmocyte, figure issue de NUTT et al. 2015

1.2 Le Myélome Multiple

Comme évoqué précédemment, le MM est caractérisé par la prolifération de plasmocytes malins au sein de la moelle osseuse. Au cours de la maladie, le MM peut évoluer et s'affranchir de l'environnement médullaire, on parle alors de leucémie à plasmocytes.

1.2.1 Épidémiologie

Le MM est un cancer relativement rare, environ 6000 nouveaux cas sont diagnostiqués chaque année en France. Cela représente approximativement 2% de l'ensemble des cancers. À l'échelle mondiale, on recense environ 150 000 nouveaux cas par an. Le principal facteur de risque du MM est l'âge, avec une médiane de 72 ans au diagnostic chez l'homme et de 74 ans chez la femme (LAPÔTRE-LEDOUX et al. 2023). Seuls 3% des MM sont diagnostiqués chez des patients de moins de 40 ans. Le sexe est aussi un facteur de risque, pour un homme le risque de développer un MM est 1,5 fois plus élevé que pour une femme. Par ailleurs, il existe des facteurs de risque génétiques, notamment chez les personnes d'origine africaine qui sont plus à risque tandis que les personnes d'origine asiatique présentent un moindre risque. Ces déterminants génétiques se retrouvent également à l'échelle familiale. Enfin, l'exposition aux pesticides augmente les risques de MM et de ses stades précurseurs (S. K. KUMAR et al. 2017; PADALA et al. 2021).

1.2.2 Symptômes et Diagnostic

Le MM est associé à différents symptômes non exclusifs de cette pathologie. Le diagnostic nécessite donc des examens supplémentaires pour être confirmé. Les principaux symptômes sont connus sous l'acronyme CRAB. Environ 10% des patients présentent une hyperCalcémie, qui est par ailleurs un symptôme de mauvais pronostic. L'hypercalcémie résulte de la diminution de la masse osseuse causée par l'activation des ostéoclastes par le MM. L'insuffisance Rénale est quant à elle provoquée par la forte production d'IGs monoclonales des cellules tumorales. La forte concentration de ces protéines dans le sang surcharge les capacités de réabsorption rénale provoquant la formation de dépôts et d'obstructions au niveau des reins. L'Anémie est une autre conséquence de l'envahissement de la moelle par la tumeur. En occupant la moelle, les plasmocytes malins réduisent l'espace et les ressources disponibles pour le bon déroulement de l'hématopoïèse et le renouvellement des cellules sanguines. Cela conduit à un déficit en hématies, plaquettes et globules

blancs, avec pour conséquences un mauvais contrôle des saignements et une susceptibilité aux infections. Enfin, les douleurs et lésions osseuses (Bone lesions) sont provoquées par l'activation des ostéoclastes par le MM. Cette activation vient perturber l'équilibre entre la dégradation et le renouvellement osseux, ce qui provoque des fragilités osseuses et une élévation des risques de fractures.

En cas de suspicion de MM, des examens approfondis sont effectués pour confirmer le diagnostic. Un examen sanguin et une analyse d'urine sont réalisés pour détecter la présence d'une IG monoclonale, d'hypercalcémie et pour évaluer l'état de fonctionnement rénal. Un myélogramme est également réalisé. Cette analyse permet de détecter la sur-représentation et les anomalies morphologiques des plasmocytes au sein d'un échantillon de ponction de moelle. Une partie du matériel biologique de la ponction peut également permettre de rechercher les anomalies chromosomiques présentes dans les cellules tumorales. Enfin, les examens d'imagerie IRM et TEP permettent d'améliorer la prise en charge des patients (DIMOPOULOS et al. 2015 ; CAVO et al. 2017).

1.2.3 Genèse et stades précurseurs du MM

1.2.3.1 Les stades du MM

Le MM fait partie des gammopathies monoclonales, ce terme regroupe les différents stades d'envahissement de la moelle par des plasmocytes produisant des IG monoclonales (figure 1.6). Le premier stade est la gammopathie monoclonale de signification indéterminée (MGUS), il n'est pas symptomatique et le taux d'infiltration de la moelle par les plasmocytes est inférieur à 10%. Ce stade est assez courant, environ 3 à 4% des personnes de plus de 50 ans sont atteintes dont environ 25% progresseront vers un MM après 25 ans (KYLE et al. 2010). Le stade myélome multiple indolent (SMM) se situe entre le MGUS et le MM. À ce stade, l'envahissement de la moelle est plus important (10 à 60%) sans pour autant provoquer l'apparition de symptômes cliniques. Enfin, le stade le plus avancé de la maladie est la leucémie à plasmocytes. À ce stade, les plasmocytes se sont affranchis des signaux de survie qui leur sont fournis par l'environnement médullaire. Ils peuvent survivre en dehors de la moelle et envahissent le sang.

1.2.3.2 Origine génomique du MM

Le MM est une maladie très hétérogène à tous les niveaux, c'est notamment le cas au niveau génomique. Il existe une variété importante d'altérations génomiques présente

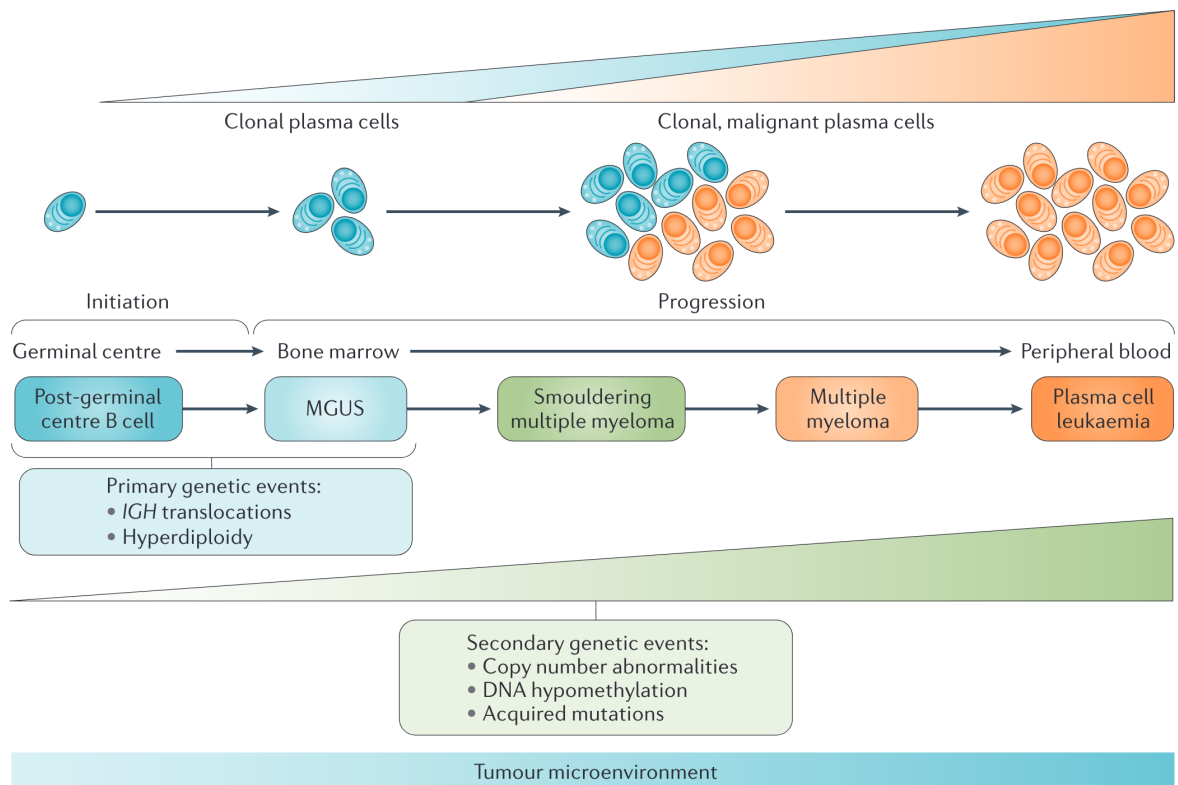


FIGURE 1.6 – Le développement de la gammopathie monoclonale, figure issue de S. K. KUMAR et al. 2017

au diagnostic dans le MM (figure 1.8). Comme dans de nombreux types de cancers, on distinguera les événements génomiques primaires et secondaires. Les événements primaires sont considérés comme le point de départ de la maladie.

Les évènements primaires Dans le MM, on regroupe les tumeurs en deux groupes basés sur les types d'altérations génomiques primaires qu'elles arborent : les tumeurs hyper-diploïdes et les tumeurs non-hyper-diploïdes. Les MM hyper-diploïdes sont caractérisés par la présence de copies sur-numéraires d'au moins deux des chromosomes impairs suivants : 3, 5, 7, 9, 11, 15, 19 et 21. Les mécanismes à l'origine de l'hyperdiploïdie sont inconnus. Les tumeurs hyperdiploïdes représentent environ 50% des MM tandis que l'autre moitié des MM présente des translocations. La majorité des translocations primaires dans le MM implique le locus de la chaîne lourde des IGs (IGH). Ces translocations trouvent leur origine dans le processus de maturation des lymphocytes B. En effet, la recombinaison VDJ, la SHM ainsi que la commutation isotypique incluent des étapes durant lesquelles une cassure double brin de l'ADN est effectuée (WALKER, WARDELL et al. 2013; S. K. KUMAR et al. 2017). Il peut alors arriver que les points de cassure fusionnent avec un autre locus du génome ayant subi une cassure fortuite, donnant lieu à une fusion aberrante. Dans la plupart des cas, ces évènements sont bénins et conduisent à la disparition de la cellule, puisqu'ils ne lui confèrent pas d'avantage. Cependant, en cas de recombinaison avec certains oncogènes, les cellules acquièrent un avantage sélectif permettant leur croissance et leur prolifération (figure 1.7).

En effet, les IGH sont destinées à être produites massivement par les plasmocytes. Leur enhancer est très puissant, et une fois juxtaposé à un oncogène, il décuple son expression. Les translocations de l'IGH impliquent un groupe limité et récurrent d'oncogènes : CCND1 t(11;14), NSD2 et FGFR3 t(4;14), MAF t(14;16), CCND3 t(6;14) et MAFB t(14;20) par ordre de fréquence. L'ensemble de ces évènements chromosomiques est observé à tous les stades de la maladie, précurseurs comme avancés. Cela permet d'affirmer qu'ils apparaissent aux prémices de la maladie et sont impliqués dans son initiation (S. K. KUMAR et al. 2017). Cependant, puisque ces évènements sont présents à tous les stades, ils n'expliquent pas la progression vers les stades symptomatiques de la pathologie. La progression nécessite donc l'apparition et l'accumulation d'autres anomalies, intrinsèques et/ou extrinsèques à la tumeur. Au niveau génomique, l'apparition d'évènements secondaires permet d'expliquer une partie des progressions.

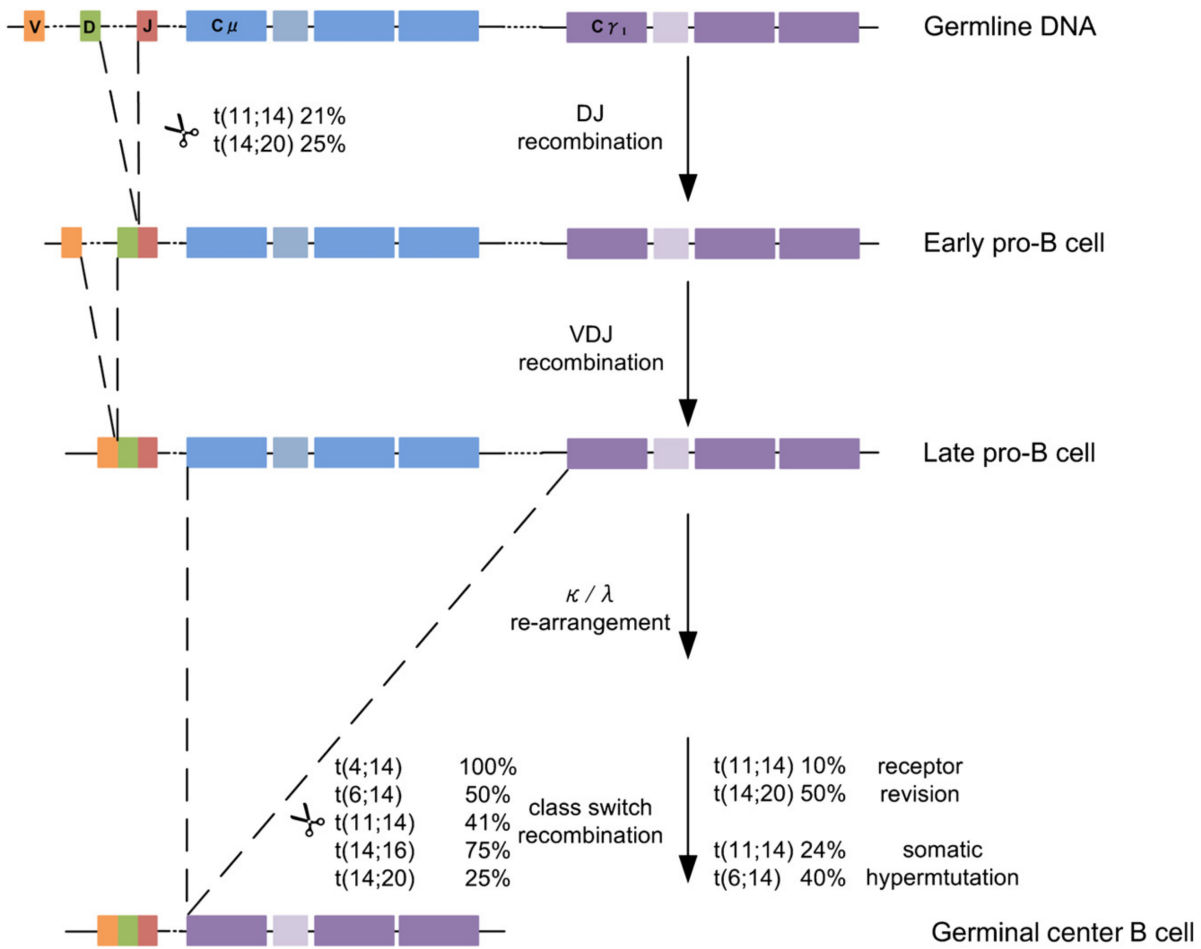


FIGURE 1.7 – Schéma des translocations du MM issues des événements génomiques impliqués dans la maturation B, figure issue de WALKER, WARDELL et al. 2013

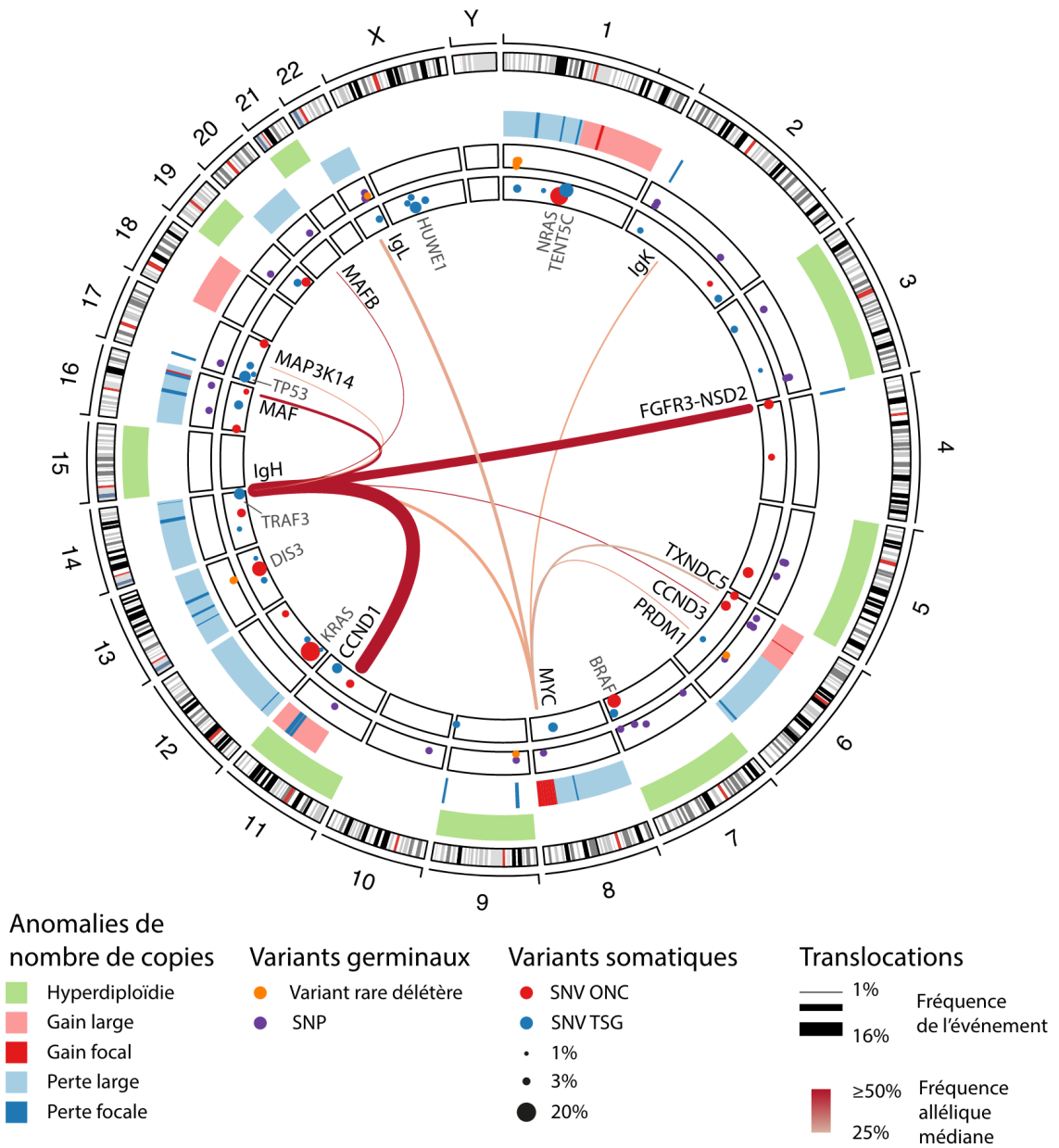


FIGURE 1.8 – Le génome du MM au diagnostic, figure tirée du manuscrit de thèse de Jean-Baptiste Alberge (ALBERGE 2021)

Les évènements secondaires Dans le MM comme dans d'autres cancers, des évènements génomiques secondaires sont observés, ceux-ci s'accumulent au fur et à mesure de la maladie. En effet, le génome tumoral est généralement instable voire très instable. Cette instabilité est directement issue de la prolifération massive, rapide et incontrôlée des cellules tumorales. Une part de cette instabilité peut également être imputée à l'expression d'oncogènes. Les cellules cancéreuses subissent donc des modifications aléatoires de leur génome. Les modifications délétères acquises par une cellule conduiront à sa disparition. Tandis que les cellules ayant acquis une modification qui leur confère un avantage sélectif seront à l'origine de l'apparition d'un nouveau clone tumoral. Le principal moteur de cette sélection est la quantité limitée de ressources à la disposition des cellules tumorales. Par ailleurs, un processus similaire est provoqué par les traitements, la pression de sélection conduira à l'apparition de nouveaux clones tumoraux résistants (voir section 1.2.4).

Les principaux évènements secondaires sont des mutations, des délétions ou amplifications de certains loci, de bras de chromosomes ou de chromosomes complets mais aussi des translocations.

Délétions du chromosome 13 L'anomalie génétique secondaire la plus courante dans le MM est la délétion du chromosome 13 (del13). Elle est présente chez 50% des patients au diagnostic. Cette anomalie est plus fortement associée aux t(4;14) et aux t(14;16) qu'aux autres événements primaires. La del13 était considérée comme de mauvais pronostic. Cependant, l'arrivée des inhibiteurs du protéasome (IPs) a permis de diminuer son impact pronostique. Cette délétion permet à la tumeur d'éliminer le gène *RB1* impliqué dans le blocage du cycle cellulaire. Les gènes *DIS3* et *BRCA2* sont également présents dans cette région et leur délétion pourrait apporter un avantage aux plasmocytes malins (BARWICK et al. 2019).

Amplification du 1q L'amplification du 1q est présente chez 40% des patients et est associée à un mauvais pronostic. L'une des signatures de risque du MM comporte d'ailleurs de nombreux gènes localisés sur le 1q. Le pronostic est d'autant plus mauvais que le nombre de copie du 1q augmente. Cette partie du génome comprend notamment *CKS1B* dont une des fonctions est de faciliter la dégradation de CDKN1B, facilitant indirectement la progression du cycle cellulaire. La présence de cette anomalie est significativement associée à la présence de la del13, ce qui est surprenant puisque ces deux anomalies semblent avoir la même cible : le cycle cellulaire. Il est possible qu'un effet de synergie

explique cette double atteinte à la régulation du cycle cellulaire. Cependant, un des autres intérêts potentiels du gain de copies 1q est la présence du gène *MCL1* (KLENER et al. 2021 ; NERI et al. 2019). Ce gène appartient à la famille de *BCL2* (voir section 1.2.4.2), dont les membres sont impliqués dans la régulation du processus apoptotique. Au sein de cette famille, *MCL1* a un rôle anti-apoptotique (BARWICK et al. 2019).

Délétion du 1p Le 1p du génome subit une délétion chez 20 à 25% des patients. Cette région comporte notamment les gènes *CDKN2C* et *FAM46C*. *FAM46C* est impliqué dans le mécanisme d'UPR (voir section 1.1.4.2) et la mort cellulaire. Le gène *FAM46C* est également altéré de façon récurrente dans le MM, ce qui renforce l'idée que c'est l'une des principales cibles de cette délétion (BARWICK et al. 2019).

MYC Les altérations de l'oncogène *MYC* sont très courantes dans le MM. Ce sont généralement des translocations accompagnées de réarrangements complexes et de gains de copie des loci impliqués. Parmi les partenaires de translocation de *MYC*, on retrouve les loci des IGH et des IGL dans 25% des cas. Du point de vue des translocations impliquant les IGs, *MYC* représente 40% des partenaires. Son rôle dans le MM est donc majeur. Ces altérations secondaires de *MYC* sont généralement sous-clonales et présentes chez un nombre croissant de patients au cours de la progression de la maladie. Bien que les altérations de *MYC* ne soient pas à l'origine de la pathologie, elles jouent un rôle important pour la progression du MM (BARWICK et al. 2019).

Délétion du 17p La délétion du 17p (del17p) est un marqueur de mauvais pronostic et, comme les altérations de *MYC*, elle est plus fréquente aux stades avancés de la maladie. Elle est par ailleurs fortement associée aux formes extra-médullaires du MM, ce qui suggère un rôle clé dans la dépendance de la tumeur à l'environnement médullaire. La principale cible de cette délétion est *TP53*. Des altérations de ce puissant gène suppresseur de tumeur sont décrites dans tous types de cancer. La perte de copie de *TP53* est très régulièrement associée à des mutations de la copie restante conduisant à une perte de fonction complète du gène. Dans ce cas, le pronostic est très mauvais pour le patient (WALKER, MAVROMMATIS et al. 2019).

Altérations des signalisations NF- κ B et RAS Dans le cas de la voie de signalisation NF- κ B, des altérations activatrices et inhibitrices des différents gènes impliqués sont observées dans le MM. La plupart du temps, les altérations favorisent la signalisation

non-conventionnelle de NF- κ B. L'activation de cette voie fournit un avantage de survie et de prolifération au MM. De façon similaire, des altérations activatrices des gènes de la signalisation RAS sont fréquemment observées, les plus touchés sont *KRAS* et *NRAS* (WALKER, BOYLE et al. 2015).

Malgré l'importance des altérations du génome pour le développement et la progression du MM, elles n'expliquent pas, à elles seules, la progression ou non des patients vers les stades symptomatiques. En effet, aux stades MGUS et SMM, des profils d'altérations similaires vont conduire à la progression vers un MM pour certains patients tandis que chez d'autres la maladie reste sous contrôle. Ce constat permet de supposer qu'il existe des facteurs extrinsèques qui participent à la progression vers les stades symptomatiques.

1.2.4 Prise en charge du MM

La prise en charge et le traitement des patients atteints de MM est une thématique complexe. En effet, la gamme de traitements disponibles est relativement large et diversifiée. De plus, les choix thérapeutiques dépendent de nombreux facteurs tels que l'âge, l'état de la fonction rénale, le type d'altération génomique primaire, la présence d'altérations génomiques de haut risque mais aussi de l'état de santé global du patient (RAJKUMAR et S. KUMAR 2020 ; MAPLES, SCOTT et LONIAL 2023). Sur la base de ces critères, le MM est généralement traité avec une combinaison de médicaments. Ici, nous n'exposerons pas en détail l'ensemble de l'algorithme de décision thérapeutique et des traitements disponibles, mais nous en aborderons les principaux aspects.

1.2.4.1 Le traitement du MM

Le traitement du Myélome Multiple nouvellement diagnostiqué (NDMM)
Pour les NDMM, la prise en charge se fait sous la forme d'une séquence de traitements. Les différentes phases de cette séquence sont l'induction, la consolidation et la maintenance. Les thérapies d'induction sont en général composées de trois types d'agents thérapeutiques : un IP, un immunomodulateur (IMiD) et un glucocorticoïde. En plus de ce triplet, des essais cliniques récents ont permis d'établir la pertinence de l'ajout d'un anti-CD38, le daratumumab à ce cocktail (MOREAU, ATTAL et al. 2019 ; MAPLES, SCOTT et LONIAL 2023).

La consolidation a pour objectif de renforcer la réponse obtenue lors de la phase

d'induction. C'est un traitement de courte durée et à forte dose. Il s'agit généralement de la même combinaison de traitement que celle utilisée pour l'induction mais cela peut aussi être une chimiothérapie (melphalan notamment) (MAPLES, SCOTT et LONIAL 2023). Ce traitement est suivi d'une auto-greffe de cellules souches hématopoïétiques (autologous stem cell transplantation en anglais) (ASCT) pour les patients éligibles. Le principal critère déterminant de l'éligibilité à l'ASCT est l'âge. L'ASCT est recommandée chez les patients de 65 ans ou moins, mais aussi chez les patients en bonne forme jusqu'à 70 ans maximum (BAZARBACHI et al. 2022).

Enfin, la maintenance a pour but de prolonger le plus possible la réponse au traitement après l'ASCT. Cependant, le choix du type de traitement à utiliser dans ce but reste controversé. La maintenance avec un IMiD (le lenalidomide) est la plus répandue, dans le cas des MM à haut-risque il est recommandé d'y ajouter un IP : bortezomib ou carfilzomib (MAPLES, SCOTT et LONIAL 2023). L'essai CASSIOPEIA a également montré l'intérêt de l'utilisation du daratumumab pour la maintenance (MOREAU, ATTAL et al. 2019; MAPLES, SCOTT et LONIAL 2023).

Le traitement du Myélome Multiple réfractaire et en rechute (RRMM) Pour le moment, il n'existe pas de standard pour le traitement des RRMM. Cependant, on considère deux groupes de RRMM : les RRMM précoces ou Early-RRMM (ERRMM) et les rechutes tardives ou Late-RRMM (LRRMM). Ces groupes distinguent les patients ayant reçu trois lignes de traitement ou moins, des patients en ayant reçu quatre ou plus. Pour les ERRMM, le traitement est généralement composé d'un anticorps monoclonal anti-CD38 couplé à un IP et un IMiD. Lorsque les patients atteignent le stade LRRMM, leur pronostic est généralement mauvais. En effet, ils répondent mal aux lignes de traitement conventionnelles et à ce stade il est nécessaire de mettre en place de nouvelles approches thérapeutiques (MAPLES, SCOTT et LONIAL 2023). L'introduction des nouvelles immunothérapies ciblées, notamment dirigées contre BCMA, donne de bons résultats. Parmi ces nouvelles approches, on retrouve les cellules CAR T, les anticorps bis-spécifiques mais aussi des anticorps couplés à des molécules toxiques pour les plasmocytes (MAPLES, SCOTT et LONIAL 2023). Malgré des résultats encourageants, ces traitements ne sont pas efficaces pour l'ensemble des patients. De plus, chez la plupart des patients qui répondent, on finit par observer l'apparition de résistances (MOREAU et TOUZEAU 2022; MOREAU, GARFALL et al. 2022; DERRIEN, GASTINEAU et al. 2023). Le MM reste donc une maladie incurable et les rechutes successives semblent inévitables pour la majorité

des patients.

1.2.4.2 Les grandes classes d'agents thérapeutiques utilisés dans le MM

Dans le MM, différentes classes de médicaments sont utilisées. Chacune d'entre elles cible la tumeur via des mécanismes différents. La susceptibilité de la tumeur à ces traitements est généralement issue de ses caractéristiques plasmocytaires. Nous détaillerons ici une partie des mécanismes d'action de ces traitements. En effet, la compréhension des modes d'action des traitements est un facteur essentiel pour une meilleure prise en charge des patients. Elle permet notamment de mieux prédire et anticiper la réponse ou la résistance des patients.

Les glucocorticoïdes Les glucocorticoïdes sont des hormones stéroïdes sécrétées par les glandes adrénales principalement en réponse à un stress de l'organisme. Ils sont impliqués dans la régulation de divers processus métaboliques mais aussi immunitaires. Ils agissent sur de nombreux types cellulaires différents avec un effet dépendant du type cellulaire considéré. Dans le cas du SI, ils ont une action anti-inflammatoire et immunosuppressive. La dexaméthasone (dex) est un glucocorticoïde de synthèse utilisé dans de nombreuses combinaisons de traitements du MM.

Mode d'action des glucocorticoïdes Le récepteur aux glucocorticoïdes (GR) est codé par le gène *NR3C1*. C'est une protéine cytoplasmique, ce récepteur est donc intracellulaire et son activation nécessite l'entrée du ligand dans la cellule. Lorsque le GR rencontre son ligand, cela provoque son homodimérisation. L'activation du GR requiert donc deux GR et deux ligands. Une fois homodimérisé, le GR se relocalise dans le noyau, où il interagit avec l'ADN et joue un rôle de facteur de transcription. Son action régule ainsi l'expression de nombreux gènes, par différents modes d'action. Il peut agir par fixation directe à l'ADN sur un motif GR-responsive element (GRE) reconnu par l'un des domaines du GR (figure 1.9). Le GR se lie principalement à l'ADN au niveau d'enhancers distaux (REDDY et al. 2009), et permet la formation de boucles chromatiniennes entre ces régions enhancer et le promoteur du gène cible. Le GR agit aussi de façon indirecte par liaison à des facteurs de transcription déjà en place sur l'ADN. Ce type de régulation n'implique pas nécessairement la présence d'un motif GRE. Les facteurs de transcription impliqués dans ce type d'interactions sont notamment NF- κ B et AP-1 (RATMAN et al. 2013). La régulation de l'expression par le GR est la plupart du temps positive mais il peut aussi

provoquer une répression de certains gènes.

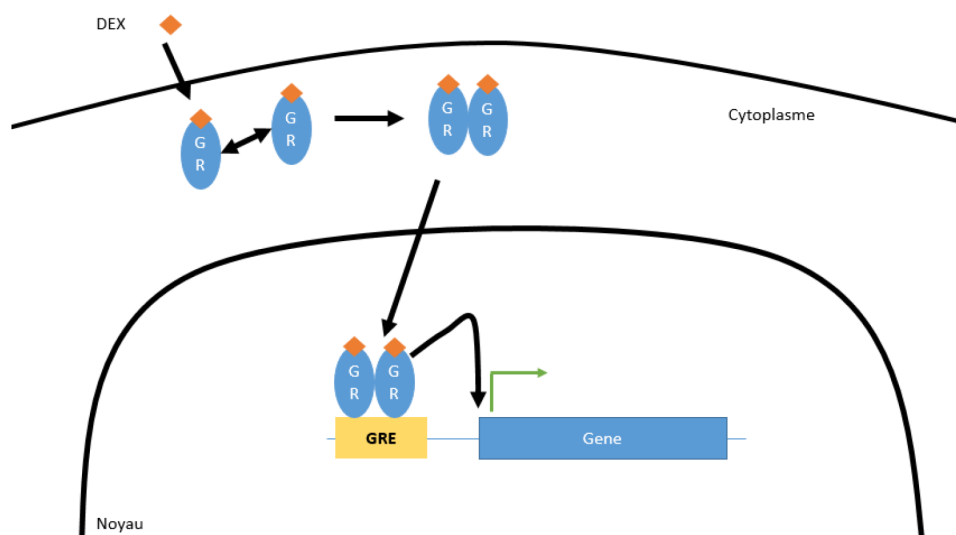


FIGURE 1.9 – Mécanismes d’action moléculaire du récepteur au glucocorticoïde sur la transcription des gènes, figure issue de la thèse de Victor Gaborit (GABORIT 2021)

Les IMiDs Les IMiDs sont un groupe de molécules dérivées de l’acide glutamique. Bien que leur mode d’action reste mal connu, leur cible principale est le céréblon (*CRBN*) (ZHU, KORTUEM et STEWART 2013). Les propriétés anti-prolifératives, anti-angiogéniques, anti-inflammatoires et de régulation des lymphocytes T et NK des IMiDs en font un traitement intéressant dans le cas du MM et du cancer en général. Le premier des IMiDs à avoir été découvert est le thalidomide. Des analogues ont été développés dans le but d’améliorer l’efficacité du traitement et de réduire sa toxicité. La plupart des effets du thalidomide sont exacerbés lors de l’utilisation de ses analogues : le lenalidomide et le pomalidomide. Seul l’effet anti-angiogénique est minoré. En plus de ce gain d’efficacité, ces analogues ont des effets supplémentaires. Ils inhibent la prolifération des Treg et augmentent la cytotoxicité des lymphocytes NK, notamment lorsqu’ils sont activés via une IG liée à son antigène (QUACH et al. 2010). Cet aspect est particulièrement intéressant dans le cadre de combinaisons des IMiDs avec des anticorps ciblant des antigènes tumoraux (voir section 1.2.4.2).

Les IPs Les IPs permettent de bloquer le processus de dégradation des protéines et notamment les protéines impliquées dans la régulation du cycle cellulaire. Cela a pour

effet d'induire l'expression de protéines pro-apoptotiques de la famille BCL2 et donc de provoquer la mort cellulaire. Du fait de leur forte prolifération, les cellules tumorales y sont donc particulièrement sensibles. Les plasmocytes produisent et sécrètent de grandes quantités d'IGs. Le blocage de la fonction de dégradation et de recyclage des protéines provoque donc une accumulation rapide de protéines mal repliées dans le RE. Le stress du RE déclenche alors l'UPR et conduit les plasmocytes à l'apoptose. Les IPs bloquent aussi l'activation de NF- κ B dont le rôle est essentiel pour la progression du MM (voir section 1.2.3.2). Le bortezomib est le premier inhibiteur du protéasome à avoir été utilisé dans le MM. Il est particulièrement prescrit dans les cas de MM avec insuffisance rénale. Le bortezomib agit également sur le MM en inhibant la production de cytokines (IL-6, VEGF, ...) nécessaires à la survie des plasmocytes, par les cellules stromales de la moelle. Le carfilzomib et l'ixasomib sont prescrits à des patients avec un RRMM. Il est à noter que les trois IPs utilisés dans le MM ont des modes d'action moléculaires relativement différents et donc des effets secondaires spécifiques. Le carfilzomib est notamment connu pour sa toxicité cardiaque (ITO 2020).

Les immunothérapies Les immunothérapies regroupent les traitements qui ont pour but de pousser le SI du patient à s'en prendre à la tumeur. Elles ciblent des protéines de surface exprimées par les tumeurs. Pour être efficaces, ces traitements doivent cibler des protéines répondant à certains critères. Dans un premier temps, il ne faut pas cibler un marqueur présent dans d'autres tissus normaux afin d'éviter de provoquer des défaillances d'organes sains. De plus, il faut cibler des protéines qui apportent un avantage pour la survie de la tumeur. En effet, la baisse ou l'arrêt de l'expression de certaines protéines est un mécanisme de résistance très courant dans les cancers (DONK et ZWEEGMAN 2023; MAPLES, SCOTT et LONIAL 2023). Si la protéine présente un avantage pour la tumeur, cela limite les chances de voir apparaître une résistance au traitement par perte d'expression.

Les anticorps monoclonaux Les anticorps monoclonaux sont produits artificiellement et sont dirigés contre une cible unique présente à la surface de la cellule tumorale. Dans le MM, le plus utilisé est le daratumumab, dirigé contre le CD38.

Les CAR T Les cellules CAR T sont des lymphocytes T modifiés génétiquement pour exprimer un récepteur CAR (Récepteur Chimérique spécifique d'un Antigène). Ce récepteur est dirigé contre un antigène présent à la surface de la tumeur et permet au

lymphocyte de s'attaquer à la cellule sans que son TCR ne lui permette de la cibler. Ce type de thérapie montre de bons résultats mais présente de nombreux désavantages. En effet, ce sont les lymphocytes T du patient qui doivent être édités afin qu'ils soient compatibles avec son SI et ne provoquent pas de réaction immunitaire. L'utilisation des CAR T provoque des effets secondaires non négligeables et notamment des atteintes neurologiques. Entre le moment où le prélèvement des lymphocytes est effectué et celui de la réinjection des CAR T, plusieurs semaines sont généralement nécessaires. Ces délais sont imposés par la logistique, le temps d'édition ainsi que les contrôles qualité des cellules obtenues. De plus, ces problématiques impactent fortement le coût du traitement. La faisabilité est donc limitée pour un passage à grande échelle de ce type de traitements (DONK et ZWEEGMAN 2023).

Les anticorps bispécifiques Les anticorps possèdent deux sites de liaison à l'antigène. En temps normal, ces deux sites sont identiques. Les anticorps bispécifiques sont élaborés afin que ces deux sites puissent se lier à des épitopes différents. Les anticorps bispecifiques sont des anticorps monoclonaux qui permettent de rediriger les lymphocytes T, en ciblant le CD3 ainsi qu'une molécule présente à la surface de la cellule tumorale (figure 1.10). Le concept de ce type de traitement est donc très similaire à celui des CAR T, avec de grands avantages logistiques puisque les lymphocytes sont redirigés vers la tumeur sans nécessiter de prélèvement ni d'édition de leur génome. De plus, les anticorps ne nécessitent pas de modifications spécifiques à chaque patient, et sont donc disponibles "off-the-shelf". Cet aspect est crucial dans le cas des patients dont la maladie progresse vite (MOREAU et TOUZEAU 2022 ; MAPLES, SCOTT et LONIAL 2023). Ce type de traitement provoque aussi moins de complications neurologiques, comparé au CAR T. Plusieurs d'entre eux ont montré de bons résultats dans le traitement des MM triple réfractaires : aux IPs, aux IMiDs et aux anticorps anti-CD38 (DONK et ZWEEGMAN 2023). Parmi les anticorps bispécifiques en phase d'essais cliniques dans le MM, on retrouve l'Elranatamab et le Teclistamab qui ciblent tous deux BCMA (*TNFRSF17*). Ce récepteur est très spécifique des plasmocytes ce qui en fait une cible idéale. Les premiers résultats des essais cliniques du Teclistamab montrent d'ailleurs de bons résultats (MOREAU, GARFALL et al. 2022). Le Talquetamab, quant à lui, cible GPRC5D. Il donne également de bons résultats. Cependant, son expression n'est pas complètement exclusive aux plasmocytes. Le traitement provoque donc des effets secondaires liés à l'expression de *GPRC5D* par certaines cellules des bulbes capillaires et des tissus producteurs de kératine.

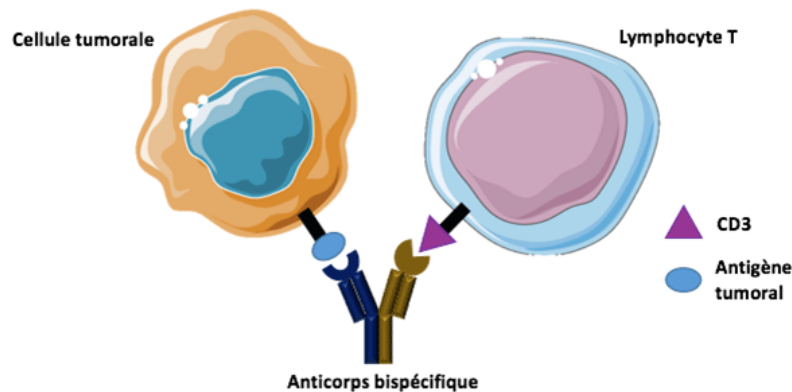


FIGURE 1.10 – Schéma simplifié de la structure et du mode d'action des anticorps bispécifiques

Les BH3 mimétiques La famille BCL2 est un groupe de protéines chargées de la régulation de l'apoptose. Les membres de cette famille sont regroupés en trois catégories basées sur leurs fonctions : les protéines anti-apoptotiques, les pro-apoptotiques multi-domaines (effecteurs) et les pro-apoptotiques "BH3-only" (figure 1.11). Les pro-apoptotiques effecteurs sont principalement BAX et BAK. Lorsqu'ils sont activés, ils s'assemblent à la surface de la mitochondrie où ils forment des pores et perméabilisent la membrane externe, ce qui déclenche le relargage de diverses protéines qui provoqueront l'apoptose par l'intermédiaire des caspases notamment. Ces effecteurs sont activés par des pro-apoptotiques BH3-only comme PUMA ou BIM. Le rôle des anti-apoptotiques tels que BCL2, MCL1 ou encore BCLxL (respectivement encodés par les gènes *BCL2*, *MCL1* et *BCL2L1*) est de séquestrer ces activateurs afin de les empêcher d'agir. Certains des BH3-only n'ont pas de rôle activateur sur BAX et BAK mais peuvent se lier aux anti-apoptotiques avec une forte affinité, permettant la libération des activateurs de moindre affinité. En temps normal, les BH3-only sont induits ou activés par un stress cellulaire. Ce stress cellulaire peut notamment être déclenché indirectement par les traitements (KLENER et al. 2021). Les cellules qui résistent à l'apoptose par l'expression des anti-apoptotiques, malgré l'expression des pro-apoptotiques déclenchée par ce stress, sont dites : "primed for death" ou prêtes à mourir. Dans le cas du cancer et du MM notamment, l'expression ou la sur-expression des anti-apoptotiques de cette famille est un facteur important pour la survie des tumeurs (BARWICK et al. 2019 ; KLENER et al. 2021). MCL1 est notamment exprimé par les plasmocytes (Voir section 1.1.4.2), et son expression peut également se trouver

renforcée dans le cas des MM avec une amplification du 1q (Voir section 1.2.3.2 et NERI et al. 2019). Par ailleurs dans le MM, la sur-expression de BCL2 est associée à la translocation t(11;14) (S. KUMAR et al. 2017; SLOMP et PEPPERZAK 2018; KITADATE et al. 2021).

