

Université de Montréal

**Expression des histones déméthylases dans les cellules hématopoïétiques
humaines et les leucémies aiguës**

**par
Lucie Pécheux**

**Sciences biomédicales
Faculté de Médecine**

Mémoire présenté à la Faculté de Médecine
en vue de l'obtention du grade de Maîtrise en Sciences (M.Sc.)
en Sciences Biomédicales, option générale

décembre, 2013

© Lucie Pécheux, 2013

RÉSUMÉ

L'importance des modificateurs de la chromatine dans la régulation de l'hématopoïèse et des hémopathies malignes est illustrée par l'histone méthyltransférase Mixed-Lineage Leukemia (MLL) qui est essentielle au maintien des cellules souches hématopoïétiques (CSH) et dont le gène correspondant, *MLL*, est réarrangé dans plus de 70% des leucémies du nourrisson. Les histones déméthylases (HDM), récemment découvertes, sont aussi impliquées dans le destin des CSH et des hémopathies malignes. Le but de ce projet est d'étudier l'expression des HDM dans les cellules hématopoïétiques normales et leucémiques afin d'identifier de potentiels régulateurs de leur destin. Nous avons réalisé un profil d'expression génique des HDM par qRT-PCR et par séquençage du transcriptome (RNA-seq) dans des cellules de sang de cordon (cellules CD34⁺ enrichies en CSH et cellules différenciées) et des cellules de leucémie aiguë myéloïde (LAM) avec réarrangement *MLL*. Les deux techniques montrent une expression différentielle des HDM entre les populations cellulaires. *KDM5B* et *KDM1A* sont surexprimés dans les cellules CD34⁺ par rapport aux cellules différenciées. De plus, *KDM4A* et *PADI2* sont surexprimés dans les cellules leucémiques par rapport aux cellules normales. Des études fonctionnelles permettront de déterminer si la modulation de ces candidats peut être utilisée dans des stratégies d'expansion des CSH, ou comme cible thérapeutique anti-leucémique. Nous avons aussi développé et validé un nouveau test diagnostique pour détecter les mutations de *GATA2* qui code pour un facteur de transcription clé de l'hématopoïèse impliqué dans les LAM. Ces travaux soulignent l'importance des facteurs nucléaires dans la régulation de l'hématopoïèse normale et leucémique.

Mots-clés: épigénétique, histones déméthylases, cellules souches hématopoïétiques, auto-renouvellement, sang de cordon, leucémie myéloïde, *GATA2*, syndrome de MonoMAC.

ABSTRACT

The importance of chromatin modifiers in regulation of hematopoiesis and hematologic malignancies is illustrated by the Mixed-Lineage Leukemia (MLL) histone methyltransferase, which is essential to maintain hematopoietic stem cells (HSC) and whose corresponding gene, *MLL*, is rearranged in over 70% of infant leukemia. The recently discovered histone demethylases (HDM) are also involved in HSC fate and in hematologic malignancies. The purpose of this project is to study the expression of HDM in normal and leukemic hematopoietic cells to identify potential regulators of their fate. We performed a comprehensive gene expression profile of HDM by qRT-PCR and transcriptome sequencing (RNA-seq) in cord blood cells (CD34⁺ cells enriched in HSC and differentiated cells) and in acute myeloid leukemia (AML) cells with *MLL* rearrangement. Both techniques revealed differential expression of HDM between these cell populations. *KDM5B* and *KDM1A* are overexpressed in CD34⁺ cells compared to differentiated cells. Moreover, *KDM4A* and *PADI2* are overexpressed in leukemic cells compared to normal cells. Functional studies will determine whether modulation of these candidates can be used in HSC expansion strategies or as anti-leukemic drug target. We have also developed and validated a new diagnostic test to detect mutations of *GATA2*, a gene encoding a key transcription factor involved in hematopoiesis and in AML. This work highlights the importance of nuclear factors in the regulation of normal and leukemic hematopoiesis.

Keywords: epigenetic, histone demethylases, hematopoietic stem cells, self-renewal, cord blood, myeloid leukemia, *GATA2*, MonoMAC syndrome.

Table des matières

RÉSUMÉ	i
ABSTRACT	ii
Table des matières	iii
Liste des tableaux	x
Liste des figures	xi
Liste des abréviations	xiv
Remerciements	xvii
CHAPITRE I: INTRODUCTION	1
1. L'HÉMATOPOÏÈSE	1
1.1. Définition	1
1.2. Ontogénie de l'hématopoïèse.....	1
1.3. Les cellules souches hématopoïétiques (CSH)	2
1.3.1. Définition des CSH.....	2
1.3.2. Différenciation et hiérarchie hématopoïétique.....	2
1.3.3. Identification des CSH et détermination de leur fréquence	3
1.3.3.1. Évaluation fonctionnelle des CSH.....	3
1.3.3.2. Marqueurs cellulaires des CSH.....	5
1.3.3.3. Purification / enrichissement en CSH.....	6
1.3.4. Auto-renouvellement et destin cellulaire	6
1.4. Régulation de l'hématopoïèse et des cellules souches hématopoïétiques	7
1.4.1. Niche et facteurs environnementaux.....	7
1.4.2. Facteurs nucléaires régulateurs de l'hématopoïèse.....	8
1.4.2.1. Facteurs de transcription.....	8
1.4.2.2. Modification de la chromatine et régulation épigénétique	11
1.4.2.2.1. Introduction.....	11
1.4.2.2.2. Méthylation de l'ADN	13
1.4.2.2.3. Modifications post-traductionnelles des résidus d'histones et méthylation des résidus d'histones.....	15

1.4.2.2.3.1.	Généralités	15
1.4.2.2.3.2.	Méthylation des histones et MLL	16
1.4.2.2.3.3.	Déméthylation des histones	18
1.4.3.	Remarques conclusives sur la régulation de l'hématopoïèse normale.....	18
1.5.	Conservation de l'hématopoïèse de la souris à l'humain et implications.....	18
2.	LES HISTONES DÉMÉTHYLASES (HDM).....	19
2.1.	Classification des histones déméthylases.....	19
2.1.1.	Réactions de déméthylation des histones.....	21
2.1.2.	Famille LSD (Lysine Specific Demethylase)	22
2.1.3.	Famille JmjC (déméthylases Jumonji C)	23
2.1.3.1.	Sous-famille KDM5/JARID	24
2.1.3.2.	Sous-famille KDM4/JMJD2	24
2.1.3.3.	Sous-famille KDM7/JHDM1	25
2.1.4.	Famille PADI (Peptidyl Arginine Déiminase).....	25
2.2.	Profil d'expression des HDM dans l'hématopoïèse de la souris	26
3.	HÉMOPATHIES MYÉLOÏDES MALIGNES DE L'ENFANT: LEUCÉMIE AIGUË ET SYNDROME MYÉLODYSPLASIQUE	28
3.1.	Les leucémies aiguës.....	28
3.1.1.	Définition et classification des leucémies aiguës.....	28
3.1.2.	Classification des leucémies aiguës de l'enfant.....	28
3.1.3.	Leucémies aiguës lymphoïdes (LAL) de l'enfant.....	28
3.1.3.1.	LAL de type B	29
3.1.3.1.1.	LAL de type B associées à des anomalies génétiques récurrentes	29
3.1.3.1.2.	LAL de type B sans anomalies génétiques récurrentes	29
3.1.3.2.	LAL de type T.....	29
3.1.4.	Leucémies aiguës myéloïdes (LAM) de l'enfant.....	30
3.1.4.1.	LAM avec anomalies génétiques récurrentes	30
3.1.4.2.	LAM avec changements myélodysplasiques	30
3.1.4.3.	Cancers myéloïdes secondaires aux traitements (LAM-t)	30
3.1.4.4.	LAM « autres »	31
3.2.	Les syndromes myélodysplasiques	31

3.3.	Leucémogénèse.....	33
3.3.1.	Cellules souches leucémiques.....	33
3.3.2.	Anomalies géniques des facteurs de transcription	35
3.3.3.	Dérégulation épigénétique	36
3.3.3.1.	Anomalies de MLL.....	36
3.3.3.2.	Autres anomalies épigénétiques impliquées dans les LAM et les SMD	38
3.3.4.	Origine prénatale de la leucémie de l'enfant	39
4.	APPLICATIONS THÉRAPEUTIQUES.....	40
4.1.	Favoriser la prolifération des cellules souches	40
4.2.	Cibler les mécanismes de régulation dysfonctionnels.	41
5.	MÉTHODES D'ANALYSE DE L'EXPRESSION GÉNIQUE	42
5.1.	Définition du transcriptome	42
5.2.	Les puces à ADN ou « micro-arrays ».....	43
5.3.	La qRT-PCR	45
5.4.	Séquençage de l'ARN.....	47
5.5.	Banques de données génomiques publiques.....	49
6.	DIAGNOSTIC DES MUTATIONS DES GÈNES RÉGULATEURS DES CSH	51
7.	HYPOTHÈSE ET OBJECTIFS	51
	CHAPITRE II: MATÉRIEL ET MÉTHODES.....	53
1.	Approche expérimentale	53
2.	Utilisation d'unités de sang de cordon comme source de cellules hématopoïétiques humaines normales	55
2.1.	Base rationnelle du choix d'utiliser le sang de cordon comme source de cellules souches hématopoïétiques humaines	55
2.2.	Critère de choix des unités de sang de cordon et manipulation initiale.....	55
2.3.	Isolement de la population cellulaire différenciée	56
2.4.	Isolement et sélection de la population cellulaire immature.....	57
2.4.1.	Étape 1: pré-enrichissement.....	57
2.4.2.	Étape 2: sélection positive	58
2.5.	Évaluation de l'homogénéité des populations cellulaires isolées	59
2.5.1.	Morphologie des populations cellulaires	59

2.5.1.1.	Cytospin	59
2.5.1.2.	Coloration des lames au GIEMSA.....	59
2.5.1.3.	Observation des lames	60
2.5.2.	Analyse par cytométrie de flux (FACS) du contenu en CD34 ⁺	60
2.5.2.1.	Marquage	60
2.5.2.2.	Récolte de données par cytométrie de flux	60
2.5.2.3.	Analyse des données	61
3.	Source de blastes de leucémie aiguë myéloïde (LAM).....	61
3.1.	Lignée cellulaire humaine MONO-MAC-1	61
3.2.	Cellules leucémiques de LAM pédiatrique de la BCLQ (Banque de Cellules Leucémiques du Québec)	62
4.	Extraction de l'ARN total	62
4.1.	Procédure d'extraction de l'ARN total	62
4.2.	Procédure de contrôle de qualité de l'ARN	62
5.	Transcription inverse.....	63
5.1.	Décontamination préliminaire des ARN totaux à la DNase I.....	63
5.2.	Réaction de transcription inverse (RT)	63
5.3.	Étape de contrôle de l'ADNc par RT-PCR du gène <i>KDM5B/JARID1B</i>	65
6.	Méthode de quantification de l'expression des gènes étudiés par qRT-PCR.....	67
6.1.	Choix des gènes à analyser	67
6.2.	Conception des tests de qRT-PCR et validation des tests amorces/sonde.....	67
6.3.	Choix des échantillons d'ADNc	70
6.4.	Choix du contrôle endogène	70
6.5.	Choix de l'échantillon calibrateur.....	71
6.6.	PCR quantitative	71
6.7.	Méthode de calcul d'expression.....	71
7.	Méthode d'analyse du profil d'expression par séquençage du transcriptome.....	73
8.	Données provenant des banques de données publiques « TCGA ».....	74
9.	Analyses statistiques	75
10.	Considérations éthiques.....	75

11. Conception et réalisation du test diagnostique pour rechercher une mutation de <i>GATA2</i>	76
11.1. Détermination de la région à tester et conception du test.....	76
11.2. Échantillons.....	77
11.3. PCR.....	78
CHAPITRE III: RÉSULTATS	81
SECTION 1: ÉTUDE D'EXPRESSION DES HISTONES DÉMÉTHYLASES DANS LES CELLULES HÉMATOPOÏÉTIQUES NORMALES ET LEUCÉMIQUES	81
1. Profil d'expression des histones déméthylases dans les cellules hématopoïétiques normales	81
1.1. Isolement de deux populations cellulaires (population immature CD34 ⁺ et cellules différenciées) à partir d'unités de sang de cordon	81
1.1.1. Caractéristiques générales des populations cellulaires isolées	81
1.1.2. Les populations cellulaires immatures isolées sont morphologiquement homogènes et enrichies en cellules CD34 ⁺	82
1.2. Les histones déméthylases sont exprimées dans la population cellulaire différenciée.....	84
1.3. Les histones déméthylases sont exprimées dans la population cellulaire immature CD34 ⁺	86
1.4. Certaines histones déméthylases sont différentiellement exprimées dans le compartiment cellulaire immature CD34 ⁺ par rapport au compartiment cellulaire différencié	88
2. Profil d'expression des histones déméthylases dans les cellules de leucémie aiguë	90
2.1. Sources de cellules de leucémie aiguë utilisées	90
2.1.1. Cellules blastiques de la lignée cellulaire MONO-MAC-1 utilisées pour l'analyse d'expression par qRT-PCR.....	90
2.1.2. Échantillons de leucémie aiguë myéloïde pédiatrique (LAM) utilisés pour l'analyse d'expression par séquençage de l'ARN	91
2.2. Profil d'expression des histones déméthylases dans les cellules de leucémie aiguë	93
2.2.1. Les histones déméthylases sont exprimées dans les cellules leucémiques de la lignée MONO-MAC-1	93
2.2.2. Les histones déméthylases sont exprimées dans les blastes de leucémie aiguë myéloïde pédiatrique.....	94

2.3.	Certaines histones déméthylases sont différentiellement exprimées dans les cellules de leucémie aiguë par rapport aux cellules hématopoïétiques immatures CD34 ⁺ de sang de cordon..	95
3.	Histones déméthylases non exprimées dans une ou plusieurs populations cellulaires	96
4.	Résumé des principales variations d'expression des histones déméthylases dans les différentes populations cellulaires analysées par séquençage de l'ARN.....	98
5.	Niveau d'expression des histones déméthylases dans le reste des tissus non-hématopoïétiques.....	99
6.	Profil d'expression des histones déméthylases dans les blastes de leucémie aiguë myéloïde de la banque de données du projet TCGA	103
	SECTION 2: MISE EN PERSPECTIVE DE L'EXPRESSION DES HISTONES DÉMÉTHYLASES (HDM) PAR RAPPORT AU TRANSCRIPTOME COMPLET DANS LES CELLULES HÉMATOPOÏÉTIQUES NORMALES ET LEUCÉMIQUES	105
1.	Profil d'expression quantitative du transcriptome complet	105
2.	Répartition de l'expression des gènes reconnus essentiels pour les cellules souches hématopoïétiques par rapport au transcriptome global	107
3.	Répartition des HDM en fonction de l'importance de leur expression par rapport au transcriptome complet.....	108
4.	Expression des gènes essentiels aux CSH dans les cellules immatures CD34⁺ normales et dans les cellules différenciées	110
1.	Histoire clinique: syndrome myélodysplasique atypique chez un jeune garçon et suspicion de syndrome de MonoMAC	112
2.	Résultat de l'analyse de séquençage du gène <i>GATA2</i>	114
	CHAPITRE IV: DISCUSSION	116
1.	L'étude des cellules souches hématopoïétiques humaines.....	116
2.	Intérêt comparé des différentes sources de cellules leucémiques	117
3.	Principales variations d'expression des histones déméthylases dans les cellules normales et leucémiques.....	117
4.	Patron de méthylation des histones et hématopoïèse : lien avec le destin cellulaire ?	122
5.	Intérêt et limites des études d'expression	123
6.	Perspectives ouvertes par notre étude d'expression des HDM.....	126
7.	Remarques conclusives sur les mutations de <i>GATA2</i>	128

CHAPITRE V: CONCLUSION	131
BIBLIOGRAPHIE.....	133

Liste des tableaux

Tableau 1. Classification des histones déméthylases (HDM).....	21
Tableau 2. Transcription inverse: réactifs et conditions	64
Tableau 3. Réaction de PCR pour étudier le gène <i>KDM5B</i> : réactifs et conditions	66
Tableau 4. Liste des gènes testés et séquences des amorces spécifiques utilisées pour la qRT-PCR	69
Tableau 5. Caractéristiques des réactions de qRT-PCR: réactifs et conditions	71
Tableau 6 Séquençage de l'ARN des échantillons étudiés: profondeur de lecture	73
Tableau 7. Réaction de PCR utilisée pour détecter les mutations du gène <i>GATA2</i> : séquences des amorces sélectionnées, réactifs et conditions.....	79
Tableau 8. Caractéristiques des unités de sang de cordon et des populations cellulaires isolées.....	82
Tableau 9. Contenu en CD34 ⁺ des populations cellulaires récoltées.....	84
Tableau 10. Caractéristiques biologiques des leucémies aiguës myéloïdes pédiatriques sélectionnées pour analyse d'expression par séquençage de l'ARN	92
Tableau 11. Histones déméthylases non-exprimées dans une des populations cellulaires analysées	97
Tableau 12. Résumé des histones déméthylases qui présentent une variation importante d'expression entre les populations cellulaires analysées	98
Tableau 13A. Niveau d'expression moyen des histones déméthylases dans les tissus hématopoïétiques et d'autres tissus humains (www.biogps.com)	101
Tableau 13B. Ratio des niveaux d'expression moyen des histones déméthylases dans le tissu hématopoïétique comparé au niveau moyen d'expression dans 79 tissus humains (www.biogps.com).....	102

Liste des figures

Figure 1. Modèle hiérarchique proposé de l'hématopoïèse et de la différenciation.....	3
Figure 2. Principe de la quantification des cellules souches hématopoïétiques par l'essai par dilution limite et la xénotransplantation de cellules humaines dans des souris NOD/SCID	4
Figure 3. Facteurs de transcription impliqués dans la régulation de l'hématopoïèse	10
Figure 4. Principe des relations entre méthylation des histones, condensation de la chromatine et expression génique.....	12
Figure 5. Découvertes épigénétiques dans la chronologie de l'hématologie.....	16
Figure 6. Antagonisme des complexes Trithorax et Polycomb dans l'expression de gènes clés des cellules souches hématopoïétiques	17
Figure 7. Mécanismes catalytiques de déméthylation	22
Figure 8. Profil d'expression différentielle chez la souris des HDM de la sous-famille JmjC dans les CSH et les progéniteurs immatures par rapport aux cellules différenciées.....	27
Figure 9. Modèle théorique du développement clonal d'une leucémie et notion de cellule souche leucémique	34
Figure 10. Représentation schématique de MLL et des fusions impliquant MLL	37
Figure 11. Effet de la surexpression de KDM4C/JMJ2C	39
Figure 12. Traitements épigénétiques en développement et leurs cibles.....	41
Figure 13. Principe de la méthode de quantification du transcriptome par puce à ADN	44
Figure 14. Exemple de méthode de quantification du transcriptome par qRT-PCR.....	46
Figure 15. Méthode de quantification du transcriptome par séquençage de l'ARN.....	48
Figure 16. Schéma résumant l'impact proposé des histones déméthylases sur les programmes transcriptionnels propres aux cellules souches hématopoïétiques, aux cellules différenciées et leur dérégulation dans les leucémies.....	52
Figure 17 Approche expérimentale du projet	54
Figure 18. Représentation schématique du principe de chaque étape de sélection	58
Figure 19. Exemple de résultats d'extraction d'ARN total.....	63
Figure 20. Exemple de résultats de l'amplification du gène <i>KDM5B/JARID1B</i>	66
Figure 21. Anomalies génétiques décrites dans le gène <i>GATA2</i>	76
Figure 22. Conception du test diagnostique pour la détection des mutations de <i>GATA2</i>	77

Figure 23. Exemple de résultat de l'amplification d'un des exons codant du gène <i>GATA2</i>	79
Figure 24. Populations cellulaires hématopoïétiques normales isolées à partir d'unités de sang de cordon	82
Figure 25. Contenu en CD34 ⁺ des populations cellulaires récoltées par cytométrie de flux.....	83
Figure 26. Profil d'expression des histones déméthylases dans le compartiment cellulaire différencié (sang de cordon)	85
Figure 27. Profil d'expression des histones déméthylases dans le compartiment cellulaire immature CD34 ⁺ (sang de cordon).....	87
Figure 28. Profil d'expression relative des histones déméthylases dans les cellules immatures CD34 ⁺ par rapport aux cellules différenciées (sang de cordon)	89
Figure 29. Cellules leucémiques de la lignée MONO-MAC-1.....	91
Figure 30. Échantillon de leucémie aiguë myéloïde pédiatrique.....	91
Figure 31. Profil d'expression des histones déméthylases dans les blastes de la lignée MONO-MAC-1 par qRT-PCR.....	93
Figure 32. Profil d'expression des histones déméthylases dans les blastes de leucémies aiguës myéloïdes pédiatriques par séquençage de l'ARN.	94
Figure 33 Profil d'expression relative des histones déméthylases dans les blastes de leucémies aiguës myéloïdes pédiatriques par rapport aux cellules sanguines immatures CD34 ⁺ de sang de cordon par séquençage de l'ARN	96
Figure 34. Profil d'expression des histones déméthylases dans les blastes de LAM adultes de la banque de données du projet TCGA.....	104
Figure 35. Répartition des gènes du transcriptome complet en fonction de leur expression quantitative par séquençage de l'ARN	106
Figure 36 Répartition des gènes clés de l'hématopoïèse en fonction de l'importance de leur expression dans le transcriptome complet	107
Figure 37. Répartition des histones déméthylases en fonction de leur expression quantitative par rapport au transcriptome complet par séquençage de l'ARN	110
Figure 38. Expression des 19 gènes essentiels aux cellules souches hématopoïétiques dans les populations cellulaires immatures CD34 ⁺ et les cellules différenciées par séquençage de l'ARN.....	110
Figure 39. Caractéristiques cliniques et biologiques du cas présenté.....	113

Figure 40. Mise en évidence d'une mutation ponctuelle hétérozygote c.1186 C→T R396W de *GATA2* dans l'échantillon du patient114

Figure 41. Le pedigree familial montre le caractère sporadique de la mutation R396W de ce cas clinique.....115

Liste des abréviations

2-OG: 2-oxoglutarate (aussi appelé α -cétoglutarate)

ADN: acide désoxyribonucléique

ADNc: ADN complémentaire

ADNg: ADN génomique

AGM: Aorte-Gonade-Mesonephros

ARN: acide ribonucléique

AR: auto-renouvellement

ATH: AT hook

BCLQ: Banque de Cellules Leucémiques du Québec

BCR: breakpoint cluster region

CAT#: numéro de catalogue

CMP: progéniteurs myéloïdes communs (« common myeloid progenitors »)

CpG: dinucléotides cytosine-guanine

CSE: cellules souches embryonnaires

CSL: cellules souches leucémiques

CSH: cellules souches hématopoïétiques

CSPH: cellules souches et progéniteurs hématopoïétiques

Ct: Cycle seuil (« treshold cycle »)

DEPC: diéthyl pyrocarbonate

DMSO: diméthyle sulfoxide

DNMT: méthyltransférase de l'ADN

FAB: French-American-British

FAD: flavine adénine dinucléotide

FACS: cytométrie de flux (« Fluorescence Activated Cell Sorting »)

FBS: sérum fœtal bovin (« Fetal Bovine Serum »)

FT: facteur de transcription

EDTA: acide éthylène diamine triacétique

GMP: progéniteurs des granulocytes et des macrophages (« granulocyte-macrophage progenitors »)

HAT: histones acétyltransférases

HDAC: histones déacétylases

HDM: histones déméthylases

HMT: histones méthyltransférases

HSC: cellules souches hématopoïétiques (« hematopoietic stem cells »)

IDH: isocitrate déshydrogénase

IL1: interleukine 1

IL6: Interleukine 6

IRIC: Institut de Recherche en Immunologie et en Cancérologie

JmjC: déméthylase jumonji C

JMJC: domaine catalytique JMJC

KDM: histone lysine déméthylase

KMT: histone lysine méthyltransférase

LA: leucémie aiguë

LAL: leucémie aiguë lymphoïde

LAM: leucémie aiguë myéloïde

LDA: méthode des dilutions limites (« Limiting-Dilution Assay »)

LSD: déméthylases spécifiques des lysines (« lysin-specific demethylases »)

MEP: progéniteurs mégacaryocytaires et érythroïdes (« megakaryocytic and erythroid progenitors »)

MLL: Mixed Lineage Leukemia

MLL-PTD: duplication en tandem de MLL

MLP: progéniteurs communs lymphoïdes (« multilymphoid progenitors »)

MM1: lignée cellulaire MONO-MAC-1

MO: moelle osseuse

MPP: progéniteurs multipotents (« multipotent progenitors »)

MPTH: modification post traductionnelle des histones

MRC: Medical Research Council

NLS: domaine signal de localisation nucléaire

NRD: domaine de régulation négative
PADI: déméthylases peptidyl arginine déiminases
pb: paire de bases
PBS: tampon phosphate (Phosphate Buffered Salin)
PcG: Polycomb
qRT-PCR: quantitative reverse transcription polymerase chain reaction
RNAi: ARN interférence
RNA-seq: séquençage de l'ARN ("RNA-sequencing")
RPKM: nombre de lectures alignées par kilobase par million de lectures alignées (« reads per kilobase of exon per million mapped reads »)
SCF: stem cell factor
SET: domaine "Su(var)3-9, enhancer of-zeste trithorax"
SHL: domaine "speckled nuclear localization"
SMD: syndrome myélodysplasique
SNG: séquençage de nouvelle génération
SP: sang périphérique
TCGA : banque de données TCGA ou « The Cancer Genome Atlas Data Portal »
TAD: domaine d'activation de la transcription (transactivation domain)
t-LAM: leucémie aiguë myéloïde secondaire à un traitement
t-LAL: leucémie aiguë lymphoïde secondaire à un traitement
t-SMD: syndrome myélodysplasique secondaire à un traitement
TRD: domaine de répression transcriptionnelle (« Transcriptional Repression Domain »)
USC: unité de sang de cordon
ZF: domaine en doigt de zinc

Remerciements

Je tiens avant tout à remercier très chaleureusement ma directrice de recherche Dr Sonia Cellot, pour m'avoir fait confiance et m'avoir donné la possibilité de réaliser ce projet. Je la remercie aussi pour ses encouragements, sa disponibilité et son mentorat qui dépasse le cadre de cette maîtrise. Je voudrais aussi remercier Dr Josée Hébert, co-directrice de recherche, pour avoir accepté de me guider avec Dr Cellot, pour la grande qualité de son encadrement et de son enseignement. Je voudrais aussi très spécifiquement remercier Louise Laramée, assistante de recherche, pour tout ce qu'elle m'a appris grâce à ses formidables compétences scientifiques, pédagogiques et surtout humaines, pour sa disponibilité et son positivisme inébranlable.

Je souhaite également remercier toutes les autres personnes qui m'ont aidée pour ce projet et particulièrement: Benjamin Turgeon, l'équipe de cytogénétique de la Banque de Cellules Leucémiques du Québec, l'équipe du laboratoire du Dr Duval, l'équipe de la plateforme de génomique de l'IRIC, Jalila Chagraoui et l'équipe du Dr Sauvageau (IRIC), Brian Wilhelm, etc.... C'est une chance de pouvoir collaborer dans de si bonnes conditions. Je voudrais aussi mentionner ici les patients dont j'ai eu le privilège de m'occuper, et leurs parents, puisqu'ils sont ma principale motivation.

Je remercie bien sûr mes très précieuses et inspirantes amies sans qui rien n'est possible, particulièrement Marjo, Rosemonde, Josianne et Mathilde, mais aussi mon demi-frère Thomas et tous ceux et celles que j'ai la chance d'avoir à mes côtés. Enfin, je remercie ma merveilleuse fille Alia, qui me donne l'envie de me dépasser quoi qu'il arrive et bien sûr mes parents pour avoir fait de moi qui je suis: ma mère, qui serait fière de moi, et mon père pour son soutien et son affection indéfectibles. Tu es un roc et un repère, papa. Merci pour tout.

*À ma fille, Alia,
À ma mère, Nicole Pécheux (1946-2004)*

CHAPITRE I: INTRODUCTION

1. L'HÉMATOPOÏÈSE

1.1. Définition

L'hématopoïèse est le processus par lequel le tissu sanguin est capable de produire continuellement toutes les cellules hématopoïétiques matures nécessaires durant la vie d'un individu. Chaque jour, à partir de précurseurs communs appelés cellules souches hématopoïétiques (CSH) qui figurent au sommet de ce processus de production hiérarchique, le corps humain produit ainsi dans la moelle osseuse environ 10^{12} cellules différenciées: les lymphocytes T, B et NK, les granulocytes, les monocytes, les macrophages, les globules rouges et les plaquettes[1]. Ces cellules constituent l'essentiel du tissu sanguin et diffèrent non seulement sur le plan fonctionnel mais aussi par leur durée de vie qui peut varier entre une journée pour les monocytes et plusieurs mois pour certains lymphocytes T[2]. Dès lors, l'hématopoïèse repose sur un équilibre permanent finement régulé entre le maintien d'un réservoir de CSH multipotentes et leur différenciation.

1.2. Ontogénie de l'hématopoïèse

La constitution du tissu sanguin et la production de CSH se font progressivement au cours de l'embryogenèse humaine au niveau de différents sites successifs[3]. Ainsi l'hématopoïèse dite «primitive» débute autour du jour 21 du développement embryonnaire dans le mésoderme du sac vitellin sous forme de cellules nucléées situées au centre d'îlots sanguins, capables d'oxygéner les tissus embryonnaires. À partir de la 4^{ème} semaine de développement apparaissent des cellules primitives qui s'apparentent à des CSH au niveau de la région dorsale de l'aorte appelée AGM pour « Aorte-Gonade-Mesonephros » et aussi au niveau du placenta, constituant le début de l'hématopoïèse « définitive ». Ces cellules vont ensuite migrer progressivement à partir du 2^{ème} mois pour coloniser principalement le foie et la rate du fœtus, et y subir une forte expansion. À partir du 4^{ème} mois du développement fœtal, ces CSH migrent de nouveau pour finalement coloniser la moelle osseuse de tous les os, qui

deviendra le lieu exclusif de l'hématopoïèse peu avant la naissance et ce, pour la vie entière de l'individu. À partir de 4 ans, les sites d'hématopoïèse se limitent aux os du squelette axial et de l'os iliaque [4].

Cette évolution spatiale au cours du développement embryonnaire, accompagnée de variations des caractéristiques des CSH primitives en fonction des sites colonisés et leur capacité de migration vers des endroits précis illustrent l'importance d'un microenvironnement spécifique, appelé niche hématopoïétique, pour le destin des CSH. Ces niches hématopoïétiques ne sont pas encore bien définies, mais il est actuellement admis que l'influence de la niche sur le destin des CSH passe par des interactions cellulaires et d'autres signaux solubles acellulaires (cf. Chap. I.1.4.1)[5]. Le microenvironnement n'est cependant pas le seul à influencer le destin des CSH durant le développement de l'hématopoïèse. Il existe également des déterminants intrinsèques essentiels à l'installation de l'hématopoïèse primitive et définitive tels que les facteurs de transcription *TALI*, *LMO2*, *MLL*, *RUNX1*, *BMP*, *GATA2*, etc.[4], et des mécanismes de régulation épigénétique, comme nous le verrons plus loin (cf. Chap. I.1.4.2).

1.3. Les cellules souches hématopoïétiques (CSH)

1.3.1. Définition des CSH

Les CSH sont la clef de voûte de l'hématopoïèse. Elles sont définies fonctionnellement par leur capacité de reconstituer le système sanguin au complet lorsqu'elles sont greffées à un individu receveur dont la moelle a été létalement irradiée. Ceci est rendu possible grâce à leurs deux caractéristiques principales: elles sont non seulement capables de se différencier en tous les types de cellules matures du sang, mais elles sont aussi capables d'auto-renouvellement pour maintenir un réservoir de CSH identiques et ainsi soutenir l'hématopoïèse pendant toute la vie de l'individu.

1.3.2. Différenciation et hiérarchie hématopoïétique

En s'engageant dans la différenciation, les cellules issues des CSH perdent progressivement leur pluripotence, diminuent leur longévité et affirment leur identité spécifique de lignée en changeant également de morphologie. La figure 1 présente un des modèles actuels de

l'organisation hiérarchique de l'hématopoïèse [6]. Les CSH deviennent tout d'abord des progéniteurs multipotents (MPP) puis ceux-ci s'engagent ensuite soit dans la lignée lymphoïde en devenant des progéniteurs lymphoïdes communs (MLP), soit dans la lignée myéloïde en devenant des progéniteurs myéloïdes communs (CMP), qui donneront eux-mêmes naissance aux progéniteurs des granulocytes et macrophages (GMP) ou des mégacaryocytes et des érythrocytes (MEP).

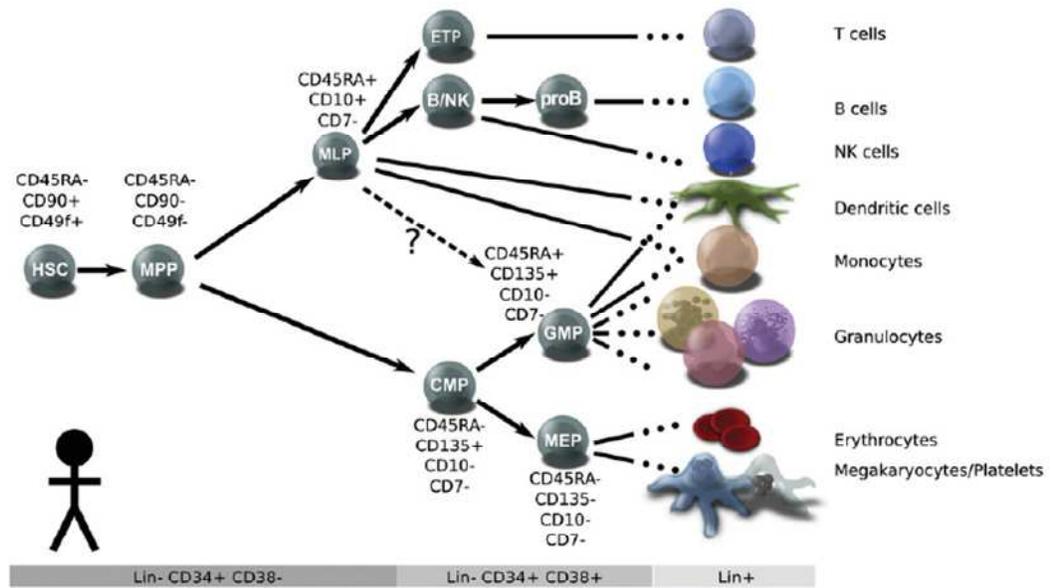


Figure 1. Modèle hiérarchique proposé de l'hématopoïèse et de la différenciation

Figure adaptée de Cell Stem Cell, 10(2), Doulatov, S., et al. (2012)[6] "*Hematopoiesis: a human perspective*", 120-136 Copyright (2012), avec la permission d'Elsevier. La figure présente les principales cellules souches et progénitrices ainsi que leurs marqueurs cellulaires de surface. Les flèches représentent l'engagement des différentes cellules dans une voie de différenciation spécifique aboutissant à la production des cellules différenciées. HSC: cellules souches hématopoïétiques; MPP: progéniteurs multipotents; MLP: progéniteurs lymphoïdes communs; CMP: progéniteurs myéloïdes communs; GMP: progéniteurs des granulocytes et macrophages; MEP: progéniteurs des mégacaryocytes et des érythrocytes.

1.3.3. Identification des CSH et détermination de leur fréquence

1.3.3.1. Évaluation fonctionnelle des CSH

La faible fréquence des cellules souches pluripotentes combinée à l'absence de marqueur immunophénotypique distinctif et spécifique rend difficile toute étude biologique et

moléculaire des CSH. La méthode de référence pour évaluer leur fréquence dans une population cellulaire hématopoïétique humaine *in vivo* est l'essai de dilution limite (ou LDA pour « limiting-dilution assay ») basé sur la xénotransplantation dans des souris irradiées de façon létale ou des souris immunodéficientes NOD/SCID, de cellules hématopoïétiques humaines à des dilutions variables afin d'observer la reconstitution correspondante chez les receveurs et d'en déduire la fréquence des CSH dans l'échantillon initial. La figure 2 en présente brièvement le principe [6, 7].

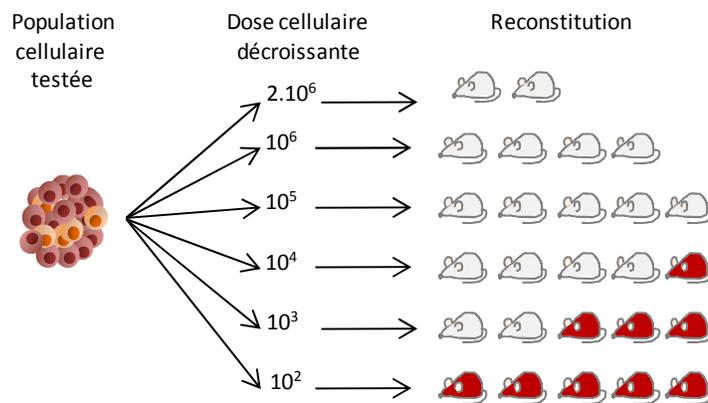


Figure 2. Principe de la quantification des cellules souches hématopoïétiques par l'essai par dilution limite et la xénotransplantation de cellules humaines dans des souris NOD/SCID

Xénotransplantation dans des souris irradiées létalement, d'une population cellulaire hématopoïétique humaine contenant une quantité inconnue de cellules souches hématopoïétiques associée à une population de cellules de moelle osseuse de souris (cellules compétitrices). La greffe se fait avec des doses cellulaires décroissantes dans chaque groupe de receveurs. L'évaluation de la reconstitution et de l'origine des cellules reconstituées se fait par cytométrie de flux après 3-4 mois. Les souris blanches sont celles pour lesquelles il y a une reconstitution et les souris rouges celles pour lesquelles il n'y a pas de reconstitution. La proportion de souris sans reconstitution après greffe dans chaque groupe de receveurs, en fonction du facteur de dilution, permet de déduire la fréquence des cellules souches hématopoïétiques dans la population initiale.

Cette méthode d'identification fonctionnelle des CSH ne permet cependant pas de les isoler pour étudier les mécanismes moléculaires et cellulaires qui sont à la base de leur identité.

1.3.3.2. Marqueurs cellulaires des CSH

De nombreux progrès ont été réalisés depuis la 1^{ère} démonstration de l'existence des CSH en 1961 par Till et McCulloch [8]. Parallèlement à l'amélioration des techniques de quantification fonctionnelle des CSH, différents marqueurs cellulaires membranaires utiles pour leur identification ont été mis en évidence. L'antigène CD34 est le premier marqueur identifié qui permet d'enrichir une population cellulaire hématopoïétique en cellules immatures capables de reconstitution [9]. On le retrouve à la surface non seulement des CSH mais aussi des cellules progénitrices (figure 1). Il est exprimé très précocement dans l'hématopoïèse humaine et son expression diminue avec la différenciation. L'antigène CD34⁺ est considéré comme un marqueur de choix pour identifier ou purifier les cellules immatures et est utilisé en clinique pour l'évaluation du contenu en CSH des greffons. La population des cellules CD34⁺ demeure cependant une population immature hétérogène composée de rares cellules souches et de progéniteurs partiellement différenciés. La fréquence des CSH dans cette population cellulaire CD34⁺, selon les données de la littérature, est estimée à environ 1 dans environ 1000 cellules CD34⁺ [10]. Différents autres marqueurs, par leur présence ou leur absence, ont successivement été démontrés utiles pour identifier avec plus de précision les cellules les plus pluripotentes parmi les CD34⁺. Ainsi, ce sont tout d'abord les antigènes CD45RA[11], CD38[12] et un groupe de marqueurs de différenciation terminale « Lineage » ou Lin[13] (qui marquent les cellules B, T, NK, les cellules dendritiques, les monocytes, les granulocytes, les mégacaryocytes et les érythrocytes) qui ont permis de définir par leur absence une population plus immature. Découvert simultanément, l'antigène CD90 ou Thy1 permet de différencier parmi ces cellules immatures CD34⁺, les CSH (Thy1⁺) des progéniteurs multipotents (Thy1⁻)[13]. Les derniers marqueurs cellulaires à avoir été identifiés pour définir la population cellulaire la plus immature des CSH sont la présence de l'antigène CD49f et la haute capacité d'efflux mitochondrial du colorant Rhodamine123-low (Rho^{lo})[14]. La meilleure définition immunophénotypique des CSH repose actuellement sur la série de marqueurs suivants: CD34⁺ CD45RA⁻ CD38⁻ Lin⁻ Thy1⁺ CD49f⁺ et Rho^{lo} [14]. Il faut cependant ajouter que, dès la fin des années 90, plusieurs auteurs ont commencé à défendre l'idée de l'existence de rares cellules souches immatures et pluripotentes qui n'exprimeraient pas l'antigène CD34, chez

l'homme comme chez la souris[15]. Il semble que cette population cellulaire rare n'exprimant pas les marqueurs CD34⁻ Lin⁻ CD38⁻ associés au marqueur CD93^{hi} se place en amont de la hiérarchie cellulaire hématologique. Ces cellules très immatures se distinguent des CSH CD34⁺ par leur caractère plus quiescent et l'utilisation de processus moléculaires et de voies de signalisation différentes des CD34⁺ [16].

1.3.3.3. Purification / enrichissement en CSH

La fréquence des CSH est estimée à environ une cellule sur 3×10^6 cellules dans la moelle osseuse et de une sur 9×10^5 cellules dans le sang de cordon [17]. En utilisant les différents marqueurs décrits précédemment (antigènes CD34, CD45RA, CD38 et Thy1), il est possible d'enrichir une population de cellules hématopoïétiques CD34⁺ jusqu'à une fréquence de CSH de 5% (cellules CD34⁺ CD38⁻ CD45RA⁻ Thy1⁺, enrichissement en CSH à 1:20 cellules), fréquence déterminée par la méthode des dilutions limites (cf. Chap.1.3.3.1). En ajoutant le marqueur CD49f, la fréquence augmente à 9,5% (Thy1⁺ CD49f⁺, enrichissement en CSH à 1:10,5). Le meilleur degré d'enrichissement atteint à partir de sang de cordon décrit dans la littérature est autour de 20% de CSH en considérant en plus l'efflux de rhodamine (Thy1⁺ CD49f⁺ Rho^{lo}, enrichissement en CSH à environ 1:5 cellules) [14]. Parvenir à isoler des cellules les plus immatures possible en nombre suffisant pour pouvoir procéder à différentes études biologiques est un défi important à relever pour mieux comprendre les processus moléculaires qui sont à la base de l'identité des CSH.

1.3.4. Auto-renouvellement et destin cellulaire

L'auto-renouvellement (AR) des CSH repose sur leur capacité de se diviser en donnant naissance soit à une cellule fille identique à la cellule mère (division asymétrique de l'AR), soit à deux cellules filles identiques à la cellule initiale (division symétrique d'AR), contrairement aux divisions cellulaires de différenciation où les deux cellules filles sont différentes de la cellule mère. L'AR est un processus crucial pour maintenir le réservoir de cellules souches et pour le développer pendant l'ontogénie par exemple, mais aussi pour le restaurer si nécessaire après un déficit transitoire pathologique. Toute anomalie du système d'AR peut conduire au développement de processus pathologiques. Ainsi, son exacerbation peut conduire au développement d'une leucémie par accumulation de blastes, et sa

disparition au profit de divisions symétriques de différenciation peut conduire à l'épuisement du réservoir de CSH comme dans l'anémie aplasique par exemple.

Les CSH sont quiescentes la plupart du temps. En effet, des études de cinétique utilisant la longueur de télomères notamment ou encore par la mesure des ratios d'inactivation du chromosome X, ont permis d'estimer leur fréquence de division à environ une fois par 45 semaines en moyenne (175 - 350 jours) [18, 19]. Lorsqu'elles entrent dans le cycle cellulaire, la détermination du but de leur division (principalement entre AR et différenciation), est associée à l'expression de programmes géniques impliqués soit dans le maintien de l'identité de cellule souche, soit dans l'entrée en différenciation. L'activation de la transcription de l'un de ces programmes d'expression génique, associée à la répression de l'autre est le fruit d'un système complexe de régulation. Les mécanismes moléculaires régulateurs de cette balance d'expression génique sont critiques pour le destin des CSH. En effet, une anomalie au niveau de la régulation de cette balance d'expression peut conduire au développement d'une leucémie par exemple si un arrêt de différenciation survient dans un contexte de cellules relativement immatures avec maintien de leur capacité d'AR. Il est donc très important de mieux comprendre quels sont ces mécanismes régulateurs qui sont encore largement méconnus [6].

1.4. Régulation de l'hématopoïèse et des cellules souches hématopoïétiques

1.4.1. Niche et facteurs environnementaux

Depuis la fin des années 70 la reconnaissance de sites anatomiques et fonctionnels spécifiques appelés niches hématopoïétiques a émergé dans la littérature. Ces niches sont capables de retenir les CSH et de leur permettre soit de rester quiescentes, soit de proliférer sans se différencier, comme nous l'avons évoqué précédemment (cf. Chap. I.1.2.). Ce n'est que lorsqu'elles sortent de leur niche que les cellules semblent capables de se différencier[20]. De plus, des CSH greffées à un individu par voie intraveineuse sont capables de se diriger et de s'établir dans les niches du receveur [21]. Bien que les niches hématopoïétiques de la moelle osseuse humaine soient encore mal définies, des études réalisées chez la souris ont permis d'identifier l'implication d'interactions cellulaires avec des cellules adjacentes (ostéoblastes, cellules endothéliales et/ou cellules souches

mésenchymateuses) ainsi que celles de signaux acellulaires solubles tel que la chémokine CXCL12/SDF-1 (Stromal cell Derived Factor-1) et son interaction avec le récepteur CXCR4[22]. Cependant, l'impact de la niche sur les divisions asymétriques et les choix afférents dans le destin des CSH ne sont pas bien connus car la niche hématopoïétique est difficile à reproduire *ex vivo* et à étudier *in vivo* [23].

Différents facteurs de croissance sont également reconnus pour soutenir spécifiquement les CSH et les progéniteurs immatures de l'hématopoïèse tels que les interleukines IL1 et IL6, ou le SCF (Stem cell factor), même si leurs rôles exacts ne sont pas encore complètement élucidés. Leurs fonctions semblent cependant être plutôt de l'ordre du soutien de la prolifération que dans la détermination du destin de ces cellules [24].

1.4.2. Facteurs nucléaires régulateurs de l'hématopoïèse

En plus des interactions avec le microenvironnement, l'hématopoïèse est régulée par un réseau complexe de différents facteurs nucléaires capable d'activer ou de réprimer les processus transcriptionnels essentiels qui déterminent notamment le destin des CSH. On distingue les facteurs de transcription et les mécanismes épigénétiques.

1.4.2.1. Facteurs de transcription

Les facteurs de transcription (FT), protéines capables de se lier à l'ADN, sont de véritables déterminants intrinsèques du phénotype cellulaire. En effet, par leurs interactions avec les régions régulatrices de gènes spécifiques, ils sont capables d'activer ou de réprimer leur transcription et ainsi influencer directement la fonction de la cellule. Ces FT sont de plus capables d'interaction avec d'autres protéines et leur action est souvent modulée par le contexte moléculaire dans lequel ils opèrent. On distingue habituellement les FT associés aux CSH et aux progéniteurs immatures de ceux impliqués dans l'une ou l'autre des lignées de différenciation, sur la base du fait qu'ils sont principalement exprimés dans une de ces populations cellulaires. La figure 3 présente les plus connus d'entre eux.

Les FT essentiels au fonctionnement des CSH sont principalement: *BM11* (« polycomb ring finger »), *ERG* (« v-ets avian erythroblastosis virus E26 oncogene homolog »), *ETV6/TEL* (« ets variant 6 »), *FL11* (« friend leukemia integration 1 »), *GATA2* (« *GATA binding protein 2* »), *GFI1B* (« growth factor independent 1B »), *LMO2* (« LIM domain only 2 »), *LYL1* («

lymphoblastic leukemia derived seq.1 »), *MEIS1* (« meis homeobox 1»), *MLL* («mixed lineage leukemia»), *PBX1*, *RUNX* («runt-related transcription factor 1», anciennement appelé *AML1*) et *TALI/SCL* («T-cell acute lymphocyte leukemia 1») [4, 25]. De plus, l'un de ces FT, *MLL*, régule l'expression des gènes homeobox de classe I ou gènes *HOX* qui codent pour des FT également déterminants pour l'identité des CSH [26]. Il existe 39 gènes *HOX* organisés en 4 groupes différents de A à D. Ces gènes ont un rôle important non seulement dans l'embryogenèse pour l'organisation spatio-temporelle des segments du corps le long d'un axe antéropostérieur [27], mais aussi dans l'hématopoïèse précoce notamment par leur soutien de l'expansion des CSH et de l'AR (membres du groupe A tels que *HOXA5*, *HOXA7*, *HOXA9* et *HOXA10*, mais aussi *HOXB4*) [28, 29].

Les facteurs plus spécifiques d'une lignée ont souvent la capacité d'inhiber les FT des autres lignées ou ceux des cellules souches pour orienter la différenciation, comme par exemple l'antagonisme entre l'expression de *GATA1* et de *SPI1/PU.1* (« spleen focus forming virus proviral integration oncogene ») qui favorisent respectivement la différenciation des érythrocytes et des mégacaryocytes ou celle des éléments de la lignée myéloïde, ou encore comme la transition entre l'expression de *GATA2* et celle de *GATA1* pendant la différenciation des CSH et l'engagement dans la lignée érythroïde et mégacaryocytaire [30]. *GATA2* est un FT important pour la prolifération des CSH et des progéniteurs, mais aussi pour la différenciation terminale des mastocytes [31]. Un autre FT important, *CEBPA* (ou *C/EBP α* ; « CCAAT/enhancer binding protein alpha »), est aussi impliqué dans la différenciation myéloïde.