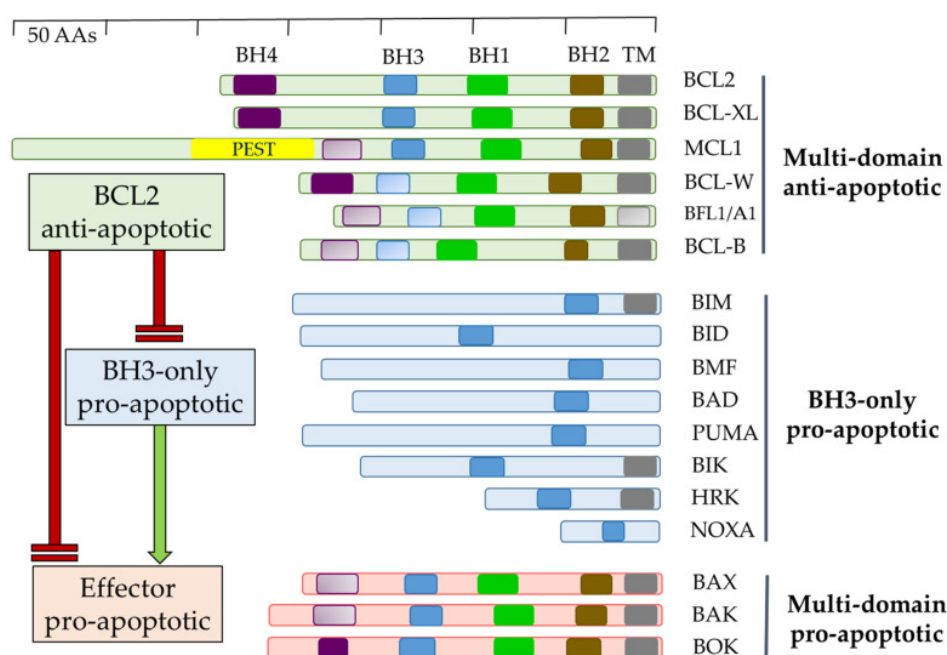


FIGURE 1.11 – Les membres de la famille BCL2, figure issue de KLENER et al. 2021

Les membres de la famille BCL2 sont donc des cibles de choix pour le traitement du MM. Le venetoclax est un inhibiteur spécifique de BCL2. Il est le premier à avoir été disponible en clinique dans divers cancers hématologiques (ROBERTS et al. 2016; DAVIDS et al. 2017). Dans le MM, il est principalement utilisé pour traiter les patients en ERRMM atteints d'une t(11;14) (S. KUMAR et al. 2017; MAPLES, SCOTT et LONIAL 2023).

1.2.4.3 La résistance du MM et ses mécanismes

Malgré cet arsenal thérapeutique important et les innovations récentes, la rechute et l'apparition de résistances sont très courantes dans le MM. La résistance est le résultat d'un changement survenu au sein de la tumeur et/ou son environnement. Ces changements affectent généralement les mécanismes biologiques liés à l'action du traitement utilisé. Parmi les résistances, certaines sont aspécifiques du traitement, c'est le cas notamment lorsque l'on observe des primo-résistances. Ces dernières touchent généralement

des mécanismes ubiquitaires du cancer, qui apportent un avantage global à la tumeur.

La résistance peut avoir des causes génétiques. On observe alors l'apparition d'un ou plusieurs clones tumoraux ayant acquis une délétion ou des mutations de gènes cibles du traitement. Dans le cas des anticorps bispécifiques, on peut notamment observer des inactivations bi-alléliques de la cible. Des altérations inactivatrices ciblées sont notamment sélectionnées dans les cas de délétions pré-existantes du second allèle du gène (MOREAU et TOUZEAU 2022; DERRIEN, GASTINEAU et al. 2023). Dans le cas du venetoclax, des résistances liées à des gains de copie du 1q et des mutations de *BCL2* ont été rapportées (NERI et al. 2019). Du côté des IMiDs, le principal mécanisme de résistance est la perte de copies et les mutations inactivantes de *CRBN* (GOODING et al. 2021). Le MM présente également une forte instabilité épigénétique (DERRIEN, GUÉRIN-CHARBONNEL et al. 2021). Cette instabilité peut notamment induire des pertes d'accessibilité de la chromatine au niveau du gène cible et/ou de ses enhanceurs, ce qui conduit à une diminution voire une perte de l'expression (DERRIEN, GASTINEAU et al. 2023). L'inactivation épigénétique de *RASD1* a été observée dans des MM résistants à la dex (ROBAK et al. 2018). Les changements du transcriptome peuvent aussi induire une résistance, la sur-expression de *BCL2L1* notamment permet l'échappement au venetoclax (BLOMBERG et al. 2019; ALBERGE 2021). Les mécanismes qui permettent cette sur-expression restent pour le moment inconnus même si ils pourraient être liés à l'instabilité épigénétique. Ce n'est pas la seule piste et d'autres signaux pourraient être impliqués. Dans le cas de la résistance aux IMiDs, des mutations de la machinerie d'épissage permettent l'expression de *CRBN* sans l'exon 10, lequel encode le domaine de liaison des IMiDs (GOODING et al. 2021). Les altérations de l'environnement immunitaire expliquent aussi une partie des résistances et plus particulièrement dans le cas des immunothérapies. Comme évoqué précédemment, l'action des anticorps bispécifiques nécessite la présence d'une population de lymphocytes effecteurs fonctionnels. Dans certains cas, ces populations présentent des caractéristiques d'exhaustion mais peuvent aussi être inhibées par d'autres types cellulaires comme les Treg ou des macrophages associés à la tumeur (MOREAU et TOUZEAU 2022). La liaison des plasmocytes aux cellules stromales leur apporte également un effet protecteur vis-à-vis de la dex (ROBAK et al. 2018).

Le rôle de l'environnement est essentiel pour certains traitements. De plus, la majorité des traitements n'a pas un effet neutre sur lui. Comme nous avons pu l'évoquer dans les sections précédentes, certains agents permettent une stimulation de l'environnement immunitaire ou une levée de l'immunosuppression, en plus de leur effet sur la tumeur

(QUACH et al. 2010; ITO 2020; RUBIO, DHUYSER et NGUYEN 2021). Il est donc nécessaire de tenir compte des modifications qui sont apportées à l'environnement par les traitements précédents afin d'assurer l'efficacité de la thérapie, mais cela permet aussi d'envisager de nouvelles combinaisons de traitements. Les résistances impliquent souvent des modifications de certaines voies de signalisations ou mécanismes, ce qui implique des vulnérabilités potentielles à d'autres traitements. Ce constat permet d'envisager des effets de synergie entre les traitements.

PROBLÉMATIQUE ET TRAVAUX DE THÈSE

La prise en charge et le traitement du MM ont connu des progrès importants au cours de la dernière décennie, permettant d'augmenter significativement l'espérance de vie des patients. Malgré tout, le MM reste une pathologie incurable et la rechute est inévitable pour la majorité des patients. Il est donc essentiel de poursuivre l'effort de recherche afin de mieux comprendre comment et pourquoi les patients rechutent.

Les causes de résistance et de progression du MM sont très diverses. Elles sont le résultat de la pression de sélection imposée aux cellules tumorales par leur environnement et par les traitements. Les mécanismes de la résistance impliquent notamment l'acquisition de modifications génétiques, épigénétiques, transcriptomiques mais aussi métaboliques par la cellule tumorale. Bien que ces modifications s'accumulent au cours de la maladie et soient plus nombreuses aux stades ERRMM puis LRRMM, elles ne suffisent pas à expliquer l'ensemble des rechutes et progressions de la maladie. En effet, l'écosystème tumoral et notamment l'environnement immunitaire évolue lui aussi au cours de la pathologie. Les traitements agissent non seulement sur la tumeur mais aussi sur son environnement. Cela est d'autant plus vrai dans le cas des traitements de dernière génération tels que les immunothérapies, puisque leur mode d'action s'appuie directement sur l'environnement. Les résistances spécifiques des traitements dépendent généralement du mode d'action du ou des agents thérapeutiques utilisés. Pour élucider ces mécanismes de résistance, il est donc essentiel de mieux comprendre comment fonctionnent les traitements. Une meilleure compréhension des évolutions subies par l'écosystème tumoral est également essentielle. Ces connaissances permettront de mieux comprendre mais aussi prédire et éventuellement prévenir l'apparition de résistance et la rechute.

L'amélioration et le développement de nouvelles méthodes de séquençage ainsi que l'essor des outils bioinformatiques permettent aujourd'hui d'accéder à de nouvelles informations tant au niveau génomique, épigénomique que transcriptomique à des échelles de résolutions de plus en plus fines. La révolution provoquée par l'apparition des technologies en cellule unique a permis de nombreuses avancées, mettant en évidence la diversité et la

variabilité présentes au sein des populations de cellules. Dans le champ de la recherche fondamentale et notamment la biologie du développement, cela a permis de capturer et de caractériser les stades transitoires de la différenciation (MEISTERMANN et al. 2021). Ces technologies ont également permis la découverte et la caractérisation de nouveaux sous-types cellulaires (VILLANI et al. 2017).

Dans cette thèse, nous présentons des travaux qui s'appuient sur les technologies de séquençage en cellule unique pour apporter un éclairage sur le MM et ses problématiques. Nous aborderons notamment deux aspects du MM assez éloignés.

Dans un premier, nous proposons d'explorer le mode d'action moléculaire de la dex. En effet, ce glucocorticoïde est l'un des agents thérapeutiques les plus utilisés dans le MM et il est testé en combinaison avec les traitements les plus récents. Nous avons donc travaillé sur les conséquences de ce médicament sur le transcriptome et l'accessibilité de la chromatine des cellules tumorales, à l'échelle de la population mais aussi pour caractériser la variabilité de ses effets entre les cellules. Dans un second temps, nous proposons une caractérisation détaillée de l'écosystème tumoral à l'échelle de la cellule unique. Dans ce but, nous avons analysé l'abondance, les phénotypes et les liens avec la tumeur des différents types cellulaires de l'environnement immunitaire de la moelle de patients. Ces échantillons sont issus de patients à tous les stades du MM, du NDMM au LRRMM. L'objectif de cette étude est de caractériser les plasmocytes et leur environnement immunitaire aux différents stades de progression du MM. Afin de montrer les évolutions qualitatives et quantitatives de toutes les cellules de la moelle qui composent l'écosystème tumoral propre au MM chez des patients avant traitement, des patients en première rechute (ERRMM) et des patients plus avancés dans la maladie (LRRMM).

MÉTHODES

3.1 Traitement et préparation des échantillons

Cette section présente les principales étapes de production des données analysées au cours de mes travaux. Ces étapes impliquent de nombreux acteurs de la recherche clinique dont les cliniciens, les ARC (attachés de recherche clinique), les chercheurs, les ingénieurs, les techniciens et les étudiants de l'équipe.

3.1.1 Prélèvement des échantillons

Les échantillons de moelle analysés en cellules uniques proviennent principalement du CHU de Nantes. Ils ont été prélevés sur des patients atteints de MM enrôlés dans l'étude MYRACLE (BENANIBA et al. 2019). Les échantillons de MM analysés dans notre étude sont issus de ponction de moelle, bien que des échantillons de sang et d'épanchement pleural aient également été réalisés. Ce choix a été fait dans le but de simplifier les comparaisons des échantillons en limitant les conditions au sein du jeu de données. Les prélèvements sont effectués dans le cadre des visites d'inclusion, de rechutes et de suivis des patients. Les moelles normales utilisées comme groupe contrôle (NBM : Normal Bone Marrow) sont issues de patients ayant subi une chirurgie de remplacement de hanche à la Nouvelle Clinique Nantaise. L'intérêt de ce groupe de référence est de réaliser une comparaison avec des patients d'un âge similaire aux patients atteints de MM.

3.1.2 Au laboratoire

3.1.2.1 Traitement des échantillons et séquençage

Une fois prélevés, les échantillons sont immédiatement pris en charge par les équipes de recherche. Tous les échantillons sauf trois sont passés par une étape de purification des cellules mononucléées. Cette étape est réalisée par centrifugation en présence d'un ficoll.

Lors de la sédimentation, les cellules mononucléées se retrouvent séparées des érythrocytes et des granulocytes et sont ensuite récupérées.

Les cellules sont ensuite mises en culture avec de l'IL-6 afin de permettre leur survie en condition *in vitro*. La moitié des cellules mises en culture a été exposée à un cocktail de BH3 mimétiques, anti-BCL2 et anti-MCL1 afin de tester leur sensibilité à ces traitements. À l'issue de la nuit en culture, la viabilité des cellules est vérifiée. Elles sont ensuite préparées pour passer au chromium. Le chromium est un équipement produit par la société 10X Genomics (ZHENG et al. 2017). Cette technologie utilise un système de micro-fluidique pour isoler les cellules avec une bille portant un code barre moléculaire. Lors de la lyse de la cellule, les ARN sont capturés et le code barre leur est apposé. Une partie de ce code barre est identique pour toutes les molécules de la bille, cette partie permettra d'identifier la cellule d'origine de la molécule d'ARN. Une autre partie du code barre est l'identifiant moléculaire unique (UMI). L'UMI permet d'identifier chaque ARN de façon unique. Les bibliothèques ARNs sont ensuite séquencées sur le NovaSeq 6000.

3.2 Traitement des données de séquençage et contrôles qualité des profils d'expression

3.2.1 Alignement des reads et création de la table de comptage

Une fois le séquençage terminé, on obtient des fichiers BCL. Ces fichiers doivent être convertis en FASTQ. Pour cela, on utilise la commande `mkfastq` du logiciel `cellranger` fourni par 10X genomics. Les fichiers FASTQ sont des fichiers texte contenant l'ensemble des séquences ARN lues par le séquenceur et reconstituées par `mkfastq`. Ces séquences sont ensuite alignées sur le génome de références afin d'identifier le gène correspondant à chaque read. La technologie utilisée capture spécifiquement l'extrémité 3' des ARNs, l'information de leurs séquences est donc très partielle. Les travaux sur ces séquences sont donc très limités. Cependant, l'objectif de la technologie est d'identifier les gènes exprimés par chaque cellule. L'extrémité 3' suffit pour accomplir cette tâche. À partir des gènes identifiés et des codes barre il est alors possible de construire une matrice d'expression. Chaque cellule est une ligne de la matrice et chaque colonne est un gène. C'est à partir de cette matrice que l'ensemble de nos analyses en cellule unique sont réalisées.

3.2.2 Contamination par les ARN ambiants

3.2.2.1 Principe de la contamination

Au cours de l'étape d'encapsulation des cellules au sein des gouttelettes permettant de les isoler avec un bille "barcodée", les ARN libres dans le milieu sont capturés. Ces ARN ambiants proviennent de cellules ayant été détruites au cours du processus de traitement de l'échantillon, en amont de l'encapsulation ou au cours de celle-ci. Ces ARN ambiants sont alors considérés comme appartenant au transcriptome de la cellule capturée. Ce faisant, ils viennent contaminer son profil d'expression, introduisant un biais inter-échantillon puisque la contamination au sein de chaque cellule est dépendante des autres cellules qui composent l'échantillon. Ce point est particulièrement important dans le cas du MM et notamment avec notre design expérimental. En effet, nos échantillons contiennent à la fois des cellules tumorales et l'environnement immunitaire. La variabilité du ratio tumeur/environnement au sein des échantillons ainsi que les variations de composition de l'environnement conduisent donc à une contamination spécifique à chaque échantillon. Les caractéristiques transcriptomiques des plasmocytes tumoraux sont également un facteur important. Ces cellules sont (en général) plutôt de grosses cellules et contiennent beaucoup d'ARN, ce qui par conséquent conduit à une contamination plus importante lorsqu'elles sont détruites avant encapsulation. Une seconde particularité est importante dans ce contexte : l'expression des gènes des immunoglobulines. Ceux-ci représentent jusqu'à 80-90% du transcriptome d'un plasmocyte tumoral, avec une variabilité importante. De plus, chez chaque patient, une combinaison unique de chaîne lourde et légère des immunoglobulines est exprimée. Cette combinaison unique et fortement exprimée introduit donc une contamination échantillon/patient spécifique dans le transcriptome des cellules de l'environnement immunitaire.

3.2.2.2 SoupX et son fonctionnement

La nécessité de prendre en compte cette contamination dépend du design expérimental. En effet, dans le cas d'échantillons composés d'un type cellulaire unique et relativement homogène, il est difficile et peu utile d'en tenir compte. Dans le cas de nos échantillons de moelle, c'est une étape essentielle. Pour décontaminer nos échantillons, nous utilisons le package R SoupX (YOUNG et BEHJATI 2018). Cette méthode procède en 3 étapes :

1. Estimation du profil transcriptomique de la contamination à partir des gouttelettes vides

2. Estimation du taux de contamination de chaque cellule
3. Correction de l'expression des cellules à partir du profil d'expression de l'ARN ambiant et du taux de contamination estimé

Alternatives à SoupX Contrairement à d'autres aspects du traitement des données de séquençage de l'ARN en cellule unique (scRNA-seq), il existe peu de méthodes consacrées à la contamination ambiante des profils d'expression. Il existe notamment Cell Bender, basé sur un modèle génératif profond et DecontX qui utilise une méthode bayésienne. Ces différentes méthodes comportent des limitations. Pour Cell Bender, l'utilisation d'un modèle génératif implique des coûts de calcul très importants. Du côté de DecontX, il est nécessaire de fournir une annotation préalable des clusters de types cellulaires de l'échantillon. De notre côté, nous avons choisi d'utiliser SoupX pour diverses raisons. Premièrement, cet outil était le seul disponible au moment de la mise en place de notre pipeline d'analyse. Deuxièmement, cet outil est simple d'utilisation et peu coûteux en temps de calcul. Enfin, nous avons retenu cette méthode car nous étions satisfaits des résultats obtenus. Nous souhaitons éviter de revenir sur une de ses premières étapes afin d'éviter de reproduire et revérifier les résultats des suivantes.

3.2.3 Présences de doublets dans les données

3.2.3.1 Origine des doublets

Durant l'étape d'encapsulation, la majorité des cellules capturées dans des gouttelettes se retrouve seule avec une bille barcodée. Cependant, une partie d'entre elles se retrouve piégées à plusieurs avec une bille barcodée. C'est ce qu'on appelle un doublet ou un multiplet, dont le transcriptome sera la somme des transcriptomes des différentes cellules. Il existe deux catégories de doublets : les doublets hétérotypiques et les doublets homotypiques. Les doublets homotypiques sont composés de cellules similaires, et par conséquent ne représentent pas un réel problème. En effet, le profil transcriptomique obtenu sera fortement similaire à celui d'une cellule unique, ce qui rend le profil obtenu valide sur le plan biologique mais rend également difficile voire impossible la détection du doublet, hormis sur la base d'un nombre d'UMI correspondant au barcode associé plus élevé. Cependant, ce critère n'est pas absolu et est dépendant du type cellulaire. Il est donc difficile de l'appliquer dans le cadre d'échantillons composés de types cellulaires multiples. En effet, il serait alors nécessaire de définir au préalable les types pour mettre en place un filtre

spécifique à chaque cluster. Les doublets hétérotypiques, par opposition, sont composés de types cellulaires différents. Il est crucial de prendre en compte leur existence puisque leur inclusion dans les analyses peut mener à des interprétations erronées des résultats. Dans de nombreux cas, il est plutôt simple de les identifier puisqu'ils forment un cluster indépendant où les cellules co-expriment des marqueurs spécifiques d'autres clusters au sein de l'échantillon. Cependant, il est parfois difficile de les isoler au sein de clusters distincts ou encore de les identifier par la co-expression de gènes spécifiques de types cellulaires distincts. De plus, dans le cas où cela est possible, c'est une tâche fastidieuse qui requiert l'examen des échantillons en profondeur après avoir réalisé les étapes de réduction de dimension et de clustering de Louvain. Ces étapes devront alors être reproduites une fois les doublets écartés des données.

Il est donc plus pertinent et plus efficace d'utiliser des méthodes computationnelles pour gérer cet aspect des données, en prenant néanmoins soin de vérifier la qualité des résultats produits.

3.2.3.2 Les méthodes de détection des doublets

Les doublets sont une problématique importante du traitement des données de scRNA-seq. Il existe donc un grand nombre d'outils dédiés à leur détection.

Les différents types de méthodes

Méthodes par simulation de doublets artificiels Une majeure partie des outils est basée sur la création de doublets artificiels. Bien que les approches de création de ces doublets divergent, le principe reste le même. Les doublets générés sont alors exploités par les méthodes pour inférer un score aux cellules de l'échantillon. Ce score reflète la probabilité que chacune des cellules soit un doublet et non une cellule unique. Les outils `Scrublet` et `doubletCell` génèrent leurs doublets en faisant la somme des profils d'expression de deux gouttelettes sélectionnées aléatoirement. Ils effectuent ensuite une analyse en composantes principales (ACP) et à partir de l'espace obtenu, ils définissent un score de doublets dépendant du nombre de doublets artificiels présents parmi les plus proches voisins de chaque cellule. Comme expliqué précédemment, malgré quelques nuances sur la façon de traiter les doublets artificiels ou le calcul du score, la plupart des méthodes fonctionnent sur un principe similaire. Le package R `scds` (BAIS et KOSTKA 2019) propose quant à lui une approche plus originale. Il combine deux approches : une basée sur des

doublés artificiels (bcds) et l'autre basée sur le calcul d'un score de compatibilité entre les paires de gènes exprimés (cxds). La méthode cxds produit donc dans un premier temps une matrice de score où les paires de gènes co-exprimés moins fréquemment qu'attendu obtiennent un score élevé et les paires co-exprimées comme attendu obtiennent un score faible. On additionne alors les scores de chaque paire de gènes exprimés par les cellules pour obtenir leur score de doublet. Afin de récapituler les deux approches, les scores de bcDs et cxds sont alors normalisés puis sommés pour obtenir un score "hybride". Une fois les résultats obtenus, une partie des méthodes produit une classification binaire singleton/doublet. Ces classifications sont obtenues en déterminant un seuil à partir duquel les cellules sont considérées comme des doublets. Les résultats dépendent donc de la manière de déterminer ce seuil. Il existe 2 principales approches : regarder la distribution des scores et établir un seuil ou retirer le nombre de doublets attendus parmi les scores les plus élevés. Le nombre de doublets attendu dépend du nombre de cellules de l'échantillon et peut donc être calculé. DoubletFinder propose, en plus, d'ajuster ce score en prenant en compte la proportion de doublets homotypiques qui seront formés.

3.2.3.3 Méthodologie choisie pour notre pipeline

Étant donné la taille de notre jeu de données, il a été nécessaire de mettre en place une méthode de filtration qui ne nécessite pas d'intervention au cas par cas pour la définition des seuils de score de doublets. Nous avons fait le choix de retirer la proportion de doublets attendus de chaque échantillon. Dans un but conservatif, nous avons cependant choisi de faire une estimation volontairement basse. Cependant, pour maximiser nos chances de détecter une majorité des doublets, nous avons choisi de retirer les doublets via deux méthodes en parallèle : scds et scDbIFinder (BAIS et KOSTKA 2019 ; P.-L. GERMAIN et al. 2022). Nous avons choisi ces deux méthodes car elles figurent parmi les plus performantes en terme de qualité de détection, mais aussi car elles sont disponibles en langage R et sont peu coûteuses en temps de calcul. Ces caractéristiques nous ont donc permis de les intégrer facilement au processus de pré-traitement des données.

3.2.4 Variables de contrôle qualité

Pour garantir la pertinence des analyses et la qualité des résultats obtenus, il est également essentiel d'écartier les cellules de mauvaise qualité qui peuvent polluer les échantillons. Pour cela, on prend en compte diverses métriques calculées à partir des données,

telles que le nombre de molécules d'ARN unique des cellules (UMIs), le nombre de gènes différents détectés mais aussi le pourcentage d'UMIs correspondants à certains groupes de gènes.

3.2.4.1 Nombre d'UMIs

Pertinence de la variable Le nombre d'UMIs d'une cellule est un des principaux indicateurs de sa qualité. Il reflète la quantité d'information disponible sur le transcriptome de cette cellule. Une trop faible quantité d'UMIs par cellule implique que la quantité d'information disponible sur son transcriptome est faible, augmentant le risque d'avoir raté des gènes dont le niveau d'expression est faible. Ces cellules de mauvaise qualité vont alors perturber les analyses suivantes, en introduisant une variabilité non contrôlée issue d'effets techniques et non de phénomènes biologiques, lors de la définition de clusters correspondant aux différents types cellulaires notamment. En effet, dans le cas où plusieurs types cellulaires partagent une partie importante de leur transcriptome et se différencient par l'expression de quelques gènes spécifiques, ces cellules de mauvaise qualité vont être difficiles à classer et vont venir "flouter" les limites entre clusters et donc entre types cellulaires.

L'origine de ces cellules de mauvaise qualité peut être multiple. Cela peut être dû à un manque de profondeur lors du séquençage de la librairie mais aussi le résultat d'un problème au cours de l'expérimentation (cellule endommagée, encapsulation d'un débris cellulaire, problème de réaction au sein de la gouttelette, etc). Il est donc d'autant plus important de filtrer ces cellules lorsque l'on cherche à identifier des sous populations précises au sein de types cellulaires.

Dans certains cas, il peut être intéressant d'écarter également les cellules dont le nombre d'UMIs est anormalement élevé. Un nombre trop élevé d'UMIs peut suggérer que l'on est en présence d'un multiplet de cellules. Ce critère peut donc, en théorie, être utilisé pour filtrer une partie des doublets. Cette approche est notamment pertinente dans le cas où l'on travaille sur des échantillons dont la composition est supposée être homogène (un type cellulaire, une lignée cellulaire). Dans ce type d'expérimentation, il sera d'ailleurs souvent pertinent d'écarter les 2 extrémités de la distribution des métriques de contrôle qualité.

Pitfalls Dans le cas où notre échantillon contient divers types cellulaires, il faut cependant rester prudent lors de la définition des seuils. En effet, la quantité d'ARN dans une

cellule est dépendante du type cellulaire et de l'état dans lequel se trouve la cellule. Ainsi, les "petits" types cellulaires se retrouvent en queue de distribution et sont par conséquent plus facilement filtrés sur ce critère. En parallèle, les cellules de mauvaise qualité appartenant à des types cellulaires "riches" en ARN ne seront pas écartées par ce type de filtre. Par ailleurs, lorsque la taille du jeu de données le permet (peu d'échantillons), il peut être pertinent de définir dans un premier temps des clusters puis d'appliquer des filtres par cluster. Comme pour la filtration des doublets ou le traitement de la contamination par les ARN ambiants, il convient la plupart du temps de définir des seuils de filtration par échantillon. En effet, les variations de la profondeur de séquençage, du nombre de cellules encapsulées et la composition cellulaire des échantillons vont faire évoluer les seuils de filtration. La définition de seuils par échantillon est notamment importante dans le cas où l'on travaille sur des échantillons générés avec des versions différentes du kit scRNA-seq du chromium de 10X comme dans notre cas. Les versions du kit ne produisent pas tout à fait les mêmes résultats en termes d'UMIs et de nombre de gènes détectés, ce qui introduit un fort effet batch. Cet effet est cependant clairement identifiable et stable, ce qui permet de réduire son impact de façon efficace avec des méthodes d'intégration par exemple (voir section 3.3).

"Notre façon de faire" Comme exposé précédemment, la définition des filtres est dépendante du design expérimental. Dans le cas de notre étude du transcriptome en cellule unique de la lignée MM.1S avec ou sans traitement à la dexaméthasone, nous avons pu définir un seuil unique pour les différents échantillons. En effet, un seul type cellulaire est représenté. On choisit alors une approche manuelle pour éliminer les cellules trop à l'écart du reste de la distribution. Dans le cas du jeu de données MYRACLE, nous avons un nombre d'échantillons important avec une variabilité de composition cellulaire non contrôlée. Cela influe sur la profondeur de séquençage par cellule. En effet, les cellules à fort contenu ARN vont absorber une majorité des reads si elles sont sur-représentées. Malgré cette hétérogénéité, nous avons décidé de fixer le minimum d'UMIs par cellule à 1000 pour l'ensemble des échantillons après avoir testé différentes alternatives. Dans un premier temps, nous avons essayé d'écartier les outliers de la distribution. Cependant, cette dernière est souvent très concentrée du côté des valeurs faibles ou encore parfois multimodale, ce qui ne permet pas de définir un seuil pertinent de cette façon.

Du côté des valeurs hautes de la distribution, nous avons décidé de ne pas établir de plafond. Bien que de trop fortes valeurs puissent être synonyme de multiplet, la va-

riabilité de composition des échantillons ne permet pas de les écarter efficacement sur ce critère (Voir section 3.2.3). De plus, la présence des cellules tumorales aux caractéristiques variables complique encore l'établissement d'un tel plafond.

3.2.4.2 Nombre de gènes détectés

Pertinence de la variable Le nombre de gènes détectés au sein d'une cellule est lui aussi un indicateur important de la qualité de son transcriptome. En effet, plus on détecte de gènes au sein d'une cellule, plus on a d'informations sur celle-ci et donc mieux les données reflètent le transcriptome réel de la cellule. Cette variable est fortement dépendante et corrélée au nombre d'UMIs des cellules. Cependant, il est important de la traiter indépendamment malgré la redondance apparente avec les filtres basés sur les UMIs. En effet, comme évoqué dans d'autres cas, la diversité des types cellulaires et de leur contenu ARN justifie cette approche. En effet, dans le cas d'un type cellulaire dont le transcriptome serait à la fois composé de nombreux ARN mais dominé par quelques gènes seulement, on observera alors des cellules avec un nombre d'UMIs suffisant pour passer les seuils mais avec un faible nombre de gènes uniques détectés. Dans ce cas, il sera difficile d'exploiter les cellules lors des analyses secondaires. Il est donc préférable de les écarter du jeu de données.

Pitfalls Comme pour la filtration sur la base du nombre d'UMIs, il faut cependant rester prudent dans les cas où plusieurs types cellulaires sont présents au sein de l'échantillon. En effet, comme pour les UMIs, la quantité de gènes uniques détectés est dépendante du type cellulaire. Dans certains cas, un seuil bas va donc permettre de filtrer efficacement les cellules de mauvaise qualité au sein de certaines populations tandis que pour des populations avec un nombre de gènes exprimés intrinsèquement bas, on observera une sur-filtration. Ce type de filtration peut introduire un biais lorsqu'on travaille sur les proportions de nos populations au sein des échantillons.

"Notre façon de faire" Dans le cas de notre jeu de données MYRACLE, nous avons mis en place des seuils de filtration dynamiques. Ces seuils dynamiques nous permettent de tenir compte des différences de profondeur de séquençage et de version du kit chromium entre nos échantillons. Pour chaque échantillon, le seuil est fixé à deux déviations absolues de la médiane (median absolute deviation en anglais) (MAD), mais jamais moins de 600 gènes. Ce seuil minimum a été mis en place empiriquement sur la base des résultats obtenus

pour certains échantillons. Pour les valeurs hautes de la distribution, nous n'appliquons pas de plafond. Un plafond de ce type aurait principalement pour but d'écarter d'éventuels doublets, hors dans notre cas, il permettrait uniquement d'écarter les doublets composés d'au moins une cellule issue d'un des types cellulaires avec le nombre de gènes le plus intrinsèquement élevé et passerait à coté des doublets composés de cellules avec un faible nombre de gènes.

3.2.4.3 Pourcentage de reads mitochondriaux

Un autre type de variables est très utilisé pour le contrôle qualité : le pourcentage du transcriptome représenté par un groupe de gènes au sein de la cellule. Le plus utilisé est le pourcentage de reads correspondant à des ARN mitochondriaux. Cette variable est utilisée de façon quasi systématique dans le traitement de la plupart des jeux de données scRNA-seq. Comme pour la plupart des variables, il est difficile de fixer une valeur plafond unique, puisqu'une fois encore celle-ci est en partie liée au type cellulaire considéré. Dans le cas de l'étude MYRACLE, nous avons décidé de fixer le plafond à cinq MAD. Cependant, pour s'assurer d'éviter des valeurs aberrantes, nous contraignons ce plafond entre 10 et 20%. Nous appliquons également un seuil minimum à 0,25% car nous avons observé que les cellules concernées ne semblaient pas être bien traitées par les étapes de réduction de dimensions et de clustering.

3.2.4.4 Pourcentage de reads ARN longs non codant (lncRNA)

Au cours de l'exploration des données et de la mise en place du pipeline, nous avons rencontré des difficultés à filtrer certaines cellules manifestement de mauvaise qualité sur les précédents critères. Les clusters formés par ces cellules étaient caractérisés principalement par une forte expression de *MALAT1*, ainsi que quelques autres gènes correspondant à des lncRNA. L'enrichissement de ces gènes suggère que la mauvaise qualité de ces cellules est due à une perte partielle ou totale du cytoplasme avant encapsulation. Les ARN obtenus sont alors principalement originaires du noyau cellulaire, ce qui conduit à un enrichissement des transcrits qui y sont sur-représentés. Afin de se débarrasser de ces cellules, nous avons initialement décidé de filtrer les cellules sur cette variable. Cependant, pour une partie des échantillons, cela nous conduisait à sur-filtrer, écartant ainsi des cellules de qualité acceptable. Par ailleurs, la redondance importante entre cette variable de QC et les autres, nous a permis de nous passer de ce filtre.

3.2.5 Discussion des filtres

Les filtres mis en place nous permettent donc d’assembler et d’analyser un jeu de données robuste et le moins biaisé possible. Un des points importants à aborder sur les contrôles qualités de nos données est la redondance observée entre les variables. En effet, et notamment pour les variables telles que le nombre d’UMI, de gènes ou encore le pourcentage mitochondrial, une part importante des cellules est discriminée par au moins deux variables. Cependant, chacune d’entre elles permet d’affiner et d’améliorer le résultat final. Les méthodes et variables présentées précédemment couvrent la majorité des aspects du contrôle qualité des données scRNA-seq. Il faut cependant rappeler qu’il existe une très large variété d’outils et d’algorithmes disponibles pour traiter ces problématiques de contrôle qualité. De plus, régulièrement, de nouvelles techniques et approches sont développées et rendues accessibles. Cette grande variété d’outils permet de plus en plus et de mieux en mieux d’adapter le traitement bio-informatique des données aux questions et au design spécifique de chaque étude. La construction et le choix des outils utilisés dans un pipeline d’analyse sont donc dépendants du type de données, du design expérimental mais aussi dans une certaine mesure de l’accessibilité. En effet, la qualité de l’implémentation, le langage utilisé ainsi que la qualité de la documentation associée à l’outil sont des critères essentiels pour leur utilisation. Par ailleurs, il est essentiel de préciser que, dans notre cas comme dans beaucoup d’autres, il serait possible d’améliorer les combinaisons et les paramétrages des outils. Cependant, il faut tenir compte de la temporalité et des contraintes de temps qui nous imposent des choix. En effet, il n’est pas possible de mettre à jour et relancer l’ensemble des analyses d’un projet à la lumière des nouveautés disponibles.

3.3 Normalisation

3.3.1 Objectif de la normalisation

Le nombre d’UMIs obtenu par cellule diffère en fonction du volume cellulaire mais aussi aléatoirement durant le séquençage. L’objectif de la normalisation est de rendre comparables les profils transcriptomiques ainsi obtenus.

L’une des caractéristiques importante du scRNAseq est la forte proportion de zéro au sein de la matrice de comptage. Ces zéros ont à la fois une origine biologique - certaines populations de cellules n’expriment simplement pas ces gènes - mais aussi une origine

technique appelée dropout, principalement pour des gènes peu exprimés. Pour les gènes moyennement ou fortement exprimés, cet effet de dropout introduit une variabilité de niveau d'expression. La présence de différents types cellulaires au sein d'un échantillon rend également les données scRNAseq très hétérogènes. Ces différents aspects rendent donc les données scRNAseq assez différentes des données de RNA-seq bulk. Par conséquent, les méthodes utilisées pour analyser les données bulk ne s'appliquent pas systématiquement aux données single-cell (VALLEJOS et al. 2017).

RÉSULTATS

4.1 Étude du mode d'action de la dexaméthasone

Nous avons étudié l'effet de dex sur le génome de cellules de myélome. Cette étude a été réalisée en collaboration avec Victor Bessonneau-Gaborit dans le cadre de nos projets de thèse respectifs.

4.1.1 Travail préliminaire, par Victor Bessonneau-Gaborit (Gaborit 2021)

L'ensemble des techniques utilisées pour caractériser les effets de la dex sur le génome de la lignée MM.1S sont schématisées en figure 4.1. L'intégration et l'analyse des données de séquençage concernant l'immunoprécipitation de chromatine (ChIP-seq), sur la marque d'histone H3K27ac, l'accessibilité de la chromatine (assay for transposable-accessible chromatin sequencing en anglais) (ATAC-seq) et le séquençage de l'ARN (RNA-seq) bulk est réalisé par Victor. Son travail a permis de montrer que seul 6% des sites de liaison du GR sont associés à des gains d'activité des enhanceurs et une augmentation des boucles de régulations impliquant les enhanceurs et les promoteurs. Il a également montré que les enhanceurs se regroupent pour former un réseau d'interactions spatiales avec un unique promoteur afin de réguler l'expression du gène cible. De plus, il a également mis en évidence l'existence d'un enhanceur prédominant (pE) au sein de chaque réseau. Celui-ci est généralement proche du promoteur, son accessibilité est plus élevée, la marque H3K27ac est plus présente et il existe un enrichissement de RAD21 à son niveau après exposition à la dex. Par ailleurs, ces pEs forment plus de boucles d'interaction que les autres et sont plus centraux dans le réseau (article figure 3E).

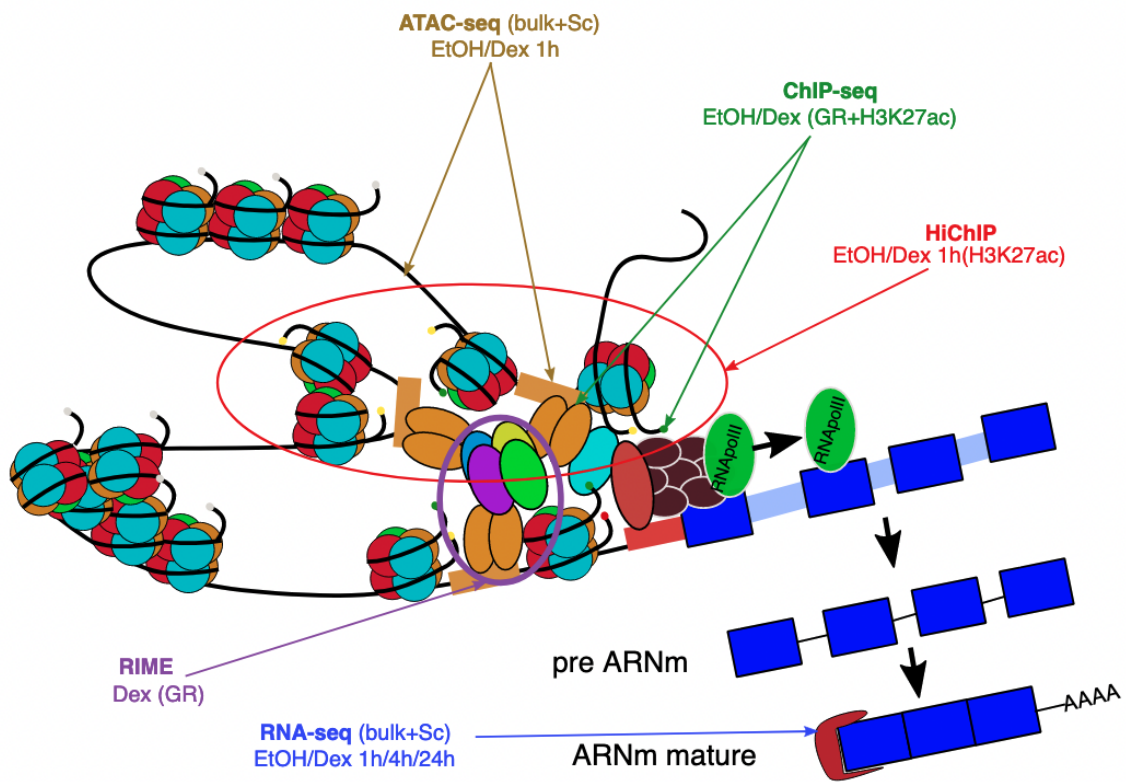


FIGURE 4.1 – Schéma des techniques utilisées pour établir les effets de la dex sur le génome de la lignée MM.1S

4.1.2 Analyses en cellule unique de la lignée MM.1S exposée à la dex

Les données de scRNA-seq et de scMultiome ont été générées et pré-analysées comme présenté dans la section 3 : Méthodes. Afin de mieux comprendre les réseaux d'interactions entre les enhancers (pEs et autres) et les promoteurs, nous avons utilisé l'aspect ATAC-seq des données scMultiome. Nous avons utilisé la méthode de prédiction d'interactions proposée par Cicero (PLINER et al. 2018). Cette méthode permet de prédire les interactions à partir de la co-accessibilité des pics ATAC-seq au sein de la population de cellules. Cette prédiction prend en compte la distance génomique des pics afin de prédire les interactions locales. Pour chaque gène d'intérêt, nous avons extrait les liens de co-accessibilité impliquant des pics correspondant aux régions utilisées pour construire le réseau d'interactions spatiales. Cette approche nous a permis de montrer des changements de co-accessibilité entre les pEs et les autres régions régulatrices lors de l'exposition à la dex (article figure 5). Un autre aspect important des résultats de ces travaux est la mise en évidence de la variabilité transcriptomique au sein de la population. Cette variabilité comporte deux aspects. Dans un premier temps, cela est perceptible au niveau de la magnitude de la réponse. En effet, au sein de la population nous avons calculé le "ratio of responding genes" (RRG), qui correspond au ratio de gènes induits par la dex (51 gènes les plus sur-exprimés en scRNA-seq après exposition à la dex) détectés pour chaque cellule. Avant traitement, la médiane du RRG est de 12% au sein de la population. Certains gènes sont donc exprimés pré-traitement et stimulés par la dex. Après traitement, la médiane du ratio of responding genes (RRG) est seulement de 51%. Seule la moitié des gènes de réponse est donc exprimé dans chaque cellule. La réponse est donc hétérogène au niveau transcriptomique, en terme de magnitude de la réponse avec une majorité des RRG compris entre 20 et 90% (article figure 6B,C), dans un premier temps, mais aussi au niveau des gènes induits en réponse à la dex. Nous avons constaté sur la projection Uniform Manifold Approximation and Projection (UMAP) la formation de deux groupes de cellules après traitement. Ces deux groupes ont des RRG différents (article figure 6C), et semblent ne pas exprimer les mêmes gènes en réponse au traitement.

L'analyse des corrélations deux à deux entre nos 51 gènes a permis de révéler l'existence de deux groupes de gènes. Ceux-ci contiennent respectivement 45 et 6 gènes (figure 4.2). Au sein du premier groupe, on retrouve principalement les gènes ubiquitaires de réponse à la dexaméthasone (*TSC22D3*, *FKBP5*, *DDIT4*, etc), mais aussi un gène *BCL2L11* codant

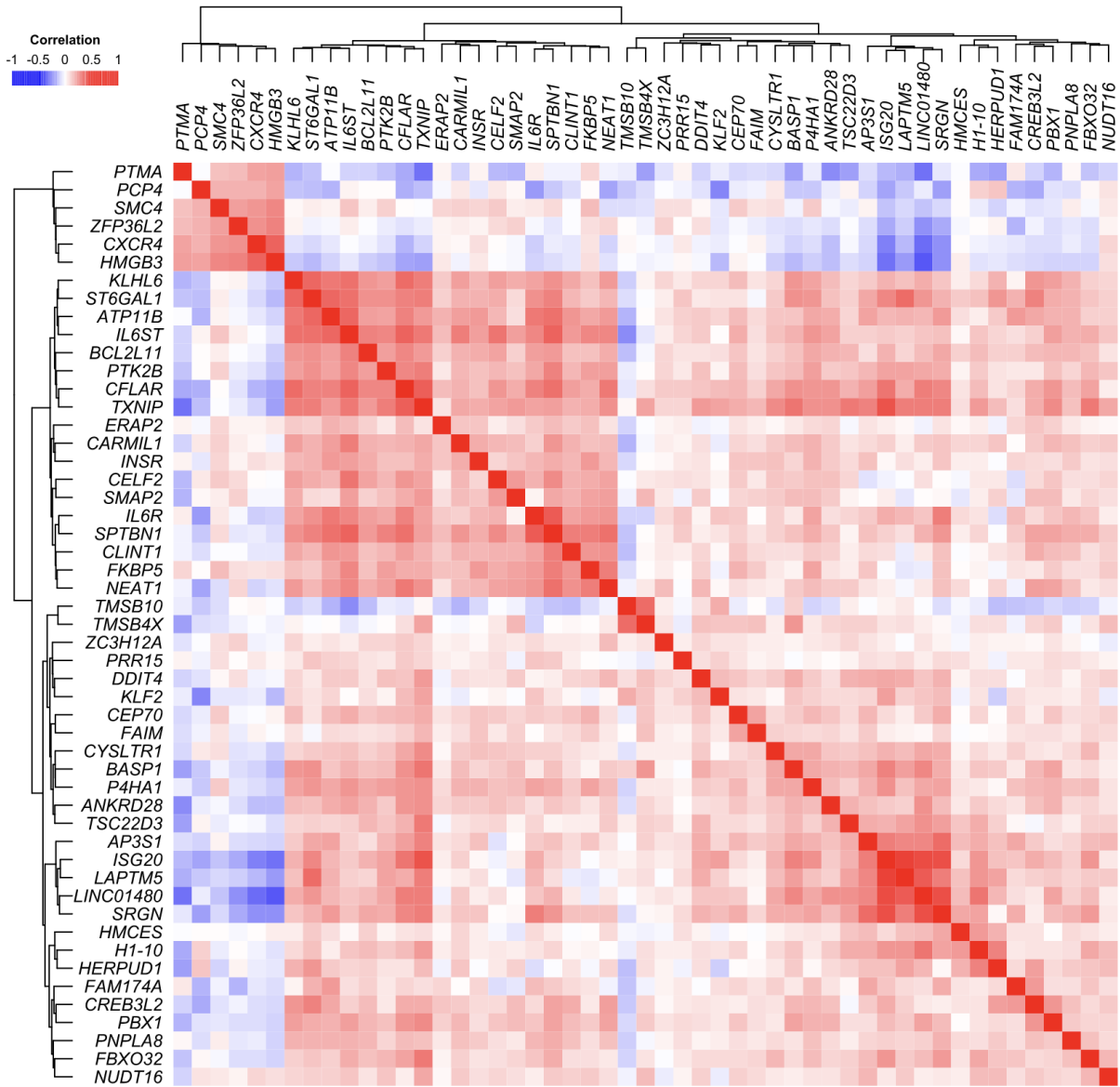


FIGURE 4.2 – Heatmap de corrélation des 51 gènes les plus sur-exprimés en réponse à la dex dans les données scRNA-seq (article figure supplémentaire S10A)

pour la protéine pro-apoptotique de la famille BCL2 : BIM. Le second groupe comprend notamment *CXCR4*. Les corrélations au sein de ces deux groupes suggèrent qu'ils ne sont pas exprimés par les mêmes cellules. En effet, ces deux groupes de gènes sont respectivement plus exprimés au sein d'un des deux clusters et sont mutuellement exclusifs. À titre d'illustration, nous avons coloré la projection UMAP en fonction de l'expression de *BCL2L11*, *CXCR4* et des deux combinés (article figure 6E). Des projections similaires de *CXCR4* avec *TSC22D3* et *FKBP5* montrent le même résultat.

L'exploitation des données en cellule unique a permis de montrer qu'en réponse à la dex, le réseau d'interactions des enhanceurs est modifié de manière hétérogène entre les cellules. En effet, les enhanceurs, où se fixe la dex qui sont accessibles varient en fonction de la cellule considérée. Nous avons également pu montrer que cette variabilité se traduit au niveau de la transcription par des différences d'expression. Enfin, nous avons pu montrer que la lignée MM.1S peut répondre à la dex de deux façons différentes. Notre étude n'a cependant pas permis de mettre en évidence l'origine précise de cette diversité de réponse. Le stade du cycle cellulaire auquel se trouve la cellule au moment du traitement pourrait être l'une des explications. En effet, l'expression de *CXCR4* est associée au cycle cellulaire. L'expression préférentielle d'un des deux modules pourrait donc conduire les cellules à des destins différents. *CXCR4* est plutôt considéré comme favorable à la survie et la prolifération de la tumeur notamment dans le cadre des traitements de chimiothérapie, tandis que BIM est un pro-apoptotique.

L'ensemble des résultats obtenus nous a permis de mettre en évidence différents mécanismes de régulation de l'expression des gènes par la dex. Le recablage permet l'activation d'un enhanceur plus efficace, ce qui permet d'améliorer l'expression du gène cible (figure 4.3A). L'activation simple d'un enhanceur par fixation du GR permet également l'activation de l'expression de certains gènes (figure 4.3B). Les deux derniers mécanismes sont plus complexes. La dex peut en effet modifier la conformation spatiale de la chromatine, permettant de créer un hub d'enhancers et de super-enhancers pour améliorer l'expression du gène cible (figure 4.3C). Enfin, la dex via le recrutement du complexe cohésine peut également permettre de stabiliser une interaction de l'enhancer avec son promoteur (figure 4.3D).

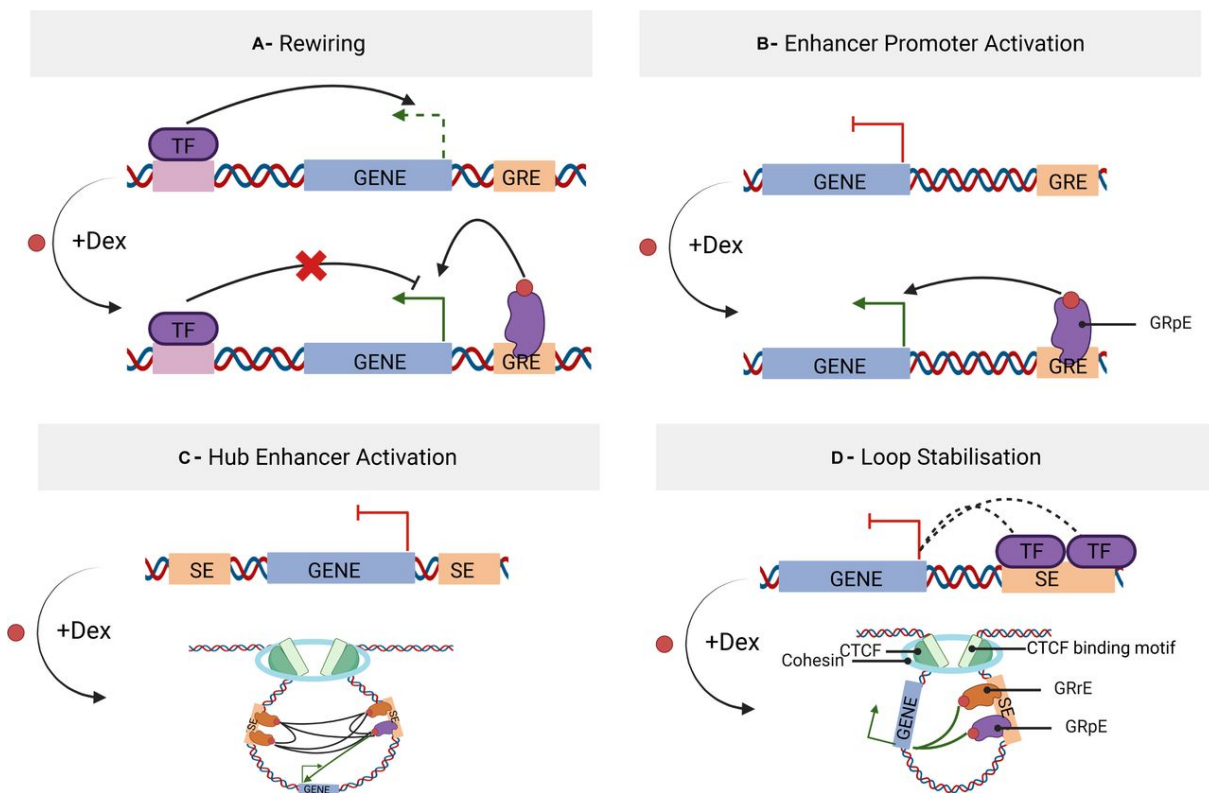


FIGURE 4.3 – Schéma des différents mode d'action de la dexaméthasone sur la régulation de l'expression des gènes cibles

4.1.3 Version finale de l'article présentant les résultats de notre étude du mode d'action de la dexaméthasone sur les cellules de myélome (Bessonneau-Gaborit et Cruard 2023)

Research Article



Exploring the impact of dexamethasone on gene regulation in myeloma cells

Victor Bessonneau-Gaborit^{1,2,*}, Jonathan Cruard^{1,*}, Catherine Guerin-Charbonnel^{1,3}, Jennifer Derrien¹, Jean-Baptiste Alberge¹, Elise Douillard^{1,2}, Magali Devic^{1,2}, Sophie Deshayes¹, Loïc Champion^{1,3}, Frank Westermann^{4,5}, Philippe Moreau^{1,2}, Carl Herrmann⁶, Jérémie Bourdon⁷, Florence Magrangeas^{1,2,†}, Stéphane Minvielle^{1,2,†}

Among glucocorticoids (GCs), dexamethasone (Dex) is widely used in treatment of multiple myelomas. However, despite a definite benefit, all patients relapse. Moreover, the molecular basis of glucocorticoid efficacy remains elusive. To determine genomic response to Dex in myeloma cells, we generated bulk and single-cell multi-omics data and high-resolution contact maps of active enhancers and target genes. We show that a minority of glucocorticoid receptor-binding sites are associated with enhancer activity gains, increased interaction loops, and transcriptional activity. We identified and characterized a predominant enhancer enriched in cohesin (RAD21) and more accessible upon Dex exposure. Analysis of four gene-specific networks revealed the importance of the CTCF-cohesin couple and the synchronization of regulatory sequence openings for efficient transcription in response to Dex. Notably, these epigenomic changes are associated with cell-to-cell transcriptional heterogeneity, in particular, lineage-specific genes. As consequences, *BCL2L1*-encoding BIM critical for Dex-induced apoptosis and *CXCR4* protective from chemotherapy-induced apoptosis are rather up-regulated in different cells. In summary, our work provides new insights into the molecular mechanisms involved in Dex escape.

DOI [10.26508/lsa.202302195](https://doi.org/10.26508/lsa.202302195) | Received 1 June 2023 | Revised 23 June 2023 | Accepted 26 June 2023 | Published online 31 July 2023

Introduction

Dexamethasone (Dex), a synthetic glucocorticoid (GC) known for its anti-inflammatory and immunosuppressive activities, in combination with immunomodulatory drugs (IMiDs) and proteasome inhibitors (PIs), is the standard induction treatment in transplant-eligible patients with newly diagnosed multiple myeloma (MM). The use of a new class of drugs, the monoclonal antibody daratumumab, in

combination with thalidomide (IMiD), bortezomib (PI), and Dex, improved depth of response and progression-free survival (Moreau et al, 2019). Recently, T-cell-engaging therapies administrating bispecific antibodies targeting CD3 and B-cell maturation antigen for patients who had relapsed or refractory MM, in which Dex is delivered in premedication, showed a promising efficacy in depth and duration of the response (Moreau et al, 2022). Despite spectacular therapeutic improvement, few patients are cured; therefore, it is necessary to better understand the precise mechanisms of action of each agent alone or in combination.

Dex exerts its biological functions by binding to the glucocorticoid receptor (GR) encoded by *NR3C1*. Upon Dex binding, the complex translocates to the nucleus, where it associates with DNA at GR-binding sites, acts as a transcription factor (TF), and regulates gene expression (Reddy et al, 2009). GR binding appears to be preprogrammed by the binding of lineage-specific TFs and chromatin accessibility before exposure (Biddie et al, 2011; John et al, 2011). At these loci, GR co-binds with cell-specific pioneer TFs including CEBPB in the liver (Grøntved et al, 2013), PU.1 in the macrophage lineage (Oh et al, 2017), and AP1 in murine hepatocytes (Biddie et al, 2011). GR binds predominantly at distal enhancers (Reddy et al, 2009) and drives transcription by interacting with gene promoters via chromosomal loops. GR binds thousands of locations across the genome but only few enhancers cooperate with each other to activate Dex-responsive genes (Vockley et al, 2016; McDowell et al, 2018). A previous study exploiting protein-directed chromatin interactions approach suggests that at GR-responsive genes, chromatin interaction loops between enhancers and promoters are preestablished, whereas in a subset of genomic loci, GR binding induces de novo interactions (Kuznetsova et al, 2015). High-resolution genome-wide maps of chromatin interactions in response to Dex confirmed and extended the model that GR binding acts predominantly through preestablished chromatin interactions

¹Université de Nantes, CNRS, INSERM, Centre de Recherche en Cancérologie et Immunologie Intégrée Nantes Angers, France ²Centre Hospitalier Universitaire, Nantes, France ³Institut de Cancérologie de l'Ouest, Nantes, France ⁴Hopp Children's Cancer Center Heidelberg, KITZ, Heidelberg, Germany ⁵Division of Neuroblastoma Genomics, German Cancer Research Center, Heidelberg, Germany ⁶Health Data Science Unit, Medical Faculty Heidelberg and BioQuant, Heidelberg, Germany ⁷Université de Nantes, LS2N, CNRS, Nantes, France

Correspondence: florence.magrangeas@chu-nantes.fr; stephane.minvielle@univ-nantes.fr

*Victor Bessonneau-Gaborit and Jonathan Cruard are co-authors

†Florence Magrangeas and Stéphane Minvielle are co-seniors

and increases their frequency (D'Ippolito et al, 2018). However, these studies did not resolve the influence of increased chromatin accessibility on chromatin loops and enhancer activity as shown by Stavreva et al (2015). Moreover, how these enhancers combine to induce gene expression is still poorly understood. Given that dex is an essential drug in the treatment landscape of MM disease course, the analysis of its molecular action on the genome of myeloma cells and on transcriptional heterogeneity is needed to better understand treatment escape.

Results

GR binds to preestablished chromatin sites in MM cells

To investigate the genomic features associated with GR binding in malignant plasma cells, we conducted chromatin immunoprecipitation sequencing (ChIP-seq) for GR, and acetylation of histone H3 at lysine 27 (H3K27ac), assay for transposable-accessible chromatin sequencing (ATAC-seq), and RNA sequencing (RNA-seq) in the Dex-sensitive human myeloma cell line MM.1S exposed to Dex (0.1 μ M) or to equal-volume ethanol (EtOH) for 1 and 4 h (Fig 1A). As expected, a large majority (84%; 15,862/18,844) of GR-binding sites fell within already accessible chromatin and 76% (14,243/18,844) of these sites were annotated as transcriptionally active chromatin at transcription start sites (TSSs) or active enhancers (Fig S1A–C). In addition, GR-binding sites were found in almost all (99%; 808/815) preestablished super-enhancers (SEs) (Hnisz et al, 2013). SE-associated genes included key genes for plasma cell development and MM biology such as *IGLL5*, *IRF4*, *XBP1*, *PRDM1*, and *IKZF1* in line with previous results (Lovén et al, 2013) (Fig S1D).

We next sought to identify TFs that could play an important role in driving GR binding to chromatin in myeloma cells. Motif discovery analysis using the MEME (Multiple Em for Motif Elicitation) (Bailey et al, 2015) algorithm revealed motif enrichment for IRF4 binding (ISRE), a TF critical for MM proliferation and/or survival (Shaffer et al, 2008), and the anticipated GR-responsive element (GRE) in GR-binding sites (Fig 1B). These findings suggest that upon Dex exposure GR binds preferentially to IRF4 chromatin complexes. To test this, Rapid Immunoprecipitation Mass spectrometry of Endogenous protein (RIME) experiments were performed in MM.1S Dex-treated cells using GR as bait (Fig 1C; Table S1). We found that IRF4 was one of the top-ranking partners along with IKZF1 and IKZF3 previously reported to be essential TFs in MM (Krönke et al, 2014). Co-immunoprecipitation (Co-IP) experiments confirmed that GR interacts with IRF4 in the nucleus of MM.1S exposed to Dex (Figs 1D and S1E). In addition, RIME identified known GR interactors including seven subunits of the SWI/SNF chromatin remodeling complex, of which the ATP-dependent remodeling enzyme SMARCA4 (alias BRG1) and ARID1A, required for GR-chromatin remodeling and transcriptional regulation (Fletcher et al, 2002; Trotter & Archer, 2004; Trotter et al, 2008). GR peaks in Dex-treated cells and IRF4 peaks in untreated cells strongly overlapped (73%) with H3K27ac-enriched regions (Fig 1E). Enrichment of the IRF4 motifs and IRF4-binding colocalization with 77% of GR-bound active regulatory regions suggest that IRF4 is the specific TF associated with GR

binding in myeloma cells. However, unlike *CEBPB*, which is strongly Dex-induced in the liver (McDowell et al, 2018), *IFR4* is repressed (\log_2 fold change = -0.55 at 4 h of Dex exposure), which is also the case for its two main responsive genes in MM.1S cells (Low et al, 2019): *TNFRS17* (alias B-cell maturation antigen) and *MANF* (\log_2 fold change = -0.59 and -0.43 at 24 h of Dex exposure, respectively) (Figs 1F and S2).

Despite the large number of GR-binding sites on the genome, only a small number of genes were deregulated in response to Dex (982 up-regulated, 3.2%; 649 down-regulated, 2.1%; abs [\log_2 fold change] > 0.6 and FDR < 0.05) in myeloma cells including well-known ubiquitous GR-responsive genes like *TSC22D3* (alias *GILZ*), *FKBP5*, and cell-specific genes like *BCL2L11* (alias *BIM*), an essential gene for Dex-induced death in MM.1S, and *CXCR4*, the chemokine receptor gene known to be associated with MM progression and poor prognosis (Fig 1F). Thus, suggesting that only a fraction of GR-binding sites is critical for transcriptional activity induced by Dex exposure.

Enhancer contact map in response to Dex treatment

As expected, GR-bound regions enriched in H3K27ac signal occurred predominantly (85%) at distal enhancers (Fig S3A and Table S2); they were closer to up-regulated genes than to stable genes (Wilcoxon test, $P < 0.0001$) (Fig S3B), closer to each other compared with GR peaks without H3K27ac changes (Wilcoxon test, $P < 0.0001$) (Fig S3C) and formed new SEs associated with known ubiquitous GR-responsive genes, including *DDIT4*, *FKBP5*, and *TSC22D3* (Fig S3D). In an attempt to decipher which GR-binding sites are mandatory to promote transcriptional regulation, we firstly drew an enhancer contact map of GR-activated enhancers and defined at high resolution the changes in chromatin topology resulting from Dex exposure. We employed H3K27ac HiChIP method (Mumbach et al, 2016, 2017) (Fig 2A), which interrogates chromatin contacts between active elements (enhancers or promoters) distal (i.e., located more than 5 kb away from closest TSS) or proximal (i.e., less than 5 kb from closest TSS). We identified 21,249 and 23,278 H3K27ac chromatin interactions across the genome in MM.1S EtOH and MM.1S Dex, respectively. Mapping of the enhancer connectome showed an overall enrichment of H3K27ac ChIP-seq signal in loop anchors compared with non-anchors (Wilcoxon test, $P < 0.0001$) (Fig S4A), a median loop distance of 125 kb (Fig S4B), and revealed that most of the loops (80%) involved a proximal anchor (Fig S4C). We were particularly interested in interaction loops that increased their frequency upon Dex exposure compared with stable and decreased interactions (Fig 2B). The distal interactions subgroup was overrepresented at the expense of the proximal interaction subgroup (Chi2 test, $P < 0.0001$), whereas distal–proximal interactions subgroup remained unchanged (Fig 2C). The anchors were enriched in GR occupancy (Wilcoxon test, $P < 0.0001$) (Fig 2D) and gained chromatin accessibility (Chi2 test, $P < 0.0001$) (Fig 2E). Finally, increased ATAC-seq peaks (27%) in these anchors were enriched in the GRE motif (P -val = 1×10^{-30}), whereas stable ATAC-seq peaks were enriched in the ISRE motif (P -val = 1×10^{-69}). A scanning-motif approach for GRE and ISRE revealed that GRE was found in 11.72% and 1.98% of increased and stable ATAC-seq peaks, respectively. We also found ISRE in 19.48% and 9.68% of increased and stable ATAC-seq peaks, respectively

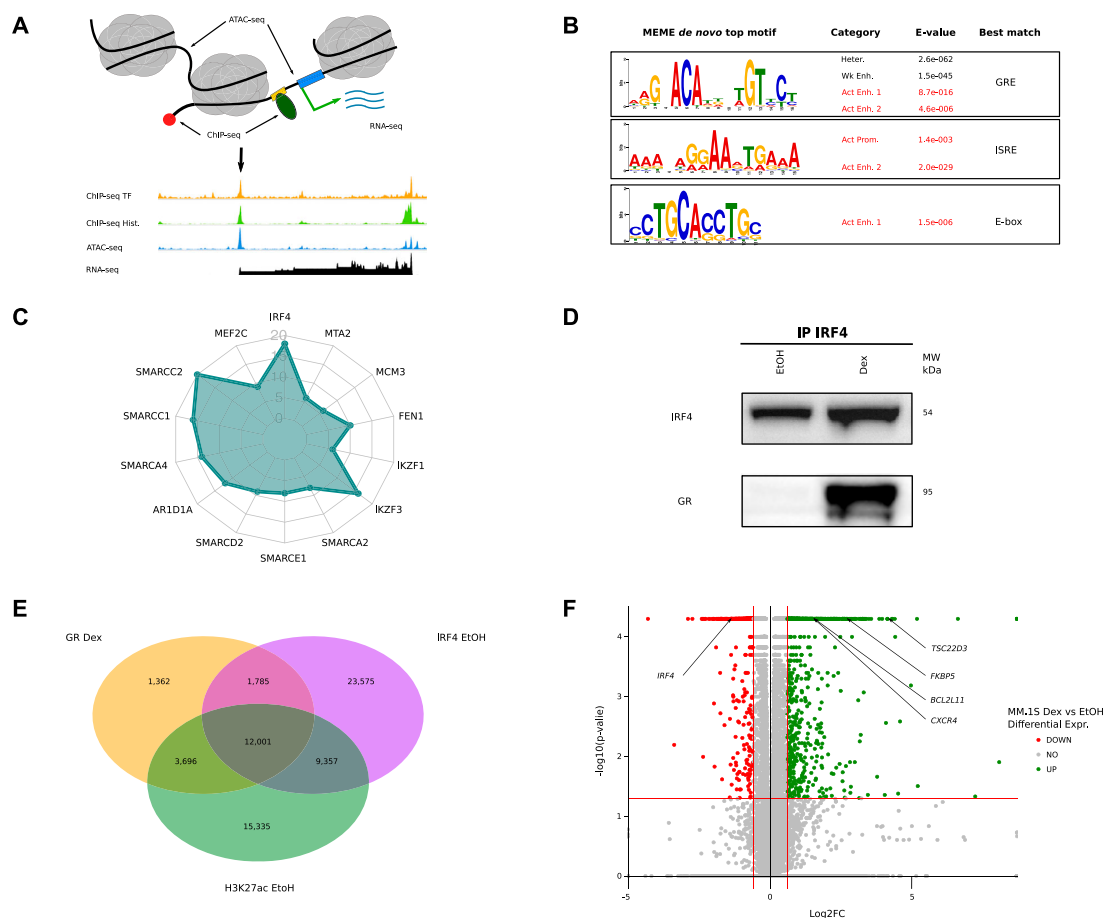


Figure 1. Preprogrammed chromatin landscape guides GR binding in malignant plasma cells.

(A) Scheme of sequencing data used to define chromatin landscape of MM.1S cell line. (B) De novo Multiple Em for Motif Elicitation top motif enriched in MM.1S chromHMM functional states. (C) Radar chart showing GR-binding partners using rapid immunoprecipitation mass spectrometry of the endogenous protein method in MM.1S (Dex 1 h); arbitrary units. (D) Immunoprecipitation of IRF4 transcription factor in MM.1S cells in control and Dex conditions. Revelation by IRF4 and GR antibodies. (E) Overlap of chromatin immunoprecipitation sequencing peaks for GR (MM.1S Dex 1 h), IRF4 (MM.1S EtOH, [Loven et al \[2013\]](#)), and H3K27ac (MM.1S EtOH). (F) Volcano plot of differential RNAseq in Dex treatment compared with control, dysregulated genes are coloured in green (up) and red (down), and key genes are highlighted.

(Fig 2F; Table S3). This reinforces the fact that for an opening to occur, the GR must bind directly to the DNA.

Predominant enhancer (pE) genomic features

To further investigate the impact of Dex exposure on regulatory networks, we selected among the Dex-increased distal-proximal interaction loops those associated with Dex-up-regulated genes (Fig S5A). We retained 62 interaction loops corresponding to 55 genes (Fig 2B), which displayed high transcriptional variability, in terms of median expression, induction intensity, and ratio of expressing cells (Fig S5B). Half of the genes appeared to be regulated by short-range promoter-enhancer interactions (<5 kb) (Fig S5C). Analysis of the enhancer connectome upon Dex exposure revealed formation of about one SE in 80% of gene networks (Fig

S5D). Of note, the density of interconnections was very varied from one network to another (Fig S5E) but was not associated with the level of gene expression, nor was the number of interactions that increased in the presence of Dex associated with the level of expression ($|r| < 0.2$). Interestingly, a careful visual inspection of increased interactions across *BCL2L11*, *CXCR4*, *FKBP5*, and *TSC22D3* revealed a particular enhancer that formed more spatial contacts than the others after Dex exposure (Fig S5F). Our observations, together with previous results ([Kuznetsova et al, 2015](#); [Vockley et al, 2016](#)), suggest that one particular enhancer could act as a regulatory hub to promote transcription of Dex-responsive genes. To explore this hypothesis, we firstly selected in each network one enhancer (as described in the Materials and Methods section), we retained 34 enhancers referred to as pE (Fig S6A and B; Table S4), among those, 79% were located within a SE. The pEs showed

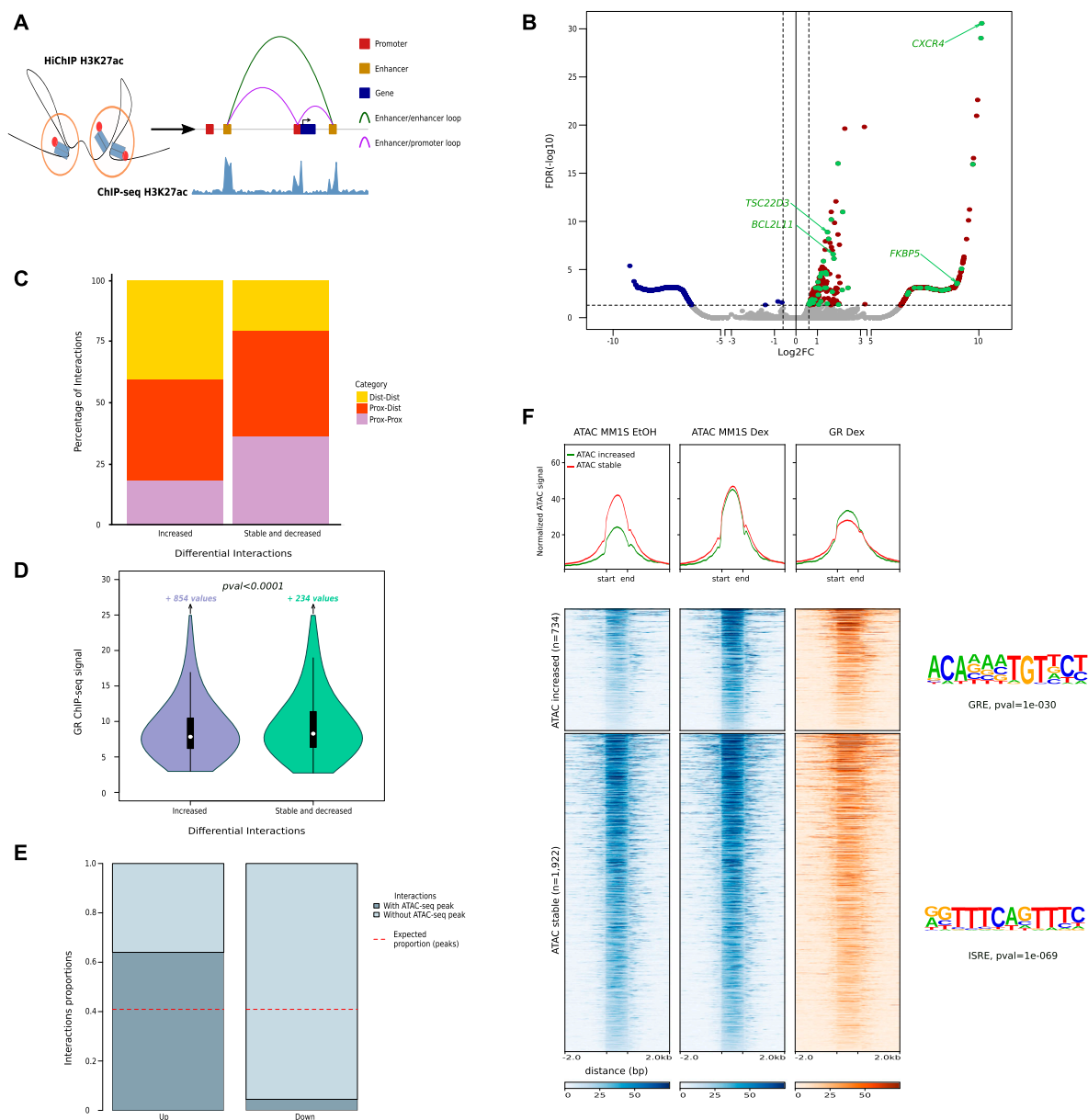


Figure 2. H3K27 acetylation chromatin interaction in response to Dex.

(A) Scheme depicting H3K27ac loop interactions between regulatory regions through linear genome. **(B)** Volcano plot of HiChIP H3K27ac differentially induced (FDR < 0.05) loops between EtOH and Dex MM1S cells, increased ($\log_2FC > 0.6$) loops in red and decreased ($\log_2FC < -0.6$) loops in blue; interactions with anchors overlapping with Dex-induced gene expression are shown with green dots. **(C)** Annotation of interactions depending on three categories: Dist-Dist (>5 kb from closest transcription start site [TSS] for each HiChIP anchors); Prox-Dist (5> kb from closest TSS for one HiChIP anchor and <5 kb for the other); prox-Prox (<5 kb from closest TSS for each HiChIP anchors). **(D)** Box plots illustrating the significantly higher levels of GR signal in increased interactions compared with GR signal in stable or decreased interactions. **(E)** Barplots showing proportion of interactions with and without transposable-accessible chromatin sequencing (ATAC-seq) peaks in increased and decreased interactions. Red dotted lines show the expected proportion of interactions with ATAC-seq peaks in case of no difference between up and down interactions. **(F)** Heatmaps illustrating ATAC-seq signal (EtOH and Dex conditions) and GR chromatin immunoprecipitation sequencing signal in ATAC-seq Dex-increased or Dex-stable peaks located within anchors of the Dex-increased H3K27ac HiChIP loop. Notable significant motifs found with de novo motif discovery approach are highlighted at right.

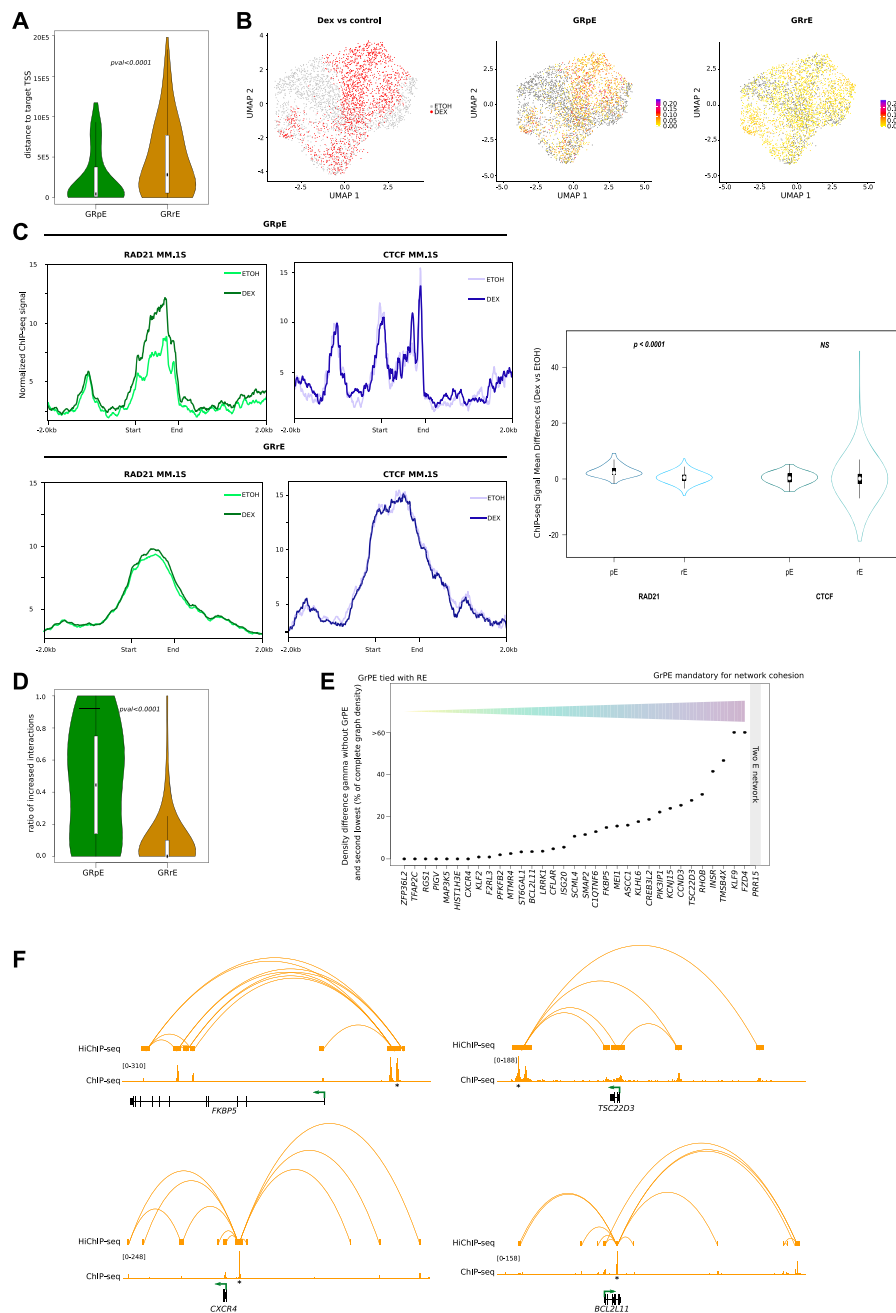


Figure 3. Comparison of molecular features between pE and rE among regulatory networks. (A) Box plot illustrating the absolute distance from closest Dex-induced transcription start site of GR predominant enhancers (pEs) (green) and GR regular enhancers (rEs) (yellow). (B) Uniform Manifold Approximation and Projection of scATAC-seq profiles in MM.1S cells, coloured by sample of origin (left), GRpE activity score (middle), and GRrE activity score (right). (C) chromatin immunoprecipitation sequencing (ChIP-seq) signal for RAD21 (left) and CTCF (right) in control (light green and light blue) and in Dex condition (green and blue) at the GRpE (top) and GRrE positions (bottom); box plots illustrating mean differences between Dex and ETOH ChIP-seq signals for RAD21 (left) and CTCF (right) at GRpE versus at GRrE. (D) Box plot illustrating the ratio of Dex-increased chromatin loops for the anchor containing a GRpE (green) or the anchors containing GRrEs (yellow). (E) Details of cohesiveness index of pE for each of the 37-gene regulatory networks. The higher the index, the more essential the pE is for the network cohesion. Index could not be computed for PRR15 because the network only has two vertices. (F) Snapshot visualisation of GR HiChIP and GR ChIP-seq data at four different loci: FKBP5, CXCR4, TSC22D3, and BCL2L11.



different molecular features than regular enhancers (rEs) (Fig 3). The pEs were closer to their target gene promoter (Fig 3A) and more accessible (Figs 3B and S7). We detected an enrichment of cohesin subunit RAD21 after Dex exposure at pEs, which correlated with gene expression ($r = 0.31$), whereas Dex exposure had no effect on CTCF binding at both pEs and rEs (Fig 3C). In line with our hypothesis, pEs formed more increased interaction loops than rEs (Fig 3D) and were essential for the cohesion of the network (Fig 3E). For example, GR HiChIP experiments, which interrogated chromatin contacts between GR-bound elements, clearly showed the central position of pEs in the *FKBP5*, *TSC22D3*, *CXCR4*, and *BCL2L11* loop networks (Fig 3F). The features of pEs are similar to those of the hub enhancers described by (Huang et al, 2018), but unlike the latter, they regulate other genes than the cell-identity genes. From these results, we conclude that in spite of a great variability of network connections and transcriptional changes, the pEs have common characteristics, they open more, bind more cohesin, and interact more than rEs upon Dex exposure.

Importance of CTCF-cohesin couple and co-accessibility in Dex response

To characterize the relationship between epigenomic events involving pEs and changes in gene expression, we focused our analysis on the four GR-responsive genes described above: *FKBP5*, *CXCR4*, *TSC22D3*, and *BCL2L11*. These four genes are well representative of the different subgroups of genes regulated by Dex: ubiquitous versus lineage-specific, densely connected versus poorly connected, robust induction versus medium induction. Regarding chromatin interaction maps assessed by H3K27ac-HiChIP experiments, we observed changes among the regulatory network of these genes (Fig 4). At the *FKBP5* gene locus, we found a strong induction of long-distance pE-rE contacts and short-distance rE-rE contacts within SEs in Dex-treated cells compared with EtOH-treated cells (Fig 4A). Increase of contacts between enhancers coincided with strong up-regulation (\log_2 fold change = 4) of *FKBP5*. In contrast, regarding the *CXCR4* gene locus, Dex induced new contacts between pE and *CXCR4* promoter and between the promoters of *CXCR4* and *DARS1* without formation of SE. These new contacts preceded a 2.9- \log_2 fold change increase of *CXCR4* expression, whereas *DARS1* expression remained unchanged (Fig 4B). However, in the case of *TSC22D3* locus, though a strong induction of gene expression after Dex exposure was found (\log_2 fold change = 4.3), interaction loops changes were only moderate (\log fold change < 1.5) (Fig 4C). This indicates that the H3K27ac-HiChIP assay is sometimes limited in defining contact maps that could explain gene expression changes, which is probably because of technical constrains. Indeed, in addition to a possible overestimation of interactions, given the high overlap between H3K27ac-increased enhancers and H3K27ac-increased loops (Fig S8A), we cannot exclude, in control conditions, an overrepresentation of labile contacts ineffective for transcription because of the fixation steps that freeze chromatin contacts. In addition, possible artifacts are introduced by bulk analysis, which averages out the effects on heterogeneous populations. To circumvent these limitations, we employed a computational method based on co-accessibility scores at single-cell resolution that predicts *cis*-regulatory DNA interactions between two regulatory elements in the same cell from scATAC-seq data (Pliner et al, 2018) (Fig S8B). As anticipated, Dex

exposure significantly modified co-accessibilities of regulatory elements ($P = 0.02$; Fig S8C). Analysis of scATAC-seq data at *TSC22D3* gene locus indicated significant gain in accessibility only at GR-bound pE (called GRpE) and the proximal rE (called GR1), both enhancers being clustered in the SE (Fig 5A top). Co-accessibility scores predicted significant Cicero links between these enhancers and the promoter, and within the SE (Fig 5A middle). These results indicate that Dex exposure leads to an increase in the number of cells with co-accessibility between enhancers within SE, and between SE and the promoter. Because we have observed CTCF occupancy at enhancers in control condition and RAD21 enrichment at pEs upon Dex exposure, we investigated binding of these molecules at *TSC22D3* gene locus. We found that the promoter and its SE were in close proximity to convergently oriented CTCF-binding sites predicted to form strong CTCF-associated loops (Oti et al, 2016). Although there were no CTCF occupancy changes, we observed a strong enrichment of cohesin subunit RAD21 at GR-binding sites of GRpE, GR1, and the promoter (Fig 5A bottom). These observations are consistent with a recent report that GR associates with cohesin complex at GR-responsive sites, strengthening preestablished chromatin loops, promoting DNA loop extrusion, and activation of *Tsc22d3* expression (Rinaldi et al, 2022).

We observed a similar phenomenon for the other ubiquitous GR-responsive gene *FKBP5* in response to Dex: an increased accessibility at GR-binding sites of GRpE, GR4, and the promoter (Fig 5B top). However, the number of Cicero-based links within SEs and also between SEs, increased significantly (Fig 5B middle), which was concordant with HiChIP data (Fig 4A). Cicero's links mostly involve GRpE, as shown by its central position in the predicted network (Fig 5B bottom right). We also found RAD21 enrichment at GR-binding sites of two neighbouring enhancers (<4 kb), GRpE and GR6 both located in the downstream SE. In addition, this SE harbored multiple CTCF sites, exhibiting inward-oriented CTCF motifs indicating a CTCF boundary, whereas chromatin configuration at its regulatory partner, the intragenic SE that clustered 4 rEs, exhibited one inward-oriented CTCF-binding motif (Fig 5B bottom). This suggests that GR binding could promote loop extrusion via cohesin mechanism, where a loop anchor (GRpE locus in this study) forms contacts with a contiguous genomic domain, such as SE; those structures are referred to as architectural stripes (Vian et al, 2018) that enhance transcription, in agreement with the recent study of Rinaldi (Rinaldi et al, 2022).