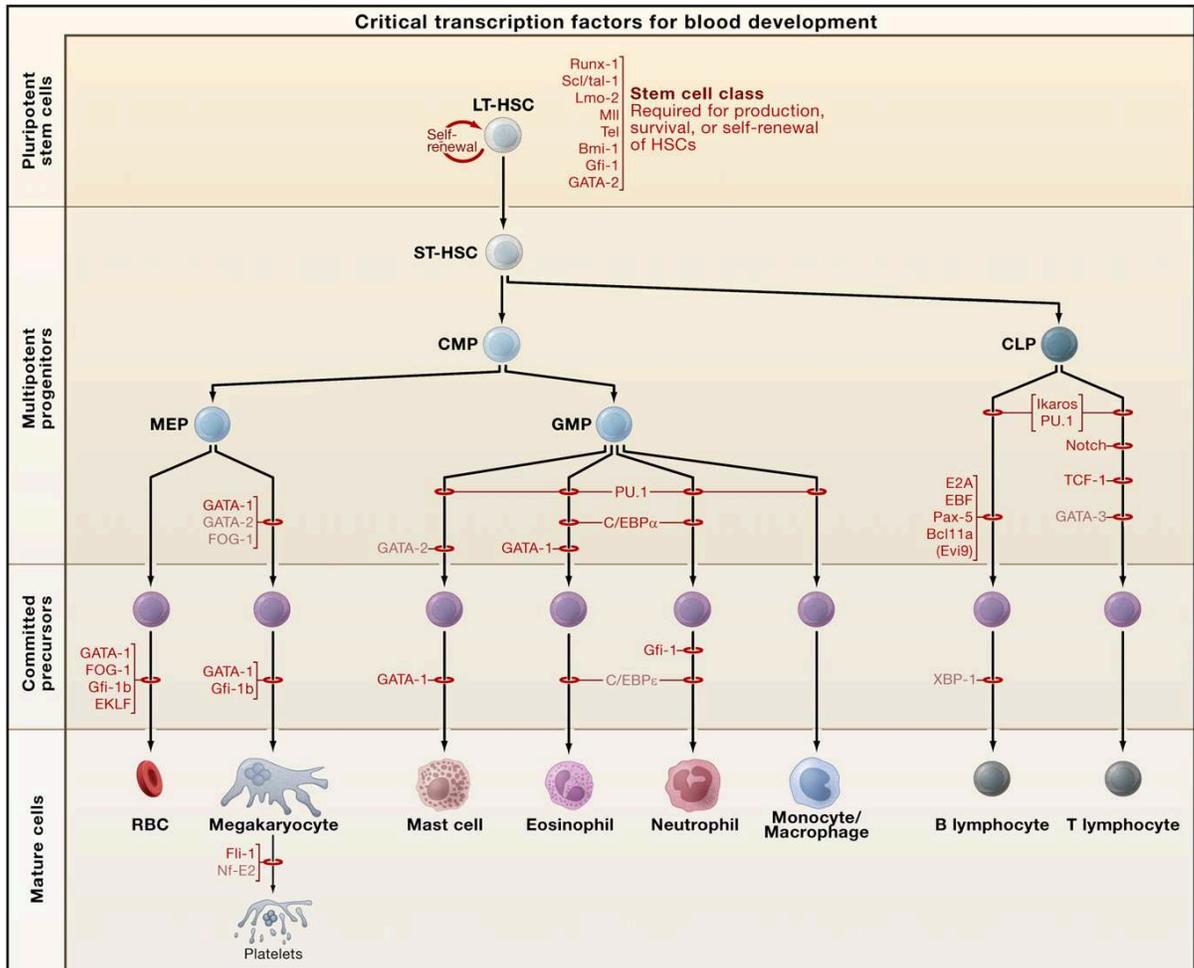


Figure 3. Facteurs de transcription impliqués dans la régulation de l'hématopoïèse

Figure adaptée de Cell 132(4), Orkin S.H., *et al.* (2008)[4] "*Hematopoiesis: an evolving paradigm for stem cell biology*", 631-644. Copyright (2008), avec la permission d'Elsevier. Les facteurs de transcription sont représentés en fonction des étapes de l'hématopoïèse où se situe l'arrêt de maturation (démonstré ou proposé) lorsqu'ils sont absents. Les facteurs de transcription dont le nom est écrit en rouge pâle sont ceux qui n'avaient pas encore été retrouvés anormaux dans des hémopathies malignes lors de la parution de l'article en 2008.

Cette classification des FT entre CSH et lignées est cependant générale. Elle reflète principalement les conséquences majeures de leur absence ou de leur dysfonction lorsqu'ils sont impliqués dans des translocations chromosomiques, par exemple en termes d'arrêt de la maturation à un niveau hématopoïétique particulier (tel que présenté dans la figure 3). En effet, ce sont souvent les données des patients présentant des hémopathies avec phénotypes particuliers qui ont permis de mettre en évidence le rôle essentiel de certains d'entre eux,

après validation par des études d'inhibition et de surexpression dans des modèles murins. Le rôle des autres FT est extrapolé à partir de l'effet de leur absence dans les modèles murins de développement de l'hématopoïèse.

De plus l'impact d'un FT ne dépend pas seulement de son niveau d'expression, mais aussi du contexte cellulaire et moléculaire dans lequel il s'exprime. En effet, la réinduction de l'expression de FT essentiels aux CSH dans des fibroblastes ne suffit pas à reprogrammer ces cellules différenciées en cellules souches [32]. En outre, il a été démontré que les cellules différenciées sont capables de maintenir leur identité même lorsque les stimuli de différenciation tels que l'expression de FT spécifiques disparaissent, suggérant qu'un autre degré de régulation suffit ensuite à maintenir ce qui peut être appelé la «mémoire» de l'identité cellulaire [33].

1.4.2.2. Modification de la chromatine et régulation épigénétique

1.4.2.2.1. Introduction

Durant la différenciation des cellules hématopoïétiques à partir des CSH et parallèlement à leur perte de pluripotence, la structure des noyaux cellulaires évolue vers une condensation progressive de la chromatine. De plus, lors de la transformation blastique, on observe une désorganisation des noyaux dont la chromatine arbore un aspect plus immature. La chromatine est constituée d'ADN enroulé autour d'octamères d'histones (le nucléosome) et associé à des histones H1 entre les nucléosomes. Lorsque la chromatine est ouverte (euchromatine), l'ADN est accessible à la machinerie transcriptionnelle. Lorsqu'elle est condensée (hétérochromatine) l'ADN ne peut être atteint et l'expression des gènes est réprimée. On appelle «épigénétique» l'ensemble des mécanismes moléculaires qui induisent une modification de l'expression des gènes sans altérer leur séquence. Ces mécanismes reposent essentiellement sur des modifications post-traductionnelles soit des acides nucléiques (méthylation de l'ADN) soit des résidus d'histones (méthylation, acétylation, phosphorylation, ubiquitination ou sumoylation)[34]. Ces modifications ont des demi-vies différentes et une cinétique variable et ceci suggère un rôle différent en termes de régulation de la transcription des gènes. Les marques avec des demi-vies plus courtes telles que l'acétylation ou la phosphorylation, ne semblent pas être associées au maintien d'un profil

d'expression stable compte tenu de leur turn-over rapide, contrairement aux marques avec des demi-vies plus longues, telle que la méthylation par exemple, qui suggèrent des modifications des profils d'expression à plus long terme avec un rôle sur la mémoire épigénétique[35]. Ce mémoire porte spécifiquement sur la méthylation des histones. L'impact de cette modification épigénétique sur la condensation de la chromatine et donc sur l'expression génique est schématisé dans la figure 4. À l'échelle tissulaire, ce système de régulation permet de maintenir l'identité cellulaire par la conservation d'un profil d'expression génique spécifique participant à la « mémoire » cellulaire non seulement au cours des divisions cellulaires (cf. chap. I.1.4.2.1) mais aussi pendant la vie entière d'un individu et même à travers les générations.

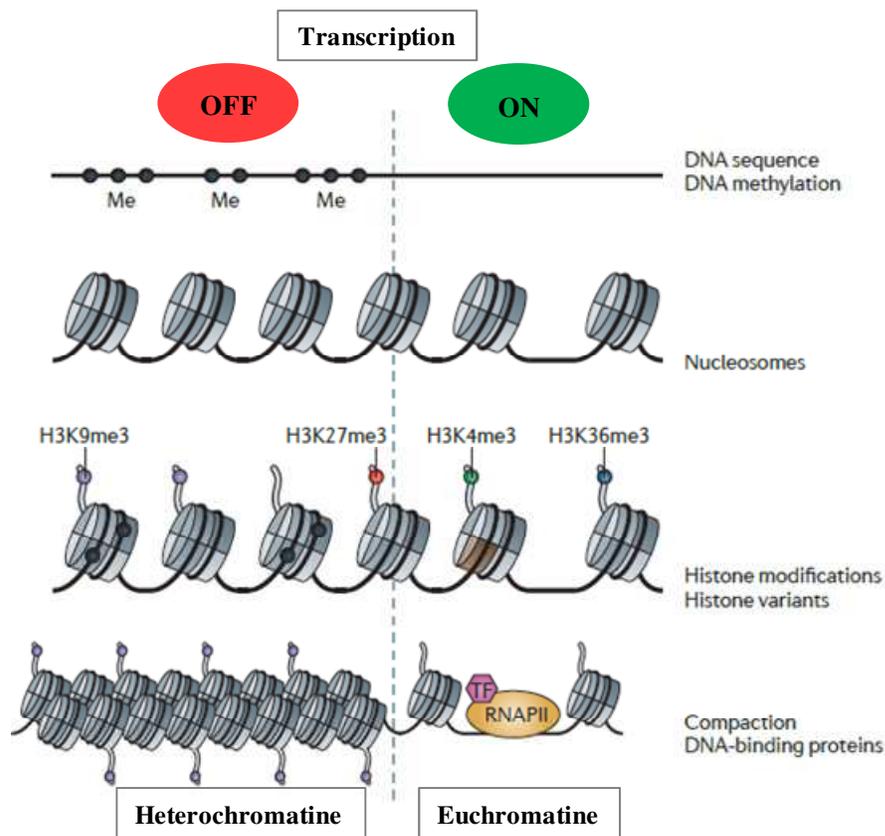


Figure 4. Principe des relations entre méthylation des histones, condensation de la chromatine et expression génique

Figure adaptée de Nat Rev Genet 12(1), Zhou, V. W., *et al.*, (2011) [36], "Charting histone modifications and the functional organization of mammalian genomes", 7-18. Copyright ©

2011, avec la permission de Nature Publishing Group. Schématisation des modifications associées à la répression ou à l'activation de la transcription, avec de haut en bas: la méthylation (« Me ») des bases de cytosine de l'ADN (marque répressive), les nucléosomes, de modifications post-traductionnelles des résidus ici de l'histone H3 en prenant l'exemple de la méthylation représentée par des billes de différentes couleurs (marques soit répressives en rouge et gris, soit activatrices en vert), et enfin la différence entre l'hétérochromatine condensée et l'euchromatine où l'ADN est accessible à l'ARN-polymérase II (RNAPII) et aux facteurs de transcription (TF).

Toutes ces modifications sont dynamiques et interagissent entre elles et avec leur environnement. Elles constituent un véritable code appelé « code des histones » où l'association de certaines marques importe autant que chacune d'elles pour déterminer le sens de la régulation transcriptionnelle dans un contexte cellulaire particulier.

1.4.2.2. Méthylation de l'ADN

La méthylation de l'ADN résulte de l'action des ADN-méthyltransférases (DNMT) qui méthylent le 5^{ème} carbone des bases cytosines le plus souvent au niveau des îlots CpG (îlots de dinucléotides cytosine-guanine) proches des régions promotrices des gènes, et est associée à une répression de la transcription. Il existe trois types de DNMT différentes: certaines sont impliquées dans la méthylation *de novo* (DNMT3A et B) alors que d'autres sont plutôt impliquées dans le maintien de la méthylation (DNMT1) [37]. Cette dernière enzyme semble particulièrement importante pour l'AR des CSH [38]. Par ailleurs, des mutations de *DNMT3A* sont présentes dans environ 20% des LAM adultes et seraient associées à un mauvais pronostic [39]. Des mutations de *DNMT3A* ont aussi été identifiées dans 8% des syndromes myélodysplasiques [40]. Ces découvertes appuient l'utilisation d'inhibiteurs des DNMT tels que l'azacitidine et la decitabine dans le traitement des LAM et des SMD. De plus, une étude récente d'inactivation conditionnelle chez la souris a démontré que DNMT1 exerce aussi son rôle dans le maintien de la méthylation de l'ADN dans les cellules leucémiques, et qu'une variation de son dosage altère la fonction des blastes, ce qui en fait une nouvelle cible thérapeutique spécifique privilégiée [41]. La méthylation de l'ADN au niveau des îlots CpG est un processus dynamique et sa distribution change durant la différenciation cellulaire [42]. Ceci est rendu possible grâce à l'action des ADN déméthylases telles que les TET (Ten-Eleven Translocation), une famille de dioxygénases nécessitant l' α -kétoglutarate et le Fe^{2+} comme co-facteurs pour leur activité enzymatique

[43]. Elles catalysent la conversion des bases cytosines méthylées (5-méthyl-cytosine ou 5mC) en 5-hydroxyméthylcytosine (5hmC), une marque épigénétique intermédiaire dans le processus de déméthylation de l'ADN. L'implication des TET dans la différenciation des CSH mais aussi dans le développement des leucémies est maintenant démontrée [44]. Plus particulièrement, des mutations de *TET2*, qui conduisent à une perte de fonction, ont été mises en évidence dans environ 15% des cancers myéloïdes de l'adulte [45] et des études de délétion chez la souris laissent suspecter un rôle de suppresseur de tumeur pour ce gène [46]. Les mutations de *TET2*, en diminuant leur activité enzymatique (protéine tronquée ou altération de leur activité catalytique), induisent une diminution du patron de 5hmC au niveau de loci importants pour la différenciation cellulaire dans les CSH/progéniteurs hématopoïétiques, notamment au niveau de la lignée érythroïde [47] résultant en une maturation anormale. Plusieurs modèles animaux étudiant diverses délétions de *Tet2* semblent converger vers un mécanisme où une diminution de l'activité de TET2 est liée à une diminution des niveaux de 5hmC, un probable phénotype d'hyperméthylation de l'ADN, un arrêt de maturation, une accumulation de progéniteurs hématopoïétiques dans la moelle osseuse et une augmentation de l'activité de repopulation des CSH *in vivo* [48-51]. On observe, dans les cellules progénitrices *Tet2*^{-/-}, l'augmentation de l'expression de régulateurs de l'auto-renouvellement tels que *Meis1* ou *Evi1* favorisant ainsi l'autorenouvellement et l'expansion des CSH et conduisent de plus à une hématopoïèse anormale *in vivo* [48]. Bien que le mécanisme exact ne soit pas parfaitement connu, ces données suggèrent que la délétion et même l'haplo-insuffisance de *TET2* résulte en une diminution de la transcription de gènes qui favorisent la différenciation ainsi que celle de gènes codant pour des régulateurs négatifs de l'auto-renouvellement. Ainsi, les mutations de *TET2* affecteraient l'hématopoïèse en conférant un avantage de croissance à certains progéniteurs, favoriseraient ainsi leur développement clonal et leur transformation maligne possiblement après l'accumulation de nouvelles mutations favorisant par exemple le développement d'une LAM comme les mutations de *FLT3* [52].

1.4.2.2.3. Modifications post-traductionnelles des résidus d’histones et méthylation des résidus d’histones

1.4.2.2.3.1. Généralités

Les modifications post-traductionnelles des histones (MPTH) surviennent au niveau des extrémités libres N-terminales des histones composant l’octamère (deux copies de H2A, H2B, H3, H4) dans le nucléosome. Alors que les MPTH sont connues depuis le début des années 70 (figure 5), c’est en 1992 que le gène *MLL* (Mixed Lineage Leukemia) localisé sur la bande chromosomique 11q23, une région fréquemment réarrangée dans les leucémies lymphoïdes et myéloïdes, a été découvert par le Dr Janet Rowley [53]. *MLL* code pour une histone méthyltransférase (HMT). Sa fonction n’était cependant pas encore connue au moment de son identification. Les fonctions des histones méthyltransférases ont été décrites à partir des années 2000. Bien qu’initialement la méthylation des histones était considérée comme irréversible, la découverte de la 1^{ère} histone déméthylase (HDM) appelée LSD1 (lysine specific demethylase 1) en 2004 [54] a prouvé le caractère dynamique de cette marque épigénétique. Enfin, l’avènement récent de technologies de séquençage de nouvelle génération et le séquençage du génome humain permettent de mettre en évidence l’implication critique des effecteurs de la méthylation des histones dans l’hématopoïèse normale et les cancers du sang.

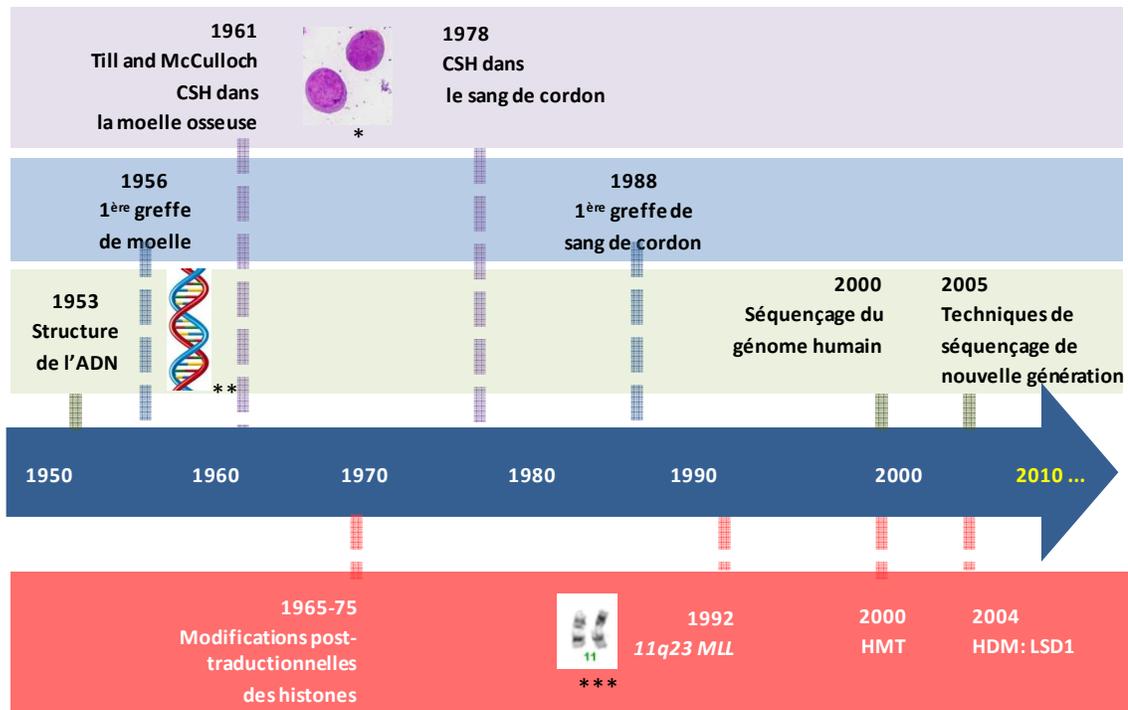


Figure 5. Découvertes épigénétiques dans la chronologie de l'hématologie

La découverte des histones déméthylases (2004) est récente comparée à la chronologie des découvertes en hématologie. Images : * cellules CD34⁺ purifiées à partir de sang d'unités de sang de cordon (figure 19B); **: schéma de la double hélice de l'ADN (découverte par Watson & Crick en 1953); *** : paire de chromosomes 11 normaux issue d'un caryotype (figure 34B). CSH: cellules souches hématopoïétiques; *MLL*: gène Mixed Lineage Leukemia; HMT: histones méthyltransférases; HDM: histones déméthylases.

1.4.2.2.3.2. Méthylation des histones et *MLL*

La méthylation post-traductionnelle de certains résidus d'histones est due à des histones méthyltransférases (HMT). Ce sont principalement des lysines (K) qui sont mono-, di- ou triméthylées (Kme, -me₂ ou -me₃) et des arginines (R) qui sont mono ou di-méthylées (Rme, -me₂) [55]. Plusieurs résidus d'une même histone peuvent être méthylés en même temps. Les résidus les plus fréquemment méthylés sont les lysines 4, 9, 27, 36 et 79 de l'histone H3, la lysine 20 de l'histone H4, les arginines 2, 8, 17, et 26 de l'histone H3 et 3 de l'histone H4. Parmi les HMT, *MLL* est une des plus étudiées en raison de son rôle important dans l'hématopoïèse normale et de son implication dans les leucémies aiguës [56]. *MLL* cible préférentiellement la lysine 4 de l'histone H3 (H3K4) et fait partie du groupe de protéines

qui composent les complexes Trithorax (Trx). Les complexes protéiques Trx sont généralement associés à l'activation de l'expression de gènes clés des CSH, tels que les gènes *HOX*, en opposition à l'effet plutôt répresseur des complexes de protéines du groupe Polycomb (PcG) appelés PRC1 et PRC2 (figure 6) [55]. L'antagonisme Trx-PcG est déterminant pour l'équilibre entre AR et différenciation dans les CSH.

Les complexes Trx contiennent les HMT de la famille MLL (MLL1 à 5) qui triméthylent H3K4 (H3K4me3, marque activatrice) au niveau de régions importantes pour l'initiation de la transcription. Cette marque permet le recrutement des HDM KDM6A / UTX et KDM6B qui elles, déméthylent H3K27. Le complexe PRC2 contient l'HMT EZH2 (« enhancer of zeste homolog 2 ») qui triméthyle H3K27 (H3K27me3, marque répressive), qui à son tour permet la liaison du complexe PRC1 capable de mono-ubiquitiner la lysine 119 de l'histone H2A (H2AK119Ub) [57].

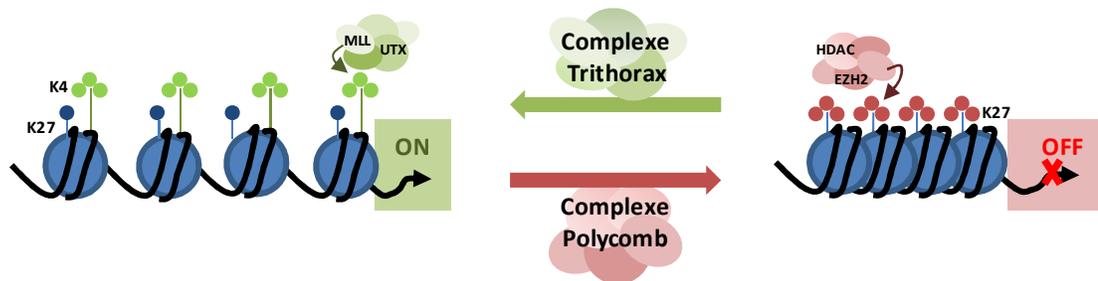


Figure 6. Antagonisme des complexes Trithorax et Polycomb dans l'expression de gènes clés des cellules souches hématopoïétiques

Représentation schématique de l'effet des complexes Trithorax et Polycomb sur la modification de la chromatine et la transcription de gènes. Le complexe Trithorax induit la méthylation de la lysine 4 de l'histone H3 (H3K4, boules vertes) notamment grâce à MLL, la déméthylation de la lysine 27 de l'histone H3 grâce à l'histone déméthylase KDM6A / UTX et son acétylation (H3K27, boules bleues) grâce à des histones acétylases pour favoriser la transcription de gènes clés (transcription « ON »). Le complexe Polycomb induit la déacétylation des résidus H3K27 grâce à une histone déacétylase HDAC et sa méthylation, notamment par la méthyltransférase EZH2 (boules rouges) pour favoriser ainsi la répression de la transcription (transcription « OFF »).

En plus de la méthylation des histones, les complexes Trx et PcG sont capables de recruter directement ou indirectement différentes protéines qui catalysent d'autres MPTH, comme par

exemple l'acétylation de H3K27 (H3K27Ac, marque activatrice) grâce aux histones acétyltransférases (HAT) recrutées par le complexe Trx, ou les DNMT recrutées par le complexe PcG (ADN-me, marque répressive).

Par ailleurs, au-delà de leurs effets antagonistes sur la transcription, les complexes Trx et PcG peuvent aussi coexister en un même locus notamment durant le développement de l'hématopoïèse, de façon à obtenir une régulation plus fine en collaborant avec d'autres modificateurs de la chromatine. Des anomalies d'expression ou des mutations de leurs composantes, diminuant l'effet de Trx en favorisant l'effet de PcG par exemple, sont de plus en plus impliquées dans le développement des néoplasies [58].

1.4.2.2.3.3. *Déméthylation des histones*

Comme nous l'avons vu plus haut, la découverte des histones déméthylases (HDM) est relativement récente (2004) et a considérablement modifié l'idée initiale d'une méthylation permanente des histones pour révéler son caractère dynamique. Il existe trois familles d'HDM dont les similitudes et les différences seront extensivement exposées dans la section 2 de ce chapitre I.

1.4.3. Remarques conclusives sur la régulation de l'hématopoïèse normale

En conclusion, c'est un équilibre dynamique et finement régulé entre stabilité et réversibilité de l'expression de différents programmes transcriptionnels qui permet la détermination de l'identité des CSH et de leur destin, entre AR et différenciation. La régulation de cet équilibre repose sur un réseau complexe de processus cellulaires, moléculaires et épigénétiques qui sont de mieux en mieux élucidés. Les dysfonctions de plusieurs effecteurs de ces mécanismes régulateurs ont été associées, voire même causalement impliquées dans les hémopathies malignes.

1.5. Conservation de l'hématopoïèse de la souris à l'humain et implications

L'hématopoïèse est généralement bien conservée chez les vertébrés. Les modèles animaux tels que la souris par exemple sont dès lors un outil accessible et très utile pour étudier fonctionnellement les mécanismes de développement du tissu sanguin et leur régulation. Ainsi, la majeure partie de notre connaissance de l'hématopoïèse humaine normale et

anormale, au-delà de ce qui est observable sur des prélèvements sanguins ou de moelle osseuse provenant de patients, repose sur des études fonctionnelles utilisant des modèles murins et l'extrapolation théorique de ces résultats chez l'humain. Cependant, en plus des évidentes variations de taille et de durée de vie entre homme et souris, il existe de nombreuses différences entre hématopoïèse humaine et murine, par exemple dans la composition du sang périphérique, dans l'organisation hiérarchique et les marqueurs de surface des cellules hématopoïétiques, dans l'implication de certains FT ou de certains modificateurs de la chromatine chez la souris et leur absence chez l'homme, etc. Ces différences rendent indispensables la confirmation des études réalisées chez la souris dans des cellules humaines primaires autant que possible. Ceci est d'autant plus important qu'un des buts principaux de la recherche en hématologie vise à mieux comprendre la pathogénèse des hémopathies afin de développer de nouveaux outils thérapeutiques [6].

2. LES HISTONES DÉMÉTHYLASES (HDM)

2.1. Classification des histones déméthylases

La découverte d'une nouvelle classe d'enzymes capables de déméthyliser les résidus lysine et arginine d'histones et appelées histones déméthylases (HDM) a démontré que la méthylation des histones, bien qu'impliquée dans l'instauration de profils d'expression particuliers stables à long terme, est en fait un processus dynamique et réversible [54]. Cette découverte a aussi attiré l'attention sur l'importance de ce mécanisme épigénétique dans la régulation de l'hématopoïèse [55]. Il est désormais admis que, selon les résidus d'histone concernés, certaines de ces marques de méthylation sont associées à l'activation de la transcription de gènes cibles (ex: H3K4, H3K36) et donc à une configuration de la chromatine ouverte (euchromatine) [59, 60] alors que d'autres sont plutôt associées à la répression de la transcription (ex: H3K27, H3K9) et donc à une configuration de la chromatine plus fermée (hétérochromatine)[61] (figure 4). Les HDM comptent actuellement 40 membres et sont réparties en trois familles définies par le type de réaction de déméthylation utilisé [55, 62]. On distingue la famille des LSD ou « lysine specific demethylases », la famille JmjC ou famille des déméthylases Jumonji C et la famille PADI ou « peptidyl arginine deiminase ».

Le tableau 1 les répertorie selon leur famille et les regroupe par sous-familles d'enzymes qui contiennent des domaines protéiques similaires.

Famille	Symbole	Synonyme	Structure	Cible	Localisation	Maladie	Anomalie génétique
LSD	KDM1A	LSD1		H3K4me ^{2/1}	1p36.12	LAM ¹	surexpression
	KDM1B	LSD2		H3K4me ^{2/1}	6p22.3		
JmjC	KDM2A	JHDM1A/FBXL11		H3K36me ^{2/1}	11q13.2		
	KDM2B	JHDM1B/FBXL10		H3K36me ^{2/1} , H3K4me ³	12q24.31	LAM ²	surexpression
	KDM3A	JMJD1A/JHDM2A/TSGA		H3K9me ^{2/1}	2p11.2		
	KDM3B	JMJD1B		H3K9me ³	5q31	LAM ³ , SMD ⁴	délétion, mutation ponctuelle
	KDM3C	JMJD1C		Inconnue	10q21.3		
	KDM4A	JMJD2A/JHDM3A		H3K9me ^{3/2} , H3K36me ^{3/2}	1p34.1		
	KDM4B	JMJD2B/JHDM3B		H3K9me ^{3/2} , H3K36me ^{3/2}	19p13.3		
	KDM4C	JMJD2C/JHDM3C/GASC1		H3K9me ^{3/2} , H3K36me ^{3/2}	9p24.1	Lymphome ⁵ , LAM ⁶	translocation, surexpression
	KDM4D	JMJD2D/JHDM3D		H3K9me ^{3/2}	11q21		
	KDM4E (DL)	JMJD2E		H3K9me ^{3/2/1}	11q21		
	KDM5A	JARID1A		H3K4me ^{3/2}	12p11	LAM-M7 pédiatrique ⁷ , Fusion avec NUP98, SMD ⁸	dél12p
	KDM5B	JARID1B		H3K4me ^{3/2/1}	1q32.1	LMC transformée ⁹ , SMD ¹⁰	dél1q32
	KDM5C	JARID1C		H3K4me ^{3/2}	Xp11.22- p11.21		
	KDM5D	JARID1D		H3K4me ^{3/2}	Yq11		
	KDM6A	UTX		H3K27me ^{3/2}	Xp11.2	Myélome ¹¹ , LAM dérivée d'une CMML ¹² ,	mutations
	UTY			Inconnue	Yq11		
	KDM6B	JMJD3		H3K27me ^{3/2}	17p13.1	Cancers hématologiques ^{13,14}	mutations, surexpression, translocation
	KDM7A	JHDM1D/KIAA1718		H3K9me ² , H3K27me ² , H4K20me ¹	7q34		
	KDM7B	JHDM1F/PHF8		H3K9me ^{2/1} , H4K20me ¹	Xp11.22		
	KDM7C	JHDM1E/PHF2		H3K9me ² , H3K27me ²	9q22.31		
	KDM8	JMJD5		H3K36me ²	16p12.1		
	C14orf169	NO66		H3K4me ³ , H3K36me ²	14q24.3		
	HIF1AN			HIF1a-N803	10q24		
HR	HAIRLESS		Inconnue	8p21.2			
HSPBAP1			Inconnue	3q21.1			
JARID2			Inconnue	6p24-p23			
JHDM1C	FBXL19		Inconnue	16p11.2			
JMJD4			Inconnue	1q42.13			
JMJD6	PTDSR/PSR		H3R2me ² , H4R3me ²	17q25			
JMJD7			Inconnue	15q15.1			
JMJD8			Inconnue	16p13.3			
MINA			H3K9me ³	3q11.2			
PLA2G4B			Inconnue	15q11.2- q21.3			
PADI	PADI1			Inconnue	1p36.13		
	PADI2			H3R26	1p36.13	Cancers solides ¹⁵	surexpression
	PADI3			Inconnue	1p36.13		
	PADI4			H3(R2 R8 R17 R26) H4R3	1p36.13	Cancers solides et hématologiques ¹⁶	surexpression
	PADI6			Inconnue	1p36.13		

Domaines protéiques:

SWIRM Pyr-redox CW-ZF FAD-binding3 JmjC JmjN PHD FBOX LRR Tudor ARID TPR PAD PAD N PAD M

Tableau 1. Classification des histones déméthylases (HDM)

Les HDM sont regroupées par famille en fonction du type de réaction de déméthylation. Leur structure protéique, leur(s) résidu(s) d'histone cible(s), ainsi que leur localisation chromosomique sont également présentés. Certaines d'entre elles ont été identifiées dans des hémopathies malignes: les références correspondantes sont notées ci-dessous.

Références: (1) Lokken et al., *Cancer Cell* 2012[63] (2) He et al., *Blood* 2011[64] (3) Lai et al., *Genomics* 2000 [65] (4) Hu et al., *Oncogene* 2001[66] (5) Black et al., *Cell* 2013[67] (6) Vinatzer et al., *Clin Cancer Res* 2008[68] (7) van Zutven et al., *Genes Chromosomes Cancer* 2006[69] (8) Streubel et al., *Br J Haematol* 1998[70] (9) Karrman et al., *Bone Marrow Transplant* 2007[71] (10) Bacher et al., *Cancer Genet Cytogenet* 2009[72] (11) van Haften et al., *Nat Genet* 2009[73] (12) Jankowska et al., *Blood* 2011[74] (13) Anderton et al., *Oncogene* 2011[75] (14) Agger et al., *Genes Dev* 2009[76] (15) McElwee et al. *BMC Cancer* 2012[77] (16) Nakashima, et al. *J Biol Chem* (1999)[78].

LSD: lysine specific demethylase; JmjC: famille Jumonji C; PADI: peptidyl arginine déiminase; LAM: leucémie aiguë myéloïde; SMD: syndrome myélodysplasique; LMC: leucémie myéloïde chronique; CMML: leucémie myélomonocytaire chronique; SWIRM: domaine Swi3p, Rsc8p et Moira; Pyr-redox: domaine pyridine nucléotide–disulfide redox; CW-ZF: domaine en doigt de zinc CW; FAD-binding 3: domaine de liaison « flavin adénine dinucléotide » de type 3 ; JmjC/N: Domaine Jumonji N et C terminaux; PHD: Domaine « Plant Homeo domain »; FBOX: domaine « F-box »; LRR: domaine « Leucine Rich Repeat »; tudor: domaine tudor; ARID: domaine « AT-Rich Interaction »; TPR: domaine « tetratricopeptid repeat »; PAD: domaine protéine arginine déiminase; PAD N ou M: domaine PAD N-terminal ou central.

Figurent également dans ce tableau, les cibles d'histones connues de ces enzymes, la localisation chromosomique du gène correspondant ainsi que les anomalies associées à une hémopathie maligne rapportée dans la littérature. Ces familles d'enzymes sont très bien conservées dans l'évolution. Les substrats connus de ces enzymes qui ne sont pas des résidus d'histones, ne sont pas présentés dans ce tableau.

2.1.1. Réactions de déméthylation des histones

Il existe trois types de réactions par lesquelles les groupements méthyliques peuvent être enlevés des lysines ou des arginines des histones. Ces réactions définissent les trois familles des HDM: l'oxydation pour la famille LSD, l'hydroxylation pour la famille Jumonji C et la déimination pour la famille PADI (figure 7).

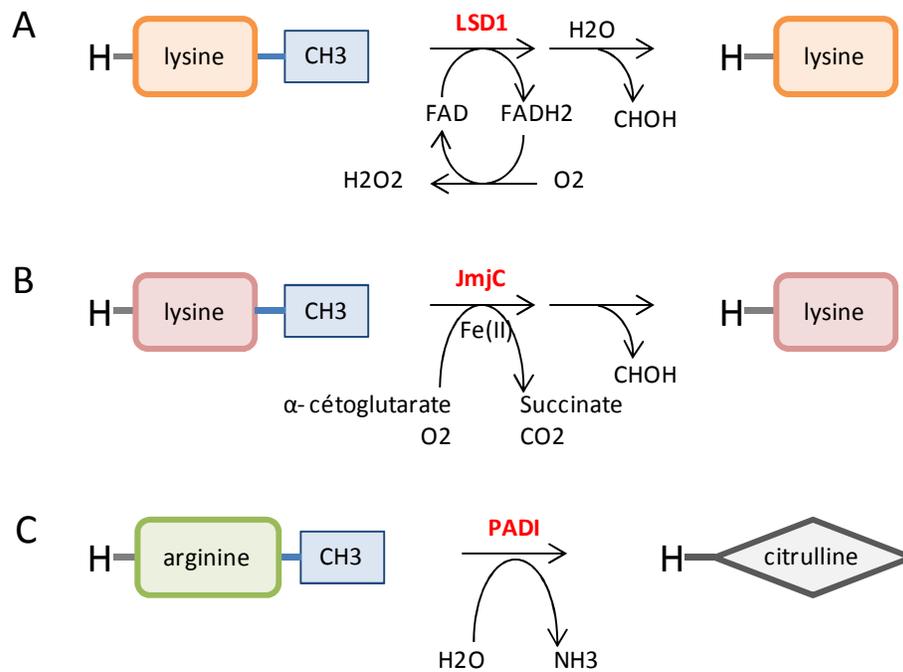


Figure 7. Mécanismes catalytiques de déméthylation

(A) Famille LSD: déméthylation des résidus lysine des histones (H) par amine oxydation dépendant du cofacteur FAD (flavine adénine dinucléotide). (B) Famille Jumonji C (JmjC): déméthylation des résidus lysine des histones (H) par hydroxylation dépendant de deux cofacteurs: l' α -cétoglutarate (appelé aussi 2-oxoglutarate) et le fer. (C) Famille PADI: déméthylation des résidus arginine des histones (H) par citrullination.

2.1.2. Famille LSD (Lysine Specific Demethylase)

La famille des LSD compte deux membres: KDM1A/LSD1 et KDM1B/LSD2. KDM1A est la première HDM à avoir été découverte [54]. Ces enzymes ne sont capables de démétyler que les lysines H3K4 mono- ou di-méthylées par oxydation. Elles ont besoin du cofacteur FAD (flavine adénine dinucléotide) pour leur réaction d'oxydation (figure 7A). De plus, KDM1A se lie fréquemment à des complexes protéiques tels que CoREST (« corepressor to REST ») et exerce une action répressive sur la transcription de gènes cibles [54]. La délétion de *Kdm1a* chez la souris provoque la mort embryonnaire [79] et *Kdm1b* est nécessaire pour l'instauration de l'empreinte génomique maternelle lors de l'ovogenèse[80]. Par ailleurs, la diminution de KDM1A provoque l'augmentation de la méthylation de H3K4 ce qui entraîne la diminution de la répression de la transcription et l'altération de la différenciation [81].

2.1.3. Famille JmjC (déméthylases Jumonji C)

La famille JmjC compte actuellement 33 membres dont la nomenclature a beaucoup évolué ces dernières années [82]. Le premier membre de cette famille à avoir été découvert peu après KDM1A/LSD1 est KDM2A [83]. Le nom *Jumonji* provient du japonais et signifie « en croix » et fait référence au phénotype morphologique des souris mutantes *Jmj^{-/-}* (gène actuellement appelé *Jarid2*) dans lesquelles une anomalie de développement forme une croix entre la plaque et la crête neurale [84]. Grâce à leur domaine catalytique commun appelé domaine JMJC, ces enzymes déméthylent différents résidus d'histones (principalement des lysines) tri-, di-, ou mono-méthylés par une réaction d'hydroxylation qui dépend de deux cofacteurs: l' α -cétoglutarate (appelé aussi 2-oxoglutarate ou 2-OG) et le fer(II) (figure 7B) [55]. Ces différentes enzymes peuvent être classées en sous-familles de structure similaires en fonction des domaines protéiques qu'elles contiennent et qui les rendent spécifiques à un ou plusieurs substrats (tableau 1). Il est important de souligner que les substrats de ces enzymes ne sont pas encore tous connus. Une autre caractéristique commune de ces enzymes est que leur action peut avoir un rôle activateur ou répresseur sur la transcription des gènes cibles en fonction des complexes protéiques au sein desquels ils agissent. En effet comme KDM1A, les HDM JmjC agissent fréquemment au sein de complexes protéiques rassemblant des facteurs de transcription, d'autres modificateurs de la chromatine (HMT, HDAC, ...) mais aussi des récepteurs nucléaires ou des cyclines, comme nous l'avons vu précédemment (complexes Trx et PcG, cf. Chap. I. 1.4.2.2.3.2) [55, 57, 58]. Au sein du complexe protéique MLL du groupe Trx, la méthylation de H3K4 par MLL qui s'associe à la déméthylation de H3K27 par KDM6A / UTX afin de favoriser l'activation de gènes *HOX* notamment est un exemple d'action combinée de différents effecteurs épigénétiques en vue de réguler la transcription de gènes cibles (figure 6) [55].

L'implication des HDM JmjC dans le destin cellulaire est de plus en plus reconnue. Depuis la démonstration du rôle important de JARID2 dans le développement neural [84], d'autres enzymes de cette famille ont montré, par leur action régulatrice de l'expression de différents programmes de transcription, leur rôle important dans le développement et les décisions concernant le destin cellulaire, et notamment entre AR et différenciation. Plusieurs auteurs ont revu les preuves de leur implication dans les processus normaux et tumoraux [55, 62,

85].

Enfin il faut noter que cette famille d'HDM JmjC est particulièrement dépendante du métabolisme du cofacteur 2-oxoglutarate (2-OG appelé aussi α -cétoglutarate), produit à partir de l'isocitrate dans le cycle de Krebs par les enzymes isocitrate déshydrogénases (IDH). Toute perturbation de ce métabolite et toute dysfonction des IDH peuvent donc affecter l'activité des HDM JmjC.

2.1.3.1. Sous-famille KDM5/JARID

La sous-famille KDM5/JARID comprend 4 membres (A-D) qui ciblent le même résidu H3K4 –tri-, -di- ou mono-méthylé, marques associées à l'activation de la transcription [86]. Plusieurs études ont souligné l'implication de différents membres de cette sous-famille dans le destin cellulaire. Par exemple KDM5A/JARID1A/RBP2 est impliquée dans le contrôle de la différenciation des cellules souches embryonnaires (CSE) [87] et cible particulièrement les résidus H3K4 méthylés de la région des gènes *HOX*. Pendant la différenciation, KDM5A n'est plus retrouvée à proximité de ces régions, les résidus H3K4 sont alors de nouveau méthylés et favorisent la transcription des gènes *HOX* [88]. Il faut noter que les gènes *HOX* sont associés à la différenciation dans les CSE. KDM5B/JARID1B a également un rôle dans la répression de gènes impliqués dans la pluripotence des cellules embryonnaires lors du développement neuronal [89]. Il a récemment été démontré que l'inhibition de *Kdm5b* par interférence avec l'ARN dans les CSH de souris augmente l'expression de différents gènes clés pour les CSH tels que les gènes *Hoxa7*, *Hoxa9*, *Hoxa10*, *Hes1* et *GATA1*, favorisant la prolifération des CSH (diminution de la répression de la transcription de ces gènes par *Kdm5b*). Cette étude prouve le rôle de régulateur négatif de *Kdm5b* pour les CSH de souris [90].

2.1.3.2. Sous-famille KDM4/JMJD2

La sous-famille KDM4 compte 5 membres (A-E). KDM4A, B et C possèdent deux domaines PHD et deux domaines Tudor qui permettent la liaison aux histones méthylées. Ces trois membres ciblent H3K9me^{3/2} (marque répressive) et H3K36me^{3/2} (marque activatrice)[91]. KDM4D, qui est le plus petit et ne possède ni le domaine Tudor ni le domaine PHD, déméthyle H3K9me^{3/2}. KDM4E (DL pour D-like) ressemble à KDM4D mais sans domaine

Tudor ni PHD. Les membres de cette famille commencent seulement à être mieux connus. Il est intéressant de mentionner qu'une de leur cible commune, H3K9me3, est une marque importante pour la sénescence des cellules [55]. De plus, la détection de la surexpression des gènes KDM4A-C dans différentes tumeurs solides laisse supposer que ces protéines puissent se comporter comme des oncogènes [92], justifiant le développement d'inhibiteurs spécifiques de ces protéines.

2.1.3.3. Sous-famille KDM7/JHDM1

La sous-famille KDM7 regroupe 3 membres et était précédemment appelée le groupe PHF en raison de leur domaine « PHD finger » qui leur permet de se lier à la marque H3K4me3. Pour KDM7A/KIA1718 qui cible H3K9me2 et H3K27me2, la liaison avec H3K4me3 diminue l'activité de déméthylation de H3K9me2 mais pas celle de H3K27me2 [93]. KDM7B/PHF8, déméthyle H4K20me1 et H3K9me2/1, et comme KDM7A, reconnaît aussi H3K4me3. Toutefois, pour cette déméthylase, cet ancrage favorise son action de déméthylation de H3K9me2 [93]. KDM7C/PHF2 cible H3K9me2 et H3K27me2, et reconnaît aussi H3K4me3. L'impact de cette liaison sur la déméthylation des cibles n'est pas connu. Des mutations des 3 gènes codant pour ces enzymes sont associées à des anomalies de développement [55]. Par ailleurs, l'augmentation de KDM7A/JHDM1D dans des conditions d'hypoxie et de carence nutritionnelle diminue la croissance tumorale en réprimant l'angiogenèse [94]. KDM7B/PHF8 régule le développement neural, et des mutations dans cette enzyme causent des retards mentaux liés au chromosome X [94]. Enfin, un rôle de régulateur positif des CSH a été démontré chez la souris pour Kdm7b/Phf8/Jhdm1f [90].

2.1.4. Famille PADI (Peptidyl Arginine Déiminase)

Cette famille d'HDM est la moins bien connue et compte actuellement cinq membres (tableau 1). Ces enzymes ont environ 50% de similitude de séquence dont leur domaine catalytique et leurs résidus de liaison au calcium [95, 96]. Le groupement méthyl des arginines de l'histone H3 est dans ce cas converti en citrulline (citrullination ou déimination) (figure 7C). Les membres de cette famille sont a priori exprimés dans différents tissus: PADI1 au niveau de l'épiderme (cible = kératine et filagrine), PADI2 au niveau des muscles

et du cerveau principalement (cible = myéline et histone H3), PADI3 au niveau des follicules pileux (cible = trichohyaline), PADI4 au niveau des granulocytes (cible = histones H2A, H3, H4 et nucléophosphamine), et finalement PADI6 au niveau des cellules souches embryonnaires et des ovocytes (cible inconnue). PADI4 est certainement le membre de cette famille le plus étudié. Jusqu'à récemment, PADI4, dont la localisation nucléaire est favorisée par son domaine NLS (domaine signal de localisation nucléaire), semblait être le seul à avoir un rôle dans la déméthylation des histones en ciblant différents résidus arginines de H2A, H3 et H4. Il agit comme co-répresseur transcriptionnel recruté au niveau des promoteurs de gènes cibles régulés par des FT tel que le récepteur à l'œstrogène, le récepteur thyroïdien et p53 [95, 97]. Le rôle de PADI4 dans l'hématopoïèse n'est pas encore bien connu. Chez la souris, *Padi4* a été récemment décrit comme étant surexprimé dans des cellules très immatures de type LSK ou Lineage⁻ Sca-1⁺ c-Kit⁺. Cette dernière étude suggère que *Padi4* régule la prolifération des progéniteurs multipotents en contrôlant l'expression de *c-myc*, et que sa diminution favorise la prolifération aberrante des progéniteurs pluripotents [98]. Un autre membre de cette famille, PADI2, semble avoir également un rôle dans la déméthylation du résidu H3R26 en interagissant avec le récepteur à l'œstrogène ER α [99]. De plus, une étude récente a montré que *PADI2* était surexprimé dans des cellules tumorales exprimant HER2/ERBB2+ de cancer du sein [77].

2.2. Profil d'expression des HDM dans l'hématopoïèse de la souris

Les HDM font présentement l'objet de nombreuses études et leur rôle de régulateurs épigénétiques essentiels dans le développement et dans la détermination du destin cellulaire commence à être étudié, parallèlement à la démonstration de l'implication des histones déméthylases dans les hémopathies malignes. L'ensemble de ces éléments laissent suspecter un rôle important de certaines de ces HDM dans les cellules hématopoïétiques. Une étude récente d'expression des HDM de la famille JmjC dans les CSH de souris en comparaison à une population de cellules plus différenciées a démontré que la majorité des HDM exceptée *Hairless* sont bien exprimées, et que plusieurs d'entre elles (*Kdm5b/Jarid1b*, *Kdm4d/Jmjd2d*, *Kdm2b/Fbxl10* et *Jmjd1c*) sont plus exprimées dans le compartiment cellulaire immature que le compartiment cellulaire différencié, et inversement pour les gènes *Kdm7a/Jhdm1d*, *Hspbap1* et *Hairless* (figure 8).

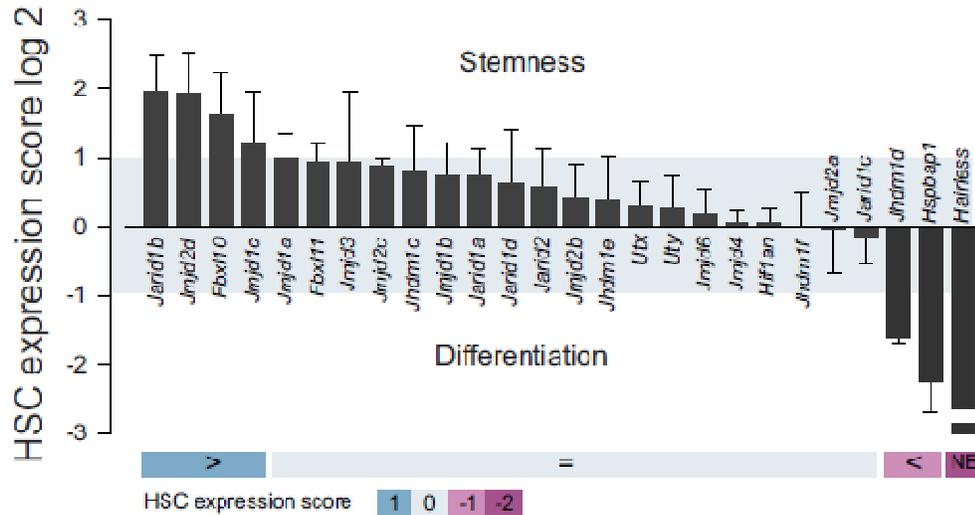


Figure 8. Profil d'expression différentielle chez la souris des HDM de la sous-famille JmjC dans les CSH et les progéniteurs immatures par rapport aux cellules différenciées (Figure adaptée de Blood 122 (9), Cellot S. et al, (2013)[90], "RNAi screen identifies Jarid1b as a major regulator of mouse HSC activity.", 1545-1555. Copyright © 2013, avec la permission de American Society of Hematology; High Wire Press. Les quantités relatives des transcrits (RQ) sont présentées en échelle de log2 et représentent le ratio ΔCT (CSH)/ ΔCT (moelle osseuse) déterminé par qRT-PCR (moyenne \pm SEM, n=3). Un score d'expression a été calculé basé sur le niveau d'expression (ΔCT) et sur l'expression différentielle (RQ) des différentes HDM pour classer les gènes cibles selon leur expression relative entre CSH et cellules différenciées: gènes non exprimés dans les CSH: -2; moins exprimés dans les CSH vs les cellules différenciées: -1, plus exprimés dans les CSH vs les cellules différenciées: +1; exprimés de façon équivalente: 0. HDM: histones déméthylases; CSH: cellules souches hématopoïétiques.