In contrast, the lineage-specific gene *BCL2L11*, already expressed in control condition, displayed a moderate up-regulation upon Dex exposure. We observed that among the 10 regulatory elements bound by GR, only GRpE was more accessible upon Dex exposure (Fig 5C top). Not surprisingly, Cicero's algorithm predicted a new co-accessibility link between the GRpE and the promoter (Fig 5C middle). The network graphs showed the switch of the promoter to a central position in the network directly connected to GRpE upon Dex exposure (Fig 5C bottom right). We also observed an enrichment of cohesin (RAD21) at both GRpE and the promoter which could stabilize contacts, suggesting that GR binding to GRpE promotes GRpE-promoter interactions at the expense of other promoter-enhancer interactions. However, given the moderate increase of transcription, we hypothesize that the "rewiring" occurs in a subset of the cell population. Likewise, in the another lineage-specific gene *CXCR4*, which displayed a mild transcriptional activation upon Dex exposure, we observed a greater accessibility at

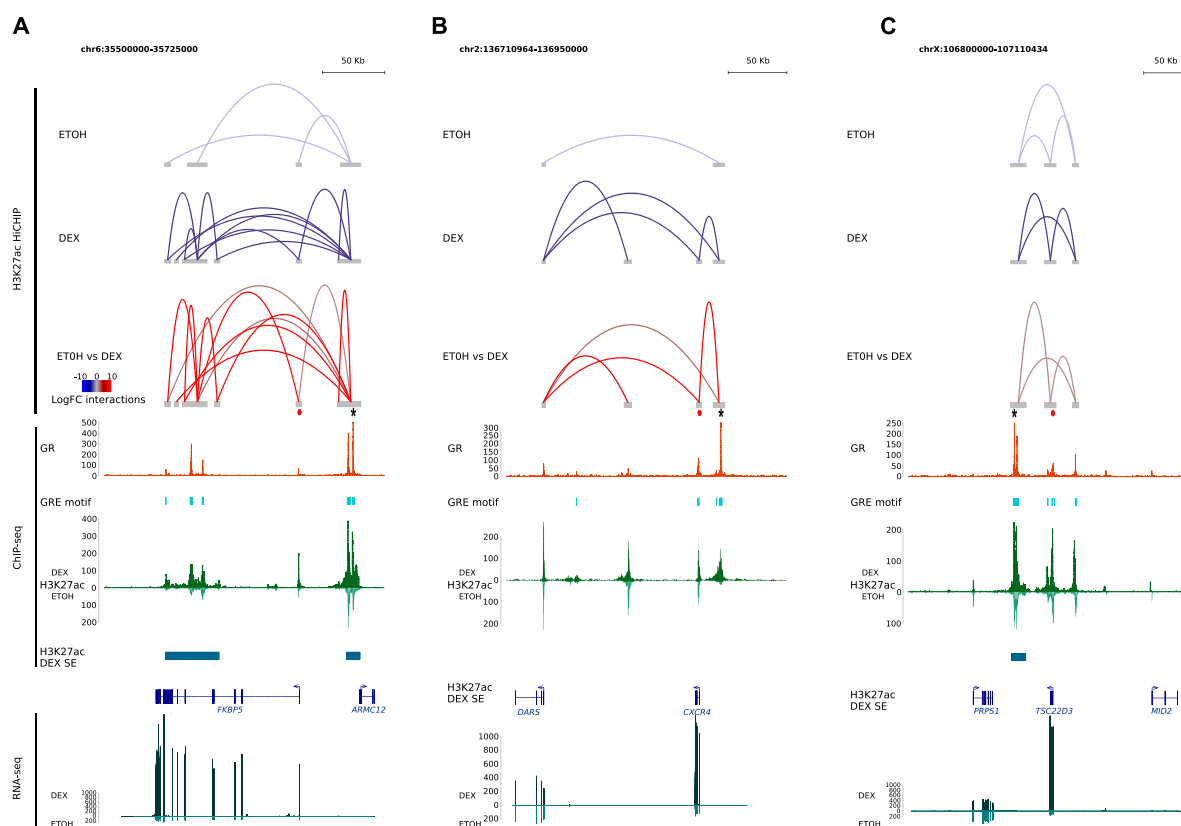


Figure 4. Epigenetic changes in response to Dex at different loci. (A, B, C) Snapshots illustrating the example of GR binding to its consensus motif GR responsive element that increases the H3K27ac chromatin immunoprecipitation sequencing signal as much as the preexistent H3K27ac chromatin interactions for (A) *FKBP5*, (B) *CXCR4*, and (C) *TSC22D3* loci. Predominant enhancer is indicated by a black star, promoter is indicated by a red dot.

GR-binding sites of GRpE and the promoter; though for the latter, it was a small trend (adjusted P -value = 0.11) (Fig 5D top). An enlargement of this region clearly revealed a new significant link between GRpE and promoter upon Dex exposure (Fig 5D bottom). Furthermore, the recruitment of RAD21 at GR-binding sites of these looping partners is consistent with a strengthening of GRpE–promoter interaction at least in a subpopulation of cells.

Together, the results show that interaction of the pE with other regulatory sequences (rE and promoter) depends on the organization of the network. Especially, efficient transcription in response to Dex coincides with synchronization of regulatory sequence openings within the cell population and recruitment of cohesin at GR-binding sites, which reinforces the importance of the CTCF–cohesin couple in the stabilization of enhancer–promoter interactions.

Cell-to-cell transcriptional heterogeneity within myeloma cells after Dex treatment

The importance of synchronized openings suggests that regions that do not open simultaneously may be associated with transcriptional

heterogeneity in the cell population. To test this hypothesis, we performed scRNA-seq assay in MM.1S cells collected at 4 and 24 h in the presence of Dex (0.1 μ M) or ETOH. We focused our analysis on the genes most strongly induced by Dex. Analysis of logfold change distribution permitted to isolate 51 highly induced genes, termed single-cell Dex-activated genes (scDAGs) (Fig S9A). More than half (29/51) were in common with the 55 selected genes involved in an increased proximal–distal interaction upon Dex exposure (Fig 2B; Table S4). To measure the level of cell-to-cell transcriptional variability, we calculated scDAG expression information entropy (Landau et al, 2014; Pastore et al, 2019) and found that, among cells exposed to Dex, the median expression increased, leading to an increased fraction of positive cells toward 1; we hence observed a decrease in entropy, that is, less heterogeneity among the cells. However, gene expression among the cells remained heterogeneous, with a high interquartile range (Fig 6A). To further examine the transcriptional variability that remained after Dex exposure, we analyzed the correlations between scDAGs. Although not very strong, we observed higher pairwise correlation coefficients between scDAGs than between random genes (Fig S9B–D). In addition, scDAGs were clustered

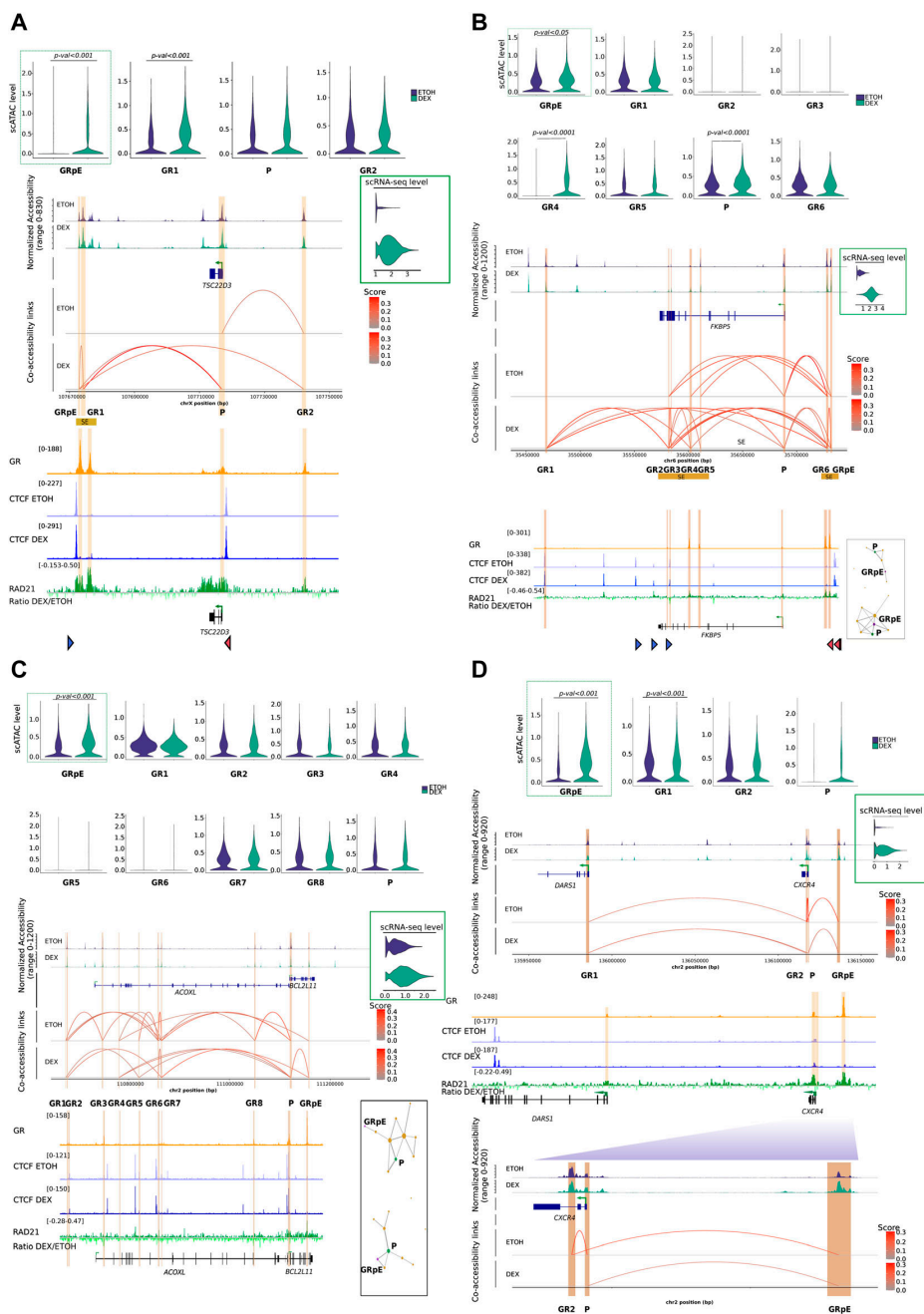


Figure 5. scMultiome sequencing in response to Dex at 4-gene type Dex-responsive shows different mechanisms.

(A) Violin plot for scATAC-seq data in GR enhancers and promoters in control (EtOH) versus treatment (Dex) (top); snapshot of scATAC-seq data and co-accessibility links at *TSC22D3* (middle), and snapshot of chromatin immunoprecipitation (ChIP-seq) for GR, CTCF (EtOH and Dex), and RAD21 (Ratio Dex/EtOH) at the corresponding locus (bottom); triangle represents predicted CTCF motif orientation (blue for forward strand and red for reverse strand); Violin plot of scRNA-seq gene expression level for *TSC22D3* (middle right). (B) Violin plot for scATAC-seq data in GR enhancers and promoters in control (EtOH) versus treatment (Dex) (top); snapshot of scATAC-seq data and co-accessibility links at *FKBP5* (middle) and snapshot of ChIP-seq for GR, CTCF (EtOH and Dex), and RAD21 (Ratio Dex/EtOH) at corresponding locus (bottom); triangle represents predicted CTCF motif orientation (blue for forward strand and red for reverse strand); Violin plot of scRNA-seq gene expression level for *FKBP5* (middle right); co-accessibility network in EtOH and Dex conditions (bottom right). (C) Violin plot for scATAC-seq data in GR enhancers and promoters in control



in two main groups: a large cluster (cluster 1; 45/51 genes), including the proapoptotic gene *BCL2L11* and ubiquitous GR-responsive genes like *TSC22D3*, *FKBP5*, and *DDIT4*, and a second cluster, encompassing six genes, including *CXCR4* (Fig S10A). These results suggest that within Dex-treated cells, two subpopulations coexist: a cell population that predominantly expresses most of scDAGs, referred to as highly Dex-responsive cells, and another population of cells expressing a reduced number of scDAGs. To further explore this possibility, we employed the method recently described by Hoffman et al (2020) to estimate the number of scDAGs expressed in each cell at 4 and 24 h of Dex or EtOH exposure. We found that an EtOH-treated cell had a median background ratio of responding genes (RRG) of 12% (6/51 scDAGs), whereas a Dex-treated cell had a median RRG of 51% (26/51 scDAGs) (Fig 6B), similar to that of 24 h Dex exposure (55%; 21/38 DAGs) (Fig S10B). These results were confirmed by Uniform Manifold Approximation and Projection (UMAP) plots colored by the RRG as a color that revealed an important cell-to-cell heterogeneity among Dex-treated cells. Notably, highly Dex-responsive cells tended to cluster together at the top of the cluster, whereas poorly responsive cells were scattered around (Fig 6C). To know how scDAGs were expressed in the population, we determined the ratio of responding cells (RRC) for each gene. In control condition, the median RRC was 13%, whereas, in the Dex condition, it rose to almost 50% (49%) at 4 h but dropped (37%) at 24 h (Figs 6D and S10C). However, RRC values in the Dex condition were very scattered: if we consider particularly the four genes of interest, the ubiquitous Dex-responsive genes like *FKBP5* and *TSC22D3* exhibited a transcriptional response in almost all cells; in contrast, cell-type specific genes like *BCL2L11* and *CXCR4* were up-regulated in only 38% and 41% of Dex-treated cells, respectively (Figs 6D and S10C). Merged UMAP plots colored according to gene expression of uncorrelated genes *BCL2L11* and *CXCR4* ($R_2 = -0.02$) clearly showed that their expressions were mutually exclusive (Fig 6E). The same applied to *CXCR4* and *TSC22D3* and to *CXCR4* and *FKBP5* Fig S11. Altogether, scRNA-seq analysis revealed that, on average, myeloma cells expressed only half of the overall Dex-responsive genes. In addition, in most of the poorly responding cells (i.e., $RRG < 50\%$), *BCL2L11*, the most important GC-induced death gene in MM, was not expressed.

Discussion

Although the efficacy of Dex in MM can largely be attributed to GR-induced apoptosis, the genomic responses to Dex treatment in malignant plasma cells' genome remain unknown. Given that Dex is used at all stages of treatment, it was crucial to investigate its molecular mode of action by using new genomic tools to better understand treatment escape and provide new insights into combination therapy options. In this study, we confirm the importance of preprogrammed chromatin landscape in guiding most

of GR binding at open and active genomic loci and we show that in plasma cell cellular context, IRF4 is probably the transcriptional factor which cooperates with GR. However, given the suppressive effect of Dex on *IRF4* expression, we cannot rule out the possibility that other factors may be involved after GR binding. As described by Vockley et al (2016), we show that despite a strong association of H3K27ac with GR binding within enhancers engaged in long-range interactions to form lineage-specific networks, the changes in enhancer activity upon Dex exposure are limited and only a small number of genes were deregulated in response to Dex (Fig 1F).

Within some gene-specific regulatory networks, we identified a particular enhancer, referred to as pE, that opens more, interacts more than rEs, and recruits cohesin subunit RAD21 upon GR binding (Fig 3). Targeted genome engineering could be used to directly test the importance of this main enhancer. However, a recent work in acute lymphoblastic leukemia has demonstrated that upon Dex treatment, this specific enhancer promotes an active chromatin interaction with the *BCL2L11* promoter to up-regulate gene transcription (Jing et al, 2018). Similarly, knockdown of a predicted internal enhancer in the *FKBP5* gene locus corresponding to the above GR6 located in the same anchor as the pE induced a 40% reduction in *FKBP5* expression in primary renal proximal tubular epithelial cells (Wilson et al, 2022). We can speculate that activation of both enhancers is essential for a maximal induction of *FKBP5* expression. The activity of critical Dex-regulated enhancers is sensitive to DNA methylation (Jing et al, 2018; Wiench et al, 2011; Wilson et al, 2022), suggesting that inhibition of chromatin accessibility by increased methylation in key regulatory regions as pE may play a critical role in Dex resistance in MM.

We also found that cohesin subunit RAD21 recruitment was associated with Dex-induced GR binding at pEs close to preoccupied CTCF-binding sites, suggesting that CTCF and cohesin are central to mediate stable chromatin loops formed with enhancers activated by GR binding (Fig 5). In addition to stabilizing long-range interactions, association of GR with cohesin complex promotes loop extrusion and long-range gene regulation (Rinaldi et al, 2022). Furthermore, recent results reported the importance of cohesin for regulating the robustness of tissue-specific enhancer-promoter interactions (Aljahani et al, 2022). All these results could explain how GR binding regulates 3D organization of the genome and transcription of both ubiquitous and lineage-specific genes.

It is now established that nucleosome shifts associated with Dex-induced GR binding to both closed nucleosomal or pre-accessible active enhancers require the action of SWI/SNF remodeling complex to regulate gene expression (Johnson et al, 2018). We show that synergistic opening of regulatory sequences within a cell population plays a more important role in transcriptional efficacy than the mere opening of novel regions (Fig 5). Nevertheless, it cannot be excluded that GR at high concentration, induced either by high dose of Dex or prolonged exposure to Dex, may act in a tetrameric state to invade closed chromatin sites and

(EtOH) versus treatment (Dex) (top); snapshot of scATAC-seq data and co-accessibility links at *BCL2L11* (middle), and snapshot of ChIP-seq for GR, CTCF (EtOH and Dex) and RAD21 (Ratio Dex/EtOH) at the corresponding locus (bottom); violin plot of scRNA-seq gene expression level for *BCL2L11* (middle right); co-accessibility network in EtOH and Dex conditions (bottom right). (D) Violin plot for scATAC-seq data in GR enhancers and promoters in control (EtOH) versus treatment (Dex) (top); snapshot of scATAC-seq data and co-accessibility links at *CXCR4* (middle) and snapshot of ChIP-seq for GR, CTCF (EtOH and Dex), and RAD21 (Ratio Dex/EtOH) at corresponding locus (middle); violin plot of scRNA-seq gene expression level for *CXCR4* (middle right); zoom on *CXCR4* locus (bottom).

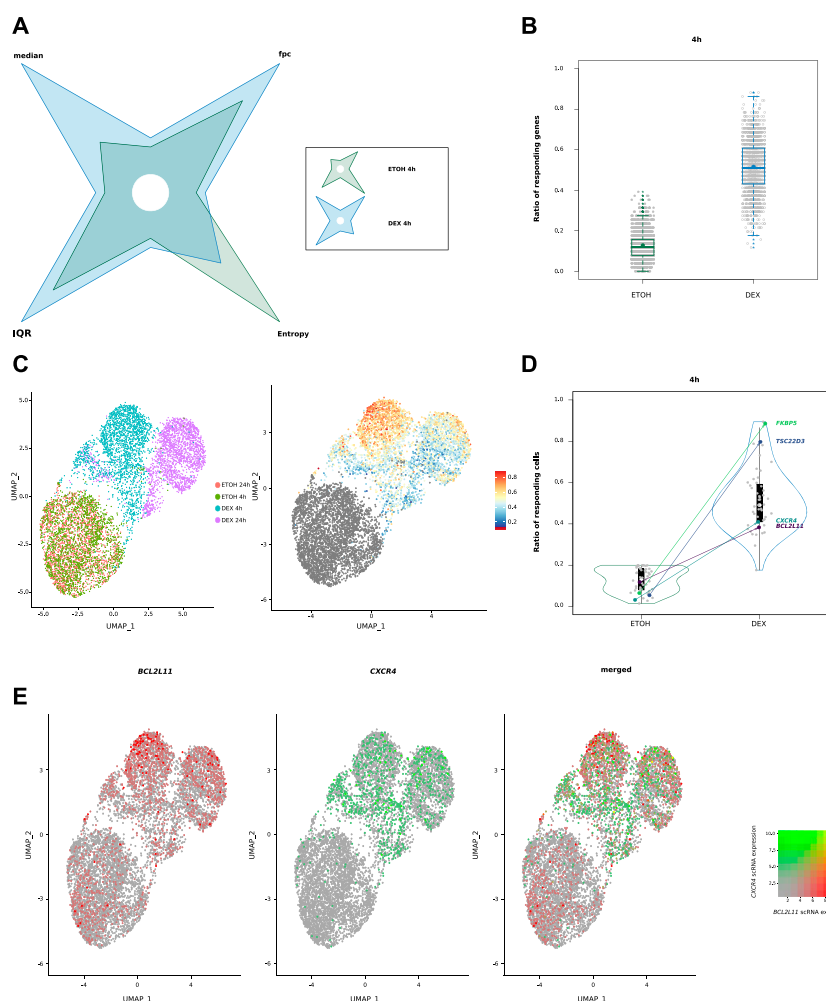


Figure 6. Cell-to-cell transcriptional heterogeneity after Dex exposure.

(A) Shuriken plot illustrating median, interquartile range, fraction of positive cells, and entropy for control and Dex-treated cells; for each parameter, maximum value is used as reference (1; second value is represented as a proportion of this maximum value). (B) Box plots illustrating the ratio of responding genes after 4 h of treatment in control and Dex conditions, with the addition of a dot indicating the mean. (C) Uniform Manifold Approximation and Projection (UMAP) plot of MM.1S cells from scRNA-seq coloured by condition (left) and UMAP plot coloured by the ratio of responding genes for 4 h of exposure (right). (D) Violin plots illustrating ratio of responding cells after 4 h of treatment. (E) UMAP plots coloured for *BCL2L11* expression (red), *CXCR4* (green) expression and merged.

potentially regulate the expression of genes that are not cell-type specific (Paakinaho et al, 2019).

Together, our findings suggest that GR must associate with at least two complexes: SWI/SNF and cohesin to accurately regulate its target genes in myeloma cells.

An important question is whether these multiple enhancer-enhancer and enhancer-promoter interactions at a given gene locus occur in the same cells or not. This is not solved at the moment because of technical limitations. Our work does not answer this question, but provides a computational prediction of the synchronized opening of pairwise anchor sites of interaction loops and gives additional clues in understanding the mechanisms of gene regulation (Fig 5). Our analysis of chromatin architecture, enhancer-promoter interactions, and transcriptional activity in four representative gene loci suggests that Dex-induced GR binding provokes, depending on the gene network, either a rewiring of the promoter-enhancer interactions (Fig S12A), an activation of pE-

promoter interaction (Fig S12B) or a stabilization of the loops inside of a hub (Fig S12C) or not (Fig S12D).

Finally, we showed that these epigenomic changes are associated with a heterogeneous response to Dex in myeloma cells. On average, cells express only 51.6% of scDAGs. Similarly, on average, scDAGs are expressed in only 51.6% of cells after 4 h of treatment, this average decreases slightly to 41.2% at 24 h (Figs 6B and S10B). Expression varies according to genes. In fact, ubiquitous genes like *FKBP5* and *TSC22D3* are expressed in almost all cells, whereas cell-specific genes like *CXCR4* and *BCL2L11* are expressed in less than half of the cells. However, it cannot be ruled out that the differences in expression between the four selected genes may be because of differences in stability or turnover of their mRNAs and cell cycle effect. In myeloma cells, GR binding could lead to a mutually exclusive expression of *BCL2L11* and *CXCR4*, and provides new insights into the mechanisms of drug escape, although considering that GR levels can be a limiting event in Dex treatment (Heuck et al, 2012; Kervoëlen et al, 2015).



Given the potential role of CXCR4 in tumor growth and dissemination (Alsayed et al, 2007; Roccaro et al, 2015), its increased expression upon Dex exposure in a subset of MM.1S cells that do not express the proapoptotic gene *BCL2L1* (Fig 6) raises the provocative possibility that minor populations of myeloma cells could proliferate in response to Dex. In this context, the three-drug combination of a human monoclonal anti-CXCR4 antibody with lenalidomide and Dex or bortezomib and Dex phase Ib/II study demonstrating a high response rate is of particular interest (Ghobrial et al, 2019).

Lastly, we show that IKZF1 and IKZF3 are among the few GR-cobound partners (Table S1), suggesting that these lineage-specific TFs could play a role in Dex response as previously described for the MegaTrans complex in the functionally active estrogen-regulated enhancers (Liu et al, 2014). A recent study showed that these TFs are degraded by IMiDs (Sievers et al, 2018). Because both drugs are combined to treat MM patients, we cannot exclude an antagonistic role of these molecules. Further studies to identify the GR-IKZF1/3 target genes, if any, are warranted.

Materials and Methods

Molecular biology

Cell line culture

MM.1S is a multiple myeloma glucocorticoid-sensitive cell line (ATCC CRL-2974). Cells were cultured in RPMI-1640 supplemented with 10% FBS, and 2 mM L-glutamine. Cell line is tested negative for mycoplasma according to the manufacturer's instructions (PCR Mycoplasma-Test Kit I, A9753; ITW Reagent). Cells were initially cultured for 24 h in reduced-serum, hormone-stripped media (RPMI 1640 medium, no glutamine, no phenol red, 32404014; Thermo Fisher Scientific) with 10% charcoal/dextran-treated FBS (Charcoal STRP FBS One Shot, A3382101; Thermo Fisher Scientific) and 2 mM L-glutamine to a concentration of 1 million cells per mL. Subsequently, Dex (D4902; Sigma-Aldrich) was added to the media at 0.1 μ M for all treatment timepoints and EtOH was used as vehicle control.

ChIP-seq procedure

MM.1S cells were exposed to Dex or EtOH for 1 h and crosslinked with freshly made 1% formaldehyde (28908; Thermo Fisher Scientific) for 15 min and quenched with 125 mM Glycine (50046; Sigma-Aldrich) for 10 min. Cells were pelleted and washed in PBS, then pelleted again and stored at -80°C .

ChIP-seq CTCF and RAD21 was performed as previously described (Jin et al, 2018) with the following modifications. Formaldehyde-fixed cells were lysed and chromatin sheared by sonication using a Bioruptor Pico (Diagenode). IP was carried out using 3 μ g of polyclonal CTCF antibody (C15410210; Diagenode) or using 5 μ g of Anti-Rad21 antibody (ab992; Abcam). DNA from protein-associated complexes and corresponding input samples were washed, eluted, and reversed crosslinking by incubation with RNase A (AM2270; Thermo Fisher Scientific) and protein digested with Proteinase K (25530049; Thermo Fisher Scientific).

Samples were purified with DNA Clean and Concentrator columns (ZD4013; Ozyme) and measured using the Qubit dsDNA HS Kit (Q32851; Thermo Fisher Scientific). Libraries were prepared using NEBNext Ultra II DNA Library Prep according to the manufacturer's instructions (E7103S; New England Biolabs). Libraries were sequenced using Miseq platform (Kit 150cycles V3-PE) with 20 million reads per sample.

ChIP-seq GR (H-300) (sc-8992; Santa Cruz Biotechnology) and H3K27ac (AM-39133; Active Motif) were performed by Active Motif Epigenetic Services. Sequencing depth was 40 million reads for CHIP-seq GR Dex, 38 million reads for CHIP-seq GR EtOH, 28 million reads for CHIP-seq H3K27ac Dex, and 27 million reads for CHIP-seq H3K27ac EtOH.

RNA-seq procedure

MM.1S cells were exposed to Dex or EtOH for 4 h. Total RNA from MM.1S cells was isolated using direct-zol RNA MicroPrep kits (ZR2060; Ozyme) with on-column DNase treatment according to the manufacturer's instructions. Before RNA-seq, RNA quality was confirmed on the Agilent Bioanalyzer 2100 using the RNA 6000 Nano Kit (5067-1511; Agilent). Total RNA-seq libraries were generated using NEBNext Poly(A) mRNA Magnetic Isolation Module (E7490S; New England Biolabs) and NEBNext Ultra II Directional RNA Library Prep (E7765S; New England Biolabs). Libraries were sequenced using the Illumina HiSeq 2500 (HiSeq Rapid SBS kit v2 2*75 cycles).

Fast-ATAC procedure

The Fast-ATAC protocol was performed as previously described (Corces et al, 2016) using 0.1 million cells. MM.1S cells were exposed to Dex or EtOH for 1 h, washed in PBS 1X, and centrifuged. The pellet was resuspended in the transposase reaction mix (25 μ l of 2x TD buffer, 5 μ l of TDE1, 0.5 μ l of 1% digitonin, 19.5 μ l of nuclease-free water) (FC-121-1030; Illumina, G9441; Promega). Transposition reactions were incubated at 37°C for 30 min in an Eppendorf ThermoMixer with agitation at 1,000 rpm. Transposed DNA was purified using the kit "DNA Clean and Concentrator"-5 (ZD4013; Ozyme). Transposed fragments were amplified and purified as described previously (Buenrostro et al, 2015) with Nextera Index Kit (FC-121-1011; Illumina). qPCR was performed to determine the optimal number of cycles to amplify the library to reduce artifacts associated with saturation PCR of complex libraries. PCR was then performed for the optimum number of cycles using the following PCR conditions: 72°C for 5 min; 98°C for 30 s; and thermocycling at 98°C for 10 s, 63°C for 30 s, and 72°C for 1 min. Libraries were amplified for a total of 11 cycles. Library amplification was followed by solid-phase reversible immobilization methodology (SPRI) size selection to exclude fragments larger than 1,200 bp. Libraries were sequenced using Illumina HiSeq 2500 (Rapid Run HiSeq paired-end 2*75 cycles).

HiChIP procedure

MM.1S cells were exposed to Dex or EtOH for 1 h, were pelleted, and resuspended in freshly made 1% formaldehyde (28908; Thermo Fisher Scientific) at a volume of 1 ml of formaldehyde for every one million cells. The cells were incubated at room temperature for 10 min with rotation. Glycine (50046; Sigma-Aldrich) was then added



to a final concentration of 125 mM to quench the formaldehyde. The cells were incubated at room temperature for 5 min with rotation. The cells were pelleted and washed in PBS, then pelleted again, and stored at -80°C .

The HiChIP protocol was performed as previously described (Mumbach et al, 2016) using 7.5 μg antibody to H3K27ac (C15410196; Diagenode) with the following modifications. Samples were sheared using Bioruptor Pico (Diagenode), the amount of Tn5 (15027865; Illumina) used and number of PCR cycles performed were based on the post-ChIP Qubit amounts. Libraries were sequenced on NovaSeq 6000 (NovaSeq 6000 S1 Reagent Kit 2*100 cycles).

Single-cell RNA-seq procedure

For scRNA-seq, MM.1S cells were exposed to Dex or EtOH for 4 and 24 h. Single-cell RNA-seq profiling was performed with the Chromium™ Single Cell Controller. A total of 6,000 cells was loaded per lane and processed for complementary DNA synthesis and library preparation, per the manufacturer's protocol using 3' v3.1 chemistry (10X Genomics—1000121). Libraries were sequenced on NovaSeq 6000 (NovaSeq 6000 S1 Reagent Kit 2*100 cycles) to a mean depth of 45,000 reads/cell using the read lengths 26 bp Read1, 8 bp i7 Index, 98 bp Read2.

Single-cell multiome ATAC + gene expression procedure

For scMultiome, MM.1S cells were exposed to Dex or EtOH for 1 and 4 h. Single-cell 3' gene expression and open chromatin libraries were simultaneously generated using Chromium Next GEM Single-Cell Multiome ATAC + Gene Expression Kit from 10x Genomics, following the protocol provided by the manufacturer. A total of 5,000 nuclei were loaded per lane on the Chromium™ Single-Cell Controller. Libraries were sequenced on NovaSeq 6000 (NovaSeq 6000 SP Reagent Kit 2*50 cycles) to a minimum depth of 24,000 reads/nucleus for Gene Expression library and 42,000 reads/nucleus for ATAC library.

RIME

RIME GR (H-300) (sc-8992; Santa Cruz Biotechnology) and IRF4 (sc-6059; Active Motif) were performed by Active Motif Epigenetic Services. MM.1S cells were exposed to Dex (GR) or EtOH (IRF4) for 1 h and fixed according to the manufacturer's instructions (RIME Cell Fixation protocol, Active Motif). Analyses were performed by an active motif and results are given as a supplementary table (see Table S1). Identified proteins in the IgG negative control were removed and background proteins with a spectral count < 5 . The final list is composed of proteins identified in both replicates; their spectral count is the average of spectral count in both replicates.

Cell protein extraction and fractionation procedure

MM.1S cells were exposed to Dex or EtOH for 1 h. We used NE-PER Nuclear and Cytoplasmic Extraction reagent (78833; Thermo Fisher Scientific) to obtain cytoplasmic and nuclear protein fractions based on the vendor's instructions. Nuclear protein fractions were desalted with Zeba™ Spin Desalting Columns (89882; Thermo Fisher Scientific). Protein was quantified with BC Assay Protein Quantitation Kit (UP40840; Interchim).

Co-IP

5 μg of anti-IRF4 Antibody (F-4) agarose conjugate (sc-48338; SantaCruz Biotechnology) was added to 50 μg of desalted nuclear protein, the mixture was incubated overnight with mixing at $2-8^{\circ}\text{C}$. The immunocomplexes were collected by centrifugation at 1,000g for 5 min at 4°C , washed twice with cold PBS, and resuspended in an electrophoresis sample buffer, 2X (sc-24945; Santa Cruz Biotechnology).

Western blotting

Proteins extracted from cells or isolated by Co-IP were separated in 4–20% SurePAGE, Bis-Tris Gels (M00655; GeneScript) at 200 V for 30 min, and then transferred onto a nitrocellulose membrane at 150 mA for 2 h. Immunoblotted proteins on the membrane, labelled with specific antibodies were imaged by autoradiography. All primary and secondary antibodies were used according to the manufacturer's instructions. The primary antibodies used were: IRF4 (3E4) (#646402; BioLegend); GR-HRP (sc-393232; Santa Cruz Biotechnology); β -actin (MAB8929; R&D systems); lamin A/C (E-1) (sc-376248; Santa Cruz Biotechnology).

Computational analysis

Chromatin state annotation

To obtain functional annotation of MM.1S cell line, we used ChromHMM (v1.11) (Ernst & Kellis, 2010, 2012). Five histone marks available from ENCODE consortium (ENCODE Project Consortium, 2012) (H3K4me1, H3K4me3, H3K27ac, H3K36me3, and H3K27me3) in three different cell lines (MM.1S, U266, and GM12878) were analyzed using hidden Markov model to identify 10 different chromatin states. Default parameters of chromHMM were used. Bam files were binarized into 200-bp genomic windows and the presence or absence of each histone mark was evaluated. Then, we employed biological analysis to annotate those chromatin states giving them biological meanings.

Treatment of ChIP-seq data

ChIP-seq sequencing quality was assessed with fastqc (v0.11.8) (Andrews, 2010). ChIP-seq read adaptors were firstly trimmed using trimmomatic (v0.39) (Bolger et al, 2014) and then reads were mapped using bowtie2 (v2.1.0) (Langmead et al, 2009) to the Human genome UCSC hg19 (GRCh37) (Kent et al, 2002). Only one mismatch was allowed. After alignment step, unmapped reads, low-quality mapped reads (mapQ < 30), and reads mapped to ENCODE blacklist regions (Amemiya et al, 2019) were removed with samtools (v1.3.1) (Li et al, 2009) for analysis. We also removed reads that were like to be optical and/or PCR duplicates using picard MarkDuplicates (v2.23.5) from GATK (McKenna et al, 2010).

ChIP-seq-enriched regions defined as peaks were called using macs2 (v2.1.1) (Zhang et al, 2008) versus input (sequencing without immunoprecipitation). We only retained peaks higher than specified P -value threshold ($P\text{-val} < 1 \times 10^{-7}$).

Treatment of ATAC-seq data

All ATAC-seq data were processed based on Kundaje laboratory-proposed pipelines (Koh et al, 2016; Liu et al, 2019) available on github. Quality of sequencing assessment, read adaptor



trimming, read mapping to human genome hg19, and filtration were performed the same way as ChIP-seq reads. Before peak-calling steps, and because of the Tn5 insertion, mapped reads were shifted with, respectively, 5 and 4 bp for strand + and strand - with samtools. Finally, enriched regions defined as ATAC-seq peaks were called using macs2 only significant peaks were retained (FDR < 0.05).

RNA-seq differential analysis

Each RNA-seq sample was mapped using Tophat2 (Trapnell et al, 2009) versus hg19 reference genome. We then employed the proposed protocol (Trapnell et al, 2012) to perform differential expression analysis with cufflinks. Only genes with a LogFC greater than or equal to 0.6 and an FDR < 0.05 were kept for analysis.

HiChIP data treatment and differential analysis

We employed HiC-pro (Servant et al, 2015) to process HiChIP data from raw data to normalized contact maps. All reads were mapped to hg19 genome using bowtie2 (global parameters: --very-sensitive -L 30 --score-min L, -0.6, -0.2 --end-to-end --reorder; local parameters: --very-sensitive -L 20 --score-min L, -0.6, -0.2 --end-to-end --reorder). Contact maps were generated at different resolutions (1, 2, 15, 20, and 40 kb) and normalized by the iterative correction and eigenvector decomposition method. HiC-pro output directory was then used as input to hichipper (Lareau & Aryee, 2018) with MboI restriction site position for loop calling. Differential analysis of chromatin loops was performed with function exactTest of package edgeR (Robinson et al, 2010), with default parameters except for dispersion, which was set to "trended." Interactions with FDR below 5% and absolute logFC above 0.60 were considered significant.

Global treatment of genomic data

Genomic data were proceeded using different genomic tools such as Bedtools (v2.28.0) (Quinlan & Hall, 2010) for manipulating genomic files, the homer suite for annotation, and motif scanning (v4.4) (Heinz et al, 2010). Data were also treated using own Python scripts (v2.7).

Motif search

De novo motif discovery was performed using the MEME suite (v4.11.2) (Bailey et al, 2015) for GR peaks centralized on peak submit and extend with 250 bp in both directions. Motif from 6 to 16 bp were searched with a maximum of five motifs were asked. To identify sequences where a specific motif is found, we employed FIMO tool from MEME. Finally, to identify a centrally enriched motif, we used centriMo from MEME.

Signal track generation

We employed the bamCoverage tool from the Deeptools (v2.0) (Ramírez et al, 2014) suite to generate bigWig files. Signal track files were normalized using Read Per Genomic Content (RPGC) method also known as the 1X normalization included in bamCoverage options. Once those files were generated, we used the bigwig-Compare tool to create a differential track between H3K27ac with or without Dex. All ChIP-seq and ATAC-seq files were generated with this method. Visualization of signal tracks was obtained using the Integrative Genome Viewer IGV (Robinson et al, 2011).

Genome ontology analysis

Genome ontology analysis was performed using GREAT (v3.0.0) (McLean et al, 2010) with default parameters (gene regulatory domain: prox. 5 kb upstream and 1 kb downstream; dist. up to 1,000 kb). Enrichment statistics were computed using binomial and hypergeometric gene-based tests. Pathways were selected as significantly enriched if the false discovery rate (FDR q-value) was lower than 0.01.

Differential analysis of ChIP-seq H3K27ac peaks

To find H3K27ac ChIP-seq responding to GR binding, we first selected all H3K27ac peaks found within GR peaks (n = 16,228). On those sites, we then estimated the normalized count (RPGC) of H3K27ac ChIP-seq in both conditions. Log₂fold changes were then calculated for each site and we consider H3K27ac Dex increased all sites with a log₂FC higher than 0.1. The H3K27ac Dex-increased peaks are given in the supplementary table (see Table S2).

Identification of the GRpE among each regulatory network of Dex-responsive genes

It has been shown that GR-binding sites with effects on expression are activated and enriched for the GR-binding motif (Vockley et al, 2016; McDowell et al, 2018). Inspecting more particularly specific networks: for example, TSC22D3 and BCL2L1 (Fig S6A and B), it appeared that activated enhancers with the GR-binding motif also displayed a strong GR signal, with the value within the 1% highest values on the whole genome. Based on these findings, we defined pE as follows: pE had to overlap with the presence of a GR motif and have a not-negative Dex/EtOH log₂ ratio for H3K27ac signal; among the peaks satisfying those conditions, the one with the highest GR signal value was selected, provided that this value was at least 80% of the maximum GR signal for the region (see Table S4).

Differential analysis of ATAC-seq peaks found within chromatin loop anchors

We collected all ATAC-seq peaks found within chromatin loop anchors and, for each peak, we estimated the RPGC count of ATAC-seq in EtOH and Dex conditions. Log₂FC was then estimated and all ATAC-seq peaks with a LogFC greater than or equal to 0.6 were considered as ATAC up. The ATAC Dex-increased peaks obtained are given as supplementary table (see Table S3).

CTCF motif orientation-based loop prediction

To predict which CTCF ChIP-seq peaks could potentially create DNA contact, we employed the algorithm and scripts proposed by Oti et al (2016) on our own CTCF ChIP-seq data in MM.1S cell line to define predicted CTCF loops based on motif orientation.

Global treatment of single-cell data

Preprocessing steps for single-cell data were done using Cell Ranger Software suite, respectively, cellranger (v5.0.0) (Zheng et al, 2017) and cellranger-arc (v1.0.1) for scRNA-seq and scMultiome-seq (Satpathy et al, 2019). For both types of data, the hg38 genome assembly provided by 10xGenomics was used



for alignment. Further analyses were performed on R (v3.6). For scRNAseq, Count matrices were loaded into R using the Seurat package (v3.9.9) (Satija et al, 2015). For each cell, we calculated the percentage of mitochondrial reads (percent.mt) and the percentage of nuclear-retained lncRNA (percent.nc). We also used the CellCycleScoring function from Seurat to assign a cell cycle state to each cell (Phase); the assignment of the cell cycle state is based on the S. score and G2M.score calculated by this function. Cells were then filtered on the following criteria: $5 < \text{percent.mt} < 25$, $\text{percent.nc} < 10$, a minimum of 2,000 reads, and 1,500 different genes expressed. Normalization was performed using the Seurat NormalizeData function with standard parameters. The function FindVariableFeatures was then used to select the 3,000 most variable features, those features have been scaled with Seurat ScaledData function, because cell cycle was a major part of the variability, we added S.score and G2M.score to the vars. to.regress argument of the function. We reduced dimension using RunPCA from Seurat; only the first 30 dimensions were used for downstream analyses. We also calculated a 2D embedding of our cells with RunUMAP; neighbor search and clustering were performed using FindNeighbors and FindClusters functions, with default parameters. For scMultiome-seq, RNA and ATAC matrices were loaded into R using Seurat and Signac (v1.1.0) (Stuart et al, 2020 Preprint) packages. For each cell, we calculated the percent. mt, percent. nc and the Phase. For the ATAC data, we also calculated TSS score and the ratio of reads overlapping with the blacklisted regions of the genome contained in the blacklist hg38 unified provided by the Signac package. Cells detected by cellranger were filtered on both RNA and ATAC data. For RNA data, we kept cells between 3,800 and 150,000 reads and more than 2,000 different genes expressed. We also kept cells with a percent. mt between 5 and 30 and a percent. nc lower than 8. For ATAC data, we kept cells with a number of reads between 10,000 and 500,000 and a number of different features between 5,000 and 60,000. We also filtered cells with a TSS enrichment between 3.5 and 15, a nucleosome signal lower than 1.5 and more than 50% reads in peaks. For normalization and dimensionality reduction, we used the RunTFIDF and RunSVD functions from the Signac packages. RunSVD was run on the features selected by FindTopFeatures with min.cutoff set to q80. UMAP embedding, neighbor search, and clustering were performed the same way as for the RNA data alone. Differential expression and accessibility were tested using the findMarkers function provided by Seurat, with test. use argument, respectively, set to "MAST" and "LR."

Assessment of scATA-seq peaks co-accessibility

Co-accessibility scores between scATACseq peaks were calculated using Cicero (v1.3.4.11) (Pliner et al, 2018) with default parameters. Co-accessibility tables were built on each treatment condition separately. From those tables of co-accessibility scores, we built two networks for each of the 34 Dex-responsive genes GRpEs on one hand and GRrEs on other hand. Nodes of those networks were defined as all peaks overlapping with the selected regions. Edges were built using the co-accessibility tables, considering only connections with a score higher than 0.1.

scDAGs

We studied the distribution of logFC above 0 and found out it was bimodal with a small part of positive logFC being far from the main part. We then used Gaussian mixture model to identify the small subpopulation of high logFC genes, that is, scDAGs.

RRG

For each gene, we computed the 80th percentile of expressed values in untreated cells; we then calculated, for each cell, the RRG as the percentage of scDAGs with expression value above the gene threshold for untreated cells, 4 h of Dex-treated cells, and 24 h Dex-treated cells.

Data Availability

ChIP-seq, HiChIP-seq, ATAC-seq, RNA-seq, scRNAseq, and scMultiome have been deposited at the European Genome-phenome Archive (EGA, <https://www.ebi.ac.uk/ega>), which is hosted by the EBI and the CRG, under dataset accession EGAD00001011136, EGAD00001011138, EGAD00001011135, EGAD00001011137, EGAD00001011139, and EGAD00001011140, respectively.

Supplementary Information

Supplementary Information is available at <https://doi.org/10.26508/lsa.202302195>.

Acknowledgements

We thank the Genomics and Bioinformatics core facility of Nantes (Geno-BIRD, Biogenouest, IFB) for its technical support. We thank the Fondation Française Pour la Recherche contre le Myélome et les gammopathies monoclonales (FFRMG), the Programme d'investissements d'Avenir I-SITE NexT (ANR-16-IDEX-0007) the Pays de la Loire, the SIRIC ILIAD (INCA-DGOS-Inserm-12558), and Celgene for supporting this study.

Author Contributions

V Bessonneau-Gaborit: data curation, formal analysis, and writing—original draft, review, and editing.
 J Cruard: data curation, formal analysis, and writing—original draft, review, and editing.
 C Guerin-Charbonnel: data curation, formal analysis, and writing—original draft, review, and editing.
 J Derrien: data curation, formal analysis, and writing—original draft, review, and editing.
 J-B Alberge: data curation.
 E Douillard: data curation.
 M Devic: data curation.
 S Deshayes: data curation.
 L Campion: formal analysis.
 F Westermann: data curation.
 P Moreau: data curation and funding acquisition.



C Herrmann: data curation, formal analysis, and writing—original draft, review, and editing.
 J Bourdon: data curation, formal analysis, and writing—original draft, review, and editing.
 F Magrangeas: conceptualization, data curation, formal analysis, supervision, investigation, methodology, and writing—original draft, review, and editing.
 S Minvielle: conceptualization, supervision, validation, investigation, methodology, project administration, and writing—original draft, review, and editing.

Conflict of Interest Statement

The authors declare that they have no conflict of interest.

References

Aljahani A, Hua P, Karpinska MA, Quililan K, Davies JO, Oudelaar AM (2022) Analysis of sub-kilobase chromatin topology reveals nano-scale regulatory interactions with variable dependence on cohesin and CTCF. *Nat Commun* 13: 2139. doi:10.1038/s41467-022-29696-5

Alsayed Y, Ngo H, Runnels J, Leleu X, Singha UK, Pitsillides CM, Spencer JA, Kimlinger T, Ghobrial JM, Jia X, et al (2007) Mechanisms of regulation of *cxc4/sdf-1* (*cxcl12*)-dependent migration and homing in multiple myeloma. *Blood* 109: 2708–2717. doi:10.1182/blood-2006-07-035857

Amemiya HM, Kundaje A, Boyle AP (2019) The encode blacklist: Identification of problematic regions of the genome. *Sci Rep* 9: 9354–9355. doi:10.1038/s41598-019-45839-z

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. In *Babraham Bioinformatics*. Cambridge, UK: Babraham Institute.

Bailey TL, Johnson J, Grant CE, Noble WS (2015) The meme suite. *Nucleic Acids Res* 43: 39–49. doi:10.1093/nar/gkv416

Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, Miranda TB, Sung M-H, Trump S, Lightman SL, et al (2011) Transcription factor ap1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell* 43: 145–155. doi:10.1016/j.molcel.2011.06.016

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for illumina sequence data. *Bioinformatics* 30: 2114–2120. doi:10.1093/bioinformatics/btu170

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ (2015) Atac-seq: A method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* 109: 21–29. doi:10.1002/0471142727.mb2129s109

Corces MR, Buenrostro JD, Wu B, Greenside PG, Chan SM, Koenig JL, Snyder MP, Pritchard JK, Kundaje A, Greenleaf WJ, et al (2016) Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet* 48: 1193–1203. doi:10.1038/ng.3646

D'Ippolito AM, McDowell IC, Barrera A, Hong LK, Leichter SM, Bartelt LC, Vockley CM, Majoros WH, Safi A, Song L, et al (2018) Pre-established chromatin interactions mediate the genomic response to glucocorticoids. *Cell Syst* 7: 146–160.e7. doi:10.1016/j.cels.2018.06.007

ENCODE Project Consortium (2012) An integrated encyclopedia of dna elements in the human genome. *Nature* 489: 57–74. doi:10.1038/nature11247

Ernst J, Kellis M (2010) Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* 28: 817–825. doi:10.1038/nbt.1662

Ernst J, Kellis M (2012) ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods* 9: 215–216. doi:10.1038/nmeth.1906

Fletcher TM, Xiao N, Mautino G, Baumann CT, Wolford R, Warren BS, Hager GL (2002) ATP-dependent mobilization of the glucocorticoid receptor during chromatin remodeling. *Mol Cell Biol* 22: 3255–3263. doi:10.1128/mcb.22.10.3255-3263.2002

Ghobrial IM, Vij R, Siegel D, Badros A, Kaufman J, Raje N, Jakubowiak A, Savona MR, Obreja M, Berdeja JG (2019) A phase Ib/II study of oprozomib in patients with advanced multiple myeloma and waldenström macroglobulinemia. *Clin Cancer Res* 25: 4907–4916. doi:10.1158/1078-0432.ccr-18-3728

Grøntved L, John S, Baek S, Liu Y, Buckley JR, Vinson C, Aguilera G, Hager GL (2013) C/ebp maintains chromatin accessibility in liver and facilitates glucocorticoid receptor recruitment to steroid response elements. *EMBO J* 32: 1568–1583. doi:10.1038/emboj.2013.106

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell* 38: 576–589. doi:10.1016/j.molcel.2010.05.004

Heuck CJ, Szymonifka J, Hansen E, Shaughnessy JD, Usmani SZ, Van Rhee F, Anaissie E, Nair B, Waheed S, Alsayed Y, et al (2012) Thalidomide in total therapy 2 overcomes inferior prognosis of myeloma with low expression of the glucocorticoid receptor gene *nr3c1*. *Clin Cancer Res* 18: 5499–5506. doi:10.1158/1078-0432.ccr-12-0019

Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-André V, Sigova AA, Hoke HA, Young RA (2013) Super-enhancers in the control of cell identity and disease. *Cell* 155: 934–947. doi:10.1016/j.cell.2013.09.053

Hoffman JA, Pappas BN, Trotter KW, Archer TK (2020) Single-cell rna sequencing reveals a heterogeneous response to glucocorticoids in breast cancer cells. *Commun Biol* 3: 126. doi:10.1038/s42003-020-0837-0

Huang J, Li K, Cai W, Liu X, Zhang Y, Orkin SH, Xu J, Yuan GC (2018) Dissecting super-enhancer hierarchy based on chromatin interactions. *Nat Commun* 9: 943. doi:10.1038/s41467-018-03279-9

Jin Y, Chen K, De Paepe A, Hellqvist E, Krstic AD, Metang L, Gustafsson C, Davis RE, Levy YM, Surapaneni R, et al (2018) Active enhancer and chromatin accessibility landscapes chart the regulatory network of primary multiple myeloma. *Blood* 131: 2138–2150. doi:10.1182/blood-2017-09-808063

Jing D, Huang Y, Liu X, Sia KC, Zhang JC, Tai X, Wang M, Toscan CE, McCalmont H, Evans K, et al (2018) Lymphocyte-specific chromatin accessibility pre-determines glucocorticoid resistance in acute lymphoblastic leukemia. *Cancer Cell* 34: 906–921.e8. doi:10.1016/j.ccell.2018.11.002

John S, Sabo PJ, Thurman RE, Sung M-H, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* 43: 264–268. doi:10.1038/ng.759

Johnson TA, Chereji RV, Stavreva DA, Morris SA, Hager GL, Clark DJ (2018) Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. *Nucleic Acids Res* 46: 203–214. doi:10.1093/nar/gkx1044

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at ucsc. *Genome Res* 12: 996–1006. doi:10.1101/gr.229102

Kervoelen C, Ménoret E, Gomez-Bougie P, Bataille R, Godon C, Marionneau-Lambot S, Moreau P, Pellat-Deceunynck C, Amiot M (2015) Dexamethasone-induced cell death is restricted to specific molecular subgroups of multiple myeloma. *Oncotarget* 6: 26922–26934. doi:10.18632/oncotarget.4616

Koh PW, Sinha R, Barkal AA, Morganti RM, Chen A, Weissman IL, Ang LT, Kundaje A, Loh KM (2016) An atlas of transcriptional, chromatin accessibility, and surface marker changes in human mesoderm development. *Sci Data* 3: 160109. doi:10.1038/sdata.2016.109

- Krönke J, Udeshi ND, Narla A, Grauman P, Hurst SN, McConkey M, Svinkina T, Heckl D, Comer E, Li X, et al (2014) Lenalidomide causes selective degradation of ikzf1 and ikzf3 in multiple myeloma cells. *Science* 343: 301–305. doi:10.1126/science.1244851
- Kuznetsova T, Wang S-Y, Rao NA, Mandoli A, Martens JH, Rother N, Aartse A, Wang L, Li S, et al (2015) Glucocorticoid receptor and nuclear factor kappa-b affect three-dimensional chromatin organization. *Genome Biol* 16: 264. doi:10.1186/s13059-015-0832-9
- Landau DA, Clement K, Ziller MJ, Boyle P, Fan J, Gu H, Stevenson K, Sougnez C, Wang L, Li S, et al (2014) Locally disordered methylation forms the basis of intratumor methylome variation in chronic lymphocytic leukemia. *Cancer Cell* 26: 813–825. doi:10.1016/j.ccell.2014.10.012
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short dna sequences to the human genome. *Genome Biol* 10: R25. doi:10.1186/gb-2009-10-3-r25
- Lareau CA, Aryee MJ (2018) Hichipper: a preprocessing pipeline for calling DNA loops from HiChIP data. *Nat Methods* 15: 155–156. doi:10.1038/nmeth.4583
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and samtools. *Bioinformatics* 25: 2078–2079. doi:10.1093/bioinformatics/btp352
- Liu Z, Merkurjev D, Yang F, Li W, Oh S, Friedman MJ, Song X, Zhang F, Ma Q, Ohgi KA, et al (2014) Enhancer activation requires trans-recruitment of a mega transcription factor complex. *Cell* 159: 358–373. doi:10.1016/j.cell.2014.08.027
- Liu C, Wang M, Wei X, Wu L, Xu J, Dai X, Xia J, Cheng M, Yuan Y, Zhang P, et al (2019) An atac-seq atlas of chromatin accessibility in mouse tissues. *Sci Data* 6: 65. doi:10.1038/s41597-019-0071-0
- Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA (2013) Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153: 320–334. doi:10.1016/j.cell.2013.03.036
- Low MSY, Brodie EJ, Fedele PL, Liao Y, Grigoriadis G, Strasser A, Kallias A, Willis SN, Tellier J, Shi W, et al (2019) IRF4 activity is required in established plasma cells to regulate gene transcription and mitochondrial homeostasis. *Cell Rep* 29: 2634–2645. doi:10.1016/j.celrep.2019.10.097
- McDowell IC, Barrera A, D'ippolito AM, Vockley CM, Hong LK, Leichter SM, Bartelt LC, Majoros WH, Song L, Safi A, et al (2018) Glucocorticoid receptor recruits to enhancers and drives activation by motif-directed binding. *Genome Res* 28: 1272–1284. doi:10.1101/gr.233346.117
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al (2010) The genome analysis toolkit: A mapreduce framework for analyzing next-generation dna sequencing data. *Genome Res* 20: 1297–1303. doi:10.1101/gr.107524.110
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) Great improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 28: 495–501. doi:10.1038/nbt.1630
- Moreau P, Attal M, Hulin C, Arnulf B, Belhadj K, Benboubker L, Béné MC, Broijl A, Caillon H, Caillot D, et al (2019) 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. *Lancet* 394: 29–38. doi:10.1016/s0140-6736(19)31240-1
- Moreau P, Garfall AL, van de Donk NW, Nahi H, San-Miguel JF, Oriol A, Nooka AK, Martin T, Rosinol L, Chari A, et al (2022) Teclistamab in relapsed or refractory multiple myeloma. *N Engl J Med* 387: 495–505. doi:10.1056/nejmoa2203478
- Mumbach MR, Rubin AJ, Flynn RA, Dai C, Khavari PA, Greenleaf WJ, Chang HY (2016) Hichip: Efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 13: 919–922. doi:10.1038/nmeth.3999
- Mumbach MR, Satpathy AT, Boyle EA, Dai C, Gowen BG, Cho SW, Nguyen ML, Rubin AJ, Granja JM, Kazane KR, et al (2017) Enhancer connectome in primary human cells identifies target genes of disease-associated dna elements. *Nat Genet* 49: 1602–1612. doi:10.1038/ng.3963
- Oh K-S, Patel H, Gottschalk RA, Lee WS, Baek S, Fraser ID, Hager GL, Sung M-H (2017) Anti-inflammatory chromatin landscape suggests alternative mechanisms of glucocorticoid receptor action. *Immunity* 47: 298–309.e5. doi:10.1016/j.immuni.2017.07.012
- Oti M, Falck J, Huynen MA, Zhou H (2016) CTCF-mediated chromatin loops enclose inducible gene regulatory domains. *BMC Genomics* 17: 252. doi:10.1186/s12864-016-2516-6
- Paakinaho V, Johnson TA, Presman DM, Hager GL (2019) Glucocorticoid receptor quaternary structure drives chromatin occupancy and transcriptional outcome. *Genome Res* 29: 1223–1234. doi:10.1101/gr.244814.118
- Pastore A, Gaiti F, Lu SX, Brand RM, Kulm S, Chaligne R, Gu H, Huang KY, Stamenova EK, Béguelin W, et al (2019) Corrupted coordination of epigenetic modifications leads to diverging chromatin states and transcriptional heterogeneity in cll. *Nat Commun* 10: 1874. doi:10.1038/s41467-019-09645-5
- Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, Srivatsan S, Qiu X, Jackson D, Minkina A, et al (2018) Cicero predicts cis-regulatory dna interactions from single-cell chromatin accessibility data. *Mol Cell* 71: 858–871.e8. doi:10.1016/j.molcel.2018.06.044
- Quinlan AR, Hall IM (2010) Bedtools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26: 841–842. doi:10.1093/bioinformatics/btq033
- Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T (2014) DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* 42: 187–191. doi:10.1093/nar/gku365
- Reddy TE, Pauli F, Sprouse RO, Neff NF, Newberry KM, Garabedian MJ, Myers RM (2009) Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation. *Genome Res* 19: 2163–2171. doi:10.1101/gr.097022.109
- Rinaldi L, Fettweis G, Kim S, Garcia DA, Fujiwara S, Johnson TA, Tettey TT, Ozbun L, Pegoraro G, Puglia M, et al (2022) The glucocorticoid receptor associates with the cohesin loader NIPBL to promote long-range gene regulation. *Sci Adv* 8: eabj8360. doi:10.1126/sciadv.abj8360
- Robinson MD, McCarthy DJ, Smyth GK (2010) Edger: A bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26: 139–140. doi:10.1093/bioinformatics/btp616
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP (2011) Integrative genomics viewer. *Nat Biotechnol* 29: 24–26. doi:10.1038/nbt.1754
- Roccaro AM, Mishima Y, Sacco A, Moschetta M, Tai Y-T, Shi J, Zhang Y, Reagan MR, Huynh D, Kawano Y, et al (2015) Cxcr4 regulates extra-medullary myeloma through epithelial-mesenchymal-transition-like transcriptional activation. *Cell Rep* 12: 622–635. doi:10.1016/j.celrep.2015.06.059
- Satija R, Farrell JA, Gennert D, Schier AF, Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 33: 495–502. doi:10.1038/nbt.3192
- Satpathy AT, Granja JM, Yost KE, Qi Y, Meschi F, McDermott GP, Olsen BN, Mumbach MR, Pierce SE, Corces MR, et al (2019) Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral t cell exhaustion. *Nat Biotechnol* 37: 925–936. doi:10.1038/s41587-019-0206-z
- Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E (2015) Hic-pro: An optimized and flexible pipeline for hi-c data processing. *Genome Biol* 16: 259. doi:10.1186/s13059-015-0831-x



- Shaffer AL, Emre NCT, Lamy L, Ngo VN, Wright G, Xiao W, Powell J, Dave S, Yu X, Zhao H, et al (2008) Irf4 addiction in multiple myeloma. *Nature* 454: 226–231. doi:[10.1038/nature07064](https://doi.org/10.1038/nature07064)
- Sievers QL, Petzold G, Bunker RD, Renneville A, Stabicki M, Liddicoat BJ, Abdulrahman W, Mikkelsen T, Ebert BL, Thomä NH (2018) Defining the human c2h2 zinc finger degrome targeted by thalidomide analogs through crbn. *Science* 362: eaat0572. doi:[10.1126/science.aat0572](https://doi.org/10.1126/science.aat0572)
- Stavreva DA, Coulon A, Baek S, Sung M-H, John S, Stixova L, Tesikova M, Hakim O, Miranda T, Hawkins M, et al (2015) Dynamics of chromatin accessibility and long-range interactions in response to glucocorticoid pulsing. *Genome Res* 25: 845–857. doi:[10.1101/gr.184168.114](https://doi.org/10.1101/gr.184168.114)
- Stuart T, Srivastava A, Lareau C, Satija R (2020) Multimodal single-cell chromatin analysis with Signac. *BioRxiv*. doi:[10.1101/2020.11.09.373613](https://doi.org/10.1101/2020.11.09.373613) (Preprint posted November 10, 2020).
- Trapnell C, Pachter L, Salzberg SL (2009) Tophat: Discovering splice junctions with rna-seq. *Bioinformatics* 25: 1105–1111. doi:[10.1093/bioinformatics/btp120](https://doi.org/10.1093/bioinformatics/btp120)
- Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of rna-seq experiments with tophat and cufflinks. *Nat Protoc* 7: 562–578. doi:[10.1038/nprot.2012.016](https://doi.org/10.1038/nprot.2012.016)
- Trotter KW, Archer TK (2004) Reconstitution of glucocorticoid receptor-dependent transcription in vivo. *Mol Cell Biol* 24: 3347–3358. doi:[10.1128/mcb.24.8.3347-3358.2004](https://doi.org/10.1128/mcb.24.8.3347-3358.2004)
- Trotter KW, Fan HY, Ivey ML, Kingston RE, Archer TK (2008) The HSA domain of BRG1 mediates critical interactions required for glucocorticoid receptor-dependent transcriptional activation in vivo. *Mol Cell Biol* 28: 1413–1426. doi:[10.1128/mcb.01301-07](https://doi.org/10.1128/mcb.01301-07)
- Vian L, Pełowska A, Rao SS, Kieffer-Kwon KR, Jung S, Baranello L, Huang SC, El Khattabi L, Dose M, Pruett N, et al (2018) The energetics and physiological impact of cohesin extrusion. *Cell* 173: 1165–1178.e20. doi:[10.1016/j.cell.2018.03.072](https://doi.org/10.1016/j.cell.2018.03.072)
- Vockley CM, D'Ippolito AM, McDowell IC, Majoros WH, Safi A, Song L, Crawford GE, Reddy TE (2016) Direct gr binding sites potentiate clusters of tf binding across the human genome. *Cell* 166: 1269–1281.e19. doi:[10.1016/j.cell.2016.07.049](https://doi.org/10.1016/j.cell.2016.07.049)
- Wiench M, John S, Baek S, Johnson TA, Sung MH, Escobar T, Simmons CA, Pearce KH, Biddie SC, Sabo PJ, et al (2011) DNA methylation status predicts cell type-specific enhancer activity. *EMBO J* 30: 3028–3039. doi:[10.1038/emboj.2011.210](https://doi.org/10.1038/emboj.2011.210)
- Wilson PC, Muto Y, Wu H, Karihaloo A, Waikar SS, Humphreys BD (2022) Multimodal single cell sequencing implicates chromatin accessibility and genetic background in diabetic kidney disease progression. *Nat Commun* 13: 5253. doi:[10.1038/s41467-022-32972-z](https://doi.org/10.1038/s41467-022-32972-z)
- Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al (2008) Model-based analysis of chip-seq (macs). *Genome Biol* 9: 137. doi:[10.1186/gb-2008-9-9-r137](https://doi.org/10.1186/gb-2008-9-9-r137)
- Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, et al (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8: 14049. doi:[10.1038/ncomms14049](https://doi.org/10.1038/ncomms14049)



License: This article is available under a Creative Commons License (Attribution 4.0 International, as described at <https://creativecommons.org/licenses/by/4.0/>).

4.2 Étude du transcriptome en cellule unique du Myélome Multiple et son environnement immunitaire

4.2.1 Le jeu de données

Nous avons utilisé la technologie de scRNA-seq afin de reconstituer la composition de la moelle de l'ensemble de nos échantillons. Le jeu de données comprend 59 échantillons, dont 8 issus de moelle normale (Normal Bone Marrow en anglais) (NBM). Les 51 échantillons restants sont issus de 34 patients atteints de MM. Parmi ces 34 patients, 11 ont subi un prélèvement au stade NDMM. Pour 12 patients, plusieurs prélèvements sont disponibles (figure 4.4 4.5). Parmi nos prélèvements, les différents stades de la maladie sont représentés. De plus, certains échantillons ont été prélevés lors du suivi des patients (Follow-up-MM : FUMM). De ce fait, ces échantillons comportent très peu de plasmocytes.

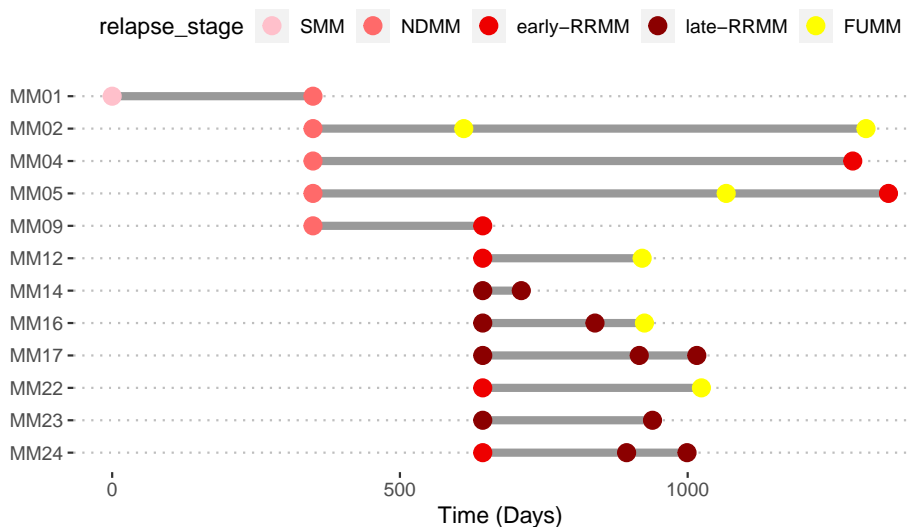


FIGURE 4.4 – Frise chronologique des prélèvements, seuls les patients avec plusieurs prélèvements sont présentés

Après les contrôles qualité et l'agrégation de l'ensemble des échantillons, nous obtenons un jeu de données de 240 833 cellules. Après traitement des données, comme attendu, la UMAP finale obtenue montre une nette séparation entre les plasmocytes malins de chaque patient, tandis que les types cellulaires de l'environnement se regroupent par type cellulaire large (figure 4.6). Les plasmocytes représentent 97 027 cellules (40%) et l'environnement immunitaire 143 806 (60%).

ID_article	ID_patient	relapse_stage	Condition	disease_status	sampling_date							
NBM01	NBM01	NBM	Control	NBM	01/06/2021	RRMM14_2	MM14	LRRMM	Control	RRMM		17/08/2020
NBM02	NBM02	NBM	Control	NBM	01/06/2021	RRMM14_2	MM14	LRRMM	BH3	RRMM		17/08/2020
NBM03	NBM03	NBM	Control	NBM	01/06/2021	RRMM15	MM15	LRRMM	Control	RRMM		27/10/2020
NBM04	NBM04	NBM	Control	NBM	13/10/2021	RRMM16	MM16	LRRMM	Control	RRMM		21/07/2020
NBM05	NBM05	NBM	Control	NBM	13/10/2021	RRMM16	MM16	LRRMM	BH3	RRMM		21/07/2020
NBM06	NBM06	NBM	Control	NBM	13/10/2021	RRMM16_2	MM16	LRRMM	Control	RRMM		01/02/2021
NBM07	NBM07	NBM	Control	NBM	20/10/2021	RRMM16_2	MM16	LRRMM	BH3	RRMM		01/02/2021
NBM08	NBM08	NBM	Control	NBM	20/10/2021	FUMM16	MM16	FUMM	Control	FUMM		28/04/2021
SMM01	MM01	SMM	Control	SMM	05/11/2019	FUMM16	MM16	FUMM	BH3	FUMM		28/04/2021
SMM01	MM01	SMM	BH3	SMM	05/11/2019	RRMM17	MM17	LRRMM	Control	RRMM		11/12/2019
NDMM01	MM01	NDMM	Control	NDMM	19/10/2020	RRMM17	MM17	LRRMM	BH3	RRMM		11/12/2019
NDMM01	MM01	NDMM	BH3	NDMM	19/10/2020	RRMM17_2	MM17	LRRMM	Control	RRMM		08/09/2020
NDMM02	MM02	NDMM	Control	NDMM	04/03/2019	RRMM17_2	MM17	LRRMM	BH3	RRMM		08/09/2020
NDMM02	MM02	NDMM	BH3	NDMM	04/03/2019	RRMM17_3	MM17	LRRMM	Control	RRMM		17/12/2020
FUMM02	MM02	FUMM	Control	FUMM	21/11/2019	RRMM17_3	MM17	LRRMM	BH3	RRMM		17/12/2020
FUMM02_2	MM02	FUMM	Control	FUMM	20/10/2021	RRMM18	MM18	ERRMM	Control	RRMM		14/01/2020
FUMM02_2	MM02	FUMM	BH3	FUMM	20/10/2021	RRMM18	MM18	ERRMM	BH3	RRMM		14/01/2020
NDMM03	MM03	NDMM	Control	NDMM	06/03/2019	RRMM19	MM19	LRRMM	Control	RRMM		14/01/2020
NDMM03	MM03	NDMM	BH3	NDMM	06/03/2019	RRMM19	MM19	LRRMM	BH3	RRMM		14/01/2020
NDMM04	MM04	NDMM	Control	NDMM	20/03/2019	RRMM20	MM20	ERRMM	Control	RRMM		12/03/2020
NDMM04	MM04	NDMM	BH3	NDMM	20/03/2019	RRMM20	MM20	ERRMM	BH3	RRMM		12/03/2020
RRMM04	MM04	ERRMM	Control	RRMM	13/10/2021	RRMM21	MM21	ERRMM	Control	RRMM		06/08/2020
RRMM04	MM04	ERRMM	BH3	RRMM	13/10/2021	RRMM21	MM21	ERRMM	BH3	RRMM		06/08/2020
NDMM05	MM05	NDMM	Control	NDMM	21/03/2019	RRMM22	MM22	ERRMM	Control	RRMM		01/09/2020
NDMM05	MM05	NDMM	BH3	NDMM	21/03/2019	RRMM22	MM22	ERRMM	BH3	RRMM		01/09/2020
FUMM05	MM05	FUMM	Control	FUMM	08/03/2021	FUMM22	MM22	FUMM	Control	FUMM		16/09/2021
FUMM05	MM05	FUMM	BH3	FUMM	08/03/2021	FUMM22	MM22	FUMM	BH3	FUMM		16/09/2021
RRMM05	MM05	ERRMM	Control	RRMM	15/12/2021	RRMM23	MM23	LRRMM	Control	RRMM		21/09/2020
RRMM05	MM05	ERRMM	BH3	RRMM	15/12/2021	RRMM23	MM23	LRRMM	BH3	RRMM		21/09/2020
NDMM06	MM06	NDMM	Control	NDMM	29/07/2019	RRMM23_2	MM23	LRRMM	Control	RRMM		13/07/2021
NDMM07	MM07	NDMM	Control	NDMM	05/08/2019	RRMM23_2	MM23	LRRMM	BH3	RRMM		13/07/2021
NDMM07	MM07	NDMM	BH3	NDMM	05/08/2019	RRMM24	MM24	ERRMM	Control	RRMM		01/10/2020
NDMM08	MM08	NDMM	Control	NDMM	04/09/2019	RRMM24_2	MM24	LRRMM	Control	RRMM		08/06/2021
NDMM08	MM08	NDMM	BH3	NDMM	04/09/2019	RRMM24_3	MM24	LRRMM	Control	RRMM		21/09/2021
NDMM09	MM09	NDMM	Control	NDMM	30/10/2019	RRMM25	MM25	ERRMM	Control	RRMM		01/10/2020
NDMM09	MM09	NDMM	BH3	NDMM	30/10/2019	RRMM25	MM25	ERRMM	BH3	RRMM		01/10/2020
RRMM09	MM09	ERRMM	Control	RRMM	20/08/2020	RRMM26	MM26	ERRMM	Control	RRMM		13/10/2020
RRMM09	MM09	ERRMM	BH3	RRMM	20/08/2020	RRMM27	MM27	LRRMM	Control	RRMM		09/12/2020
NDMM10	MM10	NDMM	Control	NDMM	21/01/2020	RRMM28	MM28	LRRMM	Control	RRMM		14/12/2020
NDMM10	MM10	NDMM	BH3	NDMM	21/01/2020	RRMM29	MM29	LRRMM	Control	RRMM		04/01/2021
NDMM11	MM11	NDMM	Control	NDMM	04/03/2020	RRMM30	MM30	LRRMM	Control	RRMM		12/01/2021
NDMM11	MM11	NDMM	BH3	NDMM	04/03/2020	RRMM31	MM31	ERRMM	Control	RRMM		02/04/2021
RRMM12	MM12	ERRMM	Control	RRMM	04/02/2021	RRMM31	MM31	ERRMM	BH3	RRMM		02/04/2021
RRMM12	MM12	ERRMM	BH3	RRMM	04/02/2021	RRMM32	MM32	LRRMM	Control	RRMM		03/06/2021
FUMM12	MM12	FUMM	Control	FUMM	08/11/2021	RRMM32	MM32	LRRMM	BH3	RRMM		03/06/2021
RRMM13	MM13	LRRMM	Control	RRMM	07/01/2020	RRMM33	MM33	LRRMM	Control	RRMM		09/07/2021
RRMM14	MM14	LRRMM	Control	RRMM	11/06/2020	RRMM33	MM33	LRRMM	BH3	RRMM		09/07/2021
RRMM14	MM14	LRRMM	BH3	RRMM	11/06/2020	RRMM34	MM34	ERRMM	Control	RRMM		01/12/2020

FIGURE 4.5 – Tableau des échantillons issus de la cohorte MYRACLE

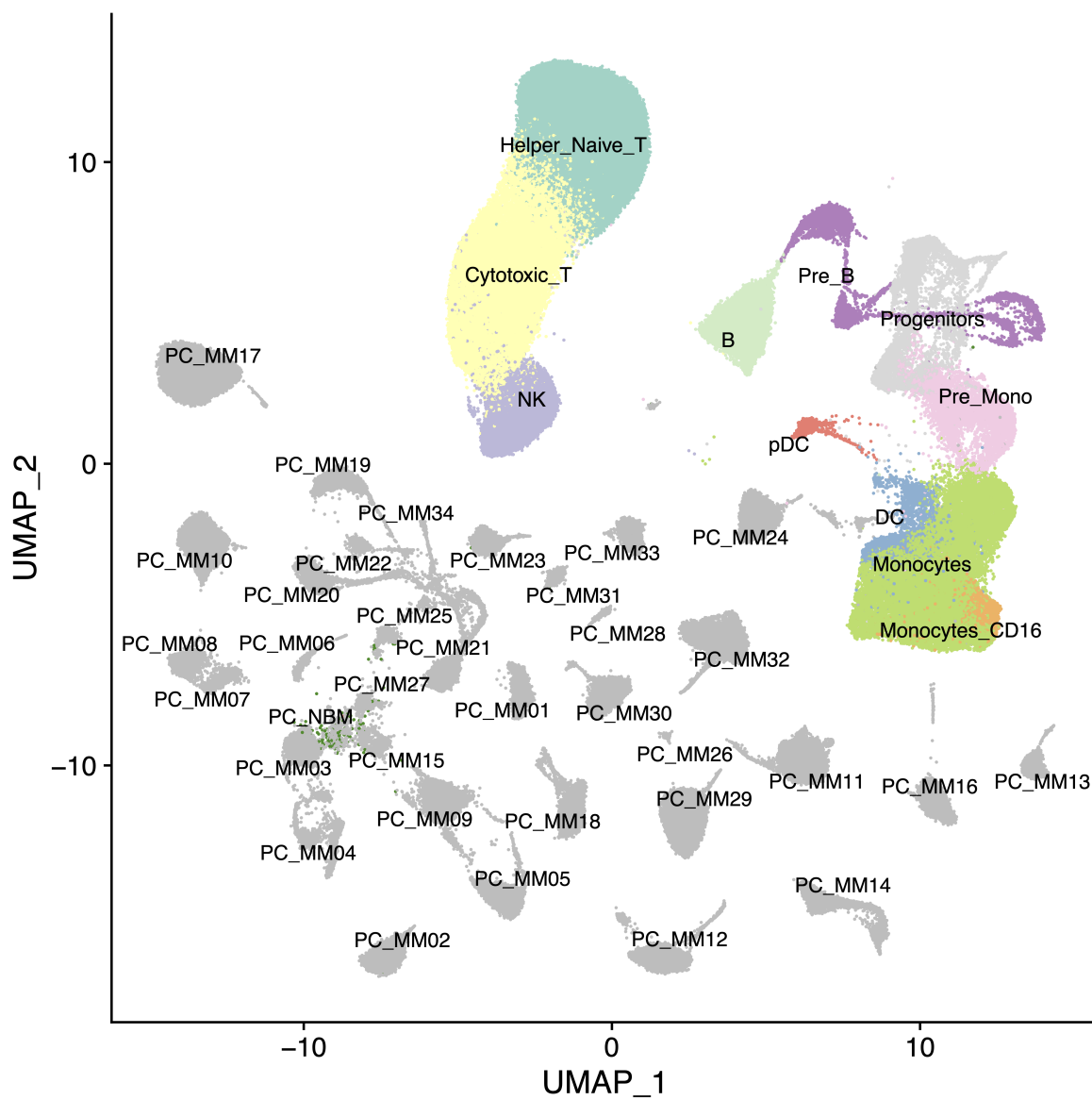


FIGURE 4.6 – UMAP du dataset complet, environnement et tumeur, l'environnement est coloré par type cellulaire large, les plasmocytes sont colorés en gris, les labels correspondent au type cellulaire ou numéro de patient pour les tumeurs

4.2.2 L'environnement immunitaire

Après cette première étape, nous avons isolé l'environnement et l'avons alors retraité séparément. Dans ce cas, nous avons cependant dû effectuer une étape d'intégration des données, afin de prendre en compte un fort effet "batch" introduit par les différentes versions du kit chromium utilisées pour la génération des données.

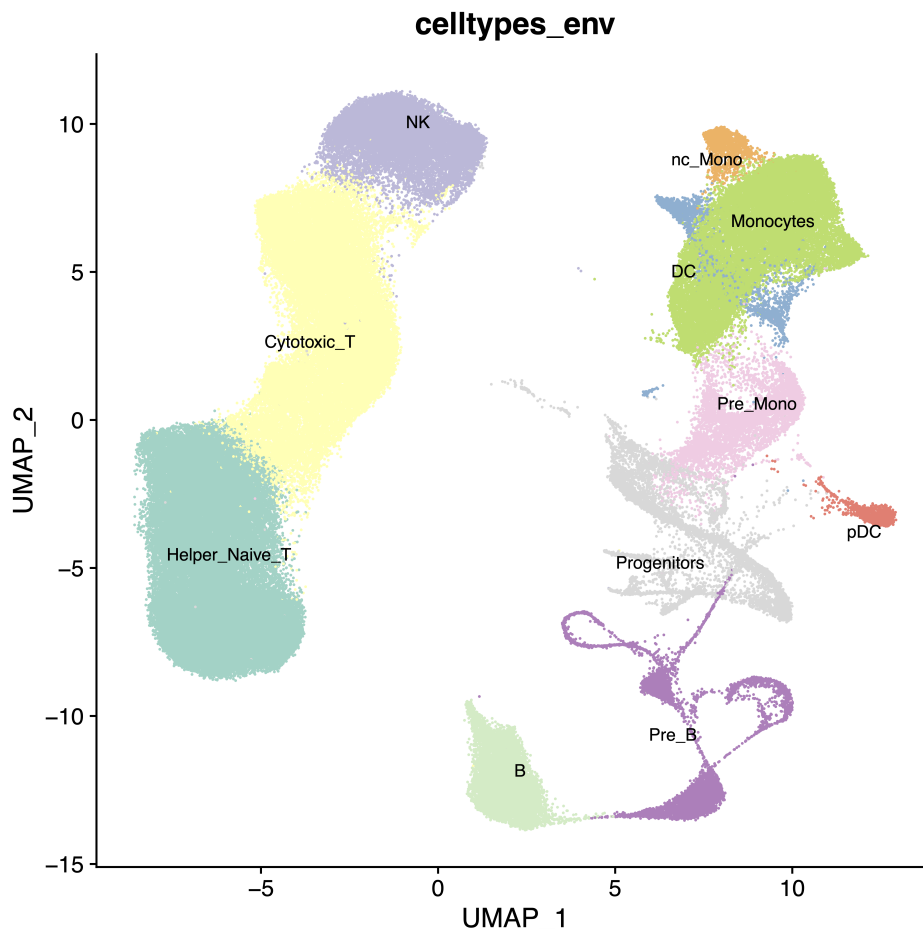


FIGURE 4.7 – Réduction UMAP de l'ensemble de l'environnement, les cellules sont colorées par type cellulaire large

Nous avons ensuite pu annoter les types cellulaires larges présents. Cette étape d'annotation est basée sur les résultats de clustering. Nous avons assigné les identités des clusters sur la base des gènes les plus différenciellement exprimés (figure 4.8) et sur l'expression de gènes marqueurs connus 4.9. Nous avons établi une classification en 11 types cellulaires majeurs (figure 4.7).

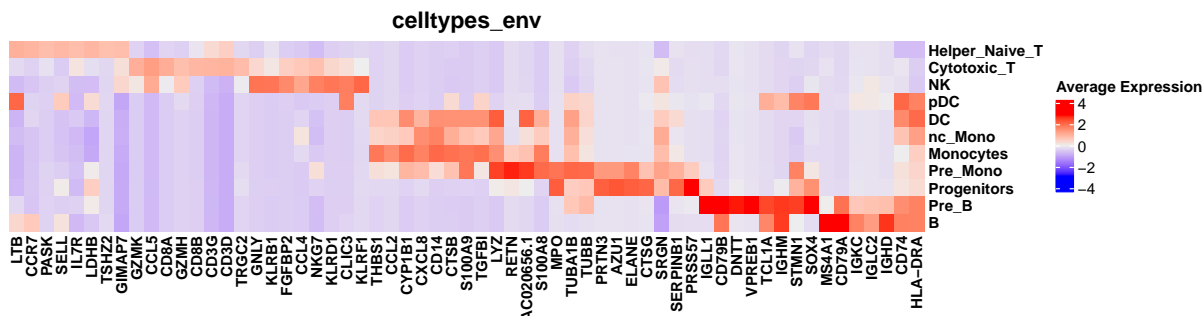


FIGURE 4.8 – Heatmap récapitulant l’expression moyenne des gènes marqueurs de nos populations

broad cell types	broad cell type markers	refined cell types	refined cell type markers
NK	CD3Dneg, cytotoxicity, CD56	CD56_bright_NK	NCAM1high, XCL1, HLA, SELL
		trNK	CXCR6
		Adaptive_NK	CD3D+/-, KLRC2, HLA
		Mature_NK	cytotoxicity
		Active_Inflamed_NK	CD69
		Intermediate_NK	intermediate cytotoxicity and NCAM1
Helper_Naive_T	CD3D, CCR7, SELL, IL7R	iNF_NK	IFN signature
		CD8_Naive	CD8A, CD8B, SELL, CCR7
		CD4_Naive	CD4, SELL, CCR7
		CD4_Mem	CD4, IL7R
		Treg	CD4, FOXP3
		CD8_Tc2	GATA3, KRT1
Cytotoxic_T	CD3D, cytotoxicity	MAIT	CD8A, CD8B, KLRB1
		CD8_Mem	CD8A, IL7R
		CD8_Mem_effector	CD8A, CD8B, GZMK
		IFN_T	IFN signature
		CD8_cytotoxic	CD3D, cytotoxicity (GZMB, GZMH)
		gdT	CD3D, TRDC, TRGC1, cytotoxicity
		NKT	KLRC2, cytotoxicity
Pre_Mono	RETN, MPO	Pre_Mono	RETN, MPO, cyclin
Monocytes	CD14	Inter_Mono	CD14, VCAN, RETN
		Normal_Macrophages	CSTB, SPINK1
		TAM_M2like_angio	CD14, LILRB1, LILRB2, FCGR3Alow, IL1B
		TAM_M2like_FCGR3A	CD14, FCGR3A, CD36, FPR3, FN1
		TAM_C1Q	SELENOP, C1Q, MHCII
		TAM_Metalo	Metallothionein genes
		IFN_Mono	IFN signature
nc_Mono	CD14, FCGR3A	nc_Mono	CD14, FCGR3Ahigh
pDC	CLEC4C, GZMB, IRF8	pDC	CD74, CLEC4C, GZMB, IRF8, MHCII
DC	CD74, FCER1A	cDC	CD74, MHCII
		iDO1_cDC	CD74, MHCII, IDO1, CCL22
Progenitors	CD34	Progenitors	CD34
B	MS4A1	B	MS4A1
Pre_B	IGLL1, RAG1, RAG2, DNIT	Pre_B	IGLL1, RAG1, RAG2, DNIT

FIGURE 4.9 – Tableau des types cellulaires et des marqueurs qui leur sont associés

4.2.3 Composition de l’environnement immunitaire

La comparaison des proportions représentées par ces types cellulaires au sein des échantillons nous a permis d’identifier des changements de composition significatifs de l’environnement aux différents stades de la maladie (figure 4.10 et 4.11).

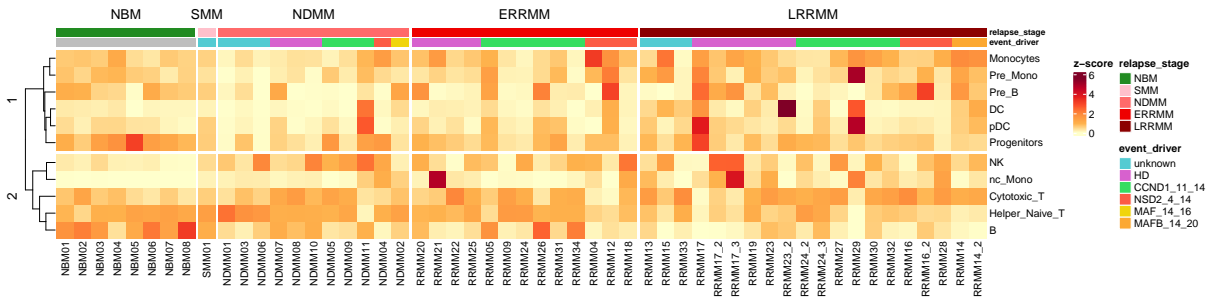


FIGURE 4.10 – Heatmap de l’abondance relative des principaux types cellulaires au sein de nos échantillons, les échantillons sont regroupés par stade de la maladie et par type d’évènement primaire

Au stade NDMM, les B, les pré-B, les progéniteurs et les pré-monocytes sont moins représentés que dans les NBM. À l’inverse, on observe un enrichissement en Monocytes non conventionnel (ncMono) et NK. Lorsqu’on compare les différents stades du MM (NDMM, ERRMM et LRRMM), on remarque que la majorité des changements de composition significatifs apparaissent au stade LRRMM, bien que des tendances similaires soient observées au stade ERRMM. Dans les LRRMM, les T helper et naïfs (groupe Helper_Naive_T) ainsi que les B sont déplétés, comparé aux NDMM. En revanche, on observe une augmentation des proportions de monocytes, pré-monocytes, cellules dendritiques (DC) et cellules dendritiques plasmacitoïdes (pDC). Dans le cas des pré-monocytes, il est intéressant de souligner qu’ils étaient diminués au NDMM.