Cette augmentation d'expression de différents gènes dans les CSH, laisse suspecter un rôle biologique de ces enzymes dans l'identité des CSH. Des études fonctionnelles d'inhibition grâce à la méthode d'interférence de l'ARN (RNAi) utilisant des petits ARN interférents en épingle à cheveux (shRNA) dirigés contre les gènes cibles ont permis de démontrer que *Kdm5b/Jarid1b* est un régulateur négatif des CSH de souris et que son inhibition permet leur expansion, alors que *Kdm7a/Jhd1d* est plutôt un régulateur positif de la reconstitution hématopoïétique [90]. Les HDM sont bien conservées entre les souris et les hommes. Nous suspectons dès lors que les HDM sont également différentiellement exprimées dans les cellules humaines immatures par rapport aux cellules différenciées. Notre étude vise à

démontrer ces différences d'expression afin de cibler par la suite les candidats les plus significatifs pour réaliser des études fonctionnelles.

3. HÉMOPATHIES MYÉLOÏDES MALIGNES DE L'ENFANT: LEUCÉMIE AIGÜE ET SYNDROME MYÉLODYSPLASIQUE

3.1. Les leucémies aiguës

3.1.1. Définition et classification des leucémies aiguës

Les leucémies aiguës sont des maladies hétérogènes caractérisées par la prolifération clonale dérégulée d'une population de cellules immatures anormales appelées blastes. Ces cellules présentent une différenciation anormale et un avantage de survie sur les cellules normales. C'est le cancer le plus fréquent chez l'enfant et représente un peu moins d'un tiers des cancers diagnostiqués chez ces patients. On distingue habituellement les leucémies aiguës lymphoblastiques (LAL) des leucémies aiguës myéloïdes (LAM) selon que le blocage de prolifération survient dans la lignée lymphoïde ou myéloïde. Ces dernières représentent 20% des leucémies de l'enfant et leur pronostic en terme de survie globale à 5 ans est beaucoup moins bon: 92% pour les LAL [100] comparativement à 60 à 70% pour les LAM. Après une première rechute de LAM, la survie globale des patients diminue à 38% [101].

3.1.2. Classification des leucémies aiguës de l'enfant

La classification des leucémies de l'enfant a évolué parallèlement à l'amélioration des techniques d'analyses phénotypiques, cytogénétiques et moléculaires. Actuellement la classification de l'OMS-2008 (« WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, 2008 ») [102] intègre ces données et distingue les LAL des LAM.

3.1.3. Leucémies aiguës lymphoïdes (LAL) de l'enfant

La classification de l'OMS-2008 prend en considération les anomalies génétiques associées à des variations importantes du pronostic.

3.1.3.1. LAL de type B

Ces LAL présentent une prolifération de progéniteurs anormaux lymphoïdes déjà engagés dans la lignée B. On distingue les groupes suivants:

3.1.3.1.1. LAL de type B associées à des anomalies génétiques récurrentes

On distingue:

- LAL avec t(9;22)(q34;q11.2) et fusion *BCR-ABL1*
- LAL avec réarrangements du gène *MLL* (11q23) dont principalement les t(4;11)(q21;q23) et t(11;19)(q23;p13.3)
- LAL avec t(12;21)(p13;q22) et fusion *ETV6-RUNX1*
- LAL avec hyperdiploïdie, (n=51 à 57 chromosomes)
- LAL avec hypodiploïdie (n<45 chromosomes)
- LAL avec t(5;14)(q31;q32) avec fusion *IL3-IGH* et surexpression d'*IL3*
- LAL avec t(1;19)(q23;p13.3) avec fusion *TCF3-PBX1*

3.1.3.1.2. LAL de type B sans anomalies génétiques récurrentes

Ces LAL de type B sont dès lors classifiées selon leur degré de différenciation grâce aux analyses d'immunophénotypage entre LAL proB, B communes ou préB.

3.1.3.2. LAL de type T

Ce groupe est identifié par l'immunophénotypage qui signe la prolifération de cellules de type T avec un arrêt de maturation à différents stades (proT, préT, corticale et médullaire).

On retrouve dans ces LAL des anomalies récurrentes impliquant:

- les gènes des récepteurs des cellules T ou *TCR* (locus α et δ en 14q11.2, locus β en 7q35 ou locus γ en 7p14-15) avec des partenaires de fusion qui codent le plus souvent pour des facteurs de transcription et qui seront surexprimés (ex: *TLX1*, *MYC*, *TLX3*, *TALI*, *LMO2*, *LMO1*, *LYL1*).
- d'autres gènes de fusion tels que *MLL-ENL* dans la t(11;19)(q23;p13.3) ou *CALM-AF10/PICALM-MLLT10* dans la t(10;11)(p13;q14).
- les délétions qui conduisent à la perte de gènes suppresseurs de tumeurs (ex. del(9p) et perte de *CDNK2A*).

- des mutations activatrices de gènes tels que *NOTCH1*.

3.1.4. Leucémies aiguës myéloïdes (LAM) de l'enfant

La classification OMS-2008 regroupe les LAM en quatre groupes:

3.1.4.1. LAM avec anomalies génétiques récurrentes

Cette entité regroupe les LAM qui présentent une anomalie génétique récurrente avec une signification pronostique spécifique. On distingue

- LAM avec t(8;21)(q22;q22) et fusion *RUNX1-RUNX1T1*
- LAM avec inv(16)(p13.1q22) ou t(16;16)(p13.1;q22) et fusion *CBFB-MYH11*
- LA promyélocytaire avec t(15;17)(q24;q21) et fusion *PML-RARA*
- LAM avec t(9;11)(p22;q23) et fusion *MLL3-MLL*
- LAM avec t(6;9)(p23;q34) et fusion *DEK-NUP214*
- LAM avec inv(3)(q21q26.2) ou t(3;3)(q21;q26.2) avec fusion *RPN1-EVII*
- LAM mégacaryoblastique avec t(1;22)(p13;q13) et fusion *RBM15-MKLI*

Les LAM avec mutations de *NPM1* ou avec mutations de *CEBPA*, compte tenu de leur fréquence dans les LAM avec caryotype normal et de leur bon pronostic en l'absence de mutation de type *FLT3-ITD*, constituent deux entités supplémentaires provisoires.

Enfin, sont regroupées également ici les autres LAM avec mutations telles que *FLT3*, *KIT*, *WT1*, *KRAS*, *NRAS* et les duplications partielles de *MLL* en tandem.

3.1.4.2. LAM avec changements myélodysplasiques

Ces LAM présentent une blastose supérieure ou égale à 20%, avec une histoire de SMD ou des anomalies cytogénétiques de SMD ou avec une myélodysplasie multilignée, sans histoire de traitement cytotoxique antérieur ou sans anomalie génétique récurrente tel que décrit précédemment.

3.1.4.3. Cancers myéloïdes secondaires aux traitements (LAM-t)

Ces hémopathies peuvent être de type LAM ou SMD et font suite à un traitement cytotoxique de chimiothérapie et /ou de radiothérapie.

3.1.4.4. LAM « autres »

Le dernier groupe de LAM rassemble toutes les LAM qui n'ont pas pu être classées dans les groupes précédents et sont différenciées entre elles par leur morphologie selon la classification classique FAB (Franco-Américano-Britannique). Elle différencie les LAM suivantes:

- FAB-M0: LAM indifférenciée,
- FAB-M1: LAM sans maturation,
- FAB-M2: LAM avec maturation,
- FAB-M3: LAM promyélocytaire,
- FAB-M4: LAM avec différenciation myélomonocytaire (et FAB-M4eo: LA myélomonocytaire avec éosinophiles anormaux)
- FAB-M5: LAM monocytaire
- FAB-M6: LAM avec différenciation érythroblastique,
- FAB-M7: LAM avec différenciation mégacaryoblastique

D'autres entités rares font aussi partie de cette catégorie mais ne seront pas détaillées ici.

3.2. Les syndromes myélodysplasiques

Les syndromes myélodysplasiques (SMD) sont des cancers myéloïdes hétérogènes caractérisés par une maturation anormale des progéniteurs myéloïdes accompagnée de dysplasie morphologique et de cytopénies. Les SMD et les leucémies aiguës myéloïdes (LAM) associées à une myélodysplasie représentent moins de 5% des cancers de l'enfant et jusqu'à 10% des hémopathies malignes myéloïdes pédiatriques [103]. Contrairement aux LAM, il n'y a pas initialement de prolifération blastique dans la moelle. Certains patients peuvent voir leur SMD évoluer en LAM. Même si les SMD et les LAM associées à une myélodysplasie sont principalement sporadiques, certains d'entre eux peuvent aussi être associés à une des formes héréditaires d'aplasie médullaire telles que l'anémie de Fanconi, la dyskératose congénitale, le syndrome Shwachman-Diamond ou la neutropénie congénitale sévère [104].

Les SMD de l'enfant constituent une entité différente des SMD de l'adulte dans la classification OMS-2008[102]. On distingue:

- les cytopénies réfractaires de l'enfant avec < 2% de blastes dans le sang périphérique (SP) et < 5% dans la moelle osseuse (MO) associés à la présence d'une dysplasie,
- les anémies réfractaires avec excès de blastes c'est-à-dire de 2 à 19% dans le SP ou 5 à 19% dans la MO (AREB). On subdivise cette catégorie en 2 groupes, les AREB-1 avec 5 - 9% de blastes dans la MO et les AREB-2 avec 10 - 19% dans la MO. La signification pronostique de ces groupes chez l'enfant n'est pas encore déterminée.

Par ailleurs, certains cas diagnostiqués comme des LAM avec 20 à 29% de blastes dans le SP et/ou dans la MO peuvent présenter en plus une myélodysplasie et/ou des anomalies cytogénétiques associées aux SMD (cf. Chap. I. 3.1.4.2). De plus, les cas de néoplasies myéloïdes qui présentent une des anomalies cytogénétiques suivantes: t(8;21)(q22;q22), inv(16)(p13.1;q22) ou t(16;16)(p13.1;q22) et t(15;17)(q24;q21) sont considérées comme des LAM, indépendamment du pourcentage de blastes.

D'autres types de SMD et de LAM associées à une myélodysplasie font partie d'une nouvelle entité clinique appelée syndrome de prédisposition familiale aux SMD/LAM (ou SPF-SMD/LAM). Cette entité est définie par la présence concomitante d'un SMD ou d'une LAM chez au moins deux membres d'une même famille associée principalement à des mutations au niveau de gènes qui codent pour des facteurs de transcription essentiels aux CSH tels que *RUNX1* (également nommé *AML1*) et *GATA2*, ou à la différenciation myéloïde tel que *CEBPA*. Des études approfondies de différentes familles atteintes ont permis d'identifier les gènes prédisposant à ce syndrome. Les mutations du gène *RUNX1* ont été les premières à être impliquées dans des cas de thrombopénie familiale avec une incidence accrue de SMD et LAM associée à une myélodysplasie [105]. Dans un deuxième temps, ce sont des mutations de *CEBPA* qui ont été détectées dans des familles à forte incidence de LAM [106]. Puis dans d'autres familles, ce sont des anomalies de *TERT* et de *TERC* qui ont été associées à une incidence anormale de LAM [107]. Plus récemment, des mutations de *GATA2* ont été impliquées dans plusieurs familles qui présentaient des phénotypes hétérogènes tels que le syndrome de MonoMAC (associant monocytopénie sévère, lymphopénie B avec diminution des cellules NK, et des infections mycobactériennes atypiques) [108], le syndrome d'Emberger (associant lymphoedème primaire, verrues cutanées et surdité) [109] ou encore une neutropénie chronique [110]. L'étude immunophénotypique de la moelle osseuse de certains de ces patients a permis de mettre en évidence un arrêt de différenciation précoce

dans l'hématopoïèse avec absence de progéniteurs communs lymphoïdes (MLP) , et donc de prouver l'importance de *GATA2* pour la différenciation des progéniteurs multipotents (MPP) en progéniteurs communs lymphoïdes (figure 1).

3.3. Leucémogénèse

3.3.1. Cellules souches leucémiques

Il est généralement admis que la plupart des cancers résultent d'une prolifération clonale au départ d'une cellule anormale ayant acquis des altérations génétiques et / ou épigénétiques déterminantes, et capables d'échapper aux mécanismes de contrôle de l'organisme. Dans les néoplasies sanguines, identifier la cellule dans laquelle s'initie la transformation maligne est un défi majeur, d'autant plus que le clone leucémique devient très souvent hétérogène avec le temps du fait de l'accumulation de nouvelles anomalies génétiques. La longue durée de vie des CSH et des progéniteurs les plus immatures les rendent plus susceptibles d'être le siège des évènements de transformation initiale (figure 9). De plus, la capacité limitée des blastes à proliférer laisse penser que ceux-ci doivent être continuellement produits par des cellules immatures anormales, les cellules souches leucémiques (CSL), selon une hiérarchie assez similaire à celle des cellules hématopoïétiques normales. Du fait de la grande hétérogénéité immunophénotypique des blastes et des CSL, il est très difficile de purifier ces cellules leucémiques primaires sur base d'un ou de quelques marqueurs cellulaires. L'existence des CSL était suspectée sans être prouvée jusque dans le milieu des années 90. L'équipe du Dr J.E. Dick a mis en évidence une rare sous-population de blastes dans le clone leucémique, qui était capable de ré-initier une LAM après avoir été greffée dans une souris immunocompromise NOD/SCID. Cette équipe a de plus estimé la fréquence de ces CSL à une cellule dans 250 000 blastes [111].

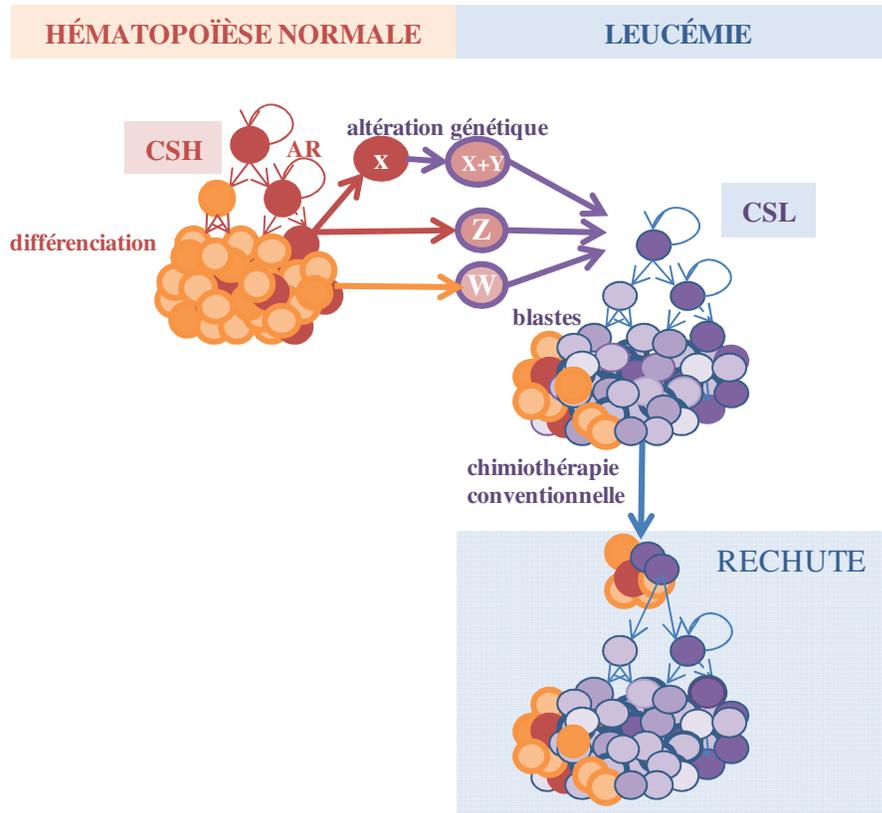


Figure 9. Modèle théorique du développement clonal d'une leucémie et notion de cellule souche leucémique

Dans ce modèle, les CSL proviennent des CSH ou des progéniteurs immatures après la survenue d'une ou de plusieurs anomalies génétiques (X, Y, Z ou W) critiques suffisantes pour maintenir ou réacquérir une capacité d'auto-renouvellement dérégulée. La rechute survient suite à la prolifération de blastes issus des CSL résistantes à la chimiothérapie conventionnelle. CSH: cellule souche hématopoïétique; AR: auto-renouvellement; CSL: cellules souches leucémiques.

Il a ensuite été démontré que le compartiment des CSL lui-même est composé de sous-populations dont le potentiel d'auto-renouvellement (AR) est variable [112] et que les CSL forment un groupe de cellules phénotypiquement très hétérogène. Bien que la biologie des CSL soit encore mal connue, il semble de plus en plus évident que les caractéristiques communes entre CSL et CSH, telles que leur capacité d'AR, de se différencier et même de migrer, reposent sur la conservation ou la ré-acquisition des mécanismes moléculaires qui sous-tendent ou régulent ces processus [113]. Une meilleure connaissance de ces mécanismes est cruciale pour le développement de thérapies capables d'éradiquer les CSL les plus quiescentes qui sont résistantes aux traitements classiques, et donc potentiellement

responsables des rechutes (figure 9). Les modèles murins de leucémies humaines, grâce à la bonne conservation des mécanismes moléculaires entre les deux espèces, sont utilisés dans ce but.

3.3.2. Anomalies géniques des facteurs de transcription

Les anomalies génétiques associées aux hémopathies malignes impliquent souvent des gènes essentiels à l'identité des CSH et à la détermination de leur destin entre AR, différenciation et apoptose. Parmi les gènes et les voies de signalisation connus pour être impliqués dans l'auto-renouvellement des CSH mais aussi des CSL, il faut citer les gènes *HOX* et notamment *HOXA9* et *HOXA10* [114], les voies de signalisation de *WNT/β-catenin* [115] ou encore de *NOTCH1* [116]. Des études d'expression dans différents modèles murins de CSL, telles que les leucémies induites par l'expression de différents oncogènes de fusion comme *NUP98-HOXA9* par exemple, montrent l'activation de l'expression de gènes similaires impliquées dans l'AR: *Meis1*, *Sox4*, *Tcf4*, *Hoxa9*, etc... [117]. Des études dans les populations cellulaires humaines enrichies en CSL, en comparaison à une population de blastes ou à une population de CSH normales, ont permis d'identifier une expression différentielle des gènes *ERG*, *MEIS1*, *EVII*, *HOXA5*, *MEF2C* et *SETBP1* dans les CSL, quel que soit leur oncogène de fusion, par rapport aux CSH [118]. Enfin, des mutations récurrentes chez les patients impliquent différents gènes essentiels pour l'AR des CSH tels que *RUNX1*, *EVII*, *ETV6*, *GATA2* ou *TALI* ou encore des gènes impliqués dans la différenciation myéloïde tels que *CEBPA* ou *SPI1* pour les LA. Dans les SMD, différentes mutations récurrentes impliquent aussi ces facteurs de transcription dont *RUNX1* et *ETV6* [119]. De plus, ce sont des mutations héréditaires de *RUNX1*, *GATA2* et *CEBPA* qui ont été impliquées dans les familles présentant des syndromes de prédisposition familiale aux SMD et aux LAM tels que le syndrome de MonoMAC [120].

3.3.3. Dérégulation épigénétique

3.3.3.1. Anomalies de MLL

Des anomalies épigénétiques sont également fréquemment impliquées dans les hémopathies malignes, dont les réarrangements du gène *MLL* (ou *MLL1*) qui est localisé sur la bande chromosomique 11q23. Comme on l'a vu précédemment (cf. Chap. I. 1.4.2.2.3.2), *MLL* code pour une HMT capable de tri-méthyler le résidu d'histone H3K4 (H3K4me3, marque activatrice). La protéine MLL est constituée de différents domaines dont, de l'extrémité N-terminale à l'extrémité C-terminale, : les domaines AT hook (ATH) permettant l'ancrage à l'ADN, les domaines SNL qui permettent sa localisation nucléaire (speckled nuclear localization), un domaine TRD de répression transcriptionnelle (Transcriptional Repression Domain), les domaines PHD qui permettent l'interaction avec la chromatine ou avec différentes protéines (Plant Homology Domain), un domaine d'activation transcriptionnelle (TA) et surtout le domaine SET (Su(var)3-9, enhancer of-zeste trithorax) à l'extrémité C-terminale qui est le siège de l'activité catalytique de méthylation de MLL (figure 10 A).

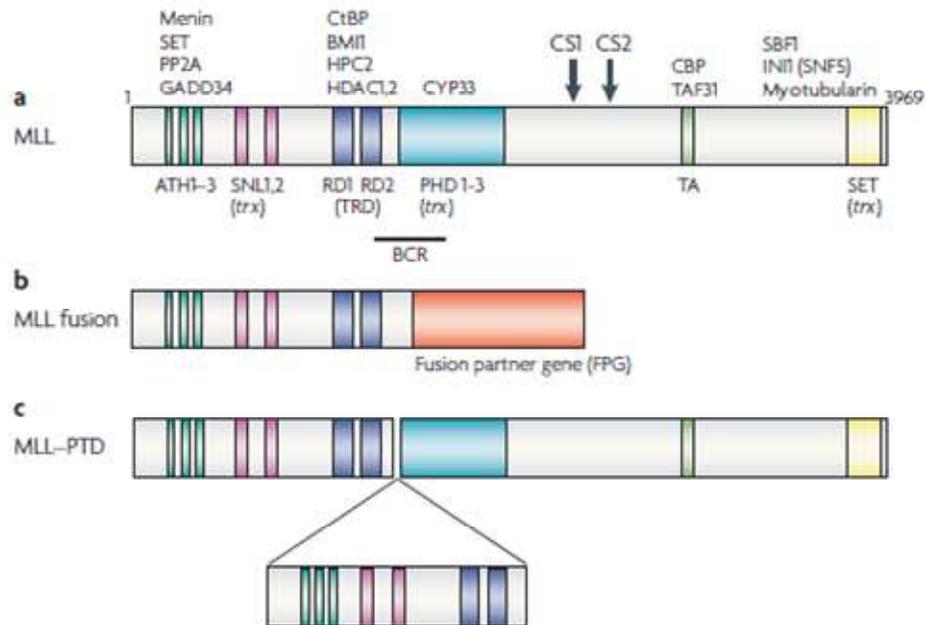


Figure 10. Représentation schématique de MLL et des fusions impliquant MLL

Figure adaptée de Nat Rev Cancer 7(11), Krivtsov, A. V., *et al.*, (2007)[121], "*MLL translocations, histone modifications and leukaemia stem-cell development*", 823-833. Copyright © 2007, avec la permission de Nature Publishing Group. (A) Structure de la protéine encodée par le gène *MLL* (Mixed Lineage Leukemia avec ses différents domaines protéiques). (B) Structure de la protéine de fusion MLL produite à partir des translocations *MLL*: association de la partie N-terminale de la protéine MLL avec la partie C-terminale d'un des partenaires de fusion. (C) schéma présentant la structure d'une duplication en tandem de *MLL* (*MLL*-PTD). ATH: AT hook; SNL: speckled nuclear localization; TRD: transcriptional repression domain; RD1: repression domain 1; RD2: repression domain 2; PHD: Plant Homology Domain; TA: transcriptional activation; SET: Su(var)3-9 enhancer-of-Zeste, trithorax; trx, trithorax; BCR: breakpoint cluster region.

MLL est réarrangé dans 15 à 20 % des LA de l'enfant. L'impact des réarrangements *MLL* sur le pronostic varie en fonction du partenaire de fusion et est généralement défavorable, à l'exception de la translocation t(1;11) qui a le plus souvent un bon pronostic [122]. On retrouve aussi ce réarrangement dans plus de 70% des LA du nourrisson [123] et également dans 9% des leucémies secondaires aux traitements cytotoxiques (inhibiteurs de la topoisomérase II, agents alkylants ou radiothérapie) [124]. Les réarrangements de *MLL* associent la partie N-terminale de la protéine avec la partie C-terminale d'un de ces partenaires de fusion (figure 10 B). Plus de 60 partenaires ont été décrits jusqu'à maintenant. La région du point de cassure (BCR ou breakpoint cluster region) est située entre les exons 8 et 13 de *MLL*, en amont du domaine SET et implique la perte de la fonction catalytique de la protéine de fusion générée. On distingue les partenaires de fusion selon leur fonction: les protéines nucléaires (*MLLT2/AF4*, *MLLT3/AF9*, *MLLT10/AF10*, *MLLT1/ENL* et *ELL*), les protéines cytoplasmiques (*MLLT4/AF6*), les septines et les histones acétyltransférases (CBP, p300) [121]. Sur le plan fonctionnel, bien que chacun de ces partenaires fusionnés à *MLL* a très probablement un mode d'action différent, ils participent tous à l'oncogénicité de la protéine de fusion par un rôle commun d'activateur de la transcription associé à l'effet du dérivé N-terminal de *MLL* qui, quant à lui, permettrait la liaison de la protéine de fusion à des zones précises du génome[125]. Certains partenaires impliquent l'action de l'enzyme DOTL1, une HMT qui triméthyle H3K79 [126]. La conséquence commune de ces différents processus oncogéniques est en effet l'activation dérégulée de la transcription de gènes *HOX* (*HOXA7* et *HOXA9*) et de *MEIS1*. Le fort potentiel oncogénique des fusions *MLL* est démontré dans des modèles murins où la transfection de certains de ces gènes de fusion dans

des progéniteurs de la moelle osseuse provoque la réactivation des capacités d'AR et induit une leucémie [127, 128].

L'autre anomalie de *MLL* rencontrée est la duplication en tandem (*MLL*-PTD). Celle-ci est retrouvée dans environ 10% des LAM à caryotype normal et est associée à un mauvais pronostic. Elle consiste en une duplication en tandem des exons 5 à 11 (figure 10 C). Le mécanisme responsable de l'oncogénicité de cette anomalie de *MLL* est moins bien connu, mais il conduit également à une expression dérégulée des gènes *HOXA* [129].

3.3.3.2. Autres anomalies épigénétiques impliquées dans les LAM et les SMD

Des anomalies de gènes ayant un rôle dans la méthylation des histones ou de l'ADN ainsi que de gènes codant pour d'autres modificateurs de la chromatine ont été mises en évidence ces dernières années dans les néoplasies myéloïdes. Parmi elles figurent des translocations impliquant des gènes tels que *MOZ* ou *KDM5A* [130] telle que la translocation fusionnant *KDM5A* à *NUP98*, la t(11;12)(p15;p13) qui a été décrite dans les LA et dans 8 à 10% des leucémies mégacaryocytaires [69, 131, 132]. On retrouve également des mutations ponctuelles ou des délétions de gènes tels que *DNMT3A*[39], *EZH2*, *TET2*, *IDH1*, *ASXL1* [133], *KDM3B* [65, 66], *KDM5A* [70] *KDM5B* [71] ou *KDM6A* [73, 74]. Plusieurs effecteurs épigénétiques sont aussi surexprimés dans différents cancers dont les leucémies: certaines *DNMT*, certaines *HDAC*, *KDM1A* [63], *KDM2B* [64], *KDM4C/JMJD2C* [134] ou encore *KDM6B* [75]. La surexpression de *KDM4C* par exemple participe à la leucémogénèse par la levée d'inhibition de la transcription de gènes de l'AR (figure 11). Ceci illustre bien l'importance du niveau d'expression des régulateurs épigénétiques dans la régulation de l'hématopoïèse. De plus, il a été démontré que les profils de méthylation de l'ADN ont une influence sur le pronostic des LAM [135]. Ces différentes données laissent suspecter également que les anomalies épigénétiques ont possiblement un rôle dans la résistance des CSL aux traitements [128].

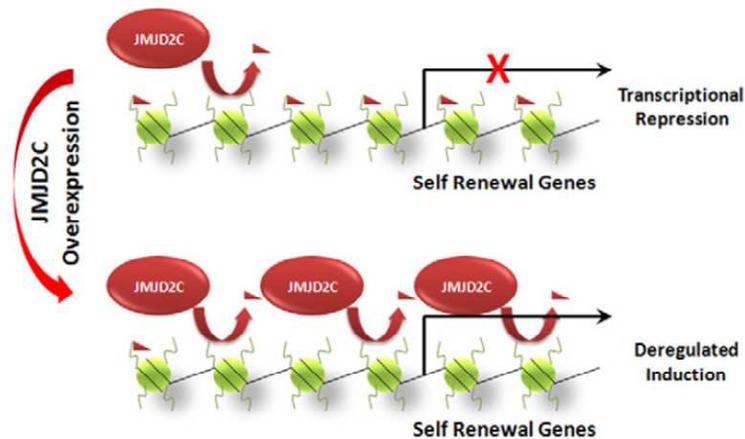


Figure 11. Effet de la surexpression de KDM4C/JMJ2C

Figure adaptée de Am J Pathol, 175(4), Muntean A.G. *et al* (2009)[134] "*Epigenetic Dysregulation in Cancer*" 1353-1361 Copyright © 2009, avec la permission de Elsevier. KDM4C/JMJD2C déméthyle H3K9me (triangles rouges) et conduit ainsi à l'activation de la transcription de gènes importants pour l'auto-renouvellement des cellules souches hématopoïétiques. Dans les cancers avec surexpression de *KDM4C/JMJD2C*, une réduction globale de la méthylation de H3K9 conduit à l'activation de gènes qui normalement seraient réprimés et contribuent alors à la cancérogénèse.

Bien que la pathogénèse des SMD ne soit pas encore bien connue, certaines anomalies génétiques récurrentes commencent à être mises en évidence dans cette maladie. Environ 20% des patients adultes avec hémopathies myéloïdes (incluant des SMD) ont une mutation de *TET2* (ADN déméthylase) [45]. D'autres mutations impliquant des gènes importants pour la régulation épigénétique tels qu'*EZH2* (HMT) [136], *DNMT3A* (méthylation de l'ADN), *ASXL1* (membre du complexe PcG), ou *IDH1/2* [119] ont également été décrites.

3.3.4. Origine prénatale de la leucémie de l'enfant

Dans la majorité des cas, les leucémies résultent de l'accumulation de multiples anomalies génétiques qui favorisent la transformation maligne et confèrent aux blastes un avantage de survie sur les cellules normales. Dans le cas des leucémies de l'enfant, il est très probable que l'anomalie initiale survienne déjà *in utero*. Cette hypothèse est soutenue par la survenue chez des jumeaux monozygotiques monochorioniques d'une leucémie aigüe similaire exprimant l'oncogène *ETV6 – RUNX1* produit de la translocation t(12;21) (p13 ;q22) [137]. Le partage prénatal du clone pré-leucémique résulte vraisemblablement de la transfusion

placentaire entre les deux jumeaux. Pour chacun d'entre eux surviennent après la naissance des anomalies génétiques additionnelles différentes qui vont permettre le développement de la leucémie [137, 138]. L'origine prénatale de ces hémopathies est aussi suggérée par la détection d'autres anomalies génétiques impliquées dans la leucémie (telles que la translocation t(4;11)(q21;q23) *MLL-AF4* par exemple) qui ont pu être mises en évidence dans des prélèvements sur papier buvard faits à la naissance chez des patients qui ont développé par la suite une leucémie [139]. Le très court temps de latence des leucémies impliquant le gène *MLL* laisse aussi présager un potentiel oncogénique puissant, requérant peu ou pas d'autres mutations pour conduire à la transformation maligne. En effet, un réarrangement *MLL* est détecté chez plus de 70% des leucémies des nourrissons (c'est-à-dire les enfants âgés de moins d'un an) et aussi associé aux leucémies secondaires à la chimiothérapie, notamment en lien avec l'exposition aux épipodophyllotoxines, pouvant induire des LMA moins de 3 ans après le traitement [140, 141].

4. APPLICATIONS THÉRAPEUTIQUES

4.1. Favoriser la prolifération des cellules souches

Les CSH sont capables de restaurer un système hématopoïétique complet lorsqu'elles sont greffées à un patient receveur après avoir été conditionné par un traitement myéloablatif. Ce type de traitement est utilisé pour des hémopathies malignes et aussi pour certaines conditions bénignes sévères, les syndromes d'immunodéficience par exemple. La qualité du greffon et son contenu en cellules souches par rapport au poids du patient sont déterminants pour la prise de la greffe. Une source de CSH telle que le sang de cordon, bien que facilement disponible, est ainsi limitée dans son utilisation pour les patients adultes car leur contenu en CSH est souvent insuffisant pour leur poids. Toute stratégie thérapeutique visant à permettre une augmentation du nombre des CSH pourrait dès lors être utilisée pour augmenter la taille des greffons. L'identification des facteurs nucléaires notamment épigénétiques importants pour la régulation de l'AR des CSH est une avenue thérapeutique intéressante, appuyée par les résultats de l'inhibition de *Kdm5b* chez la souris, si cet effet peut être reproduit avec *KDM5B* dans les cellules humaines [90].

4.2. Cibler les mécanismes de régulation dysfonctionnels.

L'implication grandissante d'anomalies génétiques de régulateurs épigénétiques et surtout des HDM dans de nombreux cancers hématologiques tels que les LAM et les SMD dont le pronostic est encore loin d'être satisfaisant même chez l'enfant, ouvre la voie à de nouvelles thérapies ciblées. Les mutations des régulateurs épigénétiques peuvent être activatrices ou inactivatrices pour un même gène dépendamment du contexte cellulaire. De plus, les anomalies épigénétiques sont potentiellement réversibles et la plupart d'entre elles concernent des enzymes. Or les enzymes sont plus faciles à cibler par des inhibiteurs chimiques que les facteurs de transcription. Toutes ces raisons en font des cibles de traitement privilégiées. Certains inhibiteurs d'effecteurs épigénétiques sont déjà approuvés pour être utilisés en clinique (déacétabine et 5-azacitidine pour les SMD ou les inhibiteurs de HDAC dans les lymphomes T). De nombreux autres inhibiteurs des enzymes épigénétiques sont en cours de développement et sont présentés dans la figure 12 [142].

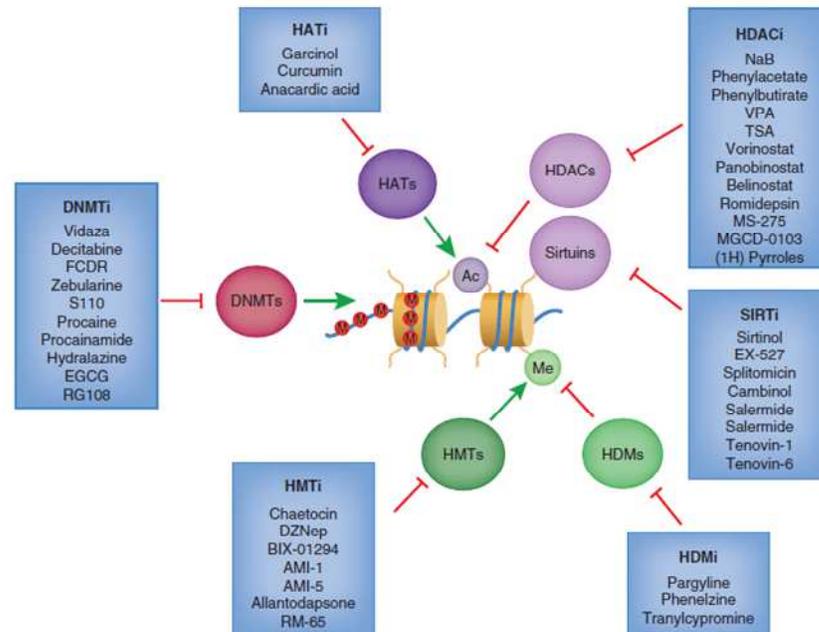


Figure 12. Traitements épigénétiques en développement et leurs cibles

Figure tirée de Nat Med, Rodrigues-Paredes *et al* (2011)[142] « *Cancer epigenetics reaches mainstream oncology* »,330-339, Copyright © 2011, avec la permission de Nature Publishing Group). Représentation schématique des différents inhibiteurs chimiques (-i) ayant pour cible

des effecteurs épigénétiques et de leur effet inhibiteur (lignes rouges) ou activateur (flèches vertes). DNMTs: DNA methyltransferases; HATs: histone acetyltransferases; HDACs: histones deacetylases; HDMs: histone demethylases ; HTMs: histones methyltransferases.

Compte tenu de son implication dans les LAM, des inhibiteurs de KDM1A (inhibiteurs de monoamine oxydase MAOi; tranlycypromine) sont en cours de développement, de même que des inhibiteurs de la famille JmjC (2,4-PDCA) ou encore des PADI tel que le Cl- amidine qui a démontré un effet cytotoxique important sur des cellules cancéreuses. Aucun effet n'a été démontré à date sur les lignées non cancéreuses [143].

L'identification de nouvelles cibles thérapeutiques des leucémies nécessite la réalisation d'études d'expression des régulateurs de l'hématopoïèse dans les cellules hématopoïétiques normales et leucémiques afin de diriger les études fonctionnelles ultérieures.

5. MÉTHODES D'ANALYSE DE L'EXPRESSION GÉNIQUE

5.1. Définition du transcriptome

Ce qui différencie deux cellules qui contiennent le même ADN mais provenant de deux tissus différents d'un même organisme est l'expression différentielle de leur génome conduisant à la production de protéines aux rôles différents, à la production de quantités différentes d'une même protéine ou encore à l'expression de transcrits différents d'une même protéine. On appelle transcriptome l'ensemble des éléments transcrits à partir d'un génome à un moment donné, par un type de cellule donné et dans des conditions environnementales spécifiques. Il est constitué d'ARN messager (ARNm ou transcrits) et d'ARN non-codants dont font partie les ARN ribosomiaux (ARNr) mais aussi les microARN et les longs ARN non-codants qui jouent un rôle important dans la régulation de la transcription et l'identité cellulaire[144]. L'analyse du transcriptome est un véritable instantané qui permet de connaître l'expression génique d'un tissu donné dans des conditions particulières. C'est l'étape initiale nécessaire de l'étude fonctionnelle d'un processus biologique particulier. Il existe plusieurs méthodes qui permettent de déterminer le profil d'expression des gènes exprimés par une population cellulaire cible: les puces à ADN, la qRT-PCR et le séquençage de l'ARN (RNA-seq). Ces trois méthodes utilisent l'ADN

complémentaire (ADNc) produit par transcription inverse à partir de l'ARN total extrait des populations cellulaires étudiées, ou encore à partir d'une fraction d'ARN enrichie en ARN messagers (ARNm). L'enrichissement en ARNm est basé sur le fait que ces ARN possèdent une extrémité poly-A et peut se faire soit avant la transcription inverse par chromatographie d'affinité de l'ARN total sur des colonnes oligodT, soit lors de la transcription inverse de l'ARN total en utilisant des oligonucléotides poly-T plutôt qu'un mélange d'amorces. La purification des ARNm est plutôt recommandée pour l'analyse par puces à ADN (ARNm avec extrémités 3' intactes pour réaliser la banque d'ADNc, cf ci-après), et n'est pas nécessaire pour la qRT-PCR ou le séquençage de l'ARN (des fragments d'ADNc sont utilisables).

5.2. Les puces à ADN ou « micro-arrays »

La méthode des puces à ADN utilise l'hybridation réversible entre des fragments d'ADN complémentaires (ADNc) provenant d'une population cellulaire et un ensemble de séquences d'ADN connues appelées sondes, qui sont chacune spécifiques d'un seul gène et associées à un support rigide. Le principe de cette méthode est illustré dans la figure 13. Les ADN complémentaires (ADNc) au transcriptome de l'échantillon cellulaire à tester et d'un échantillon témoin sont générés par transcription inverse en utilisant des fluorochromes (par exemple rouge pour l'échantillon test et vert pour l'échantillon témoin). Après hybridation des ADNc aux sondes du support solide, des images sont captées numériquement pour quantifier la fluorescence qui représente le degré d'hybridation.

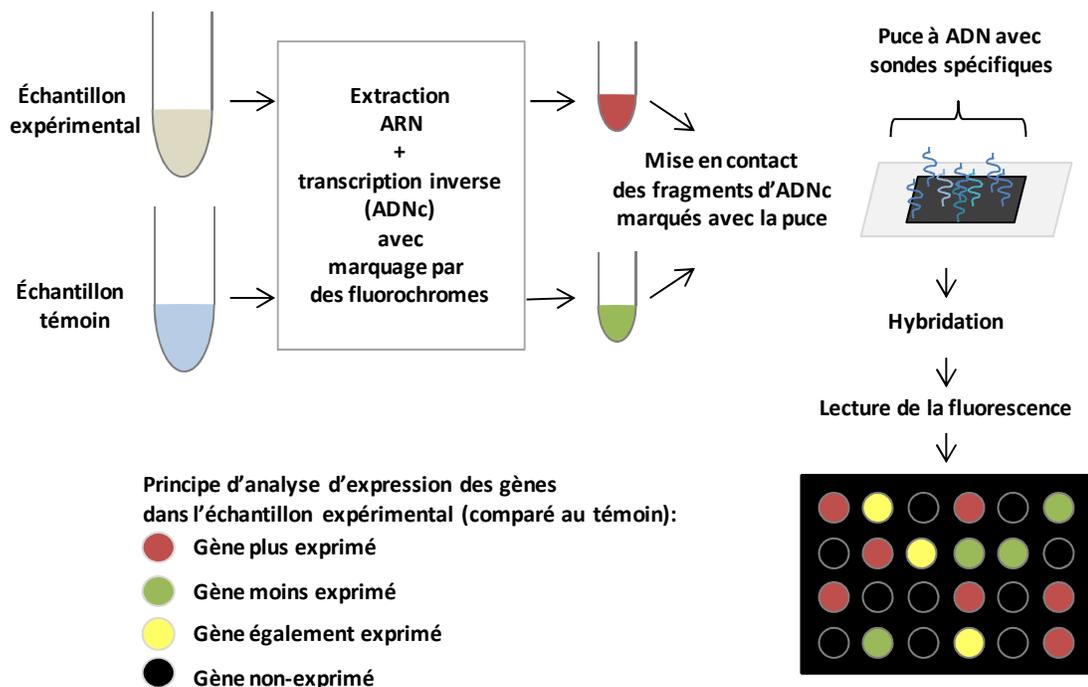


Figure 13. Principe de la méthode de quantification du transcriptome par puce à ADN
 L'ARN est extrait à partir d'un échantillon cellulaire expérimental et d'un échantillon témoin. L'ADN complémentaire (ADNc) est produit et marqué grâce à l'utilisation de fluorochromes de couleur différentes (par exemple rouge et vert). Les fragments d'ADNc marqués sont ensuite mis en contact avec une puce à ADN sur laquelle sont disposés des sondes qui sont des fragments d'ADN spécifiques de gènes connus. Les fragments d'ADNc marqués vont s'hybrider à la sonde leur correspondant si celle-ci est présente sur la puce. Après lavage, la fluorescence résultant des hybridations des fragments « expérimentaux » et/ou « témoins » est mesurée et permet d'estimer l'expression des gènes de l'échantillon expérimental en comparaison à l'échantillon témoin.

Le ratio des deux fluorescences reflète le niveau d'expression relative des gènes testés dans la population cellulaire expérimentale par rapport à la population témoin. La représentation graphique couramment utilisée est celle d'une carte de densités d'expression ou Heat-Map où les gènes les plus exprimés dans la population testée vont sembler plus rouges, et ceux moins exprimés plus verts. Une expression inchangée apparaîtra en noir. Le degré d'intensité de la « couleur » reflète l'importance de l'expression différentielle.

Les inconvénients principaux de cette méthode sont principalement la nécessité de définir le panel de gènes à étudier ainsi que la limitation dans l'amplitude de niveau d'expression mesurable. Cette technique ne permet de mesurer que les gènes représentés sur la puce et est

limitée par les contraintes physiques et techniques d'impression de la puce. Cette méthode est actuellement de plus en plus remplacée par les techniques de séquençage de nouvelle génération (SNG) qui sont en voie de devenir le nouveau standard idéal en termes de mesure d'expression.

5.3. La qRT-PCR

La PCR (« Polymerase Chain Reaction ») permet d'amplifier une séquence déterminée d'ADN par une succession d'étapes de dénaturation, d'hybridation d'amorces et élongation d'un nouveau brin d'ADN. Ces étapes sont rendues possibles par l'utilisation de variations cycliques de température grâce à un thermocycleur et de l'utilisation d'une ADN polymérase thermorésistante. Les amorces sont dessinées pour être spécifiques du gène à identifier. En utilisant l'ADN complémentaire (ADNc) produit par transcription inverse après extraction de l'ARN total à partir d'une population cellulaire choisie, il est possible de détecter les gènes qui sont exprimés. La sensibilité de la PCR est extrême: on peut détecter un ARN présent même à raison de seulement quelques copies dans une population cellulaire. C'est un outil très fréquemment utilisé en médecine diagnostique. La qRT-PCR ou PCR quantitative en temps réel (« quantitative Reverse Transcription Polymerase Chain Reaction ») permet de mesurer la quantité d'ADNc amplifié à chaque cycle de PCR par la mesure de l'intensité de la fluorescence émise par un fluorochrome intercalé entre les brins d'ADNc lors de l'amplification, comme le SYBR green, ou émise par une sonde couplée à un fluorochrome lors de sa libération suite au passage du complexe de transcription (ex. méthode Taqman, figure 14).

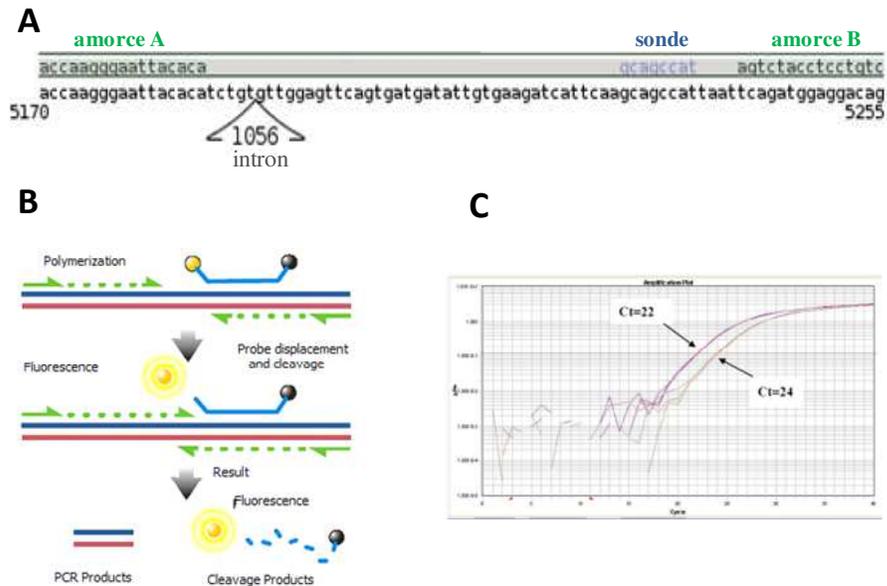


Figure 14. Exemple de méthode de quantification du transcriptome par qRT-PCR

(A) Test de qRT-PCR de type Taqman, développé et validé pour notre projet pour le gène *MLL* avec deux amorces spécifiques A et B et une sonde couplée au fluorochrome. (B) (Tiré de <http://en.wikipedia.org/wiki/File:Taqman.png>). Principe de la qRT-PCR par la méthode Taqman®: la fluorescence est émise lors de l'élongation de l'amplicon. (C) (tiré de www.genomique.irc.ca) Représentation graphique des résultats de qRT-PCR: la mesure du moment (mesure du temps en abscisse) d'émission de la fluorescence (amplitude de fluorescence en ordonnée) permet d'évaluer la quantité d'un transcrite donné par rapport à un gène contrôle dans une population cellulaire donnée en comparaison à une autre population cellulaire. Ct: cycle seuil.

Pour la méthode Taqman, que nous avons choisi d'utiliser pour ce projet, un test spécifique à chaque gène est validé à l'aide d'un couple d'amorces et d'une sonde associée à un fluorochrome. Les tests peuvent identifier soit l'ensemble des transcrits existants du gène, soit un seul transcrite du gène selon le cas. À chaque cycle, la quantité d'ADN est multipliée par deux et le degré de fluorescence augmente proportionnellement. Le paramètre déterminant est le Ct: c'est le cycle auquel l'amplitude de la fluorescence émise atteint un certain seuil. Plus le Ct est bas, plus le transcrite est abondant dans l'échantillon. Le plateau de fluorescence est atteint lorsqu'un des composants est épuisé. Il s'agit ici aussi, comme pour les puces à ADN, d'une quantification relative de l'expression génique à deux niveaux: expression d'un gène cible ($Ct_{\text{gène}}$) par rapport à un autre gène ($Ct_{\text{endogène}}$) qui sert de

contrôle endogène (Delta Ct (ΔCt) = $Ct_{\text{gène}} - Ct_{\text{endogène}}$), puis par rapport à un échantillon qui, lui, sert de calibrateur ($RQ = 2^{-\Delta\Delta Ct}$) (cf. Chap. II.6.7). Le contrôle endogène permet de normaliser les biais d'extraction de l'ARN, liés au dosage, liés à sa qualité, liés à une possible contamination par des inhibiteurs et les biais liés à l'efficacité de la transcription inverse. Les gènes habituellement utilisés comme gènes de contrôle sont *ABL1*, *TBP*, *HPRT1* ou *GAPDH*. Le choix du gène de contrôle porte généralement sur le fait que ces gènes sont bien exprimés dans tous les tissus analysés et que leur niveau d'expression n'est pas influencé par les conditions expérimentales [145]. L'échantillon calibrateur peut être l'état physiologique normal ou le tissu non-atteint par la pathologie. Cette méthode de quantification de l'expression de certains gènes est spécifique, très sensible et relativement rapide d'utilisation. Elle est utilisée de plus en plus en clinique pour le diagnostic moléculaire ainsi que pour mesurer la réponse à un traitement ou l'évaluation de la maladie résiduelle (MRD ou « minimal residual disease »). Elle ne permet cependant que de mesurer l'expression des gènes préalablement choisis et pour lesquels des tests spécifiques validés ont pu être développés, et ne permet pas de combiner plusieurs expériences, à la différence du RNA-seq.

5.4. Séquençage de l'ARN

Depuis quelques années, des techniques de séquençage de nouvelle génération (SNG) à haut débit ont été développées et révolutionnent la biologie moléculaire. Parmi elles, le séquençage de l'ARN (ou RNA-seq) permet également de quantifier le niveau d'expression d'un gène en le séquençant de façon répétitive, en mesurant le nombre de ces lectures (ou « reads ») puis en les analysant, par exemple en les localisant sur un génome de référence. Le paramètre de mesure est la fréquence de lecture d'un fragment d'ADN et est appelé RPKM (pour « Reads Per Kilobase per Million ») et correspond au nombre de ces fragments par kilobase d'acides nucléiques et par million de séquences localisées. Il est proportionnel à la quantité initiale de transcrit dans l'échantillon. La normalisation de ce paramètre en fonction de la longueur du gène et du nombre total de séquences pour chaque échantillon permet une comparaison des données de différents échantillons (figure 15). Il existe de nombreuses variations dans les méthodes utilisées pour constituer la librairie ou aligner et assembler les fragments de séquences analysées par rapport à un génome de référence [146].

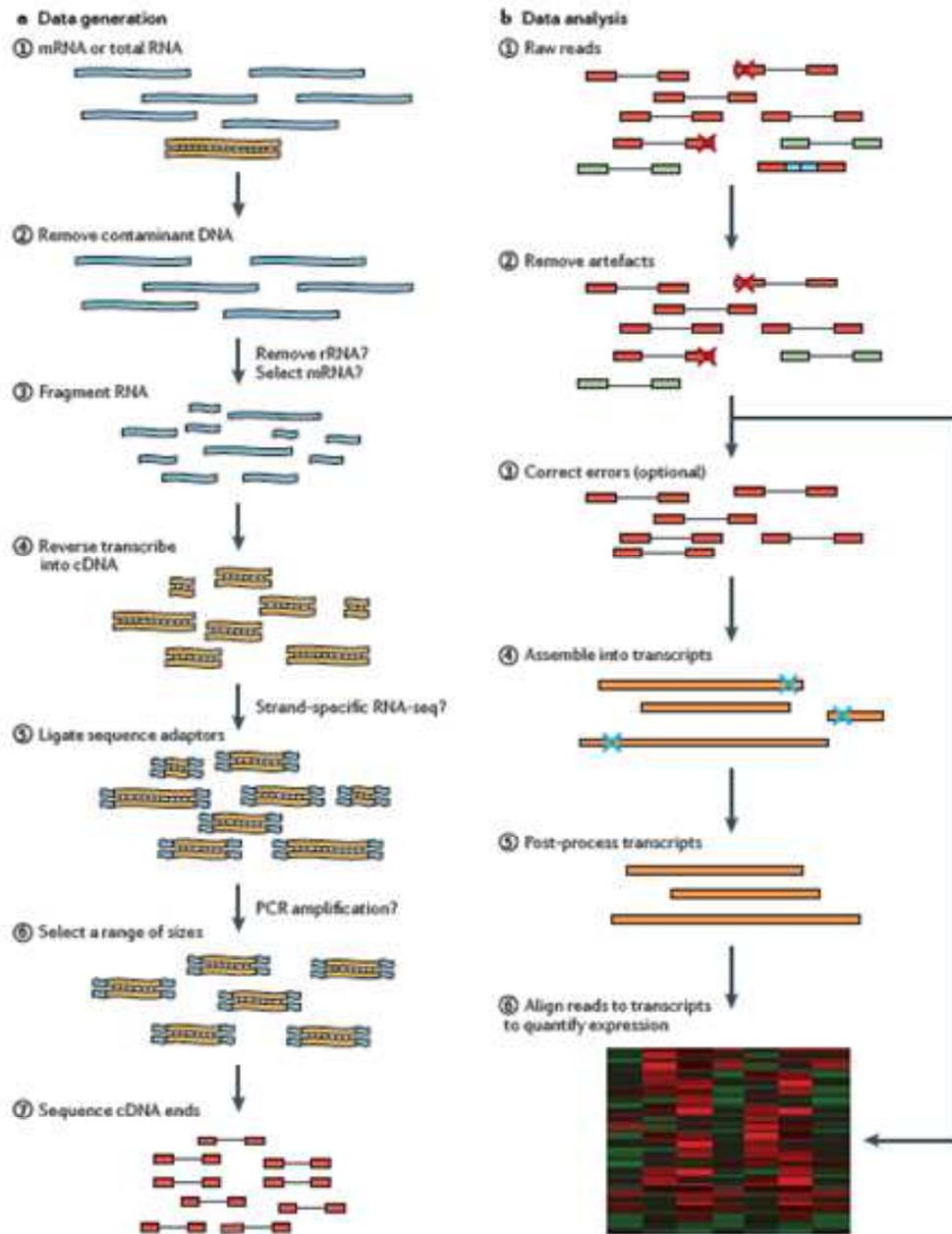


Figure 15. Méthode de quantification du transcriptome par séquençage de l'ARN

Figure adaptée de Nat Rev Genet, 12(10), Martin et Wang (2011) [146] « Next-generation transcriptome assembly », 671-682. Copyright © 2011, avec la permission de Nature Publishing Group. (A) Constitution de la librairie de données: 1- extraction de l'ARN, 2- décontamination de l'ADN génomique, 3- fragmentation de l'ARN, 4- production d'ADNc par transcription inverse, 5- liaison d'adaptateurs de séquençage aux extrémités des fragments d'ADNc, 6- sélection des fragments en fonction de leur taille, 7- appariement de deux séquences après poursuite du séquençage de leurs extrémités. (B) Analyse des données: 1, 2 et 3- les séquences sont préparées par élimination des adaptateurs (bleu), de l'ADN

contaminant (vert) et des séquences comportant des erreurs (croix rouge). 4- les séquences sont alors assemblées pour reformer un transcrit (orange), 5- les erreurs d'assemblage sont éliminées (croix bleues), 6- le niveau d'expression de chaque gène est estimé en comptant le nombre de séquences qui s'alignent sur chaque transcrit.

L'intérêt du RNA-seq est qu'il permet non seulement la quantification du niveau d'expression des gènes actifs dans un tissu donné à un moment donné, mais peut aussi donner la structure des gènes, mettre en évidence un épissage différentiel ou l'expression d'un allèle spécifique. Le RNA-seq ne se limite pas aux gènes connus et permet l'identification de nouveaux transcrits, d'un polymorphisme au niveau d'un nucléotide (SNP ou « Single Nucleotide Polymorphism ») ou encore l'identification d'un transcrit de fusion [147, 148]. C'est une méthode sensible et reproductible. Les principaux inconvénients de cette méthode sont son prix, le temps d'exécution et d'analyse. L'interprétation des résultats nécessite l'expertise de bio-informaticiens. Il est généralement admis qu'il existe une bonne corrélation entre les résultats d'analyse par qRT-PCR et par RNA-seq, sans qu'aucune étude n'ait formellement abordé cette question sur un large échantillon de gènes. Il existe cependant dans la littérature des exemples qui nous confortent dans cette idée [149]). Nous avons choisi de réaliser notre étude d'expression en utilisant ces deux dernières techniques.

5.5. Banques de données génomiques publiques

L'avènement des nouvelles techniques de séquençage à haut débit a marqué un tournant radical dans l'étude de la génétique des cellules normales et leucémiques. Différents groupes ont utilisé ces nouvelles techniques pour rendre disponibles à la communauté scientifique de volumineuses banques de données à accès contrôlé (accès autorisé sur demande). Ces banques proposent des ensembles de données génétiques et d'informations pertinentes, anonymisés et de bonne qualité, récoltés à partir de nombreux échantillons tumoraux et normaux de patients volontaires. Elles sont pour la plupart en cours de construction. La banque de données TCGA ou « The Cancer Genome Atlas Data Portal » (<https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>) compile les informations cliniques, la caractérisation génomique et les résultats d'analyse de séquençage des nombreux génomes tumoraux classés par pathologie ainsi que les génomes germinaux correspondants lorsque possible pour chaque patient. Les données du TCGA proviennent de patients adultes. Toutefois, plusieurs

anomalies génétiques des LAM sont identiques chez l'enfant et l'adulte mais de fréquence différente. Il existe également une banque de données génomiques sur les cancers pédiatriques en cours d'élaboration. Il s'agit du PCGP ou « Pediatric Cancer Genome Project » (<http://www.pediatriccancergenomeproject.org/site/>), projet dirigé par le St-Jude Children's Research Hospital et l'Université de Washington pour mieux comprendre les origines génétiques des cancers de l'enfant. Cette banque regroupe également par pathologie les données issues du séquençage du génome entier provenant d'échantillons tumoraux mais aussi des cellules normales de chaque patient afin d'identifier les anomalies impliquées dans la cancérogenèse. Les seules données actuellement disponibles dans cette banque pour les LAM, sont celles sur les leucémies aiguës mégacaryocytaires (LAM FAB-M7). Ces LAM ne représentent que 10 % des LAM, et ne sont pas pertinentes pour notre analyse.

L'intérêt majeur de ces banques de données génétiques réside dans le grand nombre d'échantillons disponibles (par ex: données de 179 LAM provenant de la banque de données du projet TCGA utilisées dans notre projet). Ceci permet de diminuer l'impact des variations individuelles et de déterminer une tendance plus représentative des profils d'expression pour chacun des gènes exprimés. De façon plus globale, ces banques de données, en stockant une image génétique complète (génome complet, exomes et transcriptome) et anonyme de chacun des échantillons tumoraux, associée à celle d'échantillons normaux pairés et parfois même de spécimens familiaux (parents pour les banques pédiatriques), offrent une masse d'informations très utiles pour la communauté scientifique. Il est à espérer, particulièrement pour les cancers de l'enfant, que les banques de données publiques continuent de se développer comme par exemple la banque de données pédiatriques existante « PCPG » (cf. Chap. I 5.5). En effet cette banque devrait rendre disponible dans un avenir rapproché les données de patients avec les autres types de LAM (et non seulement les FAB-M7 comme c'est le cas pour le moment), et aussi tous les cancers de l'enfant. Un des défis de ce type de récolte systématique d'informations génétiques tumoral et germlinal de nombreux patients des défis techniques et méthodologiques est la gestion de l'information personnelle. En effet bien qu'anonymisées, les informations issues de ce séquençage systématique peuvent mettre en évidence, chez certains patients participants, un gène prédisposant au cancer ou à d'autres pathologies graves. Même si cette éventualité est bien considérée par les banques de données, qui proposent de fournir un retour d'information et d'orienter le sujet vers un centre

qui peut assurer un suivi approprié, les conséquences pour les patients ou leurs proches peuvent être dramatiques (pour l'accès à certaines assurances par exemple). Ces banques de données génétiques sont cependant des outils désormais indispensables à l'étude de la cancérogénèse.

6. DIAGNOSTIC DES MUTATIONS DES GÈNES RÉGULATEURS DES CSH

Les anomalies génétiques récurrentes identifiées dans plusieurs hémopathies malignes permettent de mettre en lumière le rôle des régulateurs importants de l'hématopoïèse normale et de la leucémogénèse. Lorsqu'un patient ou un groupe de patients présentent un phénotype particulier d'hémopathie sévère, il est important de pouvoir diagnostiquer ces anomalies. Leur identification passe par le séquençage du gène suspecté anormal. Ceci peut se faire par séquençage traditionnel (méthode de Sanger) ou par SNG. Nous avons eu l'occasion de développer un test diagnostique par la méthode de séquençage traditionnel pour rechercher une mutation du gène *GATA2* chez un patient qui présentait un SMD avec un tableau de syndrome de MonoMAC (cf. chap. I.3.2).

7. HYPOTHÈSE ET OBJECTIFS

Une importante classe d'enzymes impliquées dans la méthylation de la chromatine, les histones déméthylases (HDM), ont fait leur apparition sur l'échiquier de la régulation des programmes transcriptionnels qui déterminent le destin des CSH entre auto-renouvellement (AR) et engagement dans la différenciation. L'implication croissante de ces enzymes, lorsqu'elles sont anormales, dans des hémopathies malignes nous laisse penser que les HDM sont des acteurs importants pour la détermination du destin cellulaire, favorisant la balance d'expression selon le contexte cellulaire, soit vers l'auto-renouvellement, soit vers la différenciation. Nous suspectons également que, dans les cellules blastiques, leur dérégulation contribue aux programmes transcriptionnels qui soutiennent l'oncogénicité (figure 16).

Il est dès lors nécessaire de décrire les profils d'expression des HDM dans les compartiments cellulaires hématopoïétiques humains différents: des cellules immatures enrichies en CD34⁺, des cellules différenciées normales et des cellules de leucémie.

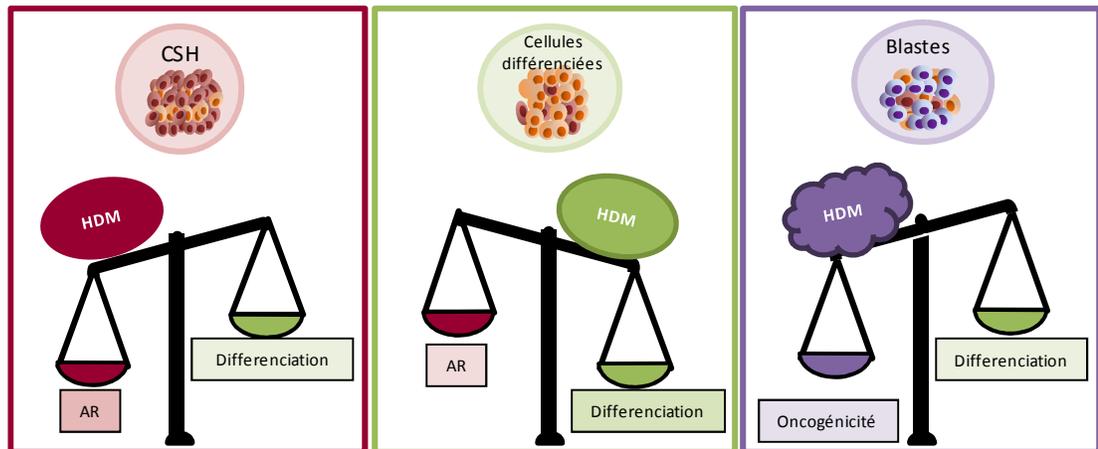


Figure 16. Schéma résumant l'impact proposé des histones déméthylases sur les programmes transcriptionnels propres aux cellules souches hématopoïétiques, aux cellules différenciées et leur dérégulation dans les leucémies

Dans les CSH, le programme transcriptionnel favorisé est celui qui soutient l'auto-renouvellement (AR), alors que dans les cellules matures, c'est celui de la différenciation qui est privilégié. Dans les cellules leucémiques, la dérégulation de différents facteurs nucléaires dont supposément les HDM favorisent l'oncogénicité. CSH: cellules souches hématopoïétiques; HDM : histones déméthylases.

L'objectif principal de ce projet de recherche est de réaliser un profil d'expression génique exhaustif des HDM dans ces trois populations cellulaires hématopoïétiques humaines différentes.

L'objectif secondaire de ce projet est d'identifier lesquelles de ces HDM sont différentiellement exprimées entre ces compartiments cellulaires hématopoïétiques (immatures versus différenciées; immatures versus leucémiques) afin d'identifier des gènes candidats pour la réalisation d'études fonctionnelles ultérieures qui permettront de déterminer si la modulation de ces gènes candidats peut être utilisée dans des stratégies d'expansion des CSH, ou comme cible thérapeutique anti-leucémique.

CHAPITRE II: MATÉRIEL ET MÉTHODES

1. Approche expérimentale

La figure 17 résume notre approche expérimentale. À partir d'unités de sang de cordon provenant d'Héma-Québec, nous avons isolé deux populations cellulaires hématopoïétiques normales: une population de cellules différenciées et une population enrichie en cellules CD34⁺ (population immature composée des cellules souches et de progéniteurs ou CSPH) dont la pureté était ensuite vérifiée morphologiquement et par cytométrie de flux (FACS). Les populations de cellules leucémiques utilisées provenaient soit de lignées cellulaires humaines de LAM MONO-MAC-1 avec translocation t(9;11)(p22;q23) / fusion *MLL-MLLT3*, soit d'échantillons de LAM pédiatriques avec la t(9;11)(p22;q23), cryopreservés à la Banque de Cellules Leucémiques du Québec (BCLQ).

Après extraction de l'ARN total, le profil d'expression des HDM et d'un gène contrôle (*TBP*) a été réalisé dans ces différentes populations cellulaires normales et anormales en utilisant la méthode de RT-PCR quantitative (cellules normales immatures CD34⁺ et différenciées, cellules leucémiques de la lignée MONO-MAC-1). Parallèlement, en collaboration avec Brian Wilhelm, chercheur à l'Institut de Recherche en Immunologie et en Cancérologie (IRIC), le séquençage du transcriptome par séquençage de nouvelle génération (cellules normales immatures CD34⁺ et différenciées, cellules leucémiques primaires de plusieurs LAM pédiatriques avec la t(9;11)(p22;q23)) a été effectué. Les résultats de ces différents profils d'expression ont ensuite été analysés. Ce projet expérimental a reçu l'approbation des différents comités d'éthique à la recherche concernés (Héma-Québec, CHU Sainte-Justine et BCLQ).

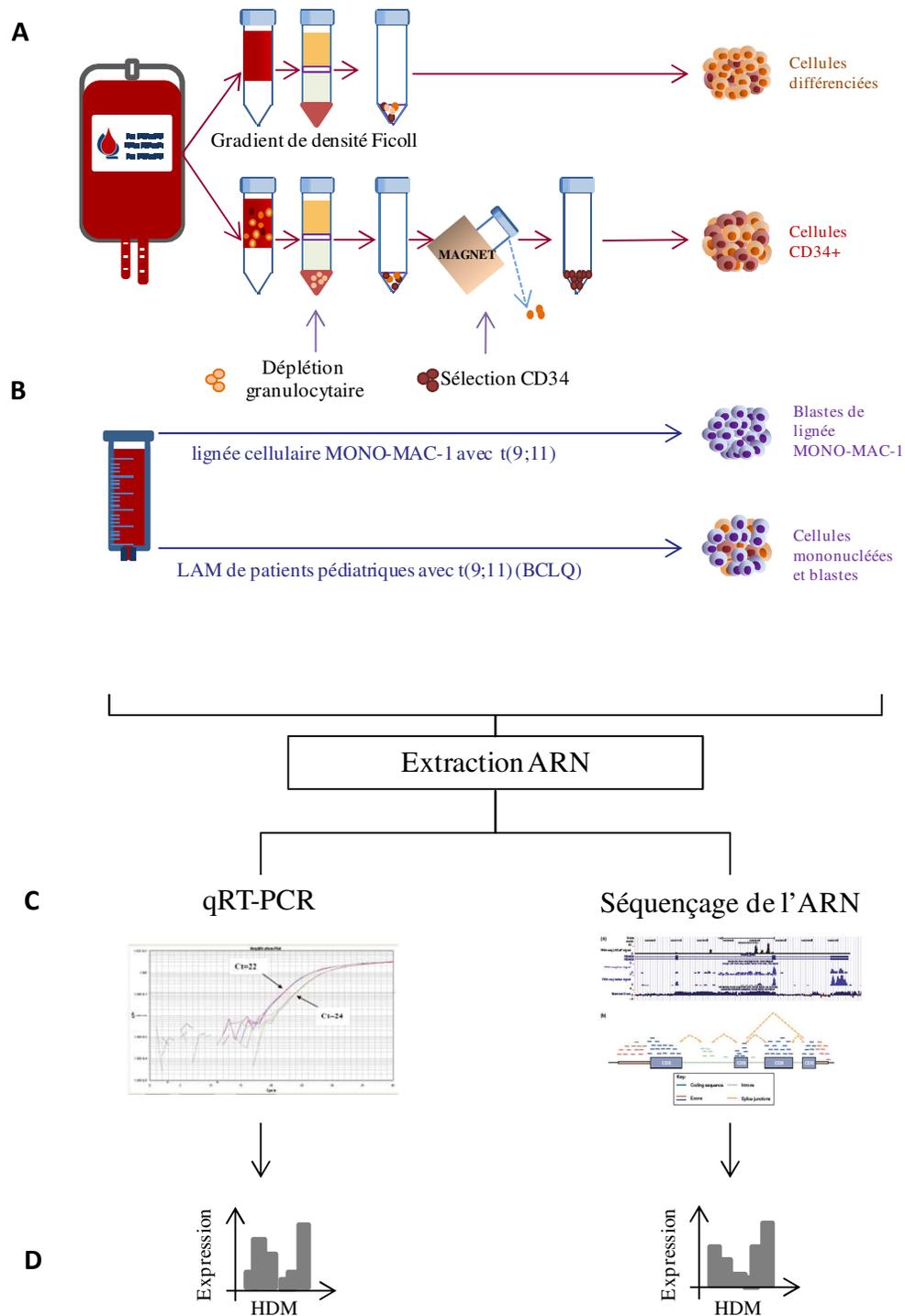


Figure 17 Approche expérimentale du projet

A. Isolation des populations cellulaires normales. B. Source des populations de cellules leucémiques C. Extraction de l'ARN total et quantification de l'expression par qRT-PCR et par séquençage de nouvelle génération. D. Analyse des résultats et comparaison aux données publiques.

2. Utilisation d'unités de sang de cordon comme source de cellules hématopoïétiques humaines normales

2.1. Base rationnelle du choix d'utiliser le sang de cordon comme source de cellules souches hématopoïétiques humaines

Bien que la moelle osseuse soit la source de cellules souches et de progéniteurs hématopoïétiques (CSPH) de référence, elle est bien moins accessible que le sang de cordon. En effet les unités de sang de cordon (USC) non retenues pour les greffes par les banques de sang de cordon, constituent un matériel privilégié pour la recherche sur la biologie des CSH. On considère généralement que les CSH sont globalement semblables quelle que soit leur source, particulièrement pour ce qui est de leur capacité d'auto-renouvellement et de leur potentiel de différenciation. Cependant, certains auteurs ont suspecté que des différences biologiques des CSH provenant des USC par rapport à celles de la moelle osseuse pourraient expliquer le délai de prise de greffe associé aux greffes d'USC: une proportion plus importante de CSPH immatures, une meilleure capacité de prolifération *in vivo*, l'expression différentielle de certains gènes importants pour différents processus cellulaires tels que l'adhésion et la migration cellulaire (par exemple moins d'expression de *CXCR4* dans les CSH issues d'USC comparées à celles de la moelle osseuse) la prolifération et la différenciation (par exemple expression plus importante de *GATA2* dans les CSH issues d'USC comparées à celles de la moelle osseuse)[150, 151]. Pour notre étude nous avons choisi d'utiliser des CSH de sang de cordon, compte tenu de leur relative facilité d'accès.

2.2. Critère de choix des unités de sang de cordon et manipulation initiale

Les unités de sang de cordon utilisées étaient celles rendues disponibles par Héma-Québec pour la recherche car elles ne remplissaient pas les critères de sélection pour être conservées dans le registre de sang de cordon en vue d'une greffe hématopoïétique. Les raisons d'exclusion du registre incluent le nombre de cellules CD34⁺ insuffisant, un volume inférieur à 85mL, un nombre de leucocytes inférieur à 1,3 x 10⁹/sac pour les dons d'origine caucasienne et de 1,1 x 10⁹/sac pour les dons non caucasiens, et des résultats sérologiques

maternels non satisfaisants. Les unités retenues pour notre étude étaient des unités les plus fraîches possibles (cf. Chap. III. 1.1.1). En effet, le délai entre la récolte du sang de cordon et la manipulation des cellules peut altérer certaines caractéristiques biologiques telles que la viabilité cellulaire[152] et pourrait possiblement également, avoir une influence sur l'expression des gènes impliqués dans le stress, le métabolisme, l'auto-renouvellement et la différenciation cellulaire. Il n'y a cependant pas de seuil optimal clairement identifié dans la littérature au-delà duquel l'utilisation de ces CSPH ne serait plus optimale pour les études d'expression. Afin de minimiser l'impact d'un délai trop long, il nous semblait approprié d'accepter un maximum de 36h. Selon notre expérience et celle de nos collaborateurs, nous n'avons pas observé de variation notable dans les profils d'expression lorsque les échantillons étaient analysés dans cet intervalle de temps, ce qui n'était plus le cas pour les échantillons analysés après plus de 48h de la naissance (données non publiées).

Les unités de sang de cordon étaient maintenues à 4°C jusqu'à leur utilisation. Sous hotte biologique, le volume total de chaque unité était mesuré: 5 ml de sang de cordon étaient prélevés pour l'isolement des cellules différenciées (cf. Chap. II.2.2) alors que le reste de l'unité était utilisé pour l'isolement de la population enrichie en cellules immatures CD34⁺ (cf. Chap. II.2.3). Les deux récoltes étaient réalisées en parallèle.

2.3. Isolement de la population cellulaire différenciée

À l'échantillon de 5 ml de sang de cordon était ajoutés 5µl de DNase I puis un volume de milieu tampon Phosphate Buffered Saline - PBS + 2% FBS + 1mM EDTA (appelé « milieu tampon enrichi » dans la suite de la description). L'échantillon de sang dilué était ensuite doucement déposé sur 17 mL de Ficoll-PaqueTM Plus. Après centrifugation pendant 30 minutes à 400 x g, à température ambiante et sans frein, la couche de cellules mononucléées était délicatement récoltée et diluée dans 50 mL de milieu tampon enrichi pour un premier lavage par centrifugation à 300 x g pendant 10 minutes (à température ambiante également sans frein). Le surnageant était alors jeté et le culot cellulaire resuspendu pour un 2^{ème} lavage avec 50mL de milieu tampon enrichi par centrifugation à 120 x g pendant 10 minutes sans frein. Avant cette dernière centrifugation, et après homogénéisation, un échantillon d'environ 20 µl était prélevé pour évaluer la concentration cellulaire. Après la centrifugation, le culot cellulaire était resuspendu dans du milieu tampon enrichi pour procéder aux étapes de

contrôle morphologique et de cytométrie de flux (cf. Chap. II.2.4.). Le reste de l'échantillon de cellules différenciées était ensuite resuspendu dans 1 mL de TRIzol® (Invitrogen, Carlsbad, CA, USA). Ce mélange était ensuite conservé à -80°C pour utilisation ultérieure (voir extraction de l'ARN cf. Chap. II.4.).

2.4. Isolement et sélection de la population cellulaire immature

Pour enrichir la population cellulaire en cellules immatures CD34⁺, nous avons utilisé le kit de sélection positive de StemCell RosetteSep® + EasySep® (Catalog #18096) selon les recommandations du fabricant, hormis l'ajout de DNase-I (permet d'éviter les agrégats cellulaires en éliminant l'ADN libéré par la dégradation de cellules mortes).

2.4.1. Étape 1: pré-enrichissement

Au volume restant de chaque unité de sang de cordon (volume moyen 87 mL) était d'abord ajouté 1µl/mL de volume de sang de DNase I puis 5µl/mL de la solution de pré-enrichissement RosetteSep® Human Cord Blood CD34 Pre-Enrichment Cocktail (Catalog #15631) suivi d'une incubation de 10 minutes à température pièce. Ce cocktail contient un mélange d'anticorps murins associés en tétramères et dirigés contre les antigènes CD66b (pan-granulocytes) et glycophorine A (globules rouges) (figure 18). L'échantillon était ensuite dilué avec un volume égal de milieu tampon enrichi (PBS + 2% FBS + 1mM EDTA). L'échantillon de sang dilué était ensuite doucement déposé à raison de 30mL sur 17mL de Ficoll-Paque™ Plus. Après centrifugation pendant 30 minutes à 400 x g, à température ambiante et sans frein, la couche de cellules mononucléées était délicatement récoltée et diluée dans 50 mL de milieu tampon enrichi pour un premier lavage par centrifugation à 300 x g pendant 10 minutes à température ambiante sans frein. Le surnageant était ensuite jeté et le culot cellulaire resuspendu pour un 2^{ème} lavage dans 50 mL de milieu tampon enrichi par centrifugation à 120 x g pendant 10 minutes sans frein. Avant cette dernière centrifugation, et après homogénéisation, un échantillon d'environ 20 µl était prélevé pour évaluer la concentration cellulaire. Après la centrifugation, les culots cellulaires étaient réunis dans un nouveau tube de 14 mL et resuspendus dans un volume de milieu requis suffisant pour obtenir une concentration de 2×10^8 cellules/mL.

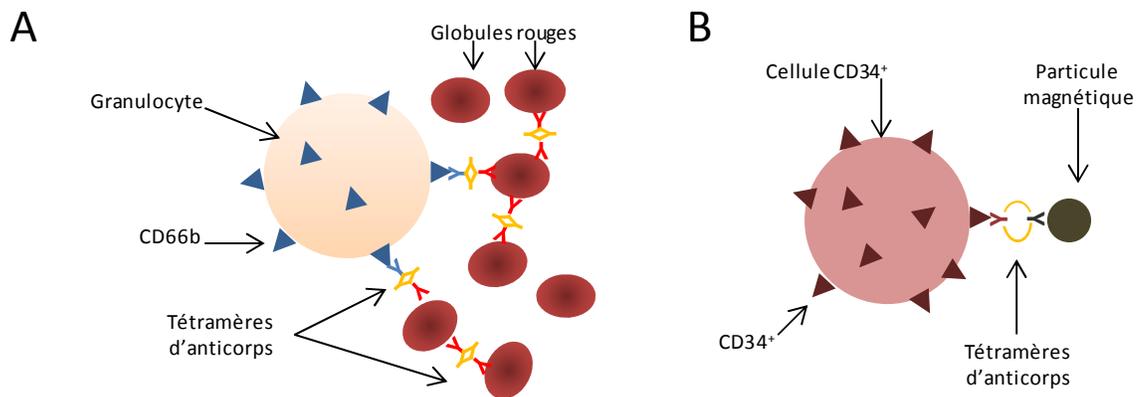


Figure 18. Représentation schématique du principe de chaque étape de sélection

A. Étape de pré-enrichissement: les tétramères d'anticorps dirigés contre le marqueur CD66b des granulocytes (bleu) et contre la glycophorine A des globules rouges (rouge) permettent d'éliminer les cellules matures non voulues. B. Étape d'enrichissement en CD34⁺: utilisation de tétramères d'anticorps dirigés contre l'antigène CD34 (rouge) et contre le dextran des particules magnétiques (noir).

2.4.2. Étape 2: sélection positive

La deuxième étape de criblage s'effectue à l'aide de l'aimant « The Big Easy » Silver EasySep® Magnet (Catalog #18001). Un volume égal de tampon de lyse des globules rouges (EasySep® RBC Lysis Buffer 1X) est ensuite ajouté à l'échantillon pré-enrichi puis 100µl de solution de sélection positive (EasySep® Positive Selection Cocktail) sont ajoutés pour chaque mL du mélange précédent (échantillon/solution de lyse des globules rouges) suivi d'une incubation à température pièce durant 15 minutes. Ensuite, 50 µl de la solution de particules magnétiques (EasySep® Magnetic Nanoparticles) préalablement homogénéisée sont ajoutés pour chaque mL du mélange précédant (échantillon/solution de lyse des globules rouges) puis incubés à température ambiante pendant 10 minutes. L'échantillon total est ensuite ramené à un volume final de 5 mL (pour un échantillon de moins de 10⁸ cellules) ou 10 mL (pour 10⁸ à 8.5 x10⁸ cellules) et homogénéisé délicatement à la pipette. Le tube est ensuite placé dans l'aimant sans bouchon pendant 5 minutes. Au terme de cette incubation, le surnageant est éliminé dans un mouvement continu d'inversion de l'ensemble tube/aimant, et maintenu 2 secondes en position inversée, sans à-coup. Puis une fois l'ensemble tube/aimant repositionné à la verticale, le tube contenant les cellules marquées collées à sa paroi pouvait être retiré délicatement de l'aimant et rempli de nouveau au même volume que précédemment pour une 2^{ème} séance de tri magnétique de 5 minutes. La même procédure est

réalisée pour un total de 5 fois. Après la dernière sélection magnétique, le tube est rempli jusqu'à 15 ml de milieu tampon enrichi et le contenu délicatement ré-homogénéisé. Après prélèvement de 20 µl pour le décompte cellulaire, le reste de l'échantillon cellulaire est centrifugé à 300 x g pour 10 minutes. Après la centrifugation, le culot cellulaire était resuspendu dans le volume de milieu requis pour procéder aux étapes de contrôle morphologique et de cytométrie de flux (cf. Chap. II. 2.4). Le reste de l'échantillon cellulaire était ensuite resuspendu dans 800 ou 1000 µl de TRIzol® selon la quantité de cellules: moins d'un million ou plus d'un million de cellules respectivement. Le tube de mélange TRIzol-cellules immatures triées était ensuite conservé à -80°C pour utilisation ultérieure (cf. Chap. II.4).

2.5. Évaluation de l'homogénéité des populations cellulaires isolées

2.5.1. Morphologie des populations cellulaires

2.5.1.1. Cytospin

Un échantillon de 50 000 à 200 000 cellules selon le type de cellules était prélevé en fin de sélection des populations cellulaires (cf. Chap. II.2.2 et 2.3.2) et resuspendu dans le milieu tampon enrichi pour les cellules de sang de cordon ou du milieu de culture complet (cf. Chap. II.3.1) pour les cellules de lignée MONO-MAC-1 afin d'obtenir une concentration de 0,1 à 1 x 10⁶ cellules par mL. Un volume maximal de 500 µl était alors déposé dans la cupule de cytopspin montée sur filtre, lame et porte lame. Après centrifugation à 18 x g (soit 450 rpm) pendant 5 minutes à température pièce, les montages étaient défaits et les lames laissées au séchage à température pièce pendant 24 heures.

2.5.1.2. Coloration des lames au GIEMSA

Une fois sèche, les lames ont été colorées au GIEMSA selon la méthode suivante: 1) les lames étaient recouvertes de méthanol pendant 2 minutes; 2) après avoir inversé les lames pour enlever le méthanol, elles étaient recouvertes de Giemsa filtré; 3) puis le tampon à pH 6,8 était ajouté jusqu'à l'apparition d'une couche métallique sur la surface de toutes les

lames et laissées 5 minutes; 4) puis les lames étaient rincées à l'eau; 5) et ensuite séchées sur une plaque chauffante.

2.5.1.3. Observation des lames

Les lames ainsi colorées étaient ensuite observées au microscope pour évaluer la morphologie des cellules. Des photos de chaque lame ont été prises au grossissement 100X avec un Olympus DP20-E.

2.5.2. Analyse par cytométrie de flux (FACS) du contenu en CD34⁺

2.5.2.1. Marquage

Chaque échantillon de population cellulaire prélevé pour l'analyse par FACS était resuspendu dans un volume total de 1,5 mL de milieu tampon enrichi (PBS + 2% FBS + 1mM EDTA) et préservé à 4°C pour la nuit après le tri. Les quantités cellulaires utilisées étaient de 100000 (cellules CD34⁺) à 250 000 (pour les cellules contrôles) pour chaque population. Les tubes étaient ensuite centrifugés pendant 10 min à 130 x g (soit 1100 rpm) à température pièce puis décantés. À chaque échantillon était ajouté du PBS+2% FBS pour obtenir une concentration finale de $0,1 \times 10^6$ cellules/10 µl suivi de l'ajout de 10µl d'agent bloquant (gammaglobulines de souris). À chaque tube étaient ensuite ajouté 1 µl de la solution d'anticorps marqués anti-CD34-APC (ou de l'eau pour le contrôle négatif), afin d'obtenir la concentration déterminée préalablement par titration sur des cellules de sang de cordon similaires. Chaque échantillon était complété jusqu'à un volume final de 50 µl par ajout de PBS + 2% FBS puis délicatement resuspendu, et enfin incubé 20 minutes à 4°C à la noirceur. Après l'incubation, chaque tube était lavé par ajout de 100µl de PBS + 2% FBS et centrifugé à 1100 rpm pendant 5 minutes à température ambiante, puis décanté jusqu'à 50 µl. Après resuspension délicate, les échantillons étaient transférés dans des tubes de FACS et le volume complété pour obtenir un volume final de 200µl.

2.5.2.2. Récolte de données par cytométrie de flux

Les analyses de cytométrie de flux étaient réalisées dans l'heure qui suivait la fin du marquage. La récolte des données a été réalisée sur 20 000 à 100 000 événements selon la

population cellulaire à l'aide de l'appareil LSR FORTESSA (BD Biosciences) et les données analysées simultanément grâce au logiciel FACS DIVA. Pour chaque analyse, des cellules de la population cellulaire différenciée non marquées étaient utilisées comme cellules contrôles pour ajuster les paramètres de voltage.

2.5.2.3. Analyse des données

L'analyse des données récoltées par FACS étaient ensuite analysées grâce au logiciel FlowJo (© TreeStar Inc).

3. Source de blastes de leucémie aiguë myéloïde (LAM)

3.1. Lignée cellulaire humaine MONO-MAC-1

La lignée MONO-MAC-1 (MM1) est une lignée de cellules de LAM humaine de type monoblastique (LAM FAB M5) provenant du sang périphérique d'un patient de 64 ans en rechute (1984). La formule chromosomique de la lignée montre: 49(43~52)<2n>XY, invdup(3)(q21q27),+8,+der(8)add(8)(q24),t(9;11)(p22;q23),t(10;12;17)(q24;q13;q11),+12,r(12)(?),+13,del(13)(q1?2q21),t(16;21)(q13;q22.2),+18,-20 avec un réarrangement de *MLL* suite à la t(9;11)(p22;q23). Les cellules de MM1 étaient cryopréservées à environ 10×10^6 cellules/mL dans une solution de congélation composée de 70% de milieu RPMI 1640, 20% de FBS et 10% de DMSO (diméthyle sulfoxyde). Après décongélation rapide, les cellules étaient maintenues en culture en suspension à la concentration de 0,5 à 1×10^6 cellules/mL dans du milieu de culture RPMI 1640 enrichi avec 10% de FBS inactivé par la chaleur et 1ml d'un mélange de pénicilline et streptomycine pour un litre de milieu. Des échantillons de 10×10^6 cellules ont été prélevés après décompte de concentration cellulaire à l'hématimètre, centrifugés à 60 x g (soit 800 rpm) pendant 8 minutes à température pièce, puis resuspendus dans 1mL de TRIzol®. Les mélanges TRIzol et cellules de lignées étaient ensuite congelés à -80°C jusqu'à l'étape d'extraction de l'ARN.

3.2. Cellules leucémiques de LAM pédiatrique de la BCLQ (Banque de Cellules Leucémiques du Québec)

Trois échantillons de cellules leucémiques avec un pourcentage de blastes supérieur à 70% provenant de patients pédiatriques atteints de LAM avec t(9;11) ont été prélevés au moment du diagnostic. Les cellules mononucléées, obtenues après séparation par le Ficoll ont été cryopreservées (vials de TRIzol® et cellules congelées en DMSO dans l'azote liquide) à la Banque de Cellules Leucémiques du Québec (www.bclq.org) après analyses morphologique et cytogénétique. Les échantillons ont ensuite été décongelés puis utilisés pour l'extraction de l'ARN total en vue d'en réaliser le transcriptome par l'équipe du laboratoire du Dr Brian Wilhelm (cf. Chap.II.7).

4. Extraction de l'ARN total

4.1. Procédure d'extraction de l'ARN total

L'ARN total des différents échantillons cellulaires collectés dans du TRIzol® et conservés à -80°C, a été extrait selon la méthode du manufacturier (TRIzol®, Invitrogen, Burlington, ON, Canada) en trois étapes:

- phase de séparation après ajout de 200 µl de chloroforme
- phase de précipitation avec ajout 10 µg de glycogène comme molécule véhicule (« carrier ») (0,5 µl d'une solution à 20 µg/µl) pour favoriser la visualisation du précipité puis de 500 µl d'isopropanol.
- phase de lavage avec ajout de 1mL d'éthanol à 95% puis séchage à l'air.

En fin de procédure, les culots d'ARN étaient dissous dans 30 à 40 µl d'eau avec 0,1% de diéthyl pyrocarbonate (DEPC).

4.2. Procédure de contrôle de qualité de l'ARN

Pour chaque ARN, un échantillon de 1 à 2 µl était utilisé pour quantification par spectrophotométrie grâce au Nanodrop ® ND-1000 spectrophotometer (Nanodrop Technologies, Wilmington, USA). Un 2^{ème} échantillon de 1 µl était utilisé pour vérification

de l'intégrité de l'ARN par électrophorèse sur un gel composé de 1% d'agarose dans un tampon de migration TAE 1X (Tris/Acétate/EDTA) (figure 19).

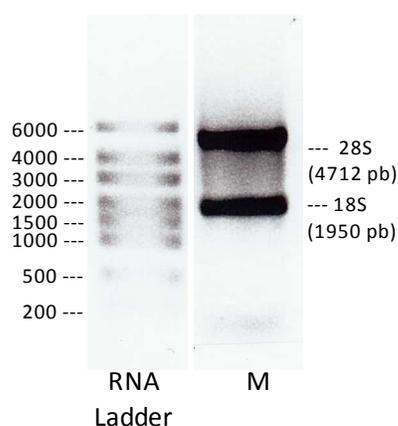


Figure 19. Exemple de résultats d'extraction d'ARN total

L'électrophorèse sur gel d'agarose (agarose 1% dans un tampon de migration TAE Tris/Acétate/EDTA 1X; migration 35 minutes à 90V) du résultat de l'extraction de l'ARN total d'un échantillon cellulaire de la lignée cellulaire Monomac-1 montre deux bandes d'ARN 28S à 4712 pb et d'ARN 18S à 1950 pb. RNA Ladder : échelle d'ARN « RiboRuler™ High Range RNA Ladder (Fermentas). M : échantillon d'ARN total des cellules Monomac-1; pb : paire de bases.

5. Transcription inverse

5.1. Décontamination préliminaire des ARN totaux à la DNase I

Afin d'éviter toute contamination par de l'ADN génomique, les échantillons d'ARN destinés à être utilisés pour l'étude d'expression par qRT-PCR ont été traités avec de la DNase I, RNase-free (Cat# EN0521, Fermentas, Thermo Fisher Scientific, Canada) selon les recommandations du fabricant en utilisant environ 5 µg d'ARN pour un volume total de 50 µl.

5.2. Réaction de transcription inverse (RT)

Nous avons utilisé un tube sans ARN (contenant de l'eau) comme contrôle pour exclure une contamination des réactifs par un des échantillons d'ARN. La réaction était réalisée grâce au kit de transcription inverse « High Capacity cDNA Reverse Transcription Kit (Cat#4368814, Applied Biosystems, California, USA) selon les recommandations du manufacturier. Il existe

3 types d'amorces utilisables pour réaliser une transcription inverse: les amorces poly-T qui permettent l'initiation de la transcription de tous les ARNm simultanément à partir de leur extrémité 3' poly-A, les amorces spécifiques d'un ARNm choisi pour ne produire que l'ADNc de cet ARNm, ou encore l'utilisation d'un mélange d'oligonucléotides, tels que dans le kit que nous avons utilisé, qui offre l'avantage de produire des fragments d'ADNc plus courts mais augmentent ainsi la probabilité de convertir les extrémités 5' des ARNm en ADNc. En effet la transcriptase inverse peut avoir de la difficulté à atteindre les extrémités 5' des ARNm les plus longs.

Nous avons utilisé 5 µg d'ARN total traités à la DNase I pour un volume final de 100 µl (tableau 2). Le programme de la machine PCR Veriti (96 wells thermal cycler, Applied Biosystem) que nous avons utilisé pour cette réaction de transcription inverse est résumé dans le tableau 2. Une fois la transcription inverse terminée, les échantillons étaient conservés à -20°C.

Réactifs transcription inverse	Concentration initiale	Quantité pour une réaction (µl)	
Tampon PCR 10X	10x	10	
mélange d'amorces 10X	10X	10	
dNTPs 100mM 25X	25X	4	
Eau PCR	-	23.5	
Transcriptase inverse MSRT	50u/µl	2.5	
ARN traité à la DNase	-	50 (5µg)	
Volume total (µl)	-	100	

Conditions	Étape 1:	Étape 2	Étape 3
Température(°C)	25	37	4
Durée (min)	10	120	∞

Tableau 2. Transcription inverse: réactifs et conditions

La partie supérieure du tableau présente les réactifs utilisés pour la réaction de transcription inverse; la partie inférieure du tableau présente les conditions des différentes étapes de cette réaction

5.3. Étape de contrôle de l'ADNc par RT-PCR du gène *KDM5B/JARID1B*

Afin de confirmer la présence d'ADNc dans nos échantillons et afin d'exclure une contamination par de l'ADNg avant de les utiliser pour la qRT-PCR, nous avons réalisé une PCR ciblant uniquement le gène *KDM5B/JARID1B* pour sa capacité à discriminer l'amplification de l'ADNc ou de l'ADNg. En effet, la séquence complémentaire de ce gène est de 147 pb contre 1048 pb pour la séquence génomique. Les séquences des amorces utilisées pour cette PCR sont indiquées dans le tableau 3. Un échantillon d'ADNc de contrôle connu pour être contaminé par de l'ADNg était utilisé comme ADN de contrôle positif. Les contrôles négatifs utilisés étaient un tube contenant de l'eau avec les réactifs de PCR *KDM5B/JARID1B* et l'échantillon d'eau utilisé comme contrôle négatif à l'étape de transcription inverse (cf. Chap.II.5.2). La préparation de la réaction utilisant 1 µl d'échantillon d'ADNc décongelé et l'enzyme ADN polymérase « Platinum® Taq DNA Polymerase » (Cat#10966-018, Invitrogen, Californie, USA) était réalisée selon les recommandations du fabricant et est résumée dans le tableau 3. L'amplification était réalisée grâce à l'appareil PCR Veriti (96 wells thermal cycler, Applied Biosystem) et le programme optimisé pour l'amplification de *KDM5B/JARID1B* est également résumé dans le tableau 3.

Amorces <i>KDM5B</i>	Séquence 5' --> 3'
Amorce sens "F2"	TGTCACAGTGGAATATGGAGCTGAC
Amorce antisens "R1"	CATCACTGGCATGTTGTTCAAATTC

Réactifs PCR <i>KDM5B</i>	Concentration initiale	Quantité (µl)
Tampon PCR	10X	3
dNTP mélange	10mM	1
MgCl ₂	25mM	1,8
Amorce sens "F2"	10µM	1
Amorce antisens "R1"	10µM	1
Eau PCR	-	20,7
Polymérase "taq PLATINUM »	50U/µl	0,5
ADNc	-	1
Volume total (µl)	-	30

Conditions PCR <i>KDM5B</i>	Étape 1	Étape 2.1	Étape 2.2	Étape 2.3	Étape 3	Étape 4
Température (°C)	95	95	60	72	72	4
Durée	5min	30sec	1min	1min	7min	∞
Répétition	1x		35x		1x	1x

Tableau 3. Réaction de PCR pour étudier le gène *KDM5B*: réactifs et conditions

Cette réaction de PCR constitue une étape de contrôle pré-qRT-PCR pour confirmer la présence d'ADNc dans les échantillons utilisés. La partie supérieure du tableau présente les amorces spécifiques utilisées pour étudier l'expression de *KDM5B*. La partie intermédiaire détaille les réactifs utilisés. La partie inférieure du tableau présente les conditions des différentes étapes de cette réaction de PCR.

Le résultat de l'amplification du gène *KDM5B/JARID1B* avec les échantillons d'ADN des différentes populations cellulaires étudiées était ensuite visualisé par électrophorèse sur gel d'agarose à 2% dans un tampon de migration TAE Tris/Acétate/EDTA 1X (migration 30 minutes à 100V). La figure 20 montre un exemple représentatif des résultats obtenus pour les différentes populations cellulaires testées.

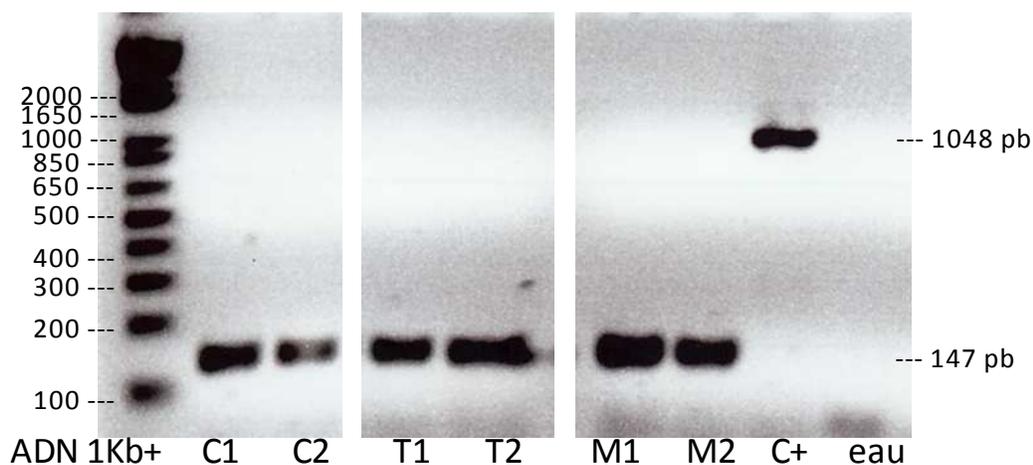


Figure 20. Exemple de résultats de l'amplification du gène *KDM5B/JARID1B*

L'électrophorèse sur gel d'agarose (agarose 2% dans un tampon de migration TAE Tris/Acétate/EDTA 1X; migration 30 minutes à 100V) du résultat de l'amplification du gène *KDM5B/JARID1B* dans différents échantillons d'ADN montre une bande à 147 pb lorsque de l'ADN complémentaire est présent, ou une bande à 1048 pb lorsque de l'ADN génomique est présent. Échelle d'ADN : 1Kb+ DNA Ladder (Life Technologies ®). Échantillons C1, C2 = populations cellulaires CD34⁺; échantillons T1, T2 = populations de cellules mononucléées totales, échantillons M1, M2 = populations cellulaires de la lignée Monomac-1; échantillon C+ = contrôle positif constitué d'une population cellulaire contenant de l'ADN génomique.