Ces résultats montrent que des changements significatifs de la composition de l’environnement existent dès le diagnostic et qu’ils se poursuivent et s’amplifient avec l’accumulation des lignes de traitement. Il faut aussi souligner qu’il existe une importante hétérogénéité de représentation des types cellulaires entre les échantillons. En effet, les NK, bien qu’ils soient enrichis significativement au diagnostic dans certains échantillons, le niveau est similaire voir inférieur à la moyenne des NBM. Pour les T cytotoxiques, cette hétérogénéité est plus flagrante encore puisque nous avons des valeurs largement supérieures et inférieures aux valeurs extrêmes des NBM aux stades RRMM. Au diagnostic, pour les types cellulaires définis de manière comparable, les résultats sont en accord avec les observations précédemment établies par ZAVIDIJ et al. 2020. Les différences observées

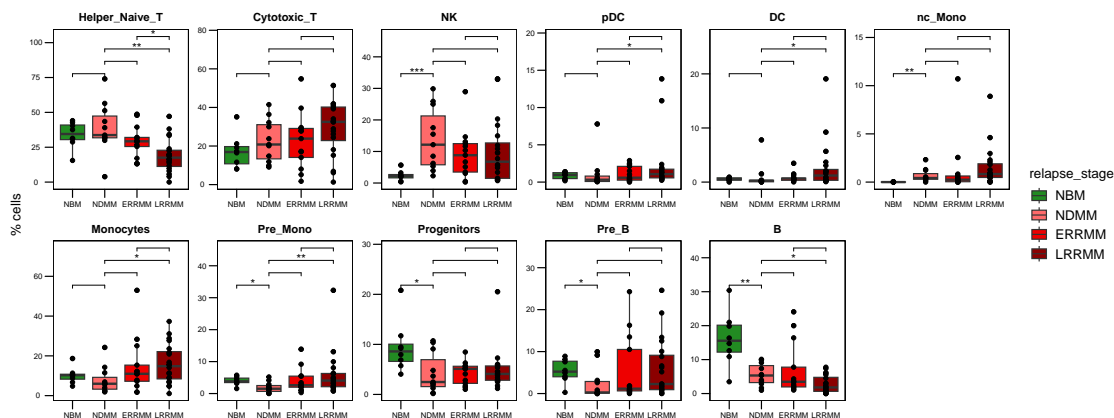


FIGURE 4.11 – Boxplots du pourcentage de cellules représenté par chaque type cellulaire au sein des échantillons, regroupés par stade

s’expliquent cependant par les différences de design entre nos études. En effet, dans cet article, les stades précurseurs MGUS et SMM sont également étudiés.

Comme nous l’avons évoqué dans l’introduction de cette thèse, il existe une grande diversité de sous-types cellulaires au sein du système immunitaire. Pour mieux rendre compte de cette diversité, nous avons réanalysé les principaux lignages cellulaires séparément.

4.2.4 Types cellulaires affinés

L’identification et l’annotation des sous-types cellulaires ont été réalisées sur le même principe que celle des types cellulaires principaux : à l’aide du clustering, de marqueurs connus et des gènes différentiellement exprimés dans les clusters (table 4.9 et figure 4.12).

Nous avons regroupé les deux populations de T pour l’analyse : Helper_Naive_T et Cytotoxic_T. En parallèle, nous avons regroupé les membres de la lignée myéloïde : pré-monocytes (Pre_Mono), Monocytes, nc_Mono, DC et pDC. Et enfin, nous avons reclustérisé les NK séparément. Avec cette approche, nous avons pu définir 12 populations de T, 12 populations myéloïdes et 7 populations de NK, ce qui donne au total 34 types cellulaires, en incluant les Progéniteurs, les pré-B et les B.

Ces 34 populations expriment des marqueurs permettant de les rattacher à un type cellulaire décrit dans la littérature. Cependant, au sein de chacun de nos trois compartiments, une population se démarque par une forte expression des gènes induits en réponse

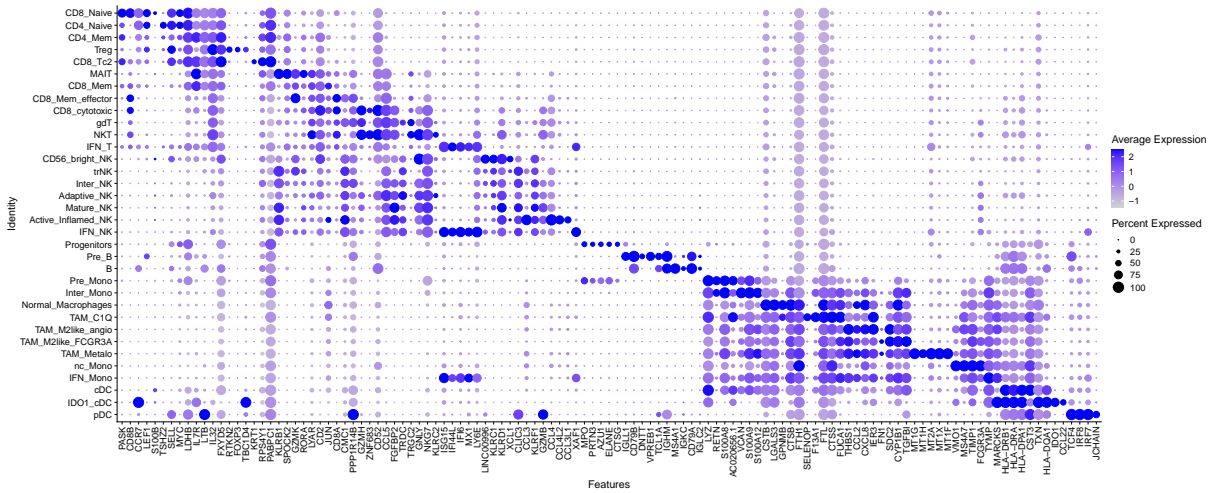
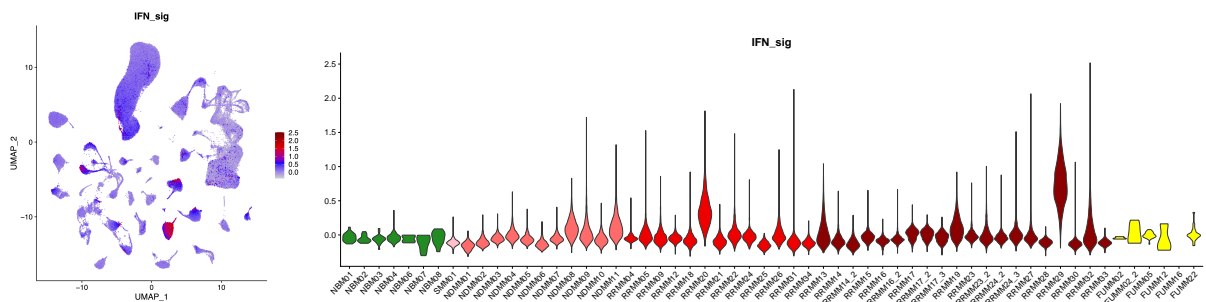


FIGURE 4.12 – Dotplot des principaux marqueurs de chaque type cellulaire fin

à l'interféron : *ISG15*, *IFI44L*, *IFI6*, *MX1*, *LY6E* notamment (figure 4.12). Ce module occupe une part importante de leur transcriptome, ce qui augmente l'effet de "dropout" et nous empêche d'observer l'expression des gènes caractéristiques de leur sous-types. Ce cluster est donc a priori constitué de différents types cellulaires qu'il nous est impossible de discriminer en raison de la signature interféron.

Ces trois populations (IFN_T, IFN_NK, IFN_Mono), comprennent des cellules provenant d'un nombre très limité d'échantillons. Ces échantillons sont principalement regroupés dans les clusters 3 et 4 de la heatmap (figure 4.14) Par ailleurs, chez ces patients, la signature interféron est également exprimée par la tumeur (figure 4.13).

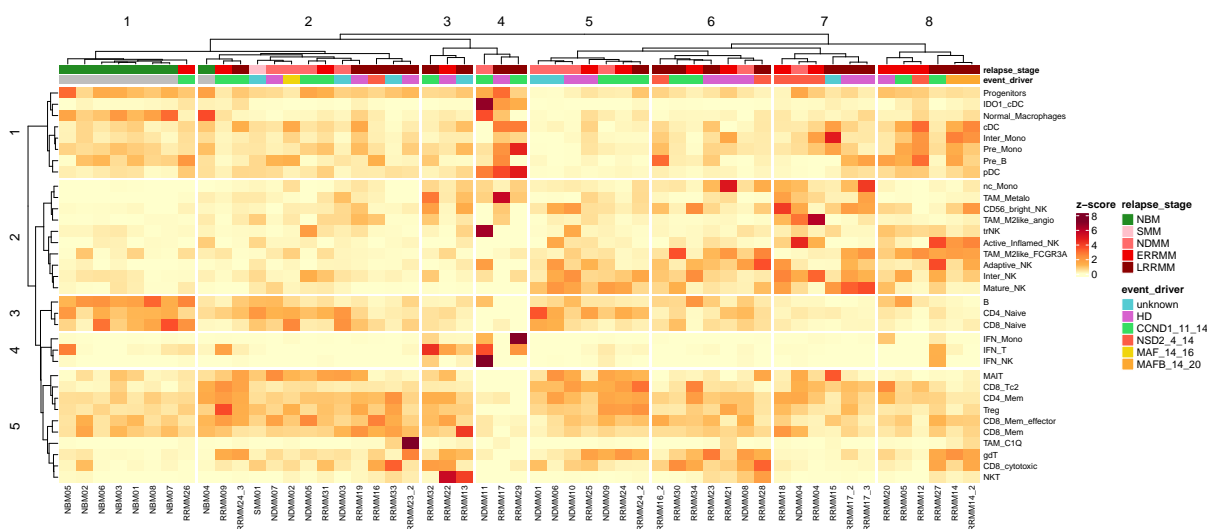


(a) UMAP globale signature IFN (b) Violins du niveau d'expression de la signature interféron au sein des tumeurs

FIGURE 4.13 – La signature interféron dans le jeu de données

4.2.4.1 Changements de la composition des types cellulaires au cours de la genèse et de l'évolution de la maladie

Une fois l'ensemble des populations de cellules identifiées, nous avons réalisé une heatmap récapitulant les proportions de chacun des sous types au sein des échantillons (figure 4.14). Les patients et les types cellulaires ont été clusterisés en fonction de leur score de corrélation de spearman. Le clustering nous permet de regrouper les patients en 8 groupes basés sur la composition de leur environnement. Le premier groupe est composé de tous les échantillons bénins sauf un, ce qui montre une composition cellulaire homogène. Comme attendu, les NBM ont une composition similaire, même si l'une d'entre elles se retrouve dans le groupe 2. La composition cellulaire du groupe 2 est proche de celle du groupe 1. Cependant, on observe une diminution des T naïfs CD4 et CD8 ainsi que des B au profit de types cellulaires plus matures, les T notamment (cluster 5 des types cellulaires). Le groupe 2 comprend une NBM et des MM à divers stades, dont le SMM et quatre NDMM. Le reste des NDMM est réparti dans les autres groupes. Ils sont plus particulièrement présents dans le groupe 5 (4/7 échantillons). Ce groupe est caractérisé par une déplétion du cluster 1 de types cellulaires et un enrichissement en NK. Les groupes restants (3,4,6,7 et 8) comprennent majoritairement des RRMM sans distinction franche entre les ERRMM et LRRMM.



exclusifs à ces échantillons. Ces 2 groupes sont distingués par la composition du reste de l'environnement : le groupe 3 est enrichi en lymphocytes T tandis que le groupe 4 a une plus forte proportion de progéniteurs. Les groupes 6, 7 et 8 sont hétérogènes, leur caractéristique commune est un enrichissement en types cellulaires issus du cluster 2 : ncMono/macrophages et NK. Le groupe 7 a en plus la particularité d'être déplété des types de T les plus cytotoxiques : les $T\gamma\delta$, les CD8 cytotoxiques et les NKT.

Les échantillons issus du même patient ont tendance à avoir une composition proche et clusteriser ensemble, mais ce n'est pas toujours le cas. Cette observation suggère que les changements observés en réponse au traitement sont progressifs et gradués.

Comme nous l'avons montré, la constitution des groupes de patients est fortement influencée par le stade de la maladie. En comparaison, le type d'altération génomique primaire de la tumeur ne semble pas avoir d'influence.

Du point de vue des types cellulaires, nous obtenons 5 clusters (figure 4.14). La composition de ces clusters suit globalement la logique des lignages et des stades de maturité des types cellulaires. Le cluster 1 comprend les types cellulaires progéniteurs et immatures ainsi que les cellules dendritiques. Le cluster 2 regroupe les différents types de NK et les macrophages associés à la tumeur (TAM), ainsi que les ncMono. Il y a une forte hétérogénéité au sein de ce cluster. Le cluster 3 regroupe les T naïfs et les B. ces trois types cellulaires varient de façon particulièrement homogène au sein des échantillons, ce qui suggère des sensibilités similaires aux traitements et aux conditions de l'environnement. Le cluster 4 regroupe nos cellules avec une forte réponse interféron. Enfin, le dernier cluster regroupe les différents lymphocytes T. Au sein de ce cluster, on observe 2 branches qui séparent les T cytotoxiques des autres types.

Ces résultats semblent montrer qu'au sein de la moelle pathologique certains sous-types co-évoluent et sont représentés de façon similaire chez des groupes de patients. De plus, les membres de ces groupes ont tendance à être issus du même type cellulaire majeur.

Les corrélations entre les types cellulaires pour l'ensemble des patients confirment cette observation et permettent de l'affiner (figure 4.15). Les variations de proportions observées sont homogènes au sein des sous-types de chaque lignage. À l'exception des T naïfs qui co-varient avec les B (figure 4.16b). L'une des plus fortes corrélations est observée entre les ncMono et les NK matures (figure 4.16a). Ce résultat est en accord avec les résultats de ZAVIDIJ et al. 2020, qui montrent une corrélation similaire aux stades précurseurs du MM et au diagnostic. Nos analyses permettent d'affiner ces résultats puisque dans cet article ils n'ont pu identifier que 3 types de NK. Les progéniteurs et les NK matures sont

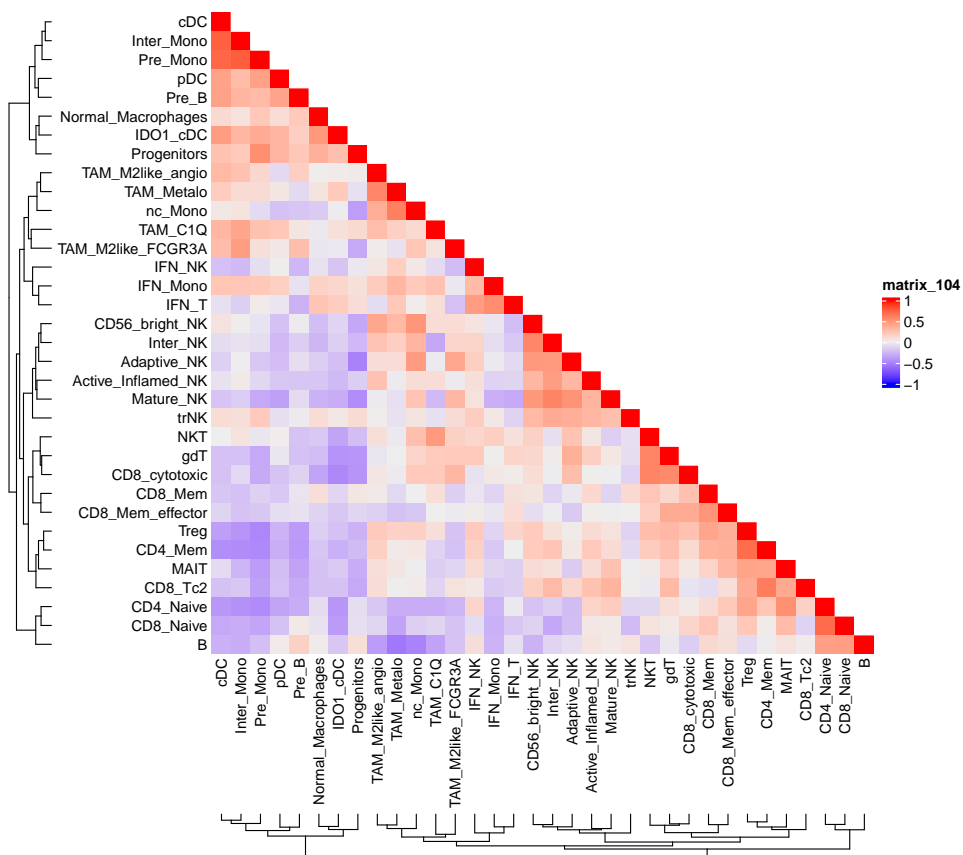
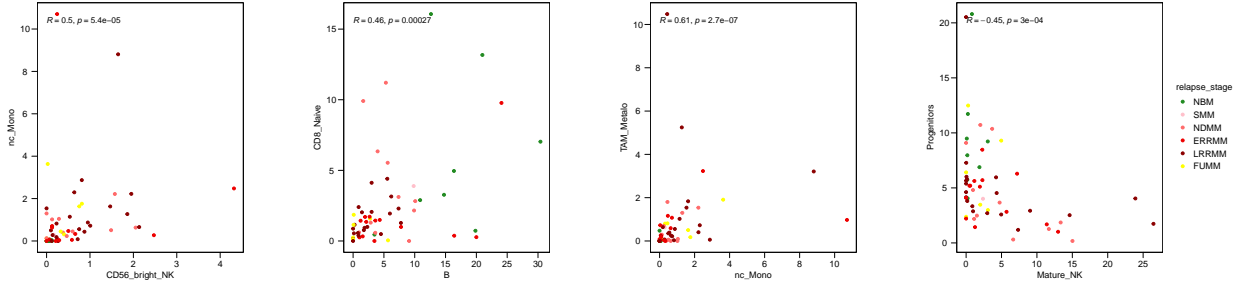


FIGURE 4.15 – Heatmap de corrélation des types cellulaires fins

quant à eux anti-corrélés (figure 4.16d).



(a) Corrélation entre les NK CD56_{bright} et les nc-Mono
 (b) Corrélation entre les B et les CD8 naïfs
 (c) Corrélation entre les ncMono et les TAM_Metallo
 (d) Corrélation entre les NK matures et les Progénérateurs

FIGURE 4.16 – Corrélations entre les proportions de certains types cellulaires

Les différents types cellulaires ne varient pas indépendamment les uns des autres. Ces co-variations suggèrent qu'ils partagent un mécanisme qui influence leur représentation. Ce lien a potentiellement différentes natures : un même profil de sensibilité à un ou des traitements, ou encore une réponse à un même signal biologique.

4.2.5 Les interactions entre l'environnement et les plasmocytes

Une fois nos types cellulaires définis avec précision, nous nous sommes interrogés sur les liens potentiels qu'ils entretiennent avec leurs tumeurs. Pour tenter de répondre à cette problématique, nous avons utilisé une méthode d'inférence d'interactions ligand-récepteur. Il existe divers algorithmes et bases de données pour ce type d'analyse. Ces algorithmes permettent de prédire les interactions ligand-récepteur sur la base de l'expression des gènes qui codent les protéines impliquées dans ces interactions. Pour notre projet, nous avons choisi d'utiliser LIANA (DIMITROV et al. 2022) qui reprend une partie d'entre eux et produit des résultats consensus.

4.2.5.1 Variations quantitatives des interactions plasmocyte-environnement

L'analyse a été réalisée sur chaque échantillon individuellement afin de ne pas obtenir d'interactions entre l'environnement d'un échantillon et la tumeur d'un autre. Pour chaque patient, on obtient donc une liste d'interactions significatives entre les types cellulaires de l'échantillon. Nous avons fait le choix de nous concentrer dans un premier temps sur les interactions impliquant la tumeur et leur quantification. Nous avons sélectionné l'ensemble

des interactions présentes dans au moins deux échantillons de MM. Nous n'avons pas pu exploiter les interactions entre les plasmocytes normaux et l'environnement de la NBM. En effet, ceux-ci sont trop peu nombreux pour obtenir des résultats pertinents. Nous avons comparé les interactions présentes aux différents stades de la maladie (Figure 4.17).

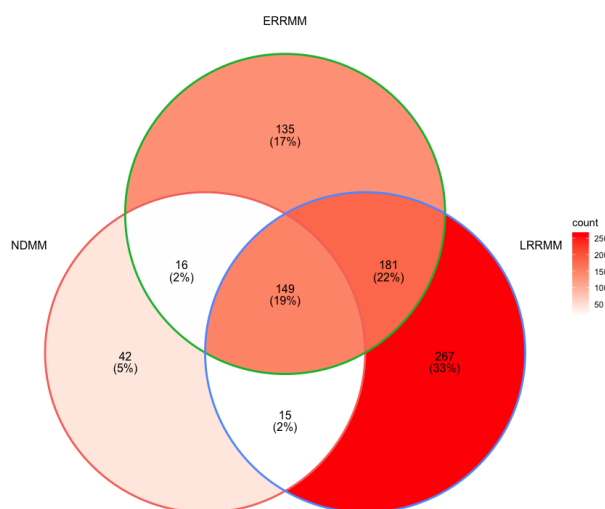


FIGURE 4.17 – Diagramme de Venn récapitulant les interactions présentes aux différents stades de la pathologie

Seules 5% des interactions sélectionnées sont présentes uniquement au diagnostic et 2% sont communes aux stades NDMM et ERRMM. Il semble donc qu'au cours de la progression, les gains d'interactions avec l'environnement soient privilégiés par la tumeur. Nous avons observé que 19% des interactions sont communes à tous les stades. Celles-ci peuvent correspondre à des interactions normales, ou bien à des interactions apparaissant tôt dans l'histoire de la maladie et conservées tout au long de sa progression. Parmi les interactions, 17% ne sont présentes qu'au stade ERRMM. Il paraît cependant surprenant qu'une interaction bénéfique acquise à ce stade ne soit pas conservée aux stades suivants. Cela s'explique peut-être par la disparition de certains types cellulaires de l'environnement sous la pression des traitements. Une autre explication est l'hétérogénéité inter et intra-patient. Les traitements influent sur la composition clonale de la tumeur. Si le paysage clonal de la tumeur change les interactions ont de bonnes chances de changer également. Par ailleurs, un seul des patients est représenté aux stades ERRMM et LRRMM. Il est donc possible que ces interactions soient le fruit de l'hétérogénéité inter-patient. Enfin, 33% des interactions sont observées exclusivement au stade LRRMM. La tumeur poursuit donc son accumulation d'altérations. Par ailleurs, il est possible que ces interactions aient

un lien avec l'utilisation de traitements réservés aux cas de MM avancés.

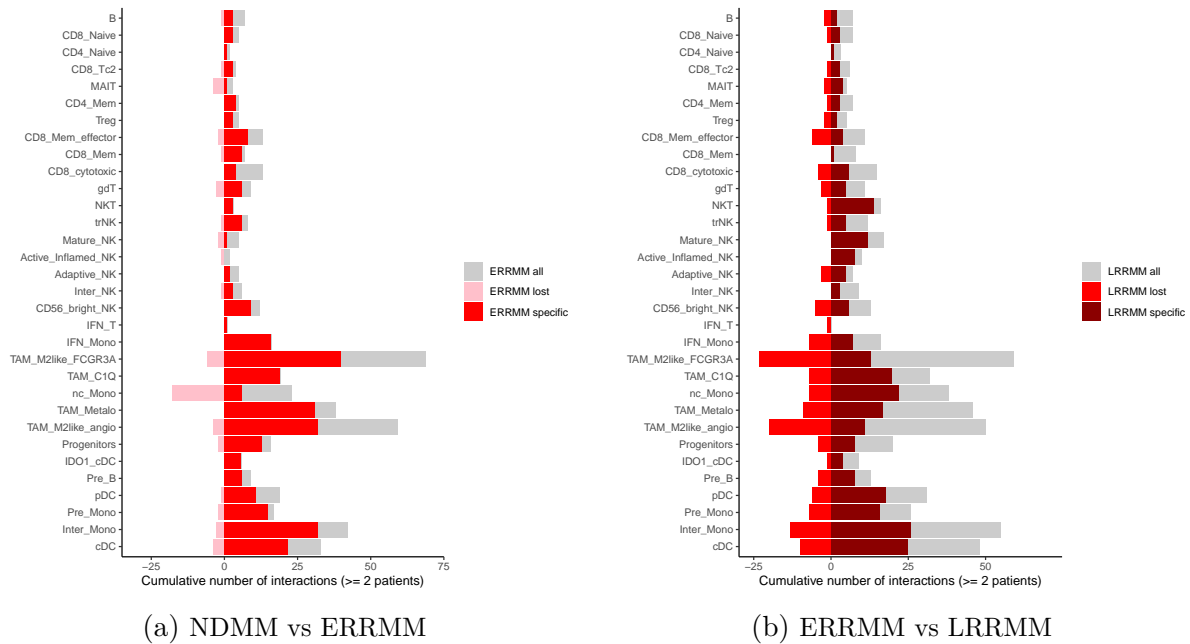


FIGURE 4.18 – Comparaison des interactions entre les différents stades de la maladie

L'analyse des interactions entre les cellules de myélome et les sous types cellulaires de la moelle (figure 4.18) montre que les traitements d'induction sont associés à une augmentation importante des interactions entre les plasmocytes et les types cellulaires de la lignée myéloïde, en particulier les populations de TAM. Il y a aussi une perte significative des interactions avec les ncMono. Dans les stades plus avancés, les gains et les pertes sont plus équilibrés mais toujours avec une forte dynamique de changement. Il est à noter des gains d'interaction plus importants à ce stade avec les cellules cytotoxiques (T, $T\gamma\delta$, NKT, Mature NK). Le nombre et la variété des interactions entre les tumeurs et les types cellulaires myéloïdes tiennent certainement de la plasticité de ces types cellulaires. En effet, en plus de la variabilité des cellules tumorales, il existe également une variabilité des phénotypes myéloïdes entre les patients. Par ailleurs, ces cellules sont fortement impliquées dans la régulation des processus biologiques. Ces processus de régulation impliquent de nombreux signaux émis et reçus par ces cellules.

4.2.5.2 Analyse de l'interaction (*CLEC2D*) LLT1-CD161 (*KLRB1*)

L'une des interactions qui a retenu notre attention et que nous avons décidé de caractériser plus en détails est l'interaction LLT1-CD161. Ces deux protéines sont respectivement

codées par les gènes *CLEC2D* et *KLRB1*.

Le CD161 est exprimé par les NK ainsi que différents types de T (figure 4.19). L'activation du CD161 à la surface des NK provoque une inhibition de leurs fonctions effectrices : cytotoxicité et production de cytokines (BULLER, P. A. MATHEW et S. O. MATHEW 2020; MATHEWSON et al. 2021; BRAUD et al. 2022). Dans le cas des lymphocytes T, l'effet de l'interaction était controversé. Cependant, il semble que le CD161 exerce une activité inhibitrice similaire à celle décrite pour les NK (MATHEWSON et al. 2021).

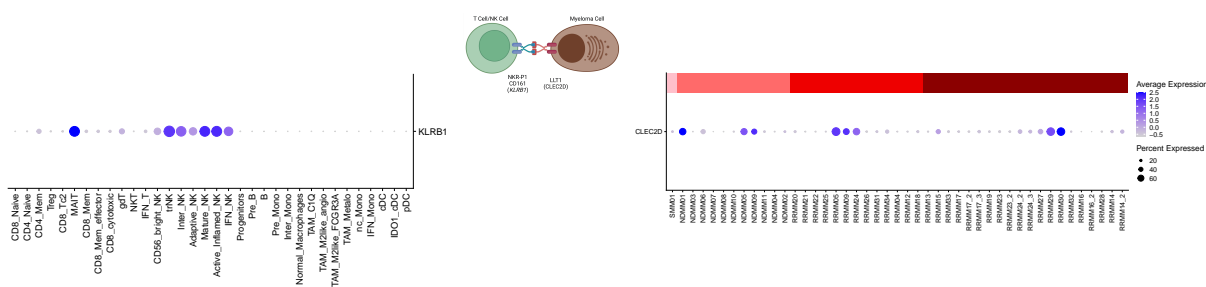
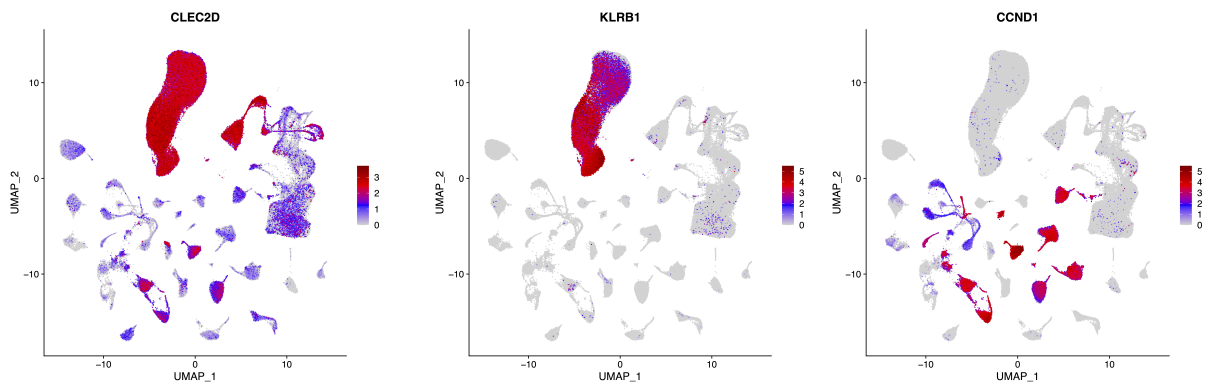


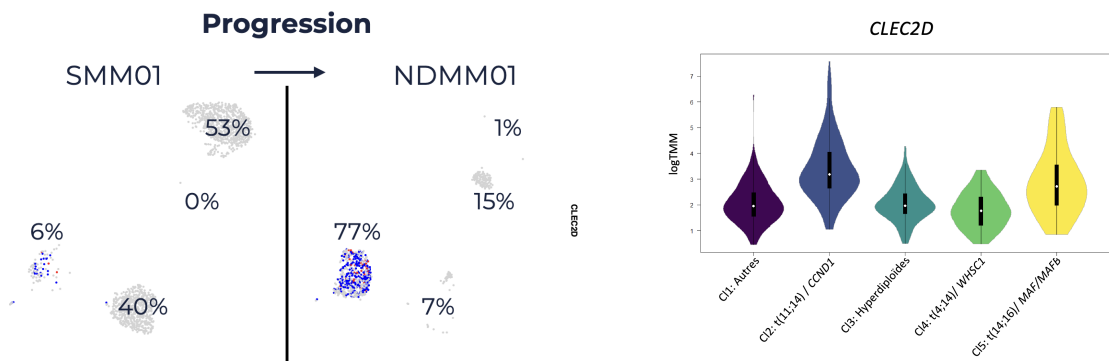
FIGURE 4.19 – Interaction *CLEC2D-KLRB1* entre l'environnement et les plasmocytes

Parmi les tumeurs, l'expression de *CLEC2D* est hétérogène (Figure 4.19 et 4.20a). Il est peu ou pas exprimé par la majorité des tumeurs. Mais, pour cinq d'entre elles, le niveau d'expression est plus élevé. Pour quatre de ces tumeurs (MM05, MM09, MM29, MM30) l'expression de *CLEC2D* est détectée dans l'ensemble des cellules. Par ailleurs, ces quatre tumeurs arborent le même évènement primaire : une t(11;14) et expriment donc fortement *CCND1* (Figure 4.20c). L'expression de *CLEC2D* semble donc liée à celle de la *CCND1* chez les patients t(11;14). L'analyse de l'expression de *CLEC2D* dans une grande cohorte de patients inclus dans l'essai clinique CASSIOPEIA (MOREAU, ATTAL et al. 2019) montre une augmentation significative dans le groupe t(11;14) et confirme cette hypothèse (figure 4.20e).

Dans le cas de notre cinquième patient (MM01), l'expression de *CLEC2D* est sous-clonale et son évènement primaire n'est pas déterminé. Au vu de son expression de la *CCND1* (Figure 4.20c), il paraît cependant clair qu'il n'a pas de t(11;14). Ces deux éléments suggèrent que l'activation de l'expression de *CLEC2D* par cette tumeur a eu lieu via un mécanisme différent des quatre autres. Ce cas est d'autant plus intéressant que nous avons deux prélèvements pour ce patient : aux stades SMM et NDMM. L'analyse conjointe des plasmocytes de ces deux échantillons a révélé la présence de quatre clones transcriptomiques. L'abondance de ces clones évolue fortement entre les deux stades de la maladie, et plus particulièrement celle d'un clone exprimant *CLEC2D*. Celui-ci est



(a) UMAP globale expression de *CLEC2D* (b) UMAP globale expression de *KLRB1* (c) UMAP globale expression de *CCND1*

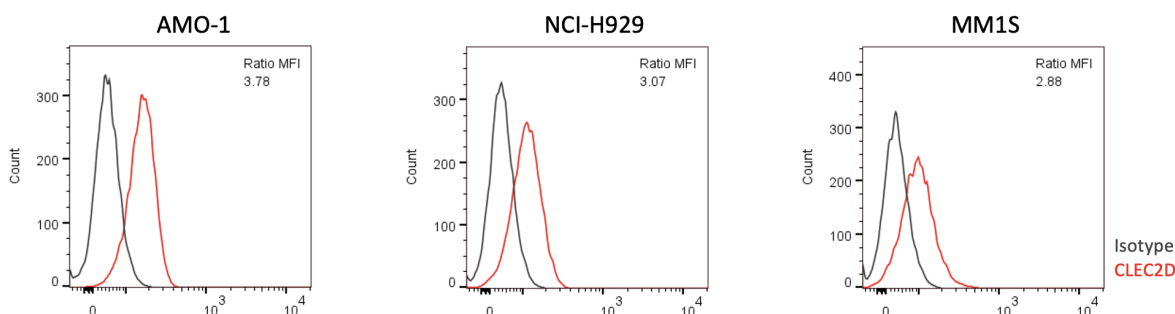


(d) Expression de *CLEC2D* chez les différents clones tumoraux du patient MM01, aux stades SMM et MM (e) Violin plots représentant l'expression de *CLEC2D* dans les différents sous groupes moléculaires établis à partir du transcriptome de 628 patients inclus dans l'essai CASSIOPEIA.

FIGURE 4.20 – L'expression des gènes impliqués dans l'interaction *CLEC2D-KLRB1*

minoritaire au stade SMM (6%) et devient majoritaire à la progression (77%) (Figure 4.20d). Bien que cela ne prouve pas que *CLEC2D* (LLT1) apporte un avantage sélectif à la tumeur, cela permet de le supposer.

Lignées cellulaires de Myélome



BMMC de patients

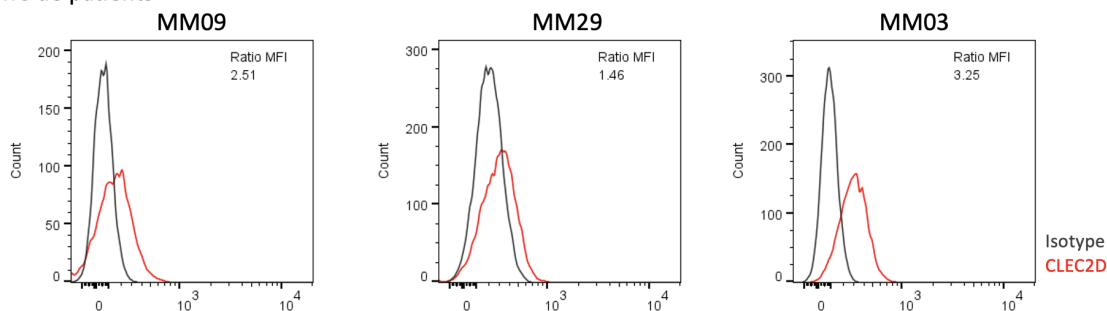


FIGURE 4.21 – Histogrammes de FACS pour les niveaux d’expression de la protéine LLT1 dans trois lignées de myélome et dans les échantillons de trois patients RRMM

Nous avons validé, par cytométrie en flux, l’expression de LLT1 à la surface des plasmocytes du patient MM09 (MFI = 2,52) (figure 4.21). Mais, il apparaît que l’expression du gène n’est pas toujours corrélée à celle de la protéine comme par exemple dans le cas du patient MM03, ce qui a déjà été rapporté (C. GERMAIN, BIHL et al. 2010), les plasmocytes du patient MM03 expriment LLT1 pourtant l’expression du gène est faible. Nous sommes en train d’analyser l’expression de la protéine CD161 codée par *KLRB1* dans les NK et T cytotoxiques. Si nous validons en cytométrie les résultats d’expression de *KLRB1*, nous pourrions faire l’hypothèse que, comme dans les lymphomes B non hodgkinien, l’interaction LLT1-CD161 peut inhiber la fonction cytotoxique des NK (C. GERMAIN, GUILLAUDEUX et al. 2015). De plus, récemment, il a été montré que cette interaction peut inhiber la cytotoxicité des T CD8+ infiltrant dans le gliome (MATHEWSON et al. 2021). Enfin, l’association de l’expression de *CLEC2D* à la t(11;14) est particulièrement intéressante, bien que ce groupe de patient soit de risque standard.

4.2.6 Analyse approfondie du compartiment T

4.2.6.1 Composition du compartiment T

Sur la UMAP du compartiment T (Figure 4.22), ces derniers s'organisent par ordre de maturité et par niveau de cytotoxicité. L'expression des gènes les plus différenciellement exprimés par ces populations est présenté en figure 4.23. Dans le coin inférieur gauche, on retrouve les populations CD4 et CD8 naïve, caractérisées par l'expression de *SELL*. Les CD4 mémoire (CD4_Mem), n'expriment plus *SELL* mais ont une plus forte expression d'*IL7R*. En marge de cette population, on retrouve les Treg et les CD8_Tc2. Les Treg ont été identifiés grâce à l'expression de *FOXP3*.

Les CD8_Tc2 sont une population de lymphocytes T helper exprimant le CD8 et non le CD4. Leurs fonctions semblent similaires. Cette population a été identifiée par son expression de *GATA3* notamment (LOYAL et al. 2020). Les sous-types de T présentés jusqu'ici ne présentent pas d'expression de gènes liés à la cytotoxicité. Ils correspondent aux cellules Helper_Naive_T de nos types cellulaires majeurs.

Du côté droit de la UMAP (Figure 4.22), on retrouve les différents sous-types qui composent notre groupe Cytotoxic_T. Les MAIT sont des T CD8, ils sont notamment caractérisés par l'expression de *KLRB1*. Les CD8 mémoires comme leur équivalent CD4, sont caractérisés par l'expression d'*IL7R* et la perte de l'expression de *SELL*. Ce type cellulaire exprime aussi faiblement les gènes liés à la cytotoxicité (*GZMA*, *GZMK*). Les CD8_Mem_effector ont un transcriptome similaire. Cependant les gènes de la cytotoxicité sont plus fortement exprimés et l'expression d'*IL7R* est perdue.

Enfin, en bas à droite de la UMAP, on retrouve les types cellulaires les plus cytotoxiques. Ceux-ci expriment des gènes associés aux cellules cytotoxiques les plus matures : *GNLY*, et les granzymes B et H (MAHNKE et al. 2013). Parmi les populations les plus cytotoxiques, on retrouve les T $\gamma\delta$ caractérisés par leur TCR gamma (TRGC) delta (TRDC), les NK T qui expriment certains marqueurs de NK (*TYROBP* par exemple) et en enfin les CD8 cytotoxiques. On notera que les NK T sont pour la plupart issus de 4 échantillons seulement (figure 4.36). Il semble donc que certains facteurs permettent ou non leur présence au sein de l'écosystème tumoral.

Nous nous sommes ensuite intéressés aux variations de ces sous-types de T au cours de la maladie (figure 4.36). Au diagnostic, les Treg, les MAIT, les T $\gamma\delta$ et les CD8 cytotoxiques sont significativement enrichis. Pour ces différents types cellulaires, l'enrichissement se maintient aux différents stades de rechute, excepté pour les MAIT qui semblent revenir

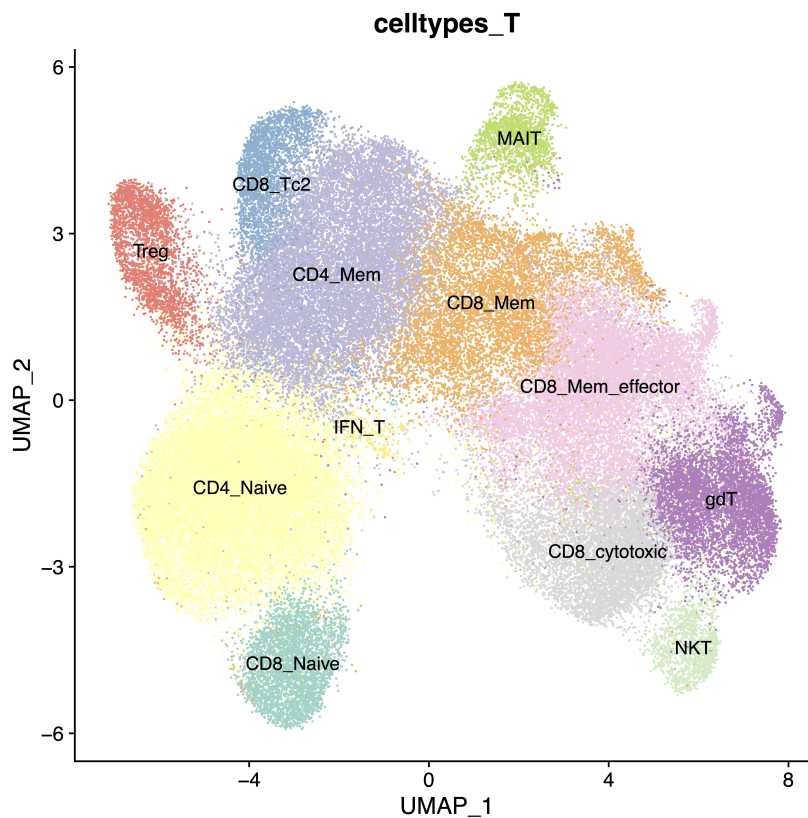


FIGURE 4.22 – UMAP du compartiment T, les cellules sont colorées par type cellulaire

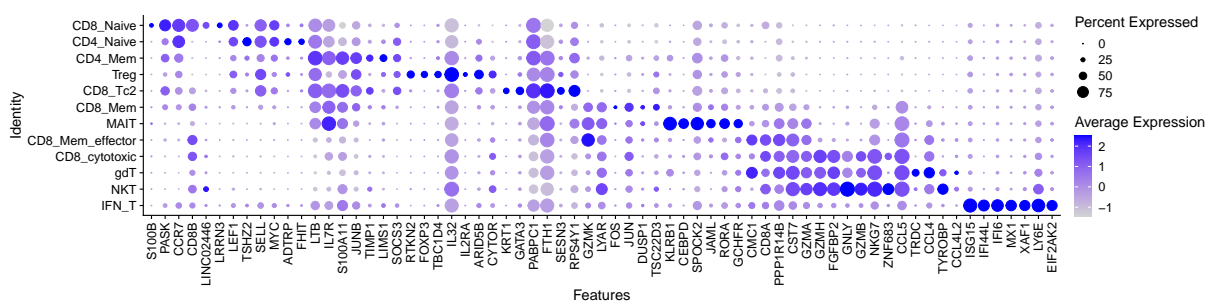


FIGURE 4.23 – Dotplots des principaux marqueurs des T

à un niveau normal malgré une certaine hétérogénéité inter-patient. Nous n'avons pas observé de dynamique d'évolution similaire à celle des MAIT. Par ailleurs, l'enrichissement observé dans notre jeu de données n'est pas cohérent avec de précédents travaux dans le MM au diagnostic (FAVREAU et al. 2017). Cependant, nos observations sont renforcées par les échantillons sériés de nos patients. En effet, pour les trois patients avec un point au NDMM et en ERRMM, nous observons une diminution de l'abondance des MAIT (figure 4.25).