6. Méthode de quantification de l'expression des gènes étudiés par qRT-PCR

6.1. Choix des gènes à analyser

Les études de qRT-PCR ont été réalisées en deux temps tel que résumé dans le tableau 4. Pour la première étape, ou étape d'analyse globale des HDM, les gènes choisis en raison des limitations techniques (quantité d'ADNc et nombre de puits disponibles sur la plaque) sont ceux codant pour toutes les histones déméthylases des familles LSD et Jumonji C, ainsi que *PADI4* (soit 36 enzymes HDM), *GATA2* comme gène témoin exprimé différemment dans les deux populations cellulaires normales, et les 2 gènes endogènes *TBP* et *HPRT1*.

Pour la deuxième étape, ou étape d'analyse de confirmation, nous avons choisi de ré-analyser les gènes (annotés avec une étoile dans le tableau 4) dont l'expression lors de la première analyse, était fortement différenciée d'une population cellulaire à l'autre ou ceux localisés sur le chromosome Y (un des sangs de cordon de la 1^{ère} analyse était d'origine XX et n'exprimait donc pas les gènes localisés sur le Y), ainsi que les autres membres de la famille *PADI* pour lesquels un test adéquat était disponible, soit 14 HDM, ainsi que *GATA2* comme gène témoin et *TBP* comme gène endogène (cf. Chap. II.6.3).

6.2. Conception des tests de qRT-PCR et validation des tests amorces/sonde

La technologie qui a été choisie pour ce projet est celle des essais de qRT-PCR Taqman®. Les amorces et les sondes ont été spécifiquement sélectionnées sur le site UPL («Universal probe library » de Roche www.universalprobelibrary.com). Un à trois tests ont été choisis pour chaque gène. Les tests devaient répondre aux critères suivants: amorces de taille entre 60 et 125 pb, tests chevauchant si possible deux exons et capables de mettre en évidence l'expression de tous les transcrits de chaque gène lorsque possible. Ces critères étaient remplis pour la majorité des 41 gènes sauf pour sept: pour 2 gènes, le test utilisé ne chevauche pas deux exons (*KDM4E(DL)* et *C14orf169*) et pour 5 gènes, le test ne couvre qu'un ou que quelques-uns des transcrits les plus importants du gène (*KDM2A*, *KDM7B*, *KDM4C*, *KDM5D* et *UTY*). Chaque test sélectionné a ensuite été évalué pour être conforme aux paramètres d'efficacité et de reproductibilité technique établis par la compagnie Roche et

appliqués à la plateforme de génomique de l'IRIC en utilisant un mélange d'ADNc provenant de différents tissus humains.

Gène	NCBI ID	Amorces	Sonde	Test
KDM1A*	NM_001009999.2 NM_015013.3	gcatggtgtaacaggtct tggccagtccatatttacttg	76	IR3889
KDM1B	NM_153042.3	ccagcaactcccactctgg tggagtaagctggattccttg	16	IR3890
KDM2A	NM_012308.2	tggacaagaactcaagagcaga gaaattactgtaaagaaccactcca	8	IR3891
KDM2B	NM_001005366 .1 NM_032590.4	acctgctcactgctgttg tgatcagtgacctattgcagt	77	IR3892
KDM3A	NM_001146688.1 NM_018433.5	ttcaagaccccaacaataag tgatgcactccagacacat	71	IR3893.2
KDM3B	NM_016604.3	ccccggtcttctctacatc ggccaagaaagttggtaggc	18	IR3894
KDM3C	NM_032776 NM_004241	gaagattgaaacccaatggag actgaacgaggaatccagt	18	IR3904
KDM4A*	NM_014663.2	gccgctagaagttcagtgag gcgtcccttgacttctatt	53	IR2268
KDM4B*	NM_015015	tcaccagccacatctaccag gatgtcccacgcttcac	33	IR2398
KDM4C*	NM_015061.3	aaagtaagaaaagcatcccgaag ctcatcagccttaggcctttt	47	IR2399
KDM4D*	NM_018039.2	caacgactcctgcttccatc ttccgaagccaactctcta	2	IR2400.2
KDM4DL*	NM_001161630.1	aatctcctgggctcctgact tggaggaaaatacctaactca	3	IR4185.3
KDM5A*	NM_001042603	ggatttccgggaaggatg caaattccaaccagaaagtgc	35	IR3551
KDM5B*	NM_006618.3	ctgcagccagaaggagatg catctctggggagacacca	47	IR3534
KDM5C*	NM_004187.3 NM_001146702.1	cgaacgcattgttatcccta tggacgtgtgttactgacac	50	IR3896
KDM5D*	NM_004653.4 NM_001146706.1 NM_001146705.1	acgggcaaaaaggctacag agctctggatgcttctcaatg	82	IR3897
KDM6A	NM_021140	catgaacacagcacagcaga accatgaatgagcttgttct	65	IR3898
UTY*	NM_007125.4 NM_182659.1 NM_182660.1	atgccccagttcaacagaag tgggtatggtgaagttggtg	10	IR2441.3
KDM6B	NM_001080424	caccactgtggtctgttgt tgtctccgctcagtaacag	3	IR3899
KDM7A*	NM_030647.1	gaattacgctctcaggtctcc catgtttctcagatatcttgtgtc	71	IR3900
KDM7B*	NM_015107	agaagttgagttgacattgagga ccaagcttgcctcaacc	31	IR3263
KDM7C	NM_005392	ctgctggaggcattcaaag ttagctcctggactagatgagg	38	IR3906

(Tableau 4 (suite) : page suivante)

KDM7C	NM_005392	ctgctggaggcattcaaag ttagctcctggactagatgagg	38	IR3906
KDM8	NM_001145348.1 NM_024773.2	agcaccagctctttgaccag gccaggctgcagtagtc	89	IR3902
C14orf169	NM_024644.2	gagtcccaaccgaggaact cgaggctgctcctgactgaag atccgagccaggctttct	42	IR4063.2
FBXL19	NM_001099784.2	aggcggctttggttctc	87	IR3905
HIF1AN	NM_017902.2	ctgctcattggcatggaag agatgcatcgttgtaacctttt	84	IR3261.3
HR*	NM_018411 NM_005144	cagcatgatgagcagaaaagg ggcatggtagtctcctgaagtc	80	IR0284
HSPBAP1	NM_024610.5	ggcgaccactcctgtgat tctctggcttaaaggtttgaca	28	IR3908
JARID2	NM_004973.2	atgttcacaacggcatgt caggttccttcccgtgt	62	IR3907
JMJD4	NM_001161465.1 NM_023007.2	cacctactggaagagtacatacagg gggtgaaaaactcctcca	55	IR3901
JMJD6	NM_001081461.1 NM_015167.2	cctggaaatcttaaaaaaccag cgagattgaggacaacatgc	75	IR3903.2
JMJD7	NM_001114632.1	aagccctgggaaagatgc tcctgagaccacgcagtaga	20	IR4182
JMJD8	NM_001005920.2	cgccaacacctactctacc gaagaagtacagggtgtcattgc	10	IR4183.2
MINA	NM_001042533.1 NM_032778.4 NM_153182.2	tcagtttcaccaacctcagaga aaggagccaaagtaacattcca	2	IR4062.2
PLA2G4B	NM_001114633.1	gaaagctgacagctcctgatcc aggtcctggcacctctgc	15	IR4184.2
PADI1	NM_013358.2	tgaggagctgtatgtgacag ttgcagttggcttcaatgt	65	IR4368.3
PADI2	NM_007365.2	acccaaagaaggcatcctg tctcggtcacagttcaccag	57	IR4369
PADI4*	NM_012387.2	tggtgaactgtgacagagacaa tcagggtcatcagcgacat	29	IR3533
GATA2*	NM_001145661.1 NM_032638.4	ttgtgcaaatgtcagacgac tcattggctggtgctgta	77	IR4212
HPRT1	NM_000194.2	tgatagatccattcctatgactgtcaagacatt ctttccagttaaagttgaga	22	IR3657
TBP*	NM_003194.4	gaacatcatggatcagaacaaca atagggattccgggagtc	87	IR3658

Tableau 4. Liste des gènes testés et séquences des amorces spécifiques utilisées pour la qRT-PCR

Le tableau présente la séquence des amorces spécifiques sélectionnées pour ce projet à partir du site UPL (Exiquon Universal Probe Library), l'identification NM_ des transcrits des gènes testés, le numéro d'identification de la sonde spécifique correspondante ainsi que le numéro d'identification du test. Gènes annotés « * »: n=5 répliquats biologiques; pour les autres n=2 ou 3. Gènes annotés « ° »: tests qui ne couvrent qu'un seul exon. Lorsque plusieurs numéros « NM_ » sont présentés pour un même gène, le test couvre les transcrits correspondants.

Nous avons donc réalisé les courbes standards des tests sélectionnés en utilisant quatre dilutions différentes du mélange amorces/sonde de chaque test (entre 1:5 et 1:625). Le test retenu pour chaque gène avait une courbe standard dont la pente de courbe était située entre -3,1 et -3,6 (entre 90% à 110% d'efficacité), avec une amplification pour la première dilution qui apparaissait à un Ct entre 20 et 33, avec des répliquats techniques similaires (variation < 0,5 Ct) et un tracé de courbe régulier. Pour deux gènes (*PADI3*, *PADI6*) nous n'avons pas pu obtenir de test acceptable selon ces critères.

6.3. Choix des échantillons d'ADNc

Pour la première étape d'analyse globale des HDM (cf. Chap. II.6.1), nous avons utilisé deux répliquats biologiques d'ADNc de chaque population cellulaire (cellules immatures CD34⁺, cellules différenciées et cellules de la lignée MONO-MAC-1) pour lesquels toutes les étapes précédentes ont été réalisées de façon identique et en parallèle. Trois répliquats technologiques ont été réalisés pour chaque échantillon biologique. Pour la deuxième étape d'analyse de confirmation (cf. Chap. II.6.1), nous avons utilisé trois échantillons biologiques d'ADNc différents pour chaque population cellulaire (cellules immatures CD34⁺, cellules différenciées et cellules de la lignée MONO-MAC-1), issus de trois unités de sang de cordon différents pour lesquels toutes les étapes précédentes ont également été réalisées de façon identique et en parallèle. Nous avons également utilisé trois échantillons biologiques différents provenant des lignées cellulaires MONO-MAC-1. Chaque échantillon biologique a été testé en trois répliquats techniques.

6.4. Choix du contrôle endogène

Le gène contrôle *TBP* (ou « TATA box Binding Protein ») qui code pour un facteur de transcription essentiel à l'initiation de la transcription a été choisi pour nos analyses plutôt que le gène *HPRT1* qui code pour une phosphoribosyl-transférase, car le niveau d'expression de *TBP* (C_{tm} ≈26.5) est semblable aux gènes testés et était plus uniforme d'un échantillon cellulaire hématopoïétique à l'autre en comparaison à *HPRT1* lors de notre première analyse.

6.5. Choix de l'échantillon calibrateur

Pour chacune des deux analyses de qRT-PCR, l'échantillon calibrateur choisi était l'une des populations cellulaires différenciées.

6.6. PCR quantitative

Nous avons utilisé des plaques de 384 puits et un volume final de 10 μ l par réaction, ce qui représentait 3 à 4 plaques pour chacune des étapes (cf. Chap.II.6.1). L'ADNc utilisé était dilué à 1:5. Les mélanges de réactif utilisés pour chacune des réactions de qRT-PCR ont été préparés selon le protocole figurant dans le tableau 5 en utilisant le mélange « Taqman® Fast qPCR MasterMix » de Applied Biosystems. Les réactions de RT-PCR quantitative ont été réalisées avec l'appareil ABI Prism 7900HT selon le programme figurant dans le tableau 5. Chaque plaque contenait une quantification du niveau d'expression du gène endogène pour chaque échantillon biologique et un contrôle négatif (réactifs sans ADNc).

Réactifs qRT-PCR	Concentration initiale	Quantité pour une réaction (μ l)
qPCR Master Mix	2X	5
Amorces (mélange)	50uM	0.05
Sonde UPL	1 μ M (diluée 1/10)	1
ADNc (environ 5-25ng)	-	1.5
H2O	-	2.45
Volume total	-	10

Conditions qRT-PCR	Étape 1	Étape 2.1	Étape 2.2	Étape 3
Température (°C)	95	95	60	4
Durée	3min	5sec	30sec	∞
Répétition	1x	40x		1x

Tableau 5. Caractéristiques des réactions de qRT-PCR: réactifs et conditions

La partie supérieure du tableau présente le mélange de réactifs utilisés pour la réaction de qPCR. La partie inférieure du tableau présente les conditions des différentes étapes de cette réaction.

6.7. Méthode de calcul d'expression

L'analyse des résultats s'est faite en utilisant le logiciel SDS2.2.2. La première étape d'analyse consistait en la vérification des contrôles négatifs pour chaque plaque, de la qualité

des courbes (forme exponentielle), de la similitude de résultats pour les trois répliquats techniques (variation tolérée inférieure à 0,5 Ct) et de l'obtention d'une valeur des Ct inférieure à 35 pour les gènes testés.

Pour chaque plaque d'analyse de qRT-PCR, l'évaluation du niveau d'expression des gènes cibles par rapport à celle du gène endogène *TBP* était déterminée par le calcul du delta Ct selon la formule: Delta Ct (ΔCt) = $Ct_{\text{gène}} - Ct_{\text{endogène}}$. Le seuil de delta CT au-delà duquel un gène peut être considéré comme peu ou pas exprimé n'est pas formellement déterminé dans la littérature même si on retrouve des valeurs de delta Ct autour de 8 par exemple pour des gènes faiblement exprimés dans des tissus humains tumoraux[153]. Nous avons choisi de considérer que des valeurs de delta Ct supérieures à 6 mettaient en évidence des gènes peu exprimés par rapport au gène de référence *TBP* dans une population donnée. Le choix de ce « seuil » était conforté par la valeur de delta CT de plus de 7 dans chacun de nos échantillons pour le gène *Hr* connu pour être peu exprimé dans le tissu hématopoïétique (il constitue dès lors un contrôle négatif). Le contrôle positif interne était le gène *GATA2* qui lui était bien exprimé dans le tissu hématopoïétique pour chacun de nos échantillons normaux (delta Ct inférieur à 0)

La quantification de l'expression relative de chaque gène dans une population donnée par rapport à celle dans l'échantillon calibrateur choisi (qui est soit la population cellulaire différenciée pour les cellules immatures, soit la population immature pour les cellules de la lignée leucémique) a été calculée selon la formule dite des « $2^{-\Delta\Delta Ct}$ »[154]:

Quantification relative: $RQ = 2^{-\Delta\Delta Ct}$

où delta delta Ct est: $\Delta\Delta Ct = \Delta Ct_{\text{échantillon testé}} - \Delta Ct_{\text{calibrateur}}$

L'expression relative dans une population cellulaire A par rapport à l'échantillon calibrateur est considérée selon la valeur du RQ en échelle log2 de la façon suivante:

- $RQ(\log 2) > +1$: gènes plus exprimés dans la population cellulaire A
- $RQ(\log 2)$ entre +1 et -1 (et $RQ = 0$): gènes similairement exprimés dans les 2 populations
- $RQ(\log 2) < -1$: gènes moins exprimés dans la population A par rapport à l'échantillon calibrateur

A noter qu'un $RQ(\log 2) = 1$ correspond à une variation de valeur 2, soit par exemple un gène testé deux fois plus exprimé dans la population cellulaire A que dans l'échantillon calibrateur. Nous considérons donc un résultat comme significatif lorsque la variation de valeur (fold-change) est de minimum deux, c'est-à-dire un RQ de plus de 2 ou moins de 0,5 car cet intervalle comprend la variation liée à la technique.

7. Méthode d'analyse du profil d'expression par séquençage du transcriptome

Des populations cellulaires immatures et différenciées provenant de deux échantillons de sang de cordon différents (isolement cellulaire réalisé en parallèle et de façon identique à celui effectué pour la qRT-PCR, cf. Chap. II. 2.) et congelées dans du TRIzol® à -80°C ont été utilisées pour l'étude du transcriptome. L'analyse du transcriptome par RNA-seq a été effectuée en collaboration avec le laboratoire du Dr Brian Wilhelm, de même que ceux des échantillons de leucémies pédiatriques de la BCLQ. L'extraction d'ARN et le séquençage des transcriptomes ont été réalisés par la plateforme de génomique de l'IRIC en utilisant la technologie « Illumina HiSeq 2000 » (San Diego, CA, USA). La lecture, la normalisation et l'analyse des données ont été réalisées par le laboratoire du Dr Brian Wilhelm. Les séquences obtenues ont été cartographiées au génome de référence hg19 en utilisant le programme Elandv2. Environ 80% des séquences ont pu être cartographiées. Les degrés de couverture pour les différents échantillons figurent dans le tableau 6.

Échantillons	Lectures brutes	Lectures cartographiées
Patient A	121,647,892	93,435,753
Patient B	126,731,194	98,591,355
Patient C	117,874,334	89,359,944
CBLP21 – CD34	41,401,812	32,090,019
CBLP22 – CD34	46,174,492	36,206,866
CBLP21 – cellules totales	139,620,268	96,649,266
CBLP22 – cellules totales	173,350,630	125,130,798

Tableau 6 Séquençage de l'ARN des échantillons étudiés: profondeur de lecture

Le tableau présente le nombre de lecture pour chaque échantillon analysé et le nombre de lectures cartographiées

La quantification de l'expression des gènes en RPKM permet la comparaison entre les différents échantillons (cf Chap.1 5.4). Bien que des seuils d'expression ne sont pas formellement décrits à notre connaissance dans la littérature, il est habituellement considéré qu'un gène est faiblement exprimé dans un tissu humain lorsque sa valeur de RPKM est inférieure à 1 [155]. Certains auteurs augmentent même ce seuil à un RPKM de 10 et considèrent de plus qu'un RPKM < 0,2 n'est pas différenciable du bruit de fond lié à la technique[156]. Nous avons choisi de considérer pour l'interprétation des résultats que les gènes avec un RPKM inférieur à 1 sont peu exprimés. De plus dans la littérature il est bien rapporté que le cluster 5' des gènes *HOXA* est exprimé dans les cellules souches/progénitrices [29] Or, dans notre expérience impliquant une analyse par transcriptome de CSH de souris arborant un ARN interférant ciblant *Kdm5b* (accompagné d'une expansion de CSH *in vitro*) l'expression du cluster 5' des gènes *HOXA* variait entre 1 et 5 RPKM, leur expression ayant été également validée en qRT-PCR[90].

L'expression différentielle des gènes d'une population A par rapport à une population B est indiquée par le ratio des valeurs de RPKM dans chaque population (par exemple entre cellules immatures CD34⁺ et cellules différenciées) et interprétée de la façon suivante:

- ratio de RPKM (A/B) \geq 2: gène plus exprimé dans la population A que dans la population B
- ratio RPKM (A/B) = 1: gène exprimé de façon similaire dans les deux populations
- ratio de RPKM (A/B) < 0,5: gène moins exprimé dans la population A que dans la population B

8. Données provenant des banques de données publiques « TCGA »

Nous avons eu accès aux résultats de séquençage du transcriptome de LAM issues de la banque de données « TCGA », de différents types morphologiques et génétiques. Parmi les 200 LAM collectées dans la banque « TCGA », l'analyse du transcriptome a pu être réalisée dans 179 échantillons. Il s'agissait de LAM *de novo* provenant de patients de plus de 18 ans (âge moyen 55 ans +/- 16 SD) parmi lesquels figuraient 4 % de LAM avec anomalie de *MLL*. Ces échantillons ont été récoltés entre novembre 2001 et mars 2010. La méthode utilisée

pour le séquençage de l'ARN a été précédemment décrite en détails [156]. En particulier, le génome humain de référence utilisé pour l'alignement était le hg18/GRCh36 et le degré de couverture était de minimum 90%.

9. Analyses statistiques

Nous avons utilisé des statistiques descriptives pour décrire les caractéristiques des unités de sang de cordon et des populations isolées (médiane +/- écart-type SD). Le nombre restreint d'échantillons utilisés pour les analyses par qRT-PCR (n=5) et ceux utilisés pour l'analyse du transcriptome (n=2) ne permettait pas de faire de comparaison statistique de leurs caractéristiques. Comme les analyses en qRT-PCR et en séquençage de l'ARN comparaient des populations différentes et non des conditions différentes dans une même population cellulaire, les ratios et RQ calculés pour chaque gène ont servi à la comparaison entre les populations (cf. Chap. II.5.7).

10. Considérations éthiques

Les unités de sang de cordons ont également été collectées avec un formulaire de consentement approuvé par le comité d'éthique d'Héma-Québec. Les cellules humaines de LAM pédiatriques ont été collectées avec un formulaire de consentement approuvé par les comités d'éthique à la recherche (CER) du CHU Sainte-Justine et de l'hôpital Maisonneuve-Rosemont tout comme les politiques de banque de la BCLQ. Ce projet de recherche a également été approuvé par les CER du CHU Saint - Justine et de l'hôpital Maisonneuve-Rosemont.

11. Conception et réalisation du test diagnostique pour rechercher une mutation de *GATA2*

11.1. Détermination de la région à tester et conception du test

À notre connaissance, 36 anomalies génétiques différentes de *GATA2* associée à une maladie hématologique ont été décrites à ce jour dans la littérature chez 57 personnes (36 patients et 21 porteurs asymptomatiques) [157]. Ce sont le plus souvent des mutations ponctuelles et 11 de ces 36 anomalies décrites chez plus de la moitié des cas (22/57), sont localisées dans la séquence qui code pour le deuxième doigt de zinc de la protéine *GATA2*, domaine de liaison à l'ADN. Les autres anomalies génétiques identifiées se répartissent dans le reste du gène comme le montre la figure 21.

L'association entre génotype et phénotype est variable. Les anomalies de *GATA2* associées au syndrome de MonoMAC et décrites dans la littérature ne lui sont pas spécifiques, et se répartissent également sur tout le gène avec une prépondérance dans le 2^{ème} domaine en doigt de zinc.

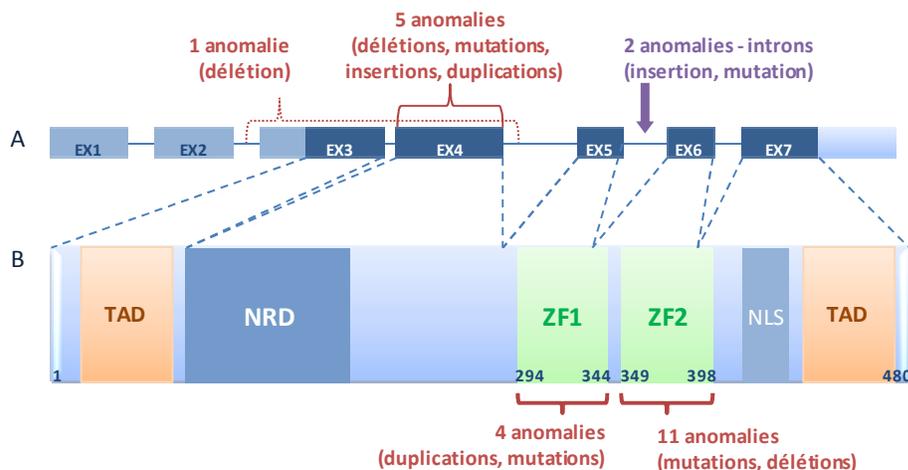


Figure 21. Anomalies génétiques décrites dans le gène *GATA2*

(A) Organisation génomique de *GATA2* avec 2 exons non codants (en 5', identifiés en bleu pâle) et 5 exons codants (identifiés en bleu foncé). (B) La protéine *GATA2* contient 2 domaines en doigt de zinc (ZF1 et ZF2), 2 domaines d'activation de la transcription (TAD), un domaine de régulation négative (NRD), un signal de localisation nucléaire (NLS). Le numéro des acides aminés correspondant au début et à la fin de la protéine et des deux domaines ZF figurent dans la partie inférieure du schéma. Les différentes anomalies

génétiques (mutations, délétions, insertions, duplications) sont indiquées en rouge. Les anomalies introniques de *GATA2* sont indiquées en mauve.

Nous avons conçu un test diagnostique pour rechercher ces anomalies par séquençage des exons codants de *GATA2* (figure 22) par amplification par PCR des séquences cibles réparties en 5 amplicons. Le test initial permet de couvrir la séquence codante de la majorité des régions décrites mutées, mais pas les régions introniques. Il était prévu de tester dans un deuxième temps l'exon 3 (en 5') dont la seule anomalie décrite est une délétion recouvrant les exons 3 et 4, et pouvait donc être suspectée pas le test décrit dans la figure 22. Au moment où nous avons conçu ce test, les anomalies introniques n'avaient pas encore été découvertes.

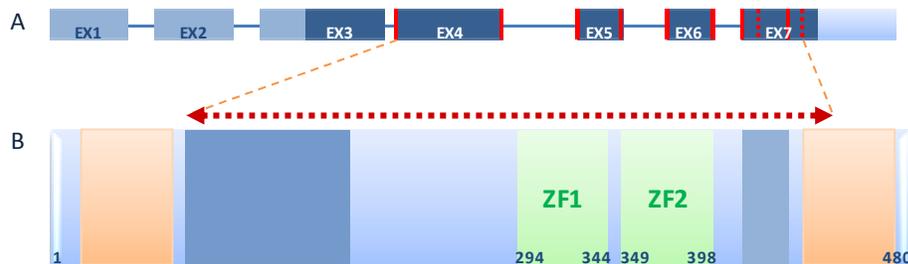


Figure 22. Conception du test diagnostique pour la détection des mutations de *GATA2*
(A) Représentation schématique du gène *GATA2* et de ses exons (bleu pâle = exons non codants, bleu foncé = exons codants). Localisation des paires d'amorces spécifiques des différents exons utilisées pour le test diagnostique (traits pleins et traits pointillés verticaux rouges). Les mutations introniques décrites entre l'exon 5 et 6 ne sont pas mises en évidence par ce test. (B) Représentation schématique de la protéine *GATA2* et des deux domaines zinc-finger ZF1 et ZF2. Les différentes paires d'amorces utilisées permettent d'étudier la région de *GATA2* où siège la majorité des mutations les plus fréquemment décrites (représentée par une flèche horizontale pointillée rouge).

11.2. Échantillons

Nous avons effectué l'étude des mutations à partir d'un échantillon de sang périphérique du patient, de son frère et d'un contrôle négatif, prélevés dans un tube contenant de l'EDTA. Nous avons utilisé 200µl de ces échantillons conservés à température pièce. Ces échantillons

permettent la détection de mutations somatiques. La présence de mutation au niveau germinale a été recherchée sur des échantillons de salive du patient et de son frère et analysés, dans un laboratoire indépendant en Australie (laboratoire du Dr Scott) [108].

11.3. PCR

L'ADN total de ces échantillons a été extrait grâce au kit d'extraction de l'ADN « QIAAmp DNA Blood minikit (50) » (Catalog #51104) selon les recommandations du fabricant, puis a été conservé à -20°C. Les séquences cibles du gène *GATA2* ont ensuite été amplifiées par PCR grâce aux amorces spécifiques et les réactifs présentés dans le tableau 7, L'amplification a été réalisée grâce à l'appareil PCR Veriti (96 wells thermal cycler, Applied Biosystem) et le programme optimisé pour l'amplification de *GATA2* est également résumé dans le tableau 6.

AMORCES <i>GATA2</i>	Séquence 5' -> 3'
Exon 4 sens	GTGAGCCCCTTCTCCAAGACG
Exon 4 antisens	ATAGTAGCTAGGCCTGGGCGCAG
Exon 5 sens	GTTAAGCAGGCCCCCGTGT
Exon 5 antisens	CCTGTAATTAACCGCCAGCTCCT
Exon 6 sens	TTTAGCCCTCCTTGACTGAGC
Exon 6 antisens	CAAGCCAAGCTGGATATTGTG
Exon 7 sens	AGGAAGGAACTGGCCCTCTGA
Exon 7 antisens	TGTCCGGAGTGGCTGAAGG
Exon 7bis sens	GGAGAAGTCATCCCCCTTCAGTG
Exon 7bis antisens	CGGTCCTCGACGTCCATCTGTT

RÉACTIFS PCR <i>GATA2</i>	Quantité (µl)
Buffer PCR 10X	2
dNTP mélange 25mM	0,5
MgCl ₂ 25mM	0,8
Eau PCR	12,45
TAQ Platinum	0,25
ADNc	3
Amorce sens 100ng/µl	0,5
Amorce antisens 100ng/µl	0,5
Total	20

CONDITIONS PCR <i>GATA2</i>	Étape 1	Étape 2.1	Étape 2.2	Étape 2.3	Étape 3	Étape 4
Température (°C)	95	95	56	72	72	4
Durée	10min	30sec	45sec	20sec	3min	∞
Répétition	1x		40x		1x	1x

Tableau 7. Réaction de PCR utilisée pour détecter les mutations du gène *GATA2* : séquences des amorces sélectionnées, réactifs et conditions

La partie supérieure du tableau détaille la séquence des amorces utilisées pour cette réaction. La partie intermédiaire du tableau présente les réactifs utilisés pour la réaction de PCR. La partie inférieure du tableau présente les conditions des différentes étapes de cette réaction.

Le résultat de l'amplification des séquences étudiées au niveau du gène *GATA2* pour les différents échantillons d'ADN était visualisé par électrophorèse sur gel d'agarose à 2% dans un tampon de migration TAE Tris/Acétate/EDTA 1X (migration 30 minutes à 100V) (figure 23).

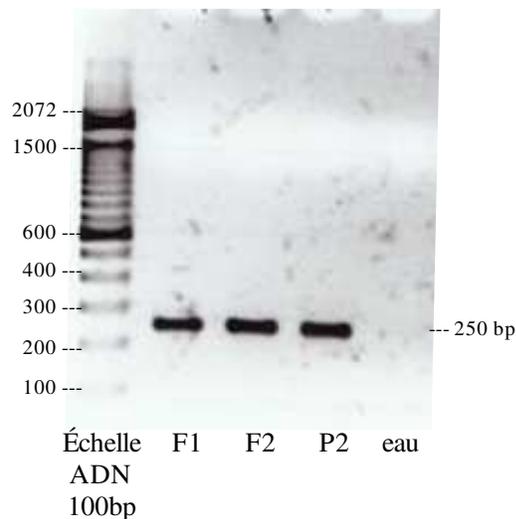


Figure 23. Exemple de résultat de l'amplification d'un des exons codant du gène *GATA2*

L'électrophorèse sur gel d'agarose (agarose 2% dans un tampon de migration TAE Tris/Acétate/EDTA 1X; migration 30 minutes à 100V) du résultat de l'amplification de l'exon 6 du gène *GATA2* dans différents échantillons d'ADN montre une bande à 250 pb correspondant au résultat attendu de l'amplification par PCR spécifique de cet exon. Échelle d'ADN : 100 bp DNA Ladder (Life Technologies ®). Échantillons : F1= frère répliquat 1 F2=frère répliquat 2, P2= patient répliquat 2.

Les produits de PCR ont ensuite été envoyés pour séquençage à la plateforme de génomique de l'IRIC sur le séquenceur 3730 ABI (Applied Biosystems, Warrington, Cheshire, UK) par

la méthode de Sanger. Chaque test a été réalisé pour chaque patient en deux répliquats biologiques indépendants et la séquence anormale a été contrôlée par le séquençage du brin complémentaire.

CHAPITRE III: RÉSULTATS

SECTION 1: ÉTUDE D'EXPRESSION DES HISTONES DÉMÉTHYLASES DANS LES CELLULES HÉMATOPOÏÉTIQUES NORMALES ET LEUCÉMIQUES

1. Profil d'expression des histones déméthylases dans les cellules hématopoïétiques normales

1.1. Isolement de deux populations cellulaires (population immature CD34⁺ et cellules différenciées) à partir d'unités de sang de cordon

1.1.1. Caractéristiques générales des populations cellulaires isolées

Le tableau 8 résume les caractéristiques des populations cellulaires étudiées qui ont été isolées à partir d'unités de sang de cordon. L'âge médian de ces unités était de 26,6h (+/- 3,4 SD). À partir de chaque unité, une petite fraction était utilisée pour isoler une population de cellules différenciées d'environ 8×10^6 cellules (médiane +/- 2,7 SD), et le reste de chaque unité a été utilisé pour récolter des populations cellulaires enrichies en cellules CD34⁺ d'environ $1,5 \times 10^6$ cellules (médiane +/- 1,1 SD).

Les deux unités de sang de cordon utilisées pour l'analyse du transcriptome par SNG étaient comparables en termes d'âge et de volume de l'unité, ainsi que de contenu cellulaire (cellules totales et CD34⁺) à celles utilisées pour l'analyse par qRT-PCR.

Unités de sang de cordon	Âge (h)	Sexe	Volume total (mL)	Population de cellules mononucléées totales ($\times 10^6$)	Population cellulaire enrichie en CD34 ⁺ ($\times 10^6$)
CBLP19	28	F	100	12,5	0,5
CBLP20	26,2	M	110	5	1,6
CBLP21*	23	F	80	11	0,9
CBLP22*	26,6	M	80	8	3,52
CBLP24	24,3	F	80	11	3
CBLP26	33,5	M	75	7,5	1,5
CBLP28	27,3	M	85	7	1

Tableau 8. Caractéristiques des unités de sang de cordon et des populations cellulaires isolées

Les échantillons marqués d'une étoile sont ceux utilisés pour l'analyse par séquençage de l'ARN par séquençage de nouvelle génération. Les autres ont été utilisés pour l'analyse par qRT-PCR.

1.1.2. Les populations cellulaires immatures isolées sont morphologiquement homogènes et enrichies en cellules CD34⁺

La figure 24 présente la morphologie des deux types de populations cellulaires récoltées et préparées sur lame par cytopspin et coloration au Giemsa (coloration bleu-mauve des constituants basophiles tels que les noyaux et certaines granulations des granulocytes, coloration rosée des globules rouges). On note que la population de cellules différenciées (A) qui constitue un ensemble très hétérogène de cellules des différentes lignées hématopoïétiques avec peu de cellules immatures et présence de neutrophiles au noyau plurilobé par exemple, annotés « N » dans la figure 24 - A.)

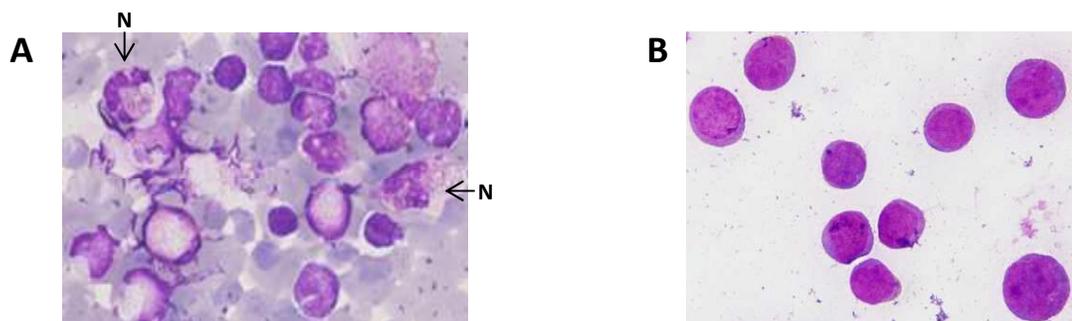


Figure 24. Populations cellulaires hématopoïétiques normales isolées à partir d'unités de sang de cordon

Morphologie sur lames de cytopspin colorées au GIEMSA. A. Cellules différenciées B. Cellules immatures enrichies en CD34⁺. Les images sont représentatives des différents échantillons observés pour chacune des unités de sang de cordon analysées (photo Olympus DP20-E, grossissement 100X). N : neutrophiles

La population enrichie en CD34⁺ montre des cellules morphologiquement homogènes et immatures i.e. possédant un rapport nucléo-cytoplasmique élevé et une chromatine peu condensée, caractéristiques des cellules souches et progénitrices hématopoïétiques (cf. figure 24-B).

L'analyse du contenu en CD34⁺ de ces deux types de populations cellulaires par cytométrie de flux confirme l'observation faite en morphologie (figure 25). On retrouve entre 1 et 2% de cellules CD34⁺ dans les populations cellulaires différenciées, alors que les populations cellulaires immatures sont composées de plus de 90% de cellules CD34⁺ (tableau 9).

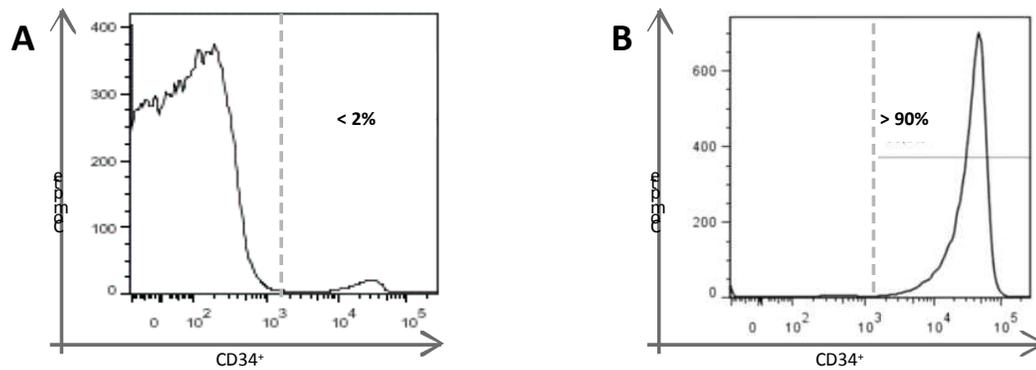


Figure 25. Contenu en CD34⁺ des populations cellulaires récoltées par cytométrie de flux

Profil d'analyse du contenu en CD34⁺ par cytométrie de flux. A. Population cellulaire différenciée. B. Cellules immatures CD34⁺. Les profils de FACS présentés sont représentatifs des différents échantillons observés pour chacune des unités de sang de cordon analysées.

L'analyse du contenu en CD34⁺ de ces deux types de populations cellulaires par cytométrie de flux confirme l'observation faite en morphologie (figure 25). On retrouve entre 1 et 2% de cellules CD34⁺ dans les populations cellulaires différenciées, alors que les populations cellulaires immatures sont composées de plus de 90% de cellules CD34⁺ (tableau 9).

Unités de sang de cordon	Population de cellules mononucléées totales		Population cellulaire enrichie en CD34 ⁺	
	Quantité cellulaire (x10 ⁶)	% CD34 ⁺	Quantité cellulaire (x10 ⁶)	% CD34 ⁺
CBLP19	12,5	1,2	0,5	97
CBLP20	5	2	1,6	93
CBLP21*	11	1	0,9	97
CBLP22*	8	1,3	3,52	95
CBLP24	11	1,7	3	93
CBLP26	7,5	1	1,5	90
CBLP28	7	1,7	1	98

Tableau 9. Contenu en CD34⁺ des populations cellulaires récoltées

Analyse des différents échantillons cellulaires isolés analysés par cytométrie de flux. Les échantillons marqués d'une étoile sont ceux utilisés pour l'analyse par séquençage de l'ARN par séquençage de nouvelle génération. Les autres ont été utilisés pour l'analyse par qRT-PCR.

1.2. Les histones déméthylases sont exprimées dans la population cellulaire différenciée

Les résultats des analyses d'expression des transcrits dans la population cellulaire différenciée de sang de cordon par qRT-PCR et par séquençage de l'ARN sont présentés dans la figure 26 A et B respectivement. Les histones déméthylases sont majoritairement exprimées dans ce compartiment cellulaire si on considère comme seuil d'expression selon la technique: $dCt < 6$ et $RPKM > 1$ (tels que définis aux Chap II. 5 et 6). De plus, on note une bonne corrélation entre les deux techniques qui montrent que les gènes *PADII*, *HR* et *KDM4E (DL)* ne sont pas exprimés (*PADII* n'est pas détectable en qRT-PCR et ne figure donc pas dans l'histogramme A). Cependant, en RPKM avec une couverture de lecture adéquate, *KDM4D* semble ne pas être exprimé alors qu'on mesure son expression avec l'analyse en qRT-PCR. Inversement, *PLA2G4B* n'est pas détectable en qRT-PCR mais l'est par séquençage de l'ARN. À titre comparatif, l'expression du gène *GATA2*, qui est bien exprimé dans les tissus hématopoïétiques, était respectivement quantifiée en moyenne à $-1 \text{ deltaCt (+/- 0,4 SD)}$ et à 7,8 en RPKM (+/- 5 SD) pour l'analyse en qRT-PCR et en séquençage de l'ARN respectivement. Par ailleurs, pour cette dernière analyse, les gènes spécifiquement exprimés dans le compartiment différencié (cf Chap.I 1.4.2.1.) étaient bien exprimés : RPKM moyen à 29,4 pour *GATA1* et RPKM moyen à 45,9 pour *SPII/PU.1* de même que le gène *EIF4H*, facteur d'initiation de la traduction, qui a été récemment identifié

comme un gène endogène particulièrement stable dans les différents tissus hématopoïétiques normaux et leucémiques [158] avec un RPKM moyen à 60.

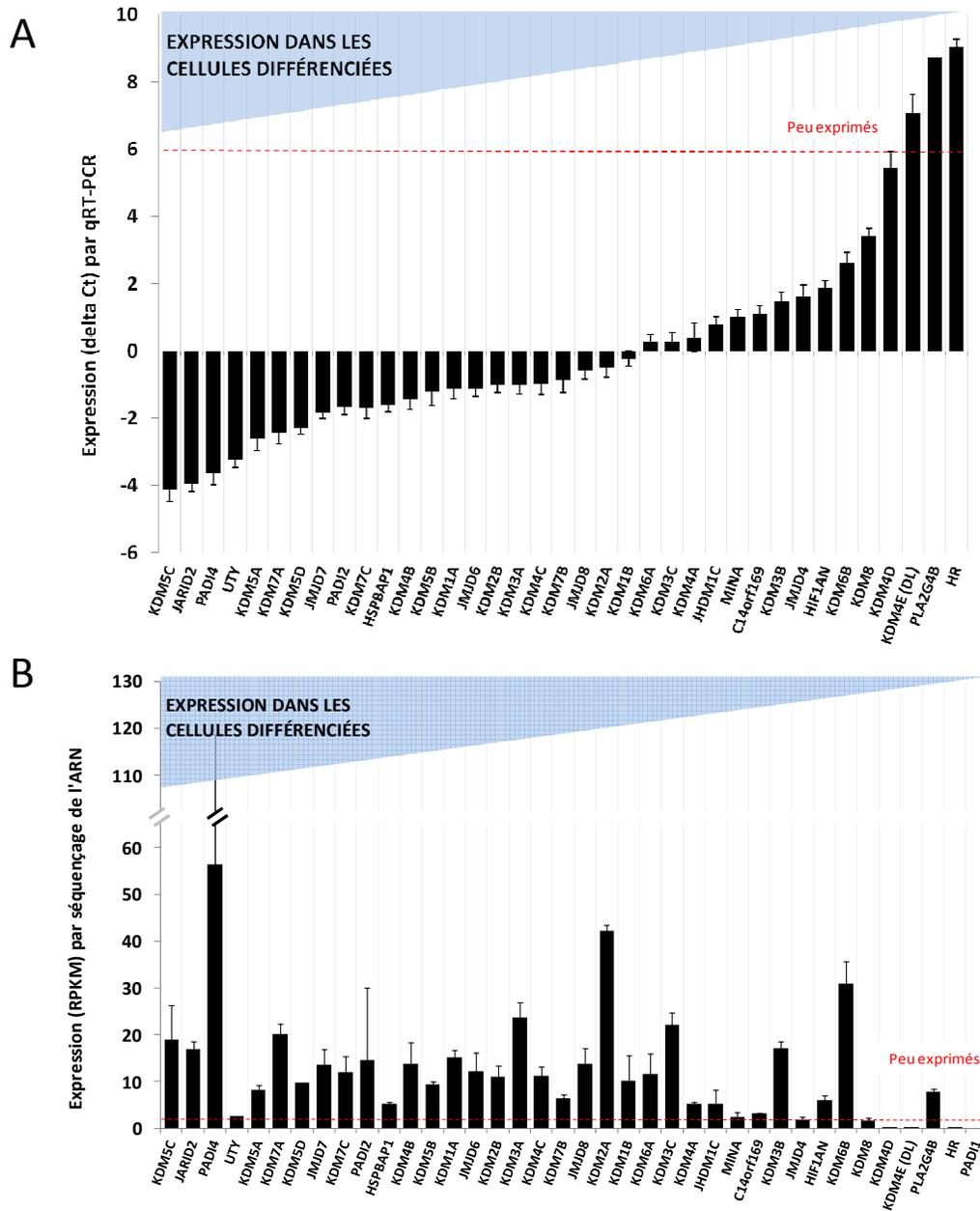


Figure 26. Profil d'expression des histones déméthylases dans le compartiment cellulaire différencié (sang de cordon)

Analyse de la quantité de transcrits des HDM dans la population cellulaire différenciée. A. Les résultats montrent les delta Ct déterminés par qRT-PCR par rapport à l'expression du gène contrôle *TBP* (Ct ~ 26.5) et représentent la moyenne \pm SD de n = 2 à 5 échantillons

cellulaires indépendants. Seuil d'expression: $dCt < 6$. B. Les résultats montrent les RPKM moyens \pm SD déterminés par séquençage du transcriptome sur $n = 2$ échantillons cellulaires indépendants. Seuil d'expression: $RPKM > 1$. Les barres d'erreurs indiquent les SD. HDM: histones déméthylases; SD: déviation standard; RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

1.3. Les histones déméthylases sont exprimées dans la population cellulaire immature CD34⁺

Les analyses par qRT-PCR et par séquençage de l'ARN des transcrits exprimés dans la population cellulaire immature CD34⁺ sont présentées dans la figure 27 et nous montrent que la majorité des histones déméthylases sont bien exprimées, si on considère comme seuil de moindre expression respectivement: $dCt < 6$ et $RPKM > 1$ (tels que définis aux Chap II. 5 et 6).. Avec les deux techniques, les gènes non exprimés sont *PADII* et *HR* (*PADII* et *PLA2G4B* ne sont pas détectables en qRT-PCR et ne figurent donc pas dans l'histogramme A). Le séquençage de l'ARN permet cependant de détecter *PLA2G4B*, mais ne permet pas de détecter les gènes *KDM4E(DL)*, *KDM4D* et *PADI2* qui sont toutefois bien détectés par qRT-PCR. Nous constatons ici pour ces 3 derniers gènes une possible différence de sensibilité de détection entre les deux techniques d'évaluation, avec une plus grande sensibilité pour la qRT-PCR pour ces gènes très faiblement exprimés. Compte tenu de cette variation de mesure pour ces gènes faiblement exprimés et/ou non exprimés dans une des populations cellulaires, nous les considérerons à part (cf. Chap. III. section I. 3). De nouveau, à titre de comparatif l'expression du gène *GATA2*, qui est bien exprimé dans les tissus hématopoïétiques et particulièrement dans le compartiment immature, était quantifiée en moyenne à $-4,2$ deltaCt ($\pm 0,3$ SD) et à 87,2 en RPKM (± 16 SD) pour l'analyse en qRT-PCR et en séquençage de l'ARN respectivement. Par ailleurs, pour cette dernière analyse, les gènes spécifiquement exprimés dans le compartiment immature (cf Chap.I 1.4.2.1.) étaient aussi bien exprimés : par exemple RPKM moyen à 53,2 pour *RUNX1/AML1* et RPKM moyen à 33,6 pour le gène *ERG*. De plus, le gène *EIF4H* était aussi bien exprimé avec un RPKM moyen de 123,9.

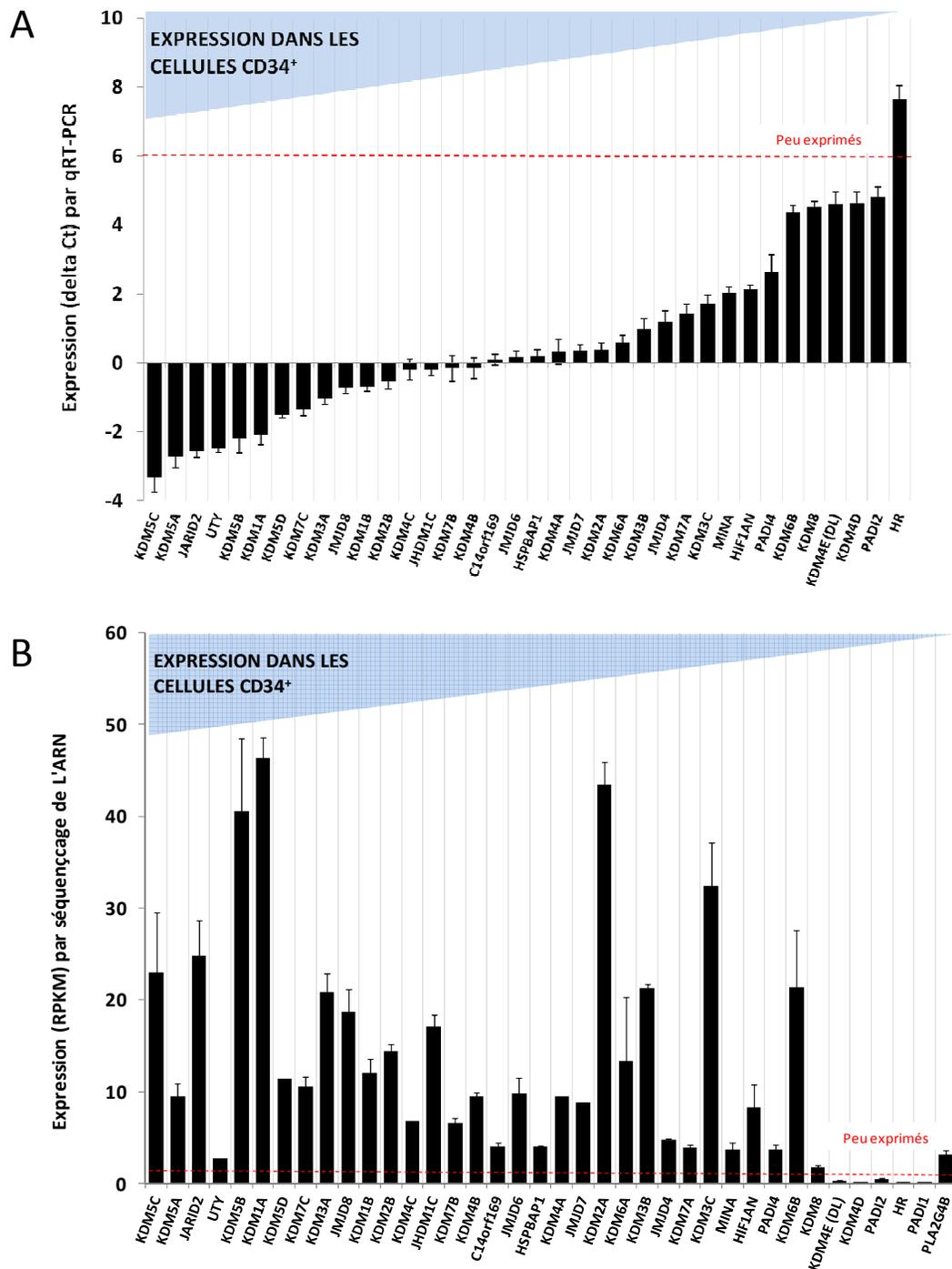


Figure 27. Profil d'expression des histones déméthylases dans le compartiment cellulaire immature CD34⁺ (sang de cordon)

Analyse de la quantité de transcrits des HDM dans la population cellulaire immature CD34⁺. A. Les résultats montrent les deltas Ct déterminés par qRT-PCR par rapport à l'expression de du gène contrôle *TBP* (Ct ~ 26.5) et représentent la moyenne \pm SD de n = 2 à 5 échantillons cellulaires indépendants. Seuil d'expression: dCt < 6. B. Les résultats montrent les RPKM

moyens \pm SD déterminés par séquençage du transcriptome sur $n = 2$ échantillons cellulaires isolés indépendamment. Seuil d'expression: RPKM > 1 . Les barres d'erreurs indiquent les SD. HDM: histones déméthylases; SD: déviation standard; RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

1.4. Certaines histones déméthylases sont différenciellement exprimées dans le compartiment cellulaire immature CD34⁺ par rapport au compartiment cellulaire différencié

Nous avons ensuite analysé les niveaux d'expression relative des HDM dans le compartiment cellulaire immature CD34⁺ par rapport au compartiment cellulaire plus différencié pour mettre en évidence les variations d'expression qui pourraient suivre la maturation cellulaire. Cette différence d'expression a d'abord été évaluée par qRT-PCR par le calcul du RQ et est représentée en log₂ dans la figure 28-A. Le RQ est déterminé à partir du delta Ct des cellules CD34⁺ par rapport au delta Ct des cellules différenciées (selon la formule du RQ présentée précédemment, cf. chap. II.5.7). Le seuil d'expression différentielle est considéré à RQ (log₂) ≥ 1 ou ≤ -1 ce qui correspond à une expression deux fois plus ou deux fois moins importante dans les cellules immatures par rapport aux cellules différenciées. La variation d'expression a ensuite été évaluée par séquençage du transcriptome et est déterminée par le ratio des RPKM (cellules CD34⁺) / RPKM (cellules différenciées) représenté dans la figure 28-B (cf. Chap. II.6). Le seuil d'expression différentielle est considéré significatif lorsque le ratio est ≥ 2 ou $\leq 0,5$, ce qui correspond à une expression deux fois plus ou deux fois moins importante dans les cellules immatures par rapport aux cellules différenciées.

Certaines HDM ont une expression différentielle mise en évidence par les deux techniques. Il s'agit de *KDM5B*, *JHDM1C*, *KDM1A* qui sont respectivement deux fois plus exprimées chacune (qRT-PCR) ou 4,2 - 3,2 et 3 fois plus exprimées (par RNA-seq) dans le compartiment cellulaire immature CD34⁺. Inversement, certains gènes sont moins exprimés dans le compartiment cellulaire immature CD34⁺ par rapport au compartiment cellulaire différencié mais le seul qui le soit avec les deux méthodes d'analyses est *KDM7A* qui est 8 fois moins (qRT-PCR) et 5 fois moins exprimé dans les cellules CD34⁺ par rapport aux cellules différenciées.

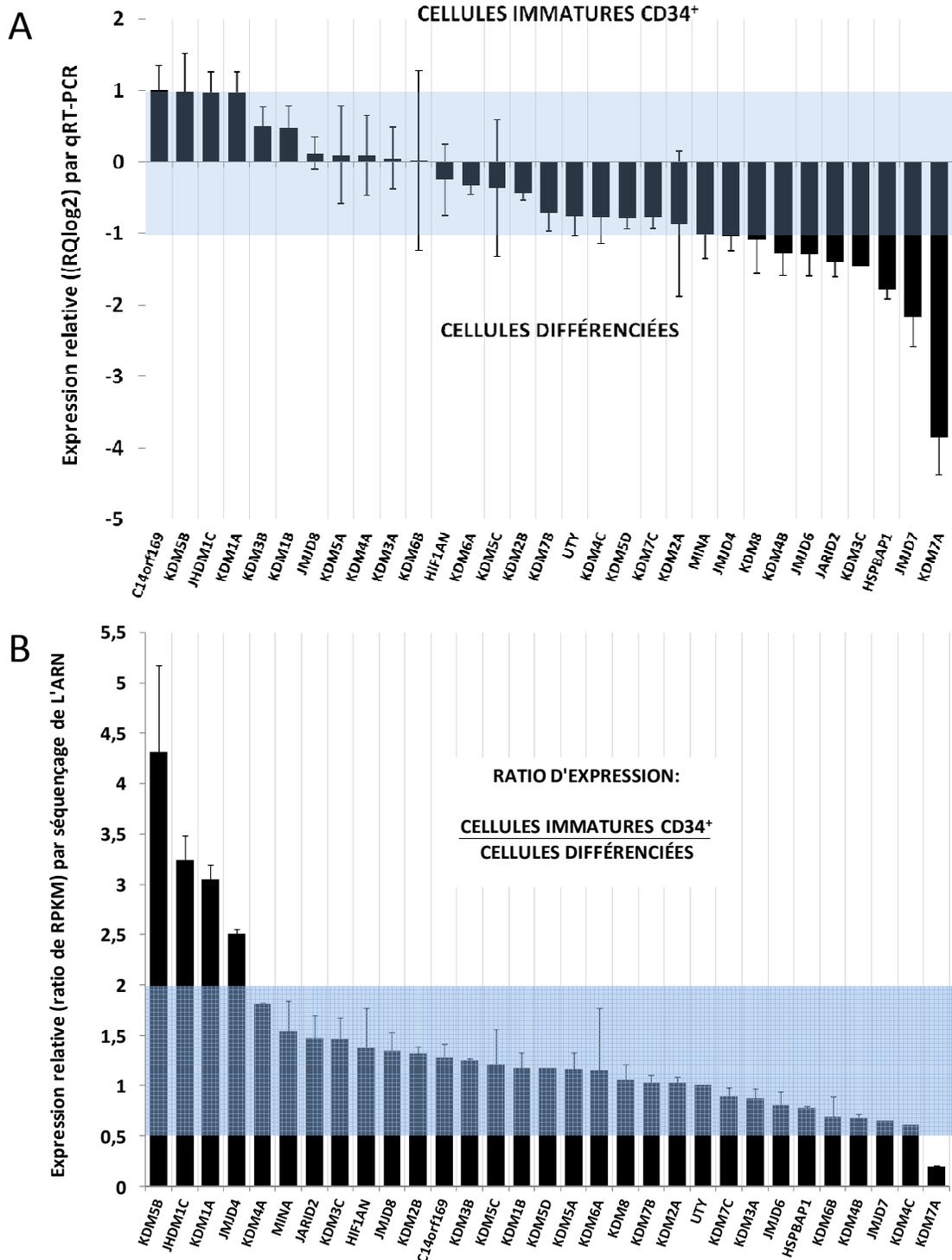


Figure 28. Profil d'expression relative des histones déméthylases dans les cellules immatures CD34⁺ par rapport aux cellules différenciées (sang de cordon)
 Comparaison du niveau de transcrits des HDM dans les cellules immatures CD34⁺ par rapport aux cellules différenciées. **A.** La quantité relative des transcrits (RQ) est montrée en

échelle log₂. Le RQ est déterminé à partir du delta Ct des cellules CD34⁺ par rapport au delta Ct des cellules différenciées déterminés par qRT-PCR (RQ(log₂) moyen ± SD sur n = 2 à 5 échantillons indépendants). Seuil d'expression différentielle: ≥1 ou ≤ 1 pour la qRT-PCR. **B.** La quantité relative des transcrits est représentée par le ratio des RPKM (cellules CD34⁺) / RPKM (cellules différenciées) détectés par séquençage du transcriptome (ratio moyen ± SD sur n = 2 échantillons indépendants). Seuil d'expression différentielle: ≥ 2 ou ≤ 0,5. Les barres d'erreurs indiquent les SD. HDM: histones déméthylases, SD: déviation standard; RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

Une expression nettement augmentée de l'expression de certains gènes dans la population cellulaire immature CD34⁺ par rapport à la population différenciée suggère un rôle biologique significatif de ces gènes au niveau des cellules souches et des cellules progénitrices. Inversement, les gènes dont l'expression est sensiblement moindre dans les cellules immatures CD34⁺ par rapport aux cellules différenciées pourraient avoir un rôle important dans la différenciation hématopoïétique.

2. Profil d'expression des histones déméthylases dans les cellules de leucémie aiguë

2.1. Sources de cellules de leucémie aiguë utilisées

2.1.1. Cellules blastiques de la lignée cellulaire MONO-MAC-1 utilisées pour l'analyse d'expression par qRT-PCR

L'analyse morphologique des cellules de la lignée cellulaire MONO-MAC-1 (LAM FAB-M5 avec translocation t(9;11)) préparées sur lame par cytopspin et coloration au Giemsa (figure 29 A) confirme l'homogénéité de cette population cellulaire et démontre une majorité de cellules blastiques, de grande taille, avec un rapport nucléo-cytoplasmique élevé et une chromatine hétérogène et déliée.

Ces cellules n'expriment pas l'antigène CD34⁺ comme le démontre l'analyse par cytométrie de flux (figure 29 B).

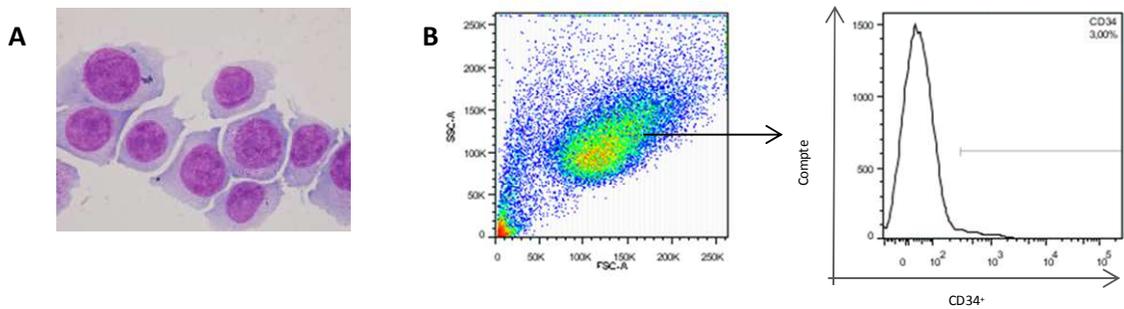


Figure 29. Cellules leucémiques de la lignée MONO-MAC-1

A. Morphologie des blastes évaluée sur lame de cytopspin après coloration au GIEMSA (photo Olympus DP20-E, grossissement 100X). B. Profil de FACS SSC-A/FSC-A et évaluation du contenu en CD34⁺. Les images sont représentatives des différents échantillons analysés de la lignée MONO-MAC-1.

2.1.2. Échantillons de leucémie aigüe myéloïde pédiatrique (LAM) utilisés pour l'analyse d'expression par séquençage de l'ARN

Les trois échantillons de leucémie de patients pédiatriques ont été prélevés au diagnostic et démontraient une LAM de type FAB-M5 avec un envahissement massif par des monoblastes. Le caryotype a confirmé un réarrangement du gène *MLL* avec présence de la translocation t(9;11)(p22;q23) (figure 30).

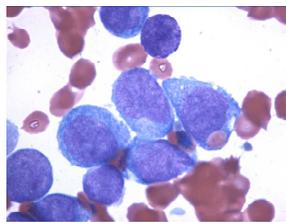


Figure 30. Échantillon de leucémie aigüe myéloïde pédiatrique

Aspect morphologique d'un des échantillons de leucémie aigüe myéloïde LAM de type FAB-M5 (frottis médullaire du patient B représentatif des deux autres patients A et C). Coloration au GIEMSA, grossissement 100X.

Le tableau 10 décrit les caractéristiques initiales de ces trois leucémies pédiatriques. Le premier patient (A) présentait une hyperleucocytose à 135×10^9 globules blancs/L et les deux autres présentaient une infiltration médullaire importante. Leur immunophénotypage, bien que sensiblement hétérogène, démontrait l'absence d'expression de l'antigène CD34

dans les 3 cas, réduisant la possibilité d'enrichir les populations cellulaires prélevées en blastes par l'intermédiaire de cet antigène.

Patient		A	B	C
Classification FAB		LAM-M5A	LAM-M5A	LAM-M5A
Réarrangement MLL		t(9;11) (p22;q23)	t(9;11) (p22;q23)	t(9;11) (p22;q23)
Formule sanguine initiale	GB (x10 ⁹ /L)	135	53,5	0,9
	Blastes (x10 ⁹ /L)	85%	45%	NA
	Hb (g/L)	69	92	65
	Plaq(x10 ⁹ /L)	80	29	116
Infiltration SNC		NA	+	-
Infiltration médullaire		NA	70%	90%
Immunophénotype	CD34 ⁺	-	-	-
	CD45 ⁺	+++	+++	-
	HLA-DR	+++	+++	+++
	CD33 ⁺	+++	+++	+++
	CD14 ⁺	+++	+++	++
	CD15 ⁺	+++	NA	+++
	CD7 ⁺	-	+	+++
	CD10 ⁺	-	-	-
	CD11b ⁺	+++	+++	++
	CD11b ⁺ /HLA-DR	+++	NA	-
	CD19 ⁺	-	-	-
	CD36 ⁺	+++	+	NA
	CD79 ⁺	-	-	NA
	CD3 ⁺	-	-	-
	CD56 ⁺	++	-	++
	CD117 ⁺	+	-	++
	CD13 ⁺	++	+	-
	CD14 ⁺ /CD56 ⁺	++	NA	NA
Caryotype patient A		48,XX,+8,+8,t(9;11)(p22;q23) [20]		
Caryotype patient B		46,XX, der(9)t(9;11)(p22;q23)del(9)(p21)[20]		
Caryotype patient C		46,XX,t(9;11)(p22;q23)[5] / 47,XX,+8,t(9;11)(p22;q23)[16]		

Tableau 10. Caractéristiques biologiques des leucémies aiguës myéloïdes pédiatriques sélectionnées pour analyse d'expression par séquençage de l'ARN

ARN: acide ribonucléique; FAB: French-American-British; LAM: leucémie aiguë myéloïde; GB: globules blancs; Hb: Hémoglobine; Plaq: plaquettes; SNC: système nerveux central.

2.2. Profil d'expression des histones déméthylases dans les cellules de leucémie aiguë

2.2.1. Les histones déméthylases sont exprimées dans les cellules leucémiques de la lignée MONO-MAC-1

Les résultats de l'analyse par qRT-PCR des niveaux d'expression des transcrits des HDM dans les cellules de la lignée MONO-MAC-1 exprimés par les deltas Ct moyen (\pm SD), sont présentés dans la figure 31. Si on considère comme seuil d'expression un $dCt < 6$ tel que définis aux Chap II. 5, la majorité des HDM sont bien exprimées dans ces cellules leucémiques, à l'exception de *KDM6B*, *PADI4*, *HR*, *KDM4E* et *PADI1*. À titre de comparatif l'expression du gène *GATA2*, qui est bien exprimé dans les tissus hématopoïétiques, était quantifiée en moyenne à 1,2 deltaCt (\pm 2,4 SD; non montré dans la figure).

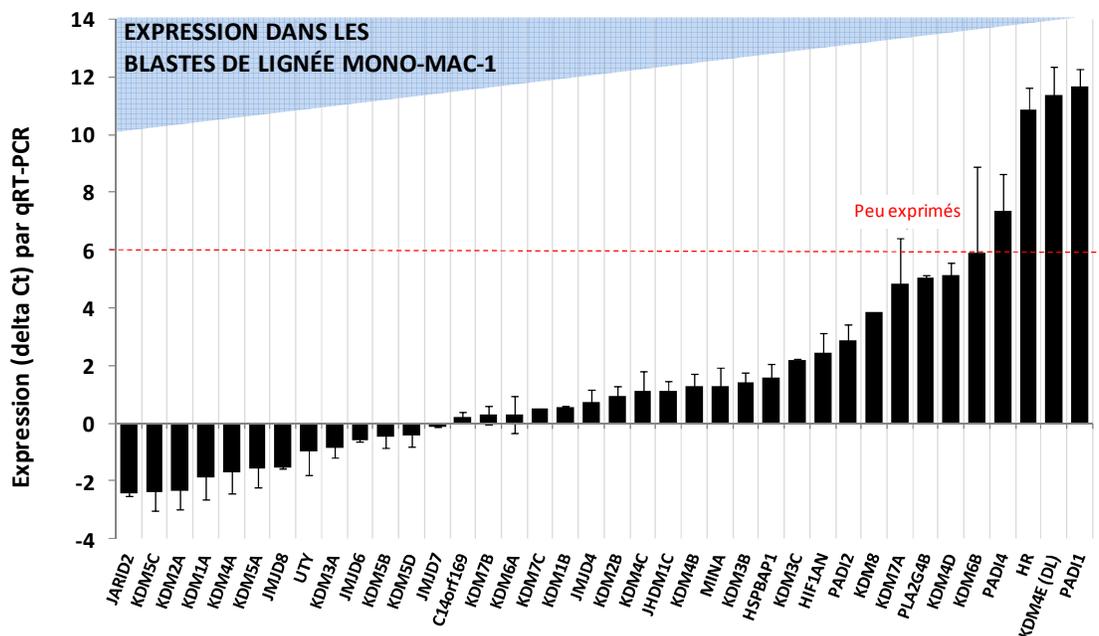


Figure 31. Profil d'expression des histones déméthylases dans les blastes de la lignée MONO-MAC-1 par qRT-PCR

Analyse de la quantité de transcrits des HDM dans la population de blastes de la lignée MONO-MAC-1. Les résultats montrent les deltas Ct moyens \pm SD déterminés par qRT-PCR par rapport à l'expression du gène contrôle *TBP* (Ct \sim 26.5) sur $n = 2$ à 5 échantillons cellulaires indépendants. Seuil d'expression: $dCt < 6$. Les barres d'erreurs indiquent les SD. HDM: histones déméthylases; SD: déviation standard.

2.2.2. Les histones déméthylases sont exprimées dans les blastes de leucémie aiguë myéloïde pédiatrique

Les résultats de l'expression des HDM des blastes de LAM pédiatriques par séquençage de l'ARN sont exprimés en RPKM et présentés dans la figure 32. Si on considère comme seuil d'expression un RPKM >1, les HDM sont majoritairement bien exprimées dans les blastes de LAM pédiatrique, à l'exception de *KDM4D*, *KDM4E(DL)*, *PADI1* et *HR*. Les gènes *KDM5D* et *UTY* ne sont pas exprimés car ils sont localisés sur le chromosome Y et les trois échantillons proviennent de patientes. Le gène *GATA2*, pour référence, est exprimé avec une grande variabilité dans les échantillons avec un RPKM moyen de 12 (+/-19 SD). Pour ce qui est du gène *EIFH4*, il est également bien exprimé dans ces échantillons de LAM pédiatriques avec un RPKM de 93,5.

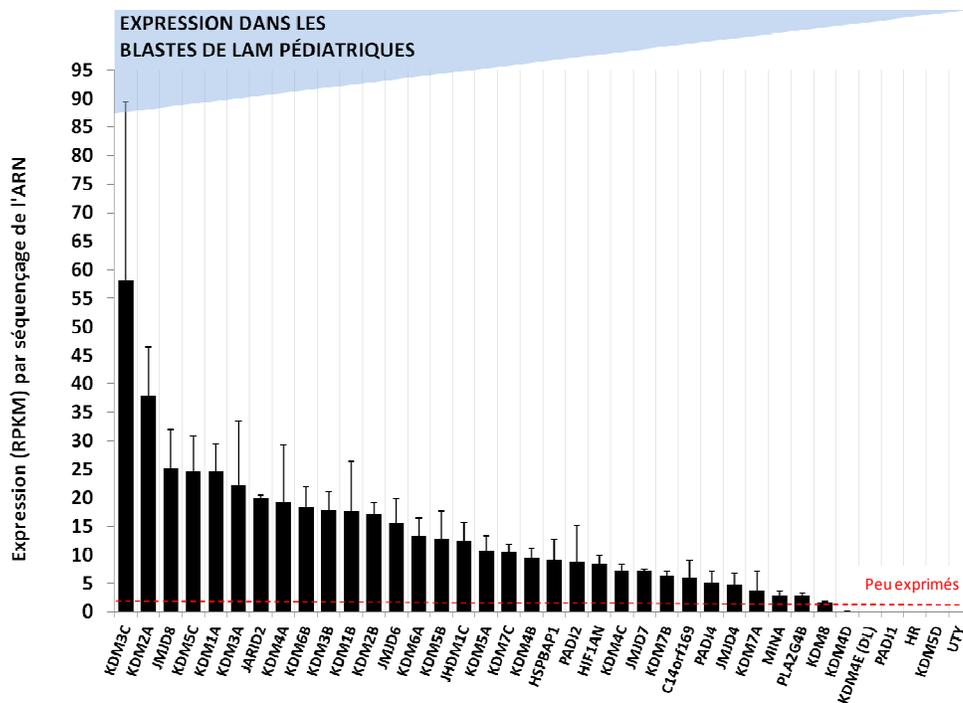


Figure 32. Profil d'expression des histones déméthylases dans les blastes de leucémies aiguës myéloïdes pédiatriques par séquençage de l'ARN.

Analyse de la quantité de transcrits des HDM dans la population de blastes de leucémies aiguës myéloïdes (LAM) pédiatriques. Les résultats montrent les RPKM moyens \pm SD déterminés par séquençage du transcriptome sur $n = 3$ échantillons indépendants (trois patients différents). Seuil d'expression: RPKM>1. Les barres d'erreurs indiquent les SD.

HDM: histones déméthylases; SD: déviation standard; RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

2.3. Certaines histones déméthylases sont différenciellement exprimées dans les cellules de leucémie aiguë par rapport aux cellules hématopoïétiques immatures CD34⁺ de sang de cordon

Nous avons ensuite évalué la variation de niveaux d'expression des HDM par séquençage du transcriptome dans les cellules de LAM de patients pédiatriques par rapport aux cellules immatures CD34⁺ de sang de cordon, détectés par séquençage de l'ARN. La quantité relative des transcrits est déterminée par le ratio des RPKM (cellules leucémiques primaires de LAM) / RPKM (cellules CD34⁺ de sang de cordon) et est représenté dans la figure 33 (cf. Chap. II.6). Un seuil d'expression différentielle a été considéré lorsque le ratio était ≥ 2 ou $\leq 0,5$.

Cette analyse nous permet de mettre en évidence des différences d'expression dans les blastes comparés aux cellules CD34⁺. Ainsi *HSPBAP1*, *KDM4A* et *KDM3C* et *JMJD6* sont entre 2,3 et 1,6 fois plus exprimées dans les blastes de LAM pédiatriques par rapport aux cellules du compartiment immature CD34⁺. Il faut cependant noter une grande variation entre les trois différents échantillons de blastes. Inversement, certains gènes sont moins exprimés dans les blastes par rapport aux cellules immatures CD34⁺: *KDM1A* est 2 fois moins exprimée dans les blastes de LAM par rapport aux cellules CD34⁺ et, de façon très intéressante, on constate que *KDM5B* est environ 3 fois moins exprimée dans les cellules leucémiques comparées aux cellules immatures. L'expression différentielle des HDM *KDM5B* et *KDM1A* entre les cellules leucémiques et les cellules primitives normales soulève la possibilité que ces modificateurs de la chromatine pourraient influencer l'identité cellulaire des CSPH.

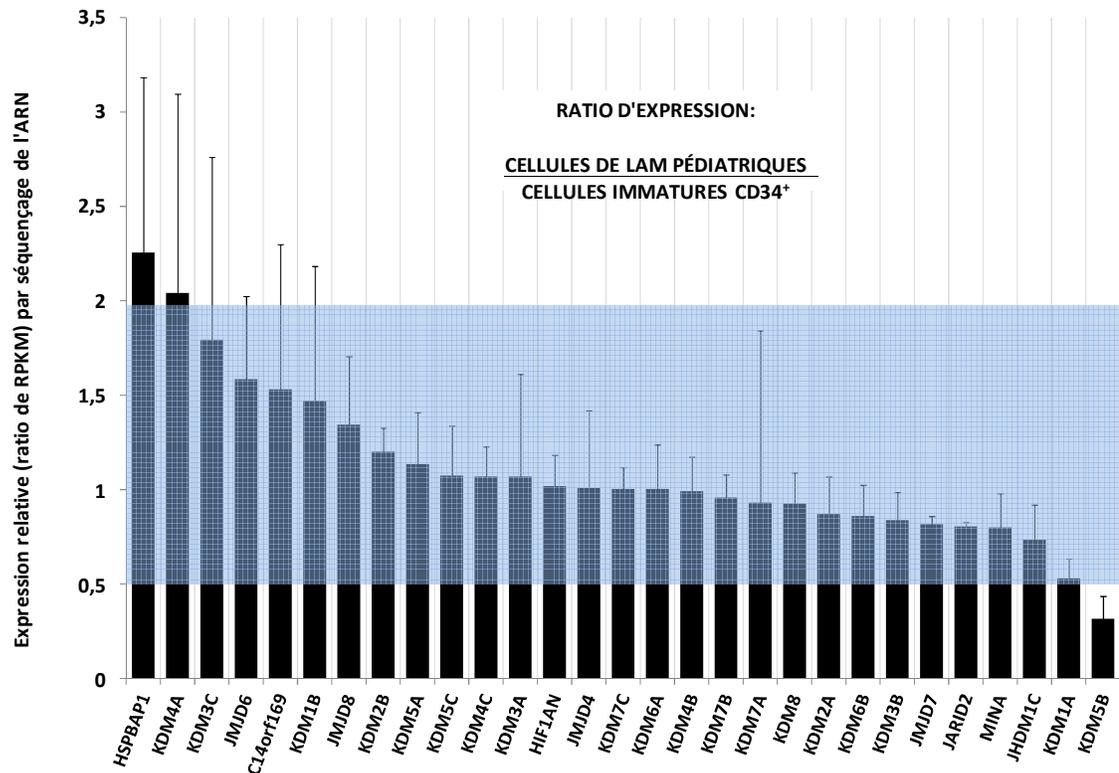


Figure 33 Profil d'expression relative des histones déméthylases dans les blastes de leucémies aiguës myéloïdes pédiatriques par rapport aux cellules sanguines immatures CD34⁺ de sang de cordon par séquençage de l'ARN

Comparaison du niveau de transcrits des HDM détectés dans les cellules de leucémies aiguës myéloïdes (LAM) pédiatriques par rapport aux cellules immatures CD34⁺. La quantité relative des transcrits montrée sur cette figure représente le ratio des RPKM (cellules LAM) / RPKM (CD34⁺) détectés par séquençage de l'ARN (ratio moyen \pm SD sur $n = 3$ échantillons indépendants pour les LAM et $n = 2$ pour la population de cellules CD34⁺). Seuil d'expression différentielle: RPKM ≥ 2 ou $< 0,5$. Les barres d'erreurs indiquent les SD. HDM: histones déméthylases; SD: déviation standard; RPKM: nombre de lectures alignées fragments par kilobase d'exon et par million de lectures alignées.

3. Histones déméthylases non exprimées dans une ou plusieurs populations cellulaires

Le tableau 11 présente certaines HDM qui ne sont pas exprimées dans l'une ou l'autre population et ne peuvent donc pas être analysées avec les calculs d'expression relative présentés précédemment. Il s'agit des gènes *KDM4E(DL)*, *KDM4D*, *PADI2*, *PADI4*, *HR*, *PLA2G4B* et *PADII*. Pour les gènes *KDM4E(DL)*, *PADI2*, *PADI4* et *PLA2G4B*, il existe une

variation d'expression entre les populations cellulaires contrairement aux gènes *KDM4D*, *HR* et *PADI1*.

Gène	Méthode	Cellules différenciées	Cellules CD34 ⁺	Cellules leucémiques
<i>KDM4E(DL)</i>	qRT-PCR (dCt)	-	+ (4,6)	-
	RNAseq (RPKM)	-	-	-
<i>KDM4D</i>	qRT-PCR (dCt)	+ (5,4)	+ (4,6)	+ (5,1)
	RNAseq (RPKM)	-	-	-
<i>PADI2</i>	qRT-PCR (dCt)	+++ (-1,7)	+ (4,8)	++ (2,8)
	RNAseq (RPKM)	+++ (14,5)	-	++ (8,9)
<i>PADI4</i>	qRT-PCR (dCt)	++++ (-3,6)	++ (2,6)	-
	RNAseq (RPKM)	+++ (14,5)	+ (3,7)	++ (5,3)
<i>HR</i>	qRT-PCR (dCt)	-	-	-
	RNAseq (RPKM)	-	-	-
<i>PLA2G4B</i>	qRT-PCR (dCt)	-	-	-
	RNAseq (RPKM)	++ (7,8)	+ (3,2)	+ (2,8)
<i>PADI1</i>	qRT-PCR (dCt)	-	-	-
	RNAseq (RPKM)	-	-	-

Tableau 11. Histones déméthylases non-exprimées dans une des populations cellulaires analysées

Score de valeur d'expression par qRT-PCR: -: dCt > 6 ou pas de donnée, +: dCt entre 6 et 3, ++: dCt entre 3 et 0, +++ entre 0 et -3, ++++ dCt < -3. Score de valeur d'expression par séquençage de l'ARN: -: RPKM < 1 ou pas de donnée, +: RPKM entre 1 et 5, ++: RPKM entre 5 et 10, +++: RPKM entre 10 et 25, ++++: RPKM > 25. Source de cellules leucémiques: lignée MONO-MAC-1 pour la qRT-PCR, blastes de leucémies aiguës myéloïdes pédiatriques pour le séquençage de l'ARN (RNA-seq). RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

Pour les gènes qui présentent une différence d'expression, entre absence et présence d'expression d'un compartiment cellulaire à l'autre comme par exemple pour *PADI2* dont l'expression n'est pas mesurable dans la population cellulaire CD34⁺ alors que son expression est mesurable dans la population cellulaire différenciée et ce avec les deux méthodes d'analyse, ceci pourrait être le reflet d'un rôle biologique différent.

4. Résumé des principales variations d'expression des histones déméthylases dans les différentes populations cellulaires analysées par séquençage de l'ARN

Ainsi, dans notre projet, nous avons pu mettre en évidence des variations importantes d'expression des HDM dans différentes populations hématopoïétiques. Le tableau 12 présente les HDM dont l'expression varie d'au moins 2 fois (augmentation ou diminution) entre les deux populations cellulaires analysées.

HDM	Famille	Cible résidu d'histone connue	Association connue à hémopathie	Expression différentielle:	
				cellules CD34 ⁺ ----- cellules différenciées	blastes ----- cellules CD34 ⁺
<i>KDM5B</i>	JmjC	H3K4me ^{3/2/1}	+	↑↑	↓/↓↓
<i>JHDM1C</i>	JmjC	inconnue	-	↑ / ↑↑	≈
<i>KDM1A</i>	LSD	H3K4me ^{2/1}	+	↑ / ↑↑	↓
<i>KDM4A</i>	JmjC	H3K9me ² H3K36me ²	+	≈ / ↑	↑
<i>HSBPAP1</i>	JmjC	inconnue	-	≈	↑
<i>KDM7A</i>	JmjC	H3K9me ² H3K27me ² H4K20me ¹	-	↓↓/↓↓↓	≈

Tableau 12. Résumé des histones déméthylases qui présentent une variation importante d'expression entre les populations cellulaires analysées

Scores pour les variations d'expression (basés sur les résultats des ratios de RPKM) par séquençage de l'ARN:

Augmentation: ↑↑: 4 fois; ↑/↑↑: entre 2 et 4 fois; ↑: 2 fois; ≈/↑: entre 1,5 et 2 fois; Équilibre: ≈ expression similaire (ratio entre 1,5 et 0,5);

Diminution: ↓: de 2 fois; ↓↓: de 4 fois; ↓↓/↓↓↓: de 4 à 8 fois; ↓↓↓: de 8 fois.

Les cellules leucémiques utilisées pour l'analyse par RNA-seq sont les blastes de leucémies aiguës myéloïdes pédiatriques. HDM: histones déméthylases; LSD: Famille d'HDM des lysine-specific déméthylases; JmjC: famille d'HDM Jumonji C; RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

Il est intéressant de noter que *KDM5B* et *KDM1A* présentent le même profil de variation d'expression augmentée dans les cellules immatures CD34⁺ par rapport aux cellules différenciées et sont par ailleurs fortement diminuées dans le contexte de prolifération cellulaire dérégulée des cellules leucémiques. Or ces déméthylases ciblent le même résidu H3K4 (H3K4me3/2/1 et H3K4me2/1 respectivement pour *KDM5B* et *KDM1A*) avec la triméthylation de H3K4 au niveau des régions promotrices qui, comme nous l'avons vu précédemment, est associé à l'activation de la transcription et donc à une structure de la chromatine plus ouverte (ou euchromatine, cf. figure 4 et cf. Chap.I. 2.1). Il est possible que ces enzymes ciblent des gènes importants pour la différenciation cellulaire en régulant leur activation épigénétique. L'expression de *KDM4A*, qui cible les résidus d'histone H3K9me² et H3K36me², est augmentée dans les cellules immatures normales et dans les cellules leucémiques. Cette augmentation d'expression relative pourrait jouer un rôle dans l'identité des CSPH et peut-être aussi dans celle des cellules leucémiques. Pour ce qui est de *KDM7A*, le profil de variation est fortement diminué dans les cellules immatures normales et montre une expression relativement similaire entre ces cellules et les cellules leucémiques, ce qui soulève l'hypothèse de son implication dans la répression d'un processus biologique commun entre cellules immatures normales et cellules leucémiques.

5. Niveau d'expression des histones déméthylases dans le reste des tissus non-hématopoïétiques

Les différences de niveau d'expression mises en évidence dans ce projet concernent des populations cellulaires hématopoïétiques normales et leucémiques. Il est dès lors pertinent de regarder également le niveau d'expression de ces HDM dans les autres tissus humains à titre comparatif. Le site «BioGPS <http://biogps.gnf.org> » est un portail publiquement accessible qui centralise différentes informations dont les données d'expression de tous les gènes dans 79 tissus humains différents (analyses d'expression réalisées par puces à ADN [159]). Le tableau 13A résume les données disponibles et montre que même si les HDM sont exprimées dans tous les tissus, plus de la moitié d'entre elles sont exprimées de façon nettement supérieure dans les tissus hématopoïétiques en comparaison à la moyenne des niveaux d'expression dans les 79 tissus humains analysés. Pour certaines d'entre elles, l'expression

est également élevée dans un ou plusieurs autres tissus tels que l'épiphyse ou les testicules en plus du tissu hématopoïétique (*KDM1A* et *KDM3A*, par exemple).

Symbole	Expression moyenne totale (79 tissus)	Expression moyenne tissu hématopoïétique	Expression la plus élevée des populations hématopoïétiques et type cellulaire	Autre tissu avec expression élevée	
KDM1A	84	75	300	lymphoblastes	Testicules (250), épiphyse (150)
KDM1B	5	=	=	=	5
KDM2A	121	200	415	cellules myéloïdes	Cortex surrénalien(177), foie (190), épiphyse (160), SNC (211)
KDM2B	6	=	=		=
KDM3A	100	300	550	lymphocytes	Testicules (360)
KDM3B	69	165	180	cellules CD34	Thymus (85), rétine (120) épiphyse (170)
KDM3C	40	120	415	cellules myéloïdes	Épiphyse(90), cerveau foetal(100), utérus (162)
KDM4A	10	12	15	cellules myéloïdes	Rein (15), SNC (10)
KDM4B	21	100	196	lymphocytes B	Thymus (56), testicule (63), prostate (109), foie (105)
KDM4C	8	24	66	lymphocytes B	Épiphyse (41)
KDM4D	7	=	=		
KDM4E (DL)	x	X	X		x
KDM5A	36	180	446	cellules NK	Thymus (85), épiphyse(50)
KDM5B	5	=	=	=	Myocytes (8)
KDM5C	121	80	122	lymphoblastes	
KDM5D	31	80	340	lymphocytes T	Ganglion lymphatiques (107), épiphyse (80)
KDM6A	8	=	=		Ganglion lymphatiques (11), ovaire (12), cardiomyocytes (11)
UTY	11	11	11		Ganglion cervical supérieur (49), appendice (33)
KDM6B	13	28	119	sang total	SNC foetal (20)
KDM7A	37	100	226	cellules érythroïdes	Épiphyse (93)
KDM7B	8	11	13	cellules CD34	Thymus (17), testicule 22), épiphyse 12
KDM7C	10	10	10		Cortex surrénalien (15), foie (15), ganglion cervical supérieur(19)
KDM8	6	=	=		Foie(21)
C14orf169	6	11	22	cellules CD34	Thymus(24), testicules (33)
HIF1AN	9	=	=		Foie(12)
HR	9	=	=		Épiphyse (30)
HSPBAP1	11	30	49	cellules myéloïdes	Épiphyse 40
JARID2	117	300	482	cellules CD34 et myéloïdes	Épiphyse (350), rétine (230), testicules(150)
JHDM1C	6	6	10	lymphocytes B	SNC foetal(11)
JMJD4	5	=	=		Thymus (7), épiphyse (10), prostate (8)
JMJD6	18	30	70	monocytes	Pancréas (33)
JMJD7	7	10	50	lymphocytes T	Ganglions lymphatiques (17), thyroïde (21)
JMJD8	18	30	50	monocytes	Pancréas (33)
MINA	56	65	133	cellules dendritiques	
PLA2G4B	7	25	50	lymphocytes T	Ganglion lymphatiques (17), thyroïde (20)
PADI1	26	30	30	sang total	Muscle (30), foie (35), cardiomyocytes (35), épiphyse(33)
PADI2	100	300	600	sang total	
PADI3	5	=	=		Cardiomyocyte (7), muscle (6)
PADI4	6	=	=		Cardiomyocytes (9,5) foie (9)
PADI6	6	=	=		

Tableau 13A. Niveau d'expression moyen des histones déméthylases dans les tissus hématopoïétiques et d'autres tissus humains (www.biogps.com)

Données d'expression analysées par la méthode des puces à ADN (Affymetrix, mesure de l'intensité de fluorescence exprimée en unité gcrma) tel que décrit par Wu *et al.*[159] et disponibles sur le site « www.biogps.com ». L'expression moyenne totale correspond à la moyenne des niveaux d'expression dans les 79 tissus analysés. L'expression moyenne du tissu hématopoïétique correspond à la moyenne des niveaux d'expression de différents compartiments cellulaires (sang total, lymphocytes T ou B, lymphoblastes, cellules myéloïdes, cellules CD34, etc.). Le compartiment cellulaire hématopoïétique dans lequel l'expression est la plus élevée est rapporté. Les rares tissus pour lesquels les valeurs d'expression sont supérieures à la moyenne tissulaire totale sont également rapportés. Les valeurs d'expression ont été mesurées par la méthode des. SNC : système nerveux central; HDM : histones déméthylases; x : pas de données disponibles, = : valeurs similaires à la moyenne tissulaire.

Le tableau 13B montre le ratio d'expression entre le niveau moyen d'expression dans l'ensemble des compartiments cellulaires hématopoïétiques (sang total, lymphocytes T ou B, lymphoblastes, cellules myéloïdes, cellules CD34, etc.) et le niveau moyen d'expression dans les 79 tissus analysés, et également le ratio d'expression entre le compartiment cellulaire hématopoïétique dans lequel l'expression est la plus élevée et le niveau moyen d'expression dans les 79 tissus analysés. Nous constatons ainsi que l'expression de certaines HDM est fortement enrichi dans les tissus hématopoïétiques : famille *KDM3*, famille *KDM4*, *KDM5A*, *KDM5D*, *KDM7A*, *HSPBAP1*, *JARID2*, *PLA2G4B*, et *PADI2*. On constate également que certaines d'entre elles sont plus de 9 fois plus exprimées dans un des compartiments cellulaires hématopoïétiques: par exemple *KDM3C* dans les cellules myéloïdes, *KDM4B* dans les lymphocytes B, ou encore *KDM5A* dans les cellules NK.

Les HDM dont la variation d'expression est la plus significative dans notre analyse, *KDM5B*, *JHDM1C*, *KDM1A*, *KDM4A*, *HSBPAP1* et *KDM7*, sont marquées d'un astérisque dans le tableau 13B. La moitié de ces candidats, *KDM1A*, *HSBPAP1* et *KDM7A* sont plus exprimées dans un des compartiments hématopoïétiques par rapport à la moyenne d'expression dans les 79 tissus humains.

Ainsi, les HDM sont exprimées partout, mais pas de façon ubiquitaire, avec une expression préférentielle dans les tissus hématopoïétiques pour certaines. Ces informations sont par exemple particulièrement utiles à connaître dans la perspective d'une future utilisation d'inhibiteurs des HDM dans un but thérapeutique (cf. Chap. 1.4.2), afin de prédire une potentielle toxicité extra-hématopoïétique

Symbole	ratio expression tissu hématopoïétique / total	ratio d'expression population hématopoïétique / total	Type cellulaire
KDM1A*	0,9	3,6	lymphoblastes
KDM1B	1,0	1,0	
KDM2A	1,7	3,4	cellules myéloïdes
KDM2B	1,0	1,0	
KDM3A	3,0	5,5	lymphocytes
KDM3B	2,4	2,6	cellules CD34
KDM3C	3,0	10,4	cellules myéloïdes
KDM4A*	1,2	1,5	cellules myéloïdes
KDM4B	4,8	9,3	lymphocytes B
KDM4C	3,0	8,3	lymphocytes B
KDM4D	1,0	1,0	
KDM4E (DL)	x	x	
KDM5A	5,0	12,4	cellules NK
KDM5B*	1,0	1,0	
KDM5C	0,7	1,0	lymphoblastes
KDM5D	2,6	11,0	lymphocytes T
KDM6A	1,0	1,0	
UTY	1,0	1,0	
KDM6B	2,2	9,2	sang total
KDM7A*	2,7	6,1	cellules érythroïdes
KDM7B	1,4	1,6	cellules CD34
KDM7C	1,0	1,0	
KDM8	1,0	1,0	
C14orf169	1,8	3,7	cellules CD34
HIF1AN	1,0	1,0	
HR	1,0	1,0	
HSPBAP1*	2,7	4,5	cellules myéloïdes
JARID2	2,6	4,1	cellules CD34 et myéloïdes
JHDM1C*	1,0	1,7	lymphocytes B
JMJD4	1,0	1,0	
JMJD6	1,7	3,9	monocytes
JMJD7	1,4	7,1	lymphocytes T
JMJD8	1,7	2,8	monocytes
MINA	1,2	2,4	cellules dendritiques
PLA2G4B	3,6	7,1	lymphocytes T
PADI1	1,2	1,2	sang total
PADI2	3,0	6,0	sang total
PADI3	1,0	1,0	
PADI4	1,0	1,0	
PADI6	1,0	1,0	

Tableau 13B. Ratio des niveaux d'expression moyen des histones déméthylases dans le tissu hématopoïétique comparé au niveau moyen d'expression dans 79 tissus humains (www.biogps.com)

Données d'expression analysées par la méthode des puces à ADN (Affymetrix, mesure de l'intensité de fluorescence exprimée en unité gcma) tel que décrit par Wu *et al.*[159] et disponibles sur le site « www.biogps.com ». Le premier ratio d'expression est réalisé entre le niveau moyen d'expression dans l'ensemble des compartiments cellulaires hématopoïétiques (sang total, lymphocytes T ou B, lymphoblastes, cellules myéloïdes, cellules CD34, etc.) et le niveau moyen d'expression dans les 79 tissus analysés. Le deuxième ratio d'expression est

réalisé entre le compartiment cellulaire hématopoïétique dans lequel l'expression est la plus élevée et le niveau moyen d'expression dans les 79 tissus analysés. HDM : histones déméthylases; x : pas de données disponibles, = : valeurs similaires à la moyenne tissulaire, * : HDM différenciellement exprimées dans notre projet entre les compartiments cellulaires immatures, totaux et/ou leucémiques.

6. Profil d'expression des histones déméthylases dans les blastes de leucémie aiguë myéloïde de la banque de données du projet TCGA

Enfin, compte tenu du faible nombre d'échantillons de LAM pédiatriques disponibles pour notre analyse, et malgré les différences qui existent entre LAM pédiatriques et LAM de l'adulte, nous avons voulu comparer certaines de nos observations d'expression des HDM dans les cellules leucémiques pédiatriques avec les données d'expression disponibles d'une série de LAM adultes de la banque de données du projet TCGA. Cette série comporte en effet les données d'expression par séquençage de l'ARN de 179 échantillons de LAM de patients adultes.

La figure 34 présente le profil d'expression des HDM dans les cellules de LAM adulte de cette banque de données. En considérant comme seuil d'expression un RPKM > 1, nous constatons que les HDM sont aussi majoritairement exprimées dans ces échantillons de LAM adulte, à l'exception de *KDM4D*, *HR* et *PAD11*, tel qu'observé dans notre étude. A noter que dans cette série, il n'y a pas de donnée disponible pour *KDM4E(DL)*. Ce profil d'expression est, de façon intéressante, très similaire à celui que nous avons obtenu dans nos 3 échantillons de LAM pédiatrique (figure 32) et particulièrement pour les 10 gènes les plus exprimés (en bleu sur la figure 34) qui sont les mêmes dans les deux populations de LAM (TCGA et pédiatrique) à l'exception du transcrit de *KDM5B* qui est plus exprimé dans les LAM du projet TCGA.

SECTION 2: MISE EN PERSPECTIVE DE L'EXPRESSION DES HISTONES DÉMÉTHYLASES (HDM) PAR RAPPORT AU TRANSCRIPTOME COMPLET DANS LES CELLULES HÉMATOPOÏÉTIQUES NORMALES ET LEUCÉMIQUES

1. Profil d'expression quantitative du transcriptome complet

Nous avons ensuite voulu mettre en perspective l'expression des HDM par rapport au transcriptome complet. Lorsqu'on analyse les résultats d'expression du transcriptome complet dans les cellules normales, immatures CD34+ ou différenciées, (figure 35-A) et les cellules leucémiques (figure 35-B), il est intéressant de constater que le profil de répartition de la quantité de gènes en fonction de l'importance de leur expression est similaire entre les trois populations cellulaires étudiées. Plus particulièrement, en prenant le seuil d'absence ou de faible expression à RPKM < 1 (cf. Chap.II. 7), il est intéressant de constater que la quantité de gènes peu ou pas exprimés dans les cellules plus différenciées n'est pas plus importante que dans les cellules plus différenciées et ce y compris dans les cellules leucémiques (environ 45 à 50% des gènes dans notre analyse). Ceci suggère que ce n'est vraisemblablement pas le nombre de gènes exprimés qui diffère entre compartiment cellulaire immature et compartiment cellulaire différencié, mais plutôt le profil d'expression des gènes c'est-à-dire quels gènes et avec quel niveau d'expression.

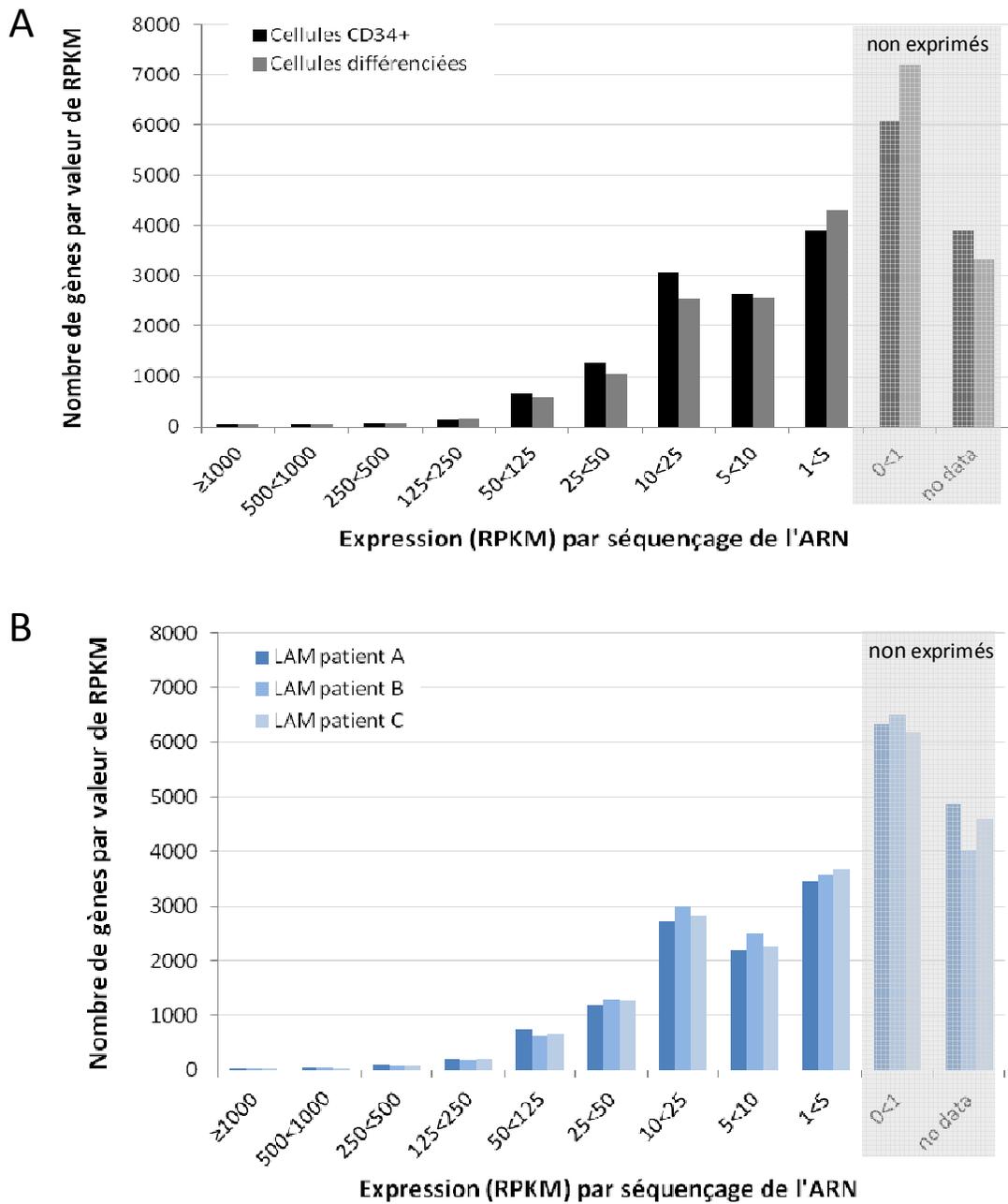


Figure 35. Répartition des gènes du transcriptome complet en fonction de leur expression quantitative par séquençage de l'ARN

Analyse par séquençage de l'ARN du transcriptome des cellules normales CD34+ ou différenciées, provenant d'unités de sang de cordon (A) et des cellules leucémiques de leucémies aiguës myéloïdes (LAM) pédiatriques (B). RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

2. Répartition de l'expression des gènes reconnus essentiels pour les cellules souches hématopoïétiques par rapport au transcriptome global

Comme nous l'avons vu dans l'introduction (cf. Chap. I. 1.4.2.1), une vingtaine de gènes codant principalement pour des facteurs de transcription sont reconnus comme étant essentiels pour l'identité ou la fonction des CSH. Nous avons dès lors voulu situer le niveau d'expression de ces gènes par rapport au profil d'expression du transcriptome complet. La figure 36 nous montre que ces gènes sont exprimés avec des RPKM entre 1 et 125 dans la population cellulaire immature CD34⁺ normale de notre projet. De plus, la grande majorité de ces gènes ($\approx 90\%$) sont fortement exprimés avec des RPKM entre 10 et 125.