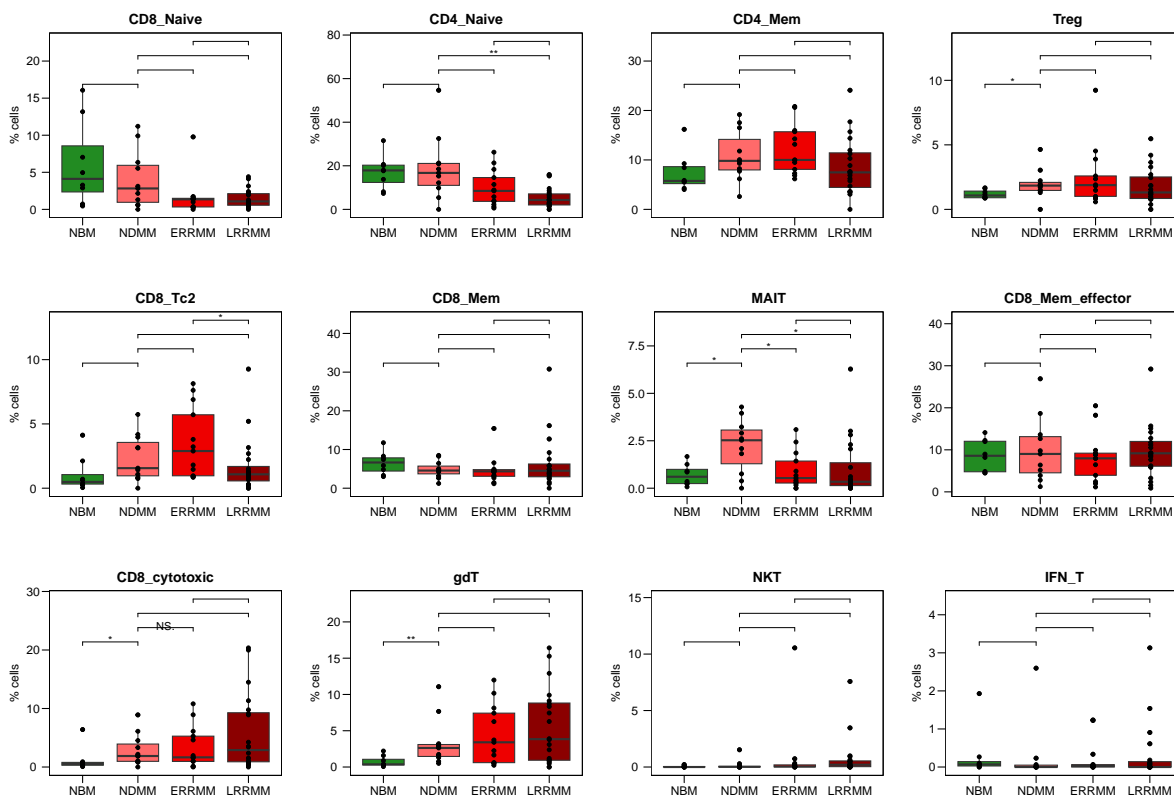


FIGURE 4.24 – Boxplots des proportions de T au sein de chaque échantillon, par stade

L'enrichissement des Tregs à tous les stades semble indiquer qu'un environnement immunosuppresseur se met en place dès le diagnostic (CRINIER et al. 2021). Du côté des T naïfs, les CD4 sont significativement déplétés seulement au stade LRRMM, bien que l'on observe une tendance à la diminution dès le stade ERRMM pour les CD4 et les CD8. Cette observation est en accord avec les précédents résultats rapportés par TIRIER et al. 2021 qui ont observé une diminution des populations T CD4 naïves et mémoires mais pas des CD8 naïfs après rechute. Ces populations semblent avoir une sensibilité particulière à

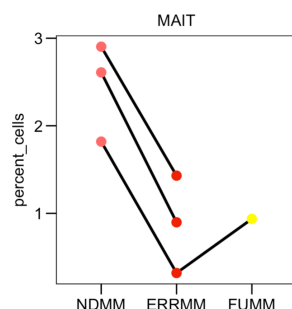


FIGURE 4.25 – Proportion de MAIT chez les patients avec des échantillons aux stades NDMM et ERRMM

l'accumulation des lignes de traitements.

Nous nous sommes aussi intéressés à l'évolution du ratio CD4/CD8 au cours de la progression. Comme cela a déjà été rapporté dans le MM, au cours de la progression le ratio évolue en faveur de la population CD8 (FRIEDRICH et al. 2023). Au NDMM, la différence avec les NBM n'est pas significative bien que la moyenne soit inférieure. En revanche, dès le stade ERRMM, le ratio diminue significativement en comparaison au NDMM. Au stade LRRMM, cette différence est encore plus significative.

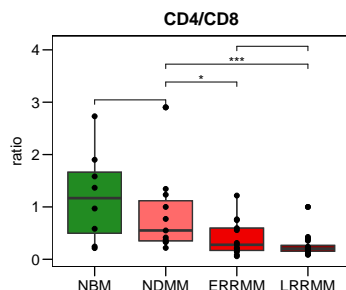


FIGURE 4.26 – Boxplot des ratios CD4/CD8 au sein des échantillons à chaque stade

4.2.6.2 État de la fonction T dans le MM

Nous nous sommes ensuite demandés si les changements observés en terme de proportion et de ratio CD4/CD8 s'accompagnent de changements phénotypiques et fonctionnels. Pour cela, nous avons travaillé sur les fonctions de cytotoxicité et l'exhaustion au sein de nos populations d'effecteurs cytotoxiques. Nous avons quantifié des signatures correspondants à ces fonctions dans l'ensemble de nos cellules (figures 4.27b et 4.27d) (cytotoxicité :

PRF1, NKG7, GNLY, GZMA, GZMB, GZMH, GZMK ; exhaustion : *PDCD1, CTLA4, LAG3, HAVCR2, CD244, VSIR*). Comme attendu, la fonction cytotoxique au sein des populations ne semble pas ou peu varier au cours de la pathologie, avec cependant une part d'hétérogénéité entre les patients (figure 4.27a). Au niveau de l'exhaustion, on remarque une hétérogénéité importante entre les patients (figure 4.27c). Comme attendu, le mécanisme d'exhaustion de la fonction T semble lié à la résistance dans le MM.

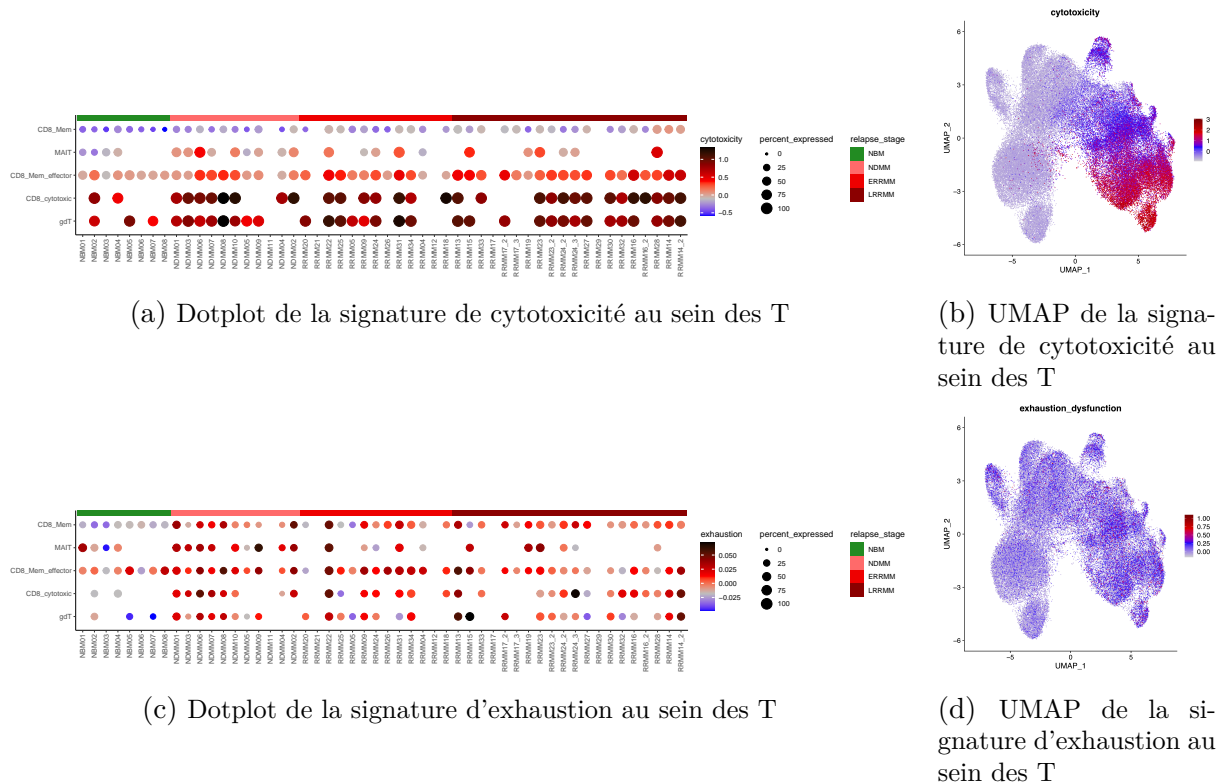
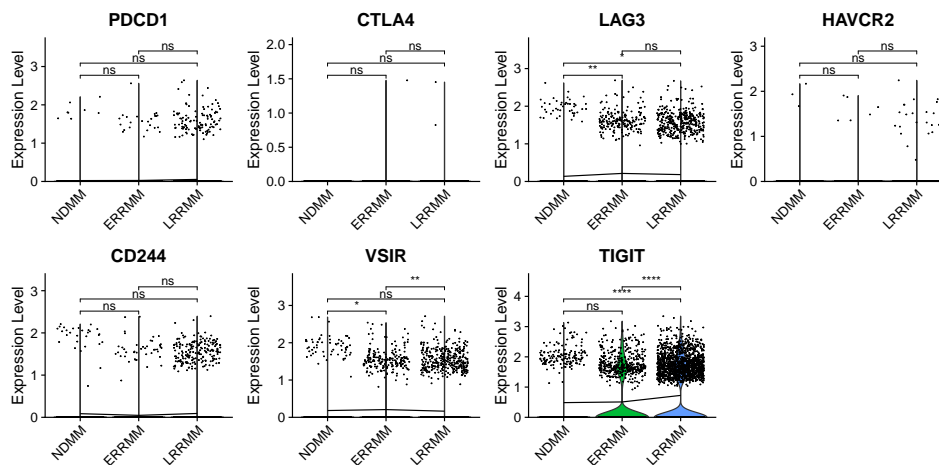


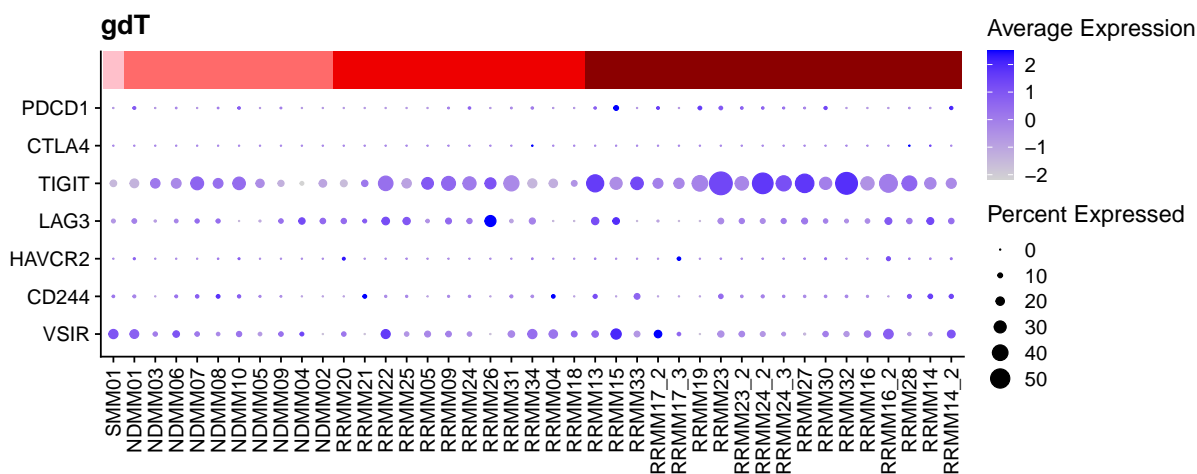
FIGURE 4.27 – Niveau d'expression des signatures d'exhaustion et de cytotoxicité au sein des populations T effectrices cytotoxiques

Pour aller plus loin sur ce phénomène d'exhaustion, nous avons décidé de regarder les variations individuelles des gènes de notre signature. Pour cela, nous nous sommes concentrés sur les $T\gamma\delta$. La proportion de cette population tend à augmenter même si cela est hétérogène entre les patients. Parmi les gènes de la signature d'exhaustion, seul *TIGIT* est significativement sur-exprimé au sein de cette population au stade LRRMM (figure 4.28a).

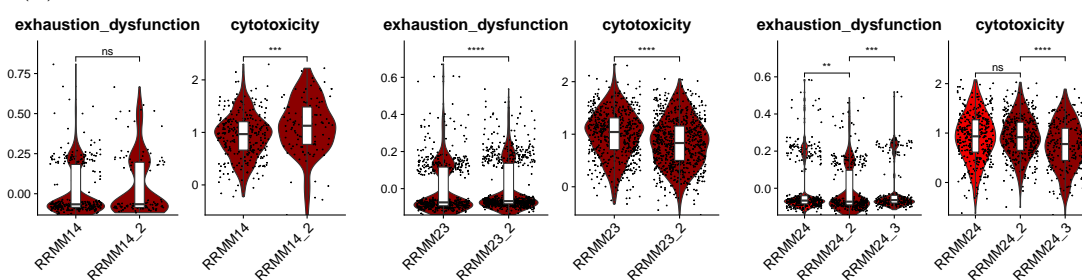
Lorsque l'on regarde les gènes de la signature individuellement et par patient pour les $T\gamma\delta$ (figure 4.28b), on remarque que malgré cette augmentation significative, ce mécanisme ne semble pas en place dans l'ensemble des échantillons. Par ailleurs, pour les autres



(a) Violin de l'expression des gènes associés à l'exhaustion et la dysfonction dans les $T\gamma\delta$



(b) Dotplot de l'expression des gènes associés à l'exhaustion et la dysfonction dans les $T\gamma\delta$



(c) Violin des signatures chez les $T\gamma\delta$ pour nos échantillons appariés

FIGURE 4.28 – Expression des gènes associés à l'exhaustion et la dysfonction dans les $T\gamma\delta$

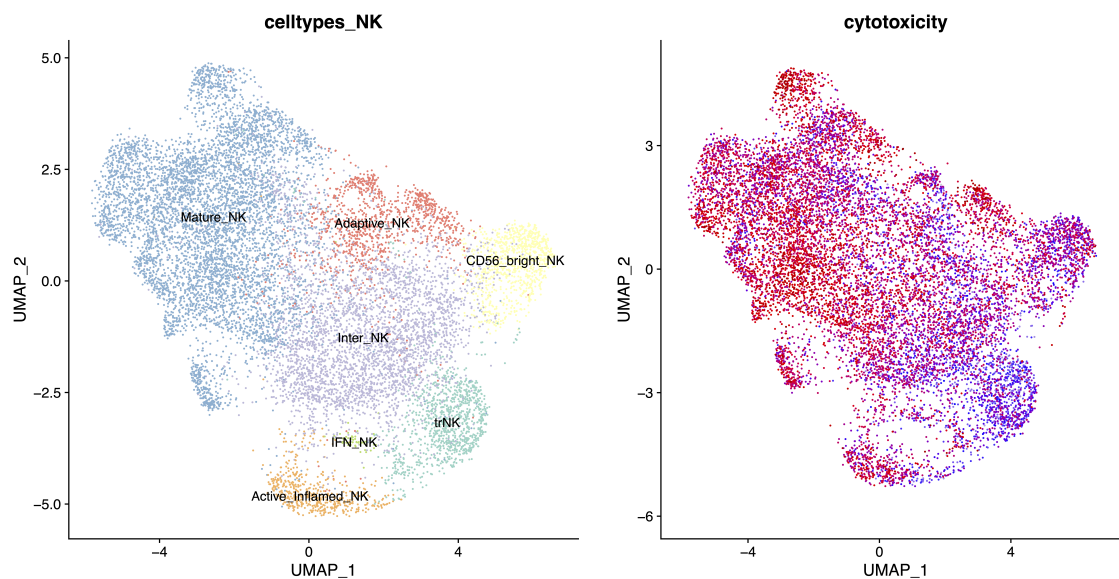
gènes de la signature, on remarque également que chez quelques patients l'expression est plus forte, c'est le cas de *PDCD1*, qui est exprimé dans les T $\gamma\delta$ de l'échantillon RRMM15.

Chez les patients avec plusieurs échantillons, nous avons observé différentes évolutions de ces 2 signatures. Pour le MM14, entre 2 LRRMM, la cytotoxicité des T $\gamma\delta$ augmente. Dans un scénario similaire pour le patient MM23, l'exhaustion augmente et la cytotoxicité diminue significativement. Enfin, pour le patient MM24, on a une chute de la cytotoxicité au troisième temps (LRRMM vers LRRMM) tandis qu'après la troisième ligne de traitement, il n'y avait pas de différence. Du côté de l'exhaustion, il existe des différences significatives entre les 2 progressions, une légère diminution puis une augmentation. Les éléments présentés précédemment tendent à montrer qu'au cours de la maladie et notamment aux stades les plus avancés, les cellules effectrices cytotoxiques deviennent dysfonctionnelles. Ces dysfonctions correspondent à une augmentation de l'expression de marqueurs d'inhibition connus mais aussi à une diminution des gènes de la cytotoxicité. Il reste cependant de l'hétérogénéité et ce n'est pas le cas pour tous les patients, comme nous l'avons montré pour le MM14 (figure 4.28c).

4.2.7 Les sous-populations de NK

Le compartiment NK est plus restreint que celui des T, néanmoins après réanalyse nous avons pu définir 7 sous-types de NK (figures 4.29 et 4.30). Comme les T, les NK sont organisés par cytotoxicité sur la UMAP. En bas à gauche, on retrouve les types cellulaires les moins cytotoxiques. Ceux-ci expriment plutôt les granzymes A et K, qui sont associés à une activité cytotoxique moindre que les granzymes B et H. Les trNK sont caractérisés notamment par l'expression du CXCR6 dans la moelle. L'expression de *CXCR6* au sein du cluster est relativement faible cependant elle n'est présente que dans celui-ci. Les NK CD56_{bright} sont une population bien caractérisée, qui exprime des marqueurs spécifiques tels que XCL1. Par ailleurs, on retrouve aussi chez eux l'expression de gènes qui correspondent aux stades immatures des populations T : *SELL* et *IL7R*. Au sein des populations de NK, ils sont les seuls à les exprimer. L'expression de ces marqueurs est l'un des facteurs qui fait que l'on considère cette population comme moins mature et précurseur des autres.

Les Inter_NK sont une population sans marqueur propre. Les gènes qui sont les plus différentiellement exprimés par cette population sont également exprimés par la plupart des autres (*PPP1R14B*, ..., *GCHFR*). Ils expriment également des niveaux intermédiaires des marqueurs de NK matures. Les Adaptive_NK, sont une population singulière. Ils



(a) UMAP du compartiment NK, les cellules sont colorées par type cellulaire (b) UMAP du compartiment NK, les cellules sont colorées par niveau de cytotoxicité

FIGURE 4.29 – UMAPs du compartiment NK

expriment des gènes du CMH II ainsi que *KLRC2*. Les NK matures sont les plus cytotoxiques, ils expriment les granzymes B et H ainsi que *FGFBP2* à un niveau élevé. La population *Active_Inflamed_NK* est caractérisée par une très forte expression des gènes codants pour des cytokines pro-inflammatoires, les *CCL3*, *CCL4* mais aussi le *TNF*. Enfin, la population *IFN_NK* est caractérisée par l'expression des gènes de la réponse interféron (*ISG15*, *IFI6*, etc). Comparé aux autres études en cellule unique de l'environnement du MM, nous avons pu établir un paysage plus détaillé de la population NK (ZAVIDIJ et al. 2020; TRIER et al. 2021). Ce niveau de détail est permis par la taille importante de notre jeu de données. Ces différences de précision ne permettent donc pas de comparer l'ensemble des résultats obtenus.

Nous nous sommes ensuite intéressés à la façon dont les proportions de ces sous-types évoluent (figure 4.31). Seules les comparaisons entre les NBM et les NDMM donnent des résultats significatifs. En effet, les NK *CD56_{bright}*, les NK matures et les NK adaptatifs sont significativement enrichis au diagnostic. Pour les NK tissue-resident et les NK intermédiaire l'augmentation n'est pas significative. Pour les 5 sous-types cités précédemment on observe une plus grande diversité de valeurs au NDMM. Cela reflète une hétérogénéité inter-patient des proportions de ces sous-types, que l'augmentation de la moyenne soit significative ou non. De manière intéressante, lorsque l'on compare les proportions entre

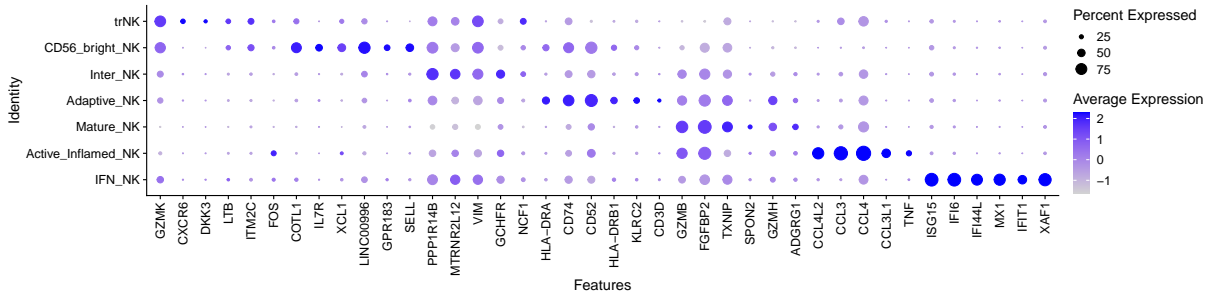


FIGURE 4.30 – Dotplots des principaux marqueurs des NK

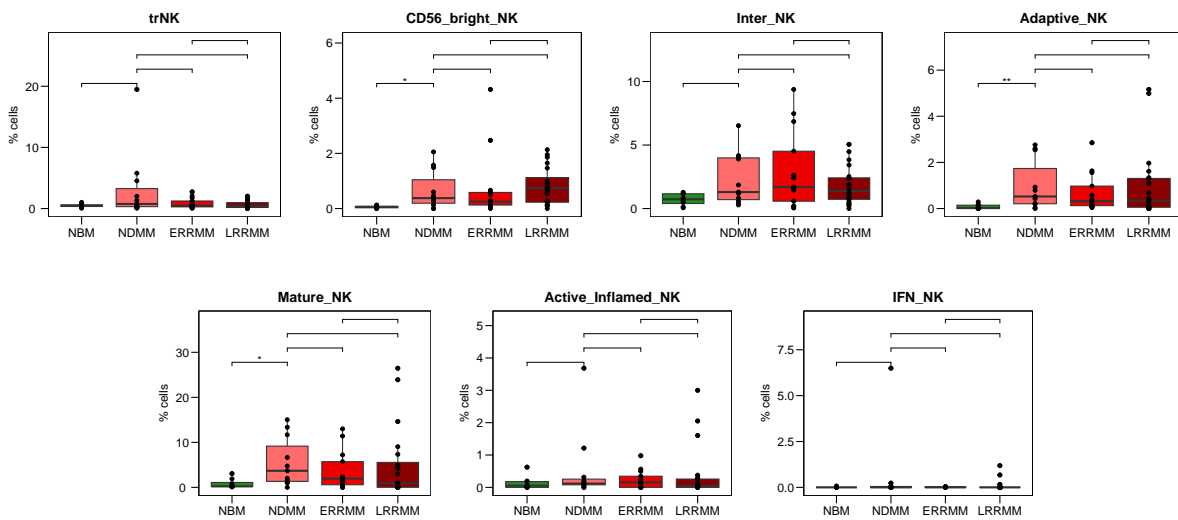
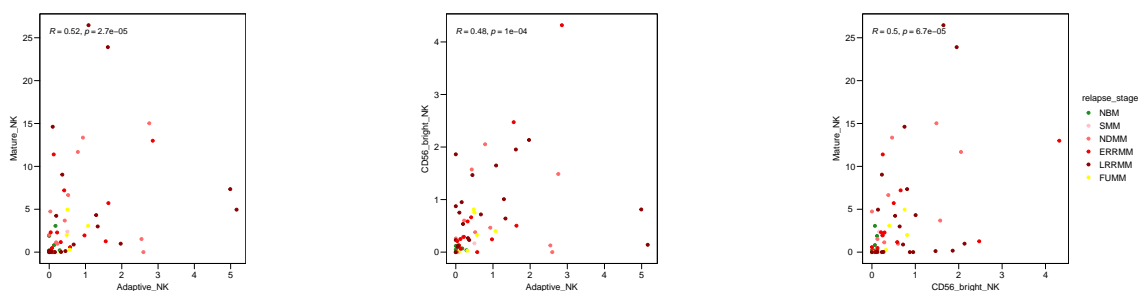


FIGURE 4.31 – Boxplots des proportions de NK au sein de chaque échantillon, par stade

les différents stades de la maladie on ne constate pas de différences significatives et l'hétérogénéité inter-patient se maintient. De leur côté les NK interféron et les NK activés sont principalement présents chez une minorité de patients. Pour les NK interféron ce résultat est en accord avec les résultats des autres populations caractérisées par l'interféron.



(a) Corrélation entre les NK adaptifs et les NK matures (b) Corrélation entre les NK adaptifs et les NK CD56_{bright} (c) Corrélation entre les NK CD56_{bright} et les NK matures

FIGURE 4.32 – Corrélations entre les proportions des types de NK enrichis au stade NDMM

Pour les populations dont l'enrichissement est significatif on observe également une corrélation entre eux (figure 4.32). Ce qui signifie que ces populations répondent à des signaux similaires induits par la présence du MM au diagnostic mais aussi aux stades suivants. Le maintien de la corrélation aux stades avancés implique des sensibilités similaires de ces populations aux traitements suivis par le patient.

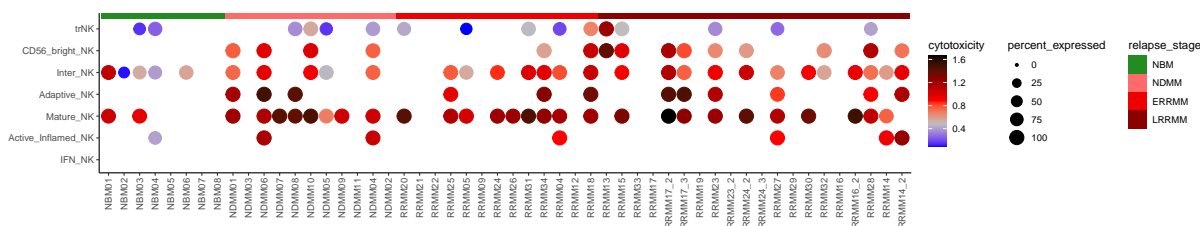


FIGURE 4.33 – Dotplots de la signature de cytotoxicité dans les NK

Comme pour les populations de lymphocytes T nous avons exploré l'aspect fonctionnel des NK. Nous avons analysé la signature de cytotoxicité au sein des sous-types. Nous n'avons pas observé de différences systématiques entre les NK aux différents stades de la maladie. Cependant comme attendu, il existe une diversité inter-échantillon du niveau de cytotoxicité (figure 4.33) qui vient s'ajouter à la diversité d'enrichissement.

4.2.8 Le compartiment myéloïde

Dans le compartiment myéloïde nous avons identifié 12 sous types (figures 4.34 et 4.35). Nous avons identifié deux populations de cellules dendritiques dont une est quasiment exclusive d'un patient. Ces deux populations ont en commun l'expression des gènes du CMH II. Cependant, la population IDO1_cDC exprime fortement *IDO1* ou encore *CCR7*. Le cluster de pDC n'a pas changé par rapport à l'annotation en types cellulaires larges. Il est caractérisé par l'expression de *LTB*, *GZMB* et *IRF8*. Les populations de monocytes/macrophages s'organisent quant à elle par ordre de maturité, de haut en bas. Les pré-monocytes sont caractérisés par l'expression de *MPO* et *ELANE* entre autres. Pour les différentes populations que nous avons appelées TAM il est difficile de définir clairement les clusters. En effet, ceux-ci forment un continuum. Cependant, comme on le constate sur la figure 4.35, il existe une réelle hétérogénéité d'expression entre ces groupes de cellules malgré leur caractéristiques communes.

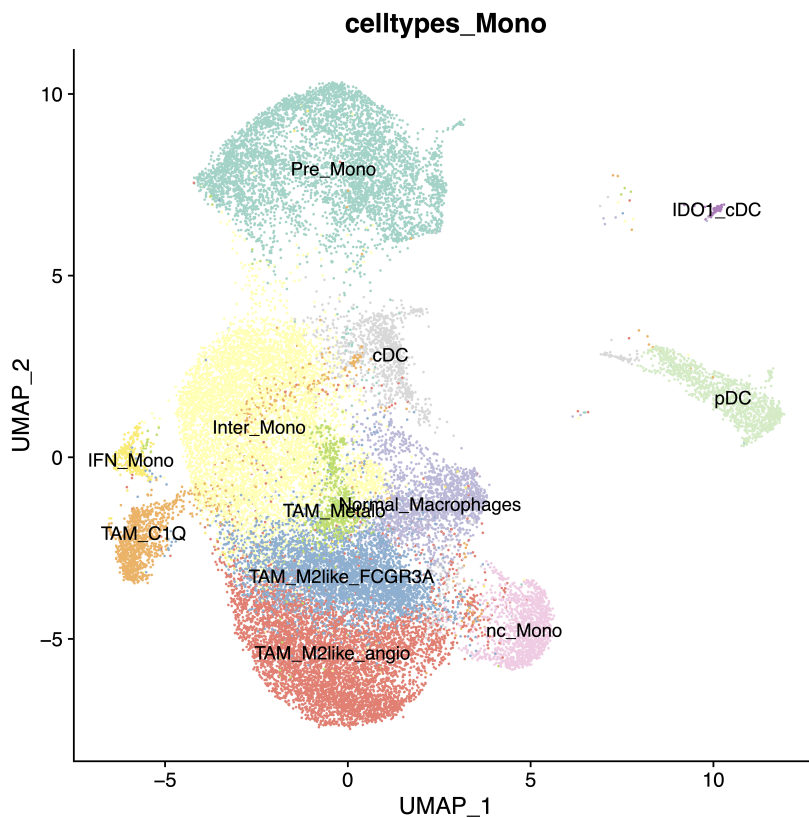


FIGURE 4.34 – UMAP du compartiment myéloïde, les cellules sont colorées par type cellulaire

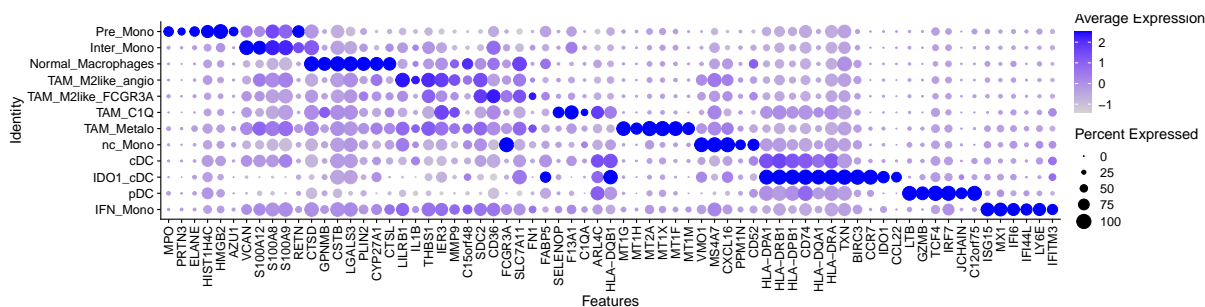


FIGURE 4.35 – Dotplots des principaux marqueurs des types cellulaires de la lignée myéloïde

Au sein de ce compartiment on observe peu de différences significatives au diagnostic. Les ncMono ainsi que les TAM_Metalo sont enrichis tandis que les pré-mono sont diminués. Lorsque l'on s'intéresse aux proportions des différentes populations du compartiment aux stades avancés, on constate que l'hétérogénéité entre les patients est forte. En effet, pour la plupart des population il y a quelques patients qui ont des valeurs très éloignées de la moyenne.

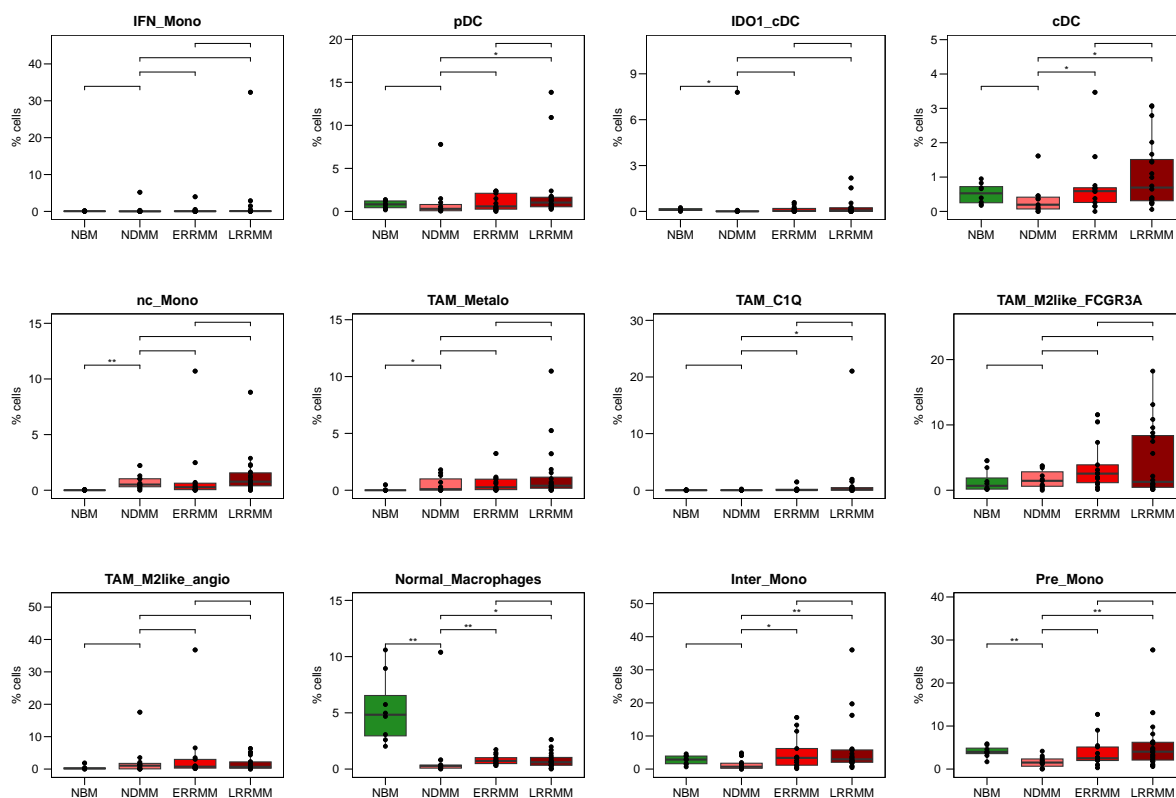


FIGURE 4.36 – Boxplots des proportions des types cellulaires de la lignée myéloïde au sein de chaque échantillon, par stade

4.2.9 Effet des BH3 mimétiques sur l'environnement immunitaire

4.2.9.1 Les BH3-mimétiques provoquent un remodelage de l'environnement immunitaire

Comme nous l'avons déjà évoqué les traitements ont un effet sur l'environnement immunitaire. Pour mieux comprendre l'effet des BH3 mimétiques nous avons analysé des échantillons dont les cellules ont été exposées à un anti-BCL2 (venetoclax) et un anti-MCL1.

L'ensemble de ces échantillons est appairé avec un des échantillons du jeu de données principal. Afin de ne pas réitérer les étapes d'identification des types cellulaires et pour éviter d'obtenir des classifications qui ne concordent pas, nous avons décidé d'utiliser une méthode de transfert d'annotation. Cette méthode consiste à projeter le second jeu de données dans l'espace issu de l'analyse du premier. Ainsi les cellules ayant des caractéristiques similaires seront annotées de la même façon, et ce même si de légers changements de transcriptomes sont induits par un effet batch ou par le traitement.

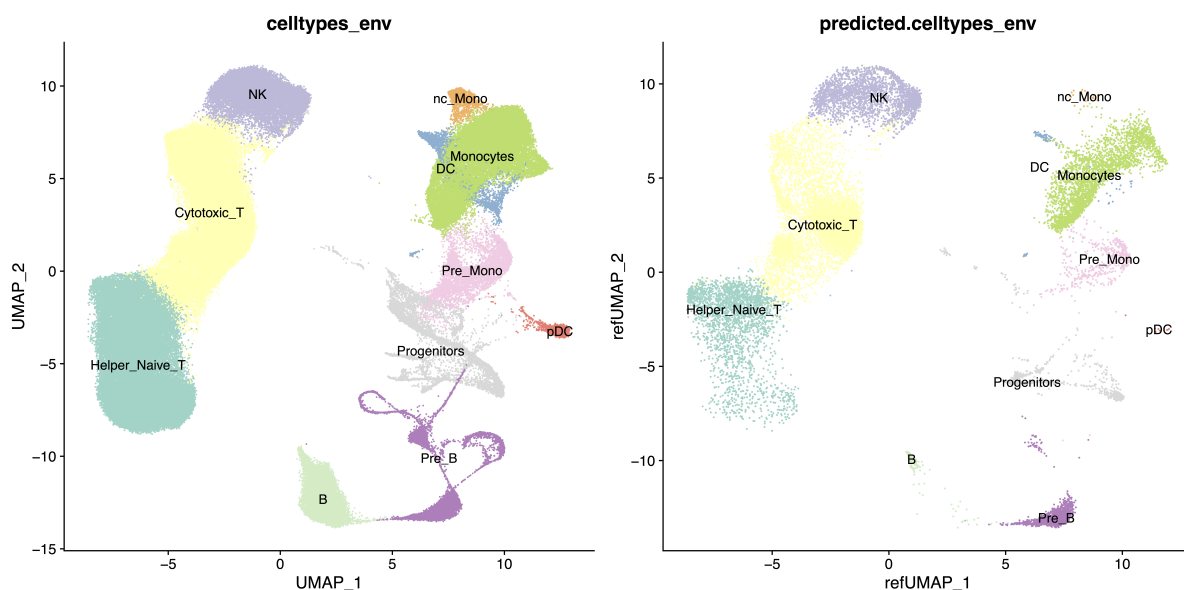


FIGURE 4.37 – UMAP de l'environnement contrôle (gauche) et BH3 (droite), les cellules sont colorées par type cellulaire large ou types cellulaire large prédit

Une fois les types cellulaires identifiés dans les 2 conditions, nous pouvons comparer leurs proportions avant et après traitement aux BH3-mimétiques (figure 4.38). À l'échelle

des types cellulaires larges, on constate déjà des conséquences significatives du traitement. Pour les populations T on observe une déplétion des types naïfs et une légère augmentation des types cytotoxiques. Ce premier constat semble indiquer que ce traitement est favorable à un environnement anti-tumoral. Les pDC, DC, ncMono et pre-monocytes sont sensibles au traitement. Les pDC et les ncMono sont particulièrement sensibles puisqu'ils tombent à zéro dans presque tous les échantillons après traitement. Pour les B la sensibilité est également importante bien que dans certains échantillons leur proportion se maintient.

La population de pré-B que nous avons définie contient en réalité plusieurs stades de différenciation qui ont été regroupés pour simplifier les analyses. Cependant, comme on le remarque sur la UMAP des échantillons traités aux BH3-mimétiques (figure 4.37 droite), les différents stades de différenciation semblent avoir une sensibilité très différente au traitement.

À l'échelle des types cellulaires fins on observe une hétérogénéité au sein des groupes issus des types principaux (Figure 4.39). Les types de lymphocytes T issus du cluster Helper_Naive_T sont tous significativement déplétés après le traitement, excepté les CD4 mémoire qui ne varient pas significativement. Parmi les types de T cytotoxiques les réponses sont variées. Les MAIT semblent être un type cellulaire très résistant au traitement, ils sont significativement enrichis. Dans une moindre mesure les CD8 mémoire effecteurs sont également enrichis. Les CD8 mémoire ainsi que les $T\gamma\delta$ sont significativement déplétés, mais sont toujours présents post-traitement. Pour le reste des sous-types de T cytotoxiques on observe pas de différence significatives. Ces résultats confirment de précédents travaux effectués sur l'exposition des populations T au venetoclax (KOHLHAPP et al. 2021).

Parmi les populations de NK on observe une légère baisse de la proportion de NK adaptifs et une augmentation de trNK. Le reste des sous-types de NK se maintient à des niveaux similaires après traitement. Du côté de la lignée myéloïde, il n'y a pas d'augmentation significative : les populations se maintiennent ou sont fortement déplétées par le traitement. Parmi les populations déplétées les cellules dendritiques (cDC et pDC) et les ncMono sont particulièrement touchées. Ces populations sont pratiquement absentes dans les échantillons traités. Deux populations de TAMs sont également affectées, les TAM_M2like_angio et les TAM_M2like_FCGR3A. Ces population ne disparaissent pas mais sont déplétées.