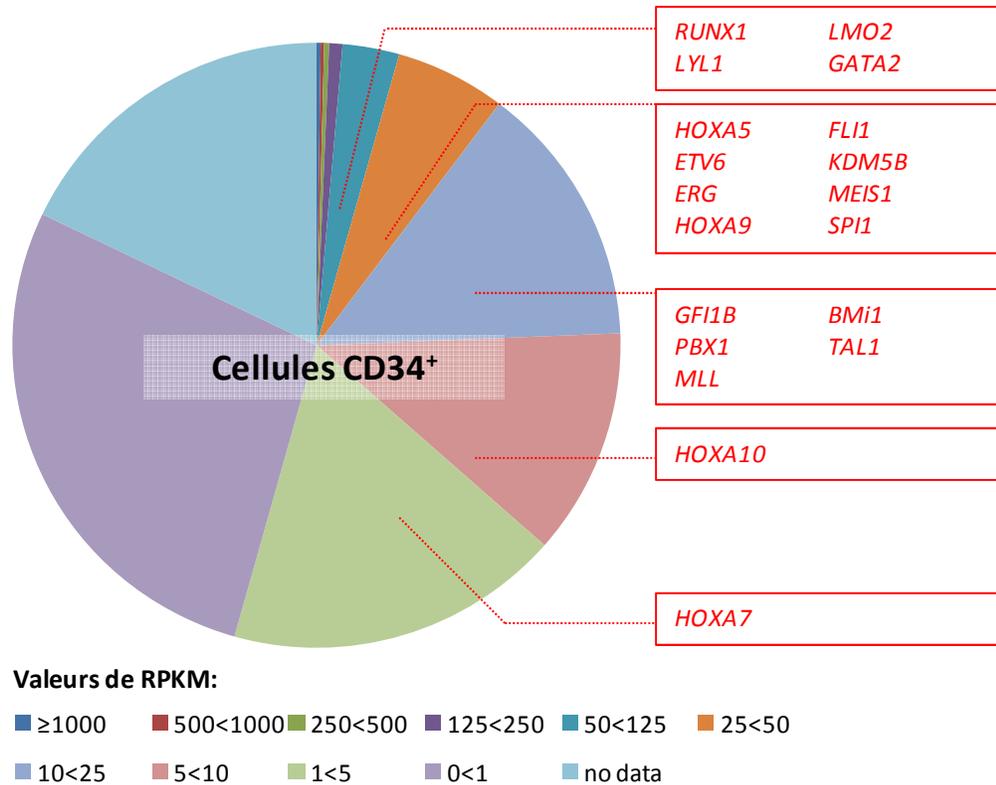


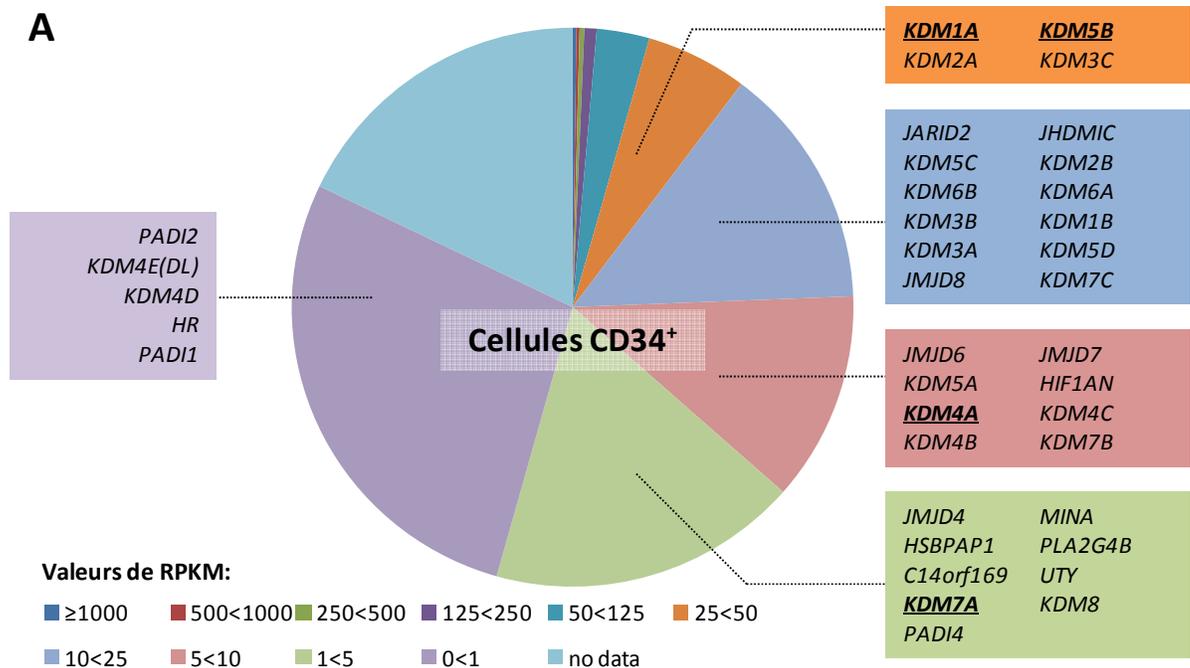
Figure 36 Répartition des gènes clés de l'hématopoïèse en fonction de l'importance de leur expression dans le transcriptome complet

Résultats de l'analyse du séquençage de l'ARN dans la population cellulaire immature normale CD34⁺, exprimés en RPKM. RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

3. Répartition des HDM en fonction de l'importance de leur expression par rapport au transcriptome complet

La figure 37 présente la répartition de l'expression des HDM dans les 3 populations cellulaires analysées dans notre projet par séquençage de l'ARN par rapport au transcriptome complet.

On constate que le profil de répartition d'expression en terme quantitatif est similaire dans ces différentes populations. Cette figure met aussi en évidence les variations d'expression des gènes *KDM5B*, *KDM1A*, *KDM4A* et *KDM7A*.



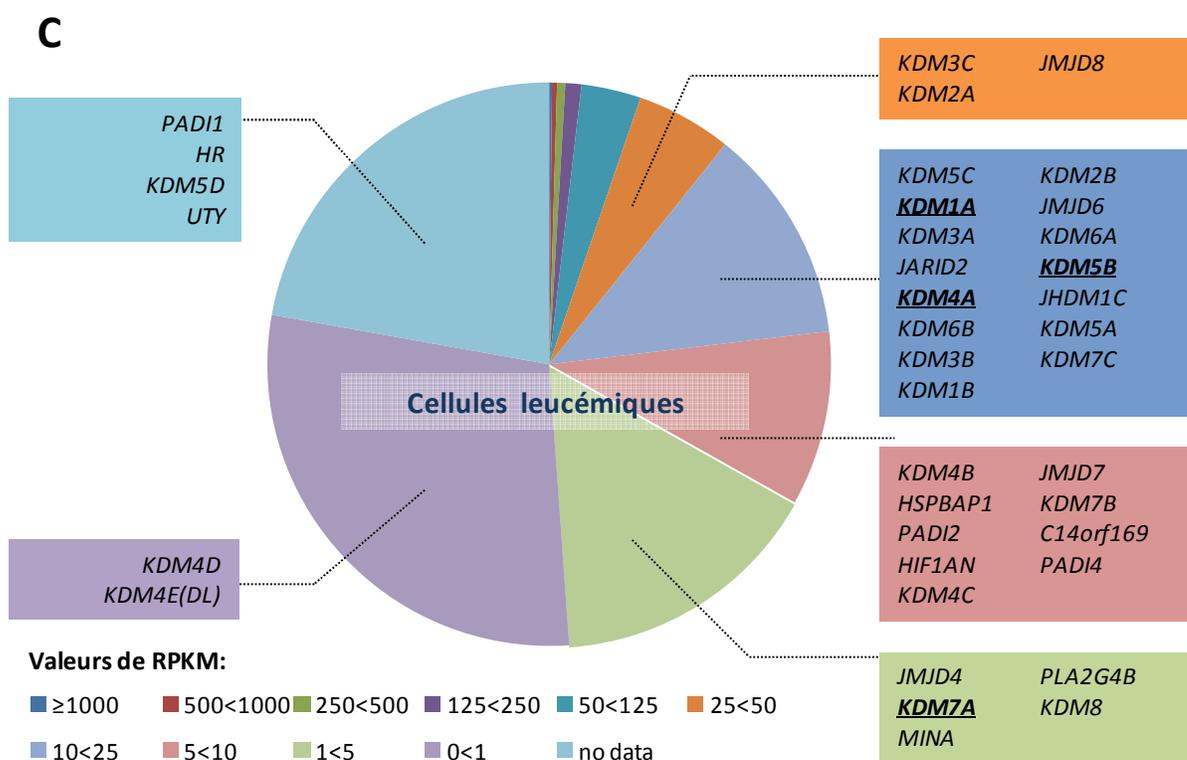
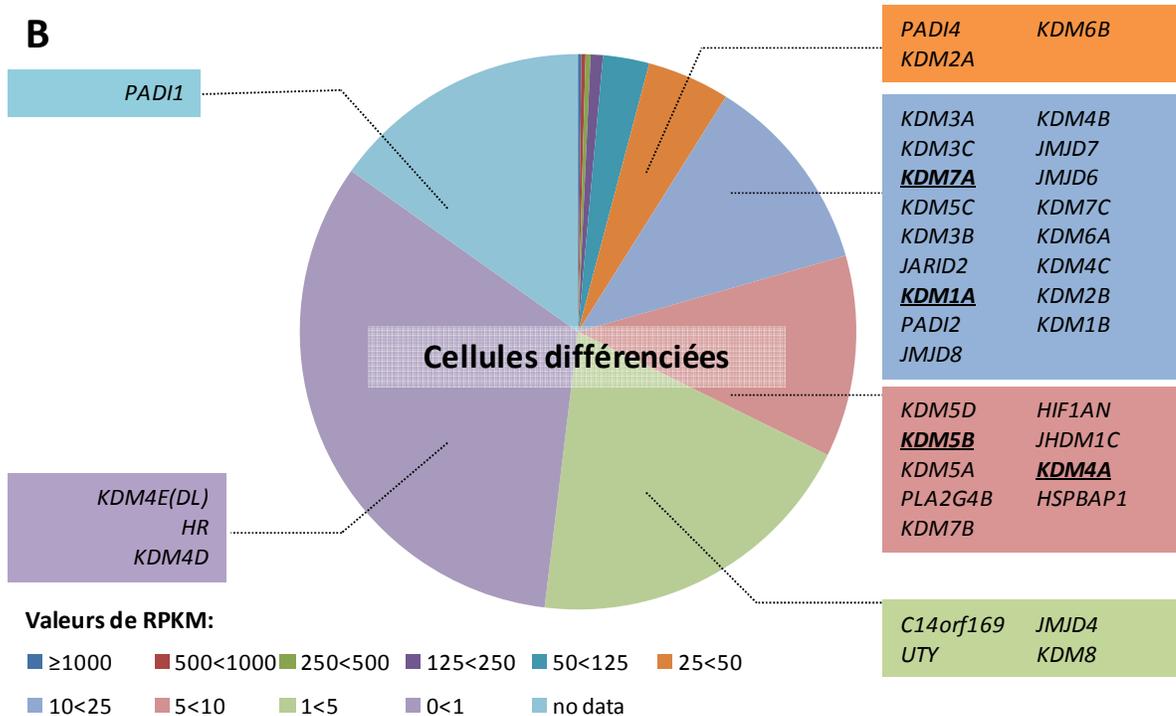


Figure 37. Répartition des histones déméthylases en fonction de leur expression quantitative par rapport au transcriptome complet par séquençage de l'ARN

Résultats de l'analyse du séquençage de l'ARN dans la population cellulaire immature normale CD34⁺ (A), dans les cellules différenciées (B) et dans les blastes de leucémies aiguës myéloïdes pédiatriques (C) exprimés en RPKM. RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

4. Expression des gènes essentiels aux CSH dans les cellules immatures CD34⁺ normales et dans les cellules différenciées

La figure 38 présente l'expression mesurée par séquençage de l'ARN de 19 gènes essentiels aux cellules souches hématopoïétiques (CSH) dans les populations cellulaires immatures normales CD34⁺ et différenciées analysées dans notre étude.

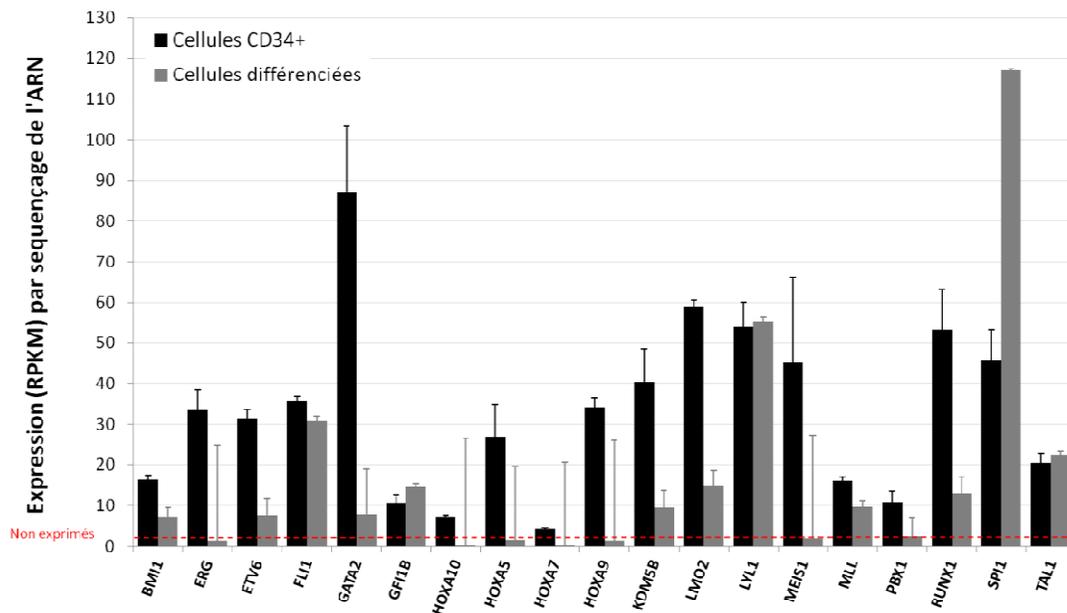


Figure 38. Expression des 19 gènes essentiels aux cellules souches hématopoïétiques dans les populations cellulaires immatures CD34⁺ et les cellules différenciées par séquençage de l'ARN

Résultats de l'analyse du séquençage de l'ARN dans la population cellulaire immature normale CD34⁺ et la population cellulaire différenciée. Seuil d'expression: RPKM>1. RPKM: nombre de lectures alignées par kilobase par million de lectures alignées.

Le niveau d'expression d'un transcrit n'est certes pas suffisant pour définir la fonction de la protéine correspondante dans une population donnée, mais il permet de suspecter un rôle spécifique de la protéine lorsque l'expression de son gène est différente entre compartiment

cellulaire immature et différencié d'un même tissu. La figure ci-dessus montre une différence importante d'expression de ces gènes essentiels aux CSH dans les cellules immatures CD34⁺ normales comparées aux cellules différenciées. Il est particulièrement intéressant de constater que cette différence d'expression existe malgré le fait que notre population cellulaire immature CD34⁺ ne soit pas une population homogène de CSH mais plutôt un mélange de CSH et de cellules progénitrices.

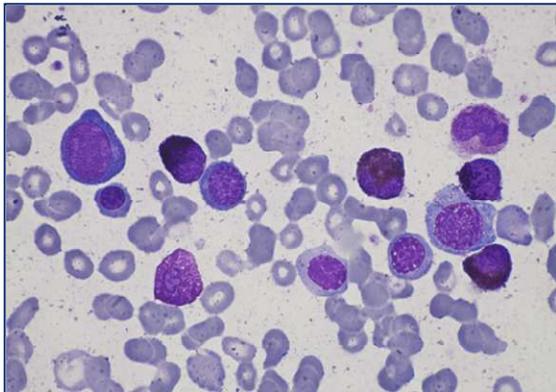
Il est aussi très intéressant de noter que parmi les gènes très différenciellement exprimés entre cellules immatures et différenciées, deux d'entre eux, *GATA2* et *RUNX1*, qui sont tous deux des gènes codant pour des facteurs de transcription importants pour les cellules souches hématopoïétiques (cf. Chap. I.1.4.2.1), sont souvent impliqués dans des hémopathies malignes (cf Chap. I.2.4). La figure 38 montre en effet que *GATA2* est 10 fois plus exprimé et *RUNX1* est plus de 4 fois plus exprimé dans les cellules immatures CD34⁺ normales par rapport aux cellules différenciées. Par ailleurs, des mutations de ces deux gènes existent et sont associées à des hémopathies malignes, comme nous l'avons évoqué précédemment (cf Chap. I.6 et Chap II. 2.11.1) et tel que mis en évidence chez un patient pour le gène *GATA2* (cf section suivante).

SECTION 3: ÉLABORATION D'UN TEST CLINIQUE POUR LA DÉTECTION DES MUTATIONS DU GÈNE *GATA2* DANS LE CADRE D'UN SYNDROME MYÉLODYSPLASIQUE ATYPIQUE

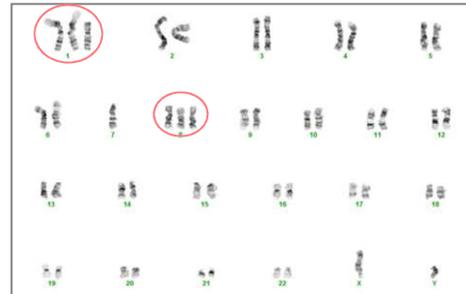
1. Histoire clinique: syndrome myélodysplasique atypique chez un jeune garçon et suspicion de syndrome de MonoMAC

Un jeune garçon de 12 ans a été référé dans notre institution pour la prise en charge d'une pancytopenie avec une histoire d'infections virales récurrentes et de verrues plantaires. Le patient présentait comme seul dysmorphisme une jambe droite plus courte que la gauche non expliquée génétiquement. L'analyse morphologique de sa moelle osseuse démontrait une dysplasie multilignée compatible avec un syndrome myélodysplasique (figure 39-A) et le caryotype réalisé était anormal: 47,XY,+1,der(1;7)(q10;p10),+8[17]/46,XY[4] (figure 39-B). En plus de la pancytopenie marquée et une profonde monocytopenie, l'immunophénotypage sanguin a démontré chez ce patient une lymphopénie T, ainsi qu'une absence de lymphocytes B et de cellules NK (figure 39-C).

A



B



47,XY,+1,der(1;7)(q10;p10),+8[17]
46,XY[4]

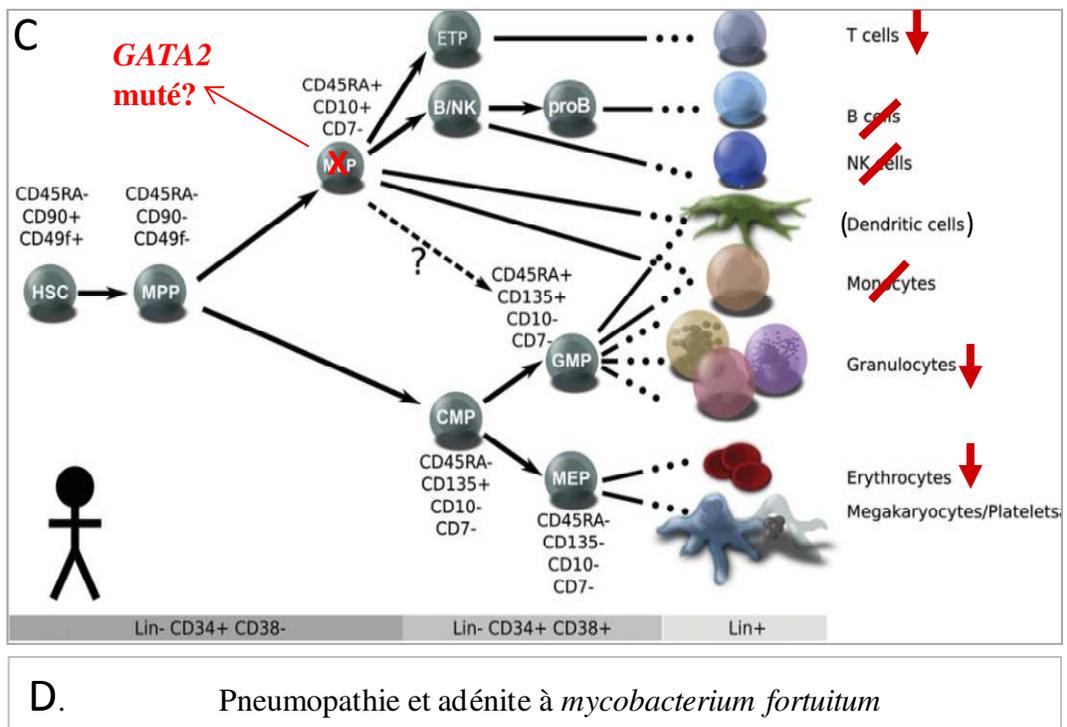


Figure 39. Caractéristiques cliniques et biologiques du cas présenté

(A) Frottis de moelle osseuse mettant en évidence un syndrome myélodysplasique (Giemsa, 100X). (B) Caryotype standard avec présence d'un chromosome dérivé der(1;7) et de trisomies des chromosomes 1 et 8. (C) Figure adaptée de Cell Stem Cell, **10**(2), Doulatov, S., et al. (2012)[6] "*Hematopoiesis: a human perspective*", 120-136 Copyright (2012), avec la permission d'Elsevier). Cytopénies mise en évidence par le bilan sanguin et l'immunophénotypage des globules blancs: lymphopénie T, absence de lymphocytes B et NK, absence de monocytes, neutropénie et anémie. Suspicion de arrêt de maturation précoce par mutation de *GATA*. HSC: cellules souches hématopoïétiques; MPP: progéniteurs multipotents; MLP: progéniteurs lymphoïdes communs; CMP: progéniteurs myéloïdes communs; GMP: progéniteurs des granulocytes et macrophages; MEP: progéniteurs des mégacaryocytes et des érythrocytes. (D) Diagnostic d'infection à *mycobacterium fortuitum*.

Cette condition se traite par une transplantation de cellules souches hématopoïétiques, et dans ce cas, un donneur de moelle osseuse compatible était identifié dans la fratrie du patient. Une allogreffe de moelle osseuse familiale a donc été programmée. Toutefois, peu avant le début du conditionnement, une infection pulmonaire asymptomatique et une adénite sous-claviculaire à *Mycobacterium Fortuitum* ont été découvertes chez ce patient. Cette présentation clinique atypique est très évocatrice du syndrome de MonoMAC qui associe une monocytopenie et une infection à mycobactérie atypique, condition qui évolue vers un SMD.

Ce syndrome est lié à une mutation du gène *GATA2* et a pour particularité d'être héréditaire, sur un mode autosomal dominant [108]. La décision de greffe familiale a donc été suspendue temporairement afin de rechercher rapidement une éventuelle mutation de *GATA2* chez ce patient et chez le donneur familial potentiel.

2. Résultat de l'analyse de séquençage du gène *GATA2*

La figure 40 montre les résultats du test diagnostique par séquençage de l'ADN à partir d'échantillon sanguin chez le patient, son frère et un contrôle normal.

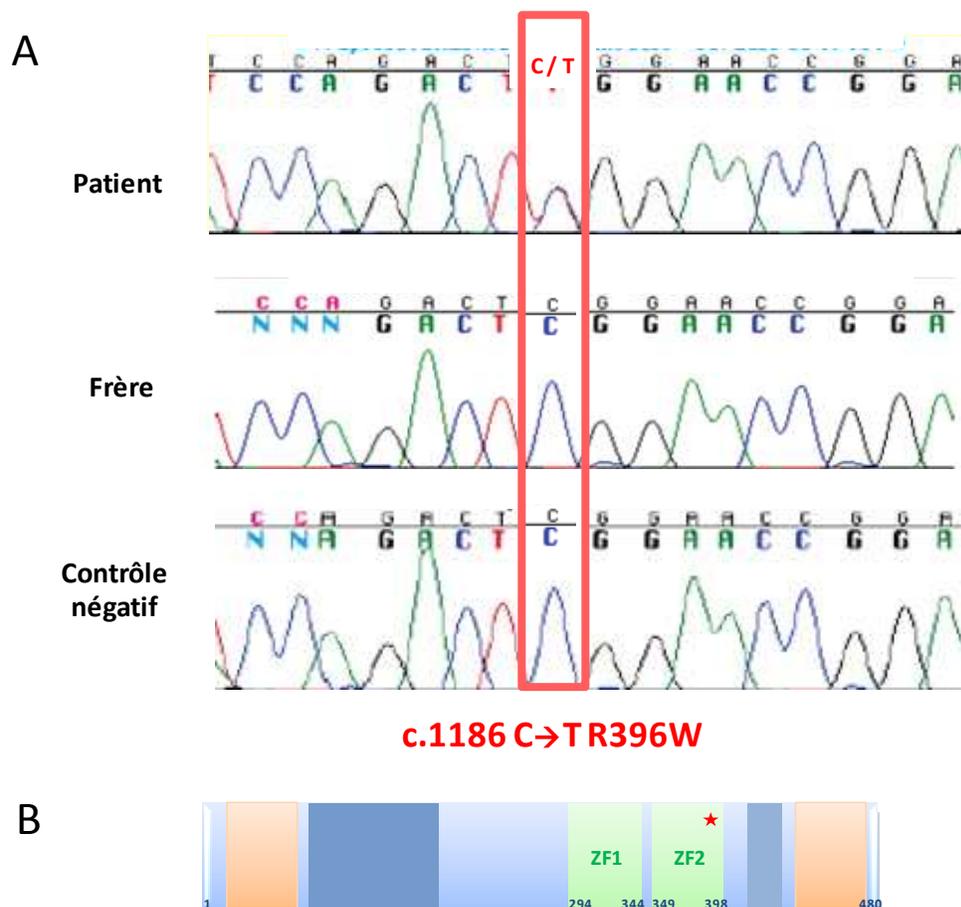


Figure 40. Mise en évidence d'une mutation ponctuelle hétérozygote c.1186 C→T R396W de *GATA2* dans l'échantillon du patient

(A) Résultats obtenus par séquençage de l'ADN chez le patient, son frère et un contrôle normal et présentés par chromatographie et séquence des nucléotides. Le patient présente un pic double C (cytosine, bleu) et T (thymine, rouge) en position c.1186 correspondant à une

mutation ponctuelle de *GATA2*. (B) Représentation schématique de la protéine *GATA2* et des domaines en doigt de zinc ZF1 et ZF2 et localisation de la mutation identifiée chez le patient à l'acide aminé 396 qui remplace une arginine (R) par un tryptophane (W) (étoile rouge).

Les résultats du séquençage par chromatographie montrent chez le patient un pic anormal rouge hétérozygote en position c.1186 qui remplace une cytosine par une thymine. Cette mutation ponctuelle conduit au remplacement d'une arginine par un tryptophane en position 396 de la protéine *GATA2*. Elle n'est pas retrouvée chez son donneur. Le test a été réalisé deux fois chez notre patient de manière indépendante. Le même test a été ensuite réalisé à partir d'échantillon sanguin des deux parents: les deux parents sont exempts de la mutation (figure 41).

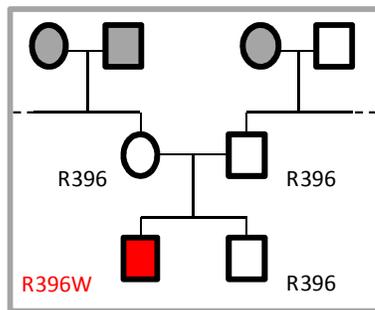


Figure 41. Le pedigree familial montre le caractère sporadique de la mutation R396W de ce cas clinique

Le patient présentant la mutation figure en rouge. Les membres de la famille testés pour la mutation de *GATA2* ont le résultat de l'analyse en regard de leur figure. Les membres de la famille atteints d'un cancer non-hématologique sont indiqués en gris.

Le résultat de cette analyse a, de plus, été confirmé par séquençage de l'ADN à partir d'échantillons de salive du patient et du donneur, traités dans un laboratoire indépendant en Australie (laboratoire du Dr Scott). Cette analyse-contrôle confirme la présence chez notre patient d'une mutation ponctuelle hétérozygote de *GATA2* en position R396W de la protéine, apparue de façon sporadique et germinale et associée cliniquement à un syndrome de MonoMAC qui a évolué en SMD. Le patient a donc pu être greffé à partir des CSH provenant du donneur familial.

CHAPITRE IV: DISCUSSION

1. L'étude des cellules souches hématopoïétiques humaines

A l'heure actuelle, il est encore difficile d'isoler des cellules souches hématopoïétiques (CSH) pour les étudier car il n'existe pas un seul marqueur spécifique permettant de les identifier (cf. Chap. I.1.3.2). Nous avons choisi d'utiliser un degré de purification nous permettant d'isoler des cellules suffisamment immatures pour présenter des caractéristiques biologiques des cellules souches et progénitrices telles que les cellules CD34⁺. Comme nous l'avons vu plus tôt, la fréquence des CSH dans la population cellulaire CD34⁺, selon les données de la littérature, est estimée à environ 1 dans environ 1000 cellules CD34⁺ [10]. Le meilleur niveau de purification de CSH à partir de sang de cordon décrit dans la littérature permet d'augmenter la fréquence des CSH à 28% (soit une fréquence de CSH de 1 pour 5 cellules) en utilisant les marqueurs CD34⁺ CD38⁻ CD45RA⁻ Thy1⁺ CD49f⁺ Rho^{lo} [14]. Au mieux, 50 cellules avec cet immunophénotype sont purifiées par unité de sang de cordon, (communication Frédéric Barabé). Opter pour l'isolement d'une population de cellules CD34⁺ dans notre projet, nous a donc permis de trier à partir de chaque USC, une population de cellules CD34⁺ en quantité suffisante pour effectuer les analyses de profils d'expression. Une vérification de l'homogénéité des populations de CSH isolées (contenu en cellules CD34⁺) a été effectuée au préalable par évaluation morphologique et immunophénotypique (FACS). Ce choix nous a permis également d'augmenter la reproductibilité de nos procédures et d'avoir ainsi plusieurs échantillons équivalents manipulés simultanément et indépendamment pour la plupart des étapes subséquentes de purification plutôt que de grouper plusieurs échantillons biologiques de CSH plus pures. L'absence de certitude concernant la quantité de CSH contenue dans nos populations de cellules immatures, puisque la seule méthode valable de quantification est l'évaluation du potentiel de reconstitution par xénotransplantation par la méthode des dilutions limites (cf. Chap. I.1.3.3) doit être prise en compte lors de l'interprétation des résultats de comparaison des profils d'expression entre la population immature et les autres.

2. Intérêt comparé des différentes sources de cellules leucémiques

Les lignées cellulaires de leucémie humaine telles que la lignée MONO-MAC-1 utilisée dans notre projet, offrent de nombreux avantages. Ces cellules sont facilement maintenues en culture, ont des caractéristiques morphologiques et phénotypiques relativement stables et permettent la manipulation d'une grande quantité de cellules. L'analyse de leur morphologie et de leur profil simple au FACS montre des cellules d'une grande homogénéité. Par ailleurs, cette lignée a été choisie particulièrement parce qu'elle possède la même translocation t(9;11)(p22;q23) que les échantillons cliniques de LAM pédiatriques auxquels nous avons accès. Il est cependant probable que ces lignées cellulaires rendues « immortelles » présentent des caractéristiques génétiques et épigénétiques différentes des cellules leucémiques primaires. L'analyse du séquençage du transcriptome d'échantillons de blastes de la lignée MONO-MAC-1 afin de comparer le profil d'expression génique des HDM et des autres enzymes épigénétiques telles que les méthyltransférases par exemple et en cas de confirmation de l'expression différentielle de certaines d'entre elles, d'identifier leur gènes cibles qui pourraient être impliqués dans « l'immortalité » des blastes de lignée cellulaire en modifiant leur capacité d'apoptose ou de prolifération.

Les trois échantillons de LAM pédiatriques ont été choisis en raison de leur similarités cytogénétiques puisque les trois sont porteurs de la translocation t(9;11)(p22;q23). Ils ont également été sélectionnés en raison de leur forte infiltration tumorale (sang ou moelle), réduisant ainsi le biais lié à la présence de cellules normales dans les échantillons de ces patients. En effet, ces cellules leucémiques n'expriment pas l'antigène CD34 (tableau 9), qui, dans le cas contraire, aurait pu être utilisé comme outil de purification [160]. Il n'existe malheureusement pas de marqueur spécifique permettant d'isoler les cellules leucémiques ou même les CSL (cf. Chap. I.3.3.1).

3. Principales variations d'expression des histones déméthylases dans les cellules normales et leucémiques

L'implication de certaines HDM dans l'hématopoïèse normale et leucémique est de plus en plus reconnue dans la littérature. Dans notre projet, nous avons voulu dresser un tableau

exhaustif de l'expression des HDM dans les cellules humaines immatures normales CD34⁺, matures et anormales, afin d'identifier celles dont l'expression différentielle pourrait être suggestive de leur implication dans les processus moléculaires et cellulaires qui garantissent l'identité cellulaire.

Nous avons ainsi pu mettre en évidence que ***KDM5B/JARID1B*** est préférentiellement exprimé dans les cellules hématopoïétiques immatures humaines, tel que précédemment démontré chez la souris (cf figure 8 et 28) [90]. Cette variation de niveau d'expression selon la maturation cellulaire suggère l'implication de ce gène soit dans l'activation de la régulation des divisions d'auto-renouvellement (AR) soit dans l'inhibition des gènes impliqués dans l'engagement vers la différenciation. Dans notre étude, le profil d'expression des HDM dans les cellules de LAM par rapport aux cellules immatures CD34⁺ normales montre que ce gène est sous-exprimé dans un contexte de prolifération dérégulée et de blocage de différenciation (cf figure 33). Ce dernier point corrèle avec la mise en évidence récente du rôle de régulateur négatif de *Kdm5b/Jarid1b* sur l'activité des CSH de souris [90]. En effet dans cette étude, la diminution du niveau de *Kdm5b/Jarid1b* par RNAi augmente la prolifération des CSH *in vitro*. De plus, chez la souris, les gènes ciblés par *Kdm5b* sont des gènes liés à l'identité souche tels que les gènes *Hoxa*, avec augmentation des marques H3K4me3 dans les régions promotrices des gènes *Hoxa* suite à l'inhibition par ARN interférence de *Kdm5b/Jarid1b*. Par ailleurs, il est intéressant de noter que ***KDM1A/LSDI*** suit les mêmes variations d'expression différentielle que ***KDM5B*** dans notre étude (cf figure 33). Or ces deux enzymes ont pour cible commune H3K4: KDM5B peut déméthylater les mono-, di- ou tri-méthylation de H3K4, alors que KDM1A voit son activité limitée à la diméthylation de H3K4 (H3K4me2/1), vraisemblablement en raison de leur différents mode de déméthylation (respectivement hydroxylation ou amine oxydation)[55]. La déméthylation de H3K4 dans les régions pronotrices s'associe généralement à la diminution de la transcription des gènes. Comme on l'a vu précédemment, ***KDM1A*** est surexprimé dans les cellules de LAM [63] et il a été démontré que l'inhibition de *Kdm1a* par une stratégie utilisant les RNAi dans les cellules leucémiques avec réarrangement *MLL* favorise la différenciation [161].

Nous émettons l'hypothèse que ***KDM5B*** est un régulateur négatif des CSH dans les cellules humaines. Notre observation d'une augmentation du niveau d'expression de ***KDM5B*** et de

KDM1A dans les cellules humaines immatures normales CD34⁺ (qui ont un plus fort potentiel d'AR) par rapport aux cellules différenciées, bien que contre-intuitif, pourrait s'expliquer par le fait que la population cellulaire immature que nous avons utilisée est un ensemble de cellules CD34⁺ qui contient probablement une proportion plus importante de progéniteurs que de CSH plus quiescentes. Il est possible que les progéniteurs, en amorçant leur engagement dans le processus de différenciation, aient besoin d'augmenter l'expression de *KDM5B* pour inhiber les gènes associés à l'identité souche tels que les gènes *HOXA*. Il serait dès lors très intéressant d'évaluer l'expression de ce gène dans une population cellulaire humaine plus immature, comme les cellules CD34⁺ CD45RA⁻ CD38⁻Thy⁻. Par ailleurs, le rôle potentiellement similaire de *KDM5B/JARID1B* comme régulateur négatif dans les CSH humaines pourrait être évalué par une étude fonctionnelle d'inhibition par RNAi dans les cellules immatures humaines. Si ce rôle est confirmé, et compte tenu de l'existence d'inhibiteurs chimiques de *KDM5B* [142] l'inhibition de cette HDM pourrait être utilisée pour favoriser la prolifération *ex vivo* des CSH de sang de cordon afin d'augmenter la taille de ces greffons et les rendre disponibles pour des patients adultes. Il faudra cependant s'assurer que cette inhibition en favorisant la prolifération, ne prédispose pas les cellules à une éventuelle transformation maligne. À noter que les données chez la souris semblent montrer l'absence de transformation à long terme [90].

Les membres de la sous-famille des *KDM4* sont de plus en plus impliqués dans des pathologies malignes (cf. Chap. I.2.1.3). Dans notre projet, nous avons pu mettre en évidence une surexpression de *KDM4A* dans les cellules de LAM pédiatrique avec t(9;11) par rapport aux cellules immatures CD34⁺ normales (cf figure 33). L'analyse des données de la banque TCGA montre par ailleurs une expression élevée de *KDM4A* dans les LAM de l'adulte (cf figure 34). À notre connaissance, *KDM4A* n'a pas été identifié à ce jour comme étant dérégulé dans les LAM. Toutefois, une amplification ou une surexpression de ce gène a été décrite dans différents cancers solides, principalement dans les cancers des ovaires (associé à un mauvais pronostic), mais aussi les cancers de la tête et du cou, des poumons, du sein et du rein [134] et [67]. Bien que *KDM4A* ait récemment été proposé comme un nouvel oncogène, les mécanismes exacts par lesquels celui-ci pourrait soutenir la croissance tumorale ne sont pas encore élucidés. Une étude récente a démontré que la surexpression de *KDM4A* empêche

la réparation des cassures double-brin de l'ADN et induit une instabilité génomique en supprimant le recrutement de 53BP1 aux sites de cassure de l'ADN [162]. De plus, KDM4A interagit avec p53 et son inhibition diminue la prolifération cellulaire des cellules de la lignée de cancer du colon HCT116 [163]. Enfin la surexpression de *KDM4A* favorise l'instabilité génomique [67] et son inhibition par ARN interférent induit la sénescence dans une lignée de cellules de cancer de poumon [164].

Par ailleurs, comme nous l'avons vu dans l'introduction, KDM4A comme les autres membres de la famille JmjC sont particulièrement dépendants du métabolisme du 2-oxoglutarate (2-OG, appelé aussi α -cétoglutarate). Or les enzymes IDH (isocitrate déshydrogénases) importantes pour ce métabolisme sont fréquemment mutées dans plusieurs types de cancers solides et dans les leucémies aiguës [165]. Lorsqu'*IDH1* est muté par exemple, la production de 2-OG est réduite et il y a production d'un autre métabolite, le 2-hydroxyglutarate (2-HG) qui est un inhibiteur des KDM4. Il est possible que la surexpression de *KDM4* soit nécessaire pour compenser l'effet inhibiteur du 2-HG produit par les IDH qui sont souvent mutées [92]. D'autres études sont nécessaires pour mieux comprendre les interactions entre IDH et les KDM4. Enfin KDM4A se lie aux récepteurs aux androgènes et aux œstrogènes et les active. La diminution de *KDM4A* par ARN interférence dans les cellules de cancer du sein de la lignée MCF7 diminue la prolifération cellulaire [166]. De nombreux efforts sont actuellement faits pour développer des inhibiteurs spécifiques contre les KDM4. Compte tenu de ce qui précède et de la mise en évidence de la surexpression de *KDM4A* dans nos échantillons de LAM pédiatrique, il est possible que *KDM4A* soit un oncogène important dans la leucémogénèse des LAM pédiatriques avec réarrangement *MLL*.

Dans notre projet, nous avons pu mettre en évidence que l'expression de *KDM7A* est 8 fois moins exprimée dans les cellules immatures CD34⁺ par rapport aux cellules différenciées (figure 28), tel que démontré précédemment chez la souris (figure 8). KDM7A enlève les marques répressives de la transcription H3K9me et H3K27me. Ce gène n'est que peu ou pas exprimé dans nos échantillons de cellules leucémiques pédiatriques ou de lignées, de même que dans ceux étudiés dans le cadre du projet TCGA. Une étude récente a identifié ce gène comme un suppresseur de tumeur, du fait de son effet inhibiteur sur la croissance tumorale par réduction de l'angiogénèse dans des conditions de limitation des apports nutritifs pour

des cancers solides [94]. Les résultats dans notre étude plaident également en faveur d'un effet suppresseur de tumeur de *KDM7A*, qui semble activer les programmes transcriptionnels qui favorisent la différenciation. Il serait dès lors important d'identifier ses gènes cibles.

Deux autres histones déméthylases, *PADI2* et *PADI4*, ont une expression fortement diminuée dans le compartiment cellulaire immature CD34⁺ comparé aux cellules différenciées (tableau 11). Ceci concorde avec les données de la littérature qui ont montré que l'expression de *PADI4* augmente avec la différenciation myéloïde des cellules leucémiques HL60 traitées notamment par l'acide rétinoïque [78]. A l'inverse, chez la souris, la surexpression de *PADI4* a récemment été mise en évidence dans des cellules très immatures de type LSK ou lineage⁻ Sca-1⁺ c-Kit⁺ (dont la fréquence est de environ une CSH pour 40) [98]. Cette dernière étude suggère que *PADI4* régule la prolifération des progéniteurs multipotents en contrôlant l'expression de c-myc. Il sera intéressant dans le futur de déterminer la fraction cellulaire spécifique dans lequel l'expression de *PADI4* est enrichie, tant chez l'homme que la souris, et d'étudier fonctionnellement le rôle de cette enzyme dans ce compartiment spécifique.

Enfin, nous avons pu observer que l'expression de *PADI2* est 4 fois plus marquée dans les cellules leucémiques que dans les cellules normales immatures CD34⁺. *PADI2* est encore très peu étudié. Il a néanmoins récemment été trouvé surexprimé dans les cancers du sein HER2/ERBB2+ [77]. Le mécanisme par lequel *PADI2* soutient la croissance tumorale dans ces cas n'est pas élucidé, mais son implication dans la tumorigenèse est démontrée par le fait que son inhibition par le Cl-amidine permet de diminuer la croissance tumorale. Nous suspectons dès lors que *PADI2* pourrait se comporter comme un oncogène dans les LAM pédiatriques.

Dans notre étude, une grande partie des HDM est exprimée de façon similaire entre les cellules immatures et les cellules différenciées. Ces enzymes pourraient dès lors avoir un rôle important tout au long de l'hématopoïèse normale et devraient éventuellement être explorées fonctionnellement, particulièrement si elles ont été impliquées dans des pathologies hématologiques. Par exemple *KDM2B* dans notre étude est bien exprimé dans les cellules normales immatures CD34⁺ et différenciées, et est également fortement exprimé dans les

cellules leucémiques. *KDM2B*, récemment impliqué dans la régulation de la prolifération cellulaire hématopoïétique et dans la sénescence, est fréquemment surexprimé dans les LAM et semble être un oncogène dans ces leucémies [64]. Un autre exemple pourrait être *KDM6A* qui est muté dans des leucémies myélo-monocytaires chroniques, laissant suspecter que l'inactivation de ce gène pourrait être impliquée dans la leucémogénèse [74]. D'autre part, dans notre étude, *KDM6A* est bien exprimé dans les compartiments cellulaires normaux immatures et différenciés, et est également bien exprimé dans nos échantillons de cellules leucémiques ainsi que dans les données de la banque TCGA. Il a été récemment démontré que cette HDM, qui cible le résidu H3K27me2/3, s'associe aux promoteurs de gènes clés tels que *MLL1*, *RUNX1* et *SCL* et fait varier leur expression et que sa déplétion diminue la prolifération cellulaire de différentes lignées leucémiques [167].

4. Patron de méthylation des histones et hématopoïèse : lien avec le destin cellulaire ?

Il est actuellement démontré qu'il existe des patrons de méthylation de l'ADN spécifiques à un cancer donné [168], et que les tumeurs peuvent être regroupées selon cette signature épigénétique. Le patron de modifications covalentes de la chromatine, dont la méthylation des résidus d'histone, est encore mal connu à l'heure actuelle et reste à explorer, particulièrement dans le système hématopoïétique. Cependant, un des meilleurs exemples de l'importance des patrons de méthylation des histones en lien avec la régulation de l'hématopoïèse normale et leucémique est sans doute l'antagonisme des marques de méthylation des H3K4 (activation) / H3K27 (répression), apposées par les méthyltransférases MLL et EZH2 au sein de leur complexes respectifs, Trx et PRC2[58], au niveau des loci codant pour des régulateurs du destin cellulaire, tel qu'évoqué plus haut (cf. Chap. I. 4.2.2.3.2).

Tout d'abord, MLL est une méthyl-transférase des résidus H3K4 dont l'implication dans l'hématopoïèse normale et la leucémogénèse est bien établie [56, 125, 169]. Lorsque *MLL* est muté, le patron de méthylation de H3K4 est modifié au niveau des loci des gènes associés à l'identité souche tels que les gènes *Hoxa7* et *Hoxa9* [129] qui sont alors surexprimés. Par ailleurs, il a été récemment démontré que EZH2, méthylase des résidus H3K27, inhibe les programmes de différenciation dans les cellules de LAM [170]. Il est aussi prouvé que

certaines mutations des histones elles-mêmes, tel le résidu H3.3K27 ou H3. 1K27, inhibent le complexe PRC2 via leur interaction avec EZH2 [171] et conduisent à une réduction globale de la méthylation des résidus H3K27. Enfin, *KDM6A*, gène codant pour une déméthylase qui cible H3K27me_{2/3}, est fréquemment retrouvé muté dans les LAM pédiatriques comme nous l'avons évoqué plus haut [172] (cf Chap. IV. 3). Peu d'études ont exploré la relation entre le patron de méthylation de certains résidus d'histone et les modifications de profils d'expression de gènes clefs de l'hématopoïèse humaine dans des cellules primaires normales ou leucémiques. Certaines études ont cependant observé des différences de méthylation de H3K9 ou de H3K4/H3K27 au niveau de loci spécifiques [173, 174]. Notre équipe a pu récemment démontrer chez la souris que l'inhibition de *Kdm5b* par une stratégie d'ARN – interférence augmente l'expression de certains régulateurs des CSH (*Hoxa7*, *Hoxa9*, *Hoxa10*, *Hes1*), et que les promoteurs de ces gènes harбораient un enrichissement pour la marque activatrice H3K4me₃, laissant suggérer une régulation épigénétique de ces loci par KDM5B[90].

Ainsi les modifications covalentes post-traductionnelles des histones semblent associées au destin cellulaire du système hématopoïétique. Il serait dès lors intéressant d'observer la distribution génomique de la méthylation des résidus H3K4 et H3K27 dans les CSH humaines et dans des cellules de LAM par une stratégie de Chip-seq afin de mieux comprendre l'impact de la régulation épigénétique via la méthylation/déméthylation de ces deux résidus d'histone, et d'identifier un réseau de gènes clés en lien avec le destin cellulaire normal, ou leucémique lorsque perturbé.

5. Intérêt et limites des études d'expression

Des différentes méthodes disponibles pour l'analyse du transcriptome (cf. Chap. I.5), nous avons choisi d'utiliser dans ce projet la qRT-PCR et le séquençage de l'ARN.

La **qRT-PCR** est un outil pratique et fréquemment utilisé du fait de sa sensibilité et de la rapidité d'obtention des résultats. Moyennant une préparation adéquate et minutieuse des échantillons et l'élaboration de tests spécifiques aux gènes cibles, elle permet de mesurer un nombre important de gènes différents. Parmi les étapes critiques de préparation, le choix du

gène de contrôle endogène revêt une importance particulière car celui-ci sert pour la normalisation des résultats et l'élimination de différents biais (cf. Chap.I.5.3). Il doit être non seulement suffisamment exprimé mais aussi l'être de façon uniforme dans les différentes populations cellulaires analysées. Une étude récente a permis d'établir quels sont les meilleurs gènes de contrôle pour les études d'expression dans les cellules hématopoïétiques. En utilisant les données d'expression par RNA-seq de 55 échantillons leucémiques de patients, les auteurs ont identifié les gènes dont l'expression est la plus uniforme dans ces échantillons. Ces gènes candidats ont ensuite été validés d'abord en analysant leur expression dans 1933 autres tissus normaux et tumoraux, puis en les testant par qRT-PCR[158]. Le gène *TBP* que nous avons utilisé dans notre projet (cf Chap.II. 6.4), est, dans cette étude, un de ceux qui varient le moins parmi les gènes classiquement utilisés comme contrôle endogène. De plus, les auteurs ont identifié les gènes *HNRNPL*, *EIF4H* et *PSMA*, comme de meilleurs choix de gènes contrôles pour la qRT-PCR. Ceux-ci seront donc choisis lors des prochaines analyses qui seront réalisées dans le cadre de ce projet de recherche. Bien que la qRT-PCR présente comme inconvénient principal le besoin de déterminer a priori la liste des gènes dont on veut quantifier l'expression, sa sensibilité et son amplitude de détection en font un outil performant, même lorsque la quantité d'ARN est très faible. Notre expérience confirme la grande sensibilité de cette technique: en effet, elle permet la détection de l'expression de certains gènes tels que *KDM4D* ou *KDM4E(DL)* dans les cellules normales différenciées alors que le RNA-seq ne le permet pas (figure 26).

Le séquençage de l'ARN, en plus de la quantification du niveau d'expression des gènes, offre de nombreuses informations additionnelles (cf. chap. I.5.4). Cette polyvalence en fait un choix stratégique avantageux, particulièrement si les échantillons testés ne sont pas reproductibles comme par exemple les échantillons tumoraux de patients et pour la constitution de banques de données telles que le TCGA ou le PCGP par exemple (cf Chap.I.5.5). Cependant le coût actuel de cette technique et l'expertise nécessaire pour l'analyse des résultats générés sont, à ce stade, un frein pour son utilisation. La baisse du coût de cette technique, une meilleure accessibilité des plateformes de séquençage de nouvelle génération et d'analyse bioinformatique, ainsi que l'évolution du RNA-seq qui permettra bientôt de faire l'analyse complète du transcriptome de populations cellulaires très

petites, voire même d'une seule cellule[175] en font un outil incontournable en recherche mais aussi en clinique.

Enfin, nous pouvons considérer que les résultats des études d'expression par qRT-PCR et par RNA-seq concordent. En effet, dans l'étude de Macrae *et al.* [158], une bonne corrélation entre RPKM et dCt pour le gène codant pour CD33 a été établie dans les cellules hématopoïétiques, et ce particulièrement si on utilise *PSMA1*, *SF3B2* et *EIF4H* comme gènes contrôles. Une autre étude [149] a démontré des résultats comparables lors de l'analyse du transcriptome par les trois techniques (qRT-PCR, RNA-seq et puces à ADN) d'un ARN de référence pour les contrôles de qualité des puces à ADN (« MAQC » ou « Microarray Quality Control »). Nous avons pu constater dans la plupart de nos analyses que les résultats obtenus par qRT-PCR ou RNA-seq sont comparables. Ainsi, selon notre expérience et les données de la littérature, la qRT-PCR devrait être utilisée comme méthode de contrôle pour les gènes cibles identifiés par la technique de RNA-seq lors de l'élaboration d'un profil d'expression d'un nombre important de gènes. De plus notre étude, en combinant ces deux techniques nous a permis de dresser un portrait réaliste du transcriptome dans les cellules hématopoïétiques normales et leucémiques. Ce crible extensif constitue une première étape d'identification des gènes candidats qui seront explorés fonctionnellement dans un deuxième temps.

À notre connaissance, il n'existe pas d'analyse du profil d'expression spécifique et exhaustif des histones déméthylases dans les cellules hématopoïétiques normales et leucémiques. Il existe cependant plusieurs sources d'information générale disponibles publiquement. Par exemple, les sites proposant l'exploration de données d'expression appelés HemaExplorer ou le site bioGPS permettent d'avoir rapidement une vision du taux d'expression d'un gène dans différents tissus et offre même la mise en parallèle des données murines et humaines[159, 176] Toutefois, dans ce type de collection de données, il faut noter que les populations cellulaires étudiées sont souvent des tissus complets plutôt que des populations cellulaires triées. Les données disponibles issues des banques telles que celle du TCGA ou des différentes études publiées dans la littérature, présentent les résultats d'analyses complètes du transcriptome de LAM mais mettent rarement en parallèle ces données avec celles de

cellules hématopoïétiques normales. De plus, peu d'études ciblent spécifiquement les LAM pédiatriques [177, 178].

Notre étude comparative, en montrant une expression différentielle de certaines HDM dans les cellules immatures normales CD34⁺, et d'autres dans les cellules leucémiques, suggère ainsi qu'elles soient impliquées dans l'activation de voies biologiques critiques pour l'hématopoïèse normale ou pour la leucémogénèse. Cependant, les variations du niveau d'expression d'un compartiment cellulaire à l'autre ne sont pas pour autant systématiquement synonymes de fonction déterminante. Et inversement, l'absence de différence de niveau d'expression entre les compartiments cellulaires ne signifie aucunement que le gène considéré n'est pas essentiel à l'hématopoïèse. Ces études sont cependant très importantes car elles dressent un portrait dynamique de l'expression des gènes dans les cellules hématopoïétiques à différentes étapes de l'hématopoïèse. Elles permettent également d'identifier des cibles moléculaires potentielles pour les études fonctionnelles, qui pourraient ainsi devenir de nouvelles cibles thérapeutiques.

6. Perspectives ouvertes par notre étude d'expression des HDM

Cette étude du profil d'expression des HDM dans différentes populations cellulaires hématopoïétiques normales et leucémiques, en mettant en évidence l'expression différentielle de certaines d'entre elles, permet de sélectionner les gènes candidats qui devront être explorés fonctionnellement. Les expériences à réaliser en priorité seront:

1. L'étude du profil d'expression de gènes candidats identifiés dans cette thèse dans des populations cellulaires hautement enrichies en cellules souches hématopoïétiques. La purification de populations cellulaires plus enrichies en CSH permettra de réaliser le profil d'expression de certaines HDM telles que *KDM5B* ou *KDM4A* sur une population plus homogène. En effet l'utilisation des marqueurs cellulaires Lin⁻ CD34⁺ CD38⁻ Thy⁺ CD49f⁺ CD45RA⁻ Rho^{lo} pour purifier les CSPH permettent d'augmenter la fréquence des CSH jusqu'à environ 1:5 [14] vs 1:1000 pour les cellules CD34⁺ de sang de cordon (cf. Chap. I. 3.3.3). Cette stratégie est réservée à un nombre plus limité de gènes candidats en

raison du très grand nombre d'unités de sang de cordons requises pour isoler un nombre suffisant de cellules (au mieux 50 cellules $\text{Lin}^- \text{CD34}^+ \text{CD38}^- \text{CD90}^+ \text{CD49f}^+ \text{CD45RA}^- \text{Rhodamine}^{\text{lo}}$ sont isolées par unité de sang de cordon). On pourra quantifier la fréquence des CSH dans ces populations en utilisant la méthode des essais de dilutions limites (cf. Chap.1.3.3.1).

2. L'étude de l'expression de *KDM4A* dans une série plus large de LAM pédiatriques permettra de confirmer sa surexpression dans ce type de leucémie avec réarrangement *MLL* ainsi que dans les différents autres sous-types morphologiques et génétiques de leucémie.
3. Des études fonctionnelles de certains gènes candidats permettront de déterminer leur rôle dans la modulation du destin des cellules souches normales ou leucémiques. Un vecteur lentiviral sera utilisé pour introduire dans une population cellulaire enrichie en CSH soit un régulateur positif ou négatif de l'hématopoïèse, en utilisant une stratégie de surexpression du gène candidat (introduction d'ADNc) ou de répression de l'expression du gène candidat (inhibition par ARN-interférence ou RNAi). Ainsi l'inhibition de *KDM5B* par RNAi dans les CSH pourrait favoriser leur expansion, comme c'est le cas chez la souris [90]. L'inhibition de *KDM4A* dans des cellules leucémiques (lignée cellulaire ou modèle murin) par la même stratégie pourrait diminuer la prolifération des blastes. La surexpression de *KDM4A* dans des CSH pourrait, au contraire, augmenter leur potentiel de prolifération, voire leur conférer un potentiel de transformation. L'utilisation d'inhibiteurs chimiques ciblant spécifiquement les HDM, lorsque disponibles, pourra aussi être envisagée pour confirmer les résultats des études par RNAi.
4. Éventuellement, l'identification des cibles directes des HDM qui démontreront un impact fonctionnel dans les CSH pourra être réalisée en utilisant une stratégie d'immunoprécipitation de la chromatine en utilisant un anticorps dirigé contre une HDM elle-même (par exemple *KDM5B*) ou contre un panel de marqueurs épigénétiques (par exemple H3K4me3), suivie de séquençage de nouvelle génération (ChIP-seq) pour identifier les gènes cibles.

Ceci permettra de mieux comprendre les bases moléculaires qui sous-tendent l'auto-renouvellement des cellules souches normales, et leur dérégulation dans la leucémogénèse.

7. Remarques conclusives sur les mutations de *GATA2*

La mutation germinale hétérozygote R396W mise en évidence chez le patient est localisée dans la région codant pour le 2^{ème} domaine en doigt de zinc (C-terminal) de la protéine (cf. Chap. II. 11.1 figure 17), région du gène *GATA2* qui est la plus fréquemment mutée. Cette mutation a été décrite chez deux autres patients dans la littérature et dans chaque cas, elle était associée à un syndrome de MonoMAC [157, 179, 180]. De part sa localisation dans le 2^{ème} domaine en doigt de zinc, cette anomalie pourrait perturber l'interaction de *GATA2* avec l'ADN ou avec les autres protéines effectrices tel que mis en évidence pour d'autres mutations localisées dans le ZF2 [108, 181, 182]. Le phénotype suspecté est alors dominant négatif. Les seules mutations qui ont été fonctionnellement étudiées à ce jour sont les mutations T354M et M355del qui conduisent à une perte de fonction [108]. En revanche, les anomalies de type délétions et insertions localisées dans la région N-terminale de la protéine sont plutôt susceptibles de conduire à la formation d'un allèle nul par dégradation de l'ARNm non-sens [183] et donc d'induire une perte de fonction par haplo-insuffisance. Des mutations introniques de *GATA2* ont aussi été décrites [184-186]. Enfin, de façon générale, les mutations de *GATA2* décrites semblent conduire à une perte de fonction avec haplo-insuffisance plutôt qu'à un phénotype dominant négatif.

À ce jour, des mutations germinales causant des syndromes de prédisposition familiale aux SMD et LAM ont été décrites dans un nombre limité de gènes, tels que *RUNX1* [105] et *CEBPA* [106]. Il est intéressant de noter que pour ces gènes également, des mutations différentes, responsables d'une perte de fonction par dominance négative ou par haplo-insuffisance conduisent au développement d'un SMD ou d'une LAM, même si le phénotype associé initialement est variable pour une même mutation entre les individus porteurs au sein d'une même famille. L'absence de corrélation entre génotype et phénotype (y compris dans l'âge de début de la maladie) n'est pas encore clairement expliquée. Elle pourrait résulter de l'effet de la mutation elle-même associé à l'effet d'anomalies additionnelles et/ou à des éléments de l'environnement (infection et stress).

Par ailleurs, des mutations ponctuelles et des délétions de *GATA2* ont également été impliquées dans l'acutisation blastique des leucémies myéloïdes chroniques chez différents patients et l'une d'elle, L359V, a été étudiée fonctionnellement [187]. Ces anomalies

génétiques étaient localisées dans le ZF1 ou le ZF2, et résultaient plutôt en un gain de fonction par une augmentation de l'activation transcriptionnelle de gènes cibles connus (*EpoR*, *EDN-1* et *P-selectin*) mais aussi par une augmentation de l'activité inhibitrice sur *PU.1/SPI.1*, gène important pour les CSH [187], résultant en un gain de fonction. De plus, une surexpression de *GATA2* a été décrite récemment dans les LAM associée à un mauvais pronostic [188]. Il a été démontré que la surexpression de *GATA2* favorise la quiescence cellulaire dans les CSH normales et dans les cellules leucémiques, et favorise ainsi la résistance des blastes aux traitements [189, 190].

Il est intéressant de noter qu'à deux reprises des mutations de *GATA2* ont été associées à une autre mutation dans un gène impliqué dans les SMD ou les LAM associées aux SMD: une mutation somatique de *GATA2* associée à une mutation germinale héréditaire de *CEBPA* [191] et une mutation germinale héréditaire de *GATA2* associée à une mutation somatique d'*ASXL1* chez deux frères [192]. Ces associations de mutations dans des gènes critiques soulèvent l'importante question de la chronologie des insultes génétiques et de leur responsabilité dans l'évolution défavorable de certains patients. Il nous semble important de souligner que l'association de deux mutations est possible et qu'elles doivent dès lors être recherchées.

Sur le plan hématologique, et bien que l'importance de *GATA2* dans l'hématopoïèse normale était déjà connue [193], c'est l'analyse par cytométrie de flux des cellules CD34⁺ de la moelle de patients atteints du syndrome de MonoMAC qui a permis de mettre en évidence que l'arrêt de maturation associé aux mutations de *GATA2* survenait à une étape très précoce de l'hématopoïèse, entre les progéniteurs multipotents et les progéniteurs lymphoïdes communs (figure 34 C) [6, 194] illustrant l'implication de ce facteur de transcription à une étape initiale de l'hématopoïèse.

En conclusion, ces études prouvent que *GATA2* est un facteur de transcription très important de l'hématopoïèse, et que son action est non seulement tributaire de sa bonne fonction moléculaire mais aussi de son niveau d'expression finement régulé.

Enfin, cette histoire clinique illustre bien, pour des jeunes patients atteint de SMD ou de LAM, l'importance de prendre en compte le fait que ces maladies peuvent être dues à une mutation germinale héréditaire. Cette information revêt une importance cruciale pour choisir

le meilleur donneur de CSH en vue d'une greffe, seul traitement curatif avec un pronostic généralement bon. Ce cas clinique illustre aussi la pertinence et la faisabilité de développer rapidement un test diagnostique pour détecter une nouvelle mutation.

CHAPITRE V: CONCLUSION

L'étude de la biologie de ces CSH permet de mieux comprendre les processus moléculaires qui déterminent le destin cellulaire tels que l'auto-renouvellement des cellules souches ou leur entrée en différenciation. Certains de ces processus, lorsque dysfonctionnels, sont impliqués dans la leucémogénèse. Leur étude est dès lors essentielle pour découvrir de nouvelles cibles thérapeutiques. Les histones déméthylases (HDM) sont une nouvelle classe d'enzymes épigénétiques qui participent à la régulation des programmes transcriptionnels qui soutiennent ces processus dans les CSH. Certaines d'entre elles sont de plus associées à des hémopathies malignes. Notre étude dresse un portrait exhaustif de l'expression comparée des HDM dans les cellules immatures normales CD34⁺, les cellules différenciées et les cellules de LAM. Nous avons d'abord démontré que la majorité de ces enzymes sont bien exprimées dans les cellules hématopoïétiques et aussi que cinq d'entre elles ont une expression différentielle entre les compartiments cellulaires.

Ainsi *KDM5B* et *KDM1A*, qui ont pour cible commune H3K4me2/1, marque activatrice de la transcription de gènes clés, sont respectivement 3 fois et 2,5 fois plus exprimées dans les cellules immatures CD34⁺. Ces enzymes sont aussi respectivement 3 fois et 2 fois moins exprimées dans les cellules leucémiques. Ces résultats concordent avec ceux d'une étude similaire réalisée chez la souris [90]. Ils laissent donc suspecter que *KDM5B* puisse être également un régulateur négatif des CSH humaines. Ce rôle sur la régulation de l'auto-renouvellement devra être validé fonctionnellement dans les CSH humaines. S'il est confirmé, l'inhibition de *KDM5B* pourrait être utilisée en clinique pour favoriser la prolifération CSH et augmenter le nombre de cellules des greffons. Notre projet a également permis de mettre en évidence une surexpression de *KDM4A* dans les cellules de LAM pédiatrique. Ces résultats appuient l'idée que *KDM4A* comme les autres membres de la sous-famille *KDM4A*, est un pro-oncogène et pourrait être une nouvelle cible thérapeutique dans les LAM de l'enfant, si les études fonctionnelles d'inhibition dans les cellules leucémiques montrent une diminution de leur prolifération. Dans notre étude, *KDM7A* est fortement diminué dans les cellules immatures CD34⁺ par rapport aux cellules différenciées

mais exprimé de façon similaire dans les blastes et les cellules immatures. Enfin, notre étude montre également que l'expression de *PADI2* est plus marquée dans les cellules leucémiques que dans les cellules normales immatures CD34⁺. *PADI2* pourrait avoir un rôle d'oncogène exploitable comme cible thérapeutique si les études fonctionnelles d'inhibition dans des cellules de lignées leucémiques ou des modèles murins de leucémies humaines montrent un effet favorable.

Ainsi cette étude exhaustive de l'expression des HDM dans les cellules hématopoïétiques normales et anormales a permis d'identifier plusieurs d'entre elles comme potentiels régulateurs du destin des CSH dont la fonction sera explorée afin de déterminer si elles peuvent être considérées comme cible thérapeutique soit pour l'expansion des CSH de greffons soit pour le traitement des LAM de l'enfant.

BIBLIOGRAPHIE

1. Fey, M.F., *Normal and malignant hematopoiesis*. Ann Oncol, 2007. **18 Suppl 1**: p. i9-i13.
2. Tough, D.F. and J. Sprent, *Life span of naive and memory T cells*. Stem Cells, 1995. **13**(3): p. 242-9.
3. Mikkola, H.K. and S.H. Orkin, *The journey of developing hematopoietic stem cells*. Development, 2006. **133**(19): p. 3733-44.
4. Orkin, S.H. and L.I. Zon, *Hematopoiesis: an evolving paradigm for stem cell biology*. Cell, 2008. **132**(4): p. 631-44.
5. Krause, D.S. and D.T. Scadden, *Deconstructing the complexity of a microenvironmental niche*. Cell, 2012. **149**(1): p. 16-7.
6. Doulatov, S., et al., *Hematopoiesis: a human perspective*. Cell Stem Cell, 2012. **10**(2): p. 120-36.
7. Purton, L.E. and D.T. Scadden, *Limiting factors in murine hematopoietic stem cell assays*. Cell Stem Cell, 2007. **1**(3): p. 263-70.
8. Till, J.E. and C.E. Mc, *A direct measurement of the radiation sensitivity of normal mouse bone marrow cells*. Radiat Res, 1961. **14**: p. 213-22.
9. Civin, C.I., et al., *Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells*. J Immunol, 1984. **133**(1): p. 157-65.
10. Wang, J., et al., *SCID-repopulating cell activity of human cord blood-derived CD34-cells assured by intra-bone marrow injection*. Blood, 2003. **101**(8): p. 2924-31.
11. Mayani, H., W. Dragowska, and P.M. Lansdorp, *Characterization of functionally distinct subpopulations of CD34+ cord blood cells in serum-free long-term cultures supplemented with hematopoietic cytokines*. Blood, 1993. **82**(9): p. 2664-72.
12. Bhatia, M., et al., *Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice*. Proc Natl Acad Sci U S A, 1997. **94**(10): p. 5320-5.
13. Murray, L., et al., *Enrichment of human hematopoietic stem cell activity in the CD34+Thy-1+Lin- subpopulation from mobilized peripheral blood*. Blood, 1995. **85**(2): p. 368-78.
14. Notta, F., et al., *Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment*. Science, 2011. **333**(6039): p. 218-21.
15. Bhatia, M., et al., *A newly discovered class of human hematopoietic cells with SCID-repopulating activity*. Nat Med, 1998. **4**(9): p. 1038-45.
16. Anjos-Afonso, F., et al., *CD34(-) cells at the apex of the human hematopoietic stem cell hierarchy have distinctive cellular and molecular signatures*. Cell Stem Cell, 2013. **13**(2): p. 161-74.
17. Wang, J.C., M. Doedens, and J.E. Dick, *Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay*. Blood, 1997. **89**(11): p. 3919-24.
18. Shepherd, B.E., et al., *Estimating human hematopoietic stem cell kinetics using granulocyte telomere lengths*. Exp Hematol, 2004. **32**(11): p. 1040-50.

19. Catlin, S.N., et al., *The replication rate of human hematopoietic stem cells in vivo*. Blood, 2011. **117**(17): p. 4460-6.
20. Martinez-Agosto, J.A., et al., *The hematopoietic stem cell and its niche: a comparative view*. Genes Dev, 2007. **21**(23): p. 3044-60.
21. Hatzimichael, E. and M. Tuthill, *Hematopoietic stem cell transplantation*. Stem Cells Cloning, 2010. **3**: p. 105-117.
22. Sugiyama, T., et al., *Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches*. Immunity, 2006. **25**(6): p. 977-88.
23. Purton, L.E. and D.T. Scadden, *The hematopoietic stem cell niche*, in *StemBook*. 2008, : 2008 Louise E. Purton and David T. Scadden.: Cambridge MA.
24. Krause, D.S., *Regulation of hematopoietic stem cell fate*. Oncogene, 2002. **21**(21): p. 3262-9.
25. Wilson, N.K., et al., *Combinatorial transcriptional control in blood stem/progenitor cells: genome-wide analysis of ten major transcriptional regulators*. Cell Stem Cell, 2010. **7**(4): p. 532-44.
26. Krivtsov, A.V., et al., *Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9*. Nature, 2006. **442**(7104): p. 818-22.
27. Krumlauf, R., *Hox genes in vertebrate development*. Cell, 1994. **78**(2): p. 191-201.
28. Argiropoulos, B. and R.K. Humphries, *Hox genes in hematopoiesis and leukemogenesis*. Oncogene, 2007. **26**(47): p. 6766-76.
29. Cellot, S., et al., *Sustained in vitro trigger of self-renewal divisions in Hoxb4hiPbx1(10) hematopoietic stem cells*. Exp Hematol, 2007. **35**(5): p. 802-16.
30. Galloway, J.L., et al., *Loss of gata1 but not gata2 converts erythropoiesis to myelopoiesis in zebrafish embryos*. Dev Cell, 2005. **8**(1): p. 109-16.
31. Tsai, F.Y. and S.H. Orkin, *Transcription factor GATA-2 is required for proliferation/survival of early hematopoietic cells and mast cell formation, but not for erythroid and myeloid terminal differentiation*. Blood, 1997. **89**(10): p. 3636-43.
32. Takahashi, K. and S. Yamanaka, *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors*. Cell, 2006. **126**(4): p. 663-76.
33. Ringrose, L. and R. Paro, *Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins*. Annu Rev Genet, 2004. **38**: p. 413-43.
34. Kouzarides, T., *Chromatin modifications and their function*. Cell, 2007. **128**(4): p. 693-705.
35. Barth, T.K. and A. Imhof, *Fast signals and slow marks: the dynamics of histone modifications*. Trends Biochem Sci, 2010. **35**(11): p. 618-26.
36. Zhou, V.W., A. Goren, and B.E. Bernstein, *Charting histone modifications and the functional organization of mammalian genomes*. Nat Rev Genet, 2011. **12**(1): p. 7-18.
37. Jaenisch, R. and A. Bird, *Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals*. Nat Genet, 2003. **33** Suppl: p. 245-54.
38. Broske, A.M., et al., *DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction*. Nat Genet, 2009. **41**(11): p. 1207-15.
39. Ley, T.J., et al., *DNMT3A mutations in acute myeloid leukemia*. N Engl J Med, 2010. **363**(25): p. 2424-33.

40. Walter, M.J., et al., *Recurrent DNMT3A mutations in patients with myelodysplastic syndromes*. *Leukemia*, 2011. **25**(7): p. 1153-8.
41. Trowbridge, J.J., et al., *Haploinsufficiency of Dnmt1 impairs leukemia stem cell function through derepression of bivalent chromatin domains*. *Genes Dev*, 2012. **26**(4): p. 344-9.
42. Meissner, A., et al., *Genome-scale DNA methylation maps of pluripotent and differentiated cells*. *Nature*, 2008. **454**(7205): p. 766-70.
43. Tahiliani, M., et al., *Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1*. *Science*, 2009. **324**(5929): p. 930-5.
44. Wu, H. and Y. Zhang, *Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation*. *Genes Dev*, 2011. **25**(23): p. 2436-52.
45. Delhommeau, F., et al., *Mutation in TET2 in myeloid cancers*. *N Engl J Med*, 2009. **360**(22): p. 2289-301.
46. Solary, E., et al., *The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases*. *Leukemia*, 2013.
47. Madzo, J., et al., *Hydroxymethylation at gene regulatory regions directs stem/early progenitor cell commitment during erythropoiesis*. *Cell Rep*, 2014. **6**(1): p. 231-44.
48. Moran-Crusio, K., et al., *Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation*. *Cancer Cell*, 2011. **20**(1): p. 11-24.
49. Ko, M., et al., *Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice*. *Proc Natl Acad Sci U S A*, 2011. **108**(35): p. 14566-71.
50. Li, Z., et al., *Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies*. *Blood*, 2011. **118**(17): p. 4509-18.
51. Quivoron, C., et al., *TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis*. *Cancer Cell*, 2011. **20**(1): p. 25-38.
52. Jan, M., et al., *Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia*. *Sci Transl Med*, 2012. **4**(149): p. 149ra118.
53. Rowley, J.D., *The der(11) chromosome contains the critical breakpoint junction in the 4;11, 9;11, and 11;19 translocations in acute leukemia*. *Genes Chromosomes Cancer*, 1992. **5**(3): p. 264-6.
54. Shi, Y., et al., *Histone demethylation mediated by the nuclear amine oxidase homolog LSD1*. *Cell*, 2004. **119**(7): p. 941-53.
55. Cloos, P.A., et al., *Erasing the methyl mark: histone demethylases at the center of cellular differentiation and disease*. *Genes Dev*, 2008. **22**(9): p. 1115-40.
56. McMahan, K.A., et al., *Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal*. *Cell Stem Cell*, 2007. **1**(3): p. 338-45.
57. Radulovic, V., G. de Haan, and K. Klauke, *Polycomb-group proteins in hematopoietic stem cell regulation and hematopoietic neoplasms*. *Leukemia*, 2013. **27**(3): p. 523-33.
58. Mills, A.A., *Throwing the cancer switch: reciprocal roles of polycomb and trithorax proteins*. *Nat Rev Cancer*, 2010. **10**(10): p. 669-82.
59. Bernstein, B.E., et al., *Methylation of histone H3 Lys 4 in coding regions of active genes*. *Proc Natl Acad Sci U S A*, 2002. **99**(13): p. 8695-700.

60. Heintzman, N.D., et al., *Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome*. Nat Genet, 2007. **39**(3): p. 311-8.
61. Narita, M., et al., *Rb-mediated heterochromatin formation and silencing of E2F target genes during cellular senescence*. Cell, 2003. **113**(6): p. 703-16.
62. Kooistra, S.M. and K. Helin, *Molecular mechanisms and potential functions of histone demethylases*. Nat Rev Mol Cell Biol, 2012. **13**(5): p. 297-311.
63. Lokken, A.A. and N.J. Zeleznik-Le, *Breaking the LSD1/KDM1A addiction: therapeutic targeting of the epigenetic modifier in AML*. Cancer Cell, 2012. **21**(4): p. 451-3.
64. He, J., A.T. Nguyen, and Y. Zhang, *KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia*. Blood, 2011. **117**(14): p. 3869-80.
65. Lai, F., et al., *cDNA cloning and genomic structure of three genes localized to human chromosome band 5q31 encoding potential nuclear proteins*. Genomics, 2000. **70**(1): p. 123-30.
66. Hu, Z., et al., *A novel nuclear protein, 5qNCA (LOC51780) is a candidate for the myeloid leukemia tumor suppressor gene on chromosome 5 band q31*. Oncogene, 2001. **20**(47): p. 6946-54.
67. Black, J.C., et al., *KDM4A lysine demethylase induces site-specific copy gain and rereplication of regions amplified in tumors*. Cell, 2013. **154**(3): p. 541-55.
68. Vinatzer, U., et al., *Mucosa-associated lymphoid tissue lymphoma: novel translocations including rearrangements of ODZ2, JMJD2C, and CNN3*. Clin Cancer Res, 2008. **14**(20): p. 6426-31.
69. van Zutven, L.J., et al., *Identification of NUP98 abnormalities in acute leukemia: JARID1A (12p13) as a new partner gene*. Genes Chromosomes Cancer, 2006. **45**(5): p. 437-46.
70. Streubel, B., et al., *Correlation of cytogenetic, molecular cytogenetic, and clinical findings in 59 patients with ANLL or MDS and abnormalities of the short arm of chromosome 12*. Br J Haematol, 1998. **100**(3): p. 521-33.
71. Karrman, K., et al., *Cytogenetic evolution patterns in CML post-SCT*. Bone Marrow Transplant, 2007. **39**(3): p. 165-71.
72. Bacher, U., et al., *Inverted duplication dup(1)(q32q21) as sole aberration in lymphoid and myeloid malignancies*. Cancer Genet Cytogenet, 2009. **188**(2): p. 108-11.
73. van Haaften, G., et al., *Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer*. Nat Genet, 2009. **41**(5): p. 521-3.
74. Jankowska, A.M., et al., *Mutational spectrum analysis of chronic myelomonocytic leukemia includes genes associated with epigenetic regulation: UTX, EZH2, and DNMT3A*. Blood, 2011. **118**(14): p. 3932-41.
75. Anderton, J.A., et al., *The H3K27me3 demethylase, KDM6B, is induced by Epstein-Barr virus and over-expressed in Hodgkin's Lymphoma*. Oncogene, 2011. **30**(17): p. 2037-43.
76. Agger, K., et al., *The H3K27me3 demethylase JMJD3 contributes to the activation of the INK4A-ARF locus in response to oncogene- and stress-induced senescence*. Genes Dev, 2009. **23**(10): p. 1171-6.

77. McElwee, J.L., et al., *Identification of PADI2 as a potential breast cancer biomarker and therapeutic target*. BMC Cancer, 2012. **12**: p. 500.
78. Nakashima, K., et al., *Molecular characterization of peptidylarginine deiminase in HL-60 cells induced by retinoic acid and 1alpha,25-dihydroxyvitamin D(3)*. J Biol Chem, 1999. **274**(39): p. 27786-92.
79. Wang, J., et al., *The lysine demethylase LSD1 (KDM1) is required for maintenance of global DNA methylation*. Nat Genet, 2009. **41**(1): p. 125-9.
80. Ciccone, D.N., et al., *KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints*. Nature, 2009. **461**(7262): p. 415-8.
81. Saleque, S., et al., *Epigenetic regulation of hematopoietic differentiation by Gfi-1 and Gfi-1b is mediated by the cofactors CoREST and LSD1*. Mol Cell, 2007. **27**(4): p. 562-72.
82. Allis, C.D., et al., *New nomenclature for chromatin-modifying enzymes*. Cell, 2007. **131**(4): p. 633-6.
83. Tsukada, Y., et al., *Histone demethylation by a family of JmjC domain-containing proteins*. Nature, 2006. **439**(7078): p. 811-6.
84. Takeuchi, T., et al., *Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation*. Genes Dev, 1995. **9**(10): p. 1211-22.
85. Pedersen, M.T. and K. Helin, *Histone demethylases in development and disease*. Trends Cell Biol, 2010. **20**(11): p. 662-71.
86. Gu, B. and M. Lee, *Histone H3 lysine 4 methyltransferases and demethylases in self-renewal and differentiation of stem cells*. Cell Biosci, 2013. **3**(1): p. 39.
87. Lopez-Bigas, N., et al., *Genome-wide analysis of the H3K4 histone demethylase RBP2 reveals a transcriptional program controlling differentiation*. Mol Cell, 2008. **31**(4): p. 520-30.
88. Christensen, J., et al., *RBP2 belongs to a family of demethylases, specific for tri- and dimethylated lysine 4 on histone 3*. Cell, 2007. **128**(6): p. 1063-76.
89. Schmitz, S.U., et al., *Jarid1b targets genes regulating development and is involved in neural differentiation*. EMBO J, 2011. **30**(22): p. 4586-600.
90. Cellot, S., et al., *RNAi screen identifies Jarid1b as a major regulator of mouse HSC activity*. Blood, 2013. **122**(9): p. 1545-55.
91. Labbe, R.M., A. Holowatyj, and Z.Q. Yang, *Histone lysine demethylase (KDM) subfamily 4: structures, functions and therapeutic potential*. Am J Transl Res, 2013. **6**(1): p. 1-15.
92. Berry, W.L. and R. Janknecht, *KDM4/JMJD2 histone demethylases: epigenetic regulators in cancer cells*. Cancer Res, 2013. **73**(10): p. 2936-42.
93. Horton, J.R., et al., *Enzymatic and structural insights for substrate specificity of a family of jumonji histone lysine demethylases*. Nat Struct Mol Biol, 2010. **17**(1): p. 38-43.
94. Osawa, T., et al., *Increased expression of histone demethylase JHDM1D under nutrient starvation suppresses tumor growth via down-regulating angiogenesis*. Proc Natl Acad Sci U S A, 2011. **108**(51): p. 20725-9.
95. Cuthbert, G.L., et al., *Histone deimination antagonizes arginine methylation*. Cell, 2004. **118**(5): p. 545-53.

96. Wang, S. and Y. Wang, *Peptidylarginine deiminases in citrullination, gene regulation, health and pathogenesis*. Biochim Biophys Acta, 2013. **1829**(10): p. 1126-35.
97. Liu, Y.L., et al., *Functional role of dimerization of human peptidylarginine deiminase 4 (PAD4)*. PLoS One, 2011. **6**(6): p. e21314.
98. Nakashima, K., et al., *PAD4 regulates proliferation of multipotent haematopoietic cells by controlling c-myc expression*. Nat Commun, 2013. **4**: p. 1836.
99. Zhang, X., et al., *Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor alpha target gene activation*. Proc Natl Acad Sci U S A, 2012. **109**(33): p. 13331-6.
100. Hunger, S.P., et al., *Children's Oncology Group's 2013 blueprint for research: acute lymphoblastic leukemia*. Pediatr Blood Cancer, 2013. **60**(6): p. 957-63.
101. Gamis, A.S., et al., *Children's Oncology Group's 2013 blueprint for research: acute myeloid leukemia*. Pediatr Blood Cancer, 2013. **60**(6): p. 964-71.
102. Swerdlow, S.H., Campo, E., Harris, N.L., Jaffe, S.J., Pileri, A.S., Stein, H., Jürge n, T., and Vardiman, J.W. , *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. Lyon: IARC press, 2008: p. 88-147.
103. Niemeyer, C.M. and C.P. Kratz, *Paediatric myelodysplastic syndromes and juvenile myelomonocytic leukaemia: molecular classification and treatment options*. Br J Haematol, 2008. **140**(6): p. 610-24.
104. Stieglitz, E. and M.L. Loh, *Genetic predispositions to childhood leukemia*. Ther Adv Hematol, 2013. **4**(4): p. 270-90.
105. Song, W.J., et al., *Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia*. Nat Genet, 1999. **23**(2): p. 166-75.
106. Smith, M.L., et al., *Mutation of CEBPA in familial acute myeloid leukemia*. N Engl J Med, 2004. **351**(23): p. 2403-7.
107. Kirwan, M., et al., *Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukemia*. Hum Mutat, 2009. **30**(11): p. 1567-73.
108. Hahn, C.N., et al., *Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia*. Nat Genet, 2011. **43**(10): p. 1012-7.
109. Ostergaard, P., et al., *Mutations in GATA2 cause primary lymphedema associated with a predisposition to acute myeloid leukemia (Emberger syndrome)*. Nat Genet, 2011. **43**(10): p. 929-31.
110. Pasquet, M., et al., *High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia*. Blood, 2013. **121**(5): p. 822-9.
111. Lapidot, T., et al., *A cell initiating human acute myeloid leukaemia after transplantation into SCID mice*. Nature, 1994. **367**(6464): p. 645-8.
112. Dick, J.E., *Complexity of the human acute myeloid leukemia stem cell compartment: implications for therapy*. Biol Blood Marrow Transplant, 2005. **11**(2 Suppl 2): p. 9-11.
113. Passegue, E., *Hematopoietic stem cells, leukemic stem cells and chronic myelogenous leukemia*. Cell Cycle, 2005. **4**(2): p. 266-8.

114. Abramovich, C. and R.K. Humphries, *Hox regulation of normal and leukemic hematopoietic stem cells*. *Curr Opin Hematol*, 2005. **12**(3): p. 210-6.
115. Krause, D.S. and R.A. Van Etten, *Right on target: eradicating leukemic stem cells*. *Trends Mol Med*, 2007. **13**(11): p. 470-81.
116. Huntly, B.J. and D.G. Gilliland, *Leukaemia stem cells and the evolution of cancer-stem-cell research*. *Nat Rev Cancer*, 2005. **5**(4): p. 311-21.
117. Kvinlaug, B.T., et al., *Common and overlapping oncogenic pathways contribute to the evolution of acute myeloid leukemias*. *Cancer Res*, 2011. **71**(12): p. 4117-29.
118. Majeti, R., et al., *Dysregulated gene expression networks in human acute myelogenous leukemia stem cells*. *Proc Natl Acad Sci U S A*, 2009. **106**(9): p. 3396-401.
119. Schlegelberger, B., et al., *Update on cytogenetic and molecular changes in myelodysplastic syndromes*. *Leuk Lymphoma*, 2012. **53**(4): p. 525-36.
120. Holme, H., et al., *Marked genetic heterogeneity in familial myelodysplasia/acute myeloid leukaemia*. *Br J Haematol*, 2012. **158**(2): p. 242-8.
121. Krivtsov, A.V. and S.A. Armstrong, *MLL translocations, histone modifications and leukaemia stem-cell development*. *Nat Rev Cancer*, 2007. **7**(11): p. 823-33.
122. Balgobind, B.V., et al., *The heterogeneity of pediatric MLL-rearranged acute myeloid leukemia*. *Leukemia*, 2011. **25**(8): p. 1239-48.
123. Biondi, A., et al., *Biological and therapeutic aspects of infant leukemia*. *Blood*, 2000. **96**(1): p. 24-33.
124. Joannides, M. and D. Grimwade, *Molecular biology of therapy-related leukaemias*. *Clin Transl Oncol*, 2010. **12**(1): p. 8-14.
125. Milne, T.A., et al., *Leukemogenic MLL fusion proteins bind across a broad region of the Hox a9 locus, promoting transcription and multiple histone modifications*. *Cancer Res*, 2005. **65**(24): p. 11367-74.
126. Nguyen, A.T. and Y. Zhang, *The diverse functions of Dot1 and H3K79 methylation*. *Genes Dev*, 2011. **25**(13): p. 1345-58.
127. Somervaille, T.C. and M.L. Cleary, *Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia*. *Cancer Cell*, 2006. **10**(4): p. 257-68.
128. Horton, S.J. and B.J. Huntly, *Recent advances in acute myeloid leukemia stem cell biology*. *Haematologica*, 2012. **97**(7): p. 966-74.
129. Dorrance, A.M., et al., *Mll partial tandem duplication induces aberrant Hox expression in vivo via specific epigenetic alterations*. *J Clin Invest*, 2006. **116**(10): p. 2707-16.
130. Martens, J.H. and H.G. Stunnenberg, *The molecular signature of oncofusion proteins in acute myeloid leukemia*. *FEBS Lett*, 2010. **584**(12): p. 2662-9.
131. Gruber, T.A., et al., *An Inv(16)(p13.3q24.3)-encoded CBFA2T3-GLIS2 fusion protein defines an aggressive subtype of pediatric acute megakaryoblastic leukemia*. *Cancer Cell*, 2012. **22**(5): p. 683-97.
132. Thiollier, C., et al., *Characterization of novel genomic alterations and therapeutic approaches using acute megakaryoblastic leukemia xenograft models*. *J Exp Med*, 2012. **209**(11): p. 2017-31.

133. Fathi, A.T. and O. Abdel-Wahab, *Mutations in epigenetic modifiers in myeloid malignancies and the prospect of novel epigenetic-targeted therapy*. *Adv Hematol*, 2012. **2012**: p. 469592.
134. Muntean, A.G. and J.L. Hess, *Epigenetic Dysregulation in Cancer*. *The American Journal of Pathology*, 2009. **175**(4): p. 1353-1361.
135. Figueroa, M.E., et al., *DNA methylation signatures identify biologically distinct subtypes in acute myeloid leukemia*. *Cancer Cell*, 2010. **17**(1): p. 13-27.
136. Nikoloski, G., et al., *Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes*. *Nat Genet*, 2010. **42**(8): p. 665-7.
137. Bateman, C.M., et al., *Acquisition of genome-wide copy number alterations in monozygotic twins with acute lymphoblastic leukemia*. *Blood*, 2010. **115**(17): p. 3553-8.
138. Ma, Y., et al., *Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia*. *Proc Natl Acad Sci U S A*, 2013. **110**(18): p. 7429-33.
139. Greaves, M.F. and J. Wiemels, *Origins of chromosome translocations in childhood leukaemia*. *Nat Rev Cancer*, 2003. **3**(9): p. 639-49.
140. Smith, S.M., et al., *Clinical-cytogenetic associations in 306 patients with therapy-related myelodysplasia and myeloid leukemia: the University of Chicago series*. *Blood*, 2003. **102**(1): p. 43-52.
141. Super, H.J., et al., *Rearrangements of the MLL gene in therapy-related acute myeloid leukemia in patients previously treated with agents targeting DNA-topoisomerase II*. *Blood*, 1993. **82**(12): p. 3705-11.
142. Rodriguez-Paredes, M. and M. Esteller, *Cancer epigenetics reaches mainstream oncology*. *Nat Med*, 2011: p. 330-339.
143. Slack, J.L., C.P. Causey, and P.R. Thompson, *Protein arginine deiminase 4: a target for an epigenetic cancer therapy*. *Cell Mol Life Sci*, 2011. **68**(4): p. 709-20.
144. Morceau, F., et al., *Long and short non-coding RNAs as regulators of hematopoietic differentiation*. *Int J Mol Sci*, 2013. **14**(7): p. 14744-70.
145. Huggett, J., et al., *Real-time RT-PCR normalisation; strategies and considerations*. *Genes Immun*, 2005. **6**(4): p. 279-284.
146. Martin, J.A. and Z. Wang, *Next-generation transcriptome assembly*. *Nat Rev Genet*, 2011. **12**(10): p. 671-82.
147. Costa, V., et al., *Uncovering the complexity of transcriptomes with RNA-Seq*. *J Biomed Biotechnol*, 2010. **2010**: p. 853916.
148. Ozsolak, F. and P.M. Milos, *RNA sequencing: advances, challenges and opportunities*. *Nat Rev Genet*, 2011. **12**(2): p. 87-98.
149. Mane, S.P., et al., *Transcriptome sequencing of the Microarray Quality Control (MAQC) RNA reference samples using next generation sequencing*. *BMC Genomics*, 2009. **10**: p. 264.
150. Brown, J.A. and V.A. Boussiotis, *Umbilical cord blood transplantation: Basic biology and clinical challenges to immune reconstitution*. *Clinical Immunology*, 2008. **127**(3): p. 286-297.
151. Ng, Y.Y., et al., *Gene-expression profiling of CD34+ cells from various hematopoietic stem-cell sources reveals functional differences in stem-cell activity*. *J Leukoc Biol*, 2004. **75**(2): p. 314-23.

152. Prat-Arrojo, I., *Influence on Time to Cryopreservation in Umbilical Cord Blood Units*. International Journal of Clinical Medicine, 2011. **02**(04): p. 399-403.
153. Simon, C., et al., *A key role for EZH2 and associated genes in mouse and human adult T-cell acute leukemia*. Genes Dev, 2012. **26**(7): p. 651-6.
154. Livak, K.J. and T.D. Schmittgen, *Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method*. Methods, 2001. **25**(4): p. 402-8.
155. Wang, H., et al., *Dynamic transcriptomes of human myeloid leukemia cells*. Genomics, 2013. **102**(4): p. 250-6.
156. *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia*. N Engl J Med, 2013. **368**(22): p. 2059-74.
157. Spinner, M.A., et al., *GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics and immunity*. Blood, 2013.
158. Macrae, T., et al., *RNA-Seq Reveals Spliceosome and Proteasome Genes as Most Consistent Transcripts in Human Cancer Cells*. PLoS One, 2013. **8**(9): p. e72884.
159. Wu, C., et al., *BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources*. Genome Biol, 2009. **10**(11): p. R130.
160. de Jonge, H.J., et al., *Gene expression profiling in the leukemic stem cell-enriched CD34+ fraction identifies target genes that predict prognosis in normal karyotype AML*. Leukemia, 2011. **25**(12): p. 1825-33.
161. Harris, W.J., et al., *The histone demethylase KDM1A sustains the oncogenic potential of MLL-AF9 leukemia stem cells*. Cancer Cell, 2012. **21**(4): p. 473-87.
162. Mallette, F.A., et al., *RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites*. EMBO J, 2012. **31**(8): p. 1865-78.
163. Kim, T.D., et al., *Regulation of tumor suppressor p53 and HCT116 cell physiology by histone demethylase JMJD2D/KDM4D*. PLoS One, 2012. **7**(4): p. e34618.
164. Mallette, F.A. and S. Richard, *JMJD2A promotes cellular transformation by blocking cellular senescence through transcriptional repression of the tumor suppressor CHD5*. Cell Rep, 2012. **2**(5): p. 1233-43.
165. Lu, C., et al., *IDH mutation impairs histone demethylation and results in a block to cell differentiation*. Nature, 2012. **483**(7390): p. 474-8.
166. Li, J., *Quiescence regulators for hematopoietic stem cell*. Exp Hematol, 2011. **39**(5): p. 511-20.
167. Liu, J., et al., *A functional role for the histone demethylase UTX in normal and malignant hematopoietic cells*. Exp Hematol, 2012. **40**(6): p. 487-98 e3.
168. Wong, N.C., et al., *A distinct DNA methylation signature defines pediatric pre-B cell acute lymphoblastic leukemia*. Epigenetics, 2012. **7**(6): p. 535-41.
169. Ballabio, E. and T.A. Milne, *Molecular and Epigenetic Mechanisms of MLL in Human Leukemogenesis*. Cancers (Basel), 2012. **4**(3): p. 904-44.
170. Tanaka, S., et al., *Ezh2 augments leukemogenicity by reinforcing differentiation blockage in acute myeloid leukemia*. Blood, 2012. **120**(5): p. 1107-17.
171. Lewis, P.W., et al., *Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma*. Science, 2013. **340**(6134): p. 857-61.
172. Huether, R., et al., *The landscape of somatic mutations in epigenetic regulators across 1,000 paediatric cancer genomes*. Nat Commun, 2014. **5**: p. 3630.

173. Muller-Tidow, C., et al., *Profiling of histone H3 lysine 9 trimethylation levels predicts transcription factor activity and survival in acute myeloid leukemia*. Blood, 2010. **116**(18): p. 3564-71.
174. Paul, T.A., et al., *Signatures of polycomb repression and reduced H3K4 trimethylation are associated with p15INK4b DNA methylation in AML*. Blood, 2010. **115**(15): p. 3098-108.
175. Metzker, M.L., *Sequencing technologies - the next generation*. Nat Rev Genet, 2010. **11**(1): p. 31-46.
176. Bagger, F.O., et al., *HemaExplorer: a database of mRNA expression profiles in normal and malignant haematopoiesis*. Nucleic Acids Res, 2013. **41**(Database issue): p. D1034-9.
177. Andersson, A., et al., *Gene expression signatures in childhood acute leukemias are largely unique and distinct from those of normal tissues and other malignancies*. BMC Med Genomics, 2010. **3**: p. 6.
178. Ebert, B.L. and T.R. Golub, *Genomic approaches to hematologic malignancies*. Blood, 2004. **104**(4): p. 923-32.
179. Vinh, D.C., et al., *Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia*. Blood, 2010. **115**(8): p. 1519-29.
180. Mace, E.M., et al., *Mutations in GATA2 cause human NK cell deficiency with specific loss of the CD56(bright) subset*. Blood, 2013. **121**(14): p. 2669-77.
181. Kazenwadel, J., et al., *Loss-of-function germline GATA2 mutations in patients with MDS/AML or MonoMAC syndrome and primary lymphedema reveal a key role for GATA2 in the lymphatic vasculature*. Blood, 2012. **119**(5): p. 1283-91.
182. Hyde, R.K. and P.P. Liu, *GATA2 mutations lead to MDS and AML*. Nat Genet, 2011. **43**(10): p. 926-7.
183. Hsu, A.P., et al., *Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome*. Blood, 2011. **118**(10): p. 2653-5.
184. Hsu, A.P., et al., *GATA2 haploinsufficiency caused by mutations in a conserved intronic element leads to MonoMAC syndrome*. Blood, 2013. **121**(19): p. 3830-7, S1-7.
185. Dickinson, R.E., et al., *Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency*. Blood, 2011. **118**(10): p. 2656-8.
186. Johnson, K.D., et al., *Cis-element mutated in GATA2-dependent immunodeficiency governs hematopoiesis and vascular integrity*. J Clin Invest, 2012. **122**(10): p. 3692-704.
187. Zhang, S.J., et al., *Gain-of-function mutation of GATA-2 in acute myeloid transformation of chronic myeloid leukemia*. Proc Natl Acad Sci U S A, 2008. **105**(6): p. 2076-81.
188. Vicente, C., et al., *Overexpression of GATA2 predicts an adverse prognosis for patients with acute myeloid leukemia and it is associated with distinct molecular abnormalities*. Leukemia, 2012. **26**(3): p. 550-4.
189. Tipping, A.J., et al., *High GATA-2 expression inhibits human hematopoietic stem and progenitor cell function by effects on cell cycle*. Blood, 2009. **113**(12): p. 2661-72.

190. Bonadies, N., et al., *Genome-wide analysis of transcriptional reprogramming in mouse models of acute myeloid leukaemia*. PLoS One, 2011. **6**(1): p. e16330.
191. Green, C.L., et al., *GATA2 mutations in sporadic and familial acute myeloid leukaemia patients with CEBPA mutations*. Br J Haematol, 2013. **161**(5): p. 701-5.
192. Bodor, C., et al., *Germ-line GATA2 p.THR354MET mutation in familial myelodysplastic syndrome with acquired monosomy 7 and ASXL1 mutation demonstrating rapid onset and poor survival*. Haematologica, 2012. **97**(6): p. 890-4.
193. Orkin, S.H., *Diversification of haematopoietic stem cells to specific lineages*. Nat Rev Genet, 2000. **1**(1): p. 57-64.
194. Bigley, V. and M. Collin, *Dendritic cell, monocyte, B and NK lymphoid deficiency defines the lost lineages of a new GATA-2 dependent myelodysplastic syndrome*. Haematologica, 2011. **96**(8): p. 1081-3.