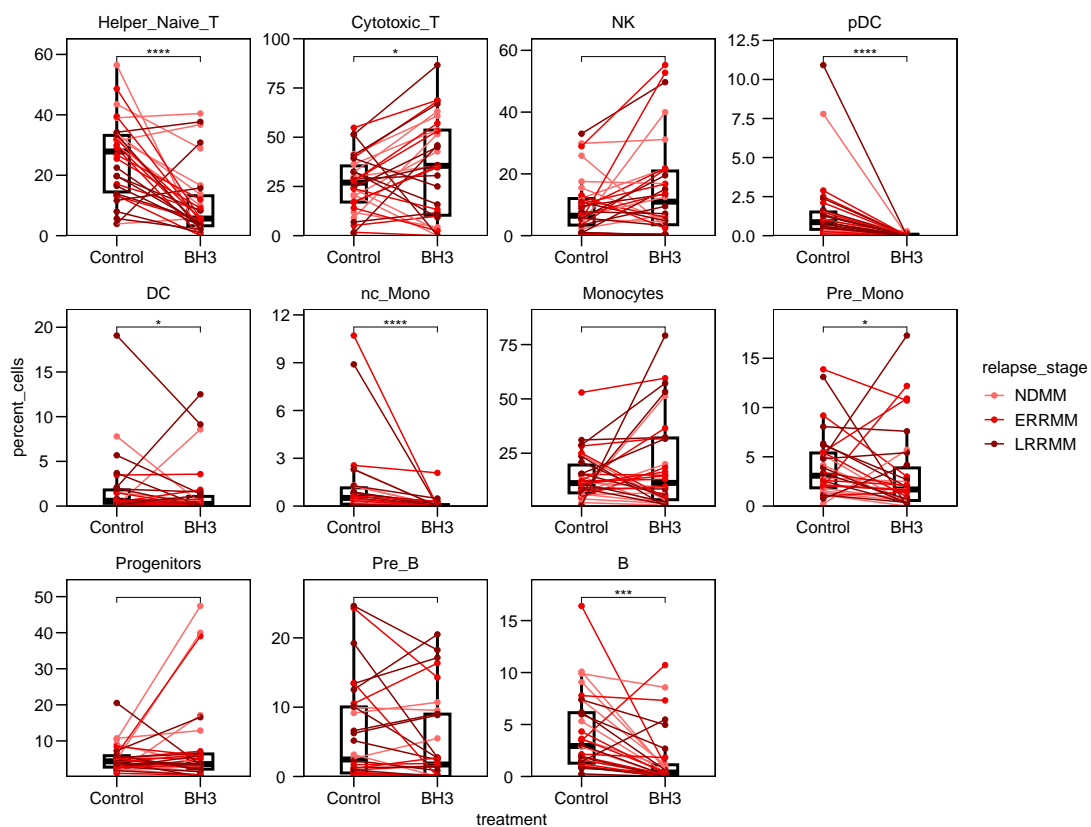


FIGURE 4.38 – Evolution des proportions des types cellulaires principaux après traitement au BH3 mimétiques

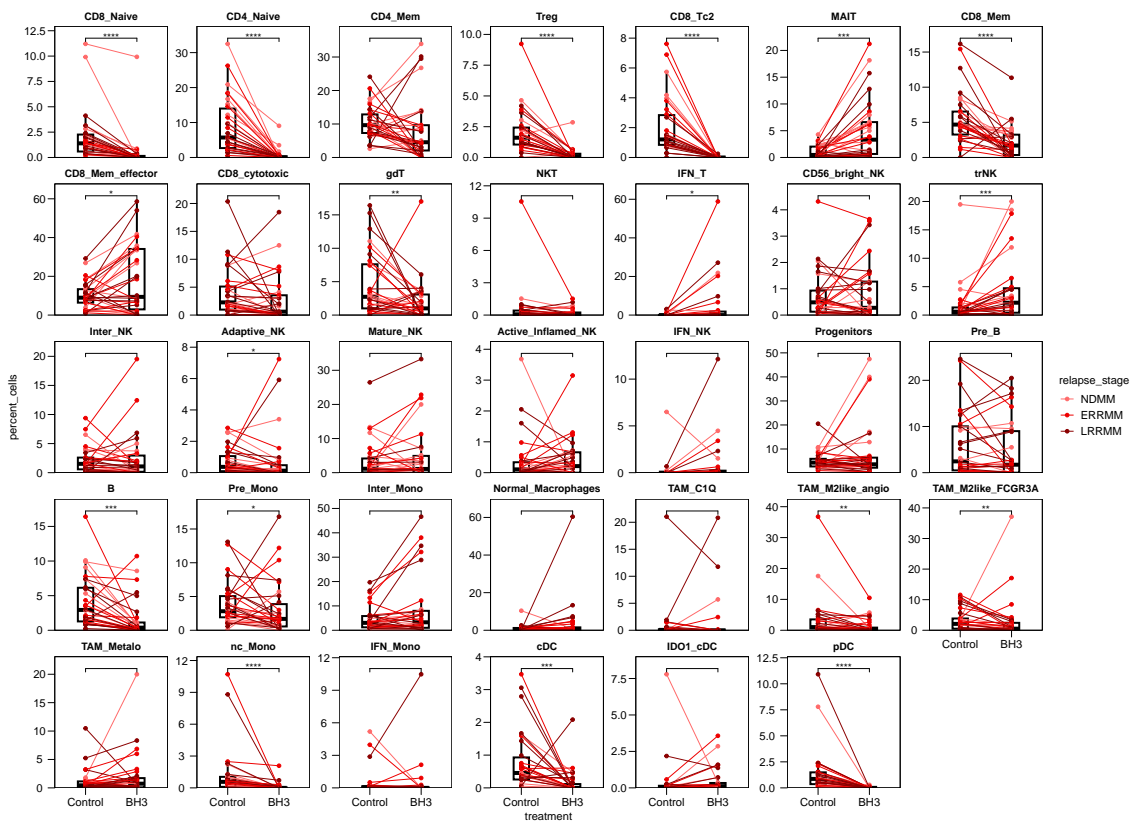


FIGURE 4.39 – Evolution des proportions des types cellulaires fins après traitement au BH3 mimétiques

4.2.9.2 La diversité de l'expression des gènes de la famille BCL2

Afin d'expliquer les différences de sensibilité entre nos types cellulaires, nous nous sommes penchés sur le niveau d'expression des membres anti-apoptotiques de la famille BCL2 : BCL2 (*BCL2*), BCL-XL (*BCL2L1*), MCL1 (*MCL1*), BCLW (*BCL2L2*) et BFL-1 (*BCL2A1*) (figure 4.40). L'expression des gènes est présentée sous forme de heatmap par compartiment. Nous avons fait le choix de présenter les compartiments séparément puisque les sous-types myéloïdes expriment plus fortement ces gènes et écrasent donc l'échelle pour les autres types. Nous les avons donc regroupés de cette façon pour rendre la comparaison plus pertinente.

L'expression des gènes de cette famille est relativement hétérogène et dépend du type cellulaire considéré. Parmi les types cellulaires proches on retrouve néanmoins des profils similaires. Dans le compartiment T, on retrouve quatre types de profils. Ces profils correspondent aux stades de maturité et fonction, les T naïfs et les CD8_Tc2 expriment faiblement les gènes de la famille. Il semble cependant que MCL1 et BCL2 jouent un rôle dans leur survie puisque ces types cellulaires sont déplétés. Les T mémoires semblent plutôt bien résister au traitement. Cela est surprenant puisqu'ils expriment *BCL2* et *MCL1* plus fortement que les trois autres gènes. Le troisième groupe de T est celui des T matures et cytotoxiques. Ces sous-types sont caractérisés par l'expression de *BCL2L1*, ce gène est induit lors de l'activation et de la maturation des lymphocytes T (KOHLHAPP et al. 2021). Le venetoclax est un BH3-mimétique spécifique de BCL2, il a été développé afin de remplacer les premières molécules qui ciblaient aussi BCL-XL. BCL-XL est essentiel pour la survie des progéniteurs mégakariocytes, son inhibition provoque donc des thrombopénies délétères pour les patients. Il semble donc que la non-spécificité pour BCL-XL permette d'épargner les populations de T effecteur en plus de limiter les toxicités. Enfin du côté des Treg, nous avons constaté une forte sensibilité au traitement. Celle-ci pourrait être expliquée par la forte expression de *BCL2* et *MCL1*, dont les Treg semblent donc dépendre pour leur survie.

On notera également que la population exprimant les gènes de réponse à l'interféron ne semble pas affectée par le traitement. Il en est de même pour les populations NK et monocytes. Chez les T et NK positifs à l'interféron on remarque un niveau d'expression de *MCL1* élevé comparé au types cellulaires de leurs compartiments respectifs. La réponse interféron pourrait donc être un facteur permettant la survie des cellules face à l'inhibition des anti-apoptotiques. Cette réponse interféron pourrait également être provoquée par le traitement, en effet en cas d'inhibition de l'activité des caspases les ARN mitochondriaux

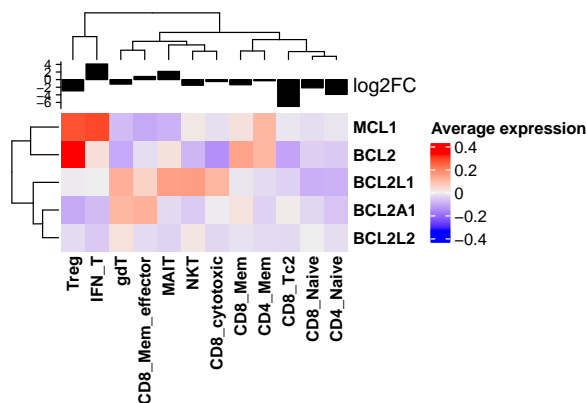
relargés dans le cytoplasme peuvent déclencher cette réponse (CZABOTAR et GARCIA-SAEZ 2023).

Les populations de NK varient peu entre nos deux conditions, hormis les trNK qui augmentent légèrement. Les différentes populations de NK expriment des niveaux assez faibles des cinq anti-apoptotiques. Il est possible que leur survie soit régulée par d'autres mécanismes. La population exprimant des gènes liés à l'activation et l'inflammation exprime plus fortement *BCL2A1*.

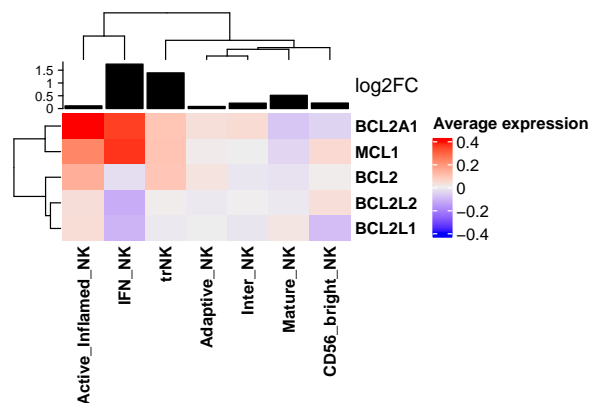
Cependant, pour les B par exemple la survie est assurée par l'expression de *BCL2*, il est donc cohérent qu'ils soient affectés par le traitement. En revanche, les pré-B n'expriment pas *BCL2* mais *BCL2L1* ce qui pourrait expliquer leur résistance au traitement.

Les niveaux d'expression de ces gènes ne permettent pas à eux seuls d'expliquer la sensibilité des différents sous-types. Il faudra donc également investiguer l'expression des autres membres de la famille. Il faut aussi tenir compte du fait que l'on travaille au niveau du transcriptome, hors il n'y a pas toujours de corrélation entre le niveau d'expression et la présence de la protéine. Ce phénomène implique donc une certaine prudence car il peut biaiser l'interprétation des résultats.

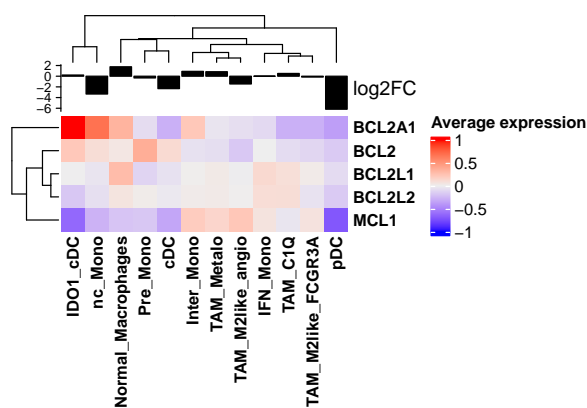
Ces résultats illustrent l'impact et les changements profonds de l'environnement qui peuvent être induits par les traitements. Ici notamment nous avons observé que les BH3-mimétiques favorisent les types de T les plus matures et cytotoxiques. Ceux-ci résistent mieux aux traitements que les types non cytotoxiques comme les helpers ou les Treg. Ainsi, ce remodelage de l'environnement paraît agir en défaveur de la tumeur en levant certains mécanismes immunosuppresseurs. Ces différences sont en partie expliquées par les différences d'expression des molécules de la famille BCL2. La régulation de l'apoptose est un phénomène complexe et il sera nécessaire de s'intéresser aux autres membres de la famille pour mieux comprendre les facteurs qui expliquent la sensibilité des types cellulaires. Par ailleurs, pour certains types cellulaires il est possible que l'effet observé soit indirect. L'élimination d'une population fournissant des signaux de survie pourrait avoir un effet en cascade sur l'écosystème tumoral.



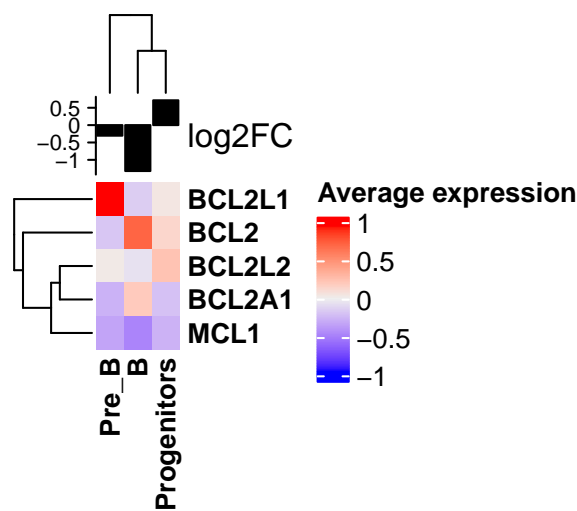
(a) Heatmap de l'expression des membres anti-apoptotiques de la famille BCL2 dans les populations de T



(b) Heatmap de l'expression des membres anti-apoptotiques de la famille BCL2 dans les populations de NK



(c) Heatmap de l'expression des membres anti-apoptotiques de la famille BCL2 dans les populations myéloïdes



(d) Heatmap de l'expression des membres anti-apoptotiques de la famille BCL2 dans les populations B et les progéniteurs

FIGURE 4.40 – Heatmaps de l'expression des gènes des membres anti-apoptotiques de la famille BCL2, les barplots au dessus des heatmap représentent les log2 fold change des moyennes des proportions dans chaque condition

DISCUSSION ET CONCLUSIONS

Les travaux présentés dans cette thèse contribuent à une meilleure compréhension de la complexité du MM. Pour apporter un nouvel éclairage sur différents aspects de cette pathologie, nous tirons partie des technologies de séquençage en cellule unique. Ces technologies sont au coeur des travaux que j'ai réalisés.

Le séquençage en cellule unique nous permet d'accéder à un niveau de résolution sans précédent. Ces méthodes et le séquençage en général ont un coût élevé. Cependant, les coûts de séquençage ont considérablement baissé ces dernières années et continuent en ce sens. L'arrivée de nouvelles alternatives moins coûteuses est également l'un des leviers qui permet et permettra une plus large utilisation de ces méthodes. Jusqu'ici, au sein de notre équipe, nous avons exclusivement utilisé le chromium proposé par la société 10X genomics. Des essais sont en cours avec la technologie Parse. Les premiers résultats obtenus sont convaincants.

Un autre aspect important des technologies en cellule unique est la complexité des analyses. Les jeux de données sont complexes et les phénomènes tels que le dropout requièrent l'utilisation d'outils mathématiques spécialisés. Le développement de ces outils est un champ de recherche très actif. De nouvelles méthodes sont régulièrement disponibles. Des environnements tels que les packages Seurat (R) (HAO et al. 2021) ou scanpy (python) rassemblent des outils pour traiter différents aspects des données. Cependant, les méthodes nouvelles sont généralement disponibles de façon indépendantes. Ces derniers aspects rendent parfois difficile d'établir les méthodes et paramètres optimaux à utiliser sur un jeu de données spécifique. D'autant plus que l'utilisation de certains outils est incompatible avec certains designs expérimentaux. Cette incompatibilité existe aussi entre certaines méthodes, qui nécessitent un pré-traitement spécifique des données. De très bons articles de revues permettent néanmoins d'orienter le choix vers les méthodes les plus adaptées au design et à la question expérimentale (LUECKEN et THEIS 2019; HEUMOS et al. 2023). Ils sont appuyés par des articles de benchmark des différents outils disponibles pour traiter les différents aspects des données. Des approches utilisant

le machine learning apparaissent de plus en plus. Ces approches sont généralement plus puissantes que les méthodes "traditionnelles". Dans le cas du jeu de données MYRACLE, nous nous intéressons plus particulièrement à EcoTyper (LUCA et al. 2021 ; STEEN et al. 2021). EcoTyper permet d'identifier les types cellulaires et d'établir des co-associations entre ces types cellulaires. Une fois le modèle établi sur un jeu de données de référence, il peut être appliqué sur de nouvelles données.

Mes travaux ont contribué à la mise en place d'un environnement et de pipelines d'analyses de données en cellule unique au sein de l'équipe. Les outils mis en place sont robustes et permettent une bonne reproductibilité des analyses. Néanmoins, comme nous l'évoquions dans le paragraphe précédent, la multiplication des outils disponibles nécessite des mises à jour régulières.

Les travaux présentés dans cette thèse illustrent l'intérêt des approches en cellule unique pour répondre à des questions biologiques liées aux modes d'action et aux mécanismes de résistance aux traitements. L'application des méthodes d'analyse mises en place sur deux problématiques a permis de mieux comprendre le MM.

Nos travaux sur le mode d'action de la dex ont permis de mieux comprendre les mécanismes impliqués dans la réponse à ce traitement. Les données d'accessibilité du génome ont permis de montrer l'importance de l'ouverture coordonnée de loci impliqués dans la régulation des gènes cibles de la dex. Nous avons également pu montrer que ces réseaux de régulations ont des caractéristiques différentes en fonction du gène ciblé et qu'ils s'organisent autour d'un enhancer prédominant. Ces résultats sont issus de l'utilisation de cicero (PLINER et al. 2018). Cet outil permet de prédire les interactions par l'intermédiaires de scores de co-accessibilité. Ces résultats viennent renforcer et approfondir ceux obtenus par les techniques "bulk". Le volet transcriptomique des données générées sur la lignée MM.1S nous a permis d'accéder à la diversité de réponses présentes au sein de cette population de cellules homogène, ce qui n'aurait pas été possible avec les méthodes "bulk". En effet, nous avons montré qu'au sein de cette population seule une partie de l'ensemble des gènes de réponse à la dex est exprimé dans chaque cellule. De plus, nous avons montré que les gènes induits au sein d'une cellule ne sont pas choisis au hasard. Deux groupes de gènes sont préférentiellement co-exprimés : un groupe de 45 gènes comprenant les gènes ubiquitaires de la réponse à la dex tels que *TSC22D3* ou *DDIT4* et l'autre comprenant 6 gènes dont *CXCR4*. Le groupe principal comprend également *BCL2L11*, qui exerce une activité pro-apoptotique. *CXCR4* quant à lui est plutôt associé à la survie et à la prolifération tumorale. Ces résultats montrent qu'il existe deux destins transcriptomiques possibles en

réponse à la dex.

Le second aspect du MM que nous avons exploré est la composition de l'écosystème tumoral. À partir d'échantillons de patients obtenus à différents stades de la maladie, nous avons pu établir le paysage détaillé de la moelle du MM.

Nous avons pu identifier finement les types cellulaires présents, ainsi que leurs variations. Nous avons identifié des altérations récurrentes aux différents stades de la maladie, bien que la composition en type cellulaire soit très hétérogène entre les patients. Nous avons également pu identifier des changements au niveau de fonctions clés de certains types cellulaires. L'expression des gènes de la cytotoxicité varie notamment chez les T. De plus, certains gènes liés à la perte des fonctions effectrices sont également plus exprimés chez les patients à des stades avancés. L'identification détaillée des types cellulaires nous permet d'utiliser des outils de prédiction des interactions entre cellules. Les résultats de ces analyses permettent de mieux comprendre comment fonctionne l'écosystème tumoral et comment il est influencé par la tumeur. L'analyse des résultats de cette étude n'est pas encore terminée et certains aspects restent à approfondir. Au niveau des populations myéloïdes notamment nous n'avons pas encore exploré les données autant que pour les lymphocytes T. Une étude spécifique de l'évolution des sous clones de plasmocytes malins en présence ou non de BH3 mimetics est réalisée sur cette même cohorte par un autre bioinformaticien en collaboration avec l'équipe 11 du centre. En parallèle des dernières analyses, nous rédigeons l'article qui présentera les résultats obtenus.

BIBLIOGRAPHIE

- AKKAYA, Munir, Kihyuck KWAK et Susan K. PIERCE (13 déc. 2019), « B cell memory : building two walls of protection against pathogens », in : *Nature Reviews Immunology*, p. 1-10, ISSN : 1474-1741, DOI : 10.1038/s41577-019-0244-2.
- ALBERGE, Jean-Baptiste (15 fév. 2021), « Étude des mécanismes de l'initiation et de la progression du myélome multiple par transcriptomique en cellule-unique et cartographie de l'hydroxyméthylome », These de doctorat, Nantes.
- BAIS, Abha S et Dennis KOSTKA (10 sept. 2019), « scds : computational annotation of doublets in single-cell RNA sequencing data », in : *Bioinformatics* 36.4, p. 1150-1158, ISSN : 1367-4803, DOI : 10.1093/bioinformatics/btz698.
- BARWICK, Benjamin G. et al. (2019), « Cell of Origin and Genetic Alterations in the Pathogenesis of Multiple Myeloma », in : *Frontiers in Immunology* 10, ISSN : 1664-3224.
- BAZARBACHI, Abdul Hamid et al. (28 mars 2022), « Induction therapy prior to autologous stem cell transplantation (ASCT) in newly diagnosed multiple myeloma : an update », in : *Blood Cancer Journal* 12.3, Number : 3 Publisher : Nature Publishing Group, p. 1-8, ISSN : 2044-5385, DOI : 10.1038/s41408-022-00645-1.
- BENANIBA, Lina et al. (29 août 2019), « The MYRACLE protocol study : a multicentric observational prospective cohort study of patients with multiple myeloma », in : *BMC Cancer* 19.1, p. 855, ISSN : 1471-2407, DOI : 10.1186/s12885-019-6080-8.
- BESSONNEAU-GABORIT, Victor et Jonathan CRUARD (1^{er} sept. 2023), « Exploring the impact of dexamethasone on gene regulation in myeloma cells », in : *Life Science Alliance* 6.9, avec la coll. de Catherine GUERIN-CHARBONNEL et al., Publisher : Life Science Alliance Section : Research Articles, ISSN : 2575-1077, DOI : 10.26508/lsa.202302195.
- BLOMBERY, Piers et al. (1^{er} mars 2019), « Acquisition of the Recurrent Gly101Val Mutation in BCL2 Confers Resistance to Venetoclax in Patients with Progressive Chronic Lymphocytic Leukemia », in : *Cancer Discovery* 9.3, p. 342-353, ISSN : 2159-8274, DOI : 10.1158/2159-8290.CD-18-1119.

BIBLIOGRAPHIE

- BRAUD, Veronique M. et al. (2022), « LLT1-CD161 Interaction in Cancer : Promises and Challenges », in : *Frontiers in Immunology* 13, p. 847576, ISSN : 1664-3224, DOI : 10.3389/fimmu.2022.847576.
- BULLER, Casey W., Porunelloor A. MATHEW et Stephen O. MATHEW (1^{er} juill. 2020), « Roles of NK Cell Receptors 2B4 (CD244), CS1 (CD319), and LLT1 (CLEC2D) in Cancer », in : *Cancers* 12.7, p. 1755, ISSN : 2072-6694, DOI : 10.3390/cancers12071755.
- CAVO, Michele et al. (avr. 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 », in : *The Lancet. Oncology* 18.4, e206-e217, ISSN : 1474-5488, DOI : 10.1016/S1470-2045(17)30189-4.
- COMAZZETTO, Stefano, Bo SHEN et Sean J. MORRISON (juill. 2021), « Niches that regulate stem cells and hematopoiesis in adult bone marrow », in : *Developmental Cell* 56.13, p. 1848-1860, ISSN : 15345807, DOI : 10.1016/j.devcel.2021.05.018.
- CRINIER, Adeline et al. (mai 2021), « Single-cell profiling reveals the trajectories of natural killer cell differentiation in bone marrow and a stress signature induced by acute myeloid leukemia », in : *Cellular & Molecular Immunology* 18.5, Number : 5 Publisher : Nature Publishing Group, p. 1290-1304, ISSN : 2042-0226, DOI : 10.1038/s41423-020-00574-8.
- CZABOTAR, Peter E. et Ana J. GARCIA-SAEZ (oct. 2023), « Mechanisms of BCL-2 family proteins in mitochondrial apoptosis », in : *Nature Reviews Molecular Cell Biology* 24.10, Number : 10 Publisher : Nature Publishing Group, p. 732-748, ISSN : 1471-0080, DOI : 10.1038/s41580-023-00629-4.
- DAVIDS, Matthew S. et al. (10 mars 2017), « Phase I First-in-Human Study of Venetoclax in Patients With Relapsed or Refractory Non-Hodgkin Lymphoma », in : *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 35.8, p. 826-833, ISSN : 1527-7755, DOI : 10.1200/JCO.2016.70.4320.
- DERRIEN, Jennifer, Sarah GASTINEAU et al. (31 août 2023), « Acquired resistance to a GPRC5D-directed T-cell engager in multiple myeloma is mediated by genetic or epigenetic target inactivation », in : *Nature Cancer*, Publisher : Nature Publishing Group, p. 1-8, ISSN : 2662-1347, DOI : 10.1038/s43018-023-00625-9.
- DERRIEN, Jennifer, Catherine GUÉRIN-CHARBONNEL et al. (9 août 2021), « The DNA methylation landscape of multiple myeloma shows extensive inter- and inpatient heterogeneity that fuels transcriptomic variability », in : *Genome Medicine* 13.1, p. 127, ISSN : 1756-994X, DOI : 10.1186/s13073-021-00938-3.

- DIMITROV, Daniel et al. (9 juin 2022), « Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data », in : *Nature Communications* 13.1, Number : 1 Publisher : Nature Publishing Group, p. 3224, ISSN : 2041-1723, DOI : 10.1038/s41467-022-30755-0.
- DIMOPOULOS, Meletios A. et al. (20 fév. 2015), « Role of magnetic resonance imaging in the management of patients with multiple myeloma : a consensus statement », in : *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology* 33.6, p. 657-664, ISSN : 1527-7755, DOI : 10.1200/JCO.2014.57.9961.
- DONK, Niels W. C. J. van de et Sonja ZWEEGMAN (8 juill. 2023), « T-cell-engaging bispecific antibodies in cancer », in : *The Lancet* 402.10396, Publisher : Elsevier, p. 142-158, ISSN : 0140-6736, 1474-547X, DOI : 10.1016/S0140-6736(23)00521-4.
- FAVREAU, Mérédis et al. (juill. 2017), « Both mucosal-associated invariant and natural killer T-cell deficiency in multiple myeloma can be countered by PD-1 inhibition », in : *Haematologica* 102.7, e266-e270, ISSN : 0390-6078, DOI : 10.3324/haematol.2017.163758.
- FREUD, Aharon G. et al. (21 nov. 2017), « The Broad Spectrum of Human Natural Killer Cell Diversity », in : *Immunity* 47.5, Publisher : Elsevier, p. 820-833, ISSN : 1074-7613, DOI : 10.1016/j.immuni.2017.10.008.
- FRIEDRICH, Mirco J. et al. (10 avr. 2023), « The pre-existing T cell landscape determines the response to bispecific T cell engagers in multiple myeloma patients », in : *Cancer Cell* 41.4, 711-725.e6, ISSN : 1535-6108, DOI : 10.1016/j.ccell.2023.02.008.
- GABORIT, Victor (30 juin 2021), « Mode d'action moléculaire de la dexaméthasone dans le myélome multiple », These de doctorat, Nantes.
- GERMAIN, Claire, Franck BIHL et al. (1^{er} nov. 2010), « Characterization of Alternatively Spliced Transcript Variants of CLEC2D Gene * », in : *Journal of Biological Chemistry* 285.46, Publisher : Elsevier, p. 36207-36215, ISSN : 0021-9258, 1083-351X, DOI : 10.1074/jbc.M110.179622.
- GERMAIN, Claire, Thierry GUILLAUDEUX et al. (3 août 2015), « Lectin-like transcript 1 is a marker of germinal center-derived B-cell non-Hodgkin's lymphomas dampening natural killer cell functions », in : *OncoImmunology* 4.8, Publisher : Taylor & Francis _eprint : <https://doi.org/10.1080/2162402X.2015.1026503>, e1026503, ISSN : null, DOI : 10.1080/2162402X.2015.1026503.

BIBLIOGRAPHIE

- GERMAIN, Pierre-Luc et al. (16 mai 2022), *Doublet identification in single-cell sequencing data using scDblFinder*, 10 :979, Type : article, F1000Research, DOI : 10.12688/f1000research.73600.2.
- GOODING, Sarah et al. (14 jan. 2021), « Multiple cereblon genetic changes are associated with acquired resistance to lenalidomide or pomalidomide in multiple myeloma », in : *Blood* 137.2, p. 232-237, ISSN : 0006-4971, DOI : 10.1182/blood.202007081.
- HAO, Yuhan et al. (24 juin 2021), « Integrated analysis of multimodal single-cell data », in : *Cell* 184.13, 3573-3587.e29, ISSN : 0092-8674, DOI : 10.1016/j.cell.2021.04.048.
- HEUMOS, Lukas et al. (1^{er} août 2023), « Best practices for single-cell analysis across modalities », in : *Nature Reviews Genetics* 24.8, Publisher : Nature Publishing Group, p. 1-23, ISSN : 1471-0064, DOI : 10.1038/s41576-023-00586-w.
- ITO, Shigeki (fév. 2020), « Proteasome Inhibitors for the Treatment of Multiple Myeloma », in : *Cancers* 12.2, Number : 2 Publisher : Multidisciplinary Digital Publishing Institute, p. 265, ISSN : 2072-6694, DOI : 10.3390/cancers12020265.
- KITADATE, Akihiro et al. (sept. 2021), « Multiple myeloma with t(11;14)-associated immature phenotype has lower CD38 expression and higher BCL2 dependence », in : *Cancer Science* 112.9, p. 3645-3654, ISSN : 1349-7006, DOI : 10.1111/cas.15073.
- KLENER, Pavel et al. (21 sept. 2021), « BH3 Mimetics in Hematologic Malignancies », in : *International Journal of Molecular Sciences* 22.18, p. 10157, ISSN : 1422-0067, DOI : 10.3390/ijms221810157.
- KOHLHAPP, Frederick J. et al. (6 jan. 2021), « Venetoclax Increases Intratumoral Effector T Cells and Antitumor Efficacy in Combination with Immune Checkpoint Blockade », in : *Cancer Discovery* 11.1, p. 68-79, ISSN : 2159-8274, DOI : 10.1158/2159-8290.CD-19-0759.
- KUMAR, Shaji et al. (30 nov. 2017), « Efficacy of venetoclax as targeted therapy for relapsed/refractory t(11;14) multiple myeloma », in : *Blood* 130.22, p. 2401-2409, ISSN : 0006-4971, DOI : 10.1182/blood-2017-06-788786.
- KUMAR, Shaji K. et al. (20 juill. 2017), « Multiple myeloma », in : *Nature Reviews Disease Primers* 3.1, p. 1-20, ISSN : 2056-676X, DOI : 10.1038/nrdp.2017.46.
- KYLE, R. A. et al. (juin 2010), « Monoclonal gammopathy of undetermined significance (MGUS) and smoldering (asymptomatic) multiple myeloma : IMWG consensus perspectives risk factors for progression and guidelines for monitoring and management », in : *Leukemia* 24.6, p. 1121-1127, ISSN : 1476-5551, DOI : 10.1038/leu.2010.60.

- LAPÔTRE-LEDOUX, Bénédicte et al. (4 juill. 2023), « Incidence des principaux cancers en France métropolitaine en 2023 et tendances depuis 1990 », in.
- LE BOURHIS, Lionel, Yvonne K MBURU et Olivier LANTZ (1^{er} avr. 2013), « MAIT cells, surveyors of a new class of antigen : development and functions », in : *Current Opinion in Immunology*, Lymphocyte development / Tumour immunology / Cancer immunology : Clinical translation 25.2, p. 174-180, ISSN : 0952-7915, DOI : 10.1016/j.coi.2013.01.005.
- LOYAL, Lucie et al. (11 déc. 2020), « SLAMF7 and IL-6R define distinct cytotoxic versus helper memory CD8+ T cells », in : *Nature Communications* 11.1, Number : 1 Publisher : Nature Publishing Group, p. 6357, ISSN : 2041-1723, DOI : 10.1038/s41467-020-19002-6.
- LUCA, Bogdan A. et al. (14 oct. 2021), « Atlas of clinically distinct cell states and ecosystems across human solid tumors », in : *Cell* 184.21, Publisher : Elsevier, 5482-5496.e28, ISSN : 0092-8674, 1097-4172, DOI : 10.1016/j.cell.2021.09.014.
- LUECKEN, Malte D. et Fabian J. THEIS (19 juin 2019), « Current best practices in single-cell RNA-seq analysis : a tutorial », in : *Molecular Systems Biology* 15.6, e8746, ISSN : 1744-4292, DOI : 10.15252/msb.20188746.
- MAHNKE, Yolanda D. et al. (2013), « The who's who of T-cell differentiation : Human memory T-cell subsets », in : *European Journal of Immunology* 43.11, _eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/eji.201343751>, p. 2797-2809, ISSN : 1521-4141, DOI : 10.1002/eji.201343751.
- MAPLES, Kathryn T., Sara A. SCOTT et Sagar LONIAL (mai 2023), « Novel approaches to the treatment of multiple myeloma », in : *Clinical Advances in Hematology & Oncology : H&O* 21.5, p. 247-256, ISSN : 1543-0790.
- MATHEWSON, Nathan D. et al. (4 mars 2021), « Inhibitory CD161 receptor identified in glioma-infiltrating T cells by single-cell analysis », in : *Cell* 184.5, 1281-1298.e26, ISSN : 0092-8674, DOI : 10.1016/j.cell.2021.01.022.
- MEISTERMANN, Dimitri et al. (2 sept. 2021), « Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification », in : *Cell Stem Cell* 28.9, 1625-1640.e6, ISSN : 1934-5909, DOI : 10.1016/j.stem.2021.04.027.
- MOREAU, Philippe, Michel ATTAL et al. (6 juill. 2019), « Bortezomib, thalidomide, and dexamethasone with or without daratumumab before and after autologous stem-cell transplantation for newly diagnosed multiple myeloma (CASSIOPEIA) : a randomised,

BIBLIOGRAPHIE

- open-label, phase 3 study », in : *Lancet (London, England)* 394.10192, p. 29-38, ISSN : 1474-547X, DOI : 10.1016/S0140-6736(19)31240-1.
- MOREAU, Philippe, Alfred L. GARFALL et al. (11 août 2022), « Teclistamab in Relapsed or Refractory Multiple Myeloma », in : *The New England Journal of Medicine* 387.6, p. 495-505, ISSN : 1533-4406, DOI : 10.1056/NEJMoa2203478.
- MOREAU, Philippe et Cyrille TOUZEAU (30 juin 2022), « T-cell-redirecting bispecific antibodies in multiple myeloma : a revolution ? », in : *Blood* 139.26, p. 3681-3687, ISSN : 1528-0020, DOI : 10.1182/blood.2021014611.
- NERI, Paola et al. (13 nov. 2019), « Mutations and Copy Number Gains of the BCL2 Family Members Mediate Resistance to Venetoclax in Multiple Myeloma (MM) Patients », in : *Blood* 134 (Supplement_1), p. 572, ISSN : 0006-4971, DOI : 10.1182/blood-2019-127593.
- NUTT, Stephen L. et al. (mars 2015), « The generation of antibody-secreting plasma cells », in : *Nature Reviews Immunology* 15.3, p. 160-171, ISSN : 1474-1741, DOI : 10.1038/nri3795.
- PADALA, Sandeep Anand et al. (mars 2021), « Epidemiology, Staging, and Management of Multiple Myeloma », in : *Medical Sciences* 9.1, Number : 1 Publisher : Multidisciplinary Digital Publishing Institute, p. 3, ISSN : 2076-3271, DOI : 10.3390/medsci9010003.
- PLINER, Hannah A. et al. (6 sept. 2018), « Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data », in : *Molecular Cell* 71.5, 858-871.e8, ISSN : 1097-2765, DOI : 10.1016/j.molcel.2018.06.044.
- QUACH, H. et al. (jan. 2010), « Mechanism of action of immunomodulatory drugs (IMiDS) in multiple myeloma », in : *Leukemia* 24.1, Number : 1 Publisher : Nature Publishing Group, p. 22-32, ISSN : 1476-5551, DOI : 10.1038/leu.2009.236.
- RAJKUMAR, S. Vincent et Shaji KUMAR (28 sept. 2020), « Multiple myeloma current treatment algorithms », in : *Blood Cancer Journal* 10.9, Number : 9 Publisher : Nature Publishing Group, p. 1-10, ISSN : 2044-5385, DOI : 10.1038/s41408-020-00359-2.
- RATMAN, Dariusz et al. (5 nov. 2013), « How glucocorticoid receptors modulate the activity of other transcription factors : a scope beyond tethering », in : *Molecular and Cellular Endocrinology* 380.1, p. 41-54, ISSN : 1872-8057, DOI : 10.1016/j.mce.2012.12.014.
- REDDY, Timothy E. et al. (12 jan. 2009), « Genomic determination of the glucocorticoid response reveals unexpected mechanisms of gene regulation », in : *Genome Research* 19.12, Company : Cold Spring Harbor Laboratory Press Distributor : Cold Spring

- Harbor Laboratory Press Institution : Cold Spring Harbor Laboratory Press Label : Cold Spring Harbor Laboratory Press Publisher : Cold Spring Harbor Lab, p. 2163-2171, ISSN : 1088-9051, 1549-5469, DOI : 10.1101/gr.097022.109.
- ROBAK, Pawel et al. (1^{er} nov. 2018), « Drug resistance in multiple myeloma », in : *Cancer Treatment Reviews* 70, p. 199-208, ISSN : 0305-7372, DOI : 10.1016/j.ctrv.2018.09.001.
- ROBERTS, Andrew W. et al. (28 jan. 2016), « Targeting BCL2 with Venetoclax in Relapsed Chronic Lymphocytic Leukemia », in : *The New England Journal of Medicine* 374.4, p. 311-322, ISSN : 1533-4406, DOI : 10.1056/NEJMoa1513257.
- RUBIO, Marie Thérèse, Adèle DHUYSER et Stéphanie NGUYEN (juin 2021), « Role and Modulation of NK Cells in Multiple Myeloma », in : *Hemato* 2.2, Number : 2 Publisher : Multidisciplinary Digital Publishing Institute, p. 167-181, ISSN : 2673-6357, DOI : 10.3390/hemato2020010.
- SLOMP, Anne et Victor PEPPERZAK (20 nov. 2018), « Role and Regulation of Pro-survival BCL-2 Proteins in Multiple Myeloma », in : *Frontiers in Oncology* 8, p. 533, ISSN : 2234-943X, DOI : 10.3389/fonc.2018.00533.
- STEEN, Chloé B. et al. (11 oct. 2021), « The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma », in : *Cancer Cell* 39.10, Publisher : Elsevier, 1422-1437.e10, ISSN : 1535-6108, 1878-3686, DOI : 10.1016/j.ccell.2021.08.011.
- TIRIER, Stephan M. et al. (29 nov. 2021), « Subclone-specific microenvironmental impact and drug response in refractory multiple myeloma revealed by single-cell transcriptomics », in : *Nature Communications* 12.1, Number : 1 Publisher : Nature Publishing Group, p. 6960, ISSN : 2041-1723, DOI : 10.1038/s41467-021-26951-z.
- VALLEJOS, Catalina A. et al. (juin 2017), « Normalizing single-cell RNA sequencing data : Challenges and opportunities », in : *Nature methods* 14.6, p. 565-571, ISSN : 1548-7091, DOI : 10.1038/nmeth.4292.
- VILLANI, Alexandra-Chloé et al. (21 avr. 2017), « Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors », in : *Science* 356.6335, Publisher : American Association for the Advancement of Science, eaah4573, DOI : 10.1126/science.aah4573.
- WALKER, Brian A., Eileen M. BOYLE et al. (17 août 2015), « Mutational Spectrum, Copy Number Changes, and Outcome : Results of a Sequencing Study of Patients With Newly Diagnosed Myeloma », in : *Journal of Clinical Oncology*, Publisher : American Society of Clinical Oncology, DOI : 10.1200/JCO.2014.59.1503.

BIBLIOGRAPHIE

- WALKER, Brian A., Konstantinos MAVROMMATIS et al. (jan. 2019), « A high-risk, Double-Hit, group of newly diagnosed myeloma identified by genomic analysis », in : *Leukemia* 33.1, p. 159-170, ISSN : 1476-5551, DOI : 10.1038/s41375-018-0196-8.
- WALKER, Brian A., Christopher P. WARDELL et al. (25 avr. 2013), « Characterization of IGH locus breakpoints in multiple myeloma indicates a subset of translocations appear to occur in pregerminal center B cells », in : *Blood* 121.17, p. 3413-3419, ISSN : 0006-4971, DOI : 10.1182/blood-2012-12-471888.
- YOUNG, Matthew D. et Sam BEHJATI (20 avr. 2018), « SoupX removes ambient RNA contamination from droplet based single cell RNA sequencing data », in : *bioRxiv*, p. 303727, DOI : 10.1101/303727.
- ZAVIDIJ, Oksana et al. (27 avr. 2020), « Single-cell RNA sequencing reveals compromised immune microenvironment in precursor stages of multiple myeloma », in : *Nature Cancer*, Publisher : Nature Publishing Group, p. 1-14, ISSN : 2662-1347, DOI : 10.1038/s43018-020-0053-3.
- ZHENG, Grace X. Y. et al. (16 jan. 2017), « Massively parallel digital transcriptional profiling of single cells », in : *Nature Communications* 8.1, Number : 1 Publisher : Nature Publishing Group, p. 14049, ISSN : 2041-1723, DOI : 10.1038/ncomms14049.
- ZHU, Yuan Xiao, K. Martin KORTUEM et A. Keith STEWART (avr. 2013), « Molecular mechanism of action of immune-modulatory drugs thalidomide, lenalidomide and pomalidomide in multiple myeloma », in : *Leukemia & lymphoma* 54.4, p. 683-687, ISSN : 1042-8194, DOI : 10.3109/10428194.2012.728597.

Titre : Le Myélome Multiple et son environnement immunitaire à l'échelle de la cellule unique

Mot clés : scRNAseq, Myélome Multiple, immune environnement, bioinformatique

Résumé : Le Myélome Multiple (MM) est un cancer hématologique dont la cellule tumorale dérive du plasmocyte à longue durée de vie. Cette pathologie est caractérisée par une forte hétérogénéité à divers niveaux. Cette hétérogénéité comprend des altérations intrinsèques et extrinsèques aux tumeurs, lesquelles ont un impact sur le pronostic des patients et leurs réponses aux traitements. L'apparition des technologies de séquençage en cellule unique nous permet aujourd'hui d'explorer de nouveaux aspects de cette diversité. Les travaux présentés ici explorent dans un premier temps la diversité de réponse à la dexaméthasone au sein de la lignée MM.1S de MM. Ces travaux montrent qu'au sein de cette population

tumorale homogène il existe une diversité de réponse au traitement. Dans un second temps nous avons travaillé sur le MM et son environnement immunitaire à l'échelle de la cellule unique. Afin de mieux comprendre comment l'environnement immunitaire évolue au cours de la maladie mais aussi sous la pression des traitements. Cet aspect est d'autant plus essentiel que les traitements les plus récents impliquent directement l'environnement immunitaire en le redirigeant contre la tumeur. Mieux caractériser l'environnement immunitaire pourrait donc permettre de mieux prédire la réponse aux traitements mais aussi leurs conséquences pour l'environnement immunitaire.

Title: Multiple myeloma and its immune environment at the single-cell level

Keywords: scRNAseq, Multiple Myeloma, immune environment, bioinformatics

Abstract: Multiple myeloma (MM) is a hematological cancer in which the tumor cell is derived from the long-lived plasma cell. This pathology is characterized by strong heterogeneity at various levels. This heterogeneity includes alterations intrinsic and extrinsic to tumors, which have an impact on patient prognosis and response to treatment. The development of single-cell sequencing technologies has enabled us to explore new aspects of this diversity. The work presented here first explores the diversity of response to dexamethasone within the MM.1S cell line of MM. This work shows that within this homogeneous tumoral population there is a diver-

sity of response to treatment. Secondly, we worked on MM together with its immune environment at the single-cell level. In order to better understand how the immune environment evolves during the course of the disease, but also under the pressure of treatment. This aspect is even more essential as the most recent treatments directly involve the immune environment by redirecting it against the tumor. A better characterization of the immune environment could therefore enable us to better predict the response to treatments, as well as their consequences for the immune environment